The role of Dichaete in transcriptional regulation during *Drosophila* embryonic development

Jelena Aleksic

University of Cambridge
Department of Genetics
Darwin College

December 2011

This dissertation is submitted for
the degree of Doctor of Philosophy

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except where specifically indicated in the text.

This dissertation does not exceed the regulation length of 60 000 words.

Acknowledgements

My PhD has been a long and interesting journey, and one I feel immensely fortunate to have experienced. Even though my contribution is just a little grain of sand in the grand universe of science, to spend a part of your life dreaming about life itself and its miracles, is a joy and a privilege. During this time, I received a huge amount of help and support from many different people, some of who I'd like to specifically thank here.

First and foremost, I would like to thank my supervisor Steve for giving me the chance to be his apprentice, sharing his love of science, always being there when I needed him, for many interesting discussions and for making me think and try harder. Boss, it's been a pleasure and an honour. I'd also like to thank him for sharing his extensive music collection (which helped ease a number of late nights in the lab), and introducing me to such cultural delights as drinking in the oldest still running brewery in the world and going to a football game for the very first time.

There are two of my colleagues in particular that this PhD would not have been possible without. I'd like to thank Bettina for huge amounts of help with both experiments and data analysis, and also for her endless positivity and wonderful company. I'd also like to thank Boris for teaching me to program, and introducing me to the weird and wonderful world of bioinformatics - a huge part of this PhD owes a debt to everything he taught me.

At various times, different people have helped me survive the traumas of labwork - so thank you to Nich, Shih-Pei and Lisa for taking the time to teach me, Richard and Sang for stepping in to help me with cloning horror, and David M and Eddie for various helpful advice. I would also like to thank John Roote for all his help with fly work.

In the somewhat happier world of bioinformatics and general biology, I would like to thank David H for writing software for comparing overlaps, Dean for many interesting biology discussions over beer, Audrey for her help with statistics as well as her many interesting questions and Robert for advice and awesome company. I'd also like to thank Enrico for many science discussions - your enthusiasm is contagious, and I feel

that our biology chats have enriched my PhD experience.

A special thank you to everyone who has trawled through my PhD thesis proofreading and editing it. In particular, I would like to thank Steve for a huge amount of help during the writing process, and my mum, who impressively managed to read through and correct the whole thing multiple times. I'd also like to thank Cas - as well as proofreading, she will forever have a special place in my heart for last minute help with reference formatting.

Also a huge thank you to my friends, old and new, for keeping me sane, supporting me when I needed it, and generally making my life here happier and more awesome. Katie and Cas for being wonderful friends, and for listening to me and always being on my side - I couldn't have done this without you. Aran for bike rides and endless cups of tea, and Ludi for her support and enthusiasm. Katharine for sharing experiences of academia and listening to my rants. Duncan for discussions and good advice. Chris for still being there for me. Ouie for knowing when it's a good time to kidnap me from work. Dave for photography, science and life chats. And Frankie for all the hours spent running our baby company while I was away trying to graduate.

I would also like to thank my partner Geraint for exciting discussions, awesome maths lessons, and all his love and support during the final stages of my PhD. Thank you for being by my side, and making the difficult times happier.

Last but not least, I would like to dedicate this work to my family, who this PhD would not have been possible without. I appreciate the huge amounts of support I got from my parents over the last 4 years, but in reality, their contribution extends so much further. I would like to thank them for instilling in me a deep love of learning, teaching me to dream without boundaries, and giving me the confidence to pursue those dreams. I would also like to thank them for their courage in leaving old worlds behind, and moving countries and continents to open up new worlds for us. Thank you for inspiring, encouraging and supporting me - this work is for you.

Abstract

Sox domain genes encode a family of developmentally important transcription factors conserved throughout the Metazoa. The subgroup B, which includes the mammalian Sox1, 2 and 3 proteins and their *Drosophila* counterparts Dichaete and SoxNeuro, are particularly important for the development of the nervous system where they appear to play conserved roles in neuronal specification and differentiation. Despite years of detailed study we still have a relatively poor idea of how Sox proteins function on a genome wide scale and the aim of my PhD work was to explore this aspect using the fly group B protein, Dichaete. A number of studies have shown that Dichaete performs a variety of critical functions during development and a few individual regulatory targets have been defined, however, at the start of my work no genome-wide data on Dichaete action were available. While such data emerged from large scale initiatives during my work, a systematic analysis of Dichaete action was lacking. Here I describe the first detailed genomic analysis of Dichaete activity, with a particular focus on three areas: finding the locations of Dichaete binding in the genome, a prediction of potential Dichaete cofactors and an analysis of Dichaete effects on gene expression.

To address the issue of where Dichaete binds in the genome, I generated whole genome DamID data for embryos and followed this with a detailed comparative analysis, combining my data with three newly published ChIP-chip datasets. The combined studies identify thousands of binding regions, mostly in the vicinity of developmentally important genes. The binding profiles were found to be consistent with Dichaete acting on enhancer regions and also suggest a role in facilitating RNA Polymerase II pausing. The analysis also identified a Dichaete binding motif closely matching that found with in vitro studies. By combined ChIP and DamID datasets I generated a very high confidence core Dichaete binding dataset, which should be of considerable use in future studies.

To identify potential Dichaete cofactors, I compiled the available embryonic transcription factor binding data from the Berkeley *Drosophila* Transcription Network and modENCODE projects, and identified significant overlaps with the core Dichaete binding data. A number of the proteins highlighted in this analysis have known roles during neuroblast development, including Hunchback and Krüppel, transcription factors in-

volved in temporal specification of neuroblast division, and Prospero, which plays a key role in neuroblast differentiation. The analysis suggests that Dichaete has a role during early neuroblast divisions, where it likely interacts with Hb and Kr to maintain neuroblast pluripotency. This is a role consistent with previous studies in *Drosophila* larval neuroblasts and is analogous to neural functions of Sox2 in mammals. My analysis suggests that Dichaete acts on the same target genes as Prospero but in an antagonistic role, with Dichaete preventing stem cell differentiation and Prospero promoting it.

To examine the effects of Dichaete on gene expression, a number of microarray transcript profiling studies were performed, including a global study with *Dichaete* null mutants, and tissue specific studies in the CNS midline and neuroblasts via the use of dominant negative constructs. Whole transcriptome expression profiling data was combined with the binding data to establish a set of high confidence potential Dichaete targets, both for specific tissues and more globally during neurogenesis. Specific high confidence targets were found, including *bancal* during nervous system development. It was also concluded that Dichaete is likely to prevent cell cycle exit by repressing the apoptosis genes *grim*, *hid* and *reaper*, as well as the differentiation genes *prospero* and *miranda*. An extensive list of potential Dichaete direct targets was generated and can be used for validation and future research.

Contents

Abbreviations

Table 1: Standard abbreviations and acronyms used throughout this thesis.

Abbreviation	Meaning
AL buffer	Qiagen lysis buffer
AP axis	Anterior-posterior axis
AW1	Qiagen wash solution 1
AW2	Qiagen wash solution 2
BDTNP	Berkeley *Drosophila* Transcription Network Project
cDNA	Complementary DNA
ChIP	Chromatin immunoprecipitation
CNS	Central nervous system
DAB	diaminobenzidine
DEPC	Diethylpyrocarbonate
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
dsAdr	Double-stranded adaptor oligonucleotide
DTT	Dithiothreitol
DV axis	Dorsal-ventral axis
EDTA	Ethylene diamine tetraacetic acid
EGTA	Ethylene glycol tetraacetic acid
EMS	Ethyl methanesulfonate
ENCODE	Encyclopedia of DNA elements
FBgn	FlyBase gene name
gDNA	Genomic DNA
GFP	Green fluorescent protein
HMG	High mobility group
MGA	Anterior midline glia
MGM	Medial midline glia
MGP	Posterior midline glia
modENCODE	Encyclopedia of DNA elements in model organisms

Continued on next page

Table 1 – continued from previous page

Abbreviation	Meaning
NMR	Nuclear magnetic resonance
PBS	Phosphate buffered saline
PBT	Phosphate buffered saline with Triton X100 added
PCR	Polymerase chain reaction
PLG	Phase lock gel
RNA	Ribonucleic acid
rpm	Rotations per minute
SI - SV	Stage 1 to stage 5 of waves of delaminating neuroblasts
TE buffer	Tris and EDTA buffer
UAS	Upstream activation sequence
vsn	Variance stabilising normalization

Chapter 1

Introduction

From so simple a beginning endless forms most beautiful and most wonderful have been, and are being evolved. – Charles Darwin, 1859

150 years on from Darwin's Origin of Species, the evolution and development of organism form is a topic that continues to puzzle and fascinate modern biology. While the diversity of living forms continues to invoke a sense of wonder, another fascinating conclusion to arise from this field of biology is that, for a wide range of animals, the early development of form follows the same blueprints. Because of this powerful idea and its implications for biomedical science, extensive developmental biology research has focused on uncovering the genetic logic that underpins this general blueprint of living beings, as well as for the specific components that make a species unique from others.

This thesis focuses on a particular transcription factor, Dichaete, involved in a number of important developmental roles, including the establishment of the basic body plan and the development of the nervous system. Dichaete belongs to the family of Sox transcription factors, which are known to have developmental roles in a wide range of metazoan species. Studying Sox function in *Drosophila* benefits from a solid groundwork from previous research to draw on to make sense of the role of Dichaete in development, while at the same time evidence of functional conservation indicates that results emerging from these studies may well be relevant to a broad range of species including mammals. The aim of this introduction is to give an overview of the available literature on Dichaete and to specify the direction in which my project aims to take this research further.

As *Dichaete* is a Group B *Sox* gene, there will be a brief overview of the evolution and function of *Sox* genes in general, with a comparison of vertebrates and invertebrates, followed by a discussion of the evolution, function and conservation of Group B *Sox* genes specifically. There will then be an overview of the literature currently available

on Dichaete and the relationship between Dichaete and SoxNeuro, as well as a brief discussion of its evolutionary and functional conservation. As a prelude to this discussion I briefly review our current understanding of aspects of eukaryotic transcription factor biology relevant to my studies.

1.1 Transcription factor theory

1.1.1 How transcription factors find their targets

In their elegant work on the *E. coli* Lac operon, Lin and Riggs (1975) observed that the *lac* repressor molecule was often found in locations not specific to its function - they estimated that 98% or more of the repressor molecules present were bound to sites other than the *lac* operator. They also observed that there was a difference in the binding affinity for the specific and non-specific binding areas. This was later verified using X-ray crystallography (Lewis et al., 1996; Kalodimos et al., 2004; Von Hippel, 2004), which made it possible to compare the binding to specific sites and to non-specific DNA. It was found that specific binding occurs in the major groove of DNA, but that this is also supported by electrostatic interactions between the negatively charged phosphates of the DNA backbone and the basic amino acid residues surrounding the transcription factor binding domain. The affinity of the transcription factor for non-specific binding sites is lower because of fewer sequence-specific hydrogen bonds, but the affinity to the DNA backbone still facilitates protein-DNA interactions. In the case of the *lac* repressor, it is thought that the repressor molecule slides along DNA and is capable of fast conformational changes when it encounters its specific binding sites (Winter et al., 1981). It is also thought that such non-specific protein-DNA interactions are crucial to the transcriptional regulation system, playing roles in balancing the overall thermodynamics of the regulatory process, facilitating the sliding mechanism for target-finding and also perhaps playing a structural role by orienting the transcription factor properly with respect to the DNA (Von Hippel, 2004). This system is very sensitive to the total concentration of DNA present, and it has even been suggested that part of the reason for the presence of large regions of 'junk' DNA in eukaryotic genomes is to optimise the thermodynamics of transcriptional regulation (Lin and Riggs, 1975).

While eukaryotes have more complex gene regulation, these conclusions nonetheless may be applicable: it is probable that all transcription factors have a general affinity to DNA and are able to bind to virtually any piece of DNA available, but that some of them have a higher affinity to their preferred sequences. Transcription factor occupancy is therefore likely to be resolved through a combination of specific affinity preferences, the availability of open chromatin and competition with other transcription factors for

the same binding sites, i.e. through balancing the thermodynamic properties of all components involved. While there is an obvious mechanism for sequence-specificity of transcription factors that bind to the major groove, minor groove binding transcription factors appear to be more of a mystery, since they do not have the same physical access to the DNA bases. However, recent research suggests that base information is still available through the changing electrostatic properties of the backbone, which are sequence dependent (Rohs et al., 2009). The shape of the minor groove changes depending on the underlying sequence, ending up more narrow in certain sections, which leads to an increase in its negative electrostatic potential in that location. It is thought that minor groove binding proteins use amino acids with positively charged side chains to 'read' the electrostatic potential of the minor groove (Tullius, 2009). This is a process different from the sequence specificity of the major groove: while it is determined by DNA sequence, it is possible for a range of DNA sequences to generate the same physical minor groove shape, making minor groove binding potentially less specific.

Sequence specificity in general, while acknowledged to be an important part of transcriptional regulation, does not appear to be a sole or sufficient predictor of binding and functionality. While matching binding motifs are frequently found within the binding intervals of sequence-specific transcription factors, they are not particularly reliable predictors of either binding or function. The largest problem is the high presence of false positives, as the position weight matrices used for these searches often match a large number of sites in the genome and rely on somewhat arbitrary thresholds (Stormo, 2000). While it may be theoretically possible for a transcription factor of interest to bind to all the sites found as far as sequence goes, in practice a lot of additional information such as DNA accessibility and competition with other factors would need to be integrated into the analysis in order to make more accurate predictions. Another problem with this type of analysis is that most DNA affinity values are measured using a single protein, whereas proteins frequently pair up in vivo and it has been found that binding to a cofactor can completely alter the sequence specificity of a protein, likely due to conformational changes (Kondoh and Kamachi, 2010). Thus it is possible that some incorrect sequences are used for motif searches. It is also not clear that sequence specificity is the only determinant of function. There is evidence of enhancers whose function is conserved despite a lack of conservation of both motif sequences and enhancer element locations, due to redundancy, and new mutations arising to stabilise the gene expression pattern (Piano et al., 1999; Ludwig et al., 2000). We do, however, know that binding location appears to be fully specified by the chromosome structure and sequence, rather than the environment of the cell - an elegant study by Wilson et al. (2008) looked at transcription factor binding to human chromosome 21 in a mouse model and found that despite it being located in mouse cells, the expected human pat-

tern of binding still occurs. However, in combination, what this tells us is that, while transcriptional regulation mechanisms are specified in the sequence and structure of the chromosomes, the logic behind it is not as simple as always having specific sequences in particular locations in order to facilitate a particular gene expression outcome. Instead, the system can be quite flexible, contain redundancy, and combinations of quite different underlying sequences can facilitate the same gene expression outcomes. This limits the utility of any conclusions based on sequence content alone, and makes the problem of understanding the role of DNA sequence in transcriptional regulation quite a complex one.

1.1.2 Different models of transcriptional control

There are a number of models for how transcription factors act on their targets once they find them. The more traditional models think of promoter and enhancer regions as logic gates, with different combinations of transcriptional regulators triggering different outcomes. Perhaps the most famous example is the *lac* operon, with its dual logic gate mechanism involving an activator and a repressor (Jacob and Monod, 1961; Zubay et al., 1970). This model has been used widely to explain more general mechanisms of transcriptional regulation (Ptashne, 2005) and various studies have attempted to study the promoter components underlying the logical framework (Ligr et al., 2006; Cox et al., 2007). Work has also been done to use the formal logic framework to express and model the qualitative aspects of mechanisms of transcriptional regulation in particular genetic systems (Gonzalez et al., 2008). However, a flaw in this approach is in the underlying model of transcriptional regulation with rigid outcomes and discrete regulation modules. Regulatory networks are complex systems, where small quantitative differences in binding may confer different regulatory outcomes and stochasticity may also play a part. Furthermore, there is considerable redundancy between different transcription factors, and the slightly varying conditions and positional information in different cells can result in different regulatory outcomes, meaning the logic gate model fails to capture much of the inherent complexity of the system.

Nonetheless it is a simple and powerful model, that has been used to great effect to uncover general organisational properties of transcriptional regulation (Babu et al., 2004; Barabasi and Oltvai, 2004). Most of these studies focus on network topology (patterns of connections between network nodes), how this relates to function and how networks are reorganised under different cellular conditions. In particular, hierarchical structures have been discovered in both prokaryotic and eukaryotic networks (Ma et al., 2004; Farkas et al., 2006; Yu and Gerstein, 2006) and have been reported to increase the adaptability of the network and to avoid conflicting constraints. Transcriptional regulatory networks display a "scale-free" topology, with a small number of global hubs

being connected to a very large number of targets, and the majority of transcription factors being connected to a small number of targets (Babu et al., 2004; Chalancon and Babu, 2010). It has been suggested that this structure makes the network more robust to failure, with the majority of the potential points of failure not crucial to the overall structure of the network. There have also been attempts to use network data as a starting point for more detailed quantitative modelling (Zaslaver et al., 2006; Martinez-Antonio et al., 2008). A recent study looking at networks and transcription factor dynamics in yeast found that different hierarchical levels of the network have distinct dynamic properties (Jothi et al., 2009). This shows the connection between network structures and biological properties, as well as that network representations can be used as a starting point, then integrated with other datasets to better reflect quantitative properties.

Another emerging view takes a more quantitative approach, claiming most transcription factors bind to most genes in the genome and that the varying levels of transcription factor occupancy have differential outcomes for gene expression (Biggin, 2011). In this model, it is thought that low occupancy binding sites are mainly non-functional, with the higher occupancy sites having an increasing impact on transcription (MacArthur et al., 2009; Biggin, 2011). This view is primarily based on in vivo crosslinking assays, which show that in a range of eukaryotes there are frequently thousands of binding sites per transcription factor in the genome and that there is considerable overlap between unrelated transcription factor families (MacArthur et al., 2009; Gerstein et al., 2010; Myers et al., 2011; Negre et al., 2011). Computational modelling also shows some support for this view, finding that the presence of a high number of multiple and sometimes overlapping sites facilitates the integration of complex signals through competition both between and within individual cis-regulatory modules (Hermsen et al., 2006).

It appears to be the case that many more transcription factor molecules are present in the nucleus at any one time than are required for direct modulation of transcription. Most of these are bound to DNA the majority of the time (Biggin, 2011), suggesting that most genuine sites of binding in the genome are non-functional. However, it is unclear whether individual molecules and functional binding sites have a key causal role in transcription, which is closer to the logic gate model, or whether active regulation relies on large quantities of a particular regulator being localized in a particular region. Either model is plausible from our current data. Links between quantitative data and functionality have been identified in several studies, but it is possible that in order for a given transcription factor to outcompete others and accurately slot into its active binding site, there need to be many molecules of that transcription factor in the region, even though most of them do not interact with the functional binding site.

1.2 Sox family proteins

1.2.1 Classification and function

Sox proteins are a developmentally important family of transcription factors found in all metazoans examined to date (Schepers et al., 2002; Koopman et al., 2004; Phochanukul and Russell, 2009; Zhong et al., 2011). Their defining feature is that they all contain a High Mobility Group (HMG) binding domain and they are classified according to their similarity to Sry, the human sex determining factor. Specifically, if a protein has an HMG-1 type domain with strong amino acid sequence similarity (usually $> 50\%$) to Sry (Sinclair et al., 1990), it is classified as a Sox family protein (McKimmie et al., 2005). Due to their important roles in various aspects of development, they have been the subject of extensive study for a number of years (Wright et al., 1993; Kamachi et al., 1995; Wegner and Stolt, 2005; Masui et al., 2007; Takahashi and Yamanaka, 2006).

Functional studies of specific *Sox* genes in model organisms have provided insights that are likely generalisable to the *Sox* family in general, and are certainly interesting to consider in the context of studying Dichaete. Most Sox proteins appear to share similar DNA binding properties, and recognise a similar heptameric sequence: 5'-(A/T)(A/T)CAA(A/T)G-3' (Harley et al., 1994). However, Sox proteins in vivo appear to be highly specific, targeting a unique repertoire of genes (Wilson and Koopman, 2002). Upon DNA binding, they bend DNA at a sharp angle (Werner, Bianchi, Gronenborn and Clore, 1995). They generally do not act on their own to alter expression, instead partner up with specific cofactors to perform their regulatory functions, for example Pax or POU proteins (Wilson and Koopman, 2002; Dailey and Basilico, 2001; Chi and Epstein, 2002). It is possible that partnering allows them to take part in a large number of diverse developmental processes, since their action can be context specific and based on the available cofactors present (Kamachi et al., 2000).

The crystal structure of the POU/HMG ternary complex on the *FGF4* enhancer has been resolved by Remenyi et al. (2003) using domains from Oct1 and Sox2 (Figure 1.1). The Sox2 domain was shown to interact with DNA sequence-specifically, with the interaction being mediated by a large number of base-pair specific hydrogen bonds. NMR studies with the HMG-domain of mammalian SRY (Murphy et al., 2001; Werner, Huth, Gronenborn and Marius Clore, 1995) demonstrate the overall conservation in HMG domain structure and have also been used to define the DNA bending properties of the domain (Werner, Bianchi, Gronenborn and Clore, 1995).

In general, *Sox* genes across a range of metazoan species are classified into groups, A-J (Bowles et al., 2000), primarily based on sequence alignment: group A contains the

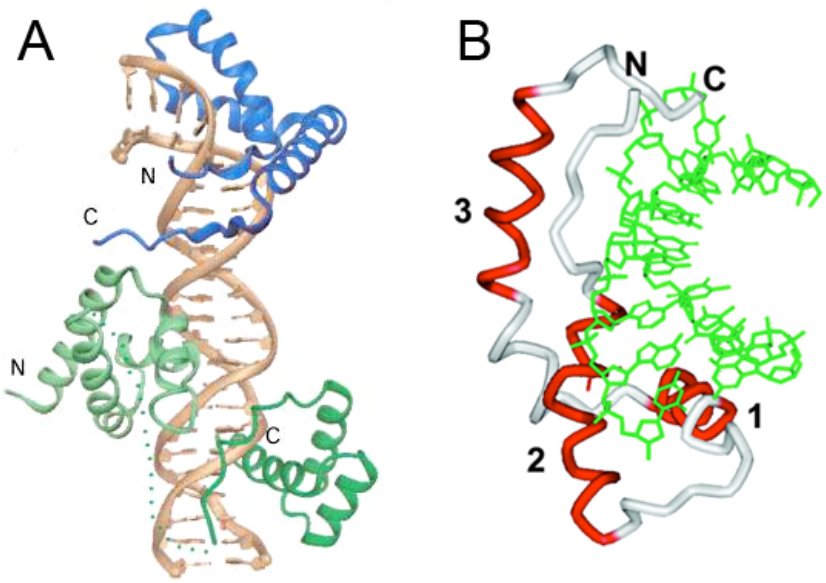

Figure 1.1: (A) Crystal structure of the Oct1/Sox2/*FGF4* enhancer ternary complex. The Sox2 HMG domain is shown in blue, and the Oct1 POU domain is shown in green. (B) NMR structure of the HMG domain of mammalian Sry bound to DNA. The model shows 3 helices in an L-shaped conformation and binding to the minor groove of DNA while bending and unwinding it (Werner, Bianchi, Gronenborn and Clore, 1995). Image from a review by Harley et al. (2003).

Sry gene used as the reference for defining the family. The first few studies used HMG domain sequences and found evidence for subdivision of the *Sox* family into groups A-H (Wright et al., 1993; Hiraoka et al., 1998; Osaki et al., 1999; Wegner, 1999), with a later study, including the full-length protein structure and the exon-intron structure of the genes, assigning groups I and J (Bowles et al., 2000). The classification is on the whole very robust for each group in higher animals: most groups were found to have characteristic amino acid signatures not present in other groups. Some of the *Sox* genes are more difficult to assign in groups such as Cnidarians (anemones, corals and jellyfish) and Ctenophorans (comb jellies), likely due to lineage specific divergence (Larroux et al., 2008). Group A is only found in eutherian mammals, groups B-F are found in all higher metazoans and groups G-J are restricted to specific lineages (Phochanukul and Russell, 2009). While the HMG domain is quite conserved between different families (by definition, Sox proteins have an HMG domain at least 50% identical to that of Sry), the sequences outside the HMG domain are quite variable. However, some common non-HMG box features can be identified among a number of them, with vertebrate orthologues in particular being very similar to each other both within and outwith the HMG domain (Bowles et al., 2000).

There is some variability in gene structure between the different *Sox* gene groups. 'HMG box introns' are introns found within the DNA sequence for the HMG domain, and the presence of this feature varies in different species. Because an intron is unlikely

to arise in the same position independently in multiple lineages, and intron gains and losses represent major genetic rearrangements, looking at intron/exon organisation can be informative to use as a measure of relatedness of homologous genes between different species (Bowles et al., 2000). No HMG box introns have been found in vertebrates for *Sox* groups A, B, C or G, or in sea urchin group B. Interestingly introns do exist in a number of *Drosophila* and *C. elegans* group B and C *Sox* genes, and there is conservation of intron position in the genes in question between the two species, while based on looking at vertebrates and sea urchin, introns appear to be missing entirely in group B *Sox* genes in the deuterostome lineage (Bowles et al., 2000). HMG box introns have been found in groups D, E, F, H and J, and are in each case conserved within the group, with the intron for groups D and F also being conserved between the two groups.

1.2.2 *Sox* genes in vertebrates

In humans and mice, there are a total of 20 *Sox* genes found in the genome, grouped into 8 subgroups (Schepers et al., 2002). At least 30 genes in total are recognised in mammals (Bowles et al., 2000; Schepers et al., 2002; Wegner, 1999). It was found that in vertebrates, the similarity of HMG domain sequences is reflective of the evolutionary relatedness of the proteins, estimated by looking at entire protein sequence (Bowles et al., 2000). Most of the *Sox* gene families found in vertebrates are represented by a single gene in invertebrates, suggesting that duplications and divergence have occured during vertebrate evolution (Koopman et al., 2004). This is consistent with 2 rounds of whole genome duplication, and not specific to the *Sox* gene family in particular.

1.2.3 *Sox* genes in invertebrates

While genetic studies often focus on a small number of model organisms, invertebrates are in fact a hugely diverse group, which includes the majority of multicellular animal species on Earth (Adoutte et al., 2000). Varying numbers of *Sox* genes have been found in a diverse range of invertebrate species (Figure 1.2). The evolution of the *Sox* gene family broadly follows the trends observed with other transcription factors in terms of family expansion and evolution (Larroux et al., 2008). The basal metazoans such as sponges contain 3-4 *Sox* genes, including a group B member (Phochanukul and Russell, 2009). There is a significant lineage specific expansion in the Radiata, with members of this group, such as the sea anemone, containing up to 14 *Sox* genes (Putnam et al., 2007). The *Sox* genes in protostomes are less numerous, with less than 10 genes found in all species examined to date. This is also the case for most bilateria in general, excluding the vertebrate lineage (Phochanukul and Russell, 2009).

Much of the study of *Sox* genes in invertebrates has focused on *Drosophila melanogaster*, where 8 *Sox* genes were found: 4 belonging to group B, and one each in groups C-F (Cremazy et al., 2001). The sequencing of 11 other Drosophila genomes indicates that this number and grouping of *Sox* genes is constant among Drosophilids (Clark et al., 2007). The same complement of *Sox* genes is also present in the mosquito *Anopheles gambiae*, while the honeybee *Apis mellifera* and parasitic wasp *Nasonia vitripennis* have an additional Group E gene (Phochanukul and Russell, 2009).

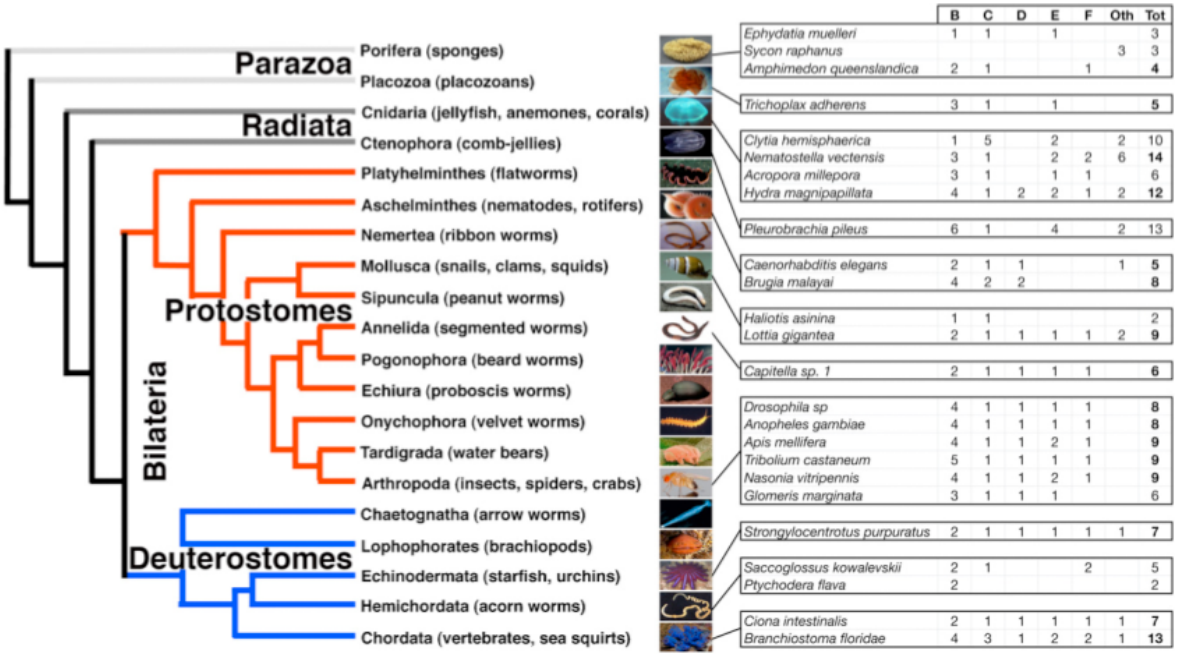

Figure 1.2: Phochanukul and Russell (2009) present an overview of the phylogeny of the metazoa, along with an overview of the identified *Sox* genes in a variety of species from the linked taxa.

1.2.4 Group B *Sox* genes

The B subgroup of *Sox* genes is the group most closely related to *Sry*, with some studies proposing that *Sry* evolved from *Sox3* (Collignon et al., 1996). They appear to have been conserved during evolution, with group B genes being found in all higher metazoans (Phochanukul and Russell, 2009). To date, a variety of studies have shown that the group B family participates in early central nervous system differentiation in a range of species, including Drosophila, Xenopus, chick and mouse (Sasai et al., 2001), as well as taking part in a number of other developmental processes, such as maintaining stem cell pluripotency (Masui et al., 2007), eye development (Ishii et al., 2009), and segmentation in insects (Russell et al., 1996), among others. Group B proteins are found to be co-expressed in specific parts of the nervous system, but their degree of cooperation and functional redundancy is still a subject of study (Uwanogho

et al., 1995; Rex et al., 1997; Overton et al., 2002).

Group B genes are further subdivided into groups B1 and B2. The two groups have a very similar HMG domain, as well as similarity in the short extension from the C-terminus of the HMG domain, but the rest of their sequences are distinct from each other (Uchikawa et al., 1999). The subdivision was suggested based on full protein sequence alignment, as well functional data. In the case of vertebrate *Sox* genes, there is a clear sequence-based and functional division: group B1 includes *Sox1*, *Sox2* and *Sox3* and these encode proteins that activate gene expression (Kamachi et al., 1995; Nishiguchi et al., 1998), whereas group B2 includes *Sox14* and *Sox21* and repress gene expression (Uchikawa et al., 1999). In *Drosophila*, the classification is less straightforward. There are 4 group B Sox genes in total, but for example, based on its effect on the expression of *ac* in proneural clusters (Overton et al., 2002), *Dichaete* appears to have the ability to act as a repressor or an activator context specifically, and there is some debate over its classification as a B1 or B2 gene (McKimmie et al., 2005; Zhong et al., 2011).

The role of group B *Sox* genes in embryonic stem cells

The vertebrate group B gene *Sox2* is essential for the maintenance of pluripotency in embryonic stem cells (Liu et al., 2008). Famously, it is one of the Yamanaka factors, Oct3/4, Sox2, Klf4 and c-Myc, the four transcription factors required for reversing differentiated cells into an induced stem cell state (Takahashi and Yamanaka, 2006). It is thought that the main role of Sox2 in ES cells is to regulate a number of transcription factors which in turn control the levels of expression of *Oct3/4* (Masui et al., 2007), thus stabilising the ES cells in a pluripotent state. It is also known that Oct3/4 and Sox2 act together on a number of enhancer regions (Boyer et al., 2005; Loh et al., 2006). However, it was found that Oct3/4 alone was sufficient for the activation of the enhancers in question, suggesting that the main role of Sox2 in maintaining ES cell pluripotency is to ensure the correct levels of Oct3/4 are present (Masui et al., 2007). Another reported role of group B1 genes is in the early stages of neural differentiation from pluripotent stem cells, where the presence of *Sox1*, *Sox2* or *Sox3* has an effect on directing stem cell fate into the neuroectodermal lineage (Avilion et al., 2003; Zhao et al., 2004; Stevanovic, 2003). Rather than interrupting proliferation, they merely convey neuroectodermal competence, which is only realised once the cell exits the cell cycle.

The role of group B *Sox* genes in nervous system development

The group B1 genes *Sox1*, *Sox2* and *Sox3* are required for stem cell maintenance in the central nervous system (CNS) and are counteracted by *Sox21*. A number of *Sox* genes from other families are also active in the same region, for example the group E genes *Sox9* and *Sox10*. *Sox9* changes the potential of stem cells from neurogenic to gliogenic, and *Sox10* is essential for terminal oligodendrocyte differentiation. Interestingly, in the peripheral nervous system, the same *Sox* genes are present, but have different functions (Wegner and Stolt, 2005).

Group B1 genes are widely expressed in the cells that are competent to form the neural primordium and are later restricted to the cells committed to the neural fate (Pevny and Placzek, 2005; Wegner and Stolt, 2005). Functional studies have been performed to determine their role during the development of the CNS in mammals. In the case of *Sox1* in mice, null mutants survive until adulthood, although the spontaneous seizure phenotype of the mice suggests some role in CNS development (Nishiguchi et al., 1998). However, since the mice are able to survive, this indicates that either *Sox1* only has a minor role, or there is some redundancy of function with the other two Group B *Sox* genes present. In contrast, *Sox2* null mutant mice die prior to implantation (Collignon et al., 1996), indicating its importance during early development.

Sox B1 genes are expressed in the presumptive vertebrate neuroectoderm. They do not interfere with the self-renewal of embryonic stem cells, but instead trigger a neuroectodermal fate after the cells have exited the cell cycle (Avilion et al., 2003; Stevanovic, 2003; Zhao et al., 2004; Wegner and Stolt, 2005). It has been reported that group B1 and group B2 genes act antagonistically (Uchikawa et al., 1999), with B1 genes maintaining pluripotency and B2 genes acting on the same gene targets, but repressing rather than activating them. *Sox21* in particular is widely coexpressed with group B1 genes and acts to counteract their activity and promote neural differentiation (Sandberg et al., 2005). Thus it has been suggested that the correct differentiation of neural cells is controlled by a precise balance between the presence of *Sox21* and of the B1 genes.

1.2.5 Group B *Sox* genes in Drosophila melanogaster

Of the eight *Sox* genes found in Drosophila (Cremazy et al., 2001), four belong to Group B: *Dichaete*, *SoxNeuro*, *Sox21a* and *Sox21b* (McKimmie et al., 2005). There is some disagreement about the subclassification into groups B1 and B2. *SoxNeuro* clearly shows sequence similarity to the B1 group, and *Sox21a* to the B2 group. However, with *Dichaete* and *Sox21b*, the classification is less straight-forward, because they have several unique amino acid mutations that differentiate them from the classic B1

and B2 groups, suggesting that they may have insect specific adaptations. Functionally *Dichaete* displays the behaviour of a B1 group *Sox* gene, so this is the classification suggested by McKimmie et al. (2005). They suggest that originally there was a single ancestral group B *Sox* gene, which duplicated to produce the ancestors of *Dichaete* and *SoxN*. Subsequently *Dichaete* then duplicated to make *Sox21a*, then acquired the insect specific amino acid mutations and underwent another duplication to form *Sox21b*. Zhong et al. (2011) disagree with this view and instead suggest that there were two ancestral group B genes, *SoxB1* and *SoxB2*, and that tandem duplications of the *SoxB2* gene occured after the arthropod-nematode divergence to create the remaining two SoxB genes. The two models of evolution are shown in Figure 1.3. The model from Zhong et al. (2011) fits better with the available sequence information, but the classification of *Dichaete* as a *SoxB2* gene seemingly goes contrary to the reported specific functional conservation between it and the vertebrate *Sox2* (Soriano and Russell, 1998).

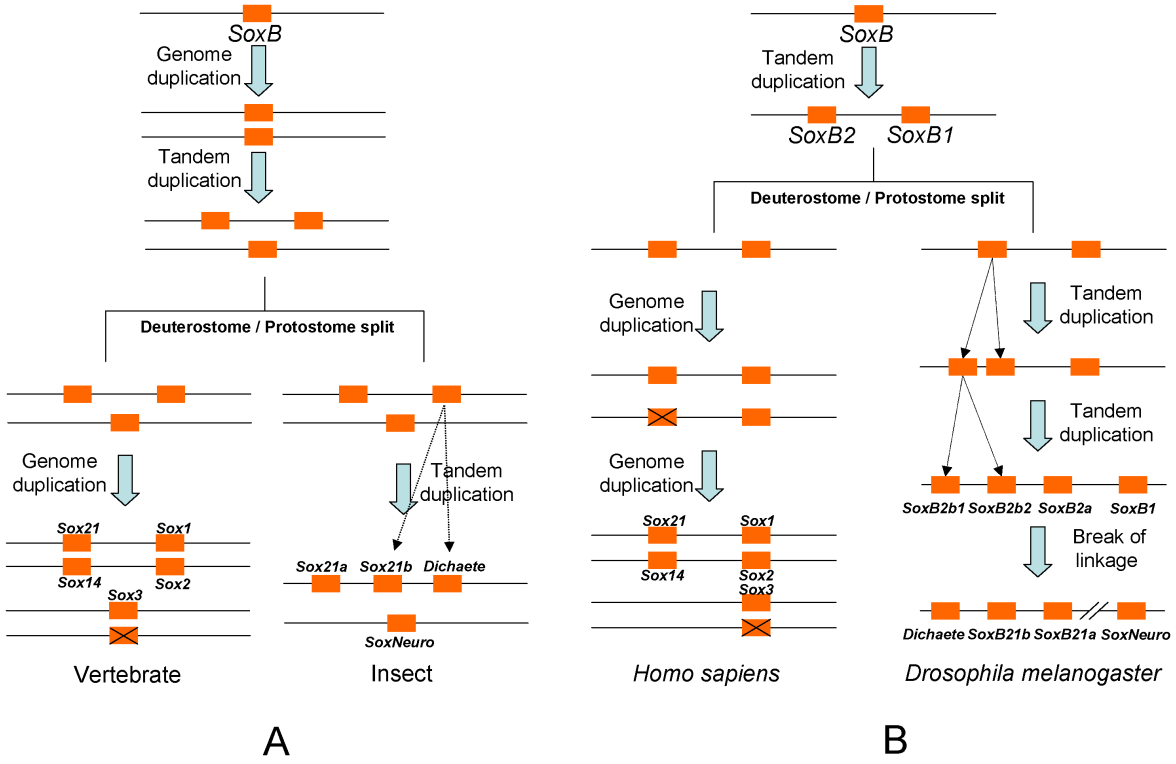

Figure 1.3: A figure from Zhong et al. (2011) showing the two competing models of *Sox* group B evolution. (A) The model proposed by McKimmie et al. (2005), in which *Dichaete* and *SoxN* are created by a duplication from a single ancestral group B *Sox* gene, and a tandem duplication created *Sox21a* before the Deuterostome/Protostome split. (B) The model proposed by Zhong et al. (2011), which argues for the existence of one *SoxB1* and one *SoxB2* gene, generated by an ancestral duplication before the Deuterostome/Protostome split.

SoxN is located on the second chromosome, while the other 3 genes are all located within the same 77 kb region of chromosome 3. *SoxN* and *Dichaete* have no introns,

while *Sox21a* and *Sox21b* have 1 and 6 introns, respectively (McKimmie et al., 2005). The structure of the *SoxB* genes in this region appears to be highly conserved across a range of insect species (Figure 1.4). *Dichaete* has a large 3' regulatory region, containing at least 30 kb of downstream regulatory sequence (Sanchez-Soriano and Russell, 2000). *SoxN* also sits alone, with no flanking genes for 35 kb proximal and 45 kb distal, a phenomenon unusual for *Drosophila*, so it is likely that it has a complex regulatory region similar to that of *Dichaete*. *Sox21a* and *Sox21b* are located 46 kb downstream of *Dichaete*, with no other genes between, and it has been suggested that the 3 genes may share at least some regulatory sequences (McKimmie et al., 2005), perhaps regulating their expression in the hindgut and the midline.

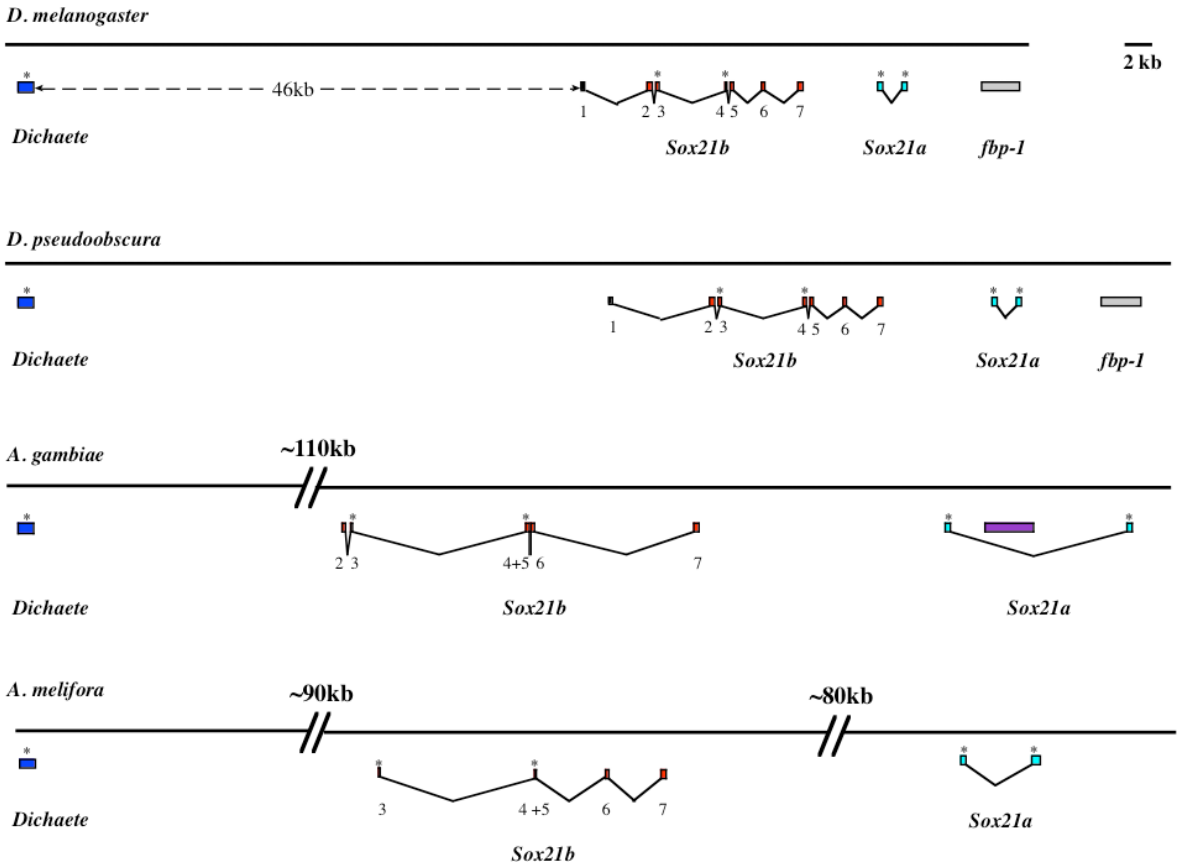

Figure 1.4: The organisation of Group B *Sox* genes on chromosome 3 in the fruit fly species *D. melanogaster* and *D. pseudoobscura*, the mosquito *A. gambiae* and the honey bee *A. melifora*. The overall genome organisation in this region is similar in all four species. Figure taken from McKimmie et al. (2005).

1.2.6 Advantages of using Drosophila as a model

As previously discussed, *Sox* genes represent an important gene family involved in a myriad of fundamental developmental processes. Group B *Sox* genes are particularly interesting due to their numerous roles as well as their involvement in the maintenance

of stem cell pluripotency and nervous system development. Functional studies suggest a remarkable degree of conservation between *SoxB* genes in *Drosophila* and vertebrates (Soriano and Russell, 1998; Blanco et al., 2005), indicating that studies in model organisms such as *Drosophila* will have broader relevance. There are general arguments for the use of *Drosophila* as a model organism, such as tractability, fast generation time and a wealth of available genetic tools. It is also particularly convenient for *Sox* genes because of the reported functional conservation and the fact that there are fewer *Sox* genes than in vertebrates. For example, things like functional redundancy are easier to resolve with model systems with two rather than three interacting genes. Furthermore, the small and dense genome makes high throughput genomics cheaper and the subsequent analysis more straight-forward. In general, studies of *Sox* genes in a large range of organisms will further our understanding of their function, evolution and variability, but these are clearly good reasons why *Drosophila* is particularly suitable for in-depth studies.

1.3 Dichaete overview

Dichaete is known to be involved in segmentation, CNS and brain development, hindgut morphogenesis and trichome development. The well known Dichaete wing phenotype (Figure 1.5), caused by ectopic expression in the wing imaginal discs, was the first dominant allele found in *Drosophila*, identifed almost 100 years ago, in 1915, by Calvin Bridges (Bridges and Morgan, 1923).

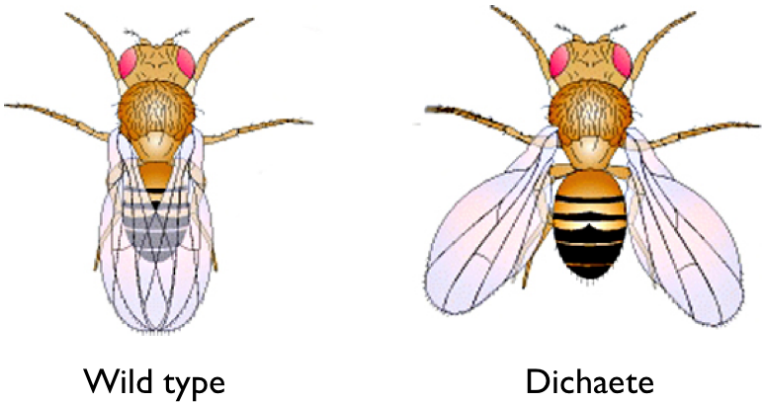

Figure 1.5: Dichaete dominant phenotype. Picture taken from Griffiths et al. (2002)

The gene responsible for the phenotype was found independently by 2 different groups in 1996. Russell et al. (1996) used *Sry* primers to look for related genes in *Drosophila* genomic and cDNA libraries, while Nambu and Nambu (1996) cloned it from a *P*-element enhancer trap insertion. An 1800 nt transcript with no introns was found, which encodes a 382 amino acid protein. Within this, a 76 amino acid stretch was found,

with 88% similarity to the DNA-binding HMG domain of Sox2 from human, mouse and chicken, suggesting possible evolutionary conservation of function. Outside of the DNA binding domain, no conservation was found, except for a 30 amino acid C-terminal stretch bearing resemblance to a potential Sox2 activation domain (Russell et al., 1996). A Dichaete protein with the C-terminal stretch removed can still rescue the *Dichaete* nervous system phenotype, but interestingly this is not the case when acting in the wing hinge, suggesting that the C-terminal region of Dichaete has a context dependent role in gene regulation, perhaps mediating specific co-factor interactions (Soriano and Russell, 1998). Another study suggests that the role of the C-terminal domain may not be in transcriptional activation: in yeast-based assays of the N-terminal region (141 amino acids), HMG domain (79 amino acids) and the C-terminal region (164 amino acids), only the N-terminal region was found to have transcriptional activation capabilities (Ma et al., 1998).

Figure 1.6: Dichaete binding motif found in a bacterial 1-hybrid experiment. The motif, published by Noyes et al. (2008), was obtained from the JASPAR database (Bryne et al., 2007).

As with other Sox family proteins, Dichaete binds to DNA in the minor groove, bending it at an 85° angle (Ma et al., 1998). As expected for a transcription factor, it localises to the nucleus in most tissues, but it has also been detected in the cytoplasm, for example in the neuroectoderm (Soriano and Russell, 1998) and it appears that the nuclear localisation signal is located in the HMG domain (Mukherjee et al., 2000). It is a sequence-specific transcription factor, which was shown to bind to the general Sox protein consensus binding sites AACAAT and AACAAAG (Ma et al., 1998). The result of a bacterial 1-hybrid experiment published by Noyes et al. (2008) was obtained from the JASPAR database (Bryne et al., 2007) with the characterised sequence motif shown in Figure 1.6.

1.4 Dichaete expression

1.4.1 Tissue-specific expression during embryonic development

Dichaete is highly expressed in a dynamic pattern during early embryogenesis, followed by expression at low levels during the rest of the life cycle (Russell et al., 1996). No maternal contribution of transcript or protein is detected even though the gene is

expressed in the ovary (Mukherjee et al., 2006; Mutsuddi et al., 2010). Expression is first initiated at late stage 4 of embryogenesis, starting with a broad central domain that is rapidly followed by the appearance of an anterior domain. The central domain then splits, and remains continuous dorsally, splitting into 7 stripes on the ventral side (Figure 1.7). By the end of stage 5, the dorsal domain and 6 out of 7 of the stripes fade, but the posterior stripe remains. During stage 6, the remaining stripe follows the pole cells, and eventually fades. At the same time, *Dichaete* is activated in the region of the neuroectoderm that will give rise to the CNS. Initially, 14 discrete stripes are observed, but the expression quickly becomes a continuum (Russell et al., 1996; Sanchez-Soriano and Russell, 2000).

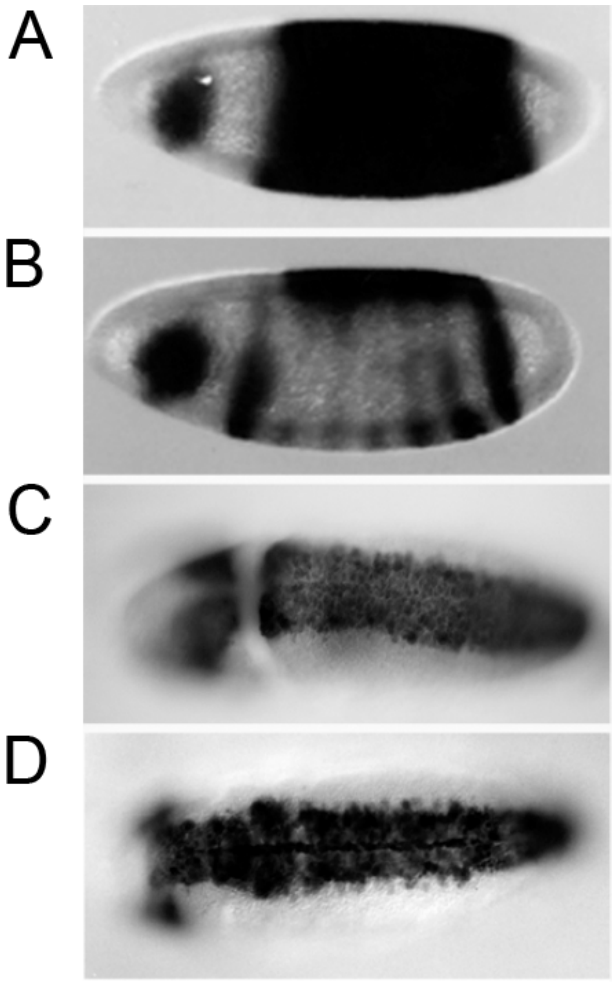

Figure 1.7: Anti-Dichaete staining shows the presence of Dichaete at different stages of embryonic development. (A) At late stage 4, a broad central domain of expression is established, rapidly followed by an anterior domain. (B) During stage 5, the central domain remains continuous dorsally, but splits into 7 stripes on the ventral side. In (C) and (D), Dichaete expression is prominent in the ventral nervous system, with waves of neuroblasts initiating Dichaete expression. Images from Sanchez-Soriano and Russell (2000).

Later in development, expression is prominent in the nervous system (Figure 1.7).

Expression is initially detected in the neuroectoderm at late stage 6 to early stage 7. Next, waves of neuroblasts in the medial and intermediate columns initiate *Dichaete* expression. *Dichaete* is also expressed in the ventral midline - it is detected weakly at stage 7, and by stage 9, it strongly labels the 2 rows of midline progenitors (Soriano and Russell, 1998). At stage 12, it is expressed strongly in the anterior and medial midline glia (MGA and MGM, respectively), and weakly in the posterior midline glia (MGP). By stage 14, this reverses, with *Dichaete* being expressed weakly in the MGA and MGM, and strongly in the MGP (Soriano and Russell, 1998).

By stage 16, the expression is restricted to 2 clusters of cells in each of the thoracic segments, and a single cell in each of the abdominal segments, as well as the hindgut, brain, and the chordotonal organs of the peripheral nervous system (Soriano and Russell, 1998). In larvae, Dichaete is expressed in discrete patterns in eye-antennal and leg imaginal discs, the CNS, the hindgut and salivary glands (Mukherjee et al., 2000).

1.4.2 Regulatory regions controlling *Dichaete* expression

As described above, *Dichaete* has a complex, dynamic expression pattern, which is widespread, temporally specific and tissue specific throughout embryogenesis (Russell et al., 1996; Soriano and Russell, 1998; Ma et al., 1998). In order to determine the regulatory regions involved in controlling Dichaete, Sanchez-Soriano and Russell (2000) used both reporter constructs and a number of mutant alleles containing chromosomal aberrations with breakpoints close to the *Dichaete* gene to characterise *Dichaete* expression in different tissues. *Dichaete* expression during embryonic development is controlled by a number of cis-regulatory elements, distributed up to 3 kb upstream and 25 kb downstream of the gene. There are a number of regions controlling the expression of *Dichaete* in specific tissues, including different regions for expression in the blastoderm, early and late midline expression, in the hindgut, and also at different times in the neuroectoderm (Sanchez-Soriano and Russell, 2000). The summary of *Dichaete* regulatory regions is shown in Figure 1.8.

1.5 The role of Dichaete in segmentation

The fruit fly, like all arthropods, has a segmented body. In *Drosophila*, there are a total of 14 segments along the AP axis of the body and their location is specified very early during embryonic development. This is a well-studied process, during which the embryo is progressively divided by a hierarchy of regulatory genes (Akam, 1987; Driever and Nusslein-Volhard, 1988). *Dichaete* null mutants exhibit a range of segmentation phenotypes, including deletions removing half of the segments, weaker partial deletions

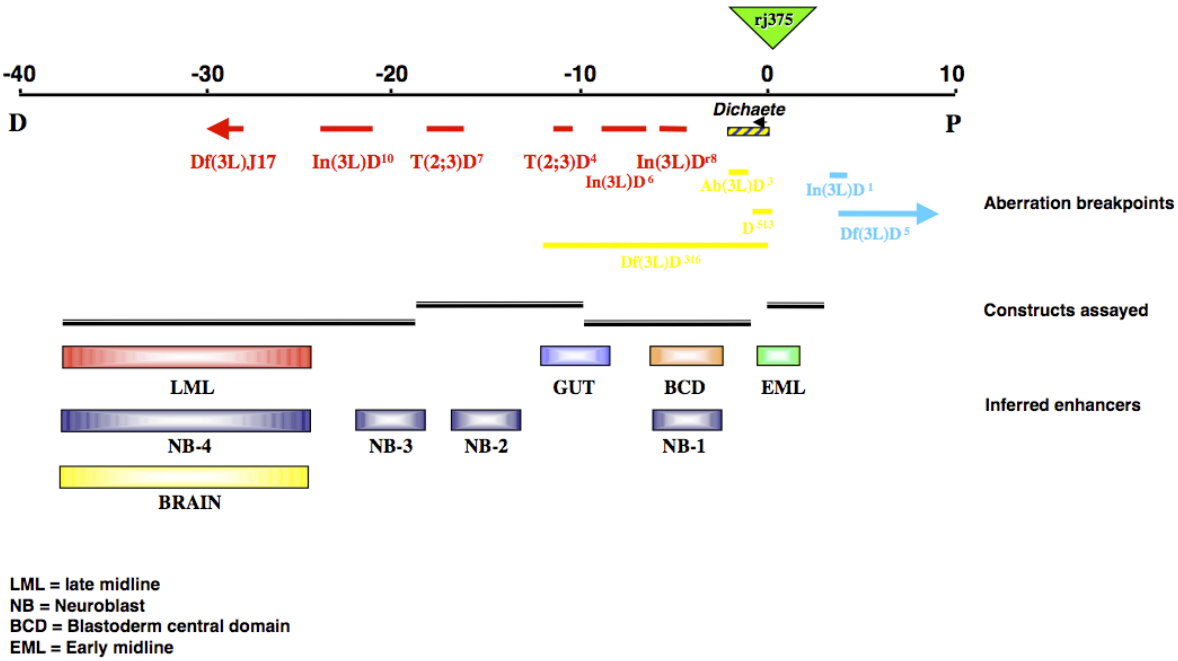

Figure 1.8: A schematic of the regulatory regions controlling *Dichaete* expression, as inferred from experiments utilizing mutant alleles containing chromosomal aberrations (adapted from Sanchez-Soriano and Russell (2000).

and segment fusions with the even numbered metameres (Russell et al., 1996). Nambu and Nambu (1996) also observed a loss and/or fusion of abdominal denticle belts, and stripe-specific defects in pair-rule and segment polarity gene expression.

Dichaete affects the expression of primary pair-rule genes *eve*, *ftz*, *hairy* and *runt* at the syncytial blastoderm stage and it can act to both activate and repress a particular stripe domain. In the particular case of *eve*, expression in stripe 4 was found to be fully repressed when ectopic expression of Dichaete was induced, which was thought to be due to direct action of Dichaete (Russell et al., 1996). Expression of *eve* in stripes 1, 5 and 6 is also directly regulated by Dichaete, and this is thought to be achieved by binding to the *eve* downstream regulatory regions (Ma et al., 1998). Some genetic interactions were detected with *Dichaete* and *Pdm* mutants, so it is thought that Dichaete may act together with the POU domain proteins Pdm-1 and Pdm-2 to regulate *eve* transcription (Ma et al., 1998).

It has been observed a number of times that the Dichaete segmentation phenotype is highly variable. In their original Dichaete paper, Russell et al. (1996) speculate that there are three potential reasons for this: maternal contribution of Dichaete augmenting zygotic Dichaete function, partial complementation by other *Sox* genes, or the variability intrinsic to Dichaete activity, for example, if it has an accessory role in transcription rather than being absolutely essential. Out of these, the maternal contribution hypothesis appears unlikely since no maternal transcripts or protein are detected. The partial complementation by other *Sox* genes has been reported to occur

in the nervous system, but is unlikely to be responsible for the variable segmentation phenotype since no other Sox gene is expressed this early. The third hypothesis is certainly possible, particularly if Dichaete acts as a stabiliser of gene expression, resulting more transcriptional variability in the null mutant. There is a possible further hypothesis, based on network theory. Transcriptional networks tend to be 'scale-free', with a small number of hub genes being connected to a very large number of target genes, and from binding data in the blastoderm (MacArthur et al., 2009), it certainly appears that *Dichaete* is one of the highly interconnected hubs. From the structure of the network, one would expect that while the network may still be partially functional with a major hub missing due to the existence of other hubs, it is likely to be destabilised and more sensitive to stochastic effects, perhaps accounting for the variability in phenotype.

1.6 The role of Dichaete in embryonic CNS development

1.6.1 Regulation of cell fate in the neuroectoderm

Some of the earliest events in the development of the nervous system have been conserved during evolution in most higher eukaryotes. The first step in neurogenesis is the formation of a region of the ectoderm that is competent to adopt a neural fate (Arendt and Nubler-Jung, 1999). In *Drosophila*, this is brought about by Sog, which directs neuroectoderm formation by antagonising Dpp, a process similar to the interaction of Chordin and BMP4 in vertebrates (Piccolo et al., 1996). The expression of proneural genes encoded by the *Achaete-Scute Complex* allows a subsection of the ectoderm to become competent to adopt a neural fate. Subsequently, neuroblasts are selected from clusters of equivalent cells through the interaction of Notch and Delta. They then delaminate and undergo a sequence of asymmetric divisions, resulting in the programmed growth of neurons and glia (for a detailed review, see Skeath, 1999).

Neuroblasts acquire fates in a position-specific manner. In *Drosophila*, there are genetic mechanisms which establish patterns and convey positional information along the AP and DV axis. Of particular relevance to neural development are the segment polarity genes, which divide each segment along the AP axis into four rows (Bhat, 1999), and the genes *vnd*, *ind*, *msh* and *Egfr*, which divide the neuroectoderm into three columns along the DV axis - the medial, intermediate and lateral columns (Skeath, 1999). These two groups of genes combined create a positional grid, providing markers that help determine neuroblast fate.

Dichaete is expressed in the medial and intermediate columns of the neuroectoderm, and phenotypic analysis indicates that its role there is to regulate cell fate and neuroblast formation. Specifically, Dichaete acts in concert with Vnd and Ind, which it has been shown to physically interact with, to establish patterning of the neuroectoderm along the DV axis (Zhao and Skeath, 2002; Zhao, Boekhoff-Falk, Wilson and Skeath, 2007). It is interesting to note that Dichaete appears to have specific functions in each domain: Dichaete and Ind act together to repress *ac* in the intermediate column, while a *Dichaete* null mutation appears to have no effect on *ac* expression in the medial column because of a degree of functional redundancy with *SoxN* in this region (Overton et al., 2002; Zhao, Boekhoff-Falk, Wilson and Skeath, 2007). This is a consequence of interactions with different cofactors, with Dichaete and Ind together having a repressor effect (Zhao, Boekhoff-Falk, Wilson and Skeath, 2007), while Dichaete and Vnd together facilitate transcriptional activation (Yu et al., 2005).

1.6.2 Midline

During *Drosophila* development, midline cells have an important organisational role in helping correctly establish the axon scaffold, as well as being essential for the differentiation of neighbouring mesodermal and ectodermal cells (Crews, 1998). Unlike embryonic neuroblasts, in which the expression domains of *Dichaete* and *SoxNeuro* partially overlap and show some functional redundancy (Overton et al., 2002), the midline is a distinct structure where only one Sox gene, *Dichaete*, is expressed at early stages of development (Soriano and Russell, 1998). This makes it a convenient model system for studying Dichaete function, in addition to the general interest in describing genetic interactions in the midline as a model neural tissue.

Defects have been found in the differentiation of midline glia in *Dichaete* mutants. Axonal defects were observed at the same time, with the longitudinal tracts becoming thinner, and most of the commissures fusing. These defects can be rescued by driving the expression of Dichaete in the ventral midline (Soriano and Russell, 1998), where *Dichaete* was also found to genetically interact with the POU domain gene *ventral veins lacking* (*vvl*). *Dichaete* and *vvl* double mutants showed significantly stronger phenotypes than either of the single mutants, indicating they have a strong synergistic effect on nerve cord development. In addition, ectopically expressing *Dichaete* and *vvl* together in neuroblasts produced a strong phenotype, whereas expression of either gene on its own produced relatively mild phenotypes, suggesting that Dichaete and Vvl work together to regulate gene expression.

In particular, Dichaete and Vvl were found to directly regulate *slit*, the expression of which is greatly reduced in double mutants (Soriano and Russell, 1998). Subsequent

work demonstrated a physical interaction between Dichaete and Vvl on a *slit* enhancer (Ma et al., 2000). The same study also found that *single-minded* (*sim*), a transcription factor required for midline development, acts in concert with *Dichaete* and *vvl* to regulate gene expression in the midline. Dichaete, Vvl and Sim were all found to bind to DNA sites present in a *slit* 1 kb enhancer region, and there is evidence from yeast that they can form a ternary complex, suggesting that the three transcription factors act as coregulators in the midline (Ma et al., 2000). The broader conclusion from this case study is also that functional interactions of Sox, bHLH-PAS and POU protein families are possible, and may happen in tissues other than the midline and as part of various developmental processes.

1.6.3 Neuroblasts

Once a neuroblast delaminates, its particular developmental sequence is specified both through positional information and through the expression of a temporal sequence of transcriptional regulators. As previously described, the positional information for each individual neuroblast is specified by distinct sets of factors along both the AP and the DV axes, which influence its subsequent cell fates during differentiation. Positional information along the DV axis is determined by the combined action of columnar genes: the Epidermal growth factor receptor (Egfr) and the transcriptional regulators Vnd, Ind and Msh, which are expressed in the medial, intermediate and lateral columns of the neuroectoderm, respectively (Skeath, 1999). The neural stem cell fate along the AP axis is regulated by segment polarity genes (such as those discovered by Nusslein-Volhard and Wieschaus (1980)), such as *gsb* and *wg* (Skeath, 1999). The combination of the AP and the DV axis determinants creates a detailed grid of positional information, controlling developmental fates of neuroblasts during early development (Figure 1.9).

As well as the positional information, there is also a temporal sequence of transcriptional regulators (Figure 1.10) that determines the sequence of divisions and the resulting developmental fates of differentiating cells (Maurange et al., 2008). The temporal sequence of transcription factors Hb, Kr, Pdm and Cas (Figure 1.10) in neuroblasts of *Drosophila* embryos and their role in specifying cell fate was first described by Isshiki et al. (2001). Subsequent work shed light on the details of timing-dependent switching of the temporal transcription factors (Grosskortenhaus et al., 2005) and also uncovered a particular competence window for responding to the temporal transcription factors, outside of which their effects on cell fate are different to the effects within it (Cleary and Doe, 2006). The latter is particularly interesting, and offers a model of how a small number of transcription factors can result in a large number of different cell fates. It is thought that Dichaete has an important role during the temporal series. Maurange

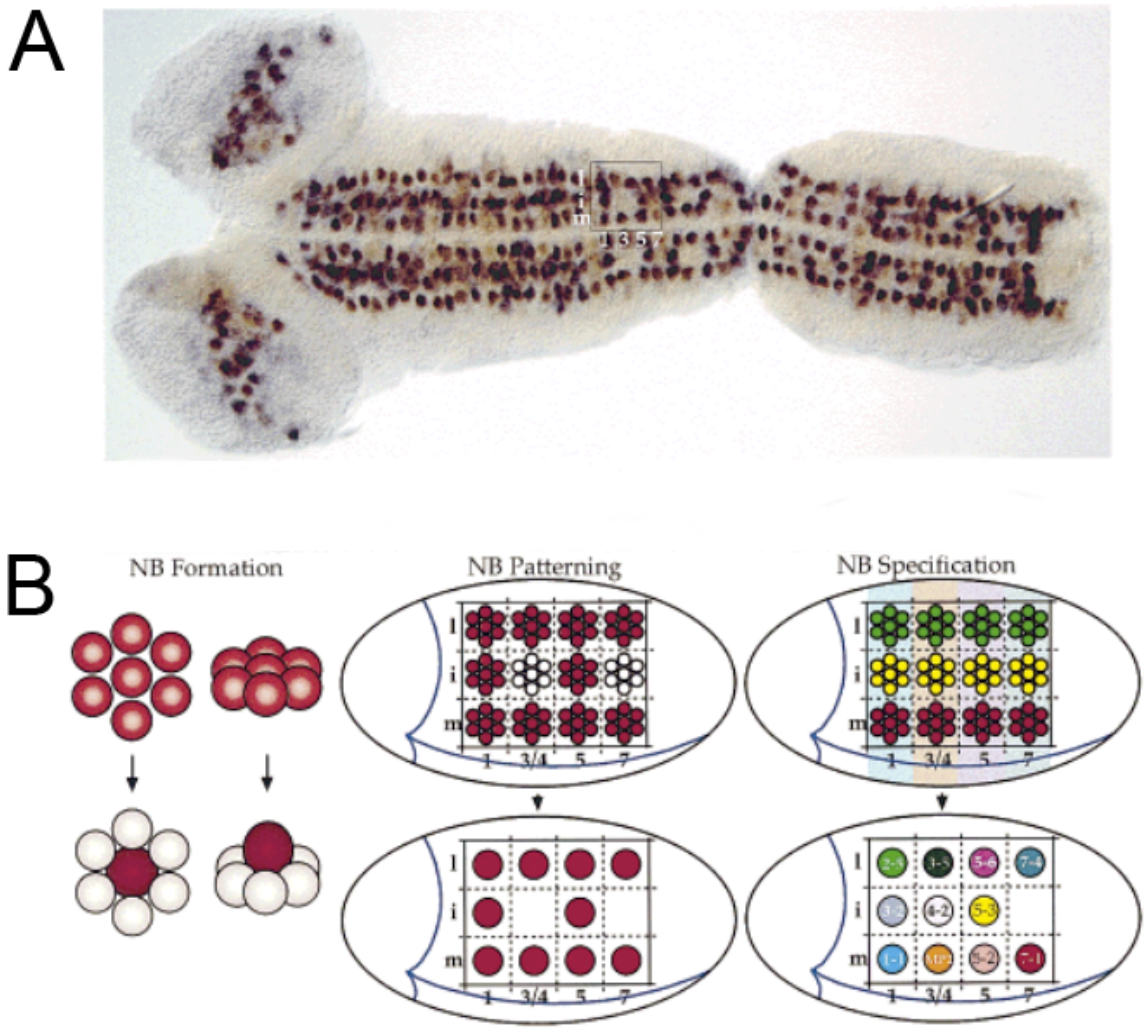

Figure 1.9: The layout of the neuroectoderm. (A) A ventral view of a *Drosophila* embryo undergoing the second wave of neuroblast divisions, stained for the Snail protein which marks all neuroblasts. A single hemisegment is labelled with a rectangle. (B) 'NB formation' shows the selection and delamination of a single neuroblast from a proneural cluster. 'NB patterning' shows the arrangement of proneural clusters in a single hemisegment (top image) and the arrangements of the neuroblasts that develop from the proneural clusters (bottom image). 'NB specification' shows how different positional information leads to different cell fates within a single hemisegment. The red, yellow and green proneural clusters correspond to the activity of medial (m), intermediate (i) and lateral (l) columnar genes, respectively (top image). The colours of neuroblasts correspond to different cell fates, and their names are shown in the middle (bottom image). Figure taken from a review by Skeath (1999).

et al. (2008) find that there is a particular point during neuroblast divisions where Dichaete is switched off and that if this event does not occur, it is not possible for neuroblasts to exit the cell cycle, indicating that the role of Dichaete in neuroblasts is to maintain pluripotency, similarly to the known role of its mammalian homologue Sox2 in stem cells.

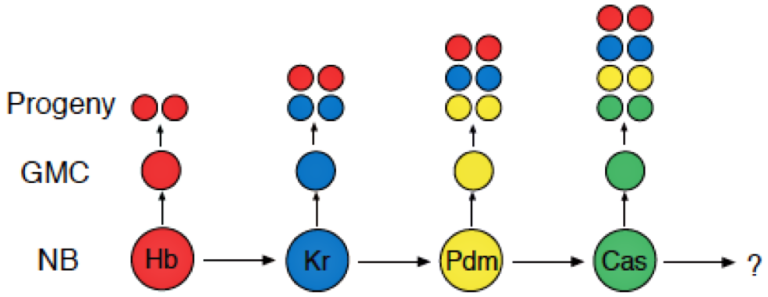

Figure 1.10: A temporal sequence of transcription factors during neuroblast development specifies cell fate. Figure taken from Maurange and Gould (2005).

In *Drosophila*, neuroblasts form in 5 distinct waves, with neuroblasts from different waves being referred to as SI to SV neuroblasts, depending on during which wave they form. In *Dichaete* null mutants, SI medial neuroblasts are not affected, but there is a loss of later delaminating SII and SIII neuroblasts from both medial and intermediate columns (Zhao and Skeath, 2002). The weakening of the neuroblast phenotype is thought to be due to a functional redundancy with SoxN, which is also expressed in medial and intermediate columns (Overton et al., 2002).

1.7 Structure of genes targeted by Dichaete

While studying identified direct Dichaete targets (in particular, referring to the primary pair rule genes *eve*, *h* and *run*, the Dichaete midline target *slit*, *ac* in the neuroectoderm, and *hedgehog* and *dpp* in the hindgut), Overton et al. (2002) observed that genes identified as being Dichaete targets had complex structure and complex regulation, with some of them having regulatory regions that extend over many kilobases. It was suggested that the complexity of the regulatory regions that Dichaete targets is linked to the variable phenotypes observed, with Dichaete perhaps having a structural role facilitated through its ability to bend DNA. Nambu and Nambu (1996) observe that in the case of the determinants involved in segmentation, there is a complicated hierarchical network in place, and that this is reflected in the regulatory regions of the genes involved: they are often very long and contain multiple binding sites for the same transcription factors, allowing for different combinations of regulators giving different outputs. If Dichaete is a master regulator and/or chromatin modifier involved in a number of these complex networks, as appears to be the case, it is plausible that

there may be a pattern in the structure of genes that such a master regulator targets. For example, if it targets a range of other regulators and important developmental genes, this might reflect in terms of the gene structure of Dichaete targets - genes with complex, dynamic patterns of expression tend to have larger areas of non-coding sequence, either in the form of long introns or surrounding sequence Nelson et al. (2004), and this is something that could be a general feature of Dichaete targets, if this is indeed the type of genes that Dichaete tends to target.

1.8 The role of Dichaete in larval development

Dichaete appears to be required for the growth and survival of imaginal cells in larvae. Ectopic expression of Dichaete at this stage causes severe phenotypes in adults, including truncated legs and antennae, and suppressed eye formation (Mukherjee et al., 2000). Dichaete was shown to be required for the expression of *engrailed* and *wingless* during larval development. There was also found to be a link between *Dichaete* ectopic expression morphological defects, and the altered expression of *wingless*, *dpp* and *bric-a-brac* (Mukherjee et al., 2000).

1.9 Dichaete and SoxNeuro

1.9.1 SoxNeuro overview

In *Drosophila*, *SoxNeuro* is the *Sox* gene most closely related to *Dichaete*. However, the majority of the sequence similarity is found in the HMG domain, with the rest of the sequence being quite divergent, so it is likely that any similarity in function found between the two is facilitated via the DNA binding domain (Overton et al., 2002).

SoxN is first expressed in the head region of the syncitial blastoderm, later followed by expression in the ventral neuroectoderm, where it is first expressed slightly earlier than *Dichaete*. It is therefore possible that SoxN regulates the initiation of *Dichaete* expression in this area, however, the study by Overton et al. (2002) showed variable effects: in approximately half of the *SoxN* null mutant embryos, the expression of *Dichaete* was unchanged, whereas in the other half, *Dichaete* levels were reduced in the anterior half of the neuroectoderm. It therefore appears that SoxN has a variable and spatially restricted effect on *Dichaete* expression. As previously discussed, variable phenotypes are observed to be a feature of *Dichaete* mutants (Russell et al., 1996), and SoxN appears to follow the same pattern. As discussed, there are a few potential reasons for this. It is possible that there is a variability intrinsic to their activity,

for example if they have accessory roles and/or act as stabilisers of gene expression, resulting in a variable range of changes in mutants. The other possible reason was that if they are hub nodes in the middle of regulatory networks, it is possible that the network can still partially function due to redundancy, but is destabilised and less robust, resulting in a variable range of phenotypes.

From embryonic stage 8 onwards, *SoxN* is expressed in the developing central nervous system (Cremazy et al., 2000). It is known to have a key role in determining the positioning of the ventral neuroectoderm relative to the dorsal non-neural ectoderm. The expression of *SoxN* is reported to be regulated maternally by Dorsal and zygotically by Dpp, Sog, Brk, Twi and Sna (Figure 1.11). The presumptive neuroectoderm requires Sog and Brk to exclude anti-neural Dpp activity. Based on their effects on SoxN, it seems probable that the expression of SoxN establishes a positive feedback loop increasing the expression of *sog* and *brk* and reducing the expression of *Dpp*, while the opposite happens in regions outside the neuroectoderm, helping to establish sharp tissue boundaries. However, the specific role of SoxN appears to be connected to maintenance rather than original establishment of the differences in gene expression, as the loss of *SoxN* was not observed to change the early expression levels of *brk*, *sog* and *Dpp* (Buescher et al., 2002).

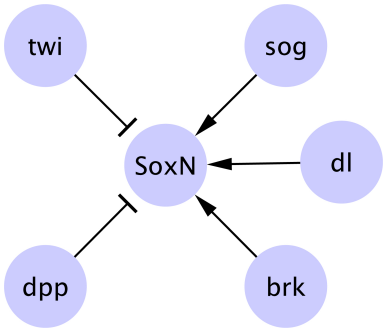

Figure 1.11: Genes regulating SoxN (data from Cremazy et al. (2000))

SoxN appears to be essential for neuroblast formation, with a reduction in lateral and intermediate column neuroblasts observed in *SoxN* null mutant embryos (Buescher et al., 2002). The ventral neuroblasts were found to be largely unaffected in the same mutants, which is thought to be due to functional redundancy with Dichaete. SoxN, like Dichaete, was found to genetically interact with Vnd and Ind, and it is thought to act upstream and parallel to proneural genes (Buescher et al., 2002). There is also evidence that SoxN is involved in later embryonic neuroblast cell fate specification (Overton et al., 2002).

1.9.2 Dichaete and SoxNeuro overlap

There is an overlap in the expression of *SoxN* and *Dichaete* in the neuroectoderm, but the SoxN expression domain extends more laterally: while both proteins are present in the medial and intermediate column, only *SoxN* is expressed in the lateral column. There is also a difference in patterns of cellular localisation between the two - while SoxN is always found in the nucleus, Dichaete displays both nuclear and cytoplasmic localization in the neuroectoderm (Cremazy et al., 2000). Based on their positioning along the DV axis, waves of forming neuroblasts initiate either Dichaete on its own, SoxN on its own, or Dichaete and SoxN together. They also colocalise in the chordotonal organs of the peripheral nervous system (Cremazy et al., 2000).

1.9.3 Differential and redundant function

Overton et al. (2002) described the first evidence for both redundant and differential function of Dichaete and SoxN in the developing CNS in *Drosophila*. The CNS phenotype of a *SoxN* null mutation was studied, and it was found that in the lateral neuroectoderm where *SoxN* is expressed on its own, the expression of *ac* was reduced and there was a reduction in the number of lateral neuroblasts. The same effects were not observed in the medial neuroectoderm, where the expression of *SoxN* and *Dichaete* overlaps; the phenotypes of both single null mutants were mild in this region. On the other hand, *Dichaete*/*SoxN* double mutants showed severe phenotypes throughout the CNS, including a loss of *ac* expressing neural clusters and medial neuroblasts, leading to the conclusion that there is partial functional compensation.

It is interesting to note that, while Dichaete and SoxN appear to cooperate and display partial functional redundancy in the medial column, *Dichaete* null mutants display a partial derepression of *ac* in the intermediate column (Overton et al., 2002; Zhao and Skeath, 2002), indicating that Dichaete and SoxN have distinct functions in this particular region of the neuroectoderm and are capable of acting antagonistically. In the double mutants, the observed effect on *ac* expression is additive: while there is a loss of expression in the lateral column, occasional ectopic expression in the intermediate column is also observed (Overton et al., 2002).

The severe neural hypoplasia observed in *Dichaete*/*SoxN* double mutants indicates that both genes control neural specification at the level of the neuroectoderm. In addition, the loss of neuroblasts in *SoxN* single mutants indicates that SoxN also has a role in later cell fate specification in neuroblasts (Overton et al., 2002), which is also the case for Dichaete (Maurange et al., 2008).

SoxN has also been implicated in epidermis development and there is reported partial redundancy with Dichaete in this function (Overton et al., 2007). To explain their

roles, I will first give a very brief overview of embryonic epidermis development in *Drosophila*. At the end of embryogenesis, a highly patterned cuticle layer is secreted by epidermal cells, with the ventral surface displaying a mixture of segmental denticle belts and naked cuticle. Wingless (Wg), a growth factor belonging to the Wnt family, regulates this process. In a *wg* null mutant, there is no naked cuticle separating the denticle belts, whereas when *wg* is overexpressed the entirety of the ventral epidermis is converted to naked cuticle (Noordermeer et al., 1992; Hays et al., 1997). SoxNeuro was identified in a screen for mutations that suppress or enhance *wg* mutation phenotypes, and it was found that SoxN negatively regulates Wg pathway activity in the embryonic epidermis, likely cooperating with the repressor form of Tcf to achieve this (Chao et al., 2007). Overton et al. (2007) found that *SoxN* is activated by Spi and repressed by Wg, and that it is necessary and sufficient for regulating the expression of *svb*, which in turn stimulates trichome formation, producing denticle belts. The same study shows that in this context, *Dichaete* is coregulated with *SoxN* and has a redundant (though slightly weaker) function in the regulation of *svb* and repression of the Wg pathway.

1.10 Conservation of structure, molecular and developmental functions

The closely matched DNA-binding domain sequence of Drosophila Dichaete and Sox2 from chick, mouse and human suggests that there may be functional conservation between the invertebrate and vertebrate proteins (Russell et al., 1996). Furthermore, Dichaete was found to interact with the POU domain protein Ventral veins lacking in the embryonic ventral midline (Soriano and Russell, 1998), in a manner analogous to the way Sox2 interacts with the POU-domain protein Oct3 (Yuan et al., 1995). There is a set of steps during neural development that is conserved across a wide range of metazoan species. For example, during early neural determination there is antagonistic action of positive and negative acting determinants that help establish the neuroectoderm: the positive action of Sog and negative action of Dpp in *Drosophila* very much mirrors the action of BMP and Chordin in vertebrates. It has been suggested that *Sox* genes also have a conserved function within this context (Buescher et al., 2002). Based on sequence, *Sox1*, *Sox2* and *Sox3* are the closest vertebrate relatives of *Dichaete* and *SoxN*, and these were also found to overlap in the developing nervous system and to display partial functional redundancy during neural development (Nishiguchi et al., 1998). It is therefore possible that the function of *Sox* genes during neural development is highly evolutionarily conserved.

The regulation of SoxN and Sox2 appears to be conserved in *Drosophila* and *Xenopus*: both are negatively regulated by Dpp / BMP4 and positively regulated by Sog

/ Chordin (Sasai et al., 2001; Buescher et al., 2002). Furthermore, Sox2 appears to have a role in the maintenance of neural tissue (Kishi et al., 2000), which is similar to the observed role of SoxN in the neuroectoderm (Buescher et al., 2002). However, experiments indicate that when expressed in the midline, *Sox2* from mouse rescues the *Dichaete* null mutant phenotype (Soriano and Russell, 1998). It has also been reported that mouse Sox1 can rescue *SoxNeuro* lateral neuroblast phenotypes, while SoxN and Sox1 are both incapable of rescuing the *Dichaete* phenotype in the midline (Overton, 2003). This appears to be contrary to the idea that SoxN and Sox2 are conserved and closely related, and instead suggests specific functional conservation between Dichaete and Sox2, as well as between SoxNeuro and Sox1. It is possible that while there is some conservation from the original ancestral *Sox* genes, there has also been some reshuffling of functions between closely related *Sox* genes, leading to only a partial conservation between individual genes in vertebrates and invertebrates.

1.11 Dichaete: a genome-wide view

When I first began my PhD, no genome-wide data was available for Dichaete, so one of the main aims of my project was to generate such a dataset, in order to gain insight into the genome-wide action of Dichaete. In the meantime, three Dichaete ChIP-chip datasets have been published as part of large scale studies. Two datasets were generated by the modENCODE project (Celniker et al., 2009; Negre et al., 2011), a large scale international effort whose aim is to elucidate a range of genomic regulatory mechanisms in the model organisms *C. elegans* and *D. melanogaster*. A third dataset was generated by the Berkeley *Drosophila* Transcription Factor Network Project (BDTNP) (MacArthur et al., 2009), a smaller high throughput project whose aim is to study the genomic regulation of *Drosophila* development. Both projects are reviewed in more detail in Chapter 6. The existence of a range of genomic datasets has been enormously helpful in progressing my thesis work and at various points in my thesis I draw data from these two projects, focusing on transcription factors, insulators, histone modifications and DNA accessibility (MacArthur et al., 2009; Negre et al., 2011; Li et al., 2011).

1.12 The aims of my project

Previous developmental biology studies have identified the roles of Dichaete in early *Drosophila* embryonic development based on phenotypic analysis and gene expression patterns, identifying a number of individual targets and genetic interactions in the process. The first few Dichaete genome-wide binding data sets have also recently become

available, but have focused more on the general patterns observed in the data from a large number of transcription factors, rather than specifically on what the genome-wide data can tell us about Dichaete action in development. In my project, I draw on the data available from previous studies, as well as generating new genome-wide binding and expression profiles, with the aim of answering the following questions:

- Where in the genome does Dichaete bind in vivo?
- Which of the genes associated with binding are likely to be direct targets?
- Which cofactors does Dichaete interact with?
- How do Dichaete and its cofactors act together to regulate gene expression during development?

These are specific questions aimed to improve and expand our understanding of Dichaete action in particular, and from there, the action of Sox genes and transcription factors in general. As covered in this chapter, Sox proteins are an interesting and highly conserved family of transcription factors whose presence spans all across the metazoa. Because of the conservation of their domain structure and the apparent conservation of their function across relatively distantly related species, further understanding of their action in the genome is generalisable in the sense that mechanisms of action observed for one Sox protein may well apply to other Sox proteins too. Since a number of Sox proteins are of medical importance, further insight into their action is therefore of general interest. There is also the broader underlying theoretical question of how transcription factors in general work together to regulate changes in gene expression. Studying specific parts of important transcriptional regulatory networks can lead to broader insights into transcription factor action in general. Finally, the developmental processes that Dichaete is involved in are interesting and fundamental ones, and understanding Dichaete action within processes such as segmentation and nervous system development also furthers our understanding of those processes.

I address these questions experimentally using genomics methods such as generating genome-wide binding data and gene expression profiling. I then use a range of bioinformatics methods to glean the underlying mechanisms of Dichaete action in the genome, and draw on a number of published high-throughput datasets to try and further elucidate the genetic network that Dichaete is involved in during early embryogenesis, with a particular focus on nervous system development.

Chapter 2

Materials and Methods

2.1 Fly work

2.1.1 Fly husbandry

Fly stocks were maintained long-term at 18°C. The stocks used in experiments were reared and maintained at 25°C. All fly stocks were fed standard cornmeal agar medium prepared by the fly media facility in the Department of Genetics, University of Cambridge. Virgin collections were performed within 8 hours after eclosion at 25°C, or 16 hours after eclosion at 18°C, as described in Greenspan (2004). Embryo collection crosses were set up in collection cages and fed fresh yeast on apple juice agar plates. All embryo collections were performed at 25°C.

2.1.2 Fly stocks

The balancers used are described in Flybase (Gelbart et al., 1997). Unless otherwise specified, the stocks were obtained from Cambridge Fly Stock Collection, Department of Genetics, University of Cambridge. An Oregon-R stock was used as the default wild type strain. The list of fly stocks used is shown in Table 2.1.

2.2 Immunohistochemistry

2.2.1 Fixation

Embryos were collected in Nytex baskets, washed with water and dechorionated for 5 mins in 50% commercial bleach. After washing thoroughly with water, they were

Name	**Genotype**	**Reference/Supplier**
D^{r72}	*w; +; D^{r72} / TM3 p{hb-LacZ}*	Stock S1-12
D^{r513}	*w; +; D^{r513} / TM3 p{hb-lacZ}*	Stock S1-2
*twi*Gal4 balancer	*w; +; Dr^{Mio}/TM3, P{w^{+mC}=GAL4-twi.G}2.3, P{UAS-2xEGFP}AH2.3, Sb^1 Ser^1*	Bloomington Drosophila Stock Centre, stock number 6663
D^{r72} / *twi*Gal4 balancer	*w; +; D^{r72} / TM3, Pw^{+mC}=GAL4-twi.G2.3, PUAS-2xEGFPAH2.3, Sb^1 Ser^1*	
D^{r513} / *twi*Gal4 balancer	*w; +; D^{r513} / TM3, Pw^{+mC}=GAL4-twi.G2.3, PUAS-2xEGFPAH2.3, Sb^1 Ser^1*	
Isogenic wildtype stock	*y^1; $Gr22b^1$ $Gr22d^1$ cn^1 $CG33964^{R4.2}$ bw^1 sp^1;* $LysC^1$ $MstProx^1$ $GstD5^1$ $Rh6^1$	Bloomington Drosophila Stock Centre, stock number 2057
*sim*Gal4	*w; p{GAL4-sim}/CyO; p{GAL4-sim2b}*	P. Overton
*pros*Gal4	*w; p{GAL4-$pros^{C20}$} C74-17; +*	Brand lab, Gurdon Institute, Cambridge
UAS GFP	*w; p{UAS GFP}; +*	Stock S3-71
HMG^-	*w; p{UAS dnDichaete (HMG^-) GFP}; +*	Shen (2006)
DWT	*w; p{UAS Dichaete (wild) GFP}; +*	Shen (2006)
EnRep	*w; +; p{UAS dnDichaete (En^{rep}) GFP};*	Shen (2006)
ED2A	*w; p{UAST-ED (E. coli Dam)}; +*	Riaz (2009)
DED3	*w; +; p{UAST-Dichaete-ED (E. coli Dam)}*	Riaz (2009)

Table 2.1: Summary of fly stocks used for experiments.

transferred to a glass scintillation vial. The embryos were fixed with 0.5 ml 37% Formaldehyde, 4 ml fixing solution (100 mM Hepes, 2 mM $MgSO_4$, 1 mM EGTA) and 5 ml Heptane, while being shaken vigorously at room temperature for 20 min. The lower phase was removed with a glass Pasteur pipette, 5 ml MeOH was added and the vial was vortexed vigorously for 30 sec. The embryos were left to sink to the bottom of the vial and were then moved to a 1.5 ml Eppendorf tube using a glass Pasteur pipette. They were washed 3 times with 1 ml MeOH, then 3 times with 1 ml PBT (PBS with 0.1% Triton-X100). They were then washed 3 times with 1 ml PBT, including a gentle rotation at room temperature for 20 min after each wash.

2.2.2 Immunohistochemistry

The antibodies used are shown in Table 2.2. The Dichaete polyclonal antibody was first preabsorbed by incubating with fixed embryos (well washed with PBT) for 2-3 hours at room temperature with gentle rotation. The BP102 monoclonal antibody, which targets the axons of the nervous system, was used without preabsorption.

Antisera	**Source / Reference**	**Host species**	**Concentration**
BP102	DSHB	Mouse	1:100
Dichaete	Soriano and Russell (1998)	Rabbit	1:1000

Table 2.2: Antibodies used

Fixed embryos were left to incubate overnight at 4°C in 100 μl of PBT with the appropriate dilution of primary antibody. The primary antibody was then removed and the embryos were washed 4 times with 1 ml PBT, each time for 10 mins at room temperature with gentle rotation. The embryos were left to incubate for 2 hours in 0.5 ml PBT with 10 μl of the appropriate biotinylated secondary antibody (Vector laboratories). The incubation was stopped with 4 times 10 min washes in 1 ml PBT. The biotinylated secondary antibody was detected using the Vectastain ABC Elite Kit (Vector Laboratories) at a 1:50 dilution in 0.5 ml PBT, left to incubate with the embryos for 30-60 min. The embryos were washed 4 times for 10 min in 1 ml PBT, then transferred to a glass box in 0.5 ml PBT. 20 μl of 10 mg/ml DAB (diaminobenzidine, Sigma) was added and left to incubate for 10 min. The staining was started by adding 5 μl of 1% hydrogen peroxide diluted in PBT, monitored under a microscope and stopped with 6-8 washes in PBT.

For mounting, the embryos were rinsed once in 1 ml PBS, once in 1:1 PBS/glycerol, then stored in 70% glycerol mountant. Sellotape was added to the sides of the slide before mounting to prevent the glass cover slip from damaging the embryos.

2.3 Dominant negative crosses

Some *Dichaete* dominant negative UAS constructs were available from previous work (Shen, 2006). One was UAS dn*Dichaete* HMG^{-} (from here referred to as HMG^{-}), the Dichaete protein missing its active High Mobility Group domain. The other was UAS dn*Dichaete* En^{rep} (from here referred to as EnRep), a Dichaete protein with the *engrailed* repressor domain attached. The third construct included in the study was UAS *Dichaete* (from here referred to as DWT), which is a wild type protein, driven by a UAS promoter and is therefore significantly overexpressed compared to the endogenous protein. All the constructs contain an in-frame N-teminal GFP fusion and they were each compared to flies expressing UAS GFP.

Because the dominant negative constructs are driven by a UAS promoter, and are thus activated wherever the Gal4 protein is expressed: consequently they are convenient for gene expression studies in specific tissues. I performed two such studies: one using the *sim*Gal4 driver (expressed in the ventral midline) and a second using the *pros*Gal4 driver (expressed in neuroblasts). In both cases, female virgins from the Gal4 strains were crossed to males from UAS strains, and stage 8-9 embryos (3.5 to 4.5 hours old) were collected. The control sample for the gene expression experiments was a UAS GFP strain crossed to the same Gal4 drivers. Since all the constructs used were GFP tagged, the embryos were inspected under a confocal microscope (Biorad MRC 1024) to ensure that they were expressed in the correct tissues. For further validation, the embryos were stained with BP102 and the structure of the nervous system inspected for phenotypes.

2.4 *Dichaete* null mutant

The *Dichaete* homozygous null mutants are embryonic lethal and homozygous D embryos were obtained from crosses using balancer chromosomes containing GFP marker - *twi*Gal4 UAS EGFP. The earliest time when fluorescence could be reliably detected was found to be stage 10, so a stage 10-11 embryo collection (5 - 7.5 hours) was used for gene expression profiling.

Two *Dichaete* null alleles were used - D^{r72}, an EMS induced null allele (Soriano and Russell, 1998) and D^{r513}, a small deletion in the coding sequence induced by P-element excision (S. Russell, unpublished data). Alleles were first balanced over the fluorescent balancer chromosome, crossed with each other and homozygous null mutants identified by a lack of fluorescence. For the purposes of studying gene expression changes, the homozygotes were compared to a heterozygous null allele. The heterozygous null allele

embryos were obtained by crossing D^{r513}/twiGal4 UAS EGFP to the isogenic wildtype stock, again identifying the embryos based on lack of fluorescence.

2.5 Gene expression experiments

2.5.1 RNA extraction

The RNA extractions were performed using the Trizol method (details available at the FlyChip website: www.flychip.org.uk) and on ice to prevent RNA degradation. Embryos were collected in PBS, then the PBS was removed and 300 μl TRIzol was added. The embryos can at this stage be stored for about a month at -20°C if required. The embryos were homogenised for 30-60 seconds using a disposable polypropylene pellet pestle. Another 700 μl of TRIzol was then added and the sample was centrifuged at 13,000 rpm (16,000 g) for 10 min. The supernatant was transferred to a fresh 1.5 ml Eppendorf tube. 200 μl of chloroform was added and the sample was vortexed for 30 sec, followed by centrifugation at 13,000 rpm (16,000 g) for 15 min. The upper phase was transfered to a new 1.5 ml tube, while taking care not to touch the interphase, as this would contaminate the RNA sample. The lower phase and interphase were discarded, 0.8 volumes of Isopropanol (around 320 μl) added and the contents mixed gently then left to precipitate for 1 hour at -20°C. The RNA was then pelleted by centrifugation at 13,000 rpm (16,000 g) in a microfuge for 15 min and the supernatant was discarded (being careful not to disturb the pellet). The RNA was washed with 500 μl of 70% ethanol/DEPC H_2O, then centrifuged at 13,000 rpm (16,000 g) for 10 min. The ethanol was discarded and the pellet was left in an open tube to dry for up to 10 min to remove any residual ethanol. The RNA was resuspended in 30 μl DEPC H_2O. The quality of RNA was verified by electrophoresis in a 1% agarose gel followed by ethidium bromide staining and quantified using a Nanodrop spectrophotometer (paying particular attention to the 260/230 ratio, to monitor contamination).

2.5.2 RNA amplification and labelling

Klenow labelling was used for all gene expression experiments performed. In brief, the procedure involves reverse transcribing RNA to cDNA, second strand synthesis, precipitating and purifying the DNA, incorporation of dCTP conjugated Cy3 and Cy5 dyes using a Klenow fragment, and finally, paired competitive hybridisation to microarrays. The standard FlyChip protocol was used for this (details available at www.flychip.org.uk).

Up to 5 μg of RNA per sample was used for the reverse transcription reaction. The volume was made up to 11 μl with DEPC H_2O and 1 μl of oligo-dT was added. The sample was heated for 10 min at 65°C, followed by incubation on ice for 5 min. A master mix containing the following quantity of reagents per sample was prepared: 4 μl of 5x 1st strand buffer, 1 μl of 10mM dNTPs, 2 μl of 0.1M DTT, 0.25 μl RNAsin and 0.75 μl of 200U/μl Superscript III. 8 μl of the master mix was added to each RNA sample, pipetting gently up and down to mix. The samples were then incubated at 46°C for 2 hours, at 65°C for 15 min, followed by snap cooling on ice for a few minutes.

For second strand synthesis. a premix was prepared on ice, containing 9.15 μl H_2O, 7.5 μl 5x second strand buffer, 0.75 μl 10 mM dNTPs, 2 μl of 10U/μl DNA Polymerase I, 0.1 μl of 5U/μl RNaseH and 0.5 μl of 10U/μl *E. coli* ligase per sample. 20 μl of premix was added to each sample, followed by incubation at 16°C for 2 hours.

The double strand cDNA was purified over Sephadex G50 columns: the samples were made up to 100 μl by the addition of 60 μl DEPC H_2O. Phase Lock Gel (PLG) tubes were centrifuged at 13,000 rpm (16,000 g) in a microfuge for 30 seconds to prepare them, then each sample was added to a PLG tube with an equal volume (100 μl) of pH 8.0 Phenol/Chloroform/Isoamyl alcohol. The tubes were shaken by hand for 15 seconds to mix them, followed by centrifugation for 5 min at 13,000 rpm (16,000 g) in a microfuge. The upper phase was removed to a new 1.5 ml Eppendorf tube and the volume reduced to 30 μl in a Speed Vac concentrator. The samples were then loaded on to prepared Sephadex G50 columns and centrifuged for 1 min at 5000 rpm (2,400 g) in a microfuge. The samples were then made up to 100 μl by the addition of 70 μl DEPC H_2O, followed by 3.5 μl of 5M NaCl, 0.5 μl of LPA and 220 μl of 100% EtOH. Samples were incubated for 2 hours at -20°C, then centrifuged for 30 min at 13,000 rpm (16,000 g) in a microfuge. The supernatant was removed and each sample was washed with 500 μl of 70% EtOH, the pellet air dried then dissolving in 10 to 20 μl DEPC H_2O, depending on pellet size. The sample concentration was measured using a Nanodrop spectrophotometer.

cDNA was labelled using the random priming method: up to 1000 ng cDNA per sample was made up to 25 μl using MilliQ water. 20 μl of 2.5x Random Primer reaction buffer was added and the samples were incubated at 100°C for 5 min in a PCR (polymerase chain reaction) block, then placed on ice to cool. A premix was prepared containing 1 μl of 10x low-C dNTP, 2 μl of the relevant fluorescent dye (dCTP conjugated Cy3 or Cy5) and 1 μl of 40U/μl Klenow per sample. 4 μl of pre-mix was added to each sample, pipetting up and down gently to mix. The samples were incubated at 37°C for 2 hours and the reaction stopped by adding 5 μl of stop buffer (0.5 M Na_2EDTA, pH 8.0). The Cy3 and Cy5 labelled sample and control pairs were combined in 1.5 ml tubes

and the volume was reduced to 25-30 μl in a SpeedVac concentrator. Sephadex G50 purification, with two G50 columns per sample, was used to eliminate unincorporated dyes. The resin in the G50 columns was resuspended by vortexing and the column was placed in a 1.5 ml tube and pre-spun in a microfuge for 1 min at 5000 rpm (2,400 g). The cap was then removed and the column was placed in a new 1.5 ml tube, ready for use. Half the sample was pippeted onto the centre of the column and centrifuged in a microfuge for 1 min at 5000 rpm (2,400 g): the process was repeated for the other half of the sample with a new column. The volume of the sample was then reduced to 2-5 μl using a SpeedVac concentrator. 2 μl of 10 mg/ml sonicated salmon sperm DNA was added, along with 140 μl of Ambion Hybridisation Buffer 1. The sample was heated at 100°C for 2 min, spun in a microfuge at 13,000 rpm (16,000 g) for 1 min, then kept at 65°C until ready to load on a slide.

2.5.3 Gene expression microarrays and hybridisation

The Flychip in-house printed FL002 and FL003 gene expression arrays were used (GEO Platform Accession numbers GPL5135 and GPL14121, respectively). Four biological replicates were performed for each expression study. Dye swaps were incorporated into the experimental design to correct for bias, with two out of four from each replica group dye swapped.

The samples were hybridised to arrays in a Genomic Solutions Hybridisation Station: 135 μl of sample was loaded onto each array, and they were left to hybridise for 16 hours at 65°C with agitation. After the hybridisation, the slides were washed to rinse off non-hybridised samples. For the rinsing, two wash solutions were used - wash solution 1 (0.2x SSC, 0.2% SDS) and wash solution 2 (0.2x SSC). 400 ml of wash solution 1 and 800 ml of wash solution 2 were preheated to 55°C in a water bath. A black staining box was filled with warm wash solution 1, with the rest poured into a glass staining dish. The hyb cassette was submerged into wash solution 1 in the glass dish, and the slides were carefully lifted and transferred into a wash rack (maximum 12 slides per rack), which was put into the black wash box and submerged in warm wash solution 1. It was then left to incubate on a shaker with gentle agitation (50 rpm, 0.24 g) for 20 mins at room temperature, keeping the lid on the wash box to shield them from light. The slides (still in the wash rack) were then transferred into a fresh staining box, and gently dipped up and down for 1 min at room temperature. The 1 min dipping was then repeated another two times using fresh wash solution. The slides were then rinsed by dipping the wash rack into MilliQ-water for 3 seconds at room temperature. The slides were transferred into a slide box lined with tissue paper, dried in a centrifuge at 1000 rpm (96 g) for 5 mins, and transferred into a dry slide box, using an AirDuster

to remove any water droplets. The slides were then scanned using an Axon GenePix 4000B scanner.

2.5.4 Data processing

In the case of the dominant negative experiments, the standard FlyChip data analysis pipeline was used (www.flychip.org.uk): Dapple (Buhler et al., 2000) was used for spot-finding (manually checking for any spot-finding mistakes and areas of high background) and variance stabilizing normalisation (vsn) (Huber et al., 2002) was performed. In brief, vsn uses an affine transformation followed by a $\log_2$ variance-stabilising transformation to normalise the data. The final results were analysed using CyberT to assess statistical significance (Baldi and Long, 2001). The thresholds used to find differentially expressed genes were average M-value < -0.5 or > 0.5 (where the M-value is the $\log_2$ of the ratio of sample vs control intensities), and p-value < 0.05.

In the case of the *Dichaete* mutant experiment, problems were encountered during the normalisation step, due to the large number of genes that changed expression. Standard normalisation methods tend to assume that 50%-75% of genes studied will not be differentially expressed and this was an inaccurate assumption for this particular experiment. To address this problem, a previously published invariants normalisation method (Pradervand et al., 2009) was suggested by Stewart MacArthur (CRUK) and implemented by Bettina Fischer. The later stage of the analysis was again performed using limma (Smyth, 2005), a more stringent alternative to CyberT. The thresholds used to find differentially expressed genes were average M-value < -1 or > 1, and p-value < 0.01, in order to filter the large number of positive hits.

2.6 DamID

As part of a PhD project by Riaz (2009), *E. coli* Dam Methylase was fused to *Dichaete* and inserted into a vector under the control of a standard UAS promoter. Thus any areas of the genome bound by the Dichaete-Dam fusion will be methylated, with the weak expression from the uninduced UAS promoter limiting the amount of fusion protein expressed, thus preventing saturation (van Steensel et al., 2001). As a control, DNA was prepared in parallel from embryos carrying an unfused Dam Methylase under the control of the same UAS promoter. The protocol used here is based on the original Vogel et al. (2007) method, with some minor adjustments.

The method involves first extracting DNA from embryos expressing the Dam methylase-Dichaete fusion (and also from control embryos only expressing Dam methylase), followed by digestion with a restriction enzyme that only cuts methylated sites (*DpnI*),

a ligation of adaptor DNA sequences to DNA overhangs left from the digestion, followed by a further digestion by a restriction enzyme (*DpnII*) that targets the same sites as *DpnI*, but only if they are unmethylated. From there, the sequences containing adaptors are amplified using PCR to generate sufficient material for labelling and hybridisation to Nimblegen genome tiling arrays (minimum material required is 1.5 μg DNA), followed be fluorescent labelling and hybridisation. Three biological replicates were performed. There is a potential for bias to be introduced during the amplification step, which is why both the sample and the control were treated the same - if bias is introduced, it should be similar for both sample and control, and therefore not a concern, since we only look at the ratios, rather than absolute values. Further to that, systematic shifts in intensity values that are different between samples and controls should show up on QC scatter plots, so there is a check in place to look for this.

2.6.1 Embryo collection and DNA extraction

The fly stocks used were the DED3 strain containing a UAS Dichaete-Dam methylase fusion protein, and the ED2A strain containing a UAS Dam methylase, originally produced by Riaz (2009). Embryos between 0-12 h were collected (stage 1 - 15) and DNA extracted using a Qiagen DNeasy Blood and Tissue kit following the manufacturers recommendations. Up to 50 mg of embryos were placed in a 1.5 ml tube, 50 μl of PBS and 1 μl of RNase added, and the sample homogenised with a disposable polypropylene pellet pestle. 130 μl of PBS, 20 μl proteinase K and 200 μl of Buffer AL (lysis buffer purchased from Qiagen) was added to the sample, which was then mixed thoroughly by vortexing and incubated at 56°C for 10 min. After the incubation, 200 μl of 96-100% ethanol was added to the sample, which was mixed thoroughly by vortexing. The sample was then pipetted into a DNeasy Mini spin column and centrifuge in a microfuge at 8000 rpm (6,100 g) for 1 min. The flow through was discarded and the column placed in a new collection tube. 500 μl of Buffer AW1 (wash solution 1, purchased from Qiagen) was added, and the column was again centrifuged in a microfuge at 8000 rpm (6,100 g) for 1 min. The flow through was discarded, 500 μl of Buffer AW2 (wash solution 2, purchased from Qiagen) was added to the column followed by centrifugation in a microfuge for 3 min at 14,000 rpm (19,000 g). 100 μl of MilliQ water was added to the column, incubated for 1 minute, then the DNA was eluted into a new 1.5 ml collection tube by centrifugation in a microfuge at 8000 rpm (6,100 g) for 1 min.

2.6.2 DNA precipitation

The concentration of each DNA extract was determined using a Nanodrop spectrophotometer. Two volumes of 100% ethanol and 0.1 volume of 3M sodium acetate (pH 5.5)

were added and mixed with the sample. The DNA was precipitated by storing on ice for 30 min, followed by centrifugation for 30 min at 4°C in a refrigerated microfuge to pellet the DNA. The supernatant was discarded and the DNA washed by adding 500 μl of 70% ethanol. The sample was then centrifuged for 5 min (4°C, 16,000g), after which the supernatant was removed. The pellet was left to air-dry for 5-10 min, to evaporate any residual ethanol. The DNA was dissolved in pH 7.5 $T_{10}E_{0.1}$ (TE buffer with reduced EDTA - 10mM Tris-HCl, pH 7.5, 0.1mM EDTA) prewarmed to 55°C, at a concentration of 1 μg/μl, based on the original Nanodrop measurement. The concentration was verified using the Nanodrop (and a 1:10 sample dilution, in order to save as much sample as possible). 500 ng of DNA was then used for the next step of the experiment.

2.6.3 Digestions

Two negative controls were included: one where a Dichaete-Dam sample was processed, but with the *DpnI* enzyme excluded, and the second where the DNA ligase was omitted. The double stranded adaptor oligonucleotide (dsAdr) was prepared from two oligonucleotides - "AdRt" (5'-CTAATACGACTCACTATAGGGCAGCGTGGTCGCGGCCGAGGA-3') and "AdRb" (5'-TCCTCGGCCG-3'). AdRt and AdRb were both dissolved in water. A 50 μM stock of dsAdr was prepared by combining equal volumes of 100 μM AdRt and 100 μM AdRb, placing the tube to float in a beaker of water at 90°C, and leaving it until the water cooled to room temperature, so that the adaptors can anneal slowly.

The *DpnI* digest was set up using 0.5 μl of DNA sample (500 ng), 1 μl 10x Buffer 4 (NEB) 0.5 μl of 20 U/μl *DpnI* (NEB) and 8 μl H_2O, and left to incubate overnight at 37°C. *DpnI* was then heat-inactivated for 20 min at 80°C.

The ligation was set up in the same tubes using the 10 μl of *DpnI* digested gDNA, 2 μl 10x ligation buffer (Roche), 1 μl of 5 U/μl T_4 ligase (Roche), 0.8 μl of 50 μM dsAdr and 6.2 μl H_2O, and left to incubate overnight at 16°C. Ligase was then heat-inactivated for 10 min at 65°C.

The *DpnII* digestion was set up in the same tubes using the 20 μl ligated DNA, 5 μl 10x DpnII Buffer (NEB), 1 μl of 10 U/μl *DpnII* (NEB) and 24 μl H_2O, and left to incubate for 1 hour at 37°C.

2.6.4 PCR amplification

The Adr-PCR primer is an oligonucleotide (5'-GGTCGCGGCCGAGGATC-3), with both the 5' and 3' ends left unphosphorylated.

Reagent	Amount
DpnII digested DNA	10 μl
10x cDNA PCR buffer (Clontech)	5 μl
Primer Adr-PCR (50 μM)	5 μl
dNTPs (10 mM)	1 μl
PCR advantage enzyme mix (Clontech, 50x)	1 μl
H_2O	31.75 μl
Total volume	50 μl

Table 2.3: PCR reagents.

The digested DNA was amplified by PCR with the reagents shown in Table 2.3. The reaction conditions were as follows: 68°C, 10 min, for 1 cycle; 94°C for 1 min, 65°C for 5 min, 68°C for 15 min, for 1 cycle; 94°C for 1 min, 65°C for 1 min, 68°C for 10 min, for 4 cycles; 94°C for 1 min, 65°C for 1 min, 68°C for 2 min, for 17 cycles. The samples and PCR controls were then analysed on a 1% agarose gel. Provided the DNA was visible on the gel and the controls came back negative, the remainder of the sample was then purified using a Qiaquick PCR purification kit, and the concentration was measured using a Nanodrop.

2.6.5 Labelling and microarrays

NimbleGen *Drosophila melanogaster* 2.1M Whole Genome Tilling Arrays from Roche were used. The labelling protocol using the Invitrogen BioPrime Kit was followed. The recommended quantity of DNA for use with NimbleGen arrays is 3-6 μg, but the protocol has been successfully performed in FlyChip with as little as 1.5 μg. Two labelling reactions, each with 1 μg of DNA each, were performed for each sample and control: 1 μg of DNA was made up to a volume of 32.1 μl by adding MilliQ water and 30 μl of 2.5x Random Primer Reaction Buffer was added. The samples were incubated at 100°C for 5 min in a PCR machine, then placed on ice. A master mix was made up with 7.5 μl of 10x low-C dNTP, 4.5 μl dCTP conjugated Cy3 or Cy5 dye (as appropriate) and 1.8 μl of 40U/μl Klenow. 13.8 μl of the master mix was added to the DNA, pipetting up and down gently to mix. From this point on, the samples were protected from light at all times, because the dyes are light sensitive. The samples were then incubated at 37°C for 2 hours. The reaction was stopped by adding 7.5 μl of stop buffer.

The two same dye sample reactions were combined in a 1.5 ml tube, and 17.25 μl 5M NaCl and 165 μl isopropanol was added. The samples were vortexed and incubated at room temperature, followed by centrifugation in a microfuge at 13,000 rpm (16,000 g) for 10 min. The supernatant was removed and the pellet rinsed with 500 μl 80% ice-cold EtOH, spun at 13,000 rpm (16,000 g) for 2 min, drained and air dried. The dried

pellet was resuspended in 50 μl water, vortexed, pulsed in a microfuge, left to incubate for 5 min, vortexed and pulse centrifuged again. The concentration was measured using a Nanodrop.

34 μg of test sample was combined with 34 μg of the matching control sample. The contents were dried in a speed vac, then resuspended in 12.3 μl of MilliQ water and vortexed. A Hybridisation Solution master mix was prepared by mixing 29.5 μl of 2x hybridisation buffer, 11.8 μl of hybridisation component A and 1.2 μl of the alignment oligo. 31.7 μl of the master mix was added to each sample pair, followed by vortexing and spinning. The samples were incubated at 95°C for 5 min, then at 42°C for a minimum of 5 min. The microarrays were prepared and placed into the hybridisation station (NimbleGen Hybridization System 4), after which 41 μl of the sample was loaded onto the array. The arrays were hybridised for 20 hours at 42°C, washed and then scanned using an Axon GenePix 4000B scanner.

2.6.6 Microarray data processing

The spot-finding and generation of pair files was carried out using the Roche proprietary software NimbleScan. The remainder of the data processing was performed using the TiMAT pipeline (http://bdtnp.lbl.gov/TiMAT/) using standard settings. The normalisation method used was quantile normalisation on samples and controls separately, since their signal distributions were quite different. Quantile normalisation on all of them together appeared to normalise away the entirety of the signal for the experiment. Peak-finding was performed as part of the TiMAT pipeline at FDR values 1%, 5%, 10% and 25%.

2.7 Bioinformatics methods

All analysis, except where indicated otherwise, was performed using custom written Perl and R scripts, with data stored in mySQL databases. Perl version 5.8.8, R version 2.12.2 and mySQL version 14.12 were used. Genome visualisation was performed using the Affymetrix Integrated Genome Browser version 5.12 (Nicol et al., 2009). *Drosophila melanogaster* Release 5 genome sequence version, and Genome Release 5.26 annotation are used consistently throughout this thesis. Significance of overlap between different binding datasets was determined using a resampling-based method implemented in the Cooccur package (Huen and Russell, 2010).

2.7.1 Gene list analysis

Gene associations for binding intervals were found using a custom method. If there was direct overlap of a gene with a binding interval, the gene or genes in question would be classed as associated. If no direct overlaps were found, the closest gene within a 10 kb region would be classed as associated with the intervals. All gene lists were subsequently uploaded into FlyMine (Lyne et al., 2007). From there, FlyMine widgets such as the ones for GO enrichment, domain enrichment and publication enrichment were routinely used for gene list analysis.

2.7.2 GO enrichment analysis

All GO enrichments were found using FlyMine (a hypergeometric test combined with a Holm and Bonferroni correction for multiple testing), unless otherwise indicated in the text. Specific analysis methods are detailed in the relevant chapters. In the case of GO enrichments performed on gene lists generated from binding data, a bias is introduced because longer genes are both more likely to be hit by binding intervals and have specific functions. The standard analysis method does not account for this bias, and it has not been corrected for anywhere in this thesis. This problem is highlighted in the relevant analysis sections, and the basis of the problem, the available solutions and the impact on the analysis in this thesis is discussed in detail at the end of Section 4.5.2 in Chapter 4.

2.7.3 Motif analysis

The NMICA software was used for motif searching (Down and Hubbard, 2005). Binding intervals were ranked according to the highest scoring window and between 100 and 200 top scoring intervals were selected (as indicated in the relevant results sections). The searches were performed for a 6-10bp long motif, using Markov background order 1 and 6 background classes. 15 motifs were retrieved from each search. The motif scanning with the Dichaete motif was performed using nmscan, a part of the NMICA package, with the cut-off score 5. A Monte Carlo method was used for estimating the overrepresentation of motifs in individual intervals, by reshuffling the columns of the existing motif and scanning the intervals with the new randomized motifs. When comparing a number of motifs of different lengths, matrix scan, a part of the RSAT package (Thomas-Chollier et al., 2008), was used, with Markov background order 1 and a score cutoff between 4 and 6, as indicated. STAMP was used for searching for matches with already known motifs (Mahony and Benos, 2007).

2.7.4 Visualisation

The software Cluster 3.0 was used for hierarchical clustering (de Hoon et al., 2004), with the default settings used being uncentered correlation and centroid linkage. The clustering output was subsequently visualised using Java TreeView (Saldanha, 2004). Cytoscape version 2.6.2 was used for network visualisation (Shannon et al., 2003). Where indicated, GO enrichments were performed and visualised using Ontologizer (Bauer et al., 2008).

Chapter 3

Dichaete genome-wide binding

The cis-regulatory systems that mediate regional specification are thereby the keys to understanding how the genome encodes development of the body plan. (Davidson, 2001)

In his treatise on genomic regulatory systems, Davidson (2001) argues, as others have done, that understanding the regulatory logic of the genome is a cornerstone of our efforts to understand the way the information in the genome is used to generate biological diversity. Whether we seek to understand the way organisms have evolved, how organisms grow and develop or how biological processes can go wrong leading to diseases such as cancer, there is a unifying thread: the combinatorial activity of transcriptional regulators at defined locations in the genome to control where and when genes are expressed. A key component of the effort to understand this process is to define where transcription factors are active in the genome, in order to identify the cis-regulatory modules controlling gene expression. The recently developed techniques for characterising in vivo protein-DNA interactions across entire genomes, such as ChIP and DamID, are powerful tools which are beginning to yield genome-wide views of transcription factor activity. This chapter focuses on an in-depth comparative analysis of 4 Dichaete genome-wide binding data sets generated from developing embryos. The aims of the chapter are to assess the quality of the available datasets, investigate the possible routes of approaching a genome-wide functional analysis and generate a preliminary set of results with respect to Dichaete biology, to be further investigated in ensuing chapters.

3.1 Introduction

Organismal development is controlled on a genetic level by the combinatorial binding of transcription factors to precise locations in the genome, allowing for transcription

complexes to form and facilitating the control of gene expression. Research suggests that few, if any, transcription factors act alone (Davidson, 2001). Instead it is through their presence/absence and the cooperative action between them, coupled with external signals and the orchestration of a precise cascade of events, which brings about the complex process of development (Mann and Carroll, 2002). Studying the binding patterns of individual transcription factors that are known to have developmental roles is an important part of solving this puzzle. High-throughput genomics methods have provided a new and powerful tool for studying the binding location of transcription factors in vivo. This has the potential to lead to insights into the action of the individual regulators studied as well as a broader understanding of the general principles of transcription factor action. What follows is an overview of the available experimental techniques, their individual pros and cons, and where they fit into what is already known about Dichaete during development.

3.1.1 ChIP-chip

While in vitro experiments can provide a useful starting point for the analysis of protein-DNA interactions by, for example, defining consensus DNA sequence motifs, a realistic picture of what happens inside living organisms can only emerge from studying interactions in vivo. Chromatin immunoprecipitation (ChIP) has emerged as a powerful technique that takes a snapshot of protein-DNA interactions in vivo, opening up new avenues for understanding aspects of transcriptional regulation (for a review, see Buck and Lieb, 2004). Provided that an appropriate antibody or tag for the protein of interest is available, ChIP provides a rapid way to gather data on where in the genome a protein is bound. This versatile technique has been used for mapping histone modifications, transcription factor binding, DNA methylation and the association of chromatin-modifying enzymes with the genome (Lieb et al., 2001; Iyer et al., 2001; Ren et al., 2000; Kwong et al., 2008; Celniker et al., 2009; MacArthur et al., 2009; Bernstein et al., 2002; Ng et al., 2002). The preparation of samples can often be adapted to allow the analysis of tissue or developmental stage specific interactions (Wilczynski and Furlong, 2010), facilitating a detailed analysis of the changing patterns of protein-DNA interactions underpinning developmental processes.

The first step in a ChIP analysis is the crosslinking reaction, which fixes proteins at whatever position in the genome they occupy upon crosslinking. Formaldehyde is the agent most commonly employed today, though other methods such as UV-crosslinking have been used (Walter and Biggin, 1997). Crosslinked nuclei are purified and the chromatin is randomly fragmented, usually by sonication, to produce chromatin fragments less than 1 kb long. With the bound DNA still attached, the protein of interest is enriched using an appropriate antibody. Crosslinks are reversed by heat treatment

and the DNA is then purified. A variety of methods can be used to identify some or all of the enriched DNA fragments: these include cloning and sequencing (Fullwood et al., 2009), specific polymerase chain reaction (PCR), Southern blotting (Solomon et al., 1988), array hybridisation (Ren et al., 2000; Iyer et al., 2001) and next generation sequencing (Robertson et al., 2007; Johnson et al., 2007). Since some regions of the genome will be unspecifically enriched by the immunopurification, a control reaction using IgG, preimmune sera or mock purification is carried out in parallel. A comparison of the specifically enriched and control samples identifies genomic regions that were crosslinked to, and by inference bound by, the protein of interest. Input DNA is kept, and can be used to help normalize the signal, based on how much DNA from each genomic region was present in the sample in the first place. As the IP reaction is known to be very noisy and to result in false positives, the presence of controls and a sufficient number of replicates is crucial to ensuring that the experimental data gained is of a high quality (Birch-Machin et al., 2005).

In the pioneering work of Solomon et al. (1988), a Southern hybridisation assay was used to identify protein-DNA interactions at specific genomic locations and this was later most commonly assayed by specific PCR reactions. ChIP entered the post-genomic era when microarray technology was first used to infer the identity and genomic location of enriched DNA (Ren et al., 2000; Iyer et al., 2001). By hybridising the enriched DNA to a microarray containing probes representing all intergenic sequences or even an entire genome, a comprehensive view of DNA binding is produced from a single assay. While the costs of such assays were initially high and the array densities limited, the prices have dropped over time and there are now a number of tiling microarray platforms available for several model organisms, including *Drosophila*. Perhaps more importantly, there are also a variety of well-developed and freely accessible data analysis pipelines for processing ChIP-chip data.

3.1.2 ChIP-seq

The ChIP-seq technique uses the same ChIP protocol, but identifies the enriched DNA using sequencing (instead of microarrays or PCR). In brief, the ChIP-seq method involves sequencing short reads from each end of the ChIP-enriched DNA fragments, mapping the sequence reads back to a reference genome and using the frequency of sequence reads in a particular region to assess enrichment. The basic concept behind ChIP-seq has existed for some time, since using sequencing to identify the composition of a mixture of DNA fragments is an obvious course of action. However, the approach only became a viable alternative to array-based methods once the available sequencing technologies were fast enough and cheap enough. The newly developed ChIP-seq approach uses second generation massively parallel DNA sequencing platforms, which

can produce several orders of magnitude more sequencing reads directly from ChIP enriched DNA at a fraction of the time and cost of traditional Sanger sequencing. The method is extremely high-throughput with, for example, the Illumina 1G Genome Analyser producing more than a gigabyte of sequence in a single run (for a review of next generation sequencing technologies, see Holt and Jones, 2008). While relatively new, initial studies using the ChIP-seq method look very promising and the method is rapidly gaining popularity for a variety of applications including studying transcription factor binding (Robertson et al., 2007; Johnson et al., 2007), histone modifications (Mikkelsen et al., 2007; Barski et al., 2007) and DNA methylation (Pomraning et al., 2009).

While ChIP-seq is a fast and powerful technique, there was no compelling rationale for using it to achieve the aims of this thesis. ChIP-seq is invaluable when dealing with large genomes containing a high proportion of repetitive DNA (which can confound array analysis) and where tiling arrays are expensive since multiple arrays are needed. However, *Drosophila* has a small compact genome and available tiling arrays are of sufficient density to provide high resolution genome coverage on a single array. The limiting factor in this case is likely to be the size of the sonicated DNA fragments rather than the microarray platform (Aleksic and Russell, 2009). Additionally, methods and analysis pipelines for ChIP-chip have been systematically tested and are well developed (Johnson et al., 2008), whereas the biases inherent in ChIP-seq are only just starting to emerge and be corrected (Cheung et al., 2011). It was therefore concluded that a combination of cost and technical reasons indicated using tiling arrays would be optimal for the purposes of this project.

3.1.3 DamID

DamID is an alternative to ChIP for studying transcription factor binding in vivo and is of particular use in instances where an antibody for the protein of interest is unavailable or unreliable, making ChIP-chip difficult. Originally developed by van Steensel (van Steensel and Henikoff, 2000; van Steensel et al., 2001), it uses adenine methylation to label locations at which a protein of interest is bound. A fusion protein is created, where *E. coli* DNA adenine methyltransferase (Dam) is attached to the transcription factor of interest and subsequently expressed ubiquitously at low levels in vivo. Thereafter, the transcription factor fusion protein leaves a methylation tag whenever it binds to DNA in the genome. While low levels of 5-methylcytosine are reported to be present at specific stages during *Drosophila* development (Lyko, 2001), no adenine methylation is found in the *Drosophila* or any other eukaryotic genome, meaning that any adenine methylation tags detected must have been produced by the fusion protein. A necessary control is a Dam only protein to account for unspecific

methylation that occurs due to the inherent Dam affinity for DNA. While DamID is mostly performed in cell lines (De Wit et al., 2008; Greil et al., 2007; Pindyurin et al., 2007; Filion et al., 2010), a number of studies have also been performed in whole organisms (Choksi et al., 2006; Pym et al., 2006). While at first sight it may appear that the DamID technique requires direct binding of the fusion protein to DNA and therefore might be unsuitable for studying features such as histone modifications or chromatin associated proteins that do not have direct access to DNA, a range of proteins including ones not directly binding to DNA has been studied in cell lines. This implies that bringing Dam into the vicinity of DNA is sufficient to allow methylation so this limitation may not be too severe in practice. Another potential limitation of the DamID technique is its reported lower resolution compared to ChIP (Holland et al., 2005). However, the reports on this are mixed: Moorman et al. (2006) use both ChIP and DamID, and do not observe a notable difference in resolution. An advantage of the DamID approach is that the ChIP-chip protocol is complex, can often be very noisy and requires many individual protocol modifications depending upon the antibody and tissue source. In contrast, the DamID protocol is comparatively straight-forward, and is reported to perform quite robustly.

3.1.4 Origin of the binding signal

The signal from a ChIP-chip reaction is generated by analysing the fragments of DNA crosslinked to the protein of interest. The bulk of the covalent links formed through formaldehyde crosslinking in a ChIP reaction are thought to be protein-protein (Schmiedeberg et al., 2009). This means that, for example, in cases of DNA looping, DNA will be captured and registered as 'bound' if it happens to localise near the protein of interest. This will also happen with pieces of DNA that happen to be packaged in a way that brings them close to the protein of interest. This is why it has been possible to successfully map chromatin associated proteins such as Polycomb and Mod(mdg4) using ChIP, despite the fact that they interact with DNA indirectly (Kwong et al., 2008; Negre et al., 2010).

In any ChIP reaction there is therefore an unknown quantity of genuine signal that is not indicative of direct DNA binding and this may account for some of the variability reported for immunoprecipitation reactions. It has also been reported that formaldehyde crosslinking fails to capture transient interactions (ones where the protein binds for less than 5 seconds) that are detected using fluorescence techniques in live cells (Schmiedeberg et al., 2009). This is not a problem if the functional role of a protein requires it to stay in one place for a while, but might be problematic if the biologically interesting protein-DNA interactions happen to be transient.

It is therefore inevitable that even with appropriate controls, in every ChIP-chip reaction there will be at least some false positives detected (pieces of DNA in the proximity of, but not bound by, the protein of interest) and some false negatives (transient interactions not picked up by the crosslinking). Furthermore, in contrast to chromatin from homogeneous sources such as cell lines, when using whole animals, such as *Drosophila* embryos, there is a mixture of tissues, and thus the final signal from a ChIP-chip reaction will be averaged over developmental time and across multiple tissues.

A DamID reaction involves isolating and amplifying the methylated DNA labelled by the Dam methylase enzyme fused to the protein of interest. This is likely to mean that some degree of direct physical proximity is required for methylation to take place, perhaps reducing some of the spurious signals from neighbouring but unbound regions. The protein being studied does not need to directly bind DNA, and DamID, much like ChIP, has also been successfully applied to Polycomb (Tolhuis et al., 2006). However, it still seems plausible that because the Dam methylase does need direct access to DNA, the signal might be more selective compared to a ChIP reaction where a range of proteins can be crosslinked to each other. However, in the absence of systematic studies, this is purely speculative, and to date, only a small number of appropriate ChIP-chip and DamID datasets have been available for comparison. The data were found to match well for GAGA factor and Polycomb group proteins (Negre et al., 2006; Tolhuis et al., 2006; Moorman et al., 2006), but a more in-depth analysis of ChIP and DamID differences would be of general interest.

E. coli Dam methylase is extremely processive (Urig et al., 2002), meaning that high levels of the Dam protein can result in substantial tracts of methylated genomic DNA and thus in poor resolution (van Steensel et al., 2001). To compensate for this, the Dam-protein fusion construct is generally regulated by an inducible promoter, such as UAS or Hsp70, with the construct used under uninduced conditions. The leaky expression from the uninduced promoter provides a suitably low quantity of enzyme to leave a specific methylation signal: induced expression results in rapid methylation of substantial portions of the genome (van Steensel and Henikoff, 2003).

The sensitivity of the enzyme may be an advantage; it has been suggested that the transient interactions not captured by ChIP may be identified with DamID (Wolffe and Leblanc, 2000). However, this has not been tested to date. On the other hand, the inability to drive the construct in a tissue-specific manner means that the signal might be unusual. The construct is expressed ubiquitously, in a range of cells and with a range of partner TFs that the target TF might be capable of interacting with, but may normally never encounter in wild type embryos: the effects of this on binding patterns are unknown. In addition, even under low level expression conditions the processivity of the enzyme means that methylated regions can be quite broad, possibly having a

negative impact on the resolution of the experiment. Another compounding factor is that the enzyme methylates the GATC sequence, which may not be evenly distributed across the genome, resulting in varying probabilities of methylation.

3.1.5 Controls

A control commonly used for ChIP-chip is an antibody that does not specifically interact with any specific *Drosophila* protein, such as β-galactosidase or IgG. This should pick up chromatin regions that are particularly prone to unspecifically binding to any antibody and provide a general picture of the stochastic noise inherent in an immunopurification. The importance of this control has been demonstrated in a number of studies (Johnson et al., 2008; Birch-Machin et al., 2005). The controls are also crucial because antibody specificity is generally not known - in a recent study of over 200 commercially available histone modification antibodies, it was found that more than 25% failed specificity tests by dot blot or western blot (Egelhofer et al., 2010). This highlights the importance of including specificity controls as part of ChIP experimental design. However, while the presence of this control undoubtedly improves data quality by removing a number of false positives, it does not remove the other previously discussed sources of noise inherent in ChIP data.

The control for DamID experiments is a Dam only protein not fused with a TF, which should leave methylation tags in areas that are likely to be unspecifically methylated, again eliminating some of the false positives. It is known that some parts of DNA will be more accessible than others and also that the fusion protein will sometimes bind to unspecific locations. However, it is thought that the methylase only control successfully cancels out these effects. It is unclear whether the methylation tags stay attached permanently after the DNA is tagged, or whether their occupancy can change more dynamically over time. This could have an impact on the signal but, in theory, as long as the developmental time periods are the same for the sample and the control, the control should still be effective. However, this is another area where a systematic study of DamID could prove helpful.

For both ChIP-chip and DamID, an equal amount of labelled DNA is used for both sample and control to make them comparable. This assumes that the quantity of DNA pulled down in an IP (or methylated, in the case of DamID) is similar for both samples and controls. In cases where, for example, the control has a lot less signal (a likely scenario for proteins which display large amounts of specific binding), whatever signal is present in the control will become saturated from amplification, and will cancel out any sample signal in those regions. While already an issue for some more active transcription factors, this is likely to be an even bigger problem in the case of histone

modification domains, which tend to display broader patterns (Liu et al., 2011). However, this is still the optimal method of matching sample and control DNA quantities. Since both the sample and the control are subject to PCR amplification, differences in amplification efficiency mean that combining the sample and control without matching the DNA quantities would produce an unknown amount of variability. Though potentially introducing false negatives due to signal saturation, the currently used technique is unlikely to result in an increase in false positives, as it is the control rather than the sample signal that is being potentially oversaturated. Ideally, a minimum of 3 replicates is used for both techniques, which allows for reliable statistical analysis and also means that dye swaps can be performed for two colour arrays, countering the different dye incorporation and detection efficiencies.

3.1.6 Experimental noise

Some of the potential sources of experimental noise for ChIP-chip and DamID have been discussed above. Human error can also have a significant impact on results; results are variable for the same experiments performed by different people or even by the same person at different times. It is therefore helpful in the case of Dichaete that several datasets are available, since while some experimental variability is expected, trends occurring consistently across datasets are likely to be genuine. Different tiling arrays have been used for different experiments, which could be another potential source of variability, though the best systematic study done to date suggests that this is not a major contributor to variability (Johnson et al., 2008). The variability in the data analysis output of different pipelines and from different starting analysis parameters is also an issue.

A difficult choice is therefore presented: either to use very stringent thresholds to eliminate false positives but generate a large number of false negatives, or to make the threshold less stringent, which introduces a number of spurious peaks. The advantage of using ChIP and DamID in combination is that the assays are fundamentally different. Therefore, while both the techniques are expected to show noise, it is also reasonable to expect that the noise profile will differ between techniques. There is no reason to suggest that the regions of chromatin that non-specifically interact with antibodies should be the exact same regions that get non-specifically methylated by the Dam fusion but not removed by the Dam only control. Therefore there is potential to use these complementary techniques in combination to identify genuine patterns of Dichaete action across the genome.

3.1.7 A systematic comparison

The situation of having multiple binding datasets available for a single protein is an interesting and fortunate one. First of all, it allows us to cross-validate, checking which binding regions are likely to be genuine and which ones are variable, and potentially a product of experimental noise. This makes multiple datasets helpful for extracting biological insights from data. The quality of the individual datasets as a whole can also be assessed by comparison with the other datasets, and by the presence or absence of expected patterns in different types of analysis. Furthermore, a systematic comparison of ChIP-chip from different labs in combination with DamID is valuable in assessing the utility of such mixed datasets in general, making an interesting technical comparison possible. In addition, results from the same analyses performed on developmental time windows of different lengths could be interesting for determining what type of ChIP-chip experimental design is most suited for future genomics studies. The main motivation for this chapter is to assess the quality of the available data, with the intention of figuring out the best way to use it for the purpose of further elucidating Dichaete biology. A variety of genomics analyses were therefore performed using all 4 datasets, partly for purposes of comparison and partly as a pilot screen to see which analysis paths are likely to be worth pursuing.

The existence of high throughput data generation initiatives such as modENCODE (Celniker et al., 2009) means that smaller labs and individual researchers worldwide have access to a large number of genome-wide datasets to use alongside the ones they generate themselves. In addition, a large number of genome-wide datasets from individual studies exist in the public domain, and this number is likely to greatly increase in years to come. Therefore the question of how to integrate and effectively use such datasets is a timely and relevant one.

3.1.8 Chapter overview

This chapter discusses the results derived from my Dichaete DamID experiment, comparing and contrasting it to 3 available Dichaete ChIP-chip datasets. For this analysis I discuss the properties of each dataset as a prelude to generating a high confidence set of Dichaete targets, which is discussed in the next chapter. I first focus on the genome-wide patterns of binding: whether the Dichaete binding sites are likely to hit genic or intergenic regions, introns or exons, and how far away they are from transcription start sites. The Dichaete sequence specificity is then studied in more detail, using motif finding and scanning tools. In the final section, Dichaete data is compared with some common genomic markers, the involvement of Dichaete in polymerase pausing is investigated, some potential coregulators are identified from enriched sequence motifs,

and the impact of chromatin accessibility on Dichaete binding is assessed.

3.2 The DamID experiment

When I first began my PhD, while some individual ChIP-PCRs had been done, no genome-wide binding datasets were available for Dichaete. Therefore, one of the initial aims of my project was to perform both ChIP-chip and DamID. However, multiple attempts at ChIP-chip yielded very noisy and unusable data due to the quality of the antibody. In the meantime, 3 separate ChIP-chip datasets became available for Dichaete, as generated by BDTNP and modENCODE (MacArthur et al., 2009; Celniker et al., 2009). Consequently, I instead focused on generating a DamID dataset, then using it together with the existing ChIP-chip datasets for an analysis of the role of Dichaete during early embryonic development.

3.2.1 Experimental design

For the DamID experiment, two genetic constructs were used. For the sample signal, *Escherichia coli* Dam methylase was attached to the Dichaete protein and for the control the Dam methylase on its own was included in the same vector. Both the vector constructs and the fly strains were created by Riaz (2009) as part of his PhD work in the Russell lab. The pUAST vector was used in both cases (Phelps and Brand, 1998), putting the transgenes under the control of a UAS promoter. A schematic of the constructs used is shown in Figure 3.1. A *Drosophila melanogaster* full-length *kayak* 3-untranslated region (3-UTR) was incorporated into the transformation constructs with the hope that its destabilizing influence will reduce the half-life of the transcripts and accordingly the total quantity of methylase present. The motivation for this was that reducing the quantity of methylase present at any one time should ameliorate the effects of Dam processivity, potentially improving the data resolution and specificity.

The fly strains UAS-D-ED-3, with the Dichaete-Dam fusion inserted on chromosome *3*, and UAS-ED-2a, with the Dam methylase control construct inserted on chromosome *2*, were used. Since the flies were homozygous for the construct, no crosses were required, and instead embryos were collected directly from the fly strains in question.. Embryos between 0 - 12 hours old (stage 1-15) were collected for the experiment. This was considered to be an interesting developmental period, covering several periods during which Dichaete is thought to be developmentally active, such as segmentation and embryonic nervous system development. It also complemented well the existing ChIP-chip datasets, all of which were generated from embryos at ages somewhere within this developmental window. DNA was extracted and specific restriction digests were used

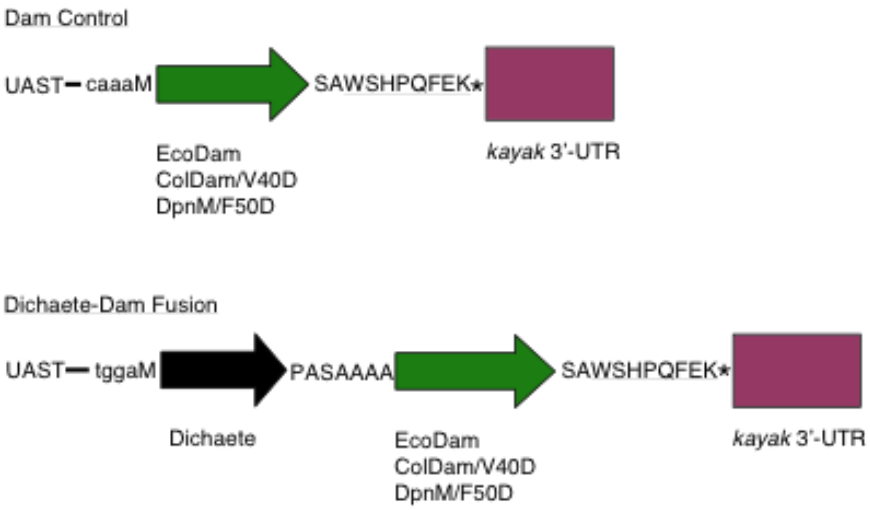

Figure 3.1: Schematic of constructs used in the DamID experiment, figure from Riaz (2009).

to enrich the methylated fragments of the genome as described in Chapter 2. The DNA was then amplified, labeled with fluorescent Cy3 and Cy5 dyes, and hybridised to NimbleGen *Drosophila melanogaster* ChIP-chip 2.1M Whole-Genome tiling arrays. The data was processed using NimbleScan and the TiMAT pipeline, using quantile normalization separately on the samples and the controls. The data was quality checked using a variety of diagnostic plots (R scripts written by Bettina Fischer).

3.2.2 Results

The raw data scatterplot (Figure 3.2) shows that the 3 Dichaete biological replicates are very similar to one another and that the 3 methylase only control replicates are also very similar to one another, which is encouraging. It also shows that the Dichaete samples are very different from controls, which is as expected: an active transcription factor leaving a specific trail in the genome should look quite different from unspecific methylation.

As shown in the density plot in Figure 3.3, the sample and control have quite distinct data distributions. The motivation for using normalization that deals with the sample and the control signals separately, is that one that normalizes them together presents a danger of normalizing away too many of the genuine differences between the two signals: as seen in the bottom part of Figure 3.3, the method used normalizes the two signals, while keeping their distinct distributions intact.

The number of intervals found (Table 3.1) is comparable to the Dichaete data from other studies (MacArthur et al., 2009), confirming the view that Dichaete binds to

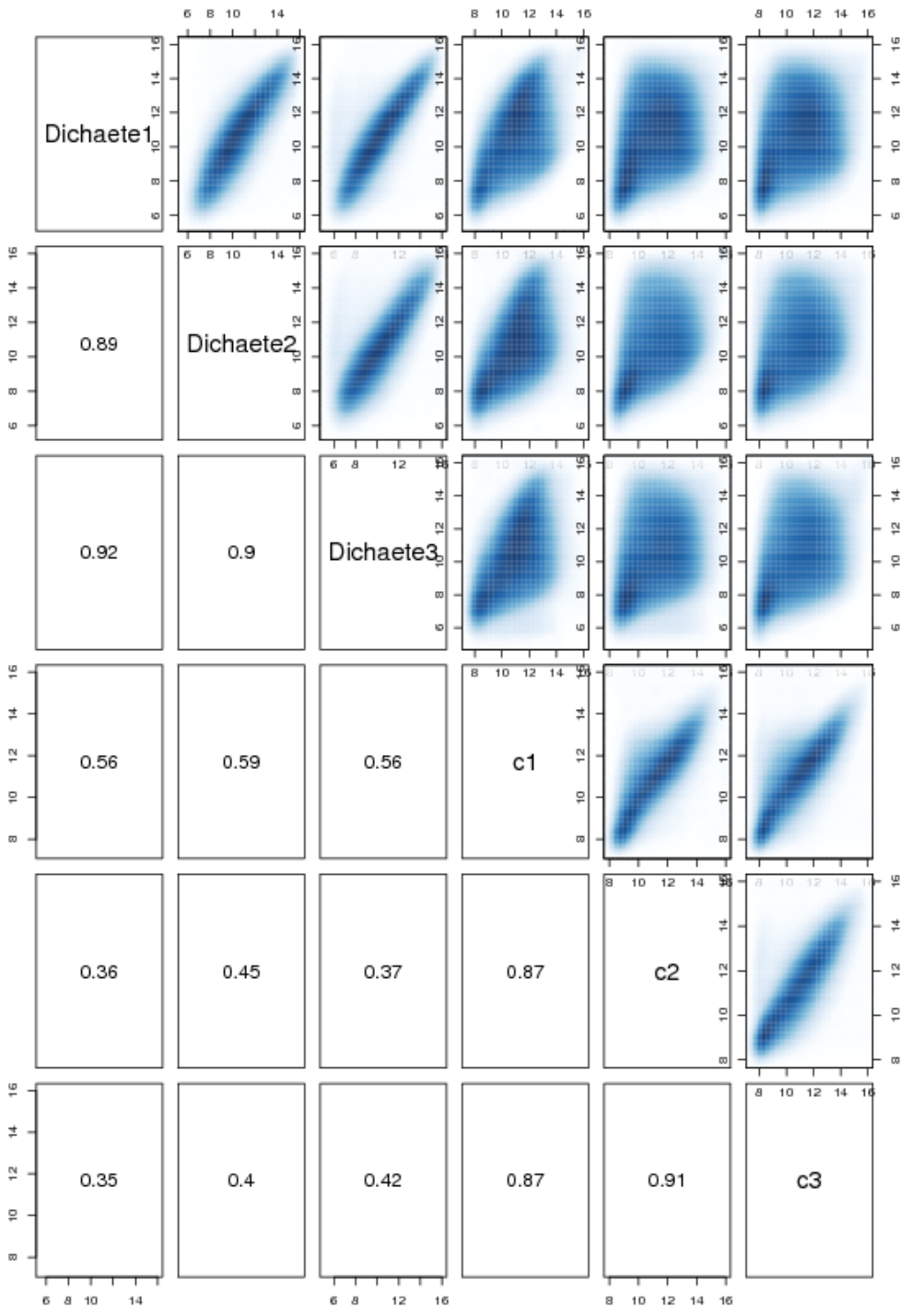

Figure 3.2: Dichaete DamID raw data scatterplot

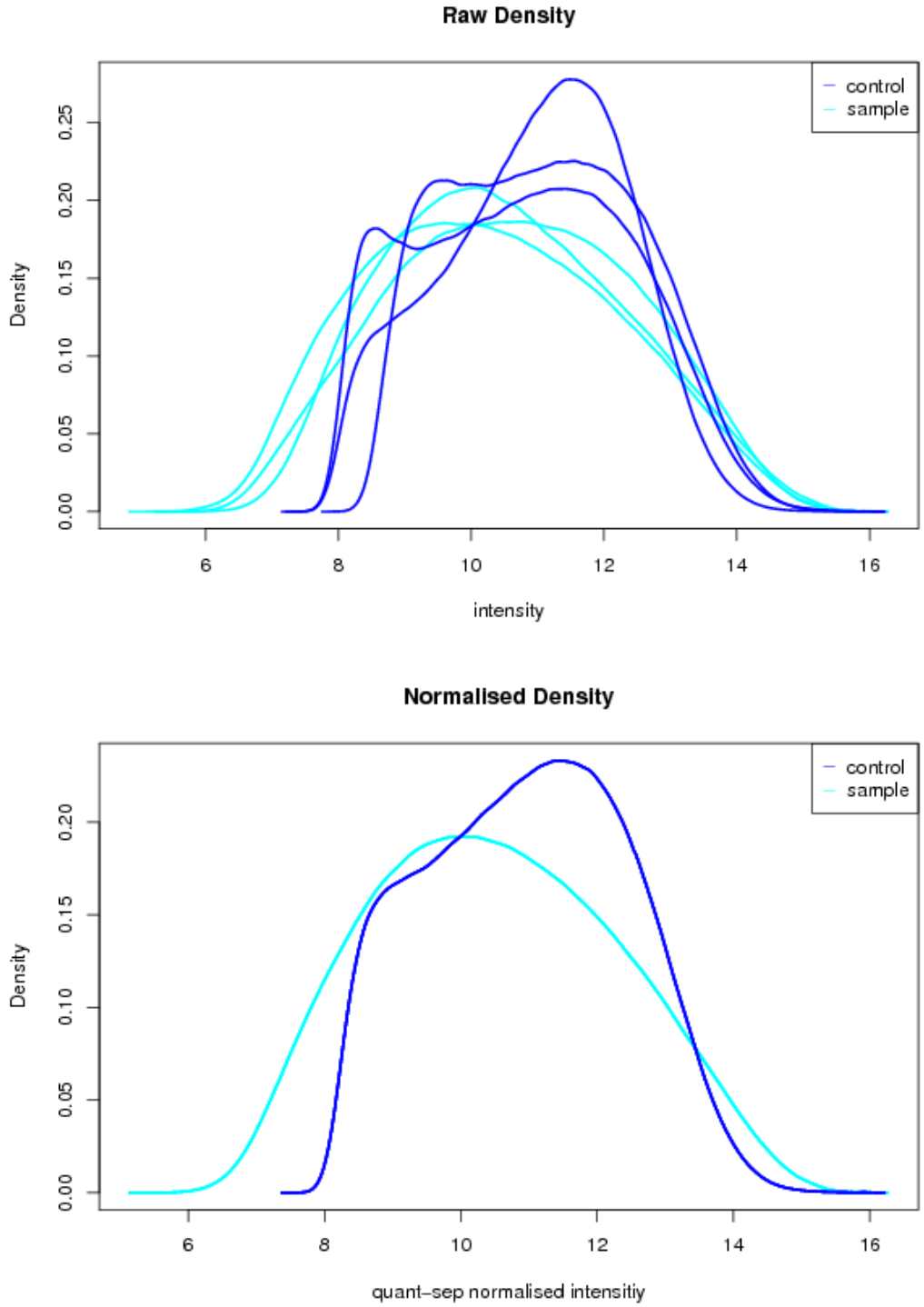

Figure 3.3: Dichaete DamID raw (above) and normalized (below) data density plot. Data from all 3 biological replicates is shown separately in the raw data plot. During normalization, the data from the different replicates is combined, with the data from the 3 samples treated together, and separately from the 3 controls. Thus the bottom chart shows two lines - one for the combined 'sample' data, the other for the combined 'control' data.

thousands of locations across the genome. The comparison of the numbers and lengths of intervals found in different Dichaete binding datasets is presented in detail in Section 3.3 of this chapter (Table 3.3 gives an overview of the figures). Encouragingly, the resolution of the DamID experiment seems to be excellent and very much comparable to ChIP-chip studies that use similar platforms, and shows an improvement to the resolution of comparable DamID studies performed in the past. For comparison, a whole embryo DamID experiment performed by Choksi et al. (2006) using a Prospero-Dam fusion was reanalysed using the same analysis settings that were used for Dichaete DamID, and the average interval length was found to be 2606 bp for FDR 1% and 3279 bp for FDR 25% - a resolution much lower than the Dichaete DamID one. The comparative improvement in resolution may be due to the addition of the destabilising *kayak* 3'-UTR region, or by the use of an improved tiling array platform (the Prospero paper uses a custom-made tiling array, with a 60-mer every 300 bp). Further analysis of the Dichaete DamID dataset will be presented later in the chapter, alongside the analysis of the Dichaete ChIP-chip datasets.

False Discovery Rate (%)	**Number of intervals**	**Average interval length**
1%	4342	1688 bp
5%	5989	1835 bp
10%	7437	1940 bp
25%	12301	2359 bp

Table 3.1: Dichaete DamID dataset overview

3.3 Integration of Dichaete binding datasets

3.3.1 Data used

The following data (shown in Table 3.2) was used for further analysis. In the case of Dichaete, 3 different ChIP-chip datasets have already been published. Dichaete ChIP-chip data was generated from 2-3h old embryos by the Berkeley *Drosophila* Transcription Network Project (MacArthur et al., 2009), and with chromatin from 0-8h and 0-12h old embryos by the modENCODE project (Celniker et al., 2009). The Dichaete DamID data generated using 0-12h old embryos provides a fourth dataset.

3.3.2 Processing methods

The Dichaete DamID (D_Dam) was run on NimbleGen arrays, processed first using NimbleScan (version 2.4) for spot-finding, then the TiMAT pipeline for peak finding,

using quant2 normalization, a 675 bp scanning window, min 5 probes and a 200 bp maximum gap. The BDTNP Dichaete ChIP data (D_Berk), generated using Affymetrix arrays, was downloaded from the BDTNP website - the intervals at FDR 1% and FDR 25% found using symmetric-null test normalization were used. The analysis method is described at length in the associated publication (MacArthur et al., 2009). For the modENCODE data (D08 and D12), also generated using Affymetrix arrays, the original plan was to use the intervals as processed by the modENCODE consortium. However, for each of the two datasets, while the raw data was available, only one interval file existed, with ambiguous labelling (it was unclear how the intervals were obtained, and whether the threshold was FDR 1% or FDR 5%). The modENCODE data was therefore reanalysed, in order to obtain binding intervals with more certainty regarding the data processing method and the contents of the interval files. Starting from the raw data (.CEL files), the modENCODE datasets were reanalysed using the TiMAT pipeline, using default window size of 675 bp, number of oligos in a valid window of 10, and maximum gap of 200 bp. The normalization method used was quant (quantile normalization across samples and controls together). The BDTNP data was not included in the reanalysis, and was instead used as processed in the original publication. The reason for this was that the data processing method used for the dataset in question was robust, well tested and well documented, so the resulting final data was deemed trustworthy enough to processed with as is. In contrast, the modENCODE data processing was badly documented, so the reanalysis was performed in order to know with certainty what methods and thresholds were used, before basing extensive downstream analysis around the processed data.

3.3.3 Parameter sensitivity analysis

For the modENCODE data, the default TiMAT settings were used: quant normalization, a 675 bp scanning window, min 10 probes (because Affymetrix rather than NimbleGen arrays were used, there is a larger number of shorter probes present) and

Dataset	Name	Method	Development time	Source
D_ChIP_2-3h	D_Berk	ChIP-chip	2-3h embryo	BDTNP (MacArthur et al., 2009)
D_DamID_0-12h	D_Dam	DamID	0-12h embryo	My data
D_ChIP_0-8h	D08	ChIP-chip	0-8h embryo	modENCODE (Celniker et al., 2009)
D_ChIP_0-12h	D12	ChIP-chip	0-12h embryo	modENCODE (Celniker et al., 2009)

Table 3.2: Dichaete datasets used for analysis

a 200 bp maximum gap. It was then thought that analysing all the datasets using the same pipeline and the same settings would be helpful in terms of making the different datasets more comparable. However, reanalysing the DamID data using quant instead of quant2 normalization unexpectedly resulted in 0 instead of the 4000 - 12000 binding intervals previously identified. The way quantile normalization works in general, is by ordering the intensity values on each array from lowest to highest, then changing the values at each rank position into an average for that rank (Russell et al., 2009). If done across all arrays, this means that the intensity distribution of all the arrays becomes identical. The difference between quant and quant2 is that quant applies the normalization across all arrays, whereas quant2 applies quantile normalization to samples and controls independently. If the sample and control signals are expected to have a similar distribution, it makes sense to normalize across all arrays. However, this is not the case with the DamID data, so quant normalization appears to normalize away much of the genuine signal. The DamID data shows high similarity between biological replicates, so it was not thought that the data quality was in question. Also, quant2 normalized data finds binding intervals with a high overlap with other Dichaete datasets and an overrepresentation of genes considered to be likely Dichaete targets. It was therefore concluded that the normalization method chosen for a particular data set is dependent upon the characteristics of the data and it is prudent to explore different methods with each data set generated.

Out of the 3 datasets reanalysed, the dramatic difference in intervals found due to alternative normalization methods was only observed with the DamID data. However, other parameters were also found to affect the number of intervals found by the pipeline. The maximum gap parameter is normally set to the default 200 bp, and it was found that varying it only results in minor fluctuations to the number of intervals found. The same was found for the minimum number of oligos, with the results of these analyses shown in Figure 3.4.

On the other hand, the size of the sliding window was found to have a large effect on the number of intervals found (as shown in Figure 3.5), most likely because using a small window results in the identification of a large number of small clustered intervals, whereas a large window registers these clusters as a single large interval.

Unfortunately, while it is clear that the initial TiMAT parameters do influence the result of the analysis, determining which parameters give the more genuine signal is not possible due to the lack of further experimental data about verified Dichaete binding sites. In addition, as previously discussed, datasets with different distributions may require different analysis tools and parameters to extract the best results. Therefore a single set of optimal settings for all datasets is unlikely to exist, and in any event it may be hard to tell which results should be considered the best. As a pragmatic solution,

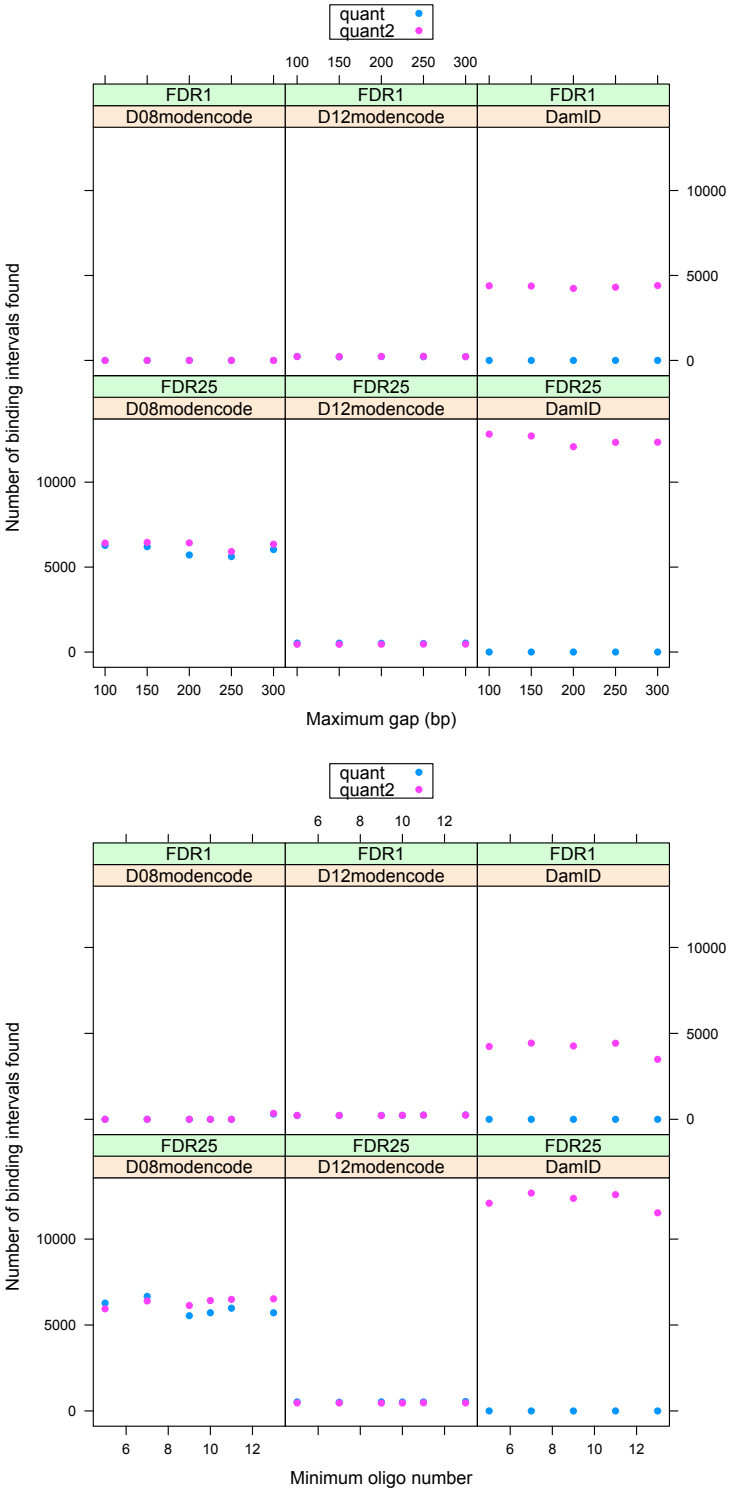

Figure 3.4: Influence of TiMAT maximum gap and minimum oligo parameters on the number of binding intervals found

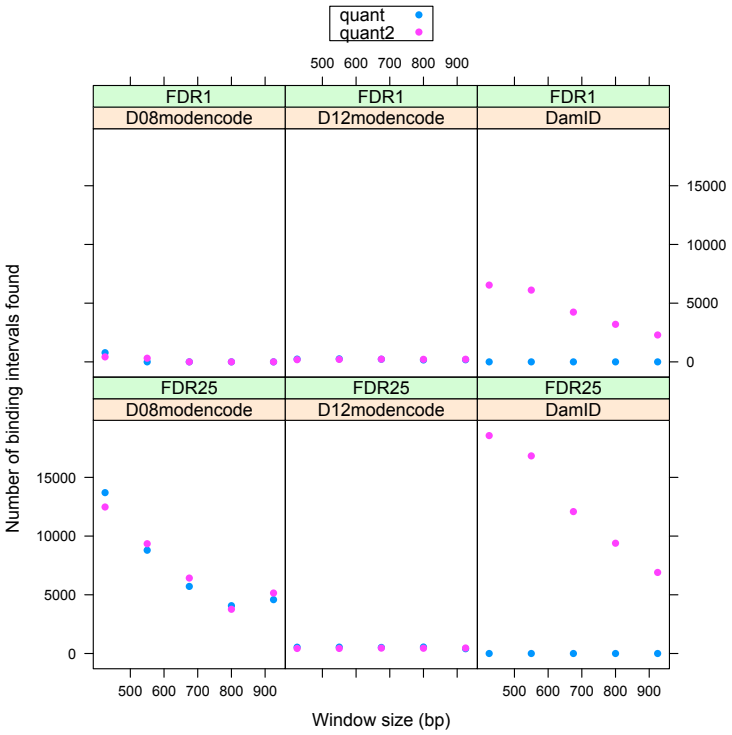

Figure 3.5: Influence of TiMAT sliding window size on the number of binding intervals found

the default settings were used wherever possible, adjusting the minimum oligo number according to the microarray platform (10 for Affymetrix, 5 for NimbleGen), and using quant2 instead of quant for DamID.

3.3.4 Processed data overview

The number of intervals identified in each dataset is shown below in Table 3.3. The high stringency intervals were found at FDR 1% for all datasets except D08, where FDR 5% data was used because no peaks were found at FDR 1%. The low stringency results for all datasets were found at FDR 25%. The high numbers of peaks found in D_Dam and D_Berk datasets, while striking, are comparable to other developmentally important transcription factors such as *Krüppel* and *hunchback* (Li et al., 2008). The modENCODE datasets (D08 and D12) give fewer peaks than the DamID and the Berkeley data, but this is thought to be due to lower signal strength (perhaps due to lower quality antibodies) rather than an abundance of false positives in the other datasets (as the high overlap between D_Dam and D_Berk suggests that they both give a reliable picture of Dichaete binding).

dataset	FDR	Number of intervals	Average interval length (bp)	% of genome covered by binding intervals
D_Berk	1	6452	1885	10.1
D_Dam	1	4342	1688	6.1
D08	5	1366	1002	1.1
D12	1	208	1025	0.2
D_Berk	25	16501	1989	27.4
D_Dam	25	12301	2359	24.2
D08	25	6493	1225	6.6
D12	25	538	1160	0.5

Table 3.3: Dichaete binding intervals

Interval lengths

The interval lengths for the 4 available Dichaete datasets at different FDRs are shown in Table 3.3. While there are reasons described above why DamID may generate lower resolution data than ChIP-chip, in this instance this does not appear to be the case at more stringent thresholds. Although the DamID does give the broadest intervals at FDR 25%, they are comparable to the length of the Berkeley ChIP-chip data at the same FDR (with differences possibly due to the use of different arrays/analysis pipelines, rather than a major difference in the resolution of the techniques). At the FDR 1% threshold, the resolution of DamID even appears to be slightly higher than that of Berkeley ChIP (though, again, differences could be due to arrays/analysis pipelines and a direct comparison is difficult because of potential differences in binding patterns at different developmental stages). The modENCODE datasets have a much narrower interval length than the Berkeley and DamID datasets, but this may be due to the fact that the signal is generally weaker and thus the peaks appear narrower.

Overlap

Because of the slightly different stages of development and different methods used, some difference in binding intervals identified is expected. However, there should still be a significant overlap between the datasets if they represent bona fide Dichaete binding locations, and this was indeed found to be the case. The resampling-based significance test implemented in the Cooccur package (Huen and Russell, 2010) was used to determine the significance of overlap between the different Dichaete datasets. All the datasets with significant overlap (p-value < 0.05) are shown as connected in Figure 3.6. All the datasets with comparable numbers of intervals were found to significantly overlap. The D12 dataset has much fewer intervals compared to all other datasets, so a possible reason for the absence of significant overlap with the other datasets is simply a lack of statistical power. A look at the actual overlap numbers

shows that between 77% and 85% of D12 overlap with D_Berk, which confirms that the lack of statistical significance is caused by the small number of intervals rather than a lack of overlap. This seems to indicate that the D12 dataset is likely to be specific to Dichaete, but just has a signal so weak that most Dichaete sites are not registered. Nevertheless, all 4 datasets were used for further analysis, with caution applied when interpreting the results from the D12 dataset.

3.4 Genome-wide view of Dichaete binding

3.4.1 Distance from transcription start sites

The 5' end of genes, as annotated in FlyBase (Gelbart et al., 1999), was used as an approximation of where the transcription start site is expected to be. The middle of the highest scoring window from each interval was used as the binding site coordinate. The analysis was only done for the high stringency intervals (FDR 1% for all datasets except D08, for which FDR 5% was used), as the low stringency intervals were thought to introduce too many false positives. On the graph, upstream locations are shown as negative values and downstream locations are shown as positive values (Figure 3.7).

The distance of the peaks from the closest transcription start site found within 5 kb is shown in Figure 3.7. Since histograms can be misleading depending on the number of bins used, the density curve is also shown. All 4 datasets show a similar trend, with peaks found close (within 500 bp) to transcription start sites, but the trend is more pronounced in the D_Berk and D08 datasets. None of the datasets are normally distributed (p-value between $9E{-}5$ and $2.2E{-}16$ for all datasets, Shapiro-Wilk test), so non-parametric tests were used to compare their similarity. First, the Kruskal-Wallis test was performed to find out whether the data distributions are the same for all datasets. This was found to not be the case (p-value $<$ 2.2E-16). The Mann-Whitney U test was then used for pairwise comparisons of the datasets. The datasets that were found to be significantly different from one another were D_Berk and D12 (p-value = 0.04), and D08 and D12 (p-value = 0.01), with the D_Dam and D12 having the borderline p-value of 0.06. This matches the previous conclusion that the D12 dataset is significantly different from the other datasets, and may be unreliable. However, after correction for multiple testing (Benjamini-Hochberg), none of the p-values are below the 0.05 cutoff threshold, so the results were in the end not found to be significantly different for different datasets.

The mean distances from the transcription start site for the D_Berk, D_Dam and D08 site were between 100 and 150 bp, suggesting that Dichaete frequently binds slightly

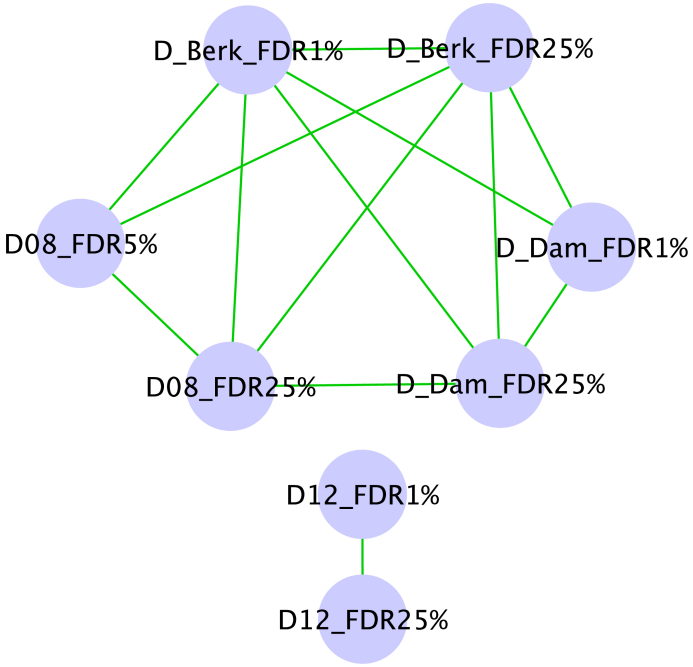

Figure 3.6: Significance of overlaps of Dichaete datasets. Datasets significantly overlapping (p-value < 0.05) according to a resampling based test are shown as connected.

Dataset 1	Dataset 2	Dataset 1 % overlap	Dataset 2 % overlap
D_Berk FDR 1%	D_Dam FDR 1%	38%	60%
D_Berk FDR 25%	D_Dam FDR 25%	50%	64%
D_Berk FDR 1%	D08 FDR 5%	17%	86%
D_Berk FDR 25%	D08 FDR 25%	26%	80%
D_Berk FDR 1%	D12 FDR 1%	2%	77%
D_Berk FDR 25%	D12 FDR 25%	3%	85%
D_Dam FDR 1%	D08 FDR 5%	15%	49%
D_Dam FDR 25%	D08 FDR 25%	31%	72%
D_Dam FDR 1%	D12 FDR 1%	2%	47%
D_Dam FDR 25%	D12 FDR 25%	3%	72%
D08 FDR 5%	D12 FDR 1%	4%	27%
D08 FDR 25%	D12 FDR 25%	5%	54%

Table 3.4: Interval overlap for different Dichaete datasets.

downstream of the transcription start site. The median values were even smaller: 69, 29 and 73 bp, respectively. In the case of the D12 data, the mean was 406 bp, and the median 268.

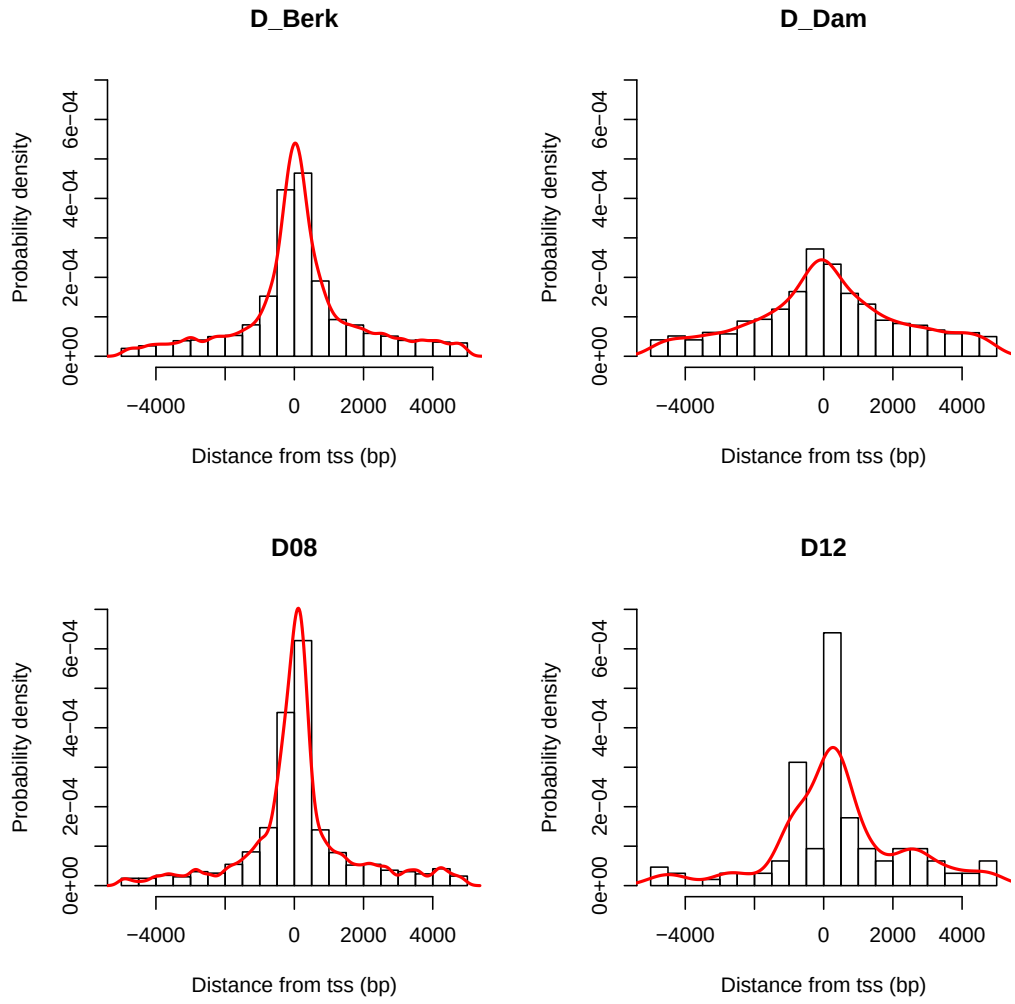

Figure 3.7: Probability distribution of the distance to the closest transcription start site

It is possible that, because the *Drosophila* genome is quite dense and gene rich, the distance to the closest transcription start site would be short by chance. The same analysis was therefore rerun to look at the distances from all transcription start sites present within the 5 kb region (Figure 3.8). The graphs flatten out somewhat but there still appears to be a trend with Dichaete binding close to the transcription start site, particularly visible in the case of the D_Berk and D08 data. The mean distance was between 100 and 200 bp for the D_Berk, D_Dam and D08 data sets, and 350 bp for the D12 dataset, with the median values being slightly smaller (between 110 and 160 bp for D_Berk, D_Dam and D08 data and 300 bp for D12).

The traditional model of transcription factor binding expects the TF to bind to a cis-

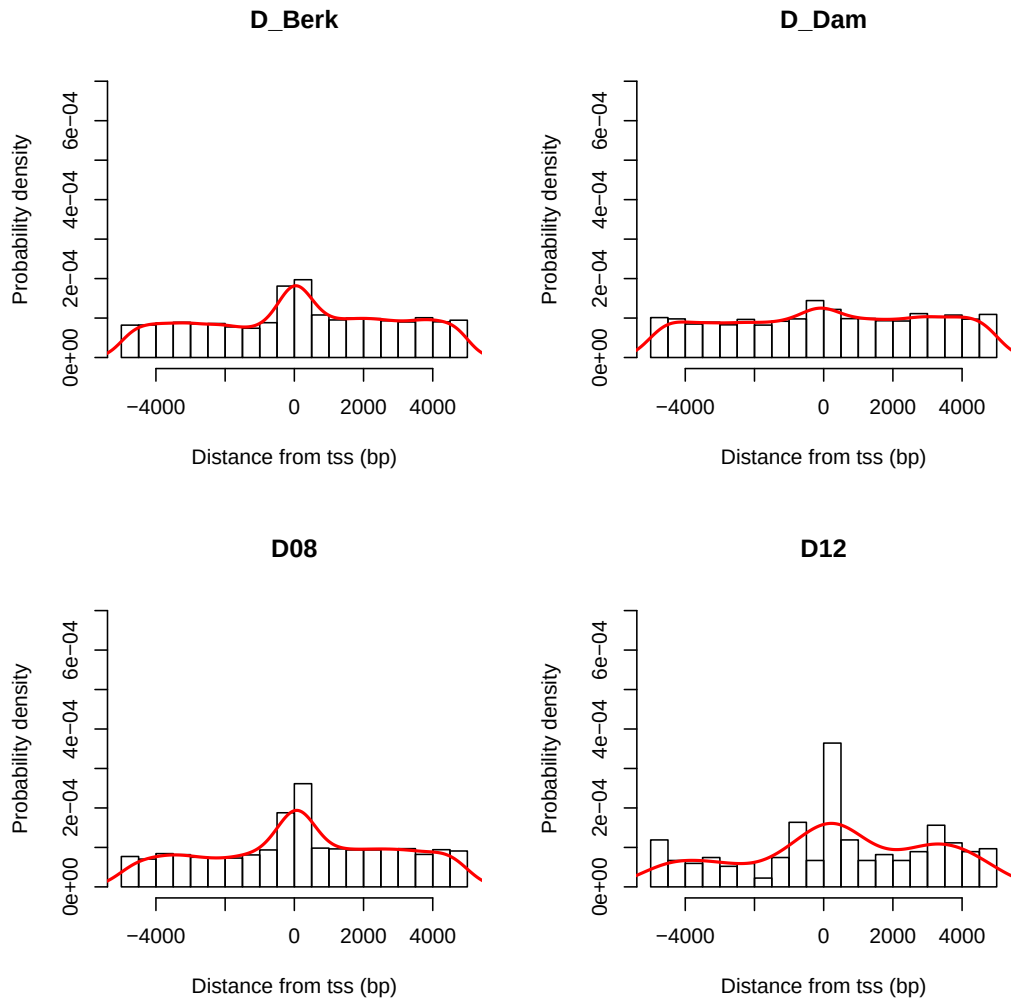

Figure 3.8: Probability distibution of the distance of Dichaete peaks from all transcription start sites within a 5 kb region

regulatory module upstream of the gene. The apparent tendency of Dichaete to bind slightly downstream of the transcription start site may suggest a different regulatory role, for example forming loops to bring other regulatory elements together, affecting chromatin accessibility or facilitating RNA polymerase pausing. In the case of binding very close downstream of the transcription start site, this may help keep the polymerase poised, or might be targeting a genic enhancer element.

3.4.2 Gene hits

For each of the intervals, all directly overlapping genes were counted as gene hits. If no direct overlaps were found for an interval, the closest gene within a 10 kb region upstream or downstream was found and registered as a gene hit instead. The complete list of *Drosophila melanogaster* genes (genome annotation version 5.26, downloaded

from FlyMine) was used and the FBgn numbers were used as identifiers. It has been observed that changes in gene names and genome annotation can lead to statistical errors, which is why the most up to date version and the most stable identifiers were selected.

dataset	% FDR	Total number of intervals	Number of genes hit
D_Berk	1	6452	6258
D_Dam	1	4342	2872
D08	5	1366	1529
D12	1	208	206
D_Berk	25	16501	10726
D_Dam	25	12301	8271
D08	25	6493	5517
D12	25	538	498

Table 3.5: Numbers of genes hit by Dichaete bound intervals

Further analysis was restricted to the high stringency intervals (FDR 5% for D_ChIP_0-8h, FDR 1% for all the other ones) in order to minimise the number of false positives: with the FDR 25% intervals, particularly in the case of the Berkeley data, the majority of the genes in the *Drosophila* genome are hit, and while this is likely to capture most Dichaete targets, it is also likely to have too many false positives to be informative.

The gene list overlap for the high stringency datasets is shown in Figure 3.9 below (graphic generated using the Venny tool. Oliveros, 2007). In total, 6918 genes were flagged in at least 1 of the 4 datasets. 3026 genes were hit by intervals from 2 or more datasets, 829 by 3 or more, and 92 by all 4 datasets.

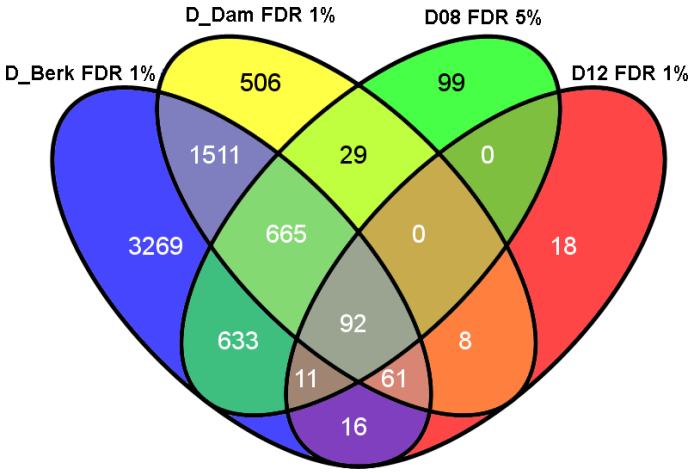

Figure 3.9: Gene list overlap

While the list of genes hit by all 4 datasets is obviously too stringent, it nonetheless gives an interesting list of candidate targets. Out of the 92 genes, 25 are annotated

as putative transcription factors in FlyTF (Adryan and Teichmann, 2006) (p-value <2.2E-16, Pearson's χ^2 test) and 30 are involved in nervous system development (p-value = 1.1E-10, as calculated by FlyMine, after Benjamini-Hochberg correction for multiple testing).

It is likely that the list of genes hit by all 4 datasets is far too stringent, and that the list of genes hit by only 1 dataset contains false positives. Nonetheless, 149 of the same GO terms were found to be statistically significantly enriched (p-value < 0.05 after correction with Holm-Bonferroni, FlyMine) in all 4 individual gene lists. The details of the GO enrichments found are shown in Table 1, Appendix 1.

In order to provide a more simplified and general overview of the processes in which Dichaete is involved, the 4 gene lists were evaluated against the subset of GO terms provided by the Panther database (Mi et al., 2005) GO slim list using Ontologizer (Bauer et al., 2008) for analysis and visualisation. Resulting p-values were corrected for multiple testing using Benjamini-Hochberg. The GO enrichments for genes hit by at least 3 of the 4 data sets are shown in Figure 3.10.

The overrepresentation of genes involved in transcriptional regulation is expected, given the high number of transcription factors hit by all of the Dichaete datasets. The overrepresentation of genes involved in developmental processes is also expected, given the known *Dichaete* mutant phenotypes. Genes for localization and cell communication could be some of the targets through which Dichaete exerts control during development. The presence of apoptosis and cell cycle regulation is particularly interesting given the role Sox2 plays in mammals to maintain stem cell pluripotency. It appears that Dichaete may perform a similar function at some point during fly development.

It has been reported that differences in gene length and the genomic spacing around them result in biases in gene assignments derived from binding data, featuring for example more transcription factor and nervous system genes, because these are on average longer and easier to hit (Taher and Ovcharenko, 2009). Since these are exactly the processes identified in this analysis, this means that the p-values presented here are certainly overinflated. However, it is still certain that these are high confidence gene hits, as they are identified a number of times in independent Dichaete datasets. Further to that, they seem to constitute a reasonable proportion of the total hits of the same stringency - about a third of the total list are nervous system development genes, and about a quarter are transcription factors. However, it is important to note that because of the gene length bias, the relative importance of transcriptional regulation and nervous system development compared to other processes that Dichaete is taking part in is likely overestimated.

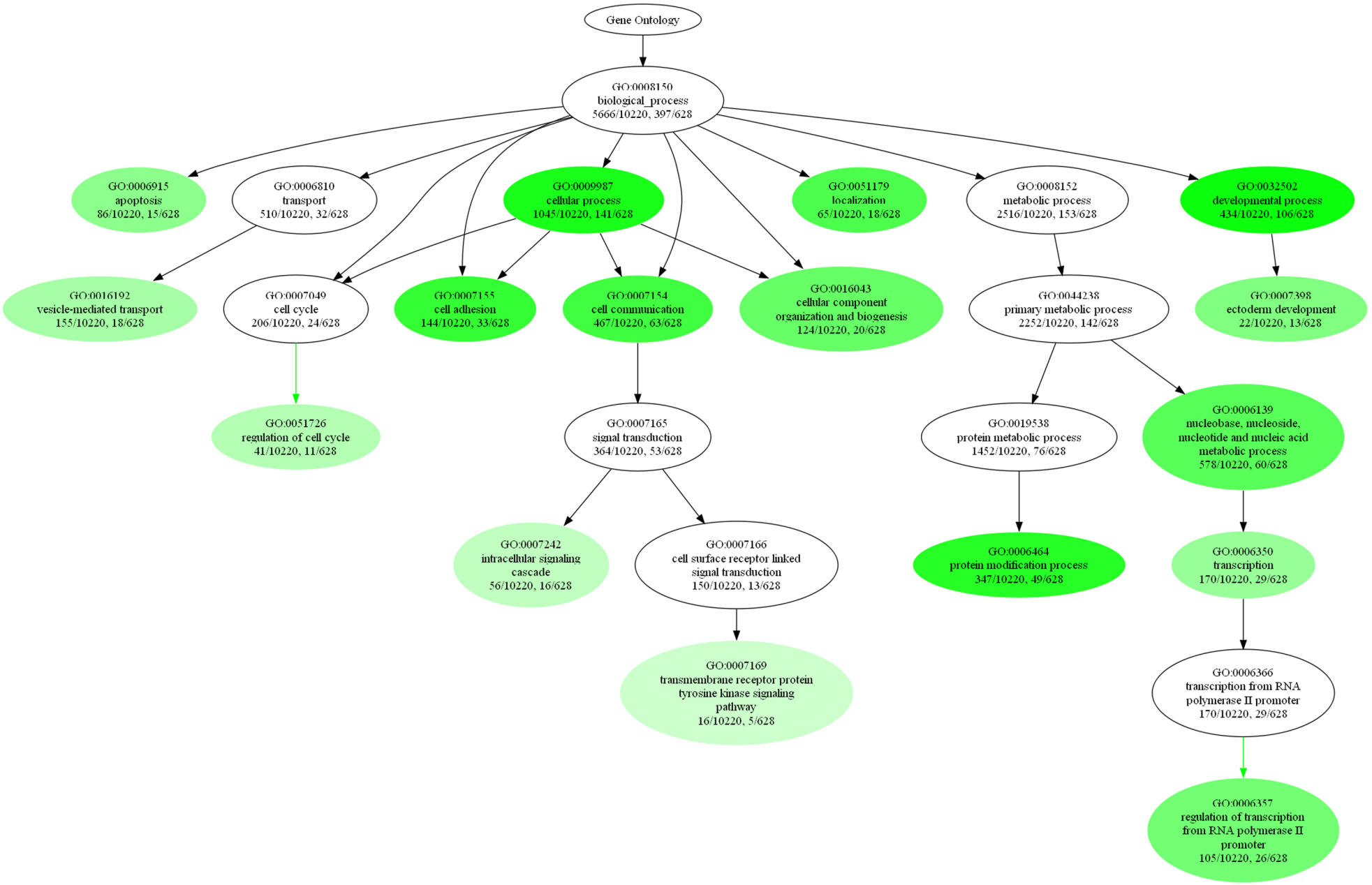

Figure 3.10: GO enrichment for genes flagged by at least 3 Dichaete binding datasets: all terms in green are significant (p-value < 0.05), with the brighter green colour corresponding to an increase in significance

3.4.3 Genomic features hit

The mean length of the regions of enrichment in each dataset was, even at stringent FDR values, still between 1 and 2 kb, which means that the span of each interval tends to encompass a range of genic and intergenic regions. Therefore for this analysis individual nucleotide positions were used instead of the whole interval. In order to use a point that is more likely to be representative of Dichaete binding site locations, the middle of the highest scoring window in each interval was used.

It was found that Dichaete binding is mostly associated with genic regions (Figure 3.11). Of the peaks mapping to genic regions, the majority were located within introns. While the traditional transcription factor binding location is just upstream of genes, it is possible that Dichaete binding in the middle of introns also has a regulatory role. A further two pieces of analysis were conducted: one to determine whether there is a link between the quantitative information from ChIP-chip experiments and the location of binding, as suggested by BDTNP (MacArthur et al., 2009), and the second to determine whether different Dichaete binding profiles clearly correspond to different classes of genes.

3.4.4 Quantitative binding information

MacArthur et al. (2009) report that high scoring peaks correlate with intergenic and intronic hits, which they claim are more likely to be functional regulatory sites. An analysis was performed to verify that the reported trend exists for the D_Berk data, and to see whether a similar phenomenon can be observed for the other available datasets. For the Berkeley data, the scores for each symmetric null analysed primary peak were used for ranking. For all the other datasets, the score of the highest ranking 675 bp window from each interval was used. The points were used in rank order to determine which genomic features were hit, with the highest scoring intervals having the lowest rank numbers. To determine whether trends for binding to particular genomic features differ based on rank, probability densities were plotted for all 4 datasets, using a window of 100 rank places for smoothing. Only features that could be unambiguously mapped to one 'type' were used for the analysis.

For the D_Berk, D_Dam and D08, the datasets do indeed show the reported trend - the higher ranking intervals are more likely to land in intergenic regions (Figure 3.12). There also appears to be a prevalence of intervals bound to genic regions in 'middle-ranking' (rank 1500-2500) peaks for D_Dam. The D12 dataset does not display the same trend, possibly because the small number of peaks and general weakness of the signal mean that the quantitative peak data is not as informative as it is for the other datasets. To support the conclusions drawn from the graphs, Wilcoxon tests were

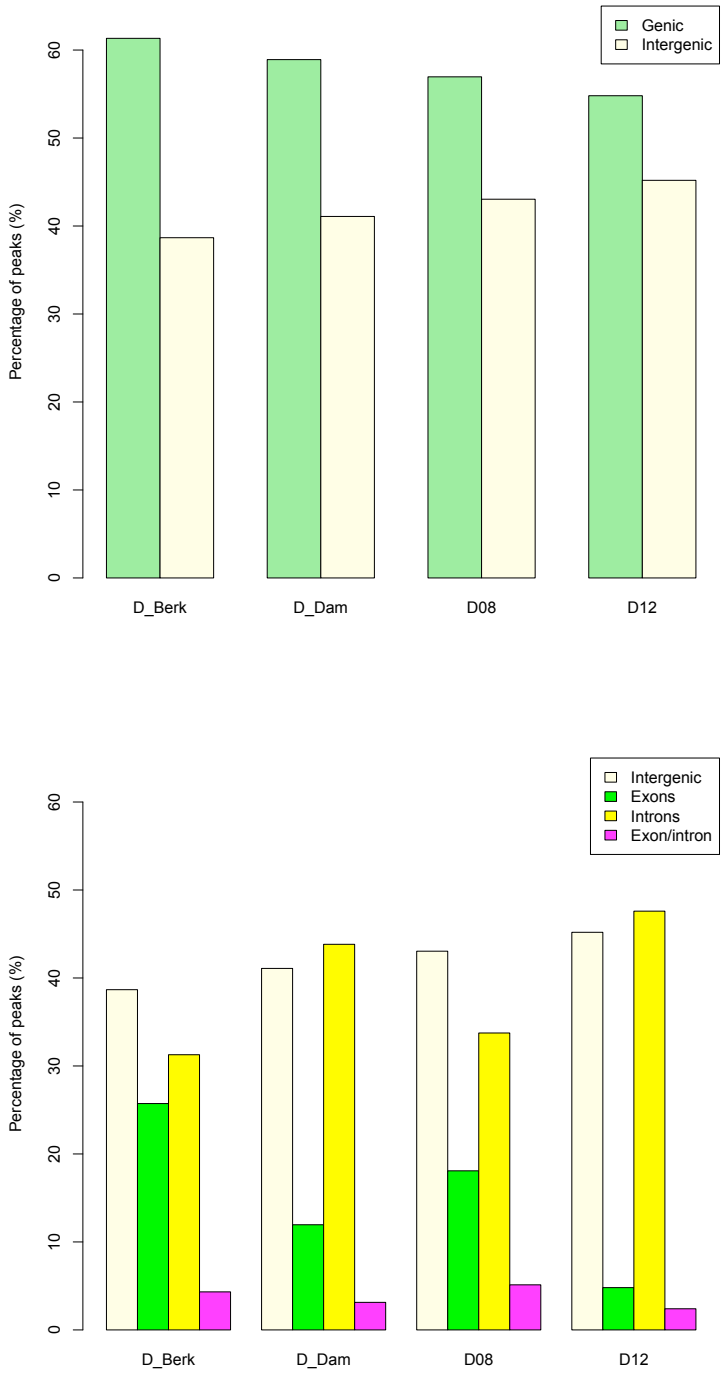

Figure 3.11: Genomic features hit by different Dichaete datasets

performed to verify whether there is a significant difference between the distributions of peak ranks for genic and intergenic regions, and the p-values were corrected for multiple testing using an estimate of FDR. The results support the conclusions drawn from the graph patterns - D_Berk, D_Dam and D08 all show significant differences between the rank distributions for genic and intergenic regions (with corrected p-values of 0.004, 0.004 and 0.01, respectively). These results mean that there may be some form of link between quantitative ChIP information and in vivo binding patterns.

The Berkeley data does display the reported trend of more highly ranked peaks binding in introns and intergenic regions, and the lower-ranking peaks binding to exons, as does D_Dam, and to a much smaller extent D08 (Figure 3.13). Interestingly, while Figure 3.12 shows significant differences between the ranks of peaks hitting genic and intergenic regions, Figure 3.13 suggests that most of the difference is due to the exon hits having much lower scores, and contributing to the lower rank of genic hits as a whole category, while the difference between the intron and intergenic hits does not appear significant. Statistical tests verify the latter observation - the only dataset showing a significant difference between intron and intergenic hits is D_Berk (Wilcoxon test, Bonferroni adjusted p-value = 0.0002). In the case of D_Berk, the trend is interesting - while the very top ranked intervals are intergenic hits, a number of the slighty lower (but still high scoring) intervals appear to be intronic hits.

What these trends mean is unclear. Assuming that exon hits are unlikely to be functional sites (an assumption that may or may not be valid), the fact that most exon hits come from low scoring peaks perhaps suggests that higher scoring peaks are more reliable 'hits', and the lower scoring peaks are false positives. If the peaks in question are not in fact false positives, then it is possible that they correspond to sites that are not of functional significance and are only occasionally bound by Dichaete when they are accessible, while higher scoring peaks correspond to more strongly preferred sites that are more frequently bound in different cells and at different developmental times. However, while we could potentially filter out non-functional hits based on low scores, it is unclear that this strategy is based on a solid premise. In terms of the difference between intronic and intergenic binding - this difference for the most part seems to be non-existent. It is clearly a significant difference in the D_Berk dataset, and this could hypothetically correspond to different profiles of binding sites generating different ranges interval scores, with the D_Berk dataset perhaps picking up on these because of the tight developmental time window used for the experiment.

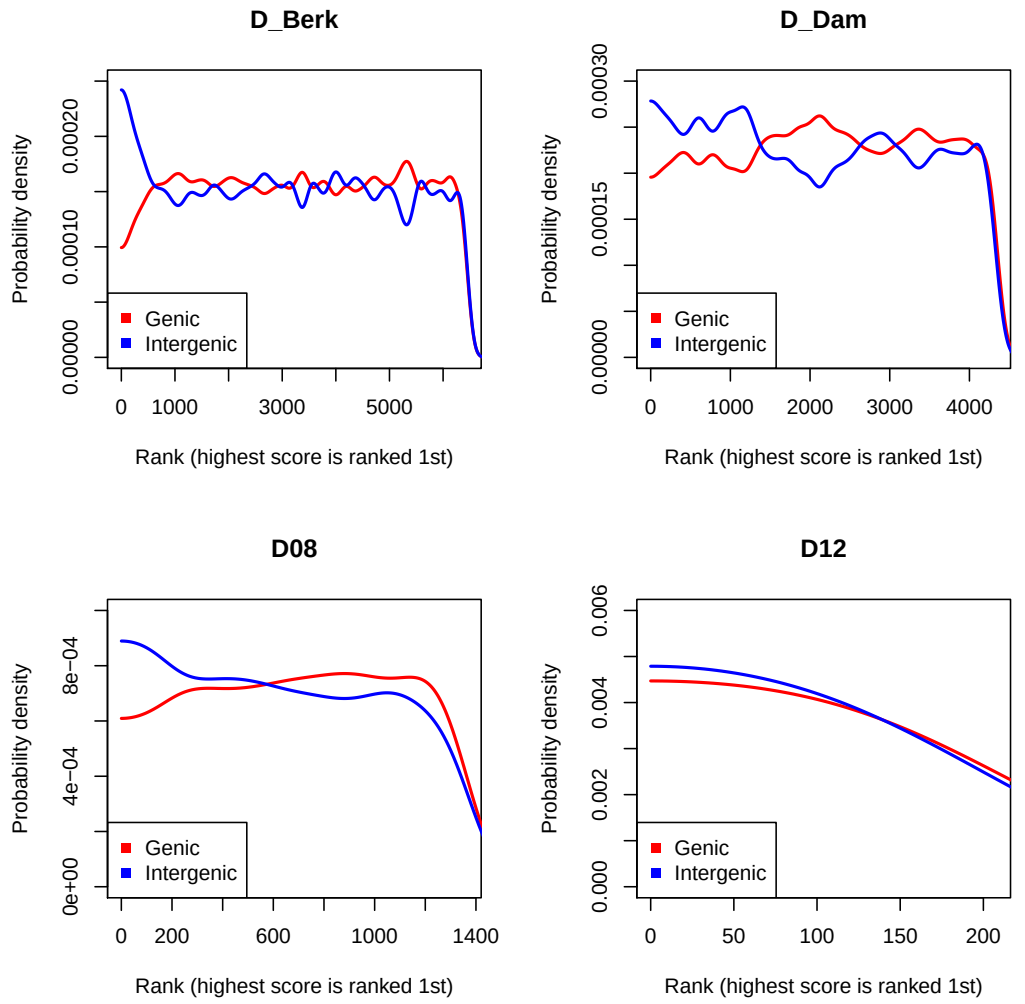

Figure 3.12: Probability density of the genomic features hit by different Dichaete datasets, based on the rank of the intervals in question. The highest scoring intervals in a set have the lowest rank numbers - for example, the highest scoring interval in each set has the rank '1'. Note that the axes are different between the different graphs - because both the numbers of intervals and the probability distibutions are very different between datasets, the axes were adjusted to visualise clearly what's happening within each individual dataset. From the graphs we can see that D_Berk, D_Dam and D08 all show the same trend, where the higher ranking intervals are more likely to land in intergenic regions. The D12 dataset does not display the same trend.

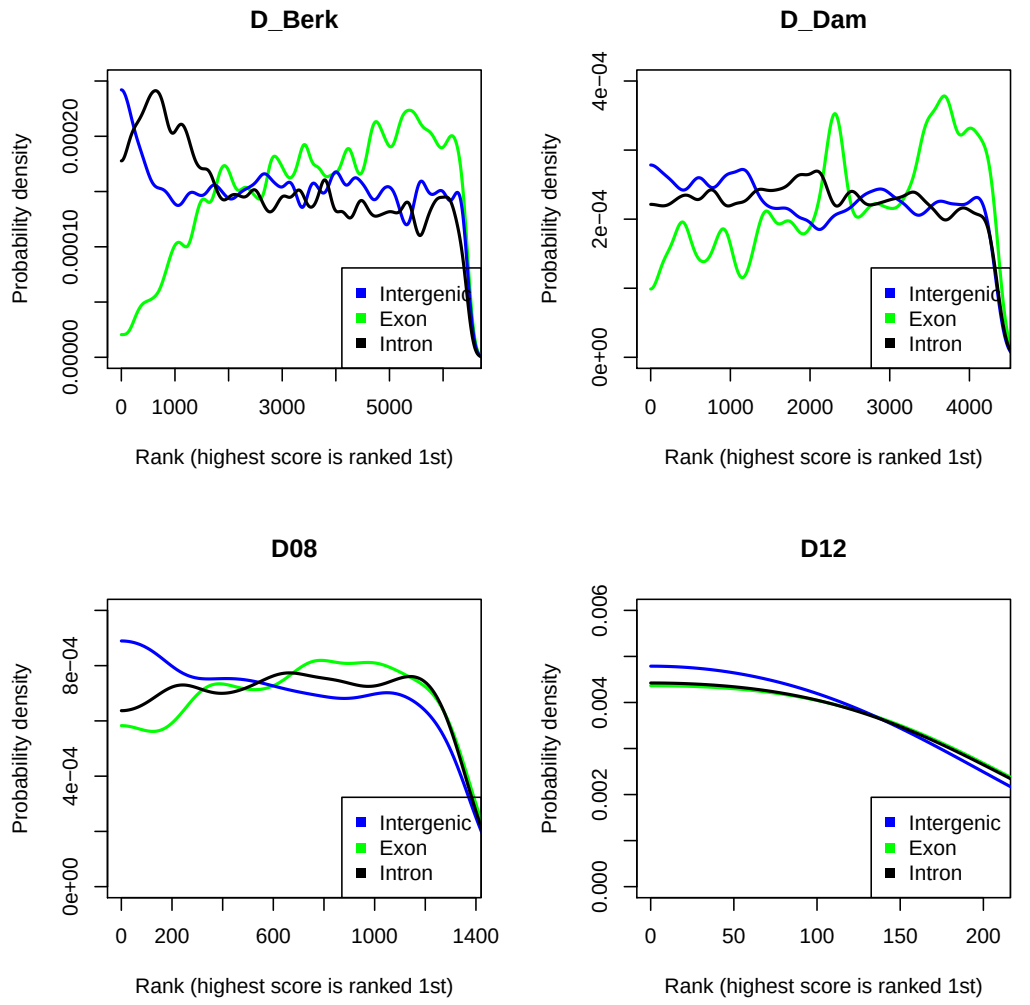

Figure 3.13: Probability density of the genomic features hit by different Dichaete datasets, based on the rank of the intervals in question. Only peaks that could be unambiguously mapped to one particular type of genomic feature were included in the analysis. The highest scoring intervals in a set have the lowest rank numbers - for example, the highest scoring interval in each set has the rank '1'. The axes are different between the different graphs, because both the numbers of intervals and the probability distibutions are very different between datasets, the axes were adjusted to visualise clearly what's happening within each individual dataset. From the graphs we can see that in D_Berk and D_Dam , and to a lesser extent in D08, exon hits tend to be associated with very low scoring peaks, compared to intergenic and intron hits. While the intergenic hits still appear to be the highest scoring, the difference between intergenic and intronic hits does not appear to be that large. The D12 dataset does not display any differences in trends for the different feature types.

3.4.5 Genes targeted by Dichaete binding to different genomic features

The peaks from each dataset were divided into 3 categories - intergenic, introns and exons. The peaks with ambiguous assignments, hitting both introns and exons, were ignored. Gene lists were then generated for each of the categories and GO enrichment analysis was performed to determine whether different genomic binding locations correspond to different classes of genes.

The GO enrichment data for D12 exon hits is not shown because no significant enrichment was found (likely due to the small number of genes in the category). For the other datasets, the heat map of enrichments is shown in Figure 3.14. The clustering and heat map results are ambiguous. The D_Dam and D_Berk datasets show similar profiles for introns and intergenic hits, with genes with intron hits performing more roles, such as cell cycle regulation and cell communication. In addition, it seems to be specifically the genes with Dichaete exon hits from these 2 datasets that perform cellular component organisation and biogenesis. For the other two datasets (D08 and D12), their introns also show a profile similar to each other, but distinct from the Berkeley and DamID data (and more similar to their intergenic hits profile). On the whole, very few genes are hit in exon regions in 3 out of 4 datasets, and therefore they show few or no GO enrichments. In the case of the Berkeley data, large numbers of exons are bound, giving a distinct enrichment profile. This raises the possibility that some blastoderm-specific regulation is happening through Dichaete exon binding. Alternatively, exon binding may be transient and tissue specific and therefore not picked up in datasets looking at longer time periods of development.

Cell cycle genes seem to be specifically targeted in D_Berk data, specifically with intergenic or exon rather than intron binding. Transient exon binding might be related to regulating transcription through PolII pausing, or perhaps to bringing regulatory elements together through DNA looping. Cellular component organisation genes seem to be targeted by both D_Berk and D_Dam exons, though the reason for this pattern is unclear. On the other hand, transcription related genes seem to be hit specifically in intronic and intergenic regions rather than in exons, suggesting that perhaps Dichaete is regulating them in a more traditional fashion via upstream regulatory regions and/or through facilitating PolII poising. Cell cycle regulation genes seem to be targeted specifically through intronic regions, as do genes for regulation of translation and ectoderm development. It is worth noting that one of the few verified Dichaete targets, the *slit* midline enhancer, is located in an intron.

It is interesting that the D_Berk and D_Dam datasets show more similarity to one another, than the D_Dam dataset shows to D08 and D12, which were done on a much

more similar time scale. This suggests that the different collection times, while perhaps contributing to, are not the main determinant of the differences between the datasets. As far as the ChIP-chip datasets go, they appear to cluster somewhat by lab - this might be due to the antibody used in each case, or to different labs using different fixation times or other distinct protocol modifications, or to variability caused by different researchers performing the experiments. The similarity between D_Berk and D_Dam is extremely encouraging, particularly given both the different techniques used and the different developmental periods studied. While it is difficult to assess the validity of any one data set without a larger pool of solid validated reference points, this nonetheless seems to suggest that D_Berk and D_Dam are reflective of bona fide Dichaete binding patterns.

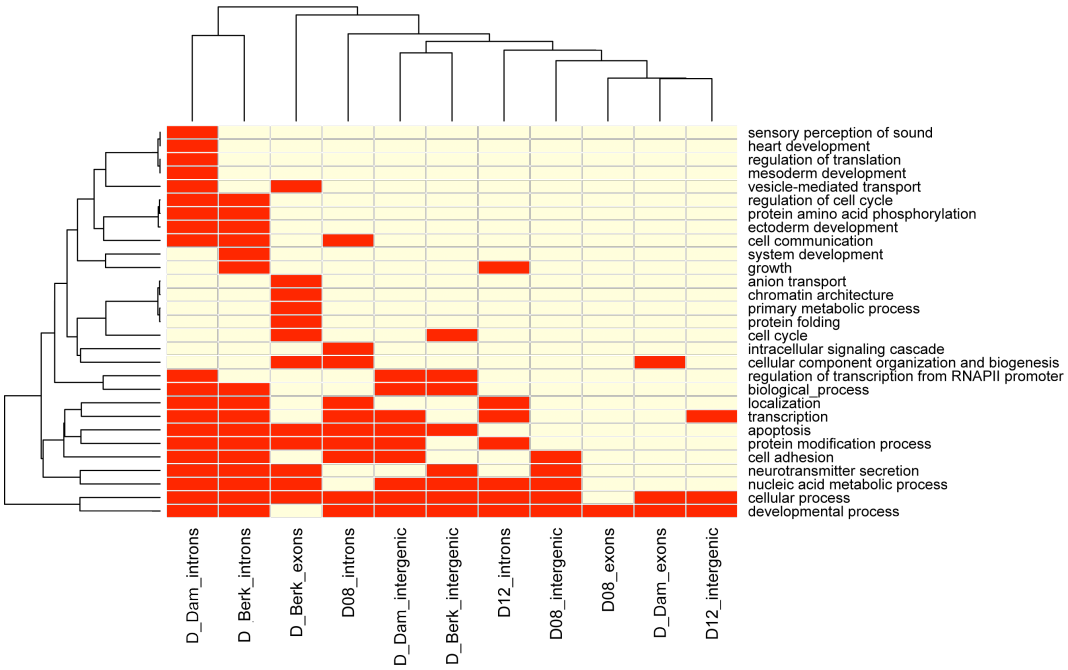

Figure 3.14: GO enrichments for Dichaete intervals hitting different genomic features (significant p-values are shown in red)

3.5 Dichaete sequence specificity

3.5.1 De novo motif finding

Dichaete is a sequence specific transcription factor and a binding motif sequence from a bacterial 1-hybrid experiment was already available in the JASPAR database (Bryne et al., 2007). De novo motif searching was also performed using the NMICA software (Down and Hubbard, 2005) with the available Dichaete datasets. NMICA is a highly sensitive but computationally intensive algorithm for finding overrepresented sequences,

so a selected subset of binding intervals was used in order to keep the computational time reasonable. The motif searches were run twice, using the best 100 and the best 200 peaks from each dataset at FDR 1% (FDR 5% in the case of D08). A subset of the peak intervals, 300 bp around the highest scoring window within each interval, was used. The top 15 motif hits of length of 6 - 10 bp were found in each search. One of the searches using the D_Berk dataset revealed a position weight matrix very similar to the reported JASPAR motif, adding validation from in vivo data to the bacterial 1-hybrid experiment: the two motifs are presented in Figure 3.15. This is therefore likely to be a genuine Dichaete binding motif.

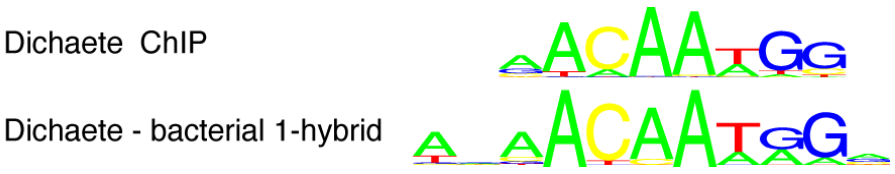

Figure 3.15: The Dichaete motif found from the top 200 peaks of the D_Berk ChIP-chip dataset, compared to the known one from JASPAR

While the motif found is very similar to the previously known sequence, which is encouraging, it is also notable that nothing at all similar to the Dichaete motif was found through searches with the other 3 datasets. There are several potential reasons for this. As previously noted, the D_Berk dataset reflects a tight developmental time and could therefore be more indicative of precise Dichaete binding sites rather than broad patterns of binding. This becomes more important when only a small selection of peaks is used: the Dichaete motif needs to be overrepresented in a 300 bp region of the selected highest scoring peaks in order to be found. Location shifts of the binding position during development may result in the peak 'centre' moving and the Dichaete motif not being found. To an extent it is also a matter of chance since NMICA has a stochastic element and the quantitative information used for selecting the top intervals may not be indicative of their reliability and functionality. Therefore, while finding a motif is encouraging, not finding one may or may not mean anything.

3.5.2 Overrepresentation of the Dichaete motif in Dichaete binding regions

It is known from literature that while transcription factors often have sequence-specific binding preferences, these are not always clearly causally connected to functionality. Indeed, there are reported examples of functional conservation despite an absence of sequence conservation. An example of this is the stripe 2 enhancer of *eve*, where it has been demonstrated that both the binding site sequences and the location of the enhancer sites have undergone a considerable degree of reshuffling between species,

with new mutations arising to keep the expression of *eve* constant (Ludwig et al., 2000), demonstrating that sequence conservation and the presence of a particular binding site sequence are not necessary for the conservation of enhancer function. Typically, while the motif of a particular transcription factor will be highly represented in genome-wide binding data, there will nonetheless be many regions where binding is detected despite the absence of the transcription factor's preferred sequence.

One explanation for these observations is that sequence specificity is not always required for the binding of a particular transcription factor, but because of the increased affinity of the TF for a particular sequence motif, it is more likely to end up in genomic locations that contain the motif. Thus, if the presence of a particular TF was required in a particular part of the genome for a critical function during development, while it is not guaranteed, it is nonetheless expected that there would be an overrepresentation of the TF binding motif at that location. On the other hand, the regions where the TF plays a less critical role may have less of an overrepresentation of the binding motif.

Dichaete is a transcription factor that binds in the minor groove, which means it does not have a physical 'view' of the bases (i.e. the specific array of hydrogen bond donors and acceptors) in the way that major groove binding transcription factors do (Tullius, 2009). Instead it is thought to rely on the electrostatic potential of the DNA backbone, which varies due to the shape of the backbone being altered by the sequence of bases attached (Rohs et al., 2009). Thus it is perhaps even more likely in this case that the presence of the motif is an aid for location-finding rather than a crucial component of Dichaete functionality. Nonetheless, looking at the patterns of motif occurrence might give an indication of the reliability of the data and their presence might also highlight some of the more important areas of Dichaete function.

Because in vivo data is thought to be more reflective of genuine transcription factor biology than in vitro studies, the motif found from the D_Berk data was used for further analysis, rather than the JASPAR one. The motif scanning software nmscan (part of the NMICA package) was used for scanning the interval regions for motif hits (at the threshold -5, which was thought appropriate for this length of motif). In order to assess the statistical significance of the overrepresentation of the Dichaete motif at each interval, the columns of the Dichaete motif were randomly shuffled 10,000 times, making sure they did not match the original motif, and the randomly generated motifs were used to scan the intervals at the same settings. The numbers were then compared to those for the original Dichaete motif and the p-values and z-scores were calculated. (p-value of an interval was calculated as the proportion of random motifs that have the same or greater number of 'hits' in that interval as Dichaete).

The summary of results is shown below in Table 3.6. Notably, the method used to simulate the 'background' in order to assess the significance is likely to massively un-

derestimate the actual significance of the Dichaete motif overrepresentation - many of the randomly generated Dichaete motifs are expected to be very similar to the original motif, and therefore are expected to be similarly overrepresented in the binding regions (giving unnecessarily low p-values). This does however mean that the intervals that are flagged as significantly enriched for the Dichaete binding motif are certain to have a substantial number of high quality matches (rather than that the intervals not flagged as significant having none or few sites).

Dataset	Total number of intervals	Number of intervals with the D motif	Number of intervals with the D motif significantly enriched
D_Berk	6452	5957 (92%)	1031
D_DamID	4342	4001 (92%)	668
D08	1366	1206 (88%)	187
D12	208	181 (87%)	7

Table 3.6: Numbers of intervals with the Dichaete binding motif found

In total, 1473 genes are associated with at least one interval significantly enriched for the Dichaete binding motif. The list includes usual targets, such as *Hox* genes, *ac*, *sim*, *prospero*, *miranda*, *cas*, *dpp*, *kuz*, a range of transcription factors and other genes involved in segmentation and nervous system development. The list contains 224 genes involved in nervous system development (p-value 7.2948E−38), 104 genes involved in cell fate commitment (p-value $1.348E-28$), 217 genes involved in the regulation of gene expression (p-value 7.3101E−21) and 78 genes involved in segmentation (p-value 1.5141E−14), amongst others. In total, 30 genes have 5 or more (to a maximum of 7) overlapping or nearby Dichaete binding intervals. This group of genes with Dichaete binding clusters contains several transcription factors and is quite specific for nervous system development: 6 of the genes in question are involved in neuroblast differentiation and 14 are involved in regulation of gene expression. A network view is shown in Figure 3.16.

Interestingly, the majority of intervals hit intergenic regions in the cases of all 4 datasets, which was not the case for the datasets as a whole. It is therefore possible, accepting the caveats discussed above, that the binding locations containing an overrepresentation of the Dichaete binding motif are active, functional binding sites, whereas a subset of the binding sites that show no significant motif overrepresentation might be non-functional sites.

3.5.3 Sequence specificity

The GC content of Dichaete binding peaks and surrounding regions was studied to determine whether there are clear patterns in sequence composition associated with

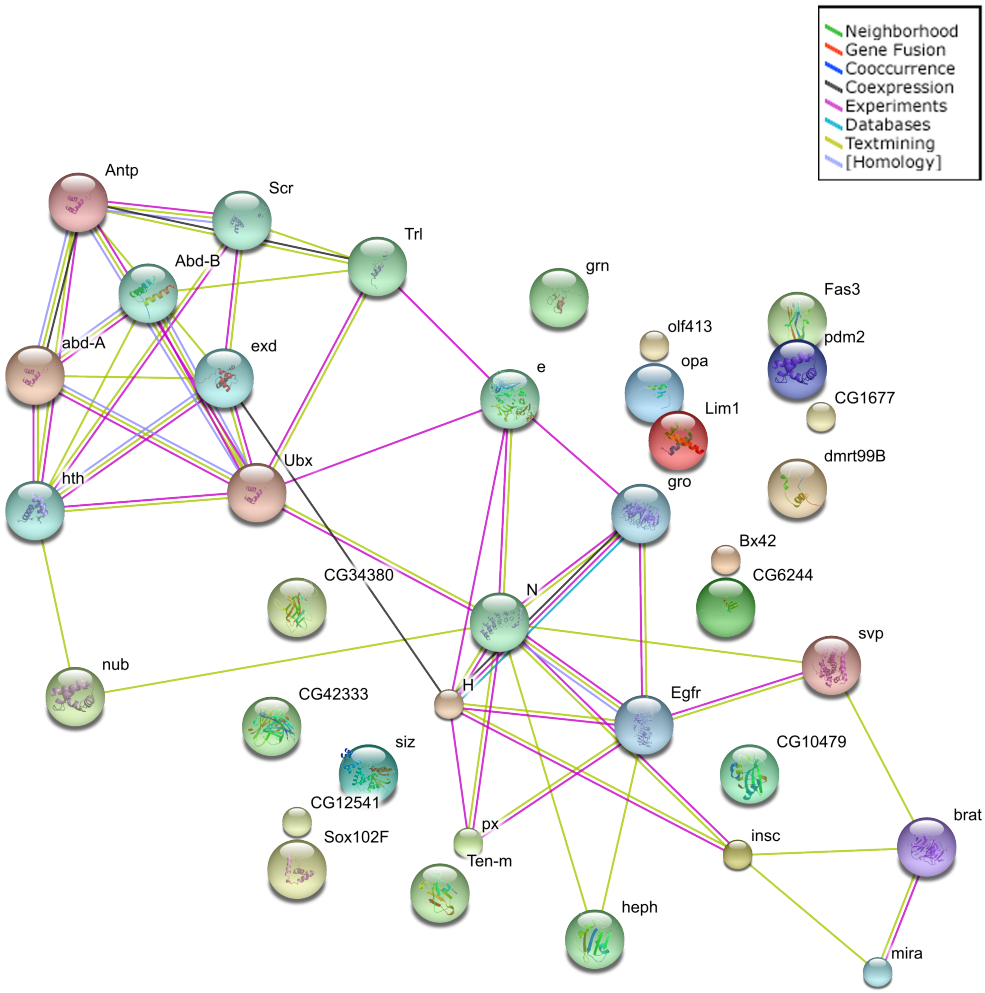

Figure 3.16: Genes hit by clusters of Dichaete binding. The genes associated with 5 or more Dichaete binding sites were uploaded to FlyMine (Lyne et al., 2007) and imported into STRING 9.0 (Search Tool for the Retrieval of Interacting Genes/Proteins Von Mering et al., 2005), the tool with which the network figure was generated. The different line colours represent the types of evidence for the association, with the legend shown in the top right corner.

Dichaete. Only the high stringency binding intervals were examined: values were found for 200 bp windows in a range of 5 kb upstream and downstream of the centre of the highest scoring window of the interval (Figure 3.17).

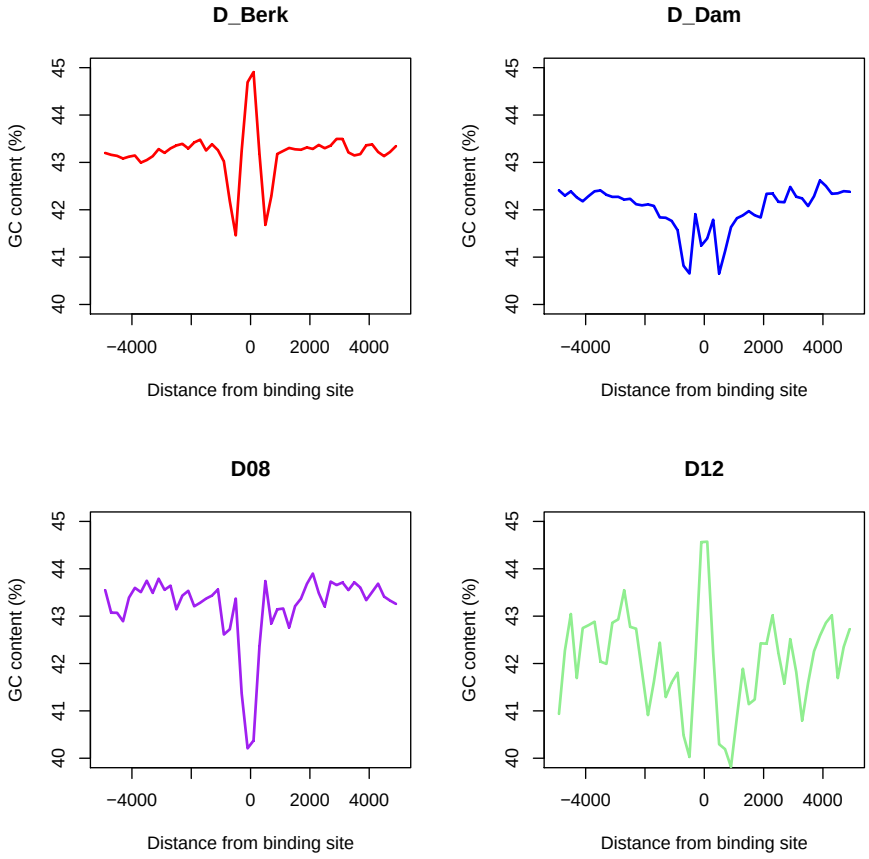

Figure 3.17: GC content of areas surrounding Dichaete binding peaks

Surprisingly, each of the datasets shows distinct but different GC profiles. The D_Berk and D12 datasets show the most similarity, with a rise in the % GC content at the estimated binding site location and dips in GC content immediately adjacent to the binding site. Interestingly, the D08 dataset shows the exact opposite trend, with a dip in the GC content right in the middle of the binding site. The DamID GC profile looks very unusual. It does not appear that the problem is in estimating the binding site location, as the profile clearly shows a distinct pattern around position zero. However, the pattern is unclear. It is possible that Dichaete binding sites at different stages have different, distinct binding profiles, with some low and some high in GC content, and the mixture of those profiles could conceivably give a graph such as that generated for DamID GC content. As previously mentioned, there is a bias in the DamID signal, in that for a signal to occur at all, GATC sites need to exist in the sequence around the location in question. However, while this may explain a bias in the GC content of the DamID data, it does not explain why different ChIP-chip experiments - D_Berk and

D08 - give very different results for GC content.

Looking at the level of variability in the data, it is also questionable whether just looking at the mean is really fully informative, since it does not capture the range of sequence variance. The collection of raw data for each interval is shown in Figure 3.18, with the GC content of each interval plotted individually (200 bp window, within a 5 kb range). While looking at the raw data suggests that there does appear to be a trend at the zero position compared to the remainder of the analysed interval, it is also clear that just taking the average flattens much of the information. Consequently, box plots were generated for each 200 bp window for each of the datasets, in order to better visualise the sequence content trends.

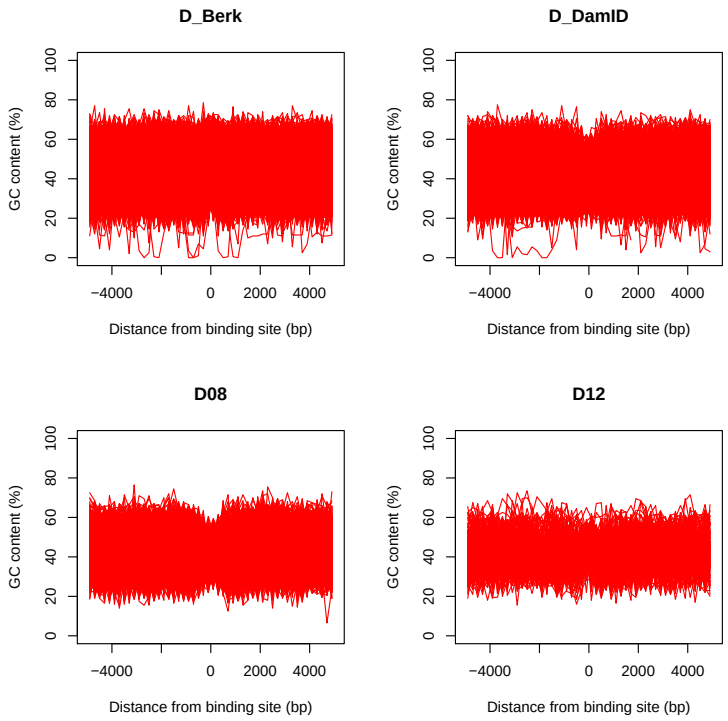

Figure 3.18: GC content of areas surrounding Dichaete binding peaks - raw data

It appears that the variance of the sequence GC content next to the binding site in data sets B_Berk, D_Dam and D08 is visibly lower than that of the surrounding sequence (Figure 3.19), suggesting sequence constraint in areas of the genome that Dichaete binds. However, the exact specificity seems to vary from dataset to dataset, with D_Berk and D12 exhibiting a peak of higher GC content in areas of Dichaete binding, but the D08 and D_Dam apparently showing a drop in GC content (though in D_Dam, it is not obvious that the median is significantly different from that of surrounding sequence). So while there appears to be a difference in the GC content of genomic DNA in the vicinity of Dichaete binding sites compared to the surrounding sequence, a clear trend cannot be deduced from these data.

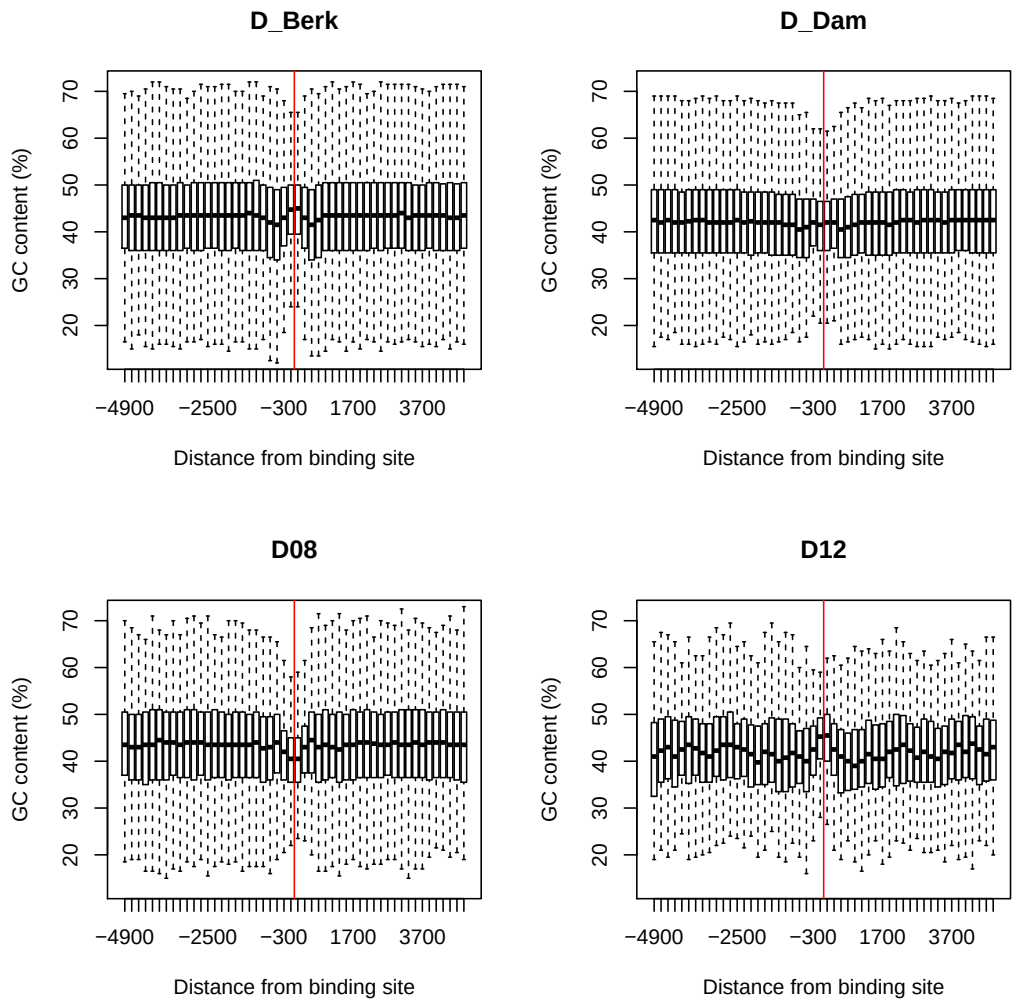

Figure 3.19: GC content of areas surrounding Dichaete binding peaks, with the coordinate of the Dichaete binding peak labelled with a red line.

3.5.4 Conservation

It has been suggested that the binding locations of important transcriptional regulators are highly conserved across closely related species, as exemplified in the case of Twist (He et al., 2011). An alternative hypothesis is that the reshuffling of transcription factor binding sites is a driving force for evolution, so some transcription factors are expected to have highly divergent binding site locations in closely related species. Examples of this have been convincingly demonstrated in mammalian genomes (Odom et al., 2007). Ludwig et al. (2000) look in depth at an example of this divergence - the enhancers regulating the stripe 2 expression of *eve*. While the expression of *eve* is strongly conserved in *Drosophila*, the enhancer sites appear to have undergone a considerable degree of reshuffling between species. Ludwig et al. (2000) experimentally show that reshuffling the enhancer sites in *D. melanogaster* has an impact on gene expression. From there, they propose that there is stabilising selection acting on the expression of *eve* across species, and that the reason the enhancer sites can move around is that new mutations arise to stabilise the effects of the reshuffling on *eve* expression.

If we assume that transcription factor binding is causally connected to the underlying genome sequence, which for sequence specific transcription factors appears to be broadly correct, then in the absence of binding data for different species, looking at the sequence conservation may act as a predictor of binding conservation. In addition it may potentially highlight the functionality of the binding sites in question. However, given the evidence for enhancer site evolution described above, the inverse is not the case - a lack of conservation does not necessarily imply non-functionality.

The phastCons scores for multiple alignments of *Drosophila* melanogaster with 14 other insects, downloaded from UCSC for Genome Release dm3, were used for this analysis. The average phastCons scores were found for 100bp windows in a range of 5 kb around the middle of the highest scoring window of each binding interval (Figure 3.20).

Dichaete peaks do not appear to be any more conserved than their surrounding regions. In fact, slightly in the D_Dam and strikingly in the D08 data, the Dichaete peak areas in fact seem to have a drop in conservation compared to the surrounding sequence. The D_Berk and the D_Dam datasets also seem to have a drop in the conservation of the surrounding sequence, which is interesting. The shapes of the graphs bear a strong resemblance to the ones for GC content, and it seems likely that GC content and conservation are connected. This may be due to the fact that the highest GC content in *Drosophila* is found in the coding regions of genes (Zhang et al., 2004), which are also the areas showing the most conservation. In terms of Dichaete conservation, it does not appear that Dichaete binding sites are more conserved than their surrounding genomic sequence. On the contrary it appears that, if anything, Dichaete and the nearby sequences tend to be significantly less conserved. This could mean that Dichaete

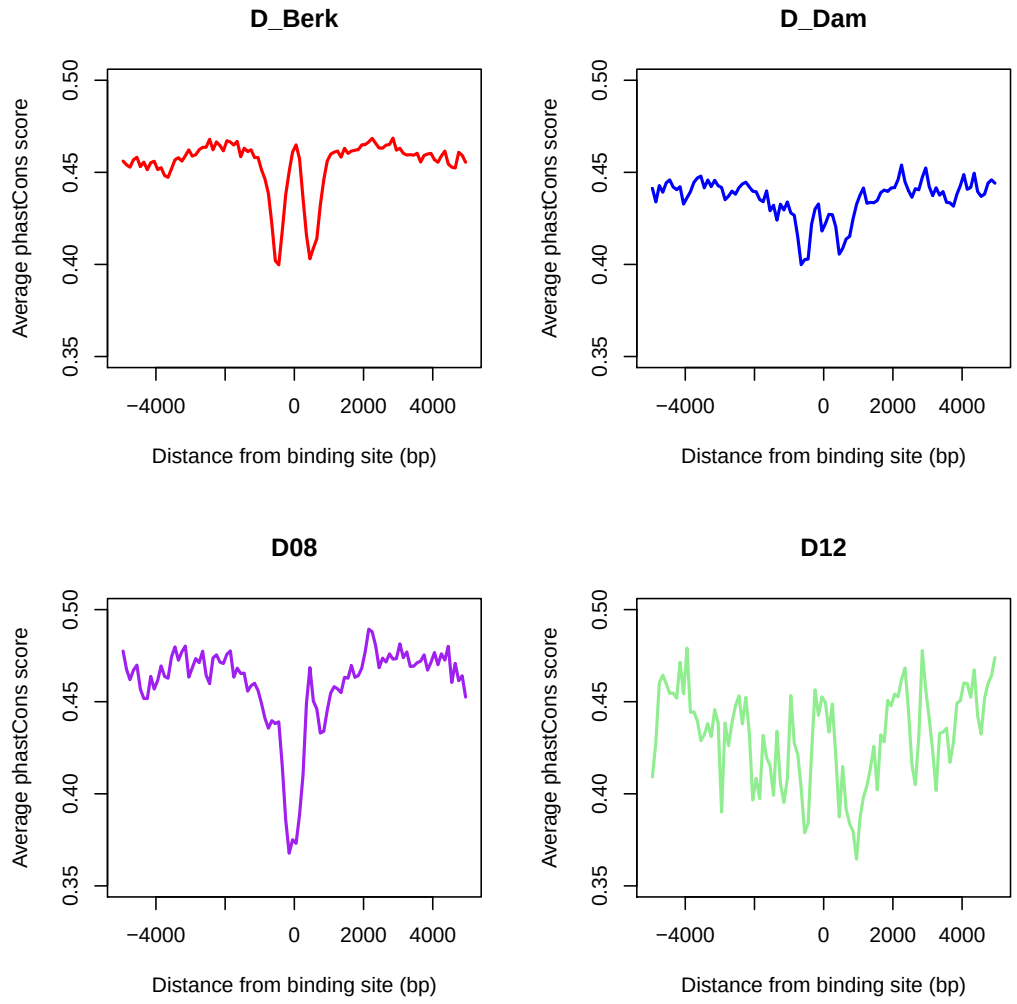

Figure 3.20: Average phastCons scores of areas surrounding Dichaete binding peaks

binding sites are fast-evolving between species (a series of ChIP-chip experiments with different *Drosophila* species could resolve this), or that binding sites perhaps stay the same between species, but are not completely determined by exact sequence specificity, but rather via some other property, such as for example the overall flexibility of the stretch of DNA in question.

3.6 Dichaete and other factors

3.6.1 Dichaete and transcriptional pausing

Looking at the Dichaete binding profile raises an interesting question - what is the purpose of apparent preference for binding in intronic regions. A plausible hypothesis would be that, as well as regulating gene expression the 'traditional' way via a cis-regulatory module, it may also have a role in influencing gene expression through regulating transcript elongation by causing or assisting RNA polymerase pausing. Alternatively, it is possible that this pattern reflects the fact that Dichaete tends to bind to developmental genes with large introns, which may contain enhancer elements: this possibility is investigated in the next section.

Transcriptional pausing has been previously reported as a regulatory mechanism for a number of developmentally important genes in *Drosophila* (Zeitlinger et al., 2007a; Muse et al., 2007). In particular, it is known that regulation of transcriptional elongation is crucial in the regulation of *slp1* (Wang et al., 2007), a gene that the binding data suggests is also regulated by Dichaete. Polymerase pausing may be an effective way of ensuring immediate transcriptional response by keeping the polymerase docked at the start of a gene until it is needed. Additionally, it may facilitate precise regulation of transcript quantity through slowing down the progress of polymerase along the gene when required. The reported effects of disrupting the polymerase pausing are varied. The obvious effect is a lowering of expression due to pausing and therefore overexpression when pausing is interrupted. However, it has also been reported that polymerase stalling can enhance gene expression (Gilchrist et al., 2008). This is thought to be because in the absence of factors that facilitate polymerase pausing, chromatin structure switches from an open to a closed configuration, resulting in apparent downregulation when pausing is disrupted.

Genes can be classified according to 3 distinct types of polymerase binding profile: actively transcribed (with no stalling), non-transcribed and genes with a stalled polymerase which show clear binding peaks in pausing positions (Zeitlinger et al., 2007a). If Dichaete were involved in facilitating polymerase pausing, a correlation between peaks of polymerase binding within genes and Dichaete binding within the same regions is

expected. While for the purposes of pausing, there is no difference between exons and introns - the polymerase can pause at any position along the transcript - the sequence of exons tends to be highly conserved due to functional constraints on the protein. Introns, on the other hand, can potentially be shaped by evolution specifically for the purpose of gene expression regulation, so they are likely to be the more promising feature to investigate.

To start with, the RNA polymerase II *Drosophila* 0-12 h embryo ChIP-chip dataset was downloaded from modENCODE and reanalysed using the TiMAT pipeline with standard settings (quant normalization, 675 bp window, 200 bp gap, 10 oligo minimum). The FDR 1% peaks were then compared with the high stringency Dichaete datasets. The rationale for using the polymerase FDR 1% data was that the peaks of interest for this analysis are the very pronounced ones, rather than general regions of low level PolII binding. Datasets for narrower time periods (0-4 h, 4-8 h) were also available, but upon visual inspection, these were found to be lower quality and noisier so only the 0-12 h dataset was used for analysis.

As an initial analysis step, a resampling-based significance test (Huen and Russell, 2010) was performed to determine whether there is a statistically significant overlap between the polymerase data and the Dichaete datasets. The only one that was flagged as significant was the overlap with D_Berk (p-value = 0.001), possibly because it corresponds more closely to specific open chromatin regions than the other 3 datasets that are an average over a longer developmental window.

The next step of the analysis was to determine whether there is anything special about Dichaete intronic binding that corresponds with spikes of PolII binding. A trend was found: the Dichaete sites in genic regions are much more likely to overlap with PolII than sites in intergenic regions (not at all surprising, given the expected location of PolII within and just upstream of transcribed regions). In the case of D_Berk data, 61% of Dichaete genic intervals overlap with PolII intervals, compared to only 43% of intergenic intervals (p-value $<$ 2.2E-16, χ^2 test). The numbers are 51% of genic intervals vs 25% of intergenic ones for D_DamID (p-value $<$ 2.2E-16, χ^2 test), and 86% genic vs 64% intergenic in the case of D08 data (p-value $<$ 2.2E-16, χ^2 test). A significant difference was not found in D12, possibly due to the low number of peaks present in these data.

Out of the genic sites, Dichaete sites in exons were much more likely to overlap with PolII than Dichaete sites in introns. In D_Berk data, 72% of the Dichaete intervals associated with exons overlap with PolII, as opposed to 49% of intervals that bind introns (p-value $<$ 2.2E-16, χ^2 test). The same trend was also present in the D_DamID and D08 datasets, with 68% of exon-binding intervals and 44% of intron binding ones overlapping with PolII in the case of D_DamID (p-value $<$ 2.2E-16, χ^2 test), and 93%

exon vs 81% intron intervals overlapping in the case of D08 (p-value $<$ 8.8E-6, χ^2 test).

Next, the distance of the PolII peak (the highest scoring window of each interval) to the Dichaete peaks was examined for all the Dichaete peaks within 1 kb of PolII sites. If more than 1 PolII site was within 1 kb of a Dichaete peak, that peak was counted multiple times. While many Dichaete peaks are found within 1 kb of a PolII peak, the trends are quite different for the different datasets (Figure 3.21). While the D08 dataset shows a clear tendency for Dichaete peaks to land right next to PolII peaks, and the D_Berk dataset shows a slight trend in that direction, the other two datasets do not show such trends at all, making the analysis inconclusive.

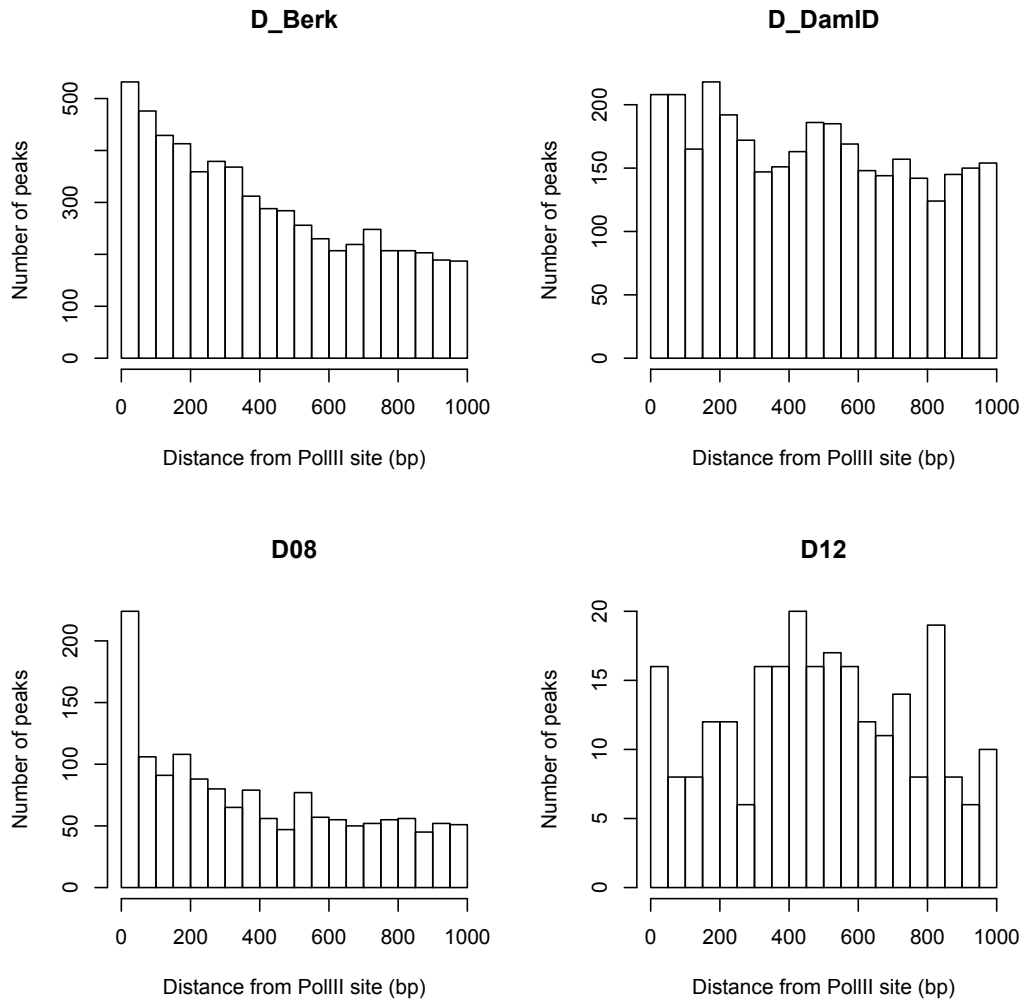

Figure 3.21: Histograms of the distance of Dichaete binding sites from PolII binding sites

3.6.2 Dichaete and enhancers

As mentioned, an alternative hypothesis for the tendency of Dichaete to land into introns is that it is targeting large introns containing enhancers. This is a difficult hypothesis to verify, since a full genome-wide dataset of enhancer regions at present does not exist, and using cloning and molecular biology, while potentially interesting, was considered outside the scope of this project. Two analyses were nonetheless run, which were thought to be, if not fully conclusive, at least informative about whether this hypothesis is worth pursuing further in the future.

Intron size

There are a number of examples from single gene studies of enhancers found in particularly long introns in Drosophila (Kohler et al., 1996; Hauck et al., 1999; Meredith and Storti, 1993; Markstein et al., 2004; Stathopoulos et al., 2002). Also, interestingly, it has been found that long introns are more evolutionarily conserved than short ones, implying that they commonly contain important regulatory elements (Haddrill et al., 2005). Since a typical metazoan enhancer is about 500 bp in length (Erives and Levine, 2004), it is possible that the length of the introns needs to be above a minimum threshold to provide enough space for the presence of enhancers.

Based on the assumption that intron length is indicative of the potential presence of enhancers in the region, an analysis was done to verify whether the introns that Dichaete binds within are particularly long, compared to those in the rest of the genome. Highest scoring windows within Dichaete binding intervals were used to estimate the Dichaete binding site location, and those mapping to introns were selected for this analysis. Intron coordinates were downloaded from UCSC (Rhead et al., 2010) for all *Drosophila* genes. Introns from the transcript with the greatest number of introns were used for this analysis. Alternative transcripts for the same genes were ignored to avoid counting the same introns multiple times. The comparison of the length of introns hit by peaks in different Dichaete datasets, compared to the rest of the genome, is shown in Table 3.7. The length of introns bound by Dichaete is strikingly different from the genome-wide average in all 4 data sets, with the average length of introns hit by Dichaete being in all cases significantly larger on average. While this is not direct evidence, these data are consistent with the hypothesis that Dichaete binds to intronic enhancer regions.

This result may also be a reflection of the fact that longer introns are more likely to be hit at random, and would therefore be more represented in the datasets. However, this assumes that the occurrence of binding intervals in question is indeed that - random.

Answering whether this is the case is tricky. Certainly, if Dichaete is just landing non-specifically to whatever regions are accessible at the time, they could be considered to be random within the constraints imposed by chromatin accessibility. Further to that, for this analysis only one specific nucleotide coordinate is chosen for each binding interval (based on the highest window score in that interval), and it is possible that the coordinate chosen is more likely to land in a large intron than a small one, just by virtue of their size. The argument against the completely random model is that there is consistency across datasets in terms of binding interval location, despite the differences in the experimental conditions used, indicating that the experiments are picking out locations that are Dichaete specific, rather than random. However, picking out a single coordinate out of the interval may still introduce bias. For this reason, while the broad observation that Dichaete lands in long introns is likely to be correct, the extremely high p-values might be a result of a length bias.

Dataset	Mean length of introns hit by Dichaete peaks	Mean length of introns in the rest of the genome	Wilcoxon rank sum test p-value
D_Berk	9199 bp	826 bp	2.2E-16
D_DamID	9372 bp	851 bp	2.2E-16
D08	9428 bp	1027 bp	2.2E-16
D12	11480 bp	1081 bp	2.2E-16

Table 3.7: Size of introns hit by Dichaete peaks compared to the rest of the genome

Overlap with known CRMs

To further verify whether it is plausible that Dichaete interacts with enhancers, a list of 997 known *Drosophila* cis-regulatory modules, associated with a total of 325 genes, was downloaded from the REDfly database (Halfon et al., 2007). The locations of the known CRMs were compared to the available Dichaete datasets (Table 3.8). Assessing the significance of overlap proved to be unexpectedly difficult. Using something like a t-test or a hypergeometric test gives implausibly low p-values for almost any degree of overlap, by assuming that binding to any site in the genome is equally likely (an assumption that is demonstrably false). On the other hand, using the resampling-based signicance test implemented in the Cooccur package (Huen and Russell, 2010) does not indicate any significance of overlap, which might be an underestimate - the numbers of sites are very different for the Dichaete data and the CRM sites, and the collection of CRMs is incomplete and non-random. Certainly, the number of CRMs overlapped by Dichaete intervals appears high in the cases of D_Berk and D_Dam data, with 55% and 43% of known CRMs being hit, respectively. Also, because the CRMs stored in REDfly regulate a range of different genes, it is not reasonable to expect a complete overlap, as we might for binding profiles of cooperating transcription factors. Despite

the thousands of binding sites, we do not actually expect Dichaete to regulate every existing CRM in the genome. Therefore the analysis outcome seems to be plausibly consistent with the hypothesis that Dichaete regulates transcription by binding to enhancer regions.

There is almost certainly a bias in what sorts of genes are studied enough to have multiple experimentally characterised enhancers, so gene list analysis in this case is not very informative - as expected both due to gene length bias and bias of being interesting enough to be studied, a number of transcriptional regulators and other important developmental genes feature. The list does contain overlaps with CRMs of genes that are expected to be targeted by Dichaete based on previous literature, which is encouraging. In particular, Dichaete overlaps the CRMs of Dichaete targets during segmentation - *eve*, *ftz* and *run*, the AS-C complex genes *ac* and *sc*, the Dichaete cofactors *ind* and *vvl*, as well as the known direct Dichaete target *slit*. It also overlaps the CRMs of interesting nervous system targets, such as *bl*, *ey*, *hb*, *Kr*, *nub*, *grh* and *sim*. The list also includes Dichaete's own CRMs, suggesting it self-regulates.

Dataset	Number of CRMs overlapping with Dichaete	Percentage of known CRMs overlapping with Dichaete
D_Berk	549 (from 185 genes)	55%
D_DamID	427 (from 158 genes)	43%
D08	98 (from 51 genes)	10%
D12	5 (from 4 genes)	0.5%

Table 3.8: Overlap of Dichaete data with verified CRMs from REDfly. The total number of genes associated with the CRMs in question is shown in brackets.

3.6.3 Matches to other TF position weight matrices

It is expected that any direct Dichaete cofactors will have binding sites physically close to a Dichaete site. For example, the distance between the Dichaete and Vvl binding sites in the *slit* regulatory region was proposed to be between 50 and 70 bp (Ma et al., 2000). In the case of mammalian Sox2 and Oct4, their crystal structure shows a distance of 3 bp between the 7 bp Sox2 and 8 bp Oct4 binding site, though distances may vary for different sites and binding partners (Remenyi et al., 2003). Cofactor motifs, if present, should therefore also be flagged as overrepresented in the NMICA searches. First motif searches were performed for each binding dataset using the NMICA tool. Two sets of 15 motifs each were acquired for each dataset - one found using a 300 bp region of the top 100 peaks, and the other one for the top 200 peaks. The search was performed for motifs that are 6-10 bp long, and all motifs found this way were used for subsequent analysis. The STAMP motif matching tool (Mahony

and Benos, 2007) was used to compare the NMICA results to the existing transcription factor position weight matrices from the JASPAR database (specifically, only the non-redundant matrices available for insects) (Bryne et al., 2007), and also to the sequences from FlyReg (Bergman et al., 2005). The default settings were used - the Pearson Correlation Coefficient was used for comparing the columns, the ungapped Smith-Waterman method was used for alignment, iterative refinement was used for generating multiple alignments and the UPGMA algorithm was used for tree building.

For each pair of motifs, an e-value was given as an approximation of the significance of the match, with the e-value representing the number of times the match is expected to occur at random given the sequence profiles of the inputs. The lower the e-value, the more significant the match. The e-value takes into account that matches to shorter sequences are more likely to be found, scaling the value accordingly.

In total, matches for 75 sequence-specific DNA binding proteins were found, with 5 of these (SuH, Sna, Adf1, Dl and BEAF-32) enriched in all 4 Dichaete data sets. The top 10 motif matches, sorted by e-value, include the strongly matching Dichaete motif, as well as motifs for Top2, Aef1, Cf2, Mad and multiple matches to the BEAF-32 motif. Two matches to Vvl, a known Dichaete cofactor, were found, though the matches were not particularly high quality ($7.54E-03$ for a match to the DamID data, and $2.43E-02$ for a match to the D08 data). This may be a reflection of the poor quality of the known Vvl binding motif. Since not all the matches found were actually high quality, a threshold was applied to produce a selection of the matching motifs considered to be reliable.

Only matches with an e-value of $< 1E-04$ were further examined. The filtered list consists of 35 DNA-binding proteins, one of them being Dichaete itself. The rest of the list is represented as a cloud in Figure 3.22, with the word size corresponding to the number of different Dichaete data sets that contained hits to the motif (minimum 1, maximum 4). The presence of Vnd is encouraging and can be associated with cooperation during CNS patterning. Dichaete is reported to physically interact with Vnd, with potential Dichaete and Vnd binding sites previously identified together in the *achaete* regulatory region (Zhao, Boekhoff-Falk, Wilson and Skeath, 2007). The presence of Grh and Hb motifs may reflect a role for Dichaete during neuroblast division.

This analysis, while not conclusively proving any direct interactions, nonetheless successfully identified some potential cofactors that could be studied in more detail in the future. The strong presence of the BEAF-32 motif is intriguing and may hint at an interaction between Dichaete and particular insulator regions. BEAF-32 is reported to be a marker of hotspots, regions of high transcription factor occupancy in the genome, where it is proposed to potentially facilitate their creation (Roy et al., 2010). Dichaete frequently binds in hotspot regions (on average, it is reported to bind in the same

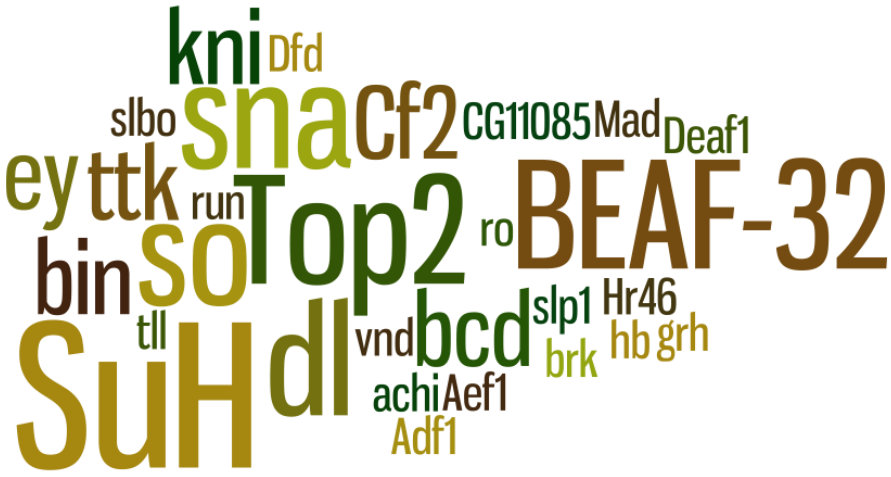

Figure 3.22: TFs with motifs flagged in NMICA searches of D binding data. Only the motif matches with an e-value $< 1E-04$ are shown. The size of the gene name is proportional to the number of Dichaete datasets that the motif for the gene was identified in. The list of genes, along with the number of Dichaete datasets the associated motifs were identified in, is shown in Table 5 in Appendix 2.

regions as 6.9 other transcription factors at a time (Roy et al., 2010)), so it is possible that this hotspot region connection is why the BEAF-32 motif is observed to be enriched so frequently in Dichaete binding intervals. The presence of binding motifs for Vnd, Grh, Hb and Ey in particular, are likely to be connected to the action of Dichaete during nervous system development. A number of transcription factors whose motifs were identified are active in the blastoderm and thus may cooperate with Dichaete during early development. For example, Bc, Hb, Kni, Run and Slp1 all play a role in the early establishment of the anterior-posterior axis and, at least in the case of the first 4, Dichaete may cooperate with them and target the same cis-regulatory modules (Russell et al., 1996).

3.6.4 Association with known genomic element markers

The modENCODE project has generated a large number of ChIP datasets for a range of DNA binding proteins and histone modifications. Much of this data was generated from embryos between 0-12 hours old, and is therefore comparable to the Dichaete data. Therefore, by looking at markers associated with different genomic features and comparing their binding profiles to that of Dichaete we can answer questions such as whether Dichaete is commonly associated with insulators, enhancers, actively transcribed genes or heterochromatin. A caveat when using these data for comparison is that it is not clear at which exact developmental point and in which tissue the observed enrichment originates from, so it is possible for regions of enrichment to

overlap perfectly, but for the proteins in question to never actually be present in the same set of cells in vivo. Nonetheless, this is still an interesting screen for potential associations, but one where the results should be approached with some caution.

In order to test the association of Dichaete with different genomic features, datasets for the following histone modifications and insulator associated proteins were used:

- H3K4me3 - active transcription start sites
- H3K36me3 - actively transcribed exons
- H3K4me1 - enhancer marker (as studied by modENCODE). H3K4me1 has a very different distribution pattern to the other histone methylations. While it is enriched around active genes, its distribution is broader (scattered about 1-2kb upstream and downstream of the TSS).
- H3K9me3 - transcriptionally inactive heterochromatin
- H3K27me3 - Polycomb-directed silencing
- Insulator proteins CTCF, Su(Hw), Mod(mdg4), Beaf-32 and CP190

The binding intervals (in Genome Release 5) were downloaded from the modENCODE website. 0-12 hour embryo datasets were used for comparison to all the Dichaete data, and, where available, 0-4 or 2-4 hour datasets were also used, because they closely match the 2-3 hour collection period of the D_Berk data. If multiple datasets were available for the same time period, they were fused to create one dataset for comparison. The full list of datasets used for analysis is shown in Appendix 1, with the original data used shown in Table 2, the unions performed shown in Table 3, and the final data used for analysis shown in Table 3.

The datasets that significantly overlapped with Dichaete (p-value < 0.05) are shown in network form in Figure 3.23. No significant overlaps were found with the D12 dataset, but as discussed, this dataset is quite different from other Dichaete data in that it has much fewer peaks. Of the other 3 datasets, D_Berk shows the largest overlap with different genomic markers - possibly reflecting the fact that data derived from a broader developmental window gets averaged out over time and space, whereas the smaller time interval appears to give a sharper signal. The less specific data sets are certainly not likely to capture all the details of Dichaete action at every point in development, and thus it is likely we may miss what may otherwise be significant overlaps. Only one genomic marker, H3K4Me1, shows a significant overlap with all 3 reliable Dichaete data sets. Since H3K4Me1 is a marker for both active genes and broader enhancer regions (Roy et al., 2010), this fits well with the observed patterns of Dichaete binding in various genomic regions, and suggests that globally, Dichaete acts by binding to enhancer regions.

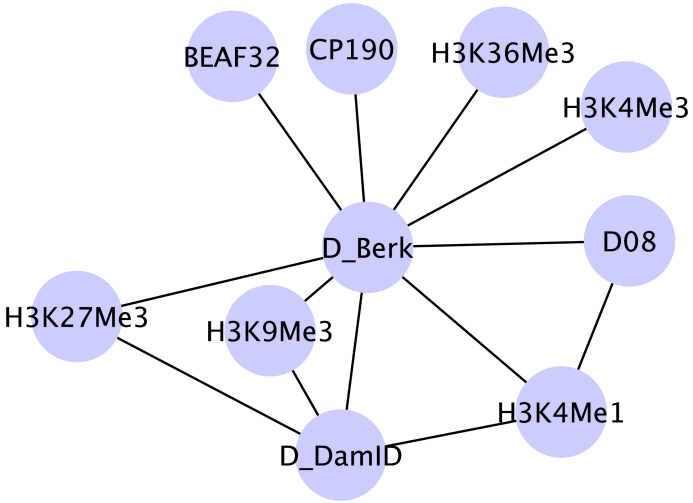

Figure 3.23: Genome-wide histone modification and insulator location datasets with a significant degree of overlap with Dichaete binding datasets (p-value < 0.05), as determined using the resampling based test Cooccur.

Interestingly, both D_Dam and D_Berk display significant overlap with H3K9me3 and H3K27me3, the markers for transcriptionally inactive heterochromatin and Polycomb-directed silencing, respectively. Polycomb group proteins are known to be regulators of *Hox* genes (Beuchle et al., 2001) and act by remodelling chromatin in a way that brings about gene silencing (Lund and van Lohuizen, 2004). Binding data suggests Dichaete also has a regulatory interaction with Hox genes, and it is not clear whether Dichaete acts antagonistically to or cooperatively with Polycomb. Interestingly, it has been found in humans and mice that when Polycomb group genes responsible for H3K27me3 methylation are disrupted, embryonic stem cells lose the ability to maintain themselves in an undifferentiated state (Lee et al., 2006; Boyer et al., 2006). This is a phenotype that is reminiscent of the role that Sox2 (a mammalian Dichaete orthologue) plays in stem cell biology, which suggests that Dichaete action and Polycomb silencing may, in the context of stem cells, have the same function. Judging by the overlap between H3K27me3 and Dichaete, it is possible that Dichaete and Polycomb may cooperate in particular genome regions, possibly through jointly altering overall chromatin architecture. An alternative hypothesis is that Dichaete and Polycomb silencing are mutually exclusive with a given genomic location occupied by Dichaete or Polycomb and the binding profiles reflecting differential occupancy in different cell populations in the embryo.

There were several other significant overlaps specific to the D_Berk dataset, including the markers of active transcription H3K4Me3 and H3K36Me3, and the insulator proteins BEAF-32 and CP190. The action of Dichaete may be context dependent - it is possible that it can bring about both activation and repression, depending on the cofactors it partners up with. In the midline, Dichaete enhances the activation

of *slit* (Ma et al., 2000), whereas *ac* is derepressed in the intermediate column of the neuroectoderm in *Dichaete* mutants (Overton et al., 2002). It is unclear whether the latter is a direct or indirect effect, but it opens the possibility that Dichaete may be capable of facilitating repression. From the data overlap it appears likely that, at least in the blastoderm, Dichaete frequently acts as an activator, a view supported by an analysis of pair rule gene expression in Dichaete loss and gain of function conditions (Russell et al., 1996). It is also possible that this is true throughout development, but that the interactions are transient, and/or stage and tissue specific, so may be lost in the broader time period used to generate the other datasets.

The association with insulator proteins is interesting, particularly in the case of BEAF-32, since the BEAF-32 binding motif was consistently found to be overrepresented in a subset of Dichaete binding peaks from all 4 available data sets, supporting the view that there is an interaction between the two (or at least, that their activity is required in the same genomic regions). It is therefore possible that Dichaete in some way affects or interacts with insulator functions. While insulator functions are poorly defined at the level of the whole genome, one model proposes they are involved in chromatin looping or architecture, intriguing in light of the known DNA bending activity of Sox proteins. The connection between Dichaete, BEAF-32 and regulatory hotspots (Roy et al., 2010) is also intriguing and potentially worth investigating further.

3.6.5 Open vs closed chromatin

It is likely that Dichaete, like other transcription factors, binds to available areas of accessible chromatin (Li et al., 2011). However, it is not yet clear what determines which areas of chromatin are open at any one time. On the one hand, it is possible that transcription factors bring about the chromatin accessibility state. Alternatively, TFs may just congregate at open regions once they become available. Perhaps the most likely explanation is that there is feedback between the two, where once the transcription factors associate with regions of the genome they help maintain an accessible state. Regardless of which of these explanations is correct, accessibility has been used before as a powerful predictive tool for modelling transcription factor binding (Kaplan et al., 2011). Therefore chromatin accessibility data can be helpful in understanding and predicting transcription factor action, even without a full understanding of the causal relationship between them.

To verify whether Dichaete tends to bind to open chromatin regions, genome-wide DNaseI chromatin accessibility data was downloaded from the Berkeley *Drosophila* Transcription Network Project (Thomas et al., 2011). Time course data was available for embryonic stages 5, 9, 10, 11 and 14. These data were converted to Genome Release

5 using the UCSC LiftOver tool (Rhead et al., 2010). Cooccur software was then used to compare the significance of overlap between the accessible regions and Dichaete binding. It is ideal to match the developmental stages of the embryos used to generate the accessibility data as closely as possible to the development times used for the ChIP-chip studies. Therefore, for an initial screening, the D_Berk (2-3 h, stage 4-5) data was compared to the stage 5 accessibility data, the D08 data (0-8 h) was compared to a union of stage 5, 9, 10 and 11 data, and the 0-12 h Dichaete datasets (D_Dam and D12) were compared to a union of all the available accessibility data.

In all cases, there was found to be significant overlap between Dichaete datasets and the relevant DNaseI datasets (p-value < 0.001). The percentage of overlap was found to be very high - between 90% and 98% of Dichaete intervals from the different datasets overlapped with DNaseI intervals. This confirms the expected hypothesis that Dichaete, like other transcription factors, is consistently found in open chromatin regions during development.

While some regions of chromatin appear to be accessible throughout development, many regions exhibit distinct and dynamic accessibility profiles during different stages of embryogenesis. Given the highly significant overlap with Dichaete, and also the predictive power that these types of data have for other transcription factors (Thomas et al., 2011), it is possible that the accessibility time course data could be used to computationally estimate what genes Dichaete is targeting at different points in development. Indeed this approach may be generally helpful for transcription factors where ChIP-chip data is only available for relatively broad time periods.

3.7 Conclusions

3.7.1 A systematic data comparison

The initial difficulties with comparison of the datasets were encountered very early in the data processing stage. It is known that using different data analysis pipelines can result in very different output results (Johnson et al., 2008). However, what the parameter sensitivity analysis concluded is that, even using the same pipeline, it is still possible to get highly divergent results from the same dataset by altering the analysis parameters. Also, while intuitively it makes sense that using the same parameters for all the datasets would reduce variability, in practice it seems that, for example, different normalization methods are optimal for different datasets, depending on the data profile. It is also possible that different peak finders and analysis pipelines are more appropriate for different data types - with, for example, some being better at recognising small sharp peaks, and others being more suited to recognising broad regions of binding. While

it is true that using different analysis methods for different datasets is introducing an unknown amount of variability, it was nonetheless concluded that matching your data analysis methods to suit your data type, rather than using a fully standardised pipeline, is optimal for the purpose of extracting the maximum biological signal from the data. However, this does also mean that independent validation becomes essential. Having multiple matched datasets is therefore recommended - it makes it possible to optimize parameters, and extract high confidence binding regions from the data, while making it easier to filter out the noise from the signal.

The consistency of replicates within individual binding studies presented is good, and the high overlap between 3 out of 4 datasets suggests that these are high quality and genuinely reflective of Dichaete binding sites in vivo rather than products of noise. The data quality, particularly in the case of D_Berk, but also to a great extent in the case of D_Dam, appears to be excellent. The datasets are both extensive, showing thousands of peaks across the genome, and specific, exhibiting various features expected of Dichaete targets. The D08 and D12 datasets seem to do well in terms of specificity, showing a lot of overlap with the other datasets, but the signal appears to be much weaker and consequently the sensitivity is much lower compared to the other two datasets. This is particularly evident in the case of D12, though the overlaps imply that the binding sites are genuine, the dataset only captures a few hundred Dichaete binding sites, compared to thousands captured by D_Berk and D_Dam.

The DamID data shown here is clearly of a comparable resolution to ChIP-chip datasets, showing a high degree of overlap in terms of both the genomic binding coordinates and the genes hit. The high overlap between the D_Berk and D_Dam datasets and their target genes is particularly encouraging since these are different experimental techniques used over different embryonic time periods, yet for the most part they give remarkably consistent results for most types of downstream analysis. It therefore appears that ChIP-chip and DamID are very much comparable and complementary, and can be used in unison as powerful tools for identifying genome-wide binding patterns.

The most interesting analysis results, including a high quality Dichaete binding motif, seemed to arise from the D_Berk dataset, probably because of the tight developmental window of the experiment - this is useful to know in general for ChIP-chip experimental design. However, it is also possible that some of the increased data quality of D_Berk compared to D08 and D12 is due to a better quality antibody and/or more optimized ChIP-chip conditions used by the laboratory. It is highly encouraging that all 4 datasets appear to have good specificity - this opens up the possibility of using all 4 datasets jointly to create a core high confidence set of Dichaete intervals, that would be both highly specific due to independent validation and highly sensitive due to the large quantity of data available.

The interval ranking method used by BDTNP only gave good results for D_Berk data - possibly because the data is sharper and more time-specific, meaning that the quantitative information from the ChIP-chip was more meaningful. However, based on the results for the other datasets, the use of this method is not widely recommended. On the other hand, differentiating between target genes based on whether they have single or multiple binding peaks (a method similar to the one used by Choo et al., 2011) was only briefly explored, but since it seemed to successfully pick out an interesting network of genes, it seems worth testing out further in the next chapter.

The extremely high overlap of all the datasets with open regions of chromatin (as assessed by a DNaseI assay) is as predicted from previous work associating TF binding with chromatin accessibility, but is nonetheless interesting as it seems to suggest that chromatin accessibility is one of the main determinants of Dichaete binding locations. As such, it should be possible to deconvolve Dichaete binding data into binding at specific stages, according to whether the chromatin was accessible at the time or not. Time series experiments are extremely interesting and informative, but also difficult, expensive and time-consuming. As we see from the D_Berk data, various features such as the Dichaete motif, and the tendency to bind to exons, were only possible to glean from the data with a tight developmental time. However, in practice, perfect time courses are not always available. Deconvolving data for different developmental times based on chromatin accessibility profiles, if proven to be reliable, could potentially provide a cheap and easy alternative when time-course experiments are not possible to perform. It is also worth noting that the data quality is not solely determined by the specificity of the time period used - while the D_Berk data does reveal some features that the other datasets do not show, the D_Berk and D_Dam datasets were in many respects very similar, which seems to suggests that it is possible to have datasets for broad developmental time windows, that are both sensitive and specific.

There were also some differences between the analysis outcomes for the different datasets. In particular, the GC profile of the surrounding sequence varied highly from dataset to dataset. A further look at the source of the differences concluded that the GC profiles of the bound regions and surrounding genome sequence was highly variable, and that perhaps the usual method of looking at the mean is in fact inadequate for this analysis in general. Looking at the whole range of data points was much more informative, and perhaps a more elaborate statistical approach would make this analysis more worthwhile. However, looking at the differences in profiles raises the question of how consistent GC profiles are in general, though there might plausibly exist sub-types of GC profiles that are in some way biologically relevant. Similar differences were found with the analysis of phastCons scores, so potentially some of the same conclusions apply.

Some differences between ChIP-chip and DamID can most likely be accounted for because DamID requires the presence of a non-randomly distributed sequence (GATC) that it methylates, whereas ChIP does not. However, figuring out the way in which this affects the results is not straight-forward - what it is likely to mean is that certain areas of the genome might be invisible to the DamID technique, due to the absence of the appropriate sequence in those regions. Since restriction enzymes are then used to isolate it, this also means that the fragment lengths should be fixed rather than continuous, which may also affect the outcome of the experiment (possibly shifting the peak centres). Analysing the consequences of these differences is beyond the scope of this work, but would make an interesting future project.

3.7.2 Dichaete biology

Previous investigations identified a number of developmental functions and potential target genes for Dichaete. However, it has only recently become possible to gain a genome-wide view of Dichaete action during *Drosophila* embryonic development. The exciting thing about looking at genome-wide data is the ability to look for broader patterns, not possible to glean from small scale experiments on individual genes.

From the binding data, it is apparent that the action of Dichaete is broad: ChIP-chip and DamID both indicate that Dichaete has thousands of binding sites spread throughout the genome. About a third of the binding sites are located in the more 'traditional' regulatory regions outside genes, whereas two thirds are located within genes, mainly within introns. This perhaps should not come as a surprise, as the verified site of Dichaete action within the *slit* CRM is located in an enhancer within the first slit intron (Ma et al., 2000). Dichaete does seem to show a tendency to bind just downstream of transcription start sites, suggesting a possible role in facilitating PolII pausing or enhancer action, but there is also generally a broad spread of Dichaete locations in relation to genes, suggesting a range of functions during transcriptional regulation. The overlap with H3K4me suggests that Dichaete binds to enhancer elements, and perhaps its DNA bending properties are used to bring more distant elements together into physical proximity, allowing a variety of regulatory functions to take place from there.

A distinct Dichaete binding motif was found, which matches previous experimental data. However, broader patterns in GC content surrounding the binding sites are less clear, and in fact seemed mixed or contradictory in different datasets. Minor groove binding proteins, which Dichaete is one of, do not have direct access to bases, and instead they rely on the differences in the electrostatic properties of the DNA backbone. While Dichaete does appear to have a clear sequence preference, it is likely that, in

principle, it can bind to any piece of DNA with its specific preferred sequence being more likely because of a higher affinity. Therefore, as well as being determined by the sequence specificity, the Dichaete binding patterns at different stages may primarily be determined by chromatin accessibility, and so the GC content patterns for Dichaete would vary in concordance with the way GC patterns of regions of open chromatin would, explaining the variability and some of the unusual graph patterns seen.

Looking at other sequence motifs found in Dichaete binding intervals reveals some interesting potential regulator partners. For Dichaete's role in segmentation, there are Kni, Hb and Bcd, a set of TFs known to regulate, as Dichaete does, the primary pair-rule genes. In terms of nervous system development, some proteins that appear interesting are Ey, Ttk, Hb, Grh and Vnd. The presence of SuH suggests a connection to the Notch signalling pathway. Also, the presence of the BEAF-32 binding motif in all data sets, as well as the overlap found with the BEAF-32 binding data, suggests a possible role in insulator function and a connection with regulatory hotspots.

3.7.3 Conclusions

All 4 Dichaete genome-wide datasets were found to have an encouraging degree of specificity and therefore will be used together in the next chapter, for generating a core high confidence Dichaete binding dataset. Because DamID is a method independent from ChIP-chip, regions of binding found using both methods can be considered highly rigorous. At the same time, because of the large amounts of data available, it is possible to build up a sensitive and extensive profile of Dichaete binding in early development.

Some methods for analysing the data further were also explored. Using the presence of accessible regions as a rough guide to which binding sites are present during which developmental times, seemed like a promising approach. The use of clusters of binding as an indicator of active regulation was also thought to be promising.

The preliminary analysis of Dichaete biology based on binding data yielded some promising starting points, including potential cofactors, an extensive catalogue of potential target genes, a high quality Dichaete motif and several possible hypotheses about the nature of Dichaete action. All of these will be explored further in the next chapter.

Chapter 4

Dichaete high confidence intervals

4.1 Introduction

As discussed in the previous chapter, chromatin immunoprecipitation (ChIP) is a powerful technique for studying in vivo protein-DNA associations. However, it is also known to be noisy and to generate false positive results, even when the experiment is designed accurately and a sufficient number of replicates are performed. There is potential to use DamID as an independent validation method to generate high confidence binding data, though this combined approach has not been used much in practice to date, perhaps in part due to time and cost constraints.

The early comparisons of ChIP and DamID suggested that ChIP gives significantly better resolution and that therefore the two techniques are not particularly suited to being used as complementary tools (Holland et al., 2005). However, subsequent comparisons suggest that the binding profiles are much more comparable than probe-based correlations suggest (Figure 4.1) and can be used for independently verifying each other (Negre et al., 2006; Moorman et al., 2006; Tolhuis et al., 2006). The evidence presented in Chapter 3 supports this conclusion.

High throughput binding assays are a common tool used for a number of exciting genomics projects, such as the large scale international ENCODE (Feingold et al., 2004) and modENCODE (Celniker et al., 2009) projects. The success of such projects and the biological insights drawn from them depend on the reliability of the experimental data generated. One of the things becoming apparent from such studies is that many transcription factors have thousands of binding sites across the genome (Li et al., 2008; MacArthur et al., 2009). However, there can be considerable variability between different IPs for the same transcription factor, as well as variability in results obtained

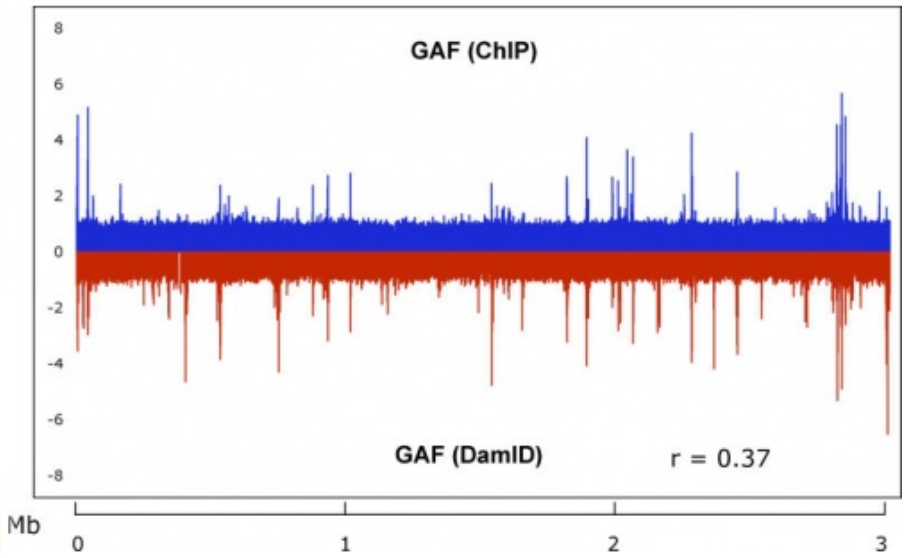

Figure 4.1: ChIP vs DamID comparison performed by Negre et al. (2006). The graph shows GAF binding in the *Adh* region obtained by ChIP at the embryonic stage (upper panel in blue) and with DamID in cultured Kc cells as described in Sun et al. (2003) (lower panel in red). Pearson's correlation coefficient between the two distributions is shown at the bottom of the graph.

using different analysis pipelines (Johnson et al., 2008). The parameter sensitivity analysis in the previous chapter highlights this problem.

It is therefore virtually inevitable that a proportion of the 'thousands' of binding sites represent false positives caused by experimental noise or biases associated with different analysis parameters. Increasingly, a number of independent datasets are becoming available for the same transcription factors, providing the potential to integrate data and produce more stringent, reliable genome-wide binding sets. While this can be done with different ChIP-chip datasets to increase reliability, ChIP-chip and DamID together are particularly well suited to the purpose, since they are completely independent experimental methods. This chapter explores the integration of ChIP and DamID data to generate a set of very high confidence intervals for Dichaete binding during embryogenesis.

Furthermore, of the thousands of binding sites, it is believed that only a subset of them are 'functional' - i.e. have an active effect on the transcription of associated genes. However, at present it is not possible to reliably differentiate between functional and non-functional binding. While a simple solution to this problem may not currently exist, this chapter explores potential strategies for filtering binding data to obtain a more meaningful collection of functional associations.

4.1.1 Chapter overview

One of the aims of this project was to develop a method for integrating different binding datasets. This chapter provides an overview of the method used and the reasoning behind it. It presents the Dichaete high confidence core binding interval dataset, followed by an analysis of the biological conclusions that can be drawn from it. I also present some ways in which the dataset can be filtered further to extract more biological information from it.

4.2 Combining datasets

When looking at peaks found at different FDR values, it is unclear which overlaps are the most meaningful to use for further analysis. Certainly, when looking at an individual data set, stringency might win out over inclusiveness, so an FDR of 1% would be preferable to an FDR of 25%, due to the latter introducing a larger number of false positives. However, it is not as intuitively clear whether this is the case when using more than one data set. For example, selecting overlapping peaks at random from two different FDR 25% peak sets, the chances of landing a 'double false positive' are 1/4 * 1/4, i.e. 1/16, so the effective FDR of two combined FDR 25% data sets is 6.25% (probability 1/16). Thus fusing 2 relatively non-stringent sets results in a data set with a reasonable FDR. On the other hand, merging an FDR 1% dataset and an FDR 25% dataset yields an FDR of 0.25% (1/100 * 1/4 = 1/400, i.e. 0.25 %) and merging two FDR1 data sets is extremely reliable (1/100 * 1/100 = 1/10000, i.e. 0.01%). The underlying assumption with these assessments is that false positives are randomly distributed, which may not be the case if all the experiments have the same experimental noise. However, combining ChIP and DamID data sets should satisfy the random noise assumption, since the noise from each technique is expected to be different. Based on this, it was thought that the FDR 1% to FDR 25% fusion of datasets would be optimal, as it gives a highly reliable FDR value, while not making it so stringent that the sensitivity of the experiment is compromised.

4.3 The high confidence Dichaete embryonic data set

4.3.1 Background

With the advance of high-throughput experimental methods, it has now become much easier to quickly obtain genome-wide binding data for transcription factors. However,

there is some concern about the reliability of these data sets. ChIP, the most commonly used experimental method for studying *in vivo* binding, is an inherently noisy experimental technique with the variability in data output from different analysis pipelines and from different starting analysis parameters compounding the issue. Thus, when using a single ChIP study there is a limit to our ability to determine how many of the identified binding sites, even at high confidence intervals, are artifactual. However, since many ChIP studies can be performed relatively quickly, and alternative techniques such as DamID are available, it is possible to independently verify binding locations using several data sources.

In the case of Dichaete, three different ChIP-chip data sets have already been published: from 2-3h embryos by the Berkeley Drosophila Transcription Network Project (MacArthur et al., 2009), and for 0-8h and 0-12h embryos by the modENCODE project(Celniker et al., 2009; Negre et al., 2011). As described in the previous chapter, I generated Dichaete DamID data for 0-12h old embryos, providing a fourth data set. Since ChIP and DamID are independent methods it is expected that, while both will generate false positives due to noise, they will be different in each case. Because of the different stages of development and different experimental methods used, some difference in identified binding intervals is expected, however there may also be a substantial overlap of direct targets. I therefore combined the available data sets described in the previous chapter into a single high confidence core Dichaete binding set. The acquisition and processing of the ChIP data were discussed in Chapter 3.

4.3.2 Creating the high confidence intervals set

Reiterating the results from the data processing performed in Chapter 3, the number of binding intervals in each dataset is presented in Table 4.1.

Data set	High stringency cut-off	Low stringency cut-off
D_ChIP_2-3h	6452	16501
D_DamID_0-12h	4342	12301
D_ChIP_0-8h	1366	6493
D_ChIP_0-12h	208	538

Table 4.1: Number of binding intervals in each data set. High stringency = FDR 1% for all except D_ChIP_0-8h (FDR 5% data was used because no peaks were found at FDR 1%). Low stringency = FDR 25%.

Based on the thought experiment presented above, FDR 1% peaks from each ChIP-chip dataset were compared to FDR 25% peaks of the DamID dataset and vice versa, fusing the resulting intersects together. In the case of the Dichaete 0-8h modENCODE data, no peaks were found at FDR 1%, so the FDR 5% data set was used instead. Once all

the overlaps were found, they were merged to generate the core set of Dichaete binding intervals. A diagram of the intersection and union schema is shown in Figure 4.2. Once this was completed, it was found that some of the resulting intervals were very small (as little as 3bp long) and were not considered meaningful overlaps. Therefore, as a final step, the intervals were filtered to remove intervals that were shorter than 100bp, reducing the set to the final 6227 intervals in total.

Figure 4.2: A representation of how the Dichaete core set was created from the original datasets. To start off with, the DamID dataset was compared to each of the 3 ChIP datasets in turn, taking the intersections of the two FDR 1% to 25% comparisons. Next, the two FDR 1% to 25% comparisons were unioned. After that, the results from all 3 ChIP to Dam comparisons were unioned to create the final set (which was subsequently filtered to remove intervals smaller than 100 bp).

4.3.3 Data strengths and weaknesses

Because of the timings of the data sets used, the core Dichaete data set generated is specific to the first 12 hours of embryonic development, with a possible bias for registering more binding in the 0-8h period (the Dichaete 0-12h ChIP data has very few peaks). It is expected that during the first 12 hours of development, binding sites change while the current data only provides a static picture of binding sites. In order to construct precise and finely dissected transcriptional networks, more detailed and precisely timed binding data is required. However, we argue that the amalgamated binding data presented here is rigorous and independently verified by two different experimental techniques, and while it is unlikely to give exact answers as to the precise mechanisms of transcriptional regulation during particular points in development, it is nonetheless a solid starting point for exploring the genome-wide action of Dichaete, identifying potential co-factors, identifying target genes, and analysing causal factors such as chromatin architecture and their potential impact on Dichaete action.

4.4 Genome-wide view of Dichaete core intervals

4.4.1 Binding interval summary

There were 6473 intervals in the unfiltered data, of which 6227 overlapped by more than 100bp and these were used for further analysis. The mean interval length is 1.6kb (1.2kb median), while the maximum length is as high as 16.8kb, with the majority of the peak intervals less than 3kb long. Thus these data show a relatively narrow binding

interval as would be expected for a transcription factor. The distribution of interval sizes is shown in Figure 4.3.

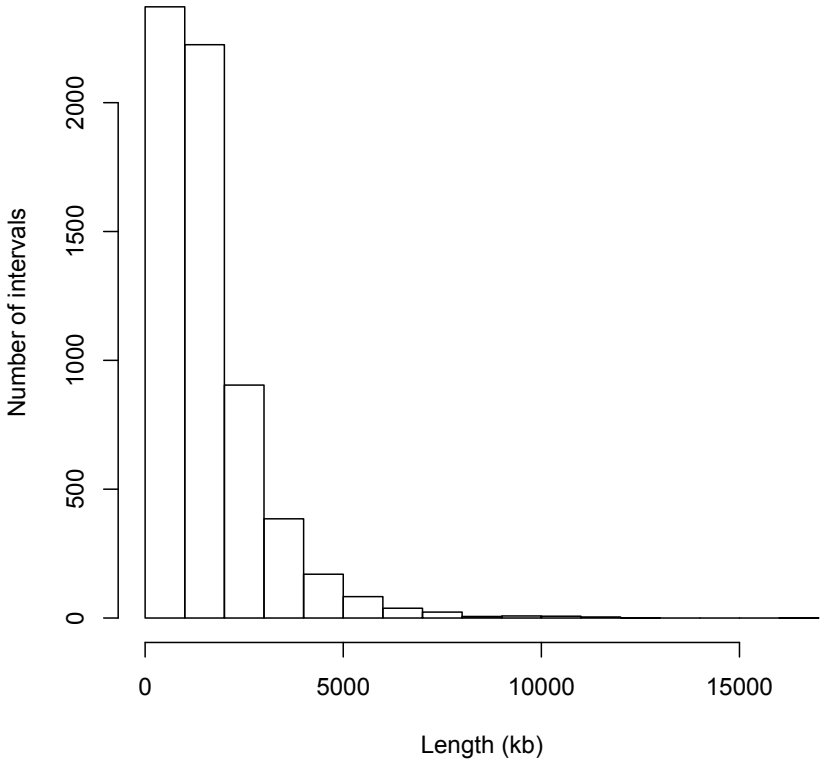

Figure 4.3: Distribution of interval sizes in the core Dichaete data set

4.4.2 Genes hit

The method described in Chapter 3 was used to identify genes hit. For each of the intervals, all directly overlapping genes were counted as gene hits. If no direct overlaps were found for an interval, the closest gene within a 10 kb region upstream or downstream was identified and registered as a gene hit. The most up to date version of the complete list of *Drosophila melanogaster* genes (annotation release 5.26) was used and the FBgn numbers were used as identifiers. The 6227 intervals were associated with 4750 genes (about a third of the total genes present in the *Drosophila* genome) and of these, 4043 genes were directly overlapping with a binding interval. As discussed in Chapter 3, two out of four datasets suggest that Dichaete has thousands of binding sites across the genome, and this is comparable to results for other developmental transcription factors. This method of finding gene associations was not used for benchmarking using the Twist and Snail data, because it was important for the gene hits there to be

unambiguous. However, this method is recommended for doing gene associations for an exploratory genomics analysis because, while it is less stringent, it takes into account the possibility of binding sites being bound to enhancers far upstream or downstream and still controlling gene expression.

Out of the 4750 genes associated with Dichaete binding intervals, a large proportion (1224 genes, p-value=2.1E−111) are involved in developmental processes, more specifically: in segmentation (165 genes, p-value=1.8E−22), the development of imaginal discs (348 genes, p-value=4.5E−63), the nervous system (558 genes, p-value=2E−87), sensory organs (292 genes, p-value=2.1E−43), the respiratory system (137 genes, p-value=1.9E−18) and muscle structures (158 genes, p-value=3.6E−14). Since Dichaete is not expressed in the mesoderm, the latter is somewhat unexpected. It is possible that there is either an overlap of genes involved in mesoderm development that perform other functions during development, or that Dichaete is repressing some mesoderm genes in the developing neuroectoderm. A significant number are implicated in the regulation of gene expression (583 genes, p-value=3.9E−53), and 294 of the genes are annotated as encoding sequence specific DNA binding transcription factor activity (p-value = 4.7E−46) with a particular enrichment for proteins containing homeodomain-like domains (107 genes, p-value = 5.1E−18). There are also a number of genes involved in the cell cycle (380 genes, p-value=6.4E−20) and cell fate commitment (201 genes, p-value=8.3E−39), perhaps connected to the role of Dichaete in neural stem cells.

The genes associated with Dichaete binding were evaluated against the subset of GO terms provided by the Panther (Mi et al., 2005) GO slim list using Ontologizer (Bauer et al., 2008) for analysis and visualisation. Resulting p-values were corrected for multiple testing using the Benjamini-Hochberg method. The GO enrichments are shown for molecular functions (Figure 4.4) and biological processes (Figure 4.5). Looking at the molecular functions of the genes, it is clear that many of the genes associated with Dichaete play a role in some form of transcriptional regulation, variously through DNA, mRNA and chromatin binding. Another highlighted major role is protein modification, with various structural and post-translational molecular activities featured.

In terms of the biological processes, the enrichment of genes involved in cell cycle regulation and apoptosis is exciting, because it hints at an involvement of Dichaete in stem cell pluripotency, reminiscent of the role of Sox2 in mammals. Genes involved in localization and cell communication may be the facilitators of some of the developmental signals regulated by Dichaete. It also seems possible that Dichaete plays a role as a global regulator, controlling a host of genes involved in transcriptional regulation and modifying chromatin architecture to facilitate broad regulation of biological processes at relevant times during development. The enrichment of genes involved in protein modification supports the results from the GO enrichment for molecular functions,

and suggests that this may be another level of regulation that Dichaete is indirectly involved in. The enrichment of developmental processes is as expected, and the enrichment of genes for neurotransmitter secretion may be connected to a role for Dichaete in later stages of nervous system development.

The list of targets in the high confidence binding data suggests that Dichaete is a global regulator, with thousands of binding sites present across the genome. Judging by the gene associations, the high confidence dataset appears to be both extensive and specific. However, the scale of the data presents a challenge: while the associated 558 genes involved in nervous system development, for example, probably provide an extensive catalogue of Dichaete nervous system targets, it is not immediately clear where to go from there. While such lists are a useful resource for anyone doing further experiments on specific Dichaete targets, or trying to validate pre-existing hypotheses, jumping from a list of thousands of genes to biological insight is challenging to say the least. For this reason I attempted to perform some filtering and deconvolution steps in an attempt to narrow down the list, allowing a more precise biological focus. The further analysis is described later in this chapter.

4.4.3 Genomic features hit

In the absence of raw binding profiles for a dataset that is a union of other datasets, the centres of the binding intervals were used as an approximation of where the Dichaete binding sites are likely to be. While this may not necessarily be the most accurate method, it was considered to be a valid pragmatic approach. The results were very similar to the ones found using the individual Dichaete datasets in Chapter 3 with the majority (59%) of the binding intervals found in genic regions. Also, as found with the individual Dichaete binding sites, most of the binding intervals mapping to genic regions are found within introns (Figure 4.6).

In the previous chapter it was found that the size of the introns containing Dichaete binding intervals was significantly larger than the average intron size for the rest of the genome. I verified that this trend holds true for the Dichaete core binding dataset. Transcripts with the greatest number of introns were used for the analysis and alternative transcripts for the same genes were ignored to avoid counting the same introns multiple times. The same trend is still present, with the mean length of the introns containing Dichaete binding intervals (8.3 kb) significantly higher than the 840 bp genome average (p-value $<$ 2.2E-16, Wilcoxon rank sum test). As discussed in Chapter 3, long introns frequently contain regulatory regions, so the mapping of Dichaete to long introns may be an indication of a tendency to bind to enhancers.

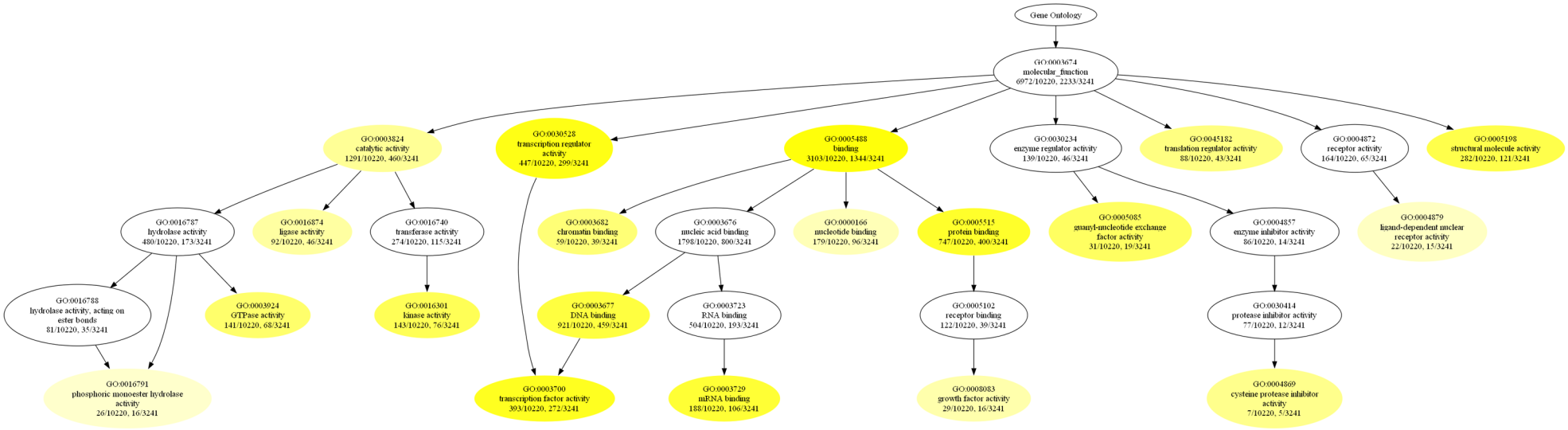

Figure 4.4: GO enrichment molecular functions for genes associated with Dichaete binding (p-value < 0.05), with the brighter yellow colour corresponding to an increase in significance

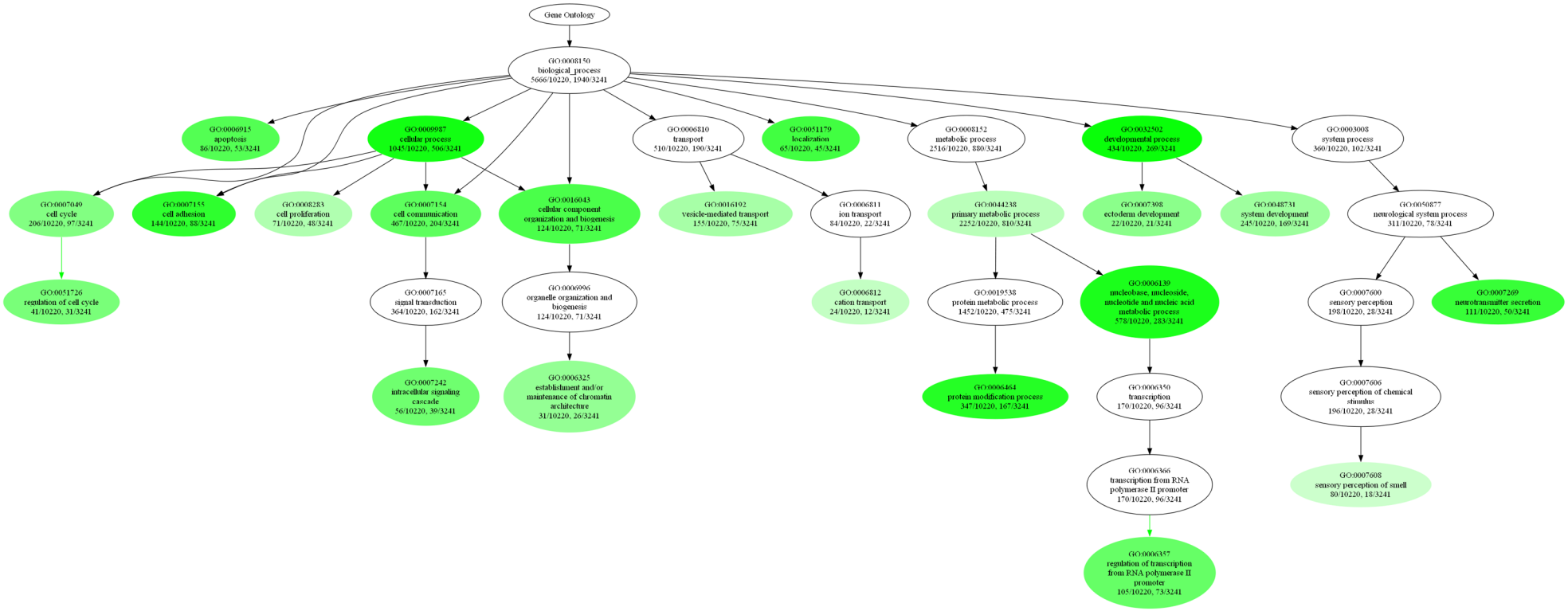

Figure 4.5: GO enrichment biological processes for genes associated with Dichaete binding (p-value < 0.05), with the brighter green colour corresponding to an increase in significance

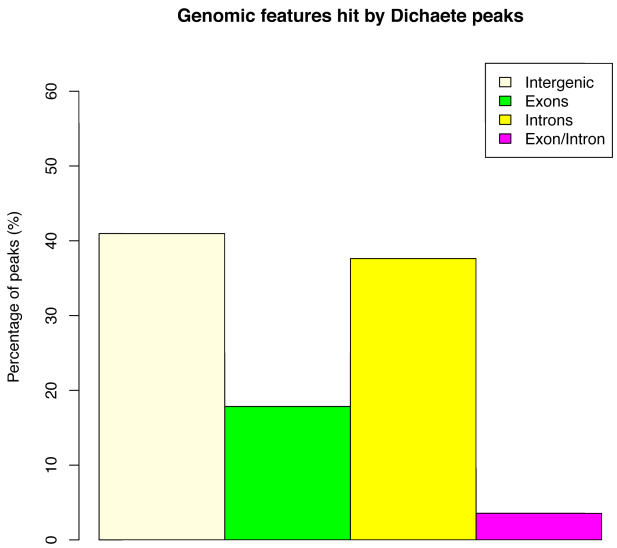

Figure 4.6: Genomic features associated with Dichaete binding sites

4.4.4 Distance from transcription start site

The middle of each core binding interval was used as an estimate of where the Dichaete binding site is likely to be and the distances from transcription start sites within a 5 kb range were analysed (Figure 4.7). Looking at the positioning of the closest transcription start site, there appears to be a trend for the majority of Dichaete binding sites to be located very close to transcription start sites, as well as both upstream and downstream. Even looking at all transcription start sites present within the 5 kb interval (to ensure the result is not accidentally caused by the high gene density of some parts of the *Drosophila* genome), the trend for Dichaete binding sites to be located within 1 kb upstream or downstream of a transcription start site remains.

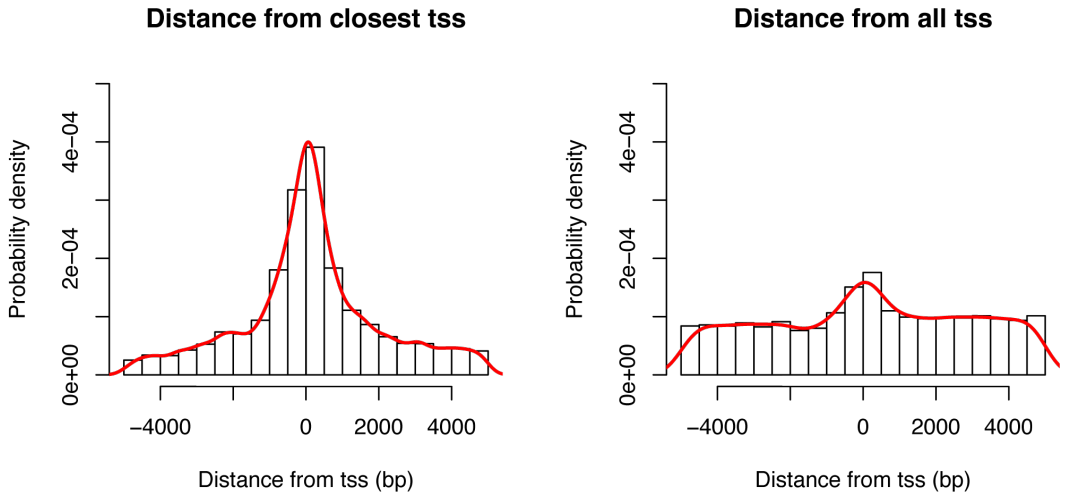

Figure 4.7: Distance of binding sites from the nearest transcription start site (left) and from all transcription start sites within 5 kb (right) expressed as a probability density function

4.4.5 GC content

The GC content around the binding sites was investigated (Figure 4.8). The first interesting observation is that there is a trend around the estimated binding site, which is different from that of the surrounding sequence. This is encouraging, as it suggests that while picking out the centres of binding intervals is arguably not the most precise way of estimating binding site location, it nonetheless seems to provide a reasonable approximation since the result is similar to those obtained using the more meaningful highest scoring windows described in Chapter 3. The actual trend resembles that found for the D_DamID dataset in Chapter 3 and may indicate that there is an increase in flexibility on the outsides of regions bound by Dichaete.

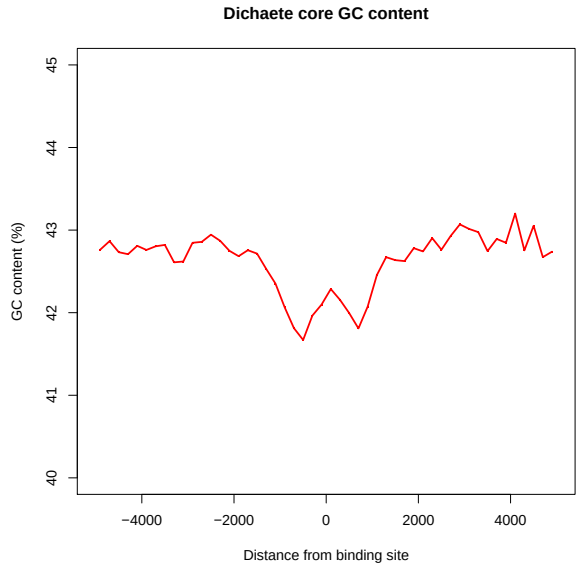

Figure 4.8: Genomic features associated with Dichaete binding sites

4.4.6 Transcriptional pausing

The potential role of Dichaete in transcriptional pausing was discussed previously and I examined this with the core dataset. Using Cooccur, the PolII 0-12h embryo ChIP-chip data set from modENCODE was found to significantly overlap with the Dichaete core 0-12h data set (p-value = 0.001). It is therefore still a plausible hypothesis that Dichaete, at least in part, regulates transcription by modulating polymerase pausing. However, further conclusions are difficult to draw without performing more specific experiments to test the hypothesis.

4.4.7 Enhancers

To verify whether it is plausible that Dichaete interacts with enhancers, the list of 997 known Drosophila cis-regulatory modules from the REDy database was compared to the Dichaete core dataset. It was found that 533 enhancers (53.5%) overlap with Dichaete binding intervals. Compared to the results from Chapter 3, this is slightly less than the number overlapping with D_Berk (549) but more than the number overlapping with DamID (427), and significantly more than the number overlapping with D08 (98) and D12 (3). The number of overlaps with CRMs seems to correspond fairly linearly to the number of binding intervals present in each dataset in the first place. The high percentage of overlap between the Dichaete core dataset and known enhancers supports the hypothesis that Dichaete acts by binding to enhancer regions in the genome.

4.5 Filtering binding data for functional analysis

As previously discussed, the collection of binding sites found in any one genome-wide binding assay can often be large. Even assuming that the dataset is perfectly reliable and that there are no false positive intervals present, that still leaves the question of which binding sites are likely to be actively regulating transcription, and which ones might be due to non-functional binding. Here I explore a few strategies for filtering and deconvolving the Dichaete dataset to focus in on more biologically meaningful hits.

4.5.1 Motif scanning

The Dichaete binding motif identified from ChIP-chip data in Chapter 3 was used to scan the binding intervals for the presence of Dichaete binding motifs. The software nmscan (part of the NMICA package) was used at threshold 5 (a medium level stringency threshold). The results for each interval were compared to the results for 10000 random motif simulations (obtained by reshuffling the columns of the existing Dichaete binding motif) and the intervals with a p-value < 0.05 were considered to have a significant enrichment of the Dichaete binding motif.

A total of 906 intervals passed this test and were found to be significantly enriched. Considering that the test used for statistical significance is extremely stringent (many of the randomly generated motifs are very similar to the genuine one), I consider this to be a high number of positives. The intervals in question were found to be associated with 917 genes. The list of genes included a number of Hox genes, as well as *ind* and *comm*, *Notch*, *kuz*, *engrailed*, several POU domain proteins, the nervous system

transcription factors *bancal*, *eyeless*, *SoxN* and *prospero*, and the known Dichaete target *slit*. Under the threshold used, 5354 of the original 6227 intervals (86%) were found to have at least one binding motif, but it should be noted that the number of motif hits relies heavily on the chosen threshold and should not be considered a reliable indication of the presence of actual binding sites.

The overall GO enrichment results are quite similar to the ones found for the entire Dichaete core dataset, with many of the same terms enriched. However, in some ways, the list enriched for the Dichaete binding motif is more specific - while 1224 out of 4750 (26%) genes in the original dataset are involved in developmental processes, 326 out of 917 (36%) are in the motif enriched set. The same is true for genes involved, for example, in nervous system development (18% compared to 12%) and cell fate commitment (8% compared to 4% in the original dataset). It is possible that these particular categories of genes are more prominent than other ones because of a gene length bias - the tendency of developmental genes and transcription factors to be long genes and/or to have a lot of genomic space around them, combined with the increased likelihood that binding intervals will be associated with long genes, compared to short ones (Nelson et al., 2004; Taher and Ovcharenko, 2009). However, what we are looking at here are extremely high confidence Dichaete intervals - ones generated from multiple independent experiments, which also contain a statistical overrepresentation of the Dichaete binding motif. The gene association method used is sensible, therefore saying that a proportion of the genes associated with the intervals are nervous system genes is merely a factual statement. However, while nervous system genes really are bound, what might be happening is that other categories of genes are, for example, short genes masked by the presence of direct Dichaete overlaps with very large genes in the same region. It is therefore possible that some additional relevant categories are being missed out here.

It is inevitable that some real functional binding sites are lost when using site enrichment (as opposed to the presence of a Dichaete site) as a filter for binding intervals, and while it appears to somewhat improve the signal to noise ratio it also filters a large number of potentially interesting targets. I concluded that filtering for enrichment of Dichaete binding sites is therefore not suitable for an exploration of genome-wide patterns of Dichaete binding. However, in cases where stringency is desired, for example when planning experimental verification of specific targets, this may well be a helpful filter to use as it is likely to more precisely zoom in on likely functional Dichaete binding sites.

In terms of the genomic features hit, the proportion of introns hit remained roughly the same as the average for the whole set (it went from 38% to 36%), but the proportion of intergenic hits increased from 41% to 48% and the proportion of exon hits decreased

from 18% to 13%. This is perhaps an indication that looking at the enrichment of the Dichaete binding motif is correctly identifying active regulatory regions, which are more likely to be situated in introns or intergenic regions than in exons.

4.5.2 Single vs multiple binding clusters

The analysis strategy which compares genes with single vs multiple associated binding sites was successfully used by Choo et al. (2011), who found that, while the genes associated with single Ubx peaks showed very few GO enrichments, the genes with multiple peaks had a set of specific GO enrichments similar to the ones found for the full set of Ubx bound genes. The approach was briefly described in Chapter 3 and identified some interesting associations in the genetic network, but here I describe a more in-depth analysis with the entire set of Dichaete associated genes. An initial screen found that of the 4750 genes associated with Dichaete binding, the majority (3476 genes, 73%) were associated with a single binding peak. Of the remainder, 682 genes (14%) had 2 associated binding peaks, 259 (5%) had 3 and 333 genes (7%) had 4 or more, up to a maximum of 25 associated binding peaks.

Hotspots of Dichaete binding

Only 22 genes were identified with 10 or more associated binding peaks (Table 4.2) and these were individually examined in detail, with the interesting potential targets described below. In the case of the gene with 25 binding sites associated, FBgn0052816 (CG32816), it was found to be an unannotated gene spanning a very large region (chrX, 201,609:364,670), encompassing the entirety of the Achaete-Scute region on chromosome *X*. Since Achaete-Scute Complex (AS-C) genes are known to be regulated by Dichaete from previous studies (Zhao and Skeath, 2002), it is likely that the binding peaks correspond to AS-C regulatory regions. This gene was therefore removed from further analysis. A second gene associated with 19 associated binding peaks - CG1677 (FBgn0029941) - comprises a long transcript (around 125 kb), which runs across a very gene rich region containing 13 other smaller genes, including the transcription factor *brk*. In this case, CG1677 was left as part of the gene list, but the analysis was also run without it to see which of the encompassed genes associate with these Dichaete binding peaks. Another potential ambiguous assignment is the gene *kirre*, which is adjacent to *Notch* in the genome, and while the 11 Dichaete peaks may be associated with *kirre* they may also be part of the Notch regulatory region. These observations highlight the general problem of associating transcription factor binding peaks with the genes they regulate.

Other approaches exist for assigning target genes based on transcription factor binding sites. For example, Sandmann et al. (2006) used BGDP in situ data to assess the gene expression patterns of the genes close to binding sites, as well as looking at distance from binding sites and differential expression in mutants. Unfortunately, this approach was considered not to be appropriate for Dichaete. Firstly, Dichaete can act context-dependently both as an activator and a repressor, so an absense of a similar expression pattern might in fact be a sign that Dichaete is doing its job, rather than that the gene in question is not a direct target. Secondly, as discussed, a number of the genes Dichaete targets are large genes with complex expression patterns, so it is unlikely that the presence of Dichaete binding sites is a sole determinant of their expression pattern, making it difficult to predict what a 'promising' expression pattern would look like. Some approaches also use functional annotations of surrounding genes to pick out the likely candidates. However, while this approach might be appropriate for picking out specific targets in a particular tissue, it is definitely inappropriate for a genomic analysis where the function of the genes identified as targets is used to draw conclusions about the roles of the transcription factor in question - picking out nervous system genes, and thus concluding that the transcription factor in question is involved in nervous system development, is definitely based on circular logic. This is why, for the purposes of this analysis, a simplistic approach for gene assignment was used - it was considered to be the best option given what we know about Dichaete and what the purpose of the analysis was.

Encouragingly the gene *slit*, which is known to be a direct Dichaete target in the developing midline, is one of the multiply bound genes and has 11 Dichaete binding peaks associated with it. This means that it is possible that multiple Dichaete binding may be an indication of direct regulation, bearing in mind that this class represents a tiny fraction of the genes associated with Dichaete binding.

A number of Hox genes were also found to be associated with multiple Dichaete binding peaks, including *Antennapedia* (18 peaks), *Ubx* (14 peaks) and *Abd-A* (10 peaks), as well as the Hox cofactor *homothorax* (15 peaks). All of these are interesting potential regulatory targets, with involvement in important developmental processes such as segmentation and neurogenesis. Since several genome-wide binding analyses of Ubx and Homothorax have recently been published (Choo et al., 2011; Slattery et al., 2011) and binding data is available, a detailed comparison of the relationship between Dichaete and Hox proteins would be interesting to explore in the future.

Other transcriptional regulators are also present in this gene subset. The transcription factor *jing* (associated with 19 peaks) is amongst other things involved in various aspects of central nervous system development and tracheal development (source: Flybase (Gelbart et al., 1999)), and there are no other genes in the same region, making

Flybase ID	Name	Number of D hits
FBgn0037153	olf413	10
FBgn0026239	gukh	10
FBgn0000014	abd-A	10
FBgn0024308	Smr	10
FBgn0261570	CG42684	10
FBgn0086364	rdx	11
FBgn0028369	kirre	11
FBgn0085403	Rapgap1	11
FBgn0003731	Egfr	11
FBgn0003425	sli	11
FBgn0016797	fz2	12
FBgn0051163	SKIP	12
FBgn0086901	cv-c	12
FBgn0040071	tara	13
FBgn0003944	Ubx	14
FBgn0001235	hth	15
FBgn0011224	heph	15
FBgn0250867	CG42238	16
FBgn0260642	Antp	18
FBgn0029941	CG1677	19
FBgn0086655	jing	19
FBgn0052816	CG32816	25

Table 4.2: Genes with 10 or more Dichaete binding sites

jing an interesting potential direct regulatory target. The gene *taranis* is involved in wing disc dorsal/ventral patterning and transcription maintenance and has 13 associated Dichaete binding peaks.

It is intriguing that a number of the genes associated with a large number of Dichaete binding peaks are transcriptional regulators involved, amongst other things, in wing development. Since the most famous Dichaete dominant phenotype is a defect in wing development (Russell, 2000) perhaps this should not come as too much of a surprise. While Dichaete itself is not normally expressed in the wing disc, many of the transcription factors that are expressed in the wing also have other functions in different tissues during development and it is likely that Dichaete normally regulates these genes elsewhere. In this view, ectopic expression of Dichaete in the wing results in defective development by missregulating genes it normally regulates elsewhere in the fly.

The list also features a number of receptors and other proteins involved in signalling pathways. The Wnt-receptor Frizzled 2 is involved in various aspects of nervous system development. The gene *cv-c* encodes a GTPase activator with a range of developmental roles, including activity at the syncytial blastoderm stage and an involvement in synaptic transmission and hindgut development. The growth factor receptor Egfr is active in the blastoderm and involved in the development of sensory organs. The gene *heph* is involved in the Notch signalling pathway.

There were several other developmentally important genes, such *gukh*, involved in protein localization and axon guidance. The apoptosis gene *roadkill* might be connected to maintenance of pluripotency vs differentiation, and *Smrtr* is a cell cycle regulator.

The results of this crude analysis look very encouraging. Virtually all of the genes associated with multiple Dichaete binding peaks are developmentally important, and the known Dichaete target *slit* is included in this set of genes. Of course, since most of the identified genes have multiple roles at various stages during development, the structure of their cis-regulatory regions is complex. It likely includes multiple CRMs, with each CRM containing binding of multiple factors, and the complete cis-regulatory region associated with the gene potentially containing multiple binding sites for each factor. These genes are likely to be direct regulatory targets of Dichaete and could be investigated further in studies focused on individual Dichaete targets. The strong connection with Hox genes that this analysis suggests is an interesting feature which may constitute an interesting topic for a future study.

GO enrichment

For this part of the analysis, the genes associated with only 1 Dichaete binding site (a total of 3474 genes after the analysis modifications described above) were classified as 'single', and the genes with 4 or more Dichaete binding sites (a total of 340 genes) were classified as 'multiply bound'. The genes with 2 or 3 Dichaete binding sites were considered unclassified and were therefore left out. As discussed in the Materials and Methods chapter and the later part of this chapter, GO enrichments are biased because longer genes tend to both belong to particular categories, and get hit more often by chance (Nelson et al., 2004; Taher and Ovcharenko, 2009). The p-values associated with these enrichments are therefore likely to be significant overestimates.

The GO enrichment profiles for the two groups of genes are very distinct. In the case of the genes bound by single peaks, while a mixture of different genes is present due to the large number of genes, the terms that are significantly enriched are cellular processes (1845 genes, p-value = 1.3E−28) such as cell cycle processes (236 genes, 4E−11), cellular metabolic processes (1163 genes, 3.6E−14), microtubule cytoskeleton organization (154 genes, 4.7E−10) and nucleic acid metabolic processes (488 genes, 9E−8). While 651 genes involved in developmental processes do feature on the list, the p-value is comparatively high (0.000016), a reflection of the fact that this is a relatively small subset (19%) of the whole gene list.

On the other hand, the multiply bound gene list is massively enriched for genes involved in developmental processes (196 genes, 1.9E−63), composing the majority (58%) of the entire gene list. 123 of the 340 genes are involved in nervous system development (2.3E−55), 97 of the 340 are involved in imaginal disc development (p-value = 2.4E−53), 139 in cell differentiation (p-value = 3.5E−52) and 205 in biological regulation (p-value = 6.4E−47), with 106 regulating transcription (p-value = 6.5E−42).

Potential analysis biases

There are several potential explanations for the observed patterns of gene enrichment. One explanation is that multiply bound genes are a distinct subset from single peak genes and are more likely to be directly regulated. Binding information is insufficient to determine whether this is the case or not, but this hypothesis will be revisited in Chapter 6, where binding data is compared with gene expression data from Dichaete mutants. The same pattern could be random, due to the fact that long genes cover a relatively large fraction of the genome and could therefore overlap many Dichaete binding peaks by chance. This might be the case for a few of the examples described in detail above, specifically those that have transcripts encompassing a region densely populated with smaller genes. However, this was not the case for most of the genes

observed. Another possibility is that the differences in GO enrichment are not due to specific subgroups in the Dichaete binding set but reflect the fact that longer genes in the Drosophila genome have particular properties, for example, they encode developmental regulators with complex cis-regulatory modules, sometimes including long introns containing enhancers. A further analysis of Dichaete binding and gene length was performed in order to shed further light on this possibility.

The first noticeable trend is that, overall, genes associated with Dichaete binding are significantly longer than genes in the rest of the genome (p-value $<$ 2.2E-16, Wilcoxon rank sum test). The mean length is 9481 bp for Dichaete associated genes, and 3916 bp for genes in the rest of the genome. This does not appear to be a by-product of a bias introduced by a small number of Dichaete bound genes, the maximum gene length for the rest of the genome is 396 kb and is much larger than the 172 kb average for the Dichaete associated class. The medians show the same trend as the means (the median is 3304 bp for Dichaete bound genes and 1556 bp for the rest of the genome). It was then investigated whether similar trends held true for different subgroups of genes bound by Dichaete. Each of the Dichaete groups, genes associated with 1, 2, 3 or multiple (4 or more) Dichaete peaks, was compared with the remainder of the genome (genes not associated with any Dichaete binding). It was found that all of the subsets contained genes significantly longer than the genome average (p-value $<$ 2.2E-16 in all cases, Wilcoxon rank sum test). However, the difference in size ranged from only slightly larger (average gene length 5.5 kb compared to the 3.9 kb genome average) in the case of genes with single Dichaete binding peaks, to the much larger: 12.2 kb average for genes with 2 peaks, 19.5 kb average for genes with 3 peaks and 35.3 kb average for genes 4 or more peaks. This is an almost exact linear trend, with on average 1 Dichaete binding peak occurring every 6 kb of gene length. The numbers of genes in each group are quite different, with the majority being associated with just one Dichaete binding peak and a much smaller total number of genes being associated with more than one peak. In order to make the comparison of the gene size distributions easier, they are shown as probability densities (Figure 4.9). It is clear both from the graphs and from the statistics that there is a correlation between an increased number of Dichaete binding peaks and an increase in gene size. However, there is also a lot of variance within this - the scatterplot in Figure 4.10 visualises this more clearly. While there does appear to be a trend, the correlation found was 0.53.

Further analysis

While the results of this analysis are clear, the conclusions are less certain. First of all, it is not obvious that the Dichaete peaks associated with long genes are really functionally relevant for gene expression or whether they have various chromatin func-

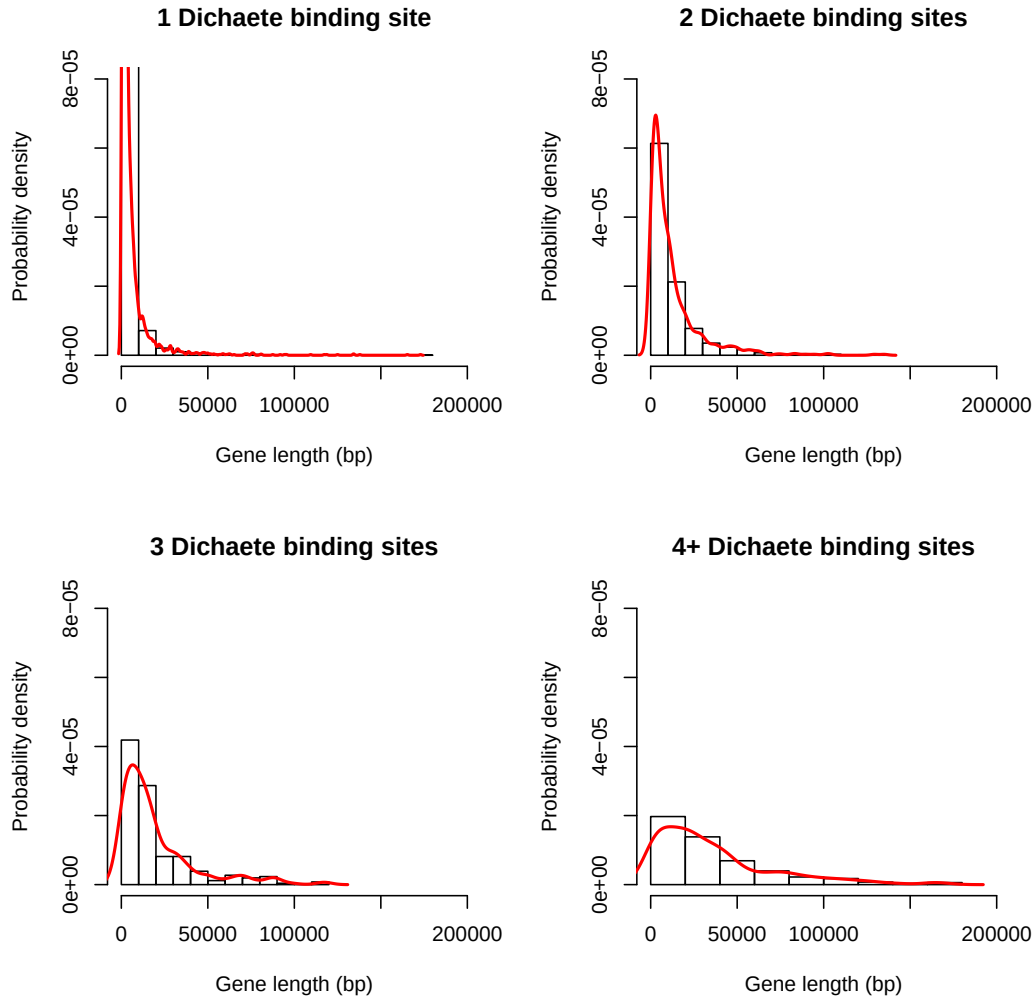

Figure 4.9: Distribution of gene sizes based on the number of Dichaete binding sites associated with the genes

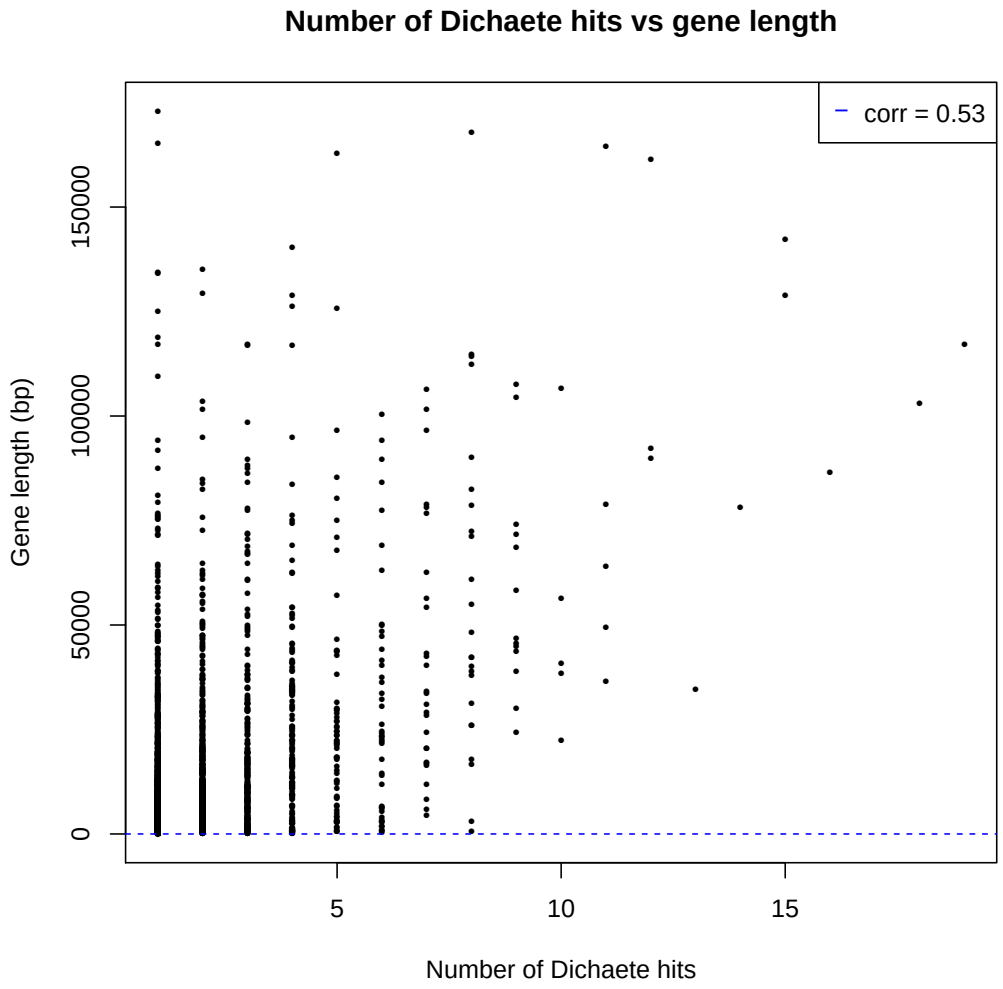

Figure 4.10: Number of associated Dichaete binding sites vs gene length

tions in different parts of the genome and large genes just happen to associate with more Dichaete peaks by chance. While the functional annotations of genes with a large number of Dichaete peaks indicates that this is unlikely to be the case for all of them, however, for genes situated in gene dense areas, a random association may be the case. Several multiple peak genes are the only annotated genes in a particular genome region and are therefore considered to be the most likely direct Dichaete targets.

An obvious question is whether there is something specific about the pattern of Dichaete binding to genes with multiple peaks or whether there are particular general features about the way long genes are regulated in the genome. It is plausible that, for example, long genes may have enhancer regions situated within introns and generally more complex cis-regulatory modules. In some developmentally important genes, the transcribed region itself may not be particularly long but there are relatively large upstream and downstream regions that contain a variety of complex regulatory modules, examples here include the primary pair-rule genes *eve*, *hairy* and *runt*. It is therefore possible that the different GO enrichments observed with the single vs multiple peak genes are due to the different modes of regulation for dynamically expressed temporally specific genes, which require complex regulation, compared to single peak genes with more mundane housekeeping functions that are likely to have more basic regulatory sequences. Indeed, this was found to be the case - Nelson et al. (2004) report that genes with complex functions and dynamic expression are flanked by significantly more non-coding DNA than genes with housekeeping functions and simple expression patterns. It is therefore possible that the correlation between gene length and number of binding peaks is caused by a subset of genes, whose complex regulation is mediated through intron-based regulatory regions. Alternatively, what may also be contributing to this is that not all binding regions are functional, and the longer genes are, the more likely they are to be bound at random. It is therefore possible that some of the increase in the frequency of Dichaete binding intervals associated with longer genes is a chance occurrence.

From a Dichaete-specific perspective, a few more questions can be investigated. One is whether the extent of correlation between multiple binding and gene length found here is Dichaete specific, or whether it is typical of most transcription factors. Looking at a few other transcription factor binding site profiles and rerunning the same analysis could potentially answer this. In addition, looking specifically at total intron and total exon length and assessing the correlation with the number of Dichaete binding peaks could indicate whether the association is due to regulatory regions within introns. Similarly, looking in more detail at where exactly within the multiply bound genes the Dichaete binding peaks are located, could provide similar indications. A brief analysis of these factors was thought to be instructive.

Binding patterns of other TFs based on gene length

To see whether the pattern of clusters correlated with gene length was Dichaete specific, several other embryonic genome-wide data sets were analysed. In order to be comparable with Dichaete data, the binding data sets were selected on the basis of the number of intervals present at FDR 1%, which were then used for the analysis. The data sets selected were the Kruppel 0-8h set (union of a Berkeley 2-3 h set and a modENCODE 0-8 h set, 4553 intervals), the Chinmo 0-12 h set (modENCODE data, 7054 intervals), the Daughterless 2-3h data set (Berkeley, 5534 intervals), the hairy 0-8 hour data set (union of a Berkeley 2-3 h set and a modENCODE 0-8 h set, 5145 intervals), the Medea 2-3 h data set (Berkeley, 5458 intervals) and the Twist 2-3 hour data set (Berkeley data, 7674 intervals). These data sets are also representative of a number of different DNA binding domains and regulatory functions (Table 4.3).

Transcription factor	DNA binding domain	Regulatory function
Kruppel	C2H2 zinc finger	A-P early gap, neurogenesis
Chinmo	C2H2 zinc finger	Neurogenesis
Daughterless	bHLH	D-V maternal
Hairy	bHLH	A-P pair rule
Medea	SMAD-MH1	D-V zygotic
Twist	bHLH	D-V zygotic

Table 4.3: Transcription factors used for comparison with Dichaete

The results (Table 4.4) clearly show that the trend Dichaete displays is very much a typical pattern, at least compared to other embryonic transcription factors with a similar number of binding sites. This leaves the question of whether the observed association occurs by chance purely because the genes are long, or whether it reflects the fact that genes of different lengths have a distinct set of functional properties.

Transcription factor	0	1	2	3	4+
Dichaete	3.9	5.5	12.2	19.5	35.3
Kruppel	3.7	7.5	17.2	25.6	44.8
Chinmo	3.4	5.7	15.3	23.2	43.9
Daughterless	3.7	6.1	14.3	23.2	41.2
Hairy	3.7	7.3	14.9	23.7	42.5
Medea	4.2	5.0	12.3	25.8	41.8
Twist	3.3	4.9	11.0	21.3	42.9

Table 4.4: Mean gene lengths (kb) associated with different numbers of binding peaks

Dichaete binding and intron size

To assess whether clusters of Dichaete binding specifically associate with genes containing long introns, the number of Dichaete binding peaks associated with a gene

was compared to the total exon and total intron length of each gene. The exon and intron coordinates from the longest transcript for each gene were used. In order to get the most up to date gene models, relevant genome annotation release 5.26 data was downloaded from FlyMine. Small discrepancies in size compared to the Dichaete gene lengths are present, due to some of the gene models changing between the two analyses, however, it was thought that the effect of this on the overall trend was negligible.

Dichaete binding sites	**0**	**1**	**2**	**3**	**4+**
Total average intron length	2.3	3.3	9.3	16.4	31.3
Total average exon length	1.7	2.3	3.1	3.7	4.4

Table 4.5: Mean total intron and exon lengths (kb) associated with different numbers of binding peaks

It was found that an increased number of Dichaete binding peaks appears to coincide with a significant increase (Wilcoxon test, p-values between 0.0003 and 2.2E−16) in both the intron size and the exon size (Table 4.5). However, it is clear from the data that, while the exons do increase in size slightly, most of the difference in size comes from a large increase in total intron length. Therefore, while it is possible that multiple Dichaete binding sites are present at random in larger genes, it is also possible that the presence of multiple binding peaks can be explained by a tendency for Dichaete to bind at multiple locations within intron regulatory regions: for example, intronic enhancers.

Gene length and gene function

One possible hypothesis is that longer genes have distinct functionality, for example that developmental genes are more likely to be longer because of complex regulatory regions sometimes present within their introns, whereas ubiquitously expressed housekeeping genes are on average shorter because their regulatory regions do not need to be as complex. This is indeed exactly what was found by Nelson et al. (2004) in their study on the subject. Here I present a short piece of analysis looking only at gene length, with the purposes of relating the conclusions of the Nelson et al. (2004) study to the currently available genome and functional annotations. The gene length information for all genes in *Drosophila melanogaster*, genome annotation release 5.26, was downloaded from FlyMine, with the gene length considered to be the length of the longest transcript.

In order to look at differences in GO enrichments of the different lengths of genes, the set of all *Drosophila* genes was divided into subsets by length. The first idea was to partition the gene set into equally divided size ranges - e.g. length from 0 to 50 kb, from 50 to 100 kb and so on. However, looking at the size distributions of the genes

revealed that the distribution is extremely skewed, with most genes falling within the 0 - 5 kb range, with the much larger genes (going up to a maximum of 396 kb) being extreme outliers. In order to have equally sized subsets of genes for comparison, the data was instead split into 4 groups using quartiles, with the ranges being 31 bp to 992 bp for the 'very small' genes, 993 bp to 1914 bp for the 'small' genes, 1915 to 4373 bp for the 'medium' genes, and 4374 to 396000 bp for the 'large' genes. This resulted in 4 subsets containing between 3719 and 3725 genes each. There is nothing specifically biologically meaningful about these cutoff points but any cutoff based on gene length alone will always be arbitrary. GO enrichments were obtained through Flymine and all p-values were corrected for multiple testing using the Holm-Bonferroni method. Any p-values of less than 0.01 after correction were considered significant.

In the case of the very small genes, about 20 GO terms were found to be enriched, most of them connected to macromolecule metabolism/biosynthesis (631 of the genes were found to be involved in macromolecule metabolic processes, p-value 5.6E−7). Interestingly, the very small subset also shows an enrichment for genes involved in gene expression (430 genes, p-value 3.1E−25) and chromatin organisation (118 genes, p-value 2.8E−29). Virtually all of the chromatin organisation genes in question were histones, which are required ubiquitously and tend to contain no introns. Many of the genes involved in gene expression are associated with mRNA processing and translation, processes that, while essential, are also likely to be required ubiquitously and not in any particularly temporal or tissue-specific fashion.

The small gene set showed comparatively little GO enrichment, with only 10 GO terms passing the statistical cut-off. Many of these genes are involved in metabolic processes (1291 genes, p-value 1.2E−11); other enriched functions are connected with either sensory perception (125 genes, p-value 1.1E−14) or spindle elongation (42 genes, p-value = 0.00005). Similarly, the medium length genes are not enriched for many specific GO annotations. A total of 13 GO terms were enriched, with p-values that, while significant, are not particularly low. Again, metabolic processes feature (1520 genes, p-value 0.001) and within this group, metabolic processes of small molecules appear to be particularly enriched (289 genes, p-value 2.6E−6). The other enriched biological processes are connected to transport (497 genes, p-value 9.1E−6) and the establishment of localization (511 genes, p-value 1.5E−5).

The large genes, on the other hand, show completely different trends compared to the other 3 groups. Two striking features immediately apparent are an abundance of enriched GO terms (over 400 biological processes were found to be enriched) and the high degree of specific enrichment, with some of the p-values as low as 2E−130. A large number of the genes are involved in developmental processes (1116 genes, p-value 4.9E−126), specifically in the development of anatomical structures (1024 genes, p-

value 2.3E−130). Many of the genes are also involved in the regulation of biological processes (1285 genes, p-value 6E−113). This group contains a large number of genes involved in cell differentiation (667 genes, p-value 9.3E−97), nervous system development (514 genes, p-value 7.1E−91) and various other specific developmental processes. In conclusion, in agreement with the Nelson et al. (2004) study, it was found that gene length is strongly correlated with gene function: smaller genes are more likely to have general housekeeping roles and ubiquitous expression, and large genes are more likely to have very time-specific and developmentally important roles, associated with dynamic expression patterns. This pattern is likely to be connected to the role of introns in the precise temporal regulation of transcription.

Analysis conclusions

The overall analysis of genes associated with Dichaete binding found that some genes in the genome are associated with single Dichaete binding sites, while others are associated with clusters of binding peaks. There is a clear difference in the functional enrichment between these two classes, with the multiply bound genes mostly associated with developmental processes. An analysis of gene length showed that the number of binding peaks associated with a gene is strongly linked with gene length, with roughly 1 Dichaete binding peak on average for every 6 kb of gene length. An analysis of other transcription factor binding data revealed that this pattern is typical of a number of different developmentally important transcription factors. It was also found the majority of the increase in target gene size can be accounted for by the size of the introns present within these genes. As previously found by Nelson et al. (2004), and redone here using the current GO enrichments, it is clear that there is a massive difference in the functional profiles of long genes, which were found to be mostly developmentally important genes, or genes involved in biological regulation.

Taken together, the conclusion of these analyses is as follows. Genes that are developmentally important, or for other reasons need to be expressed dynamically and at precise times in the organism, tend to be large, and specifically, they tend to have long introns. In addition, large genes tend to attract multiple binding peaks for a variety of transcription factors, including Dichaete. The likely explanation for this is that the presence of a large number of regulatory modules, possibly reusing the same transcription factors in different configurations, is capable of executing precise and dynamic temporal expression programmes. In contrast, shorter genes tend to be involved in housekeeping processes and are associated with a smaller number of binding sites, presumably because their regulation is more straightforward and likely to be dependent upon relatively simple regulatory elements. While Dichaete appears to be one of the transcription factors associated with multiple binding peaks within large genes, it is

likely that this is a more general trend in the regulation of gene expression during development. However, the other possibility is that longer spaces in the genome are more likely to be bound purely by chance. This does not appear to be the likely scenario for the majority of the genes in question - while one or two binding intervals could be non-functional, and located in particular genome areas by chance, this seems less likely to be the case for genes that are associated, for example, with 11 different Dichaete binding intervals. Nonetheless, this is a possibility worth bearing in mind. Perhaps an ideal method of gene association would take this into account, and for example involve a Monte Carlo simulation to assess the significance of the gene to binding interval association.

Addressing GO enrichment bias

As pointed out to me by Dr Casey Bergman, the fact that function correlates with gene length introduces a bias into GO enrichment analysis, which a number of papers describe in detail (Stanley et al., 2006; Taher and Ovcharenko, 2009; McLean et al., 2010). A consequence of this difference in regulatory locus size described by Nelson et al. (2004) is that, since high complexity genes have much bigger non-coding regions around them, it is much more likely that by chance, binding peaks will be found in these regions. Since particular 'types' of genes, such as developmentally important transcription factors, tend to have more complex gene expression patterns, and thus also bigger regulatory loci, this introduces a bias in terms of the GO terms found to be significantly enriched for genes associated with binding peaks using standard analysis tools. Judging by the observation above about the increasing size of complex regulatory loci in more complex animals, it is likely that this bias gets worse in, for example, large mammalian genomes.

Taher and Ovcharenko (2009) studied this bias in the human genome, by randomly sampling non-coding genomic locations, and assigning their functions based on the location of the nearest gene. While with a random search like this, you would expect no specific over- or under-representations, they found that the random search yielded significant enrichments of categories such as 'nervous system development' and 'transcription factor activities'. This is not a problem for gene lists generated from gene expression experiments, as these are not biased according to the surrounding regulatory regions - all genes are in theory equally likely to feature, based on the presence/absence of mRNA. However, it's a major problem for functional analysis of genome-wide binding experiments such as ChIP, because there, the availability of 'landing space' for the factors in question can greatly bias the GO analysis. In such cases, it is important to correct for the bias, in order to get an accurate assessment of GO enrichment. It is worth noting in general that care should be applied when choosing appropriate statis-

tical tests, based on what type of experiment the data comes from in the first place. While a hypergeometric test, the statistic usually applied to obtain GO enrichment p-values, is appropriate for data from gene expression studies, it becomes entirely inappropriate for gene lists generated from binding data, where the assumption of equally probable discrete data categories is violated.

While the GO enrichment bias caused by differences in gene length is definitely important (and remarkably underreported in the literature on GO enrichment), actually correcting for it turned out to be non-trivial. The study by Taher and Ovcharenko (2009) developed the theoretical framework for correcting GO enrichment biases based on the location of genes in the genome, but this method would then need to be actually implemented as software for *Drosophila*, ideally also customising the input to account for the fact that different methods and maximum distances used for gene association assignment also impact the probabilities of different GO categories being represented. While this would definitely constitute interesting future work, due to time constraints it is beyond the range of this thesis. There were also two software packages already available - GONOME (Stanley et al., 2006) and GREAT McLean et al. (2010). While GONOME is a standalone package (including a web interface) that corrects for gene length bias and works for *Drosophila*, unfortunately it is extremely out of date and no longer maintained - the last version uses a genome release dating back to 2005. GREAT looks like an excellent package that is recent enough and uses solid methods, but is unfortunately currently only implemented for human, mouse and zebrafish. Thus, while not correcting for gene length as part of a GO enrichment analysis is definitely an error, the GO enrichment analysis in this thesis was still performed using the standard methods which do not correct for this, with the caveat that therefore, the conclusions that can be drawn from GO analysis in this thesis are relatively limited.

Specifically, while counting the numbers of genes in a gene list that have a particular GO term associated with them is merely a factual statement - e.g. out of 1000 genes, 400 are annotated as having 'nervous system development' functions - what GO enrichment analysis aims to do is pick out the functionally relevant GO terms by statistically testing whether they are over- or under-represented than would be expected given the size of the gene list and the proportion of all the genes in the genome that are annotated with the function in question. So, while 400 nervous system genes might seem like a lot, it might not be if, say, 12,000 out of the total 14,000 genes in the genome have nervous system functions - in fact, there would then be a significant underrepresentation of them in the gene list in question. Since the p-values might be skewed because of gene length, it is likely that the presence of developmental genes and transcriptional regulators is deemed much more significant than is actually the case, while for example housekeeping genes are more underrepresented. If the function of

a transcription factor was being determined for the first time purely on the basis of a gene list, this would be a huge problem. In essence, all transcription factors would end up supposedly involved in the same range of developmental processes, likely purely by chance. However, with studying Dichaete specifically, I have the good fortune of drawing on a number of years of phenotypic analysis Russell et al. (1996); Soriano and Russell (1998); Sanchez-Soriano and Russell (2000); Overton et al. (2002); Shen (2006); Maurange et al. (2008). So, for example, we know from phenotype-based experiments that Dichaete is involved in nervous system development - specifically in midline and neuroblast development. Therefore, starting from a global gene list for Dichaete and identifying the subset involved in aspects of nervous system development is inherently interesting because of what we know about Dichaete function, irrespective of whether 'nervous system development' as a term is overrepresented and underrepresented. With this in mind, GO enrichment analysis results are still presented throughout this thesis, with the caveat that the p-values are biased and should not be relied on as the sole basis for further work.

4.5.3 Deconvolution for different developmental times

It was found in the previous chapter, that there is an extremely high overlap between accessible regions of chromatin (as found by the BDTNP) and Dichaete binding. This is not unexpected, as literature on transcription factor binding suggests that the accessibility of chromatin, combined with sequence specificity, is one of the strongest determinants of transcription factor binding locations (Li et al., 2011; Kaplan et al., 2011). While the correlation does not resolve the question of what comes first, chromatin accessibility or transcription factor binding, it does open up the possibility of deconvolving binding profiles collected during broad time periods into more stage-specific views, by using the available detailed chromatin accessibility time-course data. Though unlikely to be as precise as a detailed time course experiment on specific transcription factors of interest (since chromatin accessibility profiles reveal where transcription factors might be binding rather than where they actually are binding at the specified time), this method nonetheless has the potential to save time and money, in situations where a detailed time course is not readily achievable.

Method

The DNase I assay accessibility data (Li et al., 2011) was downloaded from the BDTNP website, and converted to Genome Release 5 using the UCSC LiftOver tool (Rhead et al., 2010). Data were available for stage 5, 9, 10, 11 and 14 embryos, overlapping with the 0-12 h developmental time frame of the Dichaete core binding dataset. An

intersect between the Dichaete data and the accessibility data was found for each of the available stages, and used as an estimate of where Dichaete is likely to be binding in the embryo during that developmental time. Stages 5 and 9 were of particular interest, as this is when Dichaete is likely to be active in segmentation and neurogenesis, respectively. Similarities and differences in the binding patterns between the stages, and the potential associated target genes, were then studied.

Overlap found

First of all, the Dichaete core dataset was compared to the union of all available DNase I data in the same time period to check whether the same strong correlation observed with the individual Dichaete binding datasets described in Chapter 3 was present. This was indeed found to be the case - 90.4% of the Dichaete core intervals showed an overlap with the accessibility data (p-value = 0.0009, resampling based significance test, Cooccur (Huen and Russell, 2010)). The overlap is not perfect and potential reasons for this could be: the presence of false positives in the Dichaete data (considered less likely since the binding intervals detected by independently methods); the presence of false negatives in the accessibility dataset (possibly because the accessibility of some regions changed too quickly to be captured, or because the DNase I assay is not a perfect bias-free predictor of accessibility); the possibility that Dichaete does not always bind to accessible regions; or because the accessibility datasets do not cover the entire time period that the Dichaete dataset covers. The last reason was thought to be the most likely, however, if this were the case, the D_Berk dataset from Chapter 3 would be expected to have the highest overlap, due to being the most closely temporally matched to the corresponding DNaseI dataset. This was not found to be the case: the D08 and D12 datasets display the highest overlap (96% and 98%, respectively), with D_Berk and D_Dam both showing a 90% overlap. D08 and D12 have much fewer peaks than D_Berk and D_Dam, perhaps indicating that the strongest Dichaete peaks only appear in accessible regions. If the problem is in the datasets, it is more likely to be due to the incompleteness of the DNaseI datasets (single experiment, 2 replicates, FDR 5%) than due to false positives in the Dichaete dataset (multiple experiments, between 2 and 3 replicates at FDR 1% in at least one of the experiments). Nonetheless, the overlap found was considered to be high enough to proceed with the rest of the analysis.

Time specific Dichaete datasets

The Dichaete core dataset was split into 5 stage specific datasets, by taking the intersect of the Dichaete dataset with the DNaseI dataset at each stage. To start with, previously identified potential Dichaete targets (Shen, 2006), such as the genes in the Achaete-

Scute complex, were visually inspected to verify what the deconvolved data looks like (Figure 4.11). The patterns that emerge are interesting, with some regions being continually bound throughout the 5 different stages while others appear more dynamic. A pattern that seems to emerge from looking at this subset of genes, is that positions near the transcription start site are frequently continually bound, whereas sites that are further away tend to be more dynamic, perhaps priming a gene for expression, then regulating expression at particular stages more precisely via interactions with other factors at more distant cis-regulatory modules. While this is clearly too small a subset of genes to base solid conclusions on, it is nonetheless an interesting hypothesis.

Static vs dynamic Dichaete binding

First of all, the Dichaete binding peaks present at different stages (based on the presence of DNaseI accessibility tracks in that location at the appropriate stage) were categorised into static or dynamic binding sites. The five datasets were merged, with any clusters of partially overlapping intervals treated as one large interval for the purposes of counting. The intervals in the fused dataset with overlaps at each of the developmental stages were tabulated to indicate the presence or absence of each interval at different developmental times. The fused dataset contained a total of 8140 intervals, all of which were accessible and bound at least once during the 5 stages that accessibility data is available for. The number of intervals is greater than the original 6473, because some of the original intervals got split into a number of smaller intervals based on the location of accessible regions within them. These were then split up according to the total number of stages they were bound at (Figure 4.6).

Total stages interval is bound at	Number of intervals	Percentage of total intervals
1	1433	17.6%
2	947	11.6%
3	805	9.9%
4	1121	13.8%
5	3834	47.1%

Table 4.6: An overview of static vs dynamic Dichaete binding, as estimated from DNase I accessibility data for different embryonic stages. The data was available for a total of 5 different stages and the table shows in how many of those stages the interval in question is bound.

Overall, a large number of intervals (47%) were found to be 'static' - i.e. bound at all stages. The intervals bound at only 1 out of 5 (constituting 18% of the total number) were classified as 'transient'. The intervals bound in between 2 and 4 stages (35% of the total) were classified as 'dynamic' (Figure 4.7). I then looked for differences between

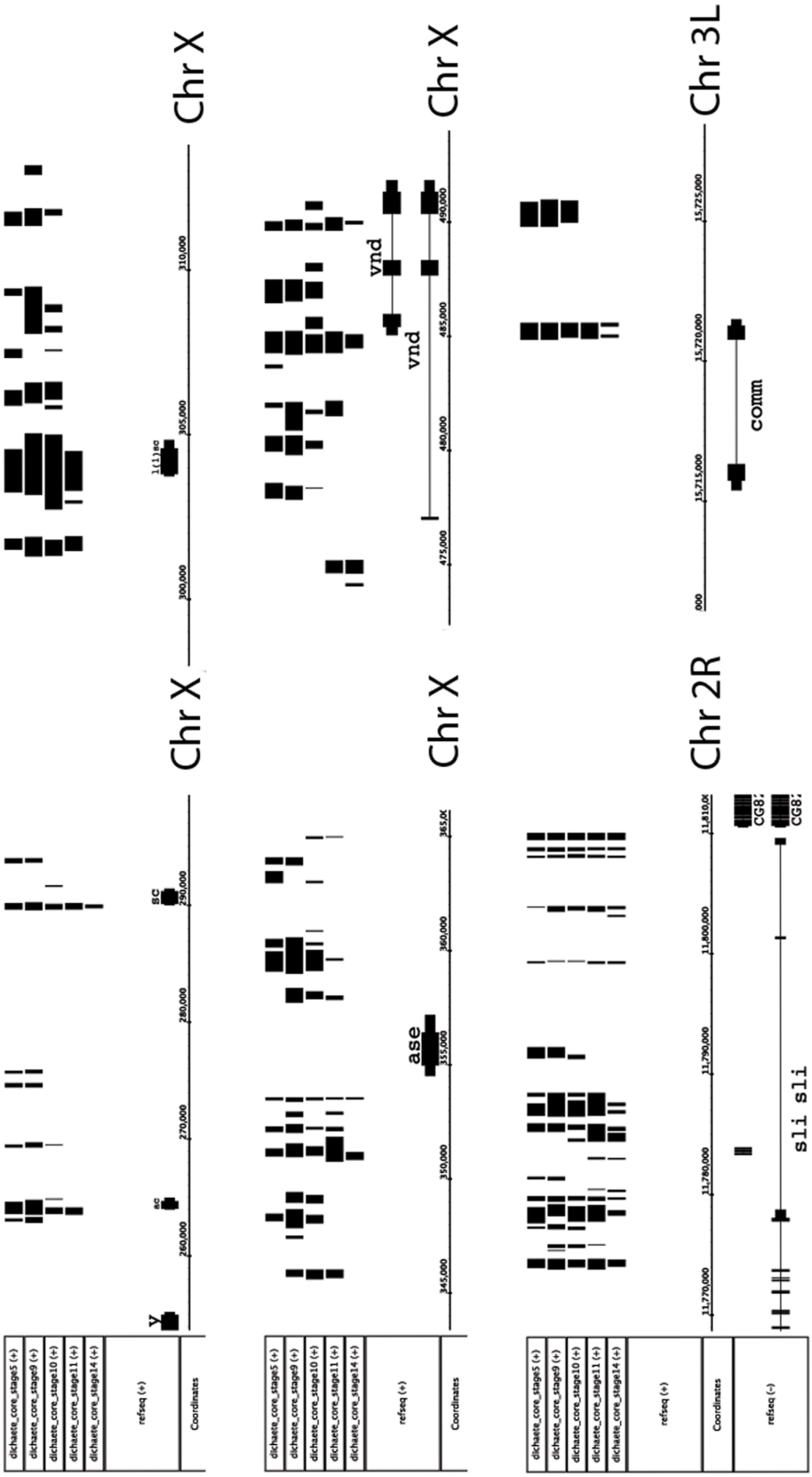

Figure 4.11: Dichaete temporal binding profiles. The diagram shows estimated Dichaete binding intervals for 5 different stages (stage 5, 9, 10, 11 and 14), with the location of Dichaete binding at each stage estimated based on the overlap with DNaseI accessibility interval locations during the stage in question.

the gene sets associated with each group, caluclating GO enrichments for each groups with Ontologizer, using the Parent-Child union method and the Benjamini-Hochberg correction for multiple testing.

Classification	Number of intervals	Percentage of total intervals	Number of genes hit
Static	3834	47%	3780
Dynamic	2873	35%	1557
Transient	1433	18%	899

Table 4.7: An overview of static vs dynamic Dichaete binding, as estimated from DNase I accessibility data. The intervals are classified as static (bound at all 5 stages), dynamic (bound during 2-4 stages) or transient (bound during 1 stage).

A number of genes were associated with more than one type of binding since they are close to multiple binding clusters (Figure 4.12). Very few genes were associated with transient binding intervals alone - instead they were frequently also associated with static or dynamic intervals, or both. The few genes only associated with transient binding show no significant enrichment and do not feature particularly interesting developmental genes. While the total set of genes associated with transient binding has a lot of interesting GO enrichments, including developmental processes (p-value $3.95E-59$, 355 genes) and regulation of gene expression (p-value $2.75E-31$, 175 genes), these enrichment classes are also found for genes associated with either static binding, dynamic binding or both. It is therefore possible that transient binding has a role in stage-specific regulation of particular genes, but it seems unlikely that transient binding is a unique feature. Rather, it is likely that transient binding sites act in unison with other more long-lasting binding regions to modulate gene expression.

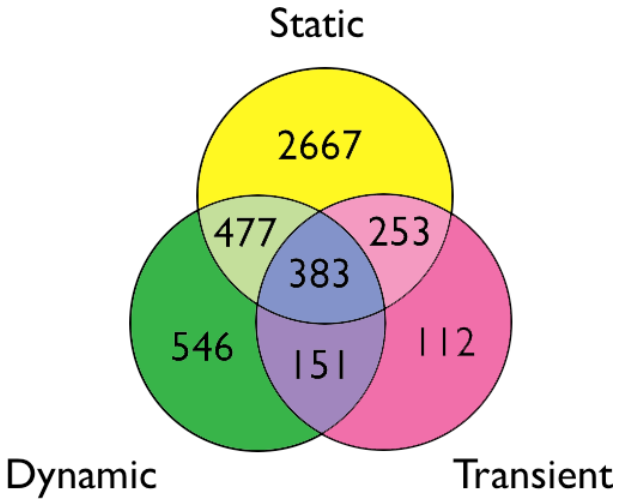

Figure 4.12: Numbers of genes hit by particular types of binding intervals.

The set of genes associated with static binding alone appears more interesting. It contains a range of genes involved in cellular (p-value 1.4E-79, 1521 genes) and metabolic

processes (p-value 1.04E-21, 1139 genes), and biological regulation (p-value 2.93E-22, 722 genes). This includes a range of cell cycle genes (p-value 2.31E-06, 251 genes), genes involved in various aspects of cytoskeleton organisation and genes involved with cell death (p-value 1E-4, 82 genes). While a large number of developmental genes are present, the enrichment for developmental processes is not particularly strong (p-value 0.01, 532 genes), and specific enrichments for particular developmental processes such as neurogenesis is notably absent in this set. It appears that static binding on its own is associated with basic cellular processes and is perhaps important in implementing some levels of Dichaete-facilitated regulation. However, it does not appear that this set is the causative factor in bringing about different developmental processes.

The genes associated with dynamic binding alone only display a small number of very weak enrichments, connected to localization (81 genes, p-value 0.42) and CNS development (15 genes, p-value 0.46), presumably because of the small number of genes featured in this group. However, the connection of these processes to development and differentiation is notable.

The set of genes associated with 2 or more different types of binding appears to be the developmentally interesting group since it is highly enriched for developmental processes (539 genes, p-value 8.6E-97) and biological regulation (562 genes, 4.5E-85), including the regulation of gene expression (253 genes, p-value 5.2E-37). It contains genes involved in signalling (271 genes, p-value 1.1E-34) and growth (96 genes, p-value 3.8E-29). It also contains genes involved in both embryonic (205 genes, p-value 4.6E-11) and post-embryonic development, and involved in specific developmental processes such as nervous system development (294 genes, p-value 1.7E-10). This finding makes sense since the presence of some form of temporally specific binding would be expected in order to facilitate the temporally specific expression of developmental regulators. An interesting observation is that a many of these regulators are associated with static Dichaete binding throughout development as well as multiple binding sites, some of which are dynamic. This raises several possibilities. The static binding sites may be effector sites that are ready to act on the nearby genes, but require particular cofactors, the presence of which is triggered by the dynamic binding sites. Alternatively, there may be some degree of non-functional binding, where Dichaete is just docking close to the genes that it will regulate later in development. Of course, these speculations rely on the datasets being considered accurate, and though they do provide an approximation of where Dichaete binding may be occurring at different developmental times, they are not actually from an experimental time course.

It was then investigated how the interval occupancy changes over different stages of development (Figure 4.13). It is interesting to note that the total number of binding intervals is fairly constant, around 6000 intervals in total during Stages 5-11, but then

drops off to approximately 4900 intervals during Stage 14, when Dichaete is thought to have less of an active developmental role. Interestingly, the numbers of static and transient binding sites remain about the same throughout the different stages, whereas the number of dynamic intervals changes dramatically, dropping from over 2000 to around 800 at Stage 14. It is possible that this is due to dynamic binding intervals being important as particular developmental triggers at set times, and then no longer being needed during later stages of embryogenesis.

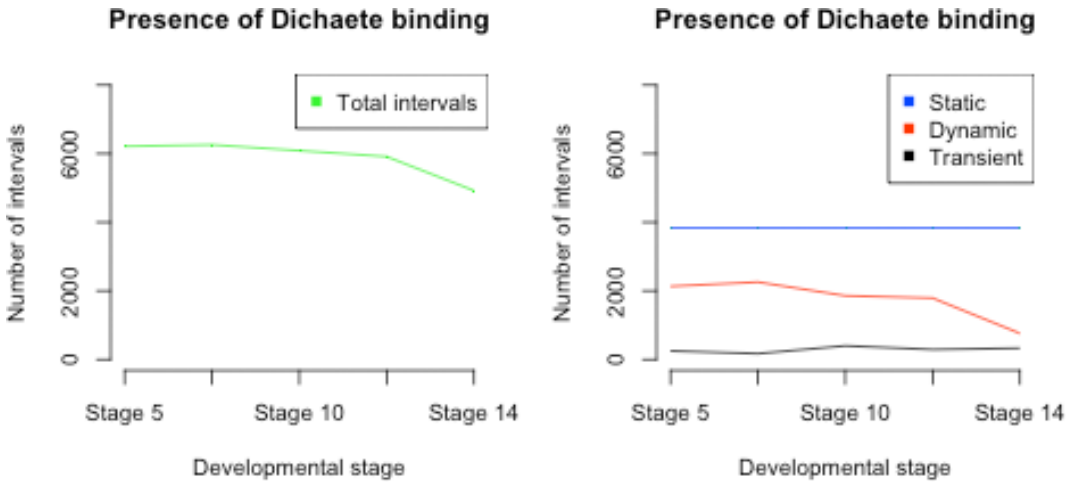

Figure 4.13: Overview of numbers of Dichaete binding sites. The total number of binding sites (left) drops off during later stages. When broken down by type of binding (right), the numbers of static and transient binding sites remain similar over time, whereas the drop in the number of dynamic binding sites causes the overall drop in the interval number.

In conclusion, the DNaseI accessibility data can be used for deconvolving ChIP-chip data into more temporally specific datasets, and the dynamically bound intervals might be of particular interest to study, as they may have the most significance in terms of developmental biology.

4.6 Conclusions

The main contribution of this chapter was the development of a method for integrating multiple genome-wide binding datasets to generate a single high confidence interval set. While simple, it is thought that this method it both useful and widely applicable. Genome-wide binding studies are becoming extremely common and using methods to increase the amount of reliable biological signal extracted from them is considered useful due to the noise levels inherent in techniques such as ChIP-chip. In terms of contribution to Dichaete biology - there is now a highly reliable binding dataset for Stages 1-15 of embryogenesis, which should be helpful for future studies.

The biological conclusions about Dichaete largely match those from Chapter 3: Dichaete binds mainly within either intergenic or intronic regions and tends to be found next to or within 1 kb upstream or downstream of transcription start sites. It is likely that it acts by binding to enhancer regions, although there is also a suggestion that it may facilitate aspects of transcriptional pausing: the binding data is consistent with both hypotheses and it is possible that Dichaete is doing both in different locations or contexts.

One of the ultimate aims of my PhD is to identify functional Dichaete sites in the genome and determine how they facilitate regulation of gene expression. One of the puzzles presented by binding data is that a lot of the binding sites present in the genome may not necessarily be functional. Therefore identifying filters for finding likely functional binding sites is extremely useful.

Using the overrepresentation of the Dichaete binding motif as a filter for functionality showed some very promising results. It is known that binding motifs do not perfectly correlate to functionality, indeed it is possible to have conservation of functionality without conservation of sequence (Hare et al., 2008). However, this seems like a promising method for isolating at least a subset of functional binding peaks, and may be a useful filter for selecting particular regions for individual validation.

Distinguishing between single and multiple binding peaks does not appear to clearly identify functional binding regions, however, it does identify interesting target genes. It appears likely that the genes associated with multiple binding peaks have complex regulation, in part facilitated by Dichaete, and are likely to be both direct and interesting targets.

The preliminary analysis of data deconvolved by time via DNAse accessibility data allows the identification of binding sites that are likely to be dynamically bound by Dichaete, and suggests that this particular class of binding sites might be functionally important. The other use of these data is in cases where other functional data, such as a gene expression profiling, is available for a particular stage of development. Here the binding data filtered by accessibility can be used as a substitute for a time course binding experiment, in order to more closely match the binding profile with the other available data. While of course time course experimental data would be preferable, this is a cheap and easy alternative that might be practical in cases when obtaining a time course is not a possibility. These filtering methods will be used alongside functional data in Chapter 6.

Chapter 5

Dichaete gene expression studies

5.1 Introduction

In Chapter 4, I presented a high confidence Dichaete binding dataset and the insights drawn from it about Dichaete biology. However, looking at binding data alone does not answer the question of which binding sites have a functional role in regulating gene expression, exactly what this role is, and what genes are being regulated. A computational analysis can go some way towards identifying potential targets, but to verify the generated hypotheses, further evidence is required.

5.1.1 Microarrays and gene expression

Gene expression studies are a commonly used tool in functional genomics. In a genome-wide gene expression study, two samples are compared; for example, cancerous and non-cancerous cells, or cells carrying a mutation vs those with a wild type gene. In the context of Drosophila developmental biology, genome wide expression profiling has been successfully employed to identify genes involved in a variety of processes, including mesoderm development (Furlong et al., 2001), early embryonic patterning (Stathopoulos et al., 2002) or aspects of CNS development (Altenhein et al., 2006) as well as studying the activity of particular transcription factors (Montalta-He et al., 2002; Hueber et al., 2007). While the main direct purpose of this type of experiment is to identify transcription factor target genes and the genetic networks they are involved in (Altenhein et al., 2006; Stathopoulos et al., 2002), further analysis can also use these data as a starting point for dissecting transcriptional regulation underlying biological processes, as was the case with uncovering the specificity and hierarchy of Hox gene action (Hueber et al., 2007). Gene expression experiments can also be designed to test evolutionary conservation between either specific genes or broad patterns of gene

expression by, for example, comparing results from the same tissues from different species (Chan et al., 2009) or seeing the effects of a transcription factor from another species on the model system of interest (Montalta-He et al., 2002). Temporal expression profiles related to the action of different transcription factors can also be obtained and can lead to new insights into biology (Furlong et al., 2001).

There are a number of different types of microarrays employed for different purposes: for gene expression, probes are commonly designed from genome annotations or sequenced cDNA libraries and only cover genic regions, whereas tiling arrays are available for use in experiments that require whole genome coverage. In general 2 types of microarray platform are in common use, the Affymetrix GeneChip, a single channel technology (Lockhart et al., 1996), and the spotted or synthesised platforms based on the cDNA array concept developed in the Brown lab, which are generally used in 2 channel studies (Schena et al., 1995). The principal difference is that for single channel microarrays, the sample and control are hybridised to different arrays, and the differences in intensities are only compared during post-processing, whereas for dual channel microarrays, both the sample and the control are labelled with different fluorophores, combined and hybridised to the same array. However, while the platforms are different, a systematic analysis indicates that the choice of microarray platform does not have a significant impact on experiment outcomes (Johnson et al., 2008). While next-generation sequencing is becoming more popular, microarray technology still offers a cost effective and convenient platform for rapidly comparing transcriptomes, particularly in species with compact genomes, where it is possible to provide high density genome coverage on a single tiling array (Aleksic and Russell, 2009). In addition, for many species gene expression arrays are readily available, comparatively inexpensive and reliable.

Here I take advantage of the long oligonucleotide platform developed by the International Drosophila Array Consortium (INDAC) (Sykacek et al., 2011), which has been shown to perform well in a number of previous studies (Choksi et al., 2006; Zeitouni et al., 2007; Meadows et al., 2010). In brief, RNA from each of the samples being compared is extracted, reverse-transcribed, sometimes amplified and then labelled with fluorescent dyes before hybridising to a microarray. After washing and scanning, the ratio of intensities for each spot on the array is quantified, providing a relative measure of the abundance for each mRNA sequence assayed by each specific array probe. Relative differences between two samples, i.e. sample and control, are calculated as the intensity ratio of each spot. In the case of dual channel arrays, which employ different fluorescent dyes for each channel, dye incorporation and detection efficiencies are known to be different and a standard experiment normally includes dye swaps to control for these effects (Russell et al., 2009).

In order to account for natural occurring gene expression variability between samples, as well as to control for technical variability, a number of different biological replicates are included in the same experiment. While the number of replicates employed will depend upon the degree of statistical power one wishes to achieve, FlyChip commonly use four biological replicates (with 2 dye swaps providing a balanced design), however, on occasion where large differences in the magnitude of expression changes are observed, three replicates may yield sufficient data and statistical power (Russell et al., 2009). Once the raw images are recorded and the intensities quantified, microarray data must be normalized within and between arrays to eliminate some of the technical variability associated with the method. There are a variety of techniques employed, depending upon the experimental design and type of data collected, ranging from relatively simple techniques, normalising the intensity distribution across arrays, to more sophisticated approaches that take into account spot intensities and utilise linear regression or probabilistic methods to better deal with the peculiarities of microarray data. After normalization, the reliability of the gene expression ratios across the replicates is assessed statistically allowing thresholding and selection of gene lists for further analysis (Russell et al., 2009).

5.1.2 Mutants used for gene expression

In biomedical research it is common to compare cells from diseased tissues with those from control normal individuals: alternatively, if the genes implicated in a particular disease are already known, they can be individually mutated or knocked-down to study the specific effects on gene expression. In more basic biology research that focuses on understanding, for example, the developmental processes shaping particular organs or tissues, a common study design involves disrupting a specific gene and comparing the mutant state to the wild type tissue or cell type. While whole organisms are frequently used in such studies, more specific experiments can also be performed by isolating particular tissues or single cell types. In the case of Drosophila, while some larval or adult tissues are relatively easy to isolate for expression studies (Pavlopoulos and Akam, 2011), the embryo is more difficult to deal with and it is only very recently that methods for isolating specific cell populations from the embryo have been developed (Salmand et al., 2011). In my studies I have used whole embryos for expression profiling. While it is clear that deconvolving data from across the entire embryo is difficult, as previous studies have shown, it can still be useful to catalogue gene expression changes across the entire animal (Hooper et al., 2007; Papatsenko et al., 2010).

There is a wealth of genetic tools available for the fly, including null alleles being available for many genes, which facilitates the identification of expression changes in the loss of function condition. If possible, it is desirable to combine two different null alleles or

to examine a null allele in combination with a deletion since chromosomes can contain other mutations that may confound the microarray analysis. Alternatively, approaches such as RNAi can be used to disrupt the expression of a specific gene in particular tissues. The Gal4 driver system available in flies makes tissue-specific expression of constructs straightforward and the availability of a variety of tissue or cell-specific drivers provide a very powerful research tool (Brand and Perrimon, 1993). Another tissue-specific approach that utilises the Gal4 system involves creating dominant negative alleles that in some way compete with and/or disrupt the action of a transcription factor of interest, for example by attaching a repressor domain to an activator or by mutating a transactivation or interaction domain of the protein (Margolin et al., 1994; John et al., 1995). In my work, I utilise both the loss of function and the dominant negative approach.

5.1.3 Chapter overview

This chapter presents the results of a set of gene expression studies I performed for a number of *Dichaete* mutant conditions. The gene expression results are both a stand-alone study in their own right as well as a useful functional validation for the Dichaete embryonic DNA binding data presented in Chapters 3 and 4. For this reason, the gene expression data will first be presented on its own, followed by a brief comparison with binding data. More in-depth integration with the binding data and the functional conclusions drawn from that will be explored in Chapter 6. During the course of the gene expression experiments, questions arose about the exact mechanisms of function of some of the dominant negative alleles used and this was investigated further. A pilot study exploring this issue is presented in this chapter, along with suggestions for future work.

5.2 Method

5.2.1 Dichaete null mutants

For expression profiling Dichaete loss of function embryos I took advantage of two characterised null alleles: D^{r72} and D^{r513}. *Dichaete* null mutants display severe segmentation and nervous system phenotypes, and do not hatch from the egg. To obtain *Dichaete* null embryos, each allele was balanced over TM3, *twi-Gal4 UAS-eGFP* and the homozygous mutants identified by the lack of eGFP fluorescence. However, it should be noted that this was not an ideal solution: while it was desirable to perform the gene expression study during the early stages of CNS development, the eGFP

construct takes some time to be expressed and is not visible until 4.5-5 hours of development (late stage 9). The earliest time when fluorescence could be reliably detected was found to be towards the end of stage 10, so a stage 10-11 embryo collection (5 - 7.5 hours) was performed and individual eGFP$^-$ embryos hand picked from timed collections. As a control population, heterozygous embryos with a single functional *Dichaete* allele were selected since it was thought that comparing homozygous null allele embryos to the wild type would poorly control for genetic background and environmental conditions. Heterozygous control embryos were collected by crossing D^{r513} / TM3, *twi-Gal4 UAS-eGFP* to wild type flies and heterozygotes were identified by the lack of eGFP. The samples were stored in Trizol and frozen, with the RNA extraction performed once a sufficient number of embryos was collected, followed by amplification, labelling and hybridisation to INDAC FL003 microarrays. 4 replicates were performed, and a dye swap was incorporated into the experimental design.

5.2.2 Dominant negative mutants

For this study, the dominant negative Dichaete constructs created by Shih-Pei Shen as part of her PhD work were used. The constructs are inserted into the pUAST vector and thus under the control of the Gal4 UAS promoter. Two different types of dominant negative Dichaete proteins were created: one (EnRep) has the *engrailed* repressor domain fused to the C-terminal of the wild type Dichaete protein, in theory turning it into a repressor wherever it binds. The second construct (HMG$^-$) has a deleted HMG DNA-binding domain, meaning that it should no longer bind to DNA but might still be able to compete with endogenous Dichaete for co-factors. A third construct, DWT, is the wild type Dichaete protein in the same vector. While this construct does not result in appreciable embryonic phenotypes compared to embryos expressing the dominant negatives, it was still thought that it might be interesting to examine gene expression when Dichaete is overexpressed. All of the constructs were tagged in frame with GFP, making visualisation, and potentially chromatin immunoprecipitation, straightforward.

5.3 Construct validation

Before embarking upon the microarray analysis, it was considered prudent to validate the dominant negative stocks by checking that they produce the expected phenotypes. For the Dichaete dominant negative constructs, the validation was performed by Aleksandra Mandic, a summer student in the lab working under joint supervision of Bettina Fischer and myself. There were two components to the validation. The first was to

verify the fly stocks, whether the Gal4 drivers were working as expected and to check that each of the constructs was still present in the stock. The second involved verifying that previously reported nervous system phenotypes were present, i.e. that the dominant negative constructs were having the expected developmental effect.

5.3.1 Gal4 driver validation

Since all the constructs were eGFP-tagged, their expression could be easily validated by fluorescence microscopy without the need for immunostaining. The expression patterns were found to be as expected, with the *sim*Gal4 crosses clearly showing fluorescence in cells of the ventral midline and *pros*Gal4 crosses displaying fluorescence in the neuroectoderm and developing neuroblasts (Figure 5.1). An interesting observation emerged from these experiments: while fluorescence was strong and correctly located with the *pros*Gal4, UAS-DWT cross, it was much weaker in the *pros*Gal4, UAS-HMG$^-$ and EnRep crosses. There are two possible explanations for these observations: either the reduction results from the general nervous system phenotype induced by the dominant negatives or, since we have evidence that Dichaete regulates *prospero*, the constructs might in fact be switching off the *pros*Gal4 driver by interfering with normal Dichaete function. Since the nervous system phenotypes do not result in a severe loss of CNS, the latter is considered more likely. A UAS-GFP strain included as a control demonstrates that GAL4-driven expression was in the expected tissues, indeed the GFP signal in this case was very strong compared to the Dichaete constructs. Some of the difference may be due to the fact that the constructs are randomly integrated into the genome via a P-element vector (a more modern approach would insert each construct at the same location using an integrase or recombinase). However, this is considered less likely since the SimGal4 crosses show much less variability.

5.3.2 Phenotype validation

Having verified by fluorescence that the constructs were present and expressed in the expected tissues, phenotypes were checked to ensure the dominant negatives disrupted Dichaete activity as expected by examining major axon tracts of the embryonic CNS with the BP102 antibody (Figure 5.2. The phenotypes observed in the dominant negatives were concordant with the *Dichaete* loss of function (Soriano and Russell, 1998). Expression of the EnRep construct resulted in a strong phenotype, with thinning and breaks of the longitudinal tracts as well as commissure fusion. The HMG$^-$ construct was also found to have a visible CNS phenotype, although not as severe as those observed with EnRep. As expected, expression of DWT and the GFP control did not result in appreciable CNS phenotypes, though occasional commissure breaks were

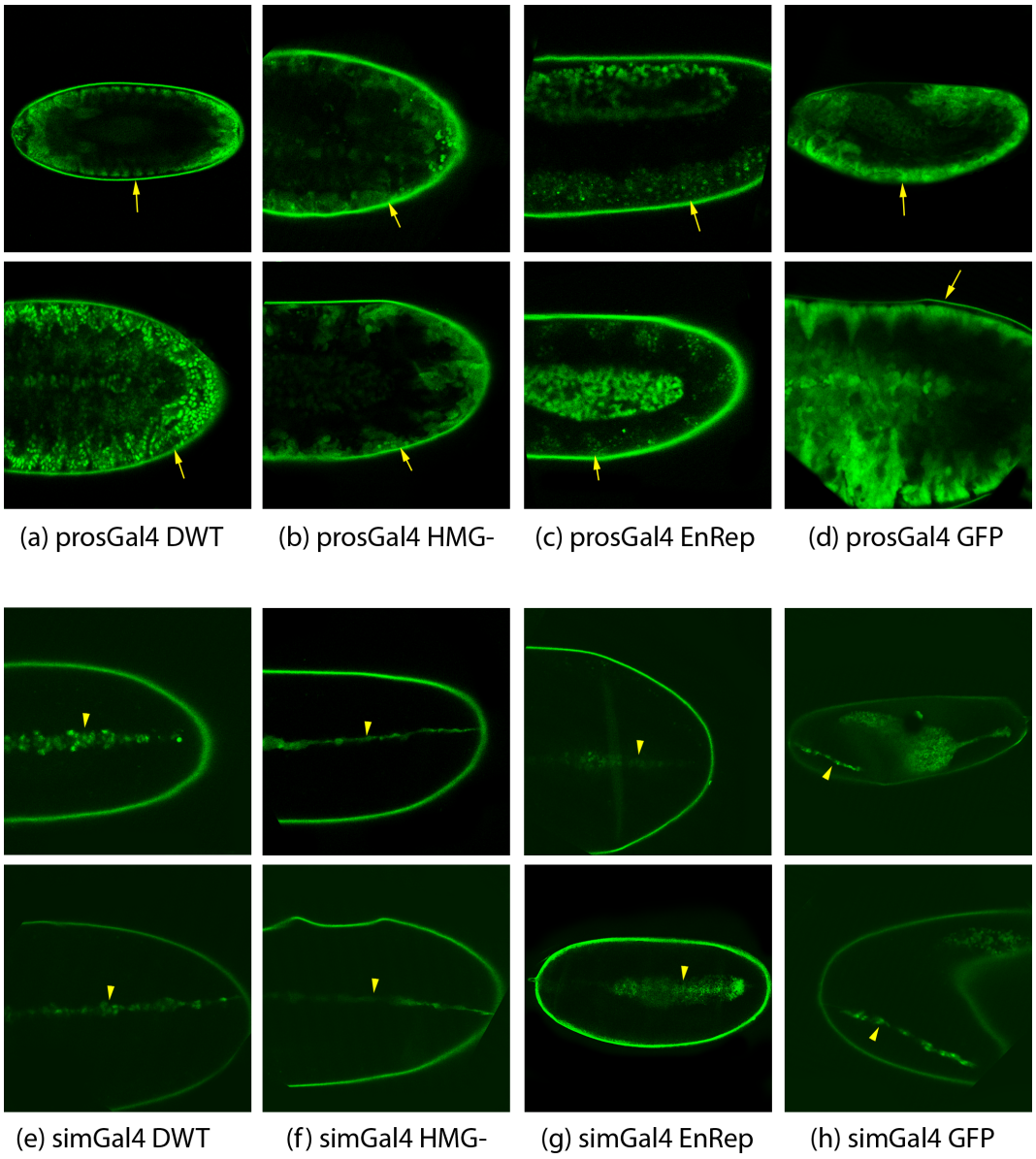

Figure 5.1: Fluorescence observed in the dominant negative crosses. Fluorescence is observed in the neuroectoderm (yellow arrows) in the *pros*Gal4 crosses (a-d) and in the midline (yellow arrowheads)in *sim*Gal4 crosses (e-h), as expected. Interestingly, the *pros*Gal4 dominant negatives (b, c) display less clear fluorescence than the wild type controls (a, d). This is also, to a lesser extent, observed for the *sim*Gal4 dominant negatives (f, g).

found in embryos expressing wild type Dichaete in the midline (arrowhead in Figure 2c). Taken together, the validation experiments reveal the expected phenotypes and confirm that the genotypes are suitable for further gene expression analysis.

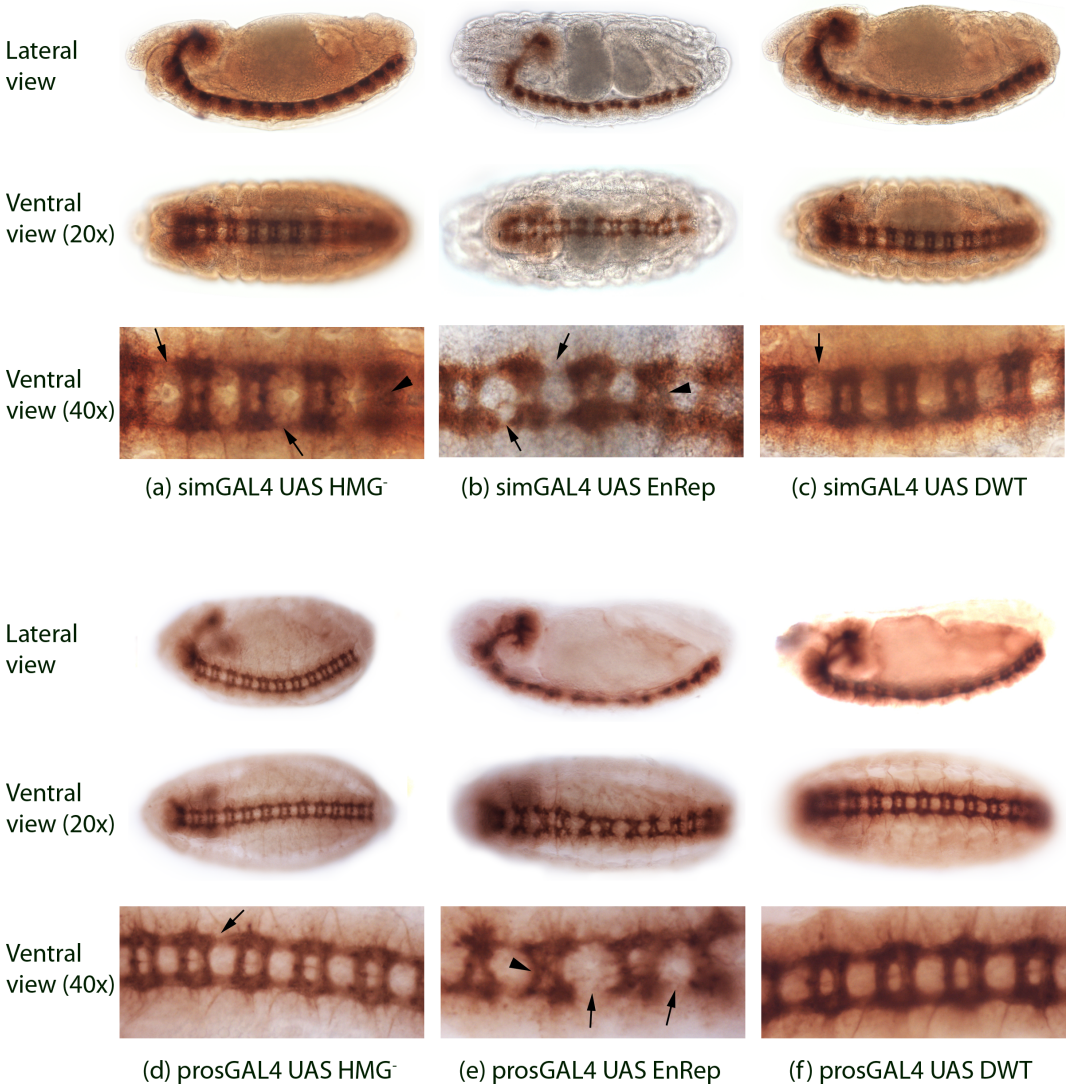

Figure 5.2: Embryonic CNS phenotypes of the dominant negative crosses, visualised with BP102 axon staining. The EnRep dominant negative (b, e) displays the most severe phenotypes, with the collapsed commisures (arrowheads) and breaks in longitudinals (arrows) clearly visible. HMG^{-} (a, d) displays a milder, but still clearly visible phenotype. *sim*Gal4 DWT (c) displays slight thinning of the longitudinals, whereas *pros*Gal4 DWT (f) appears wild type.

Microarray analysis

Two sets of experiments were performed using the dominant negative constructs: expressing each in the midline using a *sim*Gal4 driver and in neuroblasts using a *pros*Gal4 driver. In each case, HMG^{-}, EnRep and DWT were crossed with the relevant driver: sufficient embryos were collected to analyse four biological replicates per cross. In order

to keep the genetic background of the embryos as similar as possible, virgin females were always collected from the Gal4 driver stock and males taken from the dominant negative strain. A UAS GFP construct crossed in parallel with the Gal4 drivers was used to collect control embryos. Roughly 150 embryos per sample were collected between 3.5 and 4.5 hours after egg laying and stored frozen in Trizol. Following RNA extraction, reverse transcription, Klenow amplification and labelling (Chapter 2), samples were hybridised to INDAC FL002 or FL003 *Drosophila* gene expression arrays. Four biological replicates were performed for each experiment, with 2 dye swaps incorporated into the experimental design to control for bias due to different dye incorporation efficiencies.

5.4 Data processing

For all the gene expression datasets, Dapple was used for spotfinding and the arrays were quality checked manually, removing any spots that appeared to be affected by high levels of background or artefacts. The signal intensities were quantified with Dapple prior to normalization and subsequent processing. Differentially expressed genes were selected on the basis of p-values and average M values. An M-value is the $\log_2$ ratio of sample vs control for the probe in question on the array (with each probe being mapped to one particular gene). In cases where multiple probes hit the same gene, the average M value for that gene was calculated as an average across the different probes that were flagged as significant. In these cases the values from the different probes tended to be extremely similar, supporting the reliability of the INDAC microarray.

In the case of the dominant negative constructs, the standard FlyChip data analysis pipeline was employed, using the Variance Stabilisation and Normalization (vsn) package in R followed by CyberT to assess statistical significance. The thresholds used to identify differentially expressed genes were average M-value of < -0.5 or > 0.5 with a p-value < 0.05. When the limma package was used for assessing statistical significance, it filtered the majority of differentially expressed genes that CyberT finds to be significant. While limma uses a more stringent statistical test, it was thought that in this case it was too severe for an exploratory analysis. The rationale here is that the gene expression studies are not only intended as stand-alone experiments, but also to be examined together with other data and act as a screen for potential interesting targets. Therefore, while there are certainly false positives within the sets of genes identified with CyberT, these data were used with this knowledge and awareness that the reduced statistical threshold weakens evidence from each individual study. However, this approach makes it easier to compare results across studies, and it is likely that a gene flagged in multiple experiments is a strong candidate target. In addition, it

is known that *Dichaete* phenotypes are variable (Russell et al., 1996) and it is possible that this will also be reflected in variable effects on gene expression between biological replicates. If this is indeed the case, the direct Dichaete targets may end up with less significant p-values because of the variability between biological replicates despite being real targets. Taken together these arguments support the view that a less stringent statistical test for exploratory analysis of the Dichaete data is more appropriate for looking at Dichaete biology.

In the case of the *Dichaete* null mutant experiment, the data processing proved to be more complex. Most commonly used microarray normalization methods rely on the majority (50%-75%) of genes not showing significant expression changes in any given experiment (Russell et al., 2009). However, such a large number of genes changed expression in the *Dichaete* homozygous null embryos that standard normalization methods failed. To deal with this a previously published invariants normalisation method (Pradervand et al., 2009) was used (suggested by Stewart MacArthur and implemented by Bettina Fischer). In this method, the subset of genes whose expression is constant across arrays are identified and used to normalize the rest of the set. The invariant genes are identified by the invariant normalization method in question, by using mixture models of the mean and variance distributions to identify probes that have medium-high mean intensity and low variance across arrays. The later stages of the analysis were performed using limma. This normalization method seemed to work better than the standard methods (Figure 5.3), identifying 9120 differentially expressed genes using the thresholds employed for the dominant negative constructs. Since such a large number of genes were found to be differentially expressed, and because it was uncertain how accurate the normalization method was, given the small number of invariant genes, more stringent thresholds were applied to obtain a smaller number of more reliable hits. The new thresholds used to find differentially expressed genes in the *Dichaete* mutant were average M-value < -1 or > 1 and a p-value < 0.01.

5.5 Dichaete null mutant gene expression

As described above, the normalisation of the *Dichaete* mutant data was made difficult by the large number of genes that changed expression in this experiment. Indeed, if the statistical thresholds used for the dominant negatives data are applied a total of 9120 genes are found to be differentially expressed (65% of the total genes in the *Drosophila melanogaster* genome!). When more stringent cutoffs are applied a total of 4518 genes were found to be significantly differentially expressed (the data for the 4518 genes is shown in Table 8, Appendix 3). The vast majority of these, 92.6% (4182), are downregulated, with only 7.4% (336) found to be upregulated. While in a null

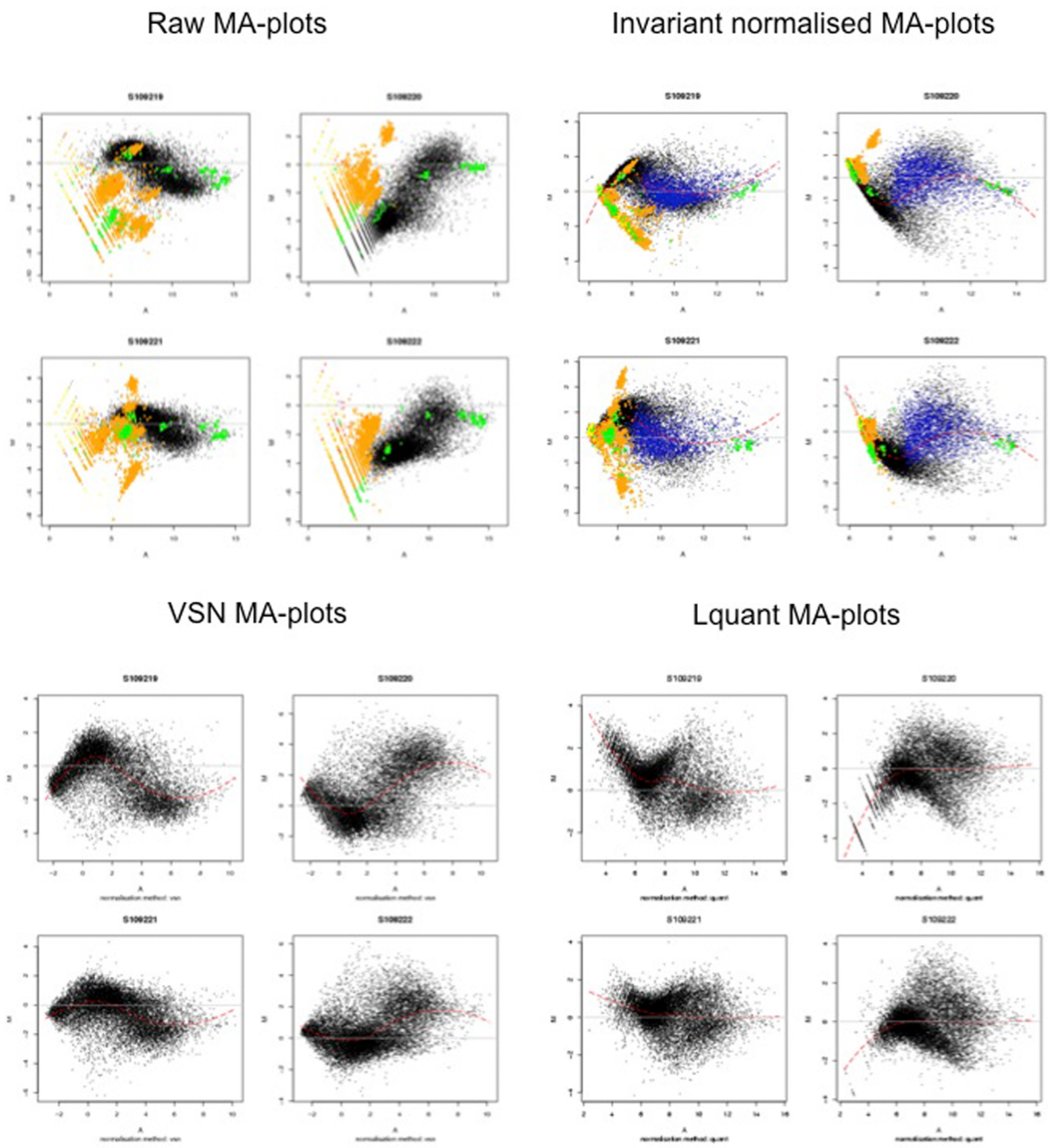

Figure 5.3: Comparison of different normalization methods applied to the Dichaete null mutant expression data. Each of the 4 panels corresponds to 1 biological replicate performed on an individual array, and dye swaps are present in 2 out of 4 replicates (and visible in the overall trends of the graphs). The gene spots are shown in black. For the raw and invariant normalized data, the controls are shown in green, and the spikes are shown in orange (hidden for the other two). For the invariant normalization, invariant genes are shown in blue. A successfully performed normalization is expected to center the data around M-value 0, which VSN and L-quant fail to do.

mutant experiment it is impossible to disentangle immediate gene regulation targets from downstream effects, these results nonetheless highlight the central role *Dichaete* plays in genetic regulatory networks during development. Encouragingly, one of the genes flagged as differentially regulated in this screen is the known Dichaete target *slit*.

The differentially expressed genes were analysed using the Genomewide Expression Maps (GEMs) tool at Fly Express (Kumar et al., 2011; Konikoff et al., 2011) and a heatmap was generated for the genes for which in situ data are available from the BDGP. Encouragingly, the majority of the genes identified have an expression pattern very similar to that of Dichaete (Figure 5.4). This indicates that, despite the large numbers of genes flagged in the gene expression analysis, many of them appear to be specific and constitute plausible Dichaete targets.

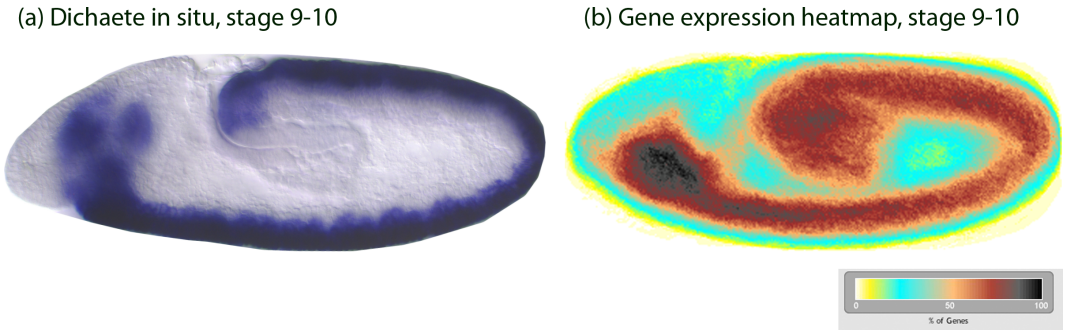

Figure 5.4: Heatmap of genes differentially expressed in the Dichaete null mutant experiment. (a) Dichaete in situ, stage 9-10, taken from BDGP, (b) Heatmap of expression of genes from the Dichaete mutant study for which BDGP in situ data was available. Generated using the GEMs tool at Fly Express

5.5.1 GO enrichment

Despite the large number of genes flagged in the *Dichaete* mutant gene expression experiment, the GO enrichments indicate that the genes are to a large extent very much specific to the reported roles of Dichaete during development. The GO enrichments found include developmental processes (802 genes, p-value $7.00E-06$), nervous system development (329 genes, p-value $1.00E-04$), cell fate commitment (129 genes, p-value$2.00E-06$), embryo development (226 genes, p-value $6.00E-04$) and digestive track development (49 genes, p-value $8.00E-04$): all of these are biological processes which Dichaete is expected to be involved in (Russell et al., 1996; Nambu and Nambu, 1996; Soriano and Russell, 1998; Sanchez-Soriano and Russell, 2000). Another process that was found to be enrichment is open tracheal system development (96 genes, p-value $1.60E-03$), possibly because a number of genes are involved in both tracheal development and neurogenesis. It was also found that 58 of the differentially expressed

genes contain a Homeobox domain (p-value $4.00E-05$). Out of the 296 genes present in the FlyTF 'trusted' list of transcription factors, 152 were found to be differentially expressed in the *Dichaete* null mutant, even with the more stringent cutoffs applied. On the whole, the GO enrichment analysis looks very encouraging and suggests that despite the very broad effects on gene expression observed, the experiment is nonetheless successfully picking out genuine components of the Dichaete genetic network.

5.5.2 Upregulation and downregulation

One of the immediate striking things about the results of this experiment is that the great majority of differentially expressed genes were found to be downregulated. The set of downregulated genes still has a very similar GO enrichment profile to the one described for the complete set of differentially expressed genes, as expected since it constitutes around 93% of the complete set. The upregulated gene set was only found to have a few biological processes significantly enriched, connected with sexual reproduction (40 genes, p-value $3.00E-03$) and macromolecule catabolic process (20 genes, p-value $7.00E-03$). However, this set does still contain a number of interesting developmental genes such as *bicoid* and *Su(H)*, so the absence of GO enrichments may be due to a smaller number of genes present resulting in a lack of statistical power, rather than necessarily because the upregulated genes are fundamentally different to the downregulated genes. From the overall profile, it nonetheless appears likely that Dichaete is mainly functioning as an activator, and possibly driving a number of other activators during development, whereas the upregulated genes are likely to be affected by downstream effects. However, it is thought that Dichaete can affect gene expression context-dependently (Phochanukul and Russell, 2009), so while this experiment gives a rough general picture of what is happening, the specifics are likely to be significantly more complex.

5.5.3 Potential direct targets

As described above, the known Dichaete target gene in the midline, *slit*, is found to be downregulated in the *Dichaete* null mutant. There are a number of other interesting targets that Dichaete may be regulating or interacting with at different stages during development: some during segmentation and axis establishment, such as *engrailed*, *bicoid* and *knirps*, and some during nervous system development, such as *huckebein*, *Kruppel*, *achaete*, *grh* and *sim*, as well as a number of Hox genes, including *Abd-A*, *Abd-b*, *Ubx*, along with the Hox gene cofactor *hth*.

On one hand, the list of genes obtained from this experiment is intriguing: it includes a large number of interesting developmental genes, including a considerable number of

transcription factors which have a range of roles at various points during development. On the other hand, this makes the analysis difficult since many interesting potential targets are highlighted in this screen and many of the affected transcription factors are known to be involved in development from early embryonic stages. This means that there are likely to be considerable downstream effects from disrupting these regulators. Thus, while the experiment is likely to identify the general genetic network that Dichaete acts within, it is impossible to determine those effects that are direct and those that are a result of the entire network being disrupted. In addition, for the transcription factors involved in a range of processes, it is not immediately clear from the data at which stage they are likely to interact with Dichaete.

In spite of these concerns, these data do provide an interesting starting point for studying the genetic network that Dichaete is involved in, particularly when analysed in combination with the high confidence Dichaete binding data, providing preliminary evidence for functional interactions. A brief comparison between binding and expression data is described below, which will be further explored in Chapter 6. A major conclusion from the whole embryo experiment is that to gain a more detailed understanding of specific Dichaete functional interactions, more focused, tissue-specific experiments are required. For these reasons I chose to focus on the nervous system, and performed a number of tissue-specific gene expression experiments using Gal4 drivers and dominant negative constructs in the hope that these studies would point to direct Dichaete targets and provide insights into the specific action of Dichaete in different tissues rather than giving a global overview.

5.5.4 Overlap with binding data

Out of the 4518 genes found to be differentially expressed in a *Dichaete* null mutant at stage 10-11, 1519 (33.6%) were also found to be associated with high confidence Dichaete binding intervals. While the majority of the genes are not associated with Dichaete binding, this is expected since much of the differential expression is likely to be due to indirect downstream effects rather than direct Dichaete action. The binding data helps identify a subset of genes that may be direct Dichaete targets and it is likely that the subset of bound and regulated genes are more focused than the entire set of differentially expressed genes. The subset includes, for example, 262 genes involved in nervous system development (p-value $1.1E-49$), which constitutes slightly over a 1/6th of the entire gene set. It also contains most of the transcription factors found in the original *Dichaete* null mutant dataset, with 131 of the FlyTF 'trusted' set of transcription factors bound and affected by Dichaete (compared with 152 TFs identified in the entire gene expression list). The overlap list also includes the known Dichaete target *slit*.

Interestingly, a reasonable proportion (144 out of 336, 43%) of the upregulated genes were found to be associated with Dichaete binding. In comparison, only 1375 out of 4182 (33%) downregulated genes are found in the high confidence Dichaete binding set. This may suggest that Dichaete action is less straightforward than simply acting as a transcriptional activator and that it is possible that Dichaete acts as a repressor in certain contexts. Alternatively, it is possible that some genes are regulated by both Dichaete and a number of other transcription factors within the Dichaete network, in these cases it is unclear which part of the network perturbation is responsible for the observed direction of expression change. Of course this is also the case for all the downregulated genes. The GO enrichments found for the upregulated genes were again mostly connected to sexual reproduction and oogenesis. The GO enrichments for downregulated genes are as expected from *Dichaete* phenotypes and contain a number of developmental processes including nervous system development.

5.6 Dichaete in the ventral midline

When the dominant negative mutants were expressed in different parts of the CNS, the ventral midline and all neuroblasts, a number of differentially expressed genes were detected (Table 5.1). Expression of the wild type Dichate construct also induced some changes in gene expression, presumably because the action of Dichaete is dosage-dependent. Focusing first on the ventral midline: 214 genes in total changed expression in at least one of the 3 screens (the data is shown in Table 6, Appendix 3). The total of differentially expressed genes was 169 for the HMG^{-} mutant (50 upregulated, 119 downregulated), 48 for the EnRep mutants (13 upregulated, 35 downregulated) and 45 for wild type (17 upregulated and 28 downregulated). Clearly this is a relatively small number compared to the total number of Dichaete binding peaks in the genome or to the results obtained in the *Dichaete* mutant study. However, it is worth bearing in mind that the ventral midline is a very small tissue (around 500 cells per embryo) and that the changes in gene expression were assessed using mRNA extracted from whole embryos. Thus only very pronounced changes in gene expression are expected to be detected in this particular study.

5.6.1 Dataset overlap

The overlap between the datasets (Figure 5.5) is reasonable: in the case of wild type and EnRep, 64% and 58% of genes flagged, respectively, are also found in at least one of the other two datasets, In the case of HMG^{-}, the proportion of shared genes is smaller, since expression of this dominant negative identifies a large number of genes

Construct used	Number of upregulated genes	Number of downregulated genes
*sim*Gal4 UAS HMG$^-$	50	119
*sim*Gal4 UAS EnRep	13	35
*sim*Gal4 UAS DWT	17	28
*pros*Gal4 UAS HMG$^-$	66	533
*pros*Gal4 UAS EnRep	119	154
*pros*Gal4 UAS DWT	18	66
D null mutant	336	4182

Table 5.1: Overview of numbers of genes differentially expressed in different screens

not flagged in the other two screens. The reason for this is not immediately obvious: while it is possible that the HMG$^-$ protein disrupts gene expression more than the other two constructs, this seems curious since the strongest CNS phenotypes are observed in EnRep expressing embryos. However, focusing only on genes that overlap in more than one dataset did not seem to meaningfully filter results to more likely targets, therefore all genes were considered further as possible targets with the obvious caveat that the genes featuring in more than one dataset are more likely to be actual targets.

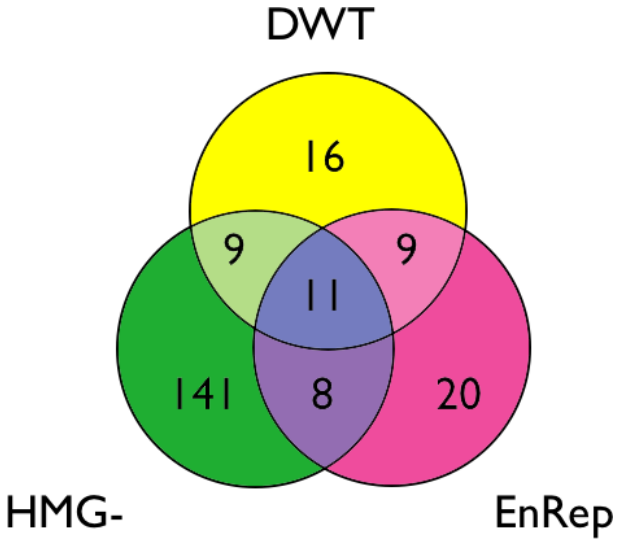

Figure 5.5: Overlaps between genes flagged in different midline gene expression datasets

One potential explanation for the extra genes disrupted by HMG$^-$ is that the Dichaete HMG$^-$ construct competes for cofactors, but since the Dichaete DNA binding domain is missing, the cofactors end up binding to DNA in different locations and disrupting a set of genes that the cofactors can affect but that Dichaete does not normally regulate in the midline. In addition, HMG$^-$ could also be sequestering the cofactors required by other transcription factors as well as Dichaete, thus disrupting several genetic networks at once. If the genes disrupted are not required for the development of the midline, this may not show up as a severe phenotype, despite considerable differences in gene

expression.

5.6.2 Functional analysis

When looking at the set of all genes changing expression in at least one of the *sim*Gal4 experiments, no GO enrichments are found. This is possibly because the number of genes flagged in the study is quite small. The known Dichaete target in the midline, *slit*, was not flagged in any of the studies, possibly because of sensitivity issues or because the dominant negatives used failed to disrupt its expression. In all cases where there was an overlap in differentially expressed genes between experiments, the change in gene expression was always in the same direction, for example, upregulated in all three. This is not an obvious finding since we would expect the consequences of Dichaete overexpression to be different, potentially opposed to the effects of attaching a repressor domain. This highlights the fact that the exact action of the dominant negatives used is not well understood and consequently a precise interpretation of the data gained from these constructs is difficult. On the other hand, these studies may provide a degree of validation for functional relationships inferred from the loss of function studies. Out of the 214 genes found to be differentially expressed, 65 were found to be upregulated and 149 downregulated. However, the two categories did not show clear functional enrichment, and it is not clear that there is a distinct difference between them.

5.6.3 Potentially interesting direct targets

Only three transcription factor genes, 'trusted' as defined by FlyTF, were found to be differentially expressed: *nejire*, *mTTF* and *slbo*. Of these, *nej* is particularly interesting as it is involved in both segmentation and neuron differentiation, which it facilitates through histone modification and transcription cofactor activity. This gene is also bound by Dichaete in the core binding interval dataset, with several upstream Dichaete binding peaks, and it is therefore an interesting potential direct target, perhaps a potential cofactor. Its expression is only disrupted using the HMG^{-} construct, where it is slightly upregulated, suggesting that perhaps Dichaete normally acts as a repressor here, and that the HMG^{-} construct disrupts its expression by competing for cofactors.

The gene *bancal*, which is reported to have transcription factor binding activity and to be involved in mRNA splicing, was found to be downregulated in all three experiments. There are four Dichaete binding intervals associated with *bancal*, which suggests it may be a direct target. However, it is unclear exactly how this gene is regulated; the

downregulation upon increased expression of wild type Dichaete suggests that it is being repressed, but the dominant negative alleles should disrupt rather than increase the usual action of Dichaete. In the case of the EnRep allele, the expectation is that targets normally activated by Dichaete should be downregulated and thus we would conclude that *bancal* should be an indirect target. However, experiments using similar Sox2 constructs in zebrafish suggest that what might be happening is that the En repressor domain is not affecting the Sox2 activation domain, but instead further repressing already repressed targets (Paul Scotting, personal communication): this may also be the case with our Dichaete-EnRep construct. As described above, the effect of the HMG$^-$ allele is the least straightforward of the three to interpret but as before we assume that it acts through sequestration of co-factors.

5.6.4 Overlap with Dichaete binding data

Out of the 214 genes identified in the midline gene expression screens, 61 are also associated with core Dichaete binding intervals. Of these 61, 24 are upregulated and 37 are downregulated, which is not significantly different from the proportions in the total set of genes ($\chi^2 = 1.37$, p-value = 0.24). The set of upregulated genes shows weak GO enrichment for the cell fate commitment of R3/R4 photoreceptors (p-value = 0.027), but this is only based on the presence of 3 genes (*nej*, *fz* and *Rala*). However, given the extreme eye phenotype associated with Sox2 in mammals, these may constitute interesting direct targets. The downregulated set contained several genes of interest such as *bl* and *Kr-h1*, but no clear enrichment patterns were found. The list of potential targets identified in this screen is varied and not all of the genes have an obvious connection with Dichaete function: in particular, a number of genes connected with chitin/cuticle development are present in the list. Some of the more interesting genes include those described above (*bl* and *nej*) as well as the Wnt receptor *frizzled*, the transcription factor *Kr-h1*, the Notch signalling pathway genes *BobA* and *Rala*, *baf*, which encodes a DNA-binding protein with an embryonic nervous system phenotype, and *pip*, which is involved in DV axis specification.

5.7 Dichaete in neuroblasts

837 genes in total changed expression in at least one of the three screens using the *pros*Gal4 driver (the data is shown in Table 7, Appendix 3). The total of differentially expressed genes was 599 for the HMG$^-$ mutant (66 upregulated, 533 downregulated), 273 for the EnRep mutants (119 upregulated, 154 downregulated) and 84 for UAS Dichaete wild type (18 upregulated and 66 downregulated). This experiment affects

about 4 times as many genes as the midline experiment does, but this is to be expected since the tissue targeted is larger and it is easier to pick up more subtle changes in gene expression. The HMG$^-$ strain again causes the largest number of gene expression changes, the reasons for which, as described above, are unclear. For the most part, the effects on gene expression were the same for all three constructs. However, there is a small subset (22 genes) which were downregulated in HMG$^-$, but upregulated in EnRep. The reason for this is unclear. It is possible that the repressor domain does not work on activated genes and that the EnRep construct is mimicking the action of Dichaete overexpression, whereas HMG$^-$ is disrupting Dichaete action. Alternatively, it is possible that these are indirect targets - only 3 out of the 22 genes are associated with Dichaete binding peaks, and none of them are clearly connected to Dichaete developmental functions.

5.7.1 Dataset overlap

The overlap between the three *pros*Gal4 datasets is reasonable but not particularly high. 39 out of 84 genes from DWT (46%) overlap with other datasets, 88 out of 273 for EnRep (32%) and 86 out of 599 for HMG$^-$ (14%). As with the *sim*Gal4 data, there is a large subset of genes from HMG$^-$ that is not identified in any other dataset. There is a large degree of overlap between the effects of the two dominant negatives, which is greater than the overlap of either of them with the Dichaete overexpression results. This could be because the effects of the dominant negatives are stronger, and it is therefore possible to identify more targets. The differences in the genes targeted could be accounted for by the dominant negative alleles having very different modes of action. The Dichaete binding data indicates that there are potentially thousands of direct targets in the genome and it is possible that the different constructs identify different targets. However, as before, potential targets are considered to be more reliable if they are identified in a minimum of two different gene expression screens.

5.7.2 GO enrichment

The neuroblast experiment provided a much larger list of differentially expressed genes than the midline experiment. However, despite this, and although a number of individual interesting targets were found, very few significant GO enrichments were identified in the total set of 837 genes. There were no GO enrichments with a p-value lower than 0.01: the enrichments found included female gamete generation (55 genes, p-value 0.017), embryonic pattern specification 932 genes, p-value 0.017), cellular process involved in reproduction (58 genes, p-value 0.019) and oogenesis (54 genes, p-value 0.022).

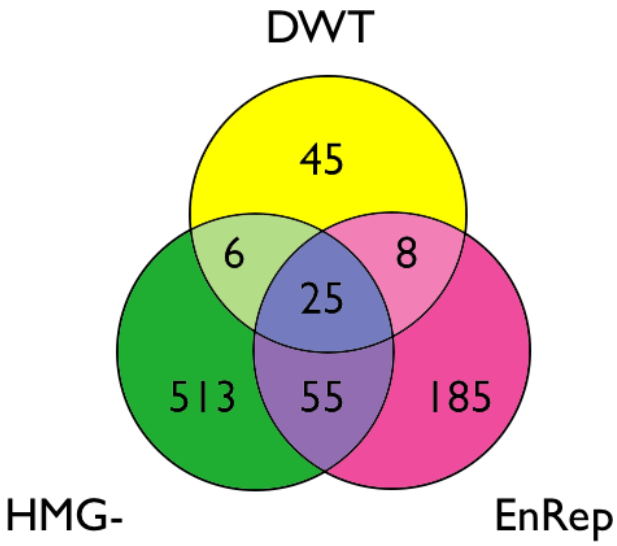

Figure 5.6: Overlaps between genes flagged in different neuroblast gene expression datasets

Embryonic pattern specification is very relevant to Dichaete function in early development, and this subset includes the genes *hunchback* and *Delta*, which are involved in nervous system development, a process associated with Dichaete function. The other results appear similar to the enrichments for the genes upregulated in a *Dichaete* mutant, and may be connected to the reported function of Dichaete in developing oocytes (Mukherjee et al., 2006). Interestingly, Dichaete does not localise in the nucleus during oogenesis, and its role appears to be performed entirely through RNA binding, so if there is indeed a connection and similar effects are being seen in *pros*Gal4 dominant negative experiments, this raises the question of where the dominant negative proteins localise and how such effects are being facilitated. The imaging data presented earlier suggests that wild type and EnRep constructs are predominantly nuclear, however the HMG- protein may also be present in the cytoplasm (Figure 5.1).

Taking a look at the HMG^{-} study individually also yielded some interesting results: with *pros*Gal4 a number of GO enrichments, mostly connected to appendage morphogenesis (32 genes, p-value 0.002) and wing disc morphogenesis (28 genes, p-value 0.01), were found. Cell cycle regulation was also highlighted (29 genes, p-value 0.047). While Dichaete does not appear to play a direct role in wing development, the original *Dichaete* phenotype is a change in wing structure caused by the ectopic expression of Dichaete (Russell, 2000). It is possible that the HMG$-$ dominant negative is pairing up with Dichaete cofactors, causing expression changes to genes that the cofactors regulate independently of Dichaete. It is therefore possible that Dichaete interacts with proteins expressed in neuroblasts that are also involved in wing morphogenesis, possibly one or more Hox genes. The individual results from expressing the wild type and EnRep constructs yielded no enrichment.

5.7.3 Upregulation and downregulation

In total, 194 genes were upregulated, and 687 genes downregulated. The upregulated genes displayed specific GO enrichments, including cellular process involved in reproduction (24 genes, p-value 0.001), positive regulation of cellular metabolic processes (13 genes, p-value 0.017), cell differentiation (36 genes, p-value 0.012), gamete generation (26 genes, p-value 0.008) and anatomical structure development (50 genes, p-value 0.013) among others. Interestingly, some of these are similar to the enrichments found for genes upregulated in the *Dichaete* null mutant, so while they could be indirect targets, it is also possible that Dichaete acts as a repressor on the subset of genes involved in reproduction when they are not required, but potentially acts as an activator in other tissues where they do need to be expressed. As described above, Dichaete has a reported role in oogenesis (Mukherjee et al., 2006), and it is unclear whether the expression experiments may be picking up on some of the same functions. While the downregulated gene list does feature a number of factors involved in neurogenesis and development, there were found to be no statistically overrepresented GO terms for this set of genes.

5.7.4 Potential direct targets

The gene *bancal* was identified in all three midline experiments and in all three neuroblast experiments, which combined with the binding data (four Dichaete core binding intervals upstream and within the gene) suggests that *bancal* is a high confidence direct target. The transcription factor *hunchback* was upregulated by wild type Dichaete in the neuroblast study, which is interesting given its role in the temporal sequence of neuroblast division (Maurange et al., 2008). This suggests that, as well as being expressed together in neuroblasts at the same time, there may be a regulatory loop between them. Since *hunchback* was not identified as differentially expressed in the midline experiments, this regulatory relationship may be specific to neuroblasts, although there is a caveat regarding the sensitivity of the midline screen. A number of other developmentally important genes were identified in the screens, including the transcription factors *bicoid*, *charlatan*, *Mad*, the insulator *mod(Mdg4)*, and a number of genes involved in neurogenesis, including *lola* and *midline*, the neuropeptide hormone *Nplp4*, and the Notch signalling pathway genes *Delta* and *Notchless*. Dichaete itself was also differentially expressed in both HMG^{-} and EnRep expressing embryos, suggesting potential self-regulation in neuroblasts. However, since it was found to be upregulated in both, it is also possible that this is an indirect effect caused by a feedback loop between Dichaete and some of its direct targets.

5.7.5 Overlap with Dichaete binding data

Out of 837 genes found to be differentially expressed in the *pros*Gal4 study, 352 genes (42%) are also associated with Dichaete binding. The GO enrichments found for these targets are highly specific, including embryo development (44 genes, p-value 0.000024), segmentation (25 genes, p-value 0.0002), cell differentiation (66 genes, p-value 0.0002) and nervous system development (49 genes, p-value 0.005) among others. This is very encouraging and it appears that, while some of the genes identified the *pros*Gal4 study may be downstream effects not directly related to Dichaete function, combining the expression data with binding data appears to give much more focused results, identifying a reasonable set of potential direct targets.

5.8 Comparison of gene expression studies

Looking at the individual analyses, the different expression datasets show some common features, for example the majority of the differentially expressed genes are downregulated in all three studies. There are also some similar GO enrichments found between sets, although judging by these, the dataset that is most reflective of global Dichaete biology is that derived from the *Dichaete* mutant. This is perhaps to be expected, as tissue-specific experiments zoom in on one particular role of a transcription factor, rather than necessarily reflecting a global picture. The purpose of this section is to take a more systematic look at the similarities and differences between the datasets, to suggest the best way of using them together to gain insights into Dichaete biology, or indeed to decide whether combining the datasets is sensible at all.

5.8.1 Overlap between datasets

The overlap found between the different gene expression datasets is reasonable (Figure 5.7): 49% of the genes differentially expressed in a *pros*Gal4 study are also found to be differentially expressed in other studies, and the percentage is even higher (59%) for *sim*Gal4. In the case of the *Dichaete* mutant, 431 out of 4518 genes (9.5%) are also differentially expressed in other studies. Due to the different nature of the gene expression experiments a perfect overlap was not expected and the level of overlap found is encouraging.

The 32 genes that are differentially expressed in all three studies seem most likely to be potential direct targets. However, a more detailed look at the list suggests that this may not be the most informative set since most of the genes have unknown functions or belong to gene families seemingly unrelated to Dichaete phenotypes (with the potential

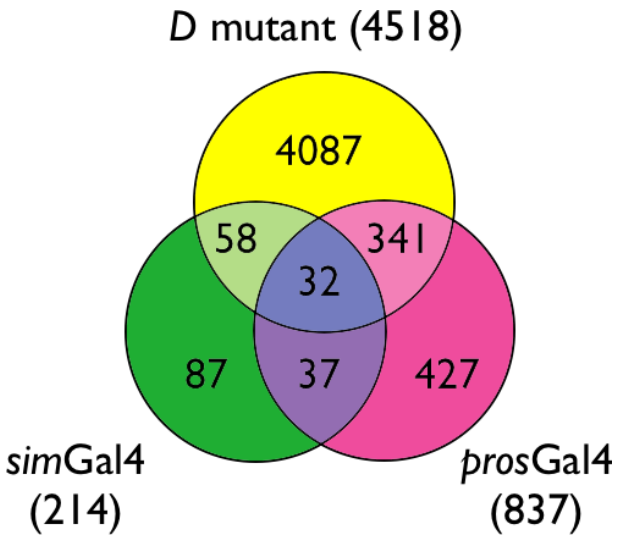

Figure 5.7: Overlaps between genes flagged in the 3 different gene expression studies

exception of the cell cycle gene *wisp*). This could be because with the dominant negative experiments, the set of genes that appear in at least one study is large and may be unspecific (as discussed before, the HMG^{-} dominant negative in particular may be causing a range of effects not directly related to Dichaete function).

Looking at these data as seven individual experiments rather than three studies was explored. However, taking all the genes that are differentially expressed in at least one of the seven studies gives a large set of 5069 genes, which is likely to include a high number of false positives. Filtering the dataset by looking at genes flagged in at least two out of seven studies narrows down the set to a more manageable 520 genes and means that each potential target is supported by at least two gene expression studies. However, the list of genes obtained in this way appears somewhat haphazard and as a whole not particularly specific to known Dichaete functions. It is therefore questionable whether the intersects of these datasets are particularly meaningful. While the fact that there is some overlap is encouraging, I concluded that taking the overlap as a guide to identifying the most reliable targets is not the optimal strategy. Looking at the individual studies suggests that the most meaningful filter for genes related to Dichaete function is actually to identify the overlap between gene expression and Dichaete binding: this strategy will be further explored as a way of finding direct Dichaete targets in Chapter 6. Prior to this however, there are a few other potential filtering methods that highlight some interesting features in the expression data.

5.8.2 Clustering

Looking at the subset of genes flagged as differentially expressed in at least four out of seven studies (31 genes; Figure 5.8) shows that, with a few exceptions, the majority of

genes change expression in the same direction. At a superficial level, it appears that interesting targets such as *bl* and *phr* are mainly found in the downregulated subset. The majority of the genes flagged in the expression experiments are downregulated, possibly because Dichaete mainly acts as an activator and thus the upregulated genes are not direct targets. However, it is also possible that Dichaete is sometimes acting as an activator and sometimes as a repressor. If this is indeed the case, some of the upregulated genes may also be direct targets. If Dichaete were acting differently dependent on the spatial context, it would be expected that certain genes would for example be consistently upregulated in the midline and downregulated in neuroblasts, but looking at the data, this does not appear to be the case. It is possible that this is because we are looking at two tissues where Dichaete acts similarly, or perhaps because the differences are more precisely tuned and arise, for example, in a specific subset of neuroblasts and are thus averaged out and not identified by this particular experimental design. Most of the differences shown here where a particular gene is sometimes upregulated and sometimes downregulated raise questions about the mode of action of particular dominant negatives, rather than providing support for a tissue-specific model of Dichaete action. While the identification of direct targets from this data alone is not possible, it is nonetheless a useful screen for targets that are connected to the Dichaete genetic network in the developing nervous system, directly regulated or otherwise.

5.8.3 Differences in expression fold-change

To explore the possibility that some of the constructs cause greater fold-changes in gene expression and hence identify larger gene lists, the average M values for each set were compared (Table 5.2). For the most part, the fold-change of downregulated genes was higher than that of the upregulated genes. In the case of the *pros*Gal4 data, this is particularly striking with the DWT construct where very few genes are upregulated and when they are the fold change is low. A much larger number of genes are downregulated with much larger fold changes. At first glance, this seems to indicate that, in the context of neuroblasts, Dichaete mostly acts as a repressor and that overexpression of Dichaete results in further repression, whereas upregulated genes are barely affected and may be indirect targets. This pattern is less clear in the context of the midline, possibly either because of the lower sensitivity of the screen or because the action of Dichaete in the midline is different from that in neuroblasts. For the *Dichaete* mutant, the thresholds used were more stringent in the first place, which is why the average M values are much higher. It is interesting to note that the averages are almost identical for upregulated and downregulated genes.

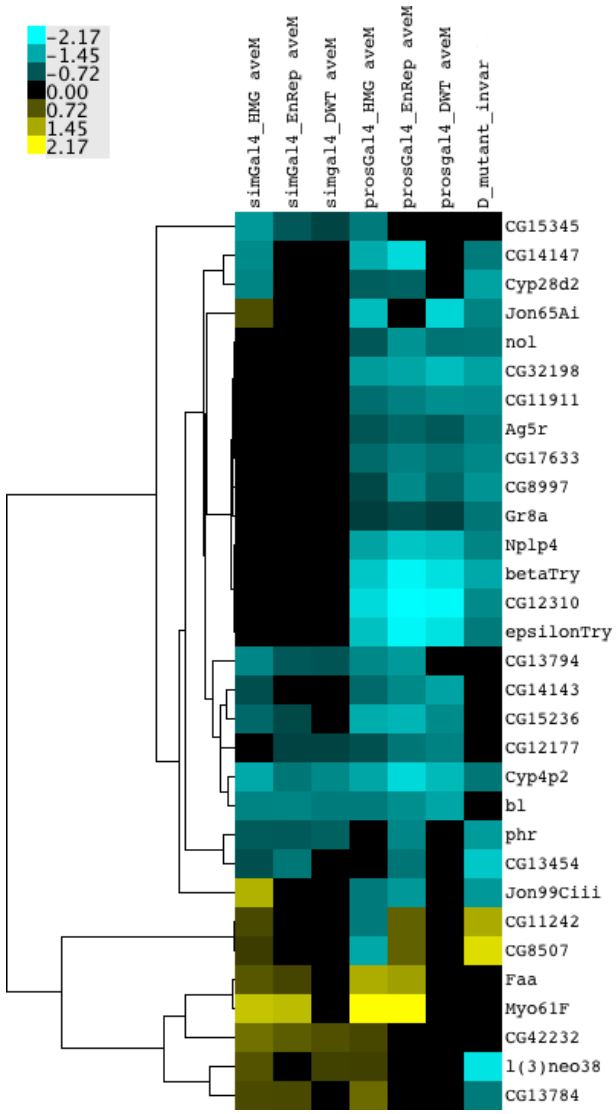

Figure 5.8: Hierarchical clustering of genes differentially expressed in at least 4 studies. Uncentered correlation, centroid linkage.

Construct used	Number of upregulated genes	Average M value	Number of downregulated genes	Average M value
*sim*Gal4 UAS HMG$^-$	50	0.74	119	-0.95
*sim*Gal4 UAS EnRep	13	0.90	35	-0.67
*sim*Gal4 UAS DWT	17	0.63	28	-0.69
*pros*Gal4 UAS HMG$^-$	66	0.72	533	-0.78
*pros*Gal4 UAS EnRep	119	0.71	154	-0.91
*pros*Gal4 UAS DWT	18	0.56	66	-1.07
D null mutant	336	1.28	4182	-1.27

Table 5.2: Average M value of genes differentially expressed in different screens

Using the average M value for finding direct targets

Following from this I investigated whether larger changes in expression are indicative of genuine Dichaete targets. The Dichaete null mutant data was ranked by average M value and split into 2 sets of 2259 genes each: the 'top' set (higher absolute average M values) and the 'bottom' set, corresponding to absolute values above and below 1.2, respectively. A higher average M value does indeed appear to be indicative of a greater potential to be a direct target. The top set features a large number of genes involved in developmental processes (510 genes, p-value 1.2E−15) and nervous system development in particular (231 genes, p-value 1E−13), alongside a host of enrichments for other specific developmental processes. In contrast, the bottom set only features two GO enrichments with p-values less than 0.01: proteolysis (157 genes, p-value 0.0009) and sensory perception (72 genes, p-value 0.005). Thus, a higher average M value does indeed appear to be indicative of a greater potential to be a direct target. Overlap with Dichaete binding data indicates a similar result: while 945 genes from the top set (42%) overlap with Dichaete binding, only 575 (25%) from the bottom set do so. These results indicate that, while all differentially expressed genes should be considered potential targets and it is unwise to filter datasets too soon in the analysis, the absolute average M value can, if required, be used as a filter to determine which genes are more likely to be direct targets.

5.8.4 Comparison to binding data

Taking all the gene expression experiments together, I conclude that while gene expression studies are a useful functional assay, on their own they are of limited utility. For example, it is difficult to determine whether overexpression of Dichaete changes the expression of genes that are already Dichaete-controlled in the tissue of interest or whether an increased concentration of Dichaete affects new targets not normally regulated. The other obvious issue that arises is whether differentially expressed genes are direct targets or whether some expression changes are due to downstream effects. This is of particular concern for the *Dichaete* mutant study since Dichaete is involved in many processes early in development and disentangling direct from downstream effects in an experiment performed as late as stage 10-11 is difficult. On the face of it, this may be less of an issue with the dominant negative experiments, since they are localized to particular tissues. In these cases it is more likely that at least some of the gene expression changes reflect direct targets, while the others may be indirect but in the same genetic network as Dichaete.

Because of the difficulty in interpreting the gene expression results it is clear that the high confidence binding data presented in Chapter 4 can provide a crucial extra piece of

the puzzle for unravelling the genome-wide action of Dichaete. Since one of the aims of my project was to uncover new direct Dichaete targets in the genome this approach will be further explored in Chapter 6. As a prelude to this, and to act as a preliminary filter for a further exploration of the expression data, the overlap between the expression and binding data was identified (Figure 5.9). This analysis identifies a reasonable number of potential direct targets (1519 genes overall during embryogenesis, and more specifically, 61 in the midline and 352 in neuroblasts) and there are a number of further areas that can be investigated from here. The binding patterns in the vicinity of genes with affected expression may be different from genes with seemingly unaffected expression. The underlying sequence and presence or absence of particular binding motifs may also play a role. The presence of binding by other transcription factors that are thought to be cofactors would also be interesting to investigate.

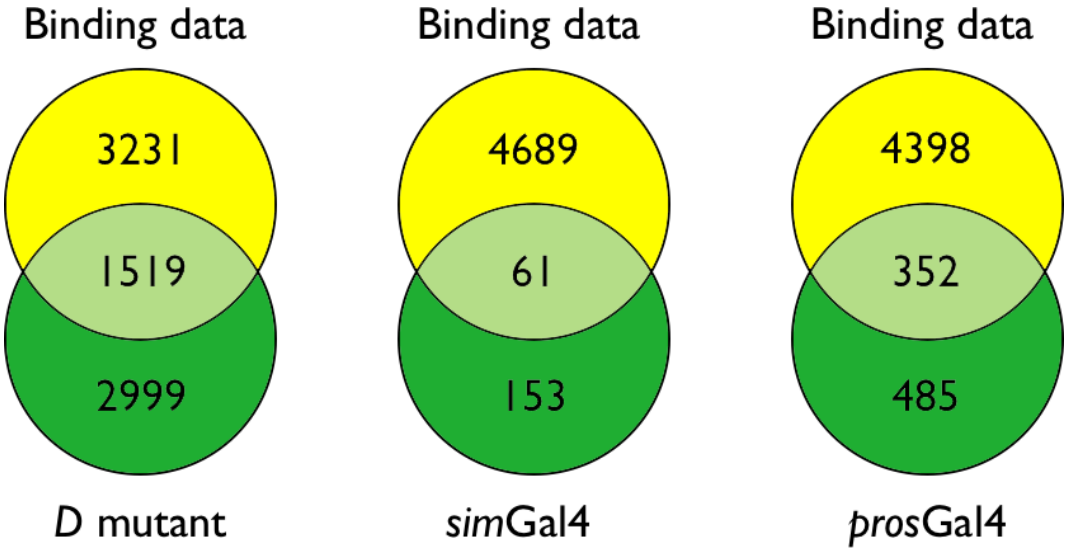

Figure 5.9: Overlap between different gene expression studies and Dichaete binding data

5.9 Variability

The phenotypic variability observed in *Dichaete* mutants (Russell et al., 1996) is unusual. One possible interpretation is that Dichaete acts as a stabilizing factor for transcriptional regulation and that when it is mutated or disrupted it leaves the whole system more open to chance and fluctuations. It was thought that one way of verifying this would be to look at whether genes that are likely to be direct Dichaete targets (i.e. those associated with Dichaete binding) display more variance in expression in a *Dichaete* null mutant than genes not associated with Dichaete binding. To obtain an estimate of expression variability the standard deviation of M values from each of the 4 replicate microarrays was calculated for each gene. If multiple probes target a

single gene the variance for each probe was used individually, rather than calculating an average for each gene. There was found to be a significant difference between the two sets (Wilcoxon rank sum test, p-value $< 2.2E-16$), with the Dichaete bound set on average showing a higher standard deviation (mean 0.32, median 0.27) than the set of genes not bound by Dichaete (mean 0.29, median 0.23). The distribution of values is shown in Figure 5.10.

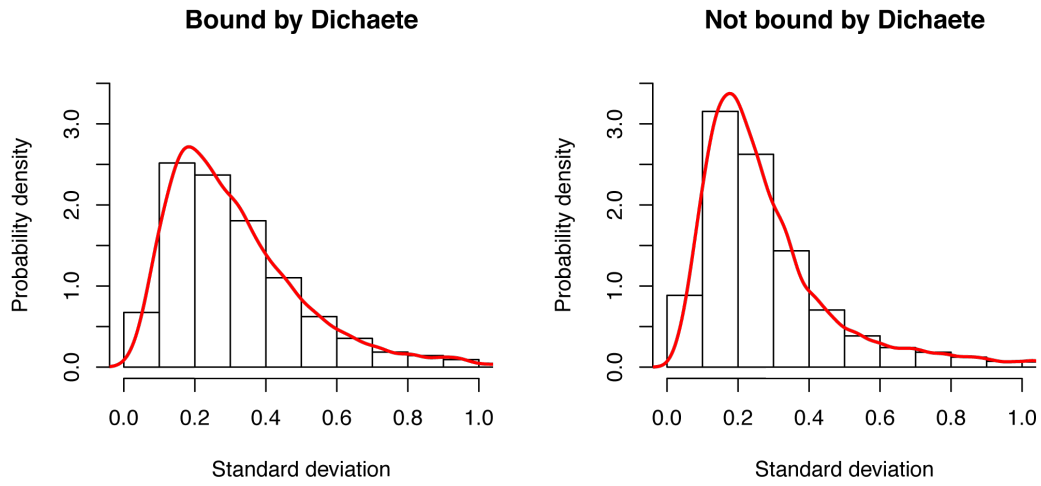

Figure 5.10: Distribution of standard deviations of differential gene expression in a Dichaete null mutant

This is an interesting result: if increased variability is an effect of disrupting *Dichaete*, it makes it more difficult to reliably identify direct targets from gene expression experiments, since higher variability between replicates is normally interpreted as an indication that data are less reliable. However, it is possible that this variability is in fact a reflection of Dichaete biology. These data are consistent with a model where Dichaete acts as a stabilising factor for gene expression, possibly required for gene expression maintenance (Soriano and Russell, 1998). Unfortunately, high variance can be caused by a number of factors, some of which are experimental noise, so on its own cannot be used as a reliable predictor of which genes are direct targets. Nonetheless, this was considered an interesting observation.

5.10 Dominant negatives as a tool for functional genomics

Dominant negative mutations such as the ones used in this study are a common tool for inferring functional relationships between transcription factors and their target genes (John et al., 1995; Margolin et al., 1994). However, it is currently not entirely clear what their precise mechanism of action is, making data analysis challenging. A pilot

study was therefore performed by Aleksandra Mandic, under the supervision of myself and Bettina Fischer, in order to shed light on the exact mechanism of action of the constructs used in this study. The HMG$^-$, EnRep and Dichaete wild type constructs were expressed in developing neuroblasts with the *pros*Gal4 driver and ChIP-chip experiments performed to determine where they bind in the genome. Since all the constructs were GFP-tagged, an anti-GFP antibody was used for the immunopurification and an anti-$\beta-$galactosidase antibody used on the same chromatin as a mock-IP control.

For the wild type and EnRep experiments technical difficulties meant that it was only possible to obtain 1 replicate each in the available time. The EnRep array identified a small number of bound regions (14 at FDR 5% and 57 at FDR 25%). With the wild type, between 211 (FDR 1%) and 6409 (FDR 25%) bound regions were identified. Obviously, a single replicate does not give a reliable enough dataset to draw any solid conclusions, however, a number of the genes associated with these bound regions were involved in various developmental processes including neural development and one of the genes associated with EnRep binding was *slit*, which is a known bona fide Dichaete target. Thus it was thought that these were encouraging preliminary results.

With the HMG$^-$ construct a more complete dataset was generated, with three biological replicates performed. The binding profiles found in each biological replicate were quite different, but the QC plots show that the microarray data was high quality (Figure 5.11), suggesting the possibility that the differences in binding profiles between the replicates are a genuine reflection of biological variability with this construct. Separating the replicates and analysing each one individually yielded a variable number of peaks; 3, 27 and 887 at FDR 25% for each of the replicates. Again, the genes associated with the binding were involved in various developmental processes, suggesting the data are at least plausible. However, overlap with the gene expression data is very limited. For the replicate that produced 887 binding intervals, these were found to be associated with 856 genes, but only 53 of these genes matched the 599 genes differentially expressed in the *pros*Gal4 HMG$^-$ experiment. Obviously, with only one replicate at FDR 25%, a very large number of noise and false positives would be expected, which might account for the difference in datasets.

The overlap with Dichaete high confidence binding data was in general low, approximately 10% on average for each of the available datasets. However, it was observed in the case of HMG$^-$ that much of the binding was in the vicinity, but not overlapping with, wild type Dichaete binding intervals. This led to the hypothesis that HMG$^-$ affects Dichaete action by competing for cofactors and associating with different DNA sequences, depending on where the cofactor is bound. This hypothesis also accounts for the variability of the data: competing for cofactors with an existing wild type protein is likely to be a stochastic, variable process, especially since the Dichaete protein with

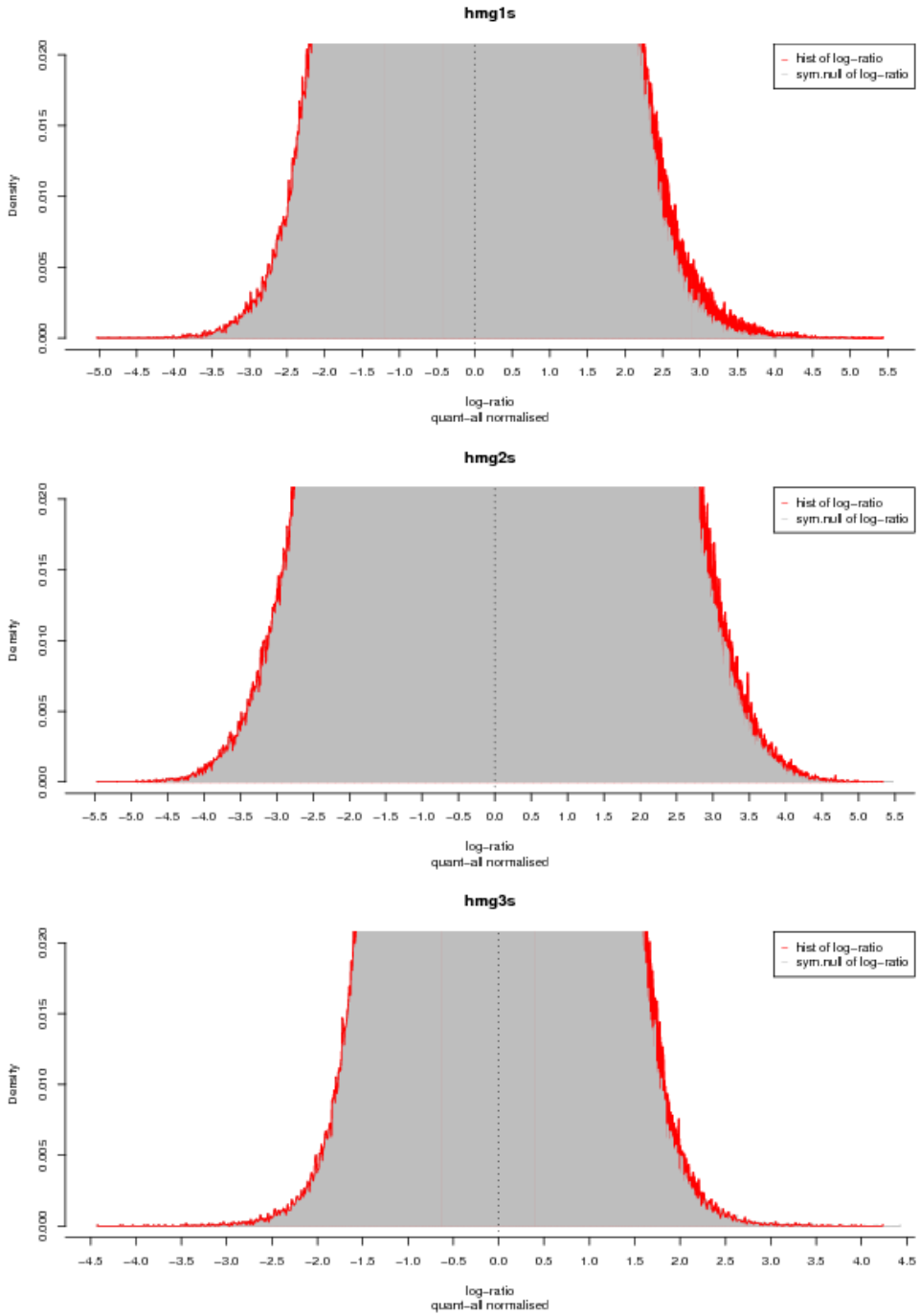

Figure 5.11: Histogram of the HMG$^-$ data from 3 biological replicates. The grey area shows the experimental noise and the red area shows the signal. The data is thought to be high quality because the signal clearly stands out from the noise, rather than overlapping with it.

its HMG domain removed is likely to be a lot less efficient at this than its wild type counterpart. In fact, the likely reason that any phenotype at all is observed with this particular mutant construct is because it is expressed much more strongly than the wild type protein, giving it a competitive advantage.

The same experiment was also performed in the midline using the HMG$^-$ and wild type constructs using the *sim*Gal4 driver, by Louis Chauviere under my supervision. While the HMG$^-$ construct did not yield any binding intervals, the wild type construct yielded much better binding data. Three replicates were performed, identifying between 740 (FDR 1%) and 5751 (FDR 25%) binding intervals. These preliminary observations suggests that even in a restricted tissue Dichaete can bind to a substantial number of genomic locations, albeit under conditions where it is overexpressed. The FDR 1% peaks from this experiment were compared to the available data from the *pros*Gal4 experiment and the overlap was found to be extremely low, possibly because the constructs are expressed in different tissues or because the datasets are unreliable. Binding studies with these dominant negative constructs are ongoing since they are likely to yield some interesting insights into the mechanisms of action of these widely used approaches.

5.11 Conclusions

The binding data and the *Dichaete* null mutant expression study indicate that Dichaete binds thousands of locations in the genome and has a knock-on effect on a similar number of genes by stage 11 of embryonic development. This makes sense, considering that Dichaete is involved in a number of crucial developmental processes (Russell et al., 1996; Soriano and Russell, 1998; Sanchez-Soriano and Russell, 2000; Overton et al., 2002; Zhao and Skeath, 2002; Zhao, Wheeler and Skeath, 2007). While the large number of binding locations and target genes is unusual, it is certainly not unique amongst transcription factors: Kr, Hairy, Daughterless, Dorsal, Mothers-against-dpp and Twist all display similar binding profiles (MacArthur et al., 2009). In addition, the GO analysis suggests that the *Dichaete* null mutant expression data appears to be reliable and quite specific to Dichaete function despite the large number of genes that change expression and the expected abundance of downstream effects. The overlap between the Dichaete binding data and the *Dichaete* null mutant expression data provide an interesting list of possible direct targets genes across the entire genome. The binding patterns associated with these genes and sequence underlying them will be explored in detail in the next chapter.

Looking at the overlaps between the three gene expression studies did not appear to give particularly illuminating results due to the confounding issues described above.

After exploring the experimental data together, I decided that it would be more effective to treat each study separately in order to answer a set of different questions. The *Dichaete* null mutant study provided insights into the broader patterns of Dichaete action, supporting the view from the binding studies that Dichaete regulates many genes in the genome. The tissue-specific experiments performed with the dominant negative constructs allowed a preliminary focus on particular areas of Dichaete influence. However, it became clear that the use of dominant negative constructs raised more questions than were answered, since the direction of gene expression changes observed did not necessarily concur with the activity expected of each dominant negative. This is certainly a useful thing to bear in mind for future studies using such approaches.

One interesting observation to emerge from the gene expression studies was that a high average M value for a gene appeared to be quite a strong indicator of whether a gene is likely to be a direct target or not. This could potentially be used to filter the strongest candidates for specific verification studies. It was intriguing to note that genes associated with Dichaete binding also appeared to display a higher variability in expression measures across replicates, which is consistent with some of the preliminary results from the dominant negative studies and supports a model of Dichaete as a factor that stabilises gene expression at specific points in development.

While Dichaete does appear to bind to a large number of different locations in the genome, the tissue-specific gene expression studies indicate that the action of Dichaete in particular tissues is much more small-scale, with a suggestion that it may switch from a being a master regulator of global phenomena, such as segmentation, to micromanaging the expression of specific neural genes in different parts of the developing CNS. Unfortunately, a precise tissue-specific binding map for Dichaete is currently unavailable, although the experiment expressing wild type protein in the midline with the Gal4 system hints that Dichaete can bind to many genomic locations in a restricted tissue. There are new experimental approaches becoming available to generate tissue-specific binding profiles (for example, ChIP with chromatin from FACS-sorted nuclei or DamID approaches using the GAL4 system to drive specific expression of reduced activity methylase constructs), but these have only recently become available and while they are technically challenging and beyond the scope of this thesis, they provide an exciting future extension to this project. In the meantime, the Dichaete core binding set generated as part of this project can be used as an approximation of where Dichaete is likely to be binding and finding overlaps between these data and tissue-specific gene expression studies appears to provide useful preliminary insights. For a more in-depth analysis, it will be interesting to take temporal chromatin accessibility data (Li et al., 2011) into account in an attempt to make the binding profile more specific to the time windows used for the gene expression analysis

The preliminary data for the dominant negative ChIP was encouraging, and this may be an interesting project to carry on with in the future. While having proper binding data for the mutant proteins would make it easier to interpret the Dichaete expression data presented here, it also has much broader relevance to research, as dominant negative constructs are frequently used in a range of model organisms, but to date their mechanisms of action are poorly understood.

Chapter 6

An integrative analysis of Dichaete genomic data

6.1 Introduction

In the era of rapid second generation sequencing and widely available genome-scale microarrays, the possibility of obtaining a detailed global overview of genome biology is, for the first time, becoming a real possibility. In particular, two large scale international projects have been launched with this aim in mind; the ENCODE project to catalogue functional elements in the human genome (Feingold et al., 2004) and modENCODE to analyse the genomes of the model organisms *C. elegans* and *D. melanogaster* (Celniker et al., 2009).

The modENCODE project is of particular interest to researchers working on the model organisms in question. The advantages of the project are two-fold: first, it broadens our current understanding of a range of genomic processes, from a broad integrative analysis of functional elements for both organisms (Gerstein et al., 2010; Roy et al., 2010; Negre et al., 2011) and full transcriptome sequencing (Graveley et al., 2010; Spencer et al., 2011), to shedding light on more specific aspects of genome regulation such as trans-splicing (Allen et al., 2011), microRNAs and mirtrons (Berezikov et al., 2011; Chung et al., 2011), transcriptional diversity of cell lines (Cherbas et al., 2011), the role of chromatin architecture in a range of processes (Eaton et al., 2011; Ercan et al., 2011; Kharchenko et al., 2010; Liu et al., 2011; Riddle et al., 2011), a genome-wide look at promoter architecture (Hoskins et al., 2011) and transcription factor binding (Niu et al., 2011).

The second advantage is that, because all the data is publicly available, a range of high quality genomics data becomes accessible worldwide, allowing smaller labs and individual researchers to do comparative analysis using datasets that would otherwise

be difficult to obtain for individual labs in terms of the time and cost required to generate them. This means that as well as providing interesting research conclusions in its own right, modENCODE enables genomics research in the wider scientific community. My own PhD project has certainly benefited from this, in particular from the ChIP-chip and ChIP-seq datasets available for a range of transcription factors and histone modifications studied during different time periods of *Drosophila* embryogenesis.

Another note-worthy project with a more specific focus, but similar aims, is the Berkeley Drosophila Transcription Factor Network Project (BDTNP) (Li et al., 2008; MacArthur et al., 2009), which has amongst other things produced very high quality ChIP-chip data for a range of developmentally important transcription factors during early stages of *Drosophila* embryogenesis. There have also been a number of smaller scale efforts by individual labs to generate genomic data for their transcription factors of interest (Choksi et al., 2006; Schwartz et al., 2006; Zeitlinger et al., 2007b; Adryan et al., 2007; Holohan et al., 2007; Kwong et al., 2008; Southall and Brand, 2009). There is considerable potential for using this large quantity of available data for mapping out the key genetic networks and pathways active during different points in development.

When looking at transcription factor binding, while determining their overlap is trivial, getting a measure of statistical significance of overlap is a much more complex problem. The question is a relevant one, as it is extremely useful to be able to look at multiple binding profiles and determine which ones may be overlapping by chance and which ones are likely to be acting in unison or antagonistically. This problem, along with a number of potential methods for solving it, is reviewed in detail by Fu and Adryan (2009). The main question being addressed is whether two transcription factors bind in the same location more commonly than would be expected by chance, and the difficulty in answering it stems from the measure of significance being highly reliant on choosing the correct null distribution model. For example, assuming a null distribution where all transcription factors can bind anywhere in the genome does not reflect existing biological constraints that limit transcription factor binding to a subset of possible genomic locations, thus this approach massively overestimates the significance of any overlap found. Failure to account for the non-random nature of transcription factor binding can result in a large number of false positives (Haiminen et al., 2008; Huen and Russell, 2010).

Resampling tests are one method of dealing with this problem, as they generate a null distribution from known binding sites, thus producing a profile that is more closely matched to specific protein binding preferences. However, Huen and Russell (2010) find that the success of this approach is highly sensitive to the manner in which the resampling is done. The standard recommended resampling method reshuffles the labels

of the binding intervals present in the datasets. This strategy is flawed as it means that the original number of overlaps is the maximum number of overlaps that can possibly be achieved after reshuffling. It therefore massively overestimates the significance of the overlaps compared to simulations. Huen and Russell have developed a hybrid resampling method for pairwise comparisons of binding profiles, which performs well and appears to be applicable to real world data. The hybrid resampling method involves splitting the sites into two groups: one group are the overlapping sites and the other group are the singleton (non-overlapping) sites. With the overlapping sites, each site gets assigned at random to one transcription factor only. With the singleton sites, they can be randomly assigned to one or both transcription factors, resulting in new overlaps. This method has been implemented as an R package (Cooccur) by David Huen, and is used extensively in this chapter.

6.2 The search for direct Dichaete targets

In previous chapters, a range of evidence was presented to address the question of where Dichaete binds in the genome and what effects its absence or disruption has on gene expression. One of the aims of my PhD was to generate a list of candidate direct targets, so the aim of this section is to combine the binding and gene expression data in order to achieve this aim. As discussed previously, the gene expression studies yielded quite different results from one another, so it was thought that they would be best treated separately, looking individually at potential direct Dichaete targets in the midline, in neuroblasts, and more globally throughout embryogenesis.

6.2.1 Dichaete targets in the midline

The transcriptome analysis in Chapter 5 revealed 61 potential Dichaete targets in the midline - genes that change expression in at least one of the *sim*Gal4 expression studies, as well as showing an overlap with Dichaete core binding data (Table 11, Appendix 5). These were already discussed in the previous chapter, but it was thought that a further look at Dichaete binding patterns around them and a more selective choice of direct targets would be interesting. This list is clearly not extensive, as the known Dichaete target in the midline, the gene *slit*, is not captured in this analysis. It was though that the reduced sensitivity was due to the midline contributing only a small fraction of the total embryo, thus all but the strongest changes in gene expression would be averaged out.

First, the gene expression data was compared to the genes associated with Dichaete binding intervals at stage 9 (estimated according to DNaseI accessibility data). Out

of the 4372 genes associated with this Dichaete dataset, 56 overlapped with expression data (Table 12, Appendix 5). The 5 genes lost were all genes of unknown function (except for *Gasp*, which is involved in chitin production), so it does not appear that much, if any, of the signal from the experiment was lost by using this filter. Instead it is likely that this filter narrowed down the list further into more likely functional targets, as these are genes where Dichaete is likely to bind at the appropriate stage of development.

Out of the 56 genes bound at stage 9, 38 are associated with a single Dichaete binding site, 5 are associated with 2 sites, 6 are associated with 3 sites, and 7 are associated with 4 or more (Table 6.1). Potentially, all 18 genes associated with more than one Dichaete binding interval at Stage 9 constitute interesting direct targets.

Number of binding peaks	**Associated genes**
2	*Cpr100A*, *BG4*, *Kr-h1*, *Rala*, *fne*
3	*bl*, *CG14688*, *Hsc70-3*, *l(3)neo38*, *CG8965*, *CR42862*
4 or more	*olf186-F*, *BobA*, *fz*, *Glut4EF*, *CG42574*,*Sema-5c*, *fog*

Table 6.1: Genes that change expression in dominant negative mutants and are associated with multiple Dichaete binding intervals at Stage 9 of embryonic development

Examining the gene expression patterns in detail reveals some interesting observations (Figure 6.1). First of all, the strongest direct target gene associated with nervous system development is *bancal*. This gene is significantly downregulated in all 6 dominant negative experiments. Further to that, while it does not get flagged as differentially expressed in the *Dichaete* mutant when the very stringent filtering is applied, it should be noted that its average M value in this experiment was -0.82, with a p-value of 0.009, so it does appear to be downregulated in all 7 experiments performed. This gene is also associated with three Dichaete binding intervals at Stage 9 (Figure 6.2), all of which are intragenic and predominantly located in intronic regions. *bancal* encodes a functionally conserved homologue of the vertebrate *hnRNP K* gene, and is reported to act in imaginal discs, promoting apoptosis to ensure the correct size of adult appendages (Charroux et al., 1999), so it is possible that it employs a similar mechanism in the nervous system. This would give it an important developmental role, given that temporally regulated apoptosis is essential for correct nervous system development (Wheeler et al., 2011). While it is definitely considered a high confidence direct target, the exact action of Dichaete on *bancal* is unclear. The downregulation in response to Dichaete overexpression in both the midline and neuroblasts suggests that Dichaete is acting as a repressor and the data from HMG^- and EnRep studies are consistent with this hypothesis. The fact that *bancal* is also downregulated in a *Dichaete* mutant seems to run contrary to this hypothesis, but this could be due to

downstream effects from various transcription factors being disrupted, and may not be indicative of direct Dichaete action. The model of Dichaete acting as a repressor of a gene promoting apoptosis is consistent with the idea of Dichaete being a factor that maintains cells in a pluripotent dividing state.

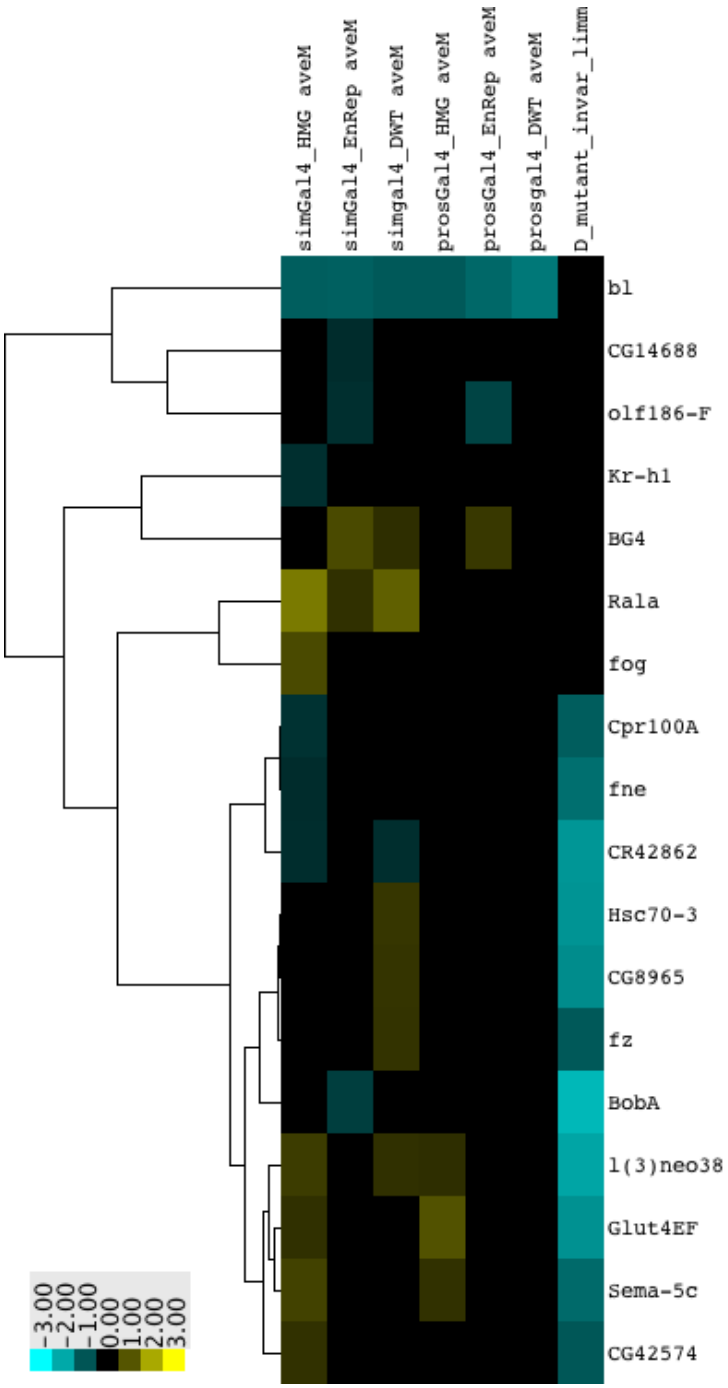

Figure 6.1: Clustering of genes differentially expressed in the midline, which are associated with 2 or more Stage 9 Dichaete binding intervals, as estimated from accessibility data.

Another interesting target from this analysis is the gene *Rala*, which is upregulated in all experiments in the midline, but not affected in any of the neuroblast experiments or

the *Dichaete* mutant. This appears to be an interesting midline-specific direct Dichaete target. It is known to be involved in regulation of the Notch signalling pathway (Eun et al., 2007) and cell fate commitment of R3/R4 photoreceptors. In an early paper, Sawamoto et al. (1999) report that *Rala* is expressed ubiquitously throughout embryogenesis and only develops specific spatial and temporal expression during larval development and later. However, the in situ data in Fly Express suggests more specific, though weak, expression in the embryonic neuroectoderm (Konikoff et al., 2009; Kumar et al., 2011), perhaps as expected, if it is indeed being repressed by Dichaete in the nervous system. It is therefore unclear whether it has any midline specific functions, or whether Dichaete merely has the potential to specifically activate it, but normally does not because it is not present in sufficient quantities. This is nonetheless an interesting general target of Dichaete, particularly in view of the severe eye phenotype observed in vertebrates with *Sox2* loss of function.

It is interesting to note that in the case of *bl*, which appears to be repressed, Dichaete binding is mainly intronic, whereas with *Rala*, which appears to be activated, Dichaete binding is in intergenic regions (Figure 6.2). While the evidence from two genes only is merely anecdotal, it would be interesting to investigate whether this is a genome-wide trend.

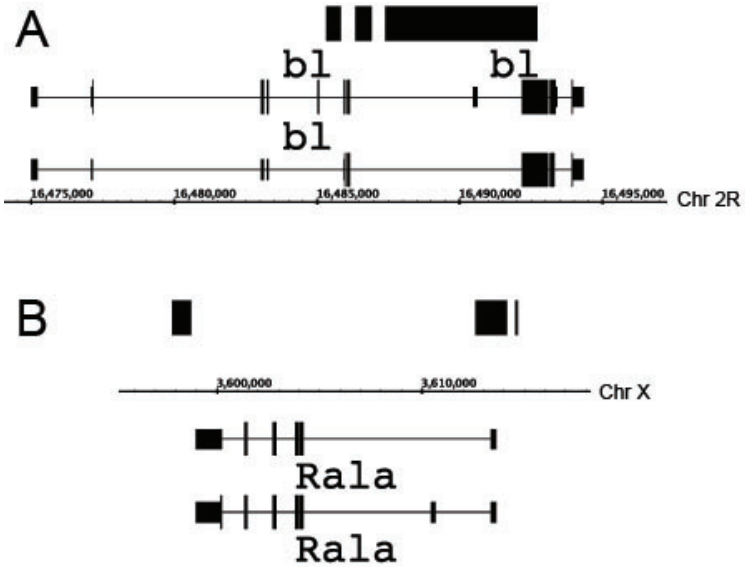

Figure 6.2: Dichaete binding patterns during embryonic Stage 9 around the direct Dichaete targets in the midline. A) *bancal* is thought to be repressed by Dichaete. B) *Rala* is thought to be activated by Dichaete.

6.2.2 Dichaete targets in neuroblasts

Out of 837 genes differentially expressed in the *pros*Gal4 study, 352 were found to overlap with Dichaete high confidence binding intervals. 325 of these genes are associated

with Dichaete Stage 9 binding, all of which can be considered potential direct Dichaete targets in neuroblasts (Table 13, Appendix 5). A large portion of these are involved in developmental processes (111 genes, p-value 0.005), and in particular in embryonic pattern specification (26 genes, p-value 2E-6) and organ development (66 genes, p-value 6E-6). Particular developmental processes that are flagged are generation of neurons (39 genes, p-value 0.004), segmentation (24 genes, p-value 3.3E-4) and female gamete generation (34 genes, p-value 0.003). The list of genes involved in neuron generation includes *SoxN*, *Delta*, *lola*, *mid*, *hb*, *ac* and *hkb* (the complete list is included in Table 14 in Appendix 5).

Out of the 325 genes in question, 209 are associated with single Dichaete binding intervals at Stage 9, 50 genes are associated with 2 Dichaete binding intervals, 23 with 3 intervals, and 43 with multiple (4 or more) intervals (the genes bound by 4 or more intervals are shown in Table 15, Appendix 5). Despite containing the greatest number of genes, the gene set associated with single Dichaete binding intervals displays no significant GO enrichments. The set associated with 2 intervals is weakly enriched for a single term (regulation of cellular component biogenesis: 6 genes, p-value 0.03) and the set associated with 3 genes contains no enrichments (possibly due to the small number of genes included). In contrast, the set of 43 genes associated with multiple Dichaete binding displays a striking level of enrichment for a range of important developmental processes, many of them related to transcription, segmentation and nervous system development (Table 16, Appendix 5). This set of potential direct targets is clearly very stringent and likely misses some high probability targets, such as the gene *bancal*, which has 3 associated Dichaete binding intervals and is flagged in all gene expression studies. However, the stringency and specificity of this gene set is also an advantage for further studies verifying Dichaete action in neuroblasts, providing a high confidence candidate list (the full set of genes is listed in Table 15, Appendix 3). For a slightly less stringent, but nonetheless high confidence set of targets, the set of genes with 2 or more associated Dichaete intervals could be used.

When looking at the gene expression patterns of the multiply bound genes (Figure 6.3, these display some unusual patterns compared to other differentially expressed genes. First of all - while in the whole neuroblast dataset genes are mostly downregulated, with this particular subset the number of upregulated genes is roughly equal to the number of downregulated genes. It is possible that this reflects the fact Dichaete can both upregulate and downregulate its direct targets, whereas its overall effect at a whole genome scale is to activate a large portion of the developmental network, thus resulting in mainly downregulated genes, not all of which are direct targets. Very few of the selected genes change expression when Dichaete is overexpressed, potentially because if the sites next to them have high affinity for Dichaete, they may already

be maximally occupied, and overexpresion of Dichaete does not alter this. Another interesting observation is that most of the dominant negative constructs caused gene expression changes in the opposite direction from those observed in the *Dichaete* null mutant, a trend contrary to the genome-wide observation.

For the few genes for which a Dichaete overexpression effect is available to indicate which way they may be regulated, a similar pattern holds to the one found in the midline: an activated gene, (*hb*), shows mainly intergenic binding, whereas repressed genes, (*mod(mdg4)* and *Jupiter*), show mainly intragenic, specifically intronic, binding (Figure 6.4). A further examination reveals that for repressed genes, particularly in the case of *mod(mdg4)*, Dichaete binding intervals closely correspond to Polymerase II binding peaks from 0-12 h embryo ChIP-chip (Figure 6.5), consistent with the hypothesis that Dichaete is attenuating gene expression by facilitating Polymerase pausing at a set of intragenic sites.

6.2.3 Global Dichaete targets during embryogenesis

Of the 4518 genes found to be differentially expressed at stage 10-11 in the *Dichaete* null mutant, 1519 were found to be associated with high confidence Dichaete binding intervals. This particular dataset is difficult to precisely deconvolve by time, since all 1519 genes in question are potentially direct targets and in principle their change in gene expression could have been caused by downstream effects from events that happened at any point before or during stage 11. Therefore identifying specific target genes in particular tissues is difficult from this dataset. Rather, these data provide more of a global look at the broad transcriptional network Dichaete is involved in during development. The purpose of this section is to identify a more narrow high confidence set of Dichaete targets, based on filters developed in previous chapters. A brief look at the binding patterns around the identified targets is also presented.

High confidence direct targets

As discussed in Chapter 5, it is possible to use the absolute M value for filtering these data meaningfully and the analysis indicated that a higher absolute M value is indicative of genes more likely to be related to Dichaete function. Accordingly, the top half of the differentially expressed genes in terms of absolute M value (2259 genes in total) were used for this part of the analysis. Of these, 176 were upregulated and 2083 were downregulated, and 945 of the genes overlap with Dichaete high confidence binding intervals. 133 of the Dichaete bound genes were associated with multiple (4 or more) Dichaete binding sites and these were considered to be a high confidence set of Dichaete targets (Table 17, Appendix 3). 83 of the genes in question were associated

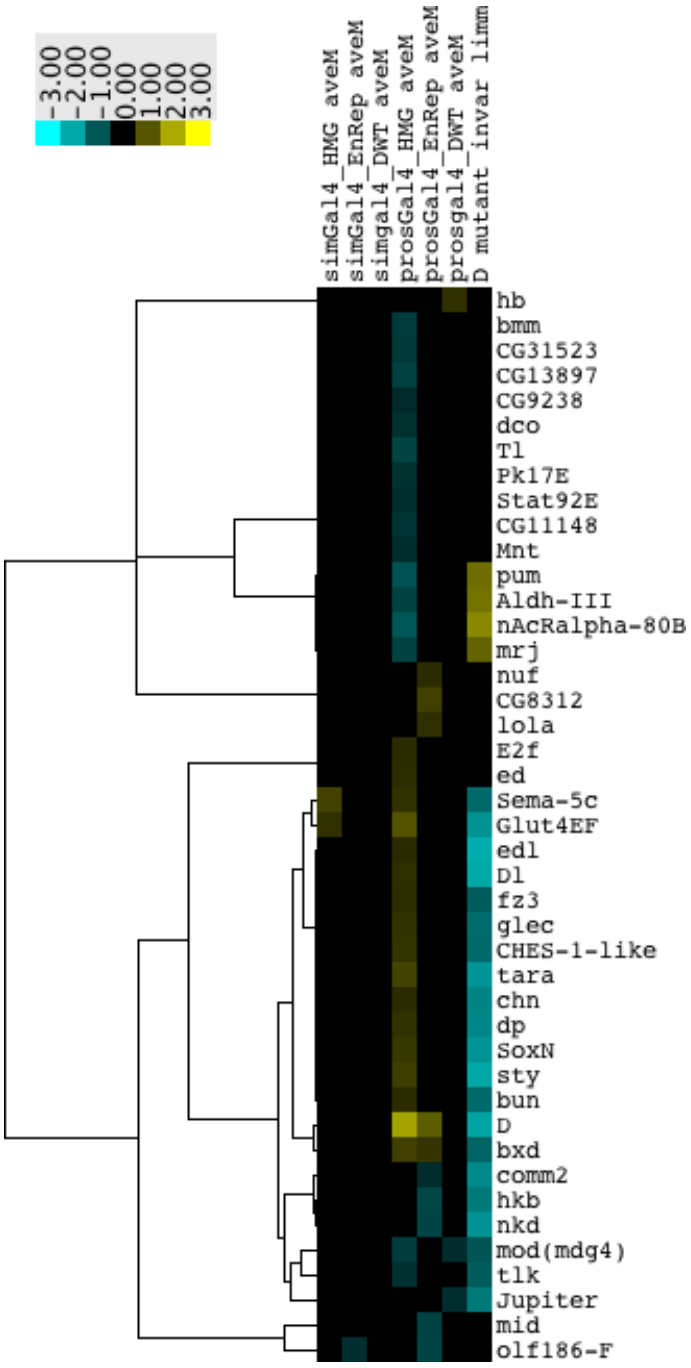

Figure 6.3: Clustering of genes differentially expressed in neuroblasts which are associated with 4 or more Stage 9 Dichaete binding intervals (as estimated from accessibility data).

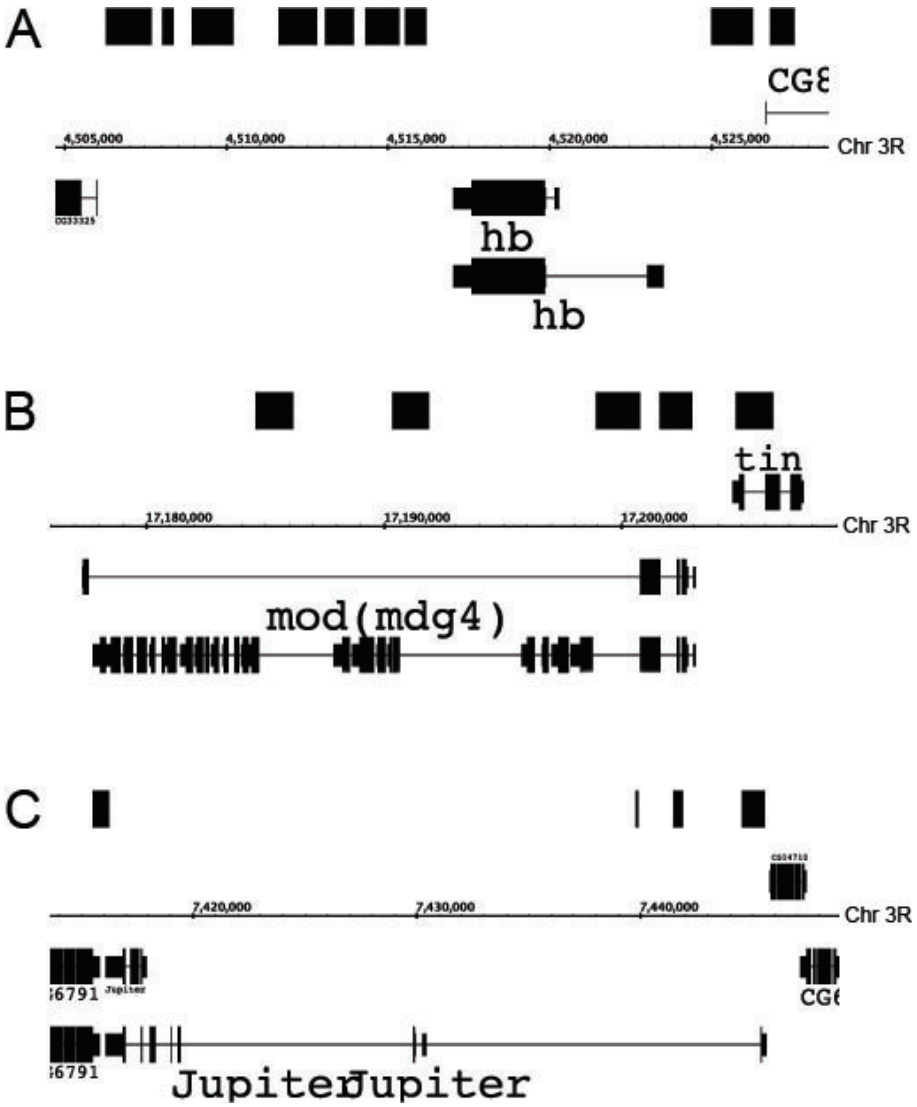

Figure 6.4: Dichaete binding during Stage 9. The gene *hb*, which is thought to be activated by Dichaete, is associated with intergenic Dichaete binding (A), whereas *mod(mdg4)* and *Jupiter*, which are thought to be repressed, show a lot of intronic Dichaete binding (B, C).

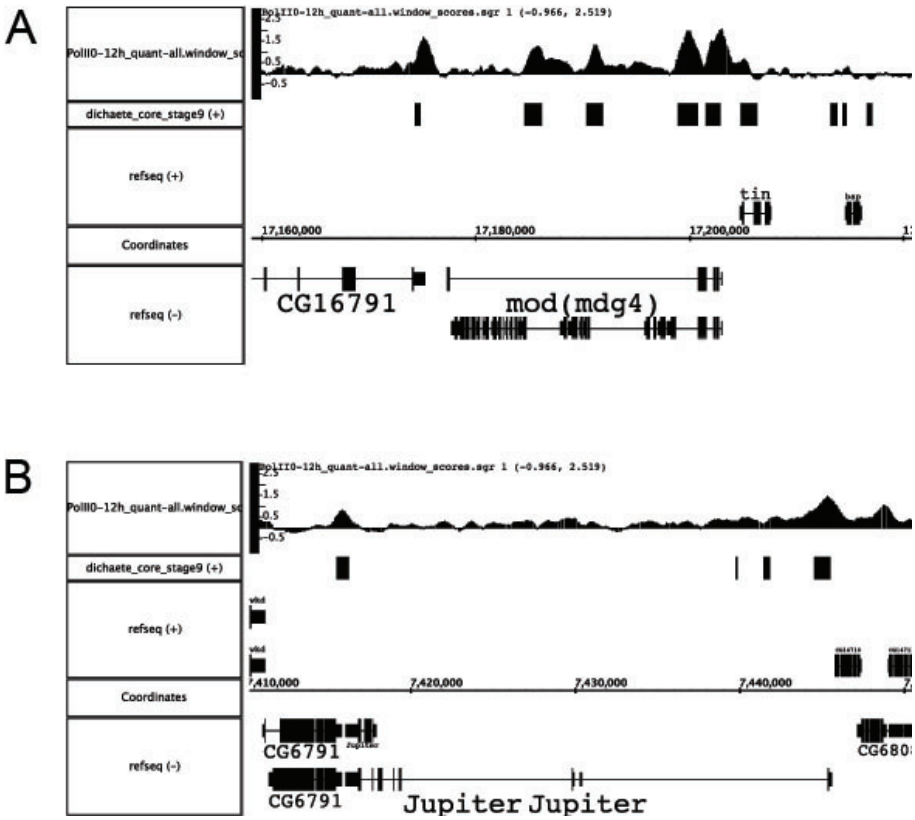

Figure 6.5: Dichaete binding during Stage 9. Dichaete binding within genes corresponds closely to Polymerase II peaks (from a 0-12h ChIP-chip) in *mod(mdg4)* (A), and partially in *Jupiter*

with binding intervals statistically enriched for the Dichaete binding motif (method discussed in Chapter 4), providing further evidence that these are indeed direct targets. Encouragingly, the known Dichaete midline target *slit* is amongst these 83 genes.

The set of 133 multiply bound genes displays a large number of significant and highly specific GO enrichments: 76% are involved in developmental processes (101 genes, p-value 5.9E-32), with specific processes highlighted including imaginal disc development (55 genes, p-value 8.7E-35), pattern specification (49 genes, p-value 6.2E-28) and nervous system development (64 genes, p-value 4.3E-19). A number of genes are involved in transcriptional regulation (56 genes, p-value 3.3E-56) and interestingly, 21 of the 133 encode a homeobox domain (p-value 2.1E-14 or a 'homeodomain-like' domain. This highlights the likely interaction of Dichaete with a number of Hox genes, and may be key to some of the regulatory activity of Dichaete during development.

33 of the genes have been flagged in a microarray screen for neural precursor genes (Brody et al., 2002), and 27 were identified in a screen profiling gene expression in CNS midline cells (Kearney et al., 2004). 16 genes were also featured in a study on direct responses to Notch activation (Krejci et al., 2009), and 33 were highlighted in a study examining the genetic network influenced by the Notch effector Mastermind (Kankel et al., 2007). *mastermind* is a particularly interesting target: it has a reported role in *Drosophila* midline development (Zhang et al., 2011), and since other direct targets are in the same genetic network, it is likely that Mastermind and Dichaete work closely together during nervous system development. While this gene was not flagged in the midline gene expression study, the lack of sensitivity of that particular screen has already been noted. Further to that, 12 genes are connected to nerve cord development (Skeath and Thor, 2003) and 14 were identified in a screen for genes regulating the specification of abdominal neurons (Gabilondo et al., 2011). 12 are identified as factors regulating segmentation (Noyes et al., 2008). Taken together, these overlaps with gene sets identified in screens for neural function highlights the specific nature of the filtered gene set in relation to Dichaete functions during development, specifically emphasizing a key role in CNS development.

Patterns of Dichaete binding at high confidence targets

The Dichaete binding data deconvolved by time (described in detail in Chapter 4) was used to look at the patterns of binding around a selection of the direct targets (Figure 6.6). The direct targets examined display a number of different binding peaks, some static and some dynamically bound at specific developmental stages. The quantity of binding present at different targets is variable, ranging from only a few peaks (minimum 4 for the high confidence set of targets, because of the selection process), to numerous peaks both upstream and within the gene, particularly striking in the case of *Delta*.

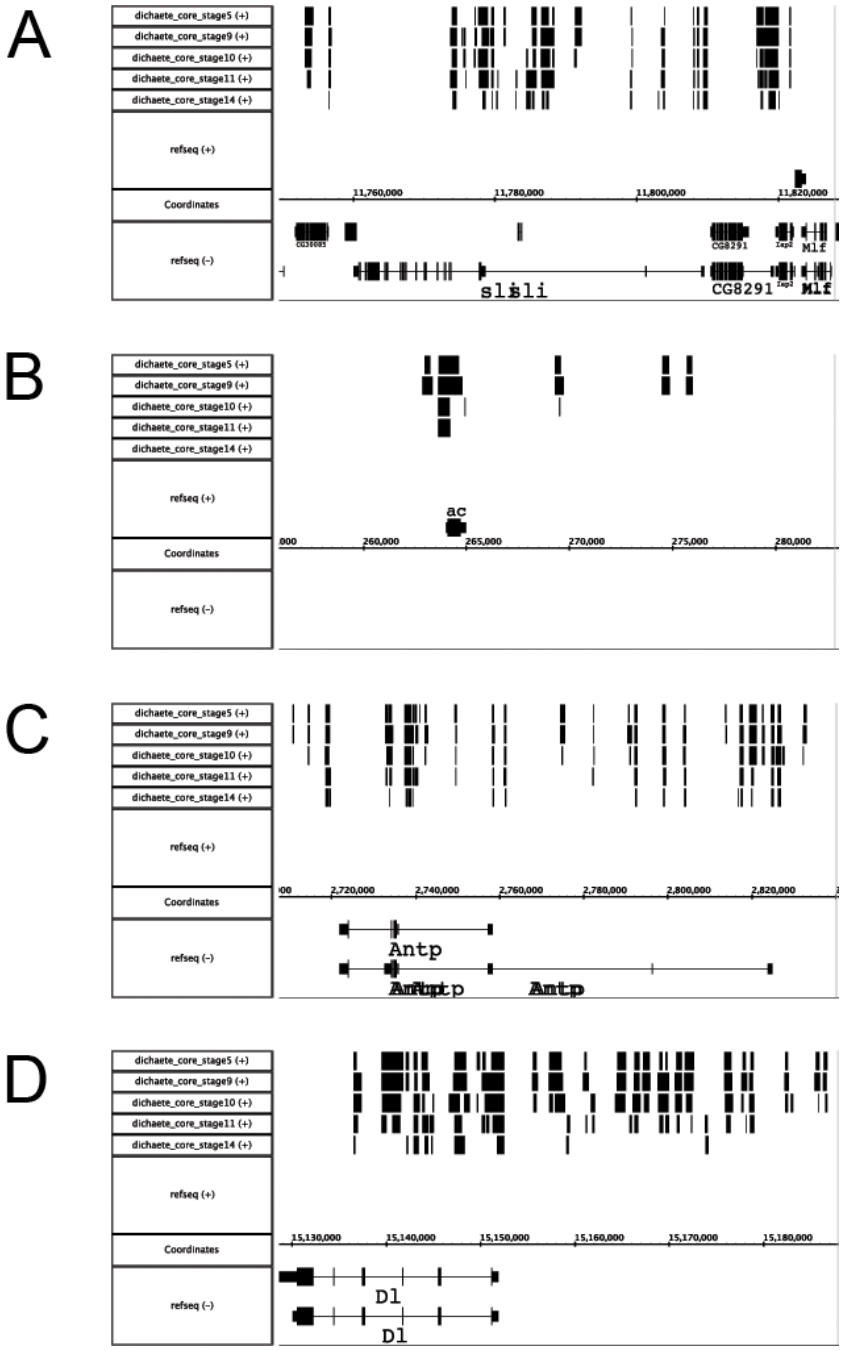

Figure 6.6: Dichaete binding at different stages of development in the vicinity of a number of selected targets. Rows in descending order show Dichaete binding at stages 5, 9, 10, 11 and 14. All targets show multiple dynamic binding, with the binding decreasing at later stages. Some binding persists at stage 14 for *sli*, *Antp* and *Dl* (A, C, D), but disappears completely in the case of *ac* (B).

There is also a trend for the genes to be more highly bound during earlier stages of development and for some of the binding to disappear by stages 11 and 14, which makes sense given the very active role of these genes during early embryogenesis in particular.

6.3 Dichaete and its cofactors

Since it is known that transcription factors rarely, if ever, act alone (Davidson, 2001), and that Sox genes in particular can change their mode of action depending on the cofactors they are paired up with (Kamachi et al., 2000), one of the main aims of my PhD was to determine which transcription factors are likely to be Dichaete cofactors during embryogenesis. Measures of genome-wide binding co-occupancy at different developmental times can be used as an indication of potential interactions between factors, since it stands to reason that if factors are binding the same targets at the same time, they may be acting in unison. The interaction can be direct, with proteins physically interacting with one another, or indirect, with different factors acting at the same cis-regulatory modules. Simply looking at factor overlap is likely to yield false positives, since it is known that transcription factors can bind non-specifically to exposed pieces of DNA. Thus it is critical to assess the statistical significance of genome-wide overlap, rather than solely focusing on individual CRMs using data from genome-wide studies. For this part of the analysis, embryonic transcription factor ChIP datasets from modENCODE (Celniker et al., 2009; Negre et al., 2011), BDTNP (Li et al., 2008; MacArthur et al., 2009) and an individual study on Prospero (Choksi et al., 2006) were used. The resampling-based method implemented by David Huen in the Cooccur R package was used for assessing the statistical significance of overlaps (Huen and Russell, 2010).

6.3.1 Data collection and parsing

There were a number of difficulties in analysing the modENCODE data from scratch. In particular, the original raw data files were often ambiguously labelled and my reanalysis would frequently generate very different numbers of peaks from that reported on the modENCODE website: the reasons for this were not immediately obvious. Compounding this problem, as discussed in Chapter 3, datasets can be sensitive to specific parameters used for analysis and optimal parameters can vary between datasets. This makes a complete reanalysis of a large number of datasets non-trivial. Therefore I decided that, for the initial analysis, a pragmatic approach would be to use the intervals

generated by modENCODE and subsequently reanalyse a selection of data highlighted during the initial screen.

Datasets with less than 10 peak intervals were ignored since independent replicates of the ChIP studies in question showed thousands of binding intervals: thus the datasets with a small number of binding intervals were considered uninformative and likely to be noisy and unreliable. In total 126 datasets were downloaded from the modENCODE website: all data was in gff3 format and all the coordinates were in Drosophila Genome Release 5. The complete list of datasets, along with the number of peak intervals present in each file, is shown in Table 9 in Appendix 4.

Along with the modENCODE data, 36 ChIP-chip datasets from 2-3h of embryonic development were downloaded from the Berkeley Drosophila Transcription Network Project and remapped to Genome Release 5 using the UCSC LiftOver tool and custom-written Perl scripts. Raw Prospero embryo DamID data were taken from Choksi et al. (2006), reanalysed with TiMAT and used in the analysis.

When multiple datasets were available for a particular transcription factor or protein modification and time period (e.g. HDAC1, 0-12h), the union of all available intervals was used. This was done with both modENCODE and Berkeley data. After the unions were performed, there were a total of 102 unique datasets, used for further analysis.

6.3.2 Gene associations

The modENCODE and Berkeley datasets were used to construct a genetic interaction network. The list of all 14833 currently known *Drosophila melanogaster* genes was downloaded from Flymine (including the start and end coordinates, FBgn number, gene name and strand orientation). The genes encoded by the mitochondrial genome were excluded, leaving 14795 genes. It was verified that all FBgn numbers in the list are indeed unique identifiers, and these were used as gene names for further analysis.

The gene associations were then determined for each binding dataset using the method described in previous Chapters. If there were 1 or more genes directly hit by the interval, all of those were counted as hits (as it is impossible to disambiguate between 'meaningful' and 'non-meaningful' hits in this scenario). A direct hit is defined as the gene coordinates and the binding intervals overlapping by 1 nucleotide or more. If an interval had no direct gene hits, the nearest gene was found and counted as a hit, up to a maximum distance of 10kb. If both a gene upstream and a gene downstream were at exactly the same distance from the interval, the downstream gene was chosen as more likely to be regulated. All the results were then tabulated, with each row being a gene and each of the columns a ChIP dataset (list detailed above). '1' signifies that

the gene is hit in the dataset, '0' signifies that it is not. This is considered a useful resource for studying binding clustering at individual genes.

6.3.3 Overlap analysis

The resampling-based Cooccur software was used to do a series of pairwise comparisons to assess the probability that the binding patterns of particular transcription factors significantly overlap. A list of transcription factors was selected for the initial analysis, with the datasets used shown in Table 10, Appendix 4. The focus of the analysis is on Dichaete, for which the high confidence dataset covers 0-12 hours of embryogenesis, so it was thought that, at least for the initial analysis, keeping the more precise time points for the other datasets was not necessarily useful. Therefore when multiple datasets falling within the 0-12 hour period were available for the same transcription factor they were unified into a single dataset. Multiple datasets were only available for a minority of the transcription factors (6 transcription factors, 13 datasets in total: Cad, Hairy, Hkb, Kr, Run and Ubx), so the merging process should not significantly change the analysis.

6.3.4 The Dichaete regulatory network

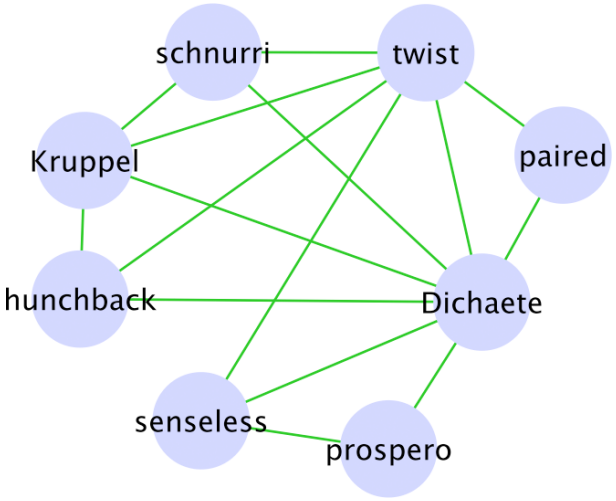

Figure 6.7: The network of transcription factors whose binding profiles significantly overlap (p-value < 0.05), as determined by a resampling based test. This particular figure shows only the network in the proximity of Dichaete - 4 significant overlaps were left out to increase the clarity of the figure.

Initially, a number of transcription factors were found to significantly overlap (Table 6.2, Figure 6.7). It is worth noting that this summary of results was obtained by doing all pairwise comparisons between transcription factors, rather than just comparing their

binding profiles to Dichaete. Therefore the presence of a number of significant overlaps between Dichaete and other factors is a reflection of Dichaete sharing a high degree of significant cooccupancy with other factors, rather than being biased towards identifying Dichaete-specific results. Once the p-values were corrected for multiple testing (using an FDR based p-value adjustment), the number of significant overlaps decreased from 19 to 12. It was decided that the original list of 19 would still be interesting to proceed with for the purposes of this analysis. The argument for this was that, while the p-values are not stringent enough to properly prove a significance of overlap in a statistical sense, these were still highly overlapping datasets, and the fact that they were initially flagged in the screen indicates that they were (more significantly so than the other 542 combinations of pairs of datasets tested). Using the more stringent set of candidate interactions is more solid proof, but is done at the danger of filtering out potentially interesting interactions by increasing the stringency of the scan. Instead, the initial 19 interactions were used as a starting point for an exploratory analysis, with the caveat that not all of them are statistically significantly overlapping in a strict sense.

Interestingly, the list of significant overlaps includes Hb and Kr (also included in the more stringent list using corrected p-values), which are the first two transcription factors active in a temporal series of transcription factors active in neuroblasts to give rise to a diverse set of neurons/glia. Since Dichaete is also expressed in neuroblasts during this period it is likely that it may interact with Hb and Kr in the modulation of neuroblast identity. The Hb dataset used for analysis is from a different period of development (blastoderm, rather than the developing neuroblasts), so it is possible that the pattern of binding is actually quite different in the neuroblasts, but nonetheless it seemed like an interesting connection worth investigating. The proteins encoded by *prospero* (*pros*) and *senseless* (*sens*) also have a role in developing neuroblasts, but specifically during differentiation, and the connection between them and Dichaete is also interesting, suggestive of an interaction in the developing nervous system that may be cooperative or antagonistic.

Of the datasets found to significantly overlap with Dichaete, Prd is a strong candidate for directly physically interacting with Dichaete. Prd is a member of the *Pax* gene family, and Pax6 has been shown to interact with Sox2 in mammals (Kamachi et al., 2001). Additionally, it has been found that the Sox/Pax interaction is conserved between *Drosophila* and chicken: it was found that the same enhancer region regulated by Sox2 and Pax6 in chicken can successfully be activated by SoxN and the *Drosophila* Pax2 homologue Shaven (Blanco et al., 2005). Thus it is possible that a similar interaction is taking place between Dichaete and Prd, most likely during segmentation in the blastoderm.

Data set 1	Data set 2	Uncorrected p-value	FDR corrected p-value
bcd_2_3h_ChIP_Berk_fdr1_union_rel5	twi_2_3h_ChIP_Berk_fdr1_union_rel5	0.0070	0.245
cad_0_8h_ChIP_timeunion_rel5	sens_4_8h_ChIP_union_rel5	0.0010	0.047
Dichaete_core_uberset	hb_2_3h_ChIP_Berk_fdr1_union_rel5	0.0010	0.047
Dichaete_core_uberset	Kr_0_8h_ChIP_timeunion_rel5	0.0010	0.047
Dichaete_core_uberset	prd_2_3h_ChIP_Berk_fdr1_union_rel5	0.0070	0.245
Dichaete_core_uberset	prospero_4_7h_FDR25_DamID_Brandlab_rel5	0.018	0.593
Dichaete_core_uberset	sens_4_8h_ChIP_union_rel5	0.0010	0.047
Dichaete_core_uberset	shn_2_3h_ChIP_Berk_fdr1_union_rel5	0.0030	0.120
Dichaete_core_uberset	twi_2_3h_ChIP_Berk_fdr1_union_rel5	0.0010	0.047
h_0_8h_ChIP_timeunion_rel5	twi_2_3h_ChIP_Berk_fdr1_union_rel5	0.029	0.903
hb_2_3h_ChIP_Berk_fdr1_union_rel5	Kr_0_8h_ChIP_timeunion_rel5	0.0010	0.047
hb_2_3h_ChIP_Berk_fdr1_union_rel5	twi_2_3h_ChIP_Berk_fdr1_union_rel5	0.0010	0.047
kni_2_3h_ChIP_Berk_fdr1_union_rel5	Kr_0_8h_ChIP_timeunion_rel5	0.044	1.0
Kr_0_8h_ChIP_timeunion_rel5	shn_2_3h_ChIP_Berk_fdr1_union_rel5	0.0010	0.047
Kr_0_8h_ChIP_timeunion_rel5	twi_2_3h_ChIP_Berk_fdr1_union_rel5	0.0010	0.047
prd_2_3h_ChIP_Berk_fdr1_union_rel5	twi_2_3h_ChIP_Berk_fdr1_union_rel5	0.0010	0.047
prospero_4_7h_FDR25_DamID_Brandlab_rel5	sens_4_8h_ChIP_union_rel5	0.0010	0.047
sens_4_8h_ChIP_union_rel5	twi_2_3h_ChIP_Berk_fdr1_union_rel5	0.0010	0.047
shn_2_3h_ChIP_Berk_fdr1_union_rel5	twi_2_3h_ChIP_Berk_fdr1_union_rel5	0.0020	0.086

Table 6.2: Results of pairwise comparisons of overlap using a resampling based test (Cooccur). All pairs of binding datasets found to be significantly overlapping (p-value < 0.05) are shown in the table.

6.3.5 The role of Dichaete during neuroblast divisions

It is known that one of the mechanisms for specification of neuroblast fates is the temporal sequence of transcription factor expression that different neuroblasts go through (Maurange and Gould, 2005; Maurange et al., 2008). The first two transcription factors in the sequence are Hb and Kr, which were found to significantly overlap with Dichaete binding (Figure 6.8). While binding data is not available for Pdm (a.k.a. Nubbin), the third factor in the sequence, this is a POU protein which has a reported interaction with Dichaete and is likely to be a genuine cofactor (Ma et al., 1998; Mace et al., 2010). In the larval brain, which appears to generate neuroblasts using the same temporal mechanism as the embryo, Maurange et al. (2008) report that Dichaete is present during the first three neuroblast divisions and is then switched off by Cas. They also find that unless *Dichaete* is switched off, neuroblasts keep on dividing instead of exiting the cell cycle. This is consistent with Dichaete maintaining neuroblasts in a self-renewing state, which is reminiscent of the role of Sox2 in vertebrate stem cells.

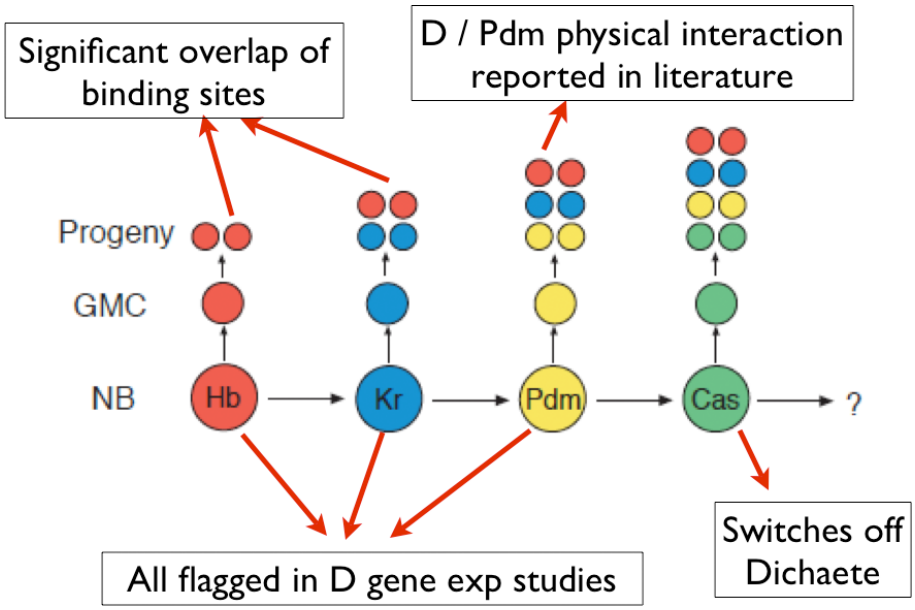

Figure 6.8: Temporal sequence of transcription factors in dividing neuroblasts and its relation to Dichaete function. Dichaete is present during the first 3 divisions, then is switched off by Cas. Original figure taken from Maurange and Gould (2005).

The presence in the same cells along with significant overlap of binding suggests that Hb and Kr may be cooperating with Dichaete during early neuroblast divisions to regulate the same genes. This is also likely to be the case for Pdm, though the data is not currently available. One of the potential ways in which Dichaete may be acting in neuroblasts is by regulating both differentiation factors and apoptosis genes (Figure 6.9). Since *hb* was one of the high confidence targets identified in neuroblasts, it is also likely that there are feedback loops between Dichaete and the neuroblast temporal transcription factors.

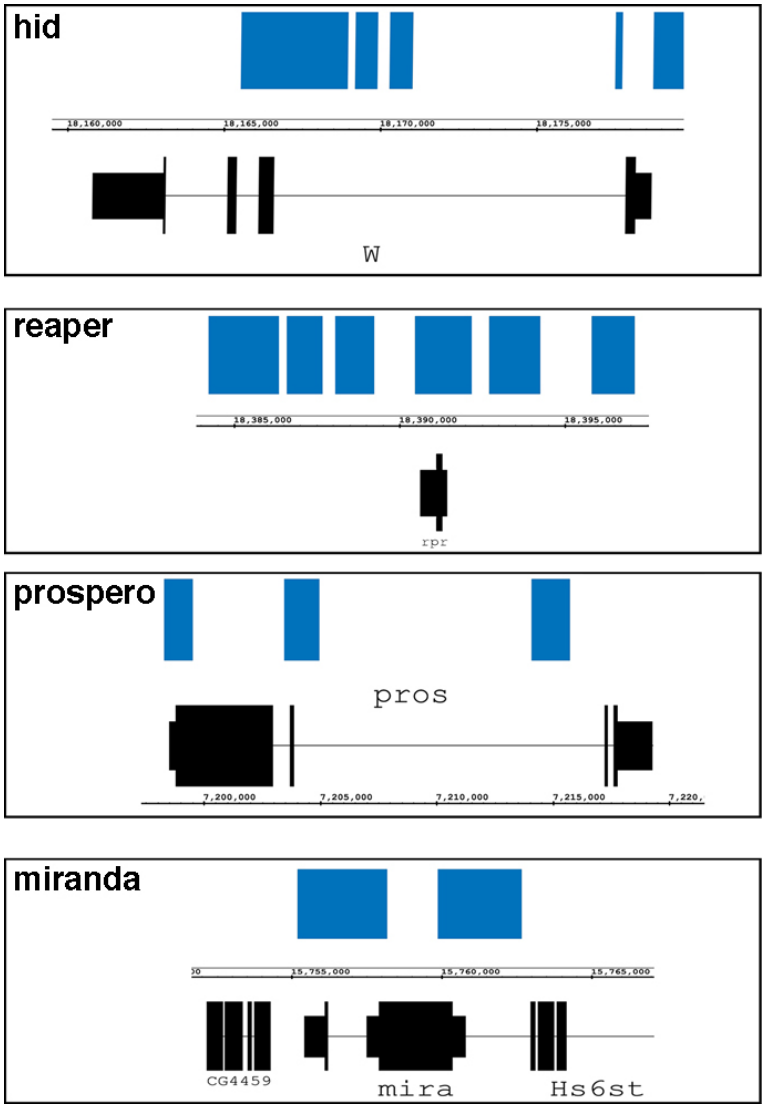

Figure 6.9: Dichaete may act in neuroblast by binding to and regulating both apoptosis and differentiation genes. Dichaete binding to apoptosis genes *hid* and *reaper* and differentiation genes *prospero* and *miranda* are shown.

The initial involvement of Dichaete in neuroblast development likely starts quite early, probably at the selection and delamination stage, due to its involvement with the Notch signalling pathway. The role of Dichaete at this stage is already documented, as it is known to act together with Vnd and Ind to establish differences in the positional information along the DV axis (Zhao and Skeath, 2002). It is likely that additional neurogenesis targets at this time help establish the exact positional information that later specifies neuroblast identity.

Subsequently, since *hb* was upregulated when Dichaete was overexpressed in neuroblasts, it seems likely that Dichaete helps activate the first step of the transcription factor temporal sequence in neuroblasts. It is then possible that Dichaete cooperates with each of the transcription factors in turn to activate the next one in the sequence - *Kr* and *Pdm* were both strongly downregulated (average M value < -1.5) in the *Dichaete* null mutant. Dichaete may also help activate *Svp*, which facilitates the transition between *hb* and *Kr* expression in the temporal sequence.

6.3.6 The role of Dichaete in neuroblast differentiation

In terms of maintaining pluripotency, it is possible that Dichaete acts to maintain the expression of *miranda* to ensure Prospero remains in the cytoplasm, since the expression of *miranda* is strongly downregulated in the *Dichaete* null mutant (average M-value < -2). It is likely that at the same time it blocks the expression of apoptosis genes (Figure 6.9) to ensure that neuroblasts do not get a chance to exit the cell cycle in any way. However, while this hypothesis makes sense in terms of the biology, in the *Dichaete* mutant experiment, *hid* and *rpr* are both strongly downregulated, contrary to the hypothesis. A number of possibilities exist. First of all, the hypothesis may be incorrect and Dichaete might be activating apoptosis genes instead. However, this seems unlikely given that Dichaete prevents neuroblast cell cycle exit rather than promoting it. The other possibilities are that either the changes in expression are due to downstream effects from a number of other key transcription factors factors being affected in a *Dichaete* mutant, or that Dichaete is acting context-specifically and its effect on apoptosis genes in neuroblasts is contrary to the global effect in the whole embryo, and so is not picked up by the *Dichaete* mutant study. Verifying the changes in expression of *hid* and *rpr* in neuroblasts of *Dichaete* mutant would resolve at least the latter theory.

There is clearly a link between Dichaete and Prospero, given the high occurrence of binding in the same genomic sites (Figure 6.10). However, it is unclear what the exact relationship between them is. It is possible that Dichaete also activates *prospero*, but makes sure that it is out of the nucleus by regulating *miranda*. Using the standard

cut-offs discussed in the previous chapter, there are no significant changes of *prospero* expression in the Dichaete gene expression experiments, despite multiple binding in and around it. However, an expression change in the *Dichaete* mutant does exist, with an M-value of -1.35 and a p-value of 0.02, which would count as significant other a less stringent threshold. This leaves it unclear whether Dichaete is a main regulator of *prospero*, and some further experiments would be useful here. Interestingly, the immunostainings performed on *Dichaete* mutants using a Prospero antibody by Shen (2006) find that the expression of *prospero* is stronger, contrary to the gene expression data. The reasons for this difference may be similar to the ones already discussed for apoptosis genes.

It is possible that Dichaete and Prospero act antagonistically at the same sites. It is not known whether they are ever present at the same time in the same nucleus, but if they are, it is possible that they compete directly for the same sites. It is also possible that they act cooperatively (for example, by Prospero changing the mode of action of Dichaete from a repressor into an activator at selected sites), but given the role of Dichaete in maintaining pluripotency, this option seems less likely.

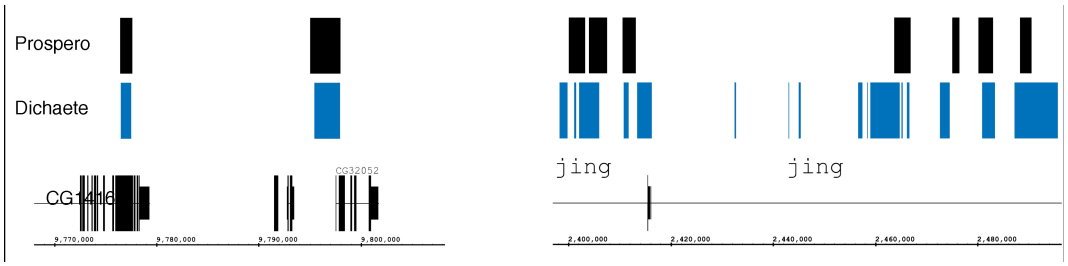

Figure 6.10: Examples of Dichaete and Prospero binding co-occurrence.

6.4 Investigation of Dichaete controlled cis-regulatory modules

It was interesting to explore in more detail the binding patterns of Dichaete and its proposed cofactors around high confidence direct targets. In particular, a list of high confidence direct targets in neuroblasts has already been generated, so was explored further. In case the sensitivity of that particular screen was doubtful, the same analysis was also performed with the direct targets flagged in the genome-wide screen. Two areas were considered of most interest: first, the interaction of Dichaete with Hb, Kr and Pdm during early neuroblast divisions, and second, the interaction of Dichaete and Prospero in the decision to self-renew or exit the cell cycle to differentiate.

6.4.1 Dichaete targets in neuroblasts

For studying Dichaete interactions in neuroblasts, three different binding profiles were available - the Dichaete core dataset, Kr 0-8 hour ChIP-chip and Hb 2-3 h ChIP-chip. Of these, the Dichaete and Kr datasets cover the correct developmental time window, whereas the Hb data is from earlier in development and may be considered an educated guess at where Hb is binding at later stages. The original overlap numbers are shown in Table 6.3. Since Hunchback has the smallest binding dataset, it shows the most overlap with the other two (73% of Hunchback binding intervals overlap with Dichaete, and 76% overlap with Kruppel). A large portion of Kruppel binding intervals (61.5%) also overlap with Dichaete. The percentage of Dichaete overlap appears relatively low, but this is due to the greater total number of intervals than the other two datasets.

Dataset	**Overlap with Dichaete**	**Overlap with Hunchback**	**Overlap with Kruppel**
Dichaete (6227)		1598 (25.7%)	2889 (46.4%)
Hunchback (2180)	1674 (76.8%)		1596 (73.2%)
Krüppel (4553)	2801 (61.5%)	1460 (32.1%)	

Table 6.3: Binding overlap between Dichaete and temporal neuroblast transcription factors Hunchback and Krüppel

The overlap in binding between all 3 datasets was found by taking intersects of binding intervals. There was a total of 1449 binding intervals present in all 3 datasets associated with 1019 genes. As expected, the list of targets was highly enriched for genes involved in developmental processes (460 genes, p-value 1.2E-74) and biological regulation (461 genes, p-value 1.7E-59), with 26% (268 genes) involved in nervous system development (p-value 6.6E-46) and specific components of it such as neuroblast differentiation (27 genes, p-value 6E-18) and axon guidance (72 genes, p-value 7.3E-26).

This list was compared to the list of Dichaete neuroblast targets with 2 or more associated Dichaete binding peaks. It was found that 56 out of 116 direct Dichaete targets in neuroblasts are also joint targets with Hb and Kr. This includes 4 of the 27 neuroblast differentiation genes (*nkd*, *mid*, *hb* and *hkb*) and the binding patterns around these 4 genes was further investigated. In general, the binding intervals for all 3 datasets were quite broad, so it was considered useful to filter them based on accessibility data, focusing only on regions that are likely to be bound at Stage 9. The genes in question were found to have complex cis-regulatory modules, with some sites being bound by all three transcription factors and other sites in the same region showing varying combinations of one or two of them (Figure 6.11). It is likely that there are specific enhancer regions being targeted at different times by different combinations of transcription factors.

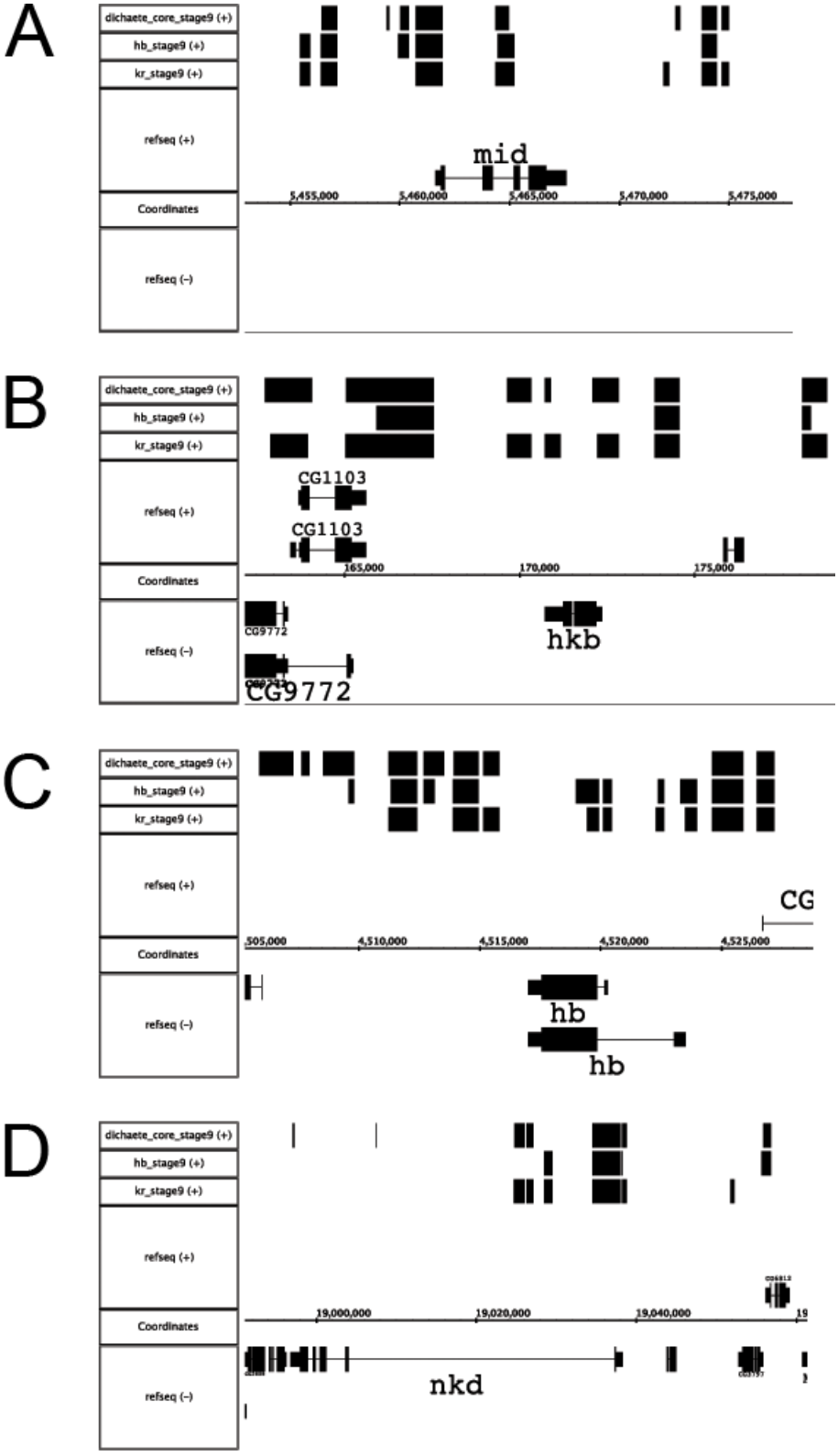

Figure 6.11: Binding of Dichaete, Hb and Kr in the regions of the neural differentiation genes *mid* (A), *hkb* (B), *hb* (C) and *nkd* (D). The transcription factor binding intervals show significant overlap, as well as complex patterns of co-occurrence.

6.4.2 *hunchback* regulation

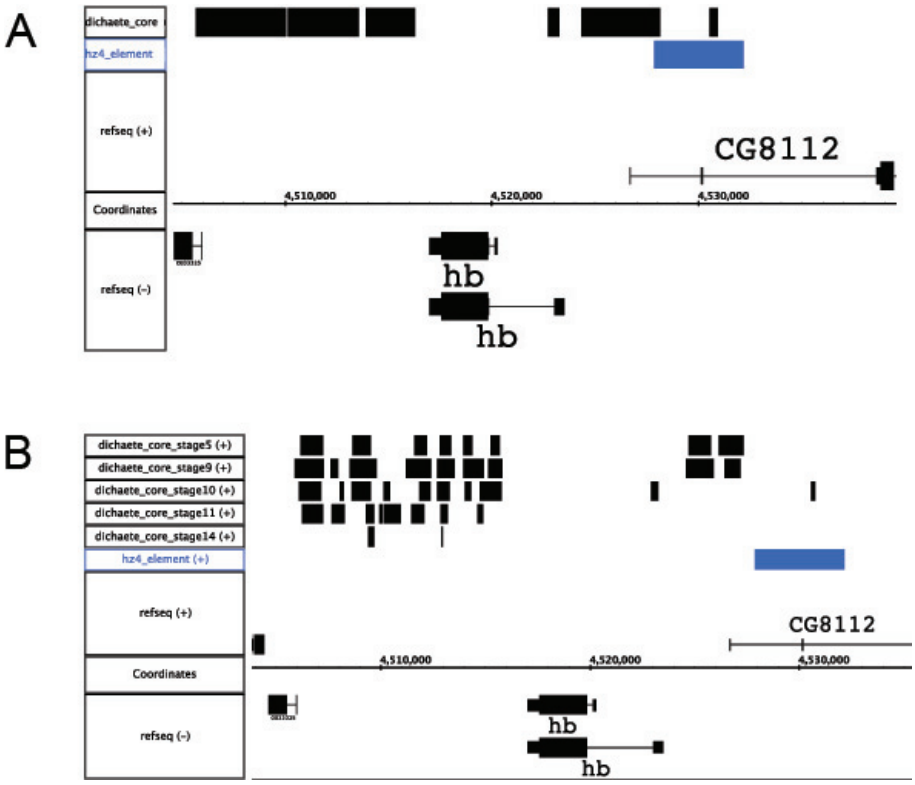

Figure 6.12: (A) Dichaete core binding in the region of *hb* (shown in black) overlaps with the HZ4 enhancer element (blue) which activates the expression of *hb* in neuroblasts and GMCs. (B) Dichaete temporal data indicates that the enhancer is specifically bound at Stage 10 only.

For *hb* in particular, there is published data on specific enhancer regions and their functions (Hirono et al., 2011), therefore I examined Dichaete binding at these enhancers and the co-occurrence of Dichaete and its cofactors. Originally there was a 4 kb enhancer element (the HZ4 element) discovered upstream of the *hb* gene (from coordinates 4,527,893 to 4,532,199 on chromosome 3R) which conferred *hb* expression in neuroblasts and GMCs, but not neurons (Margolis, 1992; Margolis et al., 1994, 1995). Dichaete core binding is found around this *hb* region, and some of it overlaps directly with the HZ4 element (Figure 6.12), thus this element may be associated with Dichaete regulation in neuroblasts. Examining the estimated Dichaete temporal data (Figure 6.12) indicates that part of the HZ4 enhancer is only accessible and likely to be bound at Stage 10 only. Looking at the overall pattern of binding, it does not appear that this enhancer is the main regulatory target of Dichaete since there is apparently more binding both upstream and downstream to *hb*. However, because HZ4 is well characterised and there is an overlap, this was thought worth investigating further.

A further study provided insight into the precise structure of the HZ4 enhancer by looking at the expression of subfragments of the 4kb region (Hirono et al., 2011) and

identifying a number of smaller regions with specific regulatory effects. Six different constructs were created (selected based on conservation) and tested. HG4-7, the 691 bp construct found to be sufficient for driving *hb* expression in neuroblasts and GMCs, partially overlaps with Dichaete core binding data. No overlap is found with Dichaete temporal data, but that may be due to the incomplete nature of the temporal data. The HG4-7 region bound by Dichaete contains a conserved binding site for Svp and a conserved homeodomain core motif as well as Pou domain binding sites slightly further upstream (the closest one being around 150 bp away from the end of the Dichaete binding interval). It seems likely that Dichaete may be interacting with Nubbin close to the Pou domain at some point, and it is possible that the Dichaete binding data comes from a time period that is too broad, since there may be important transient interactions taking place specifically in neuroblasts.

RSAT matrix-scan was used at weight threshold 4 to look for Dichaete binding sites within HG4-7 using the Dichaete position weight matrix derived from ChIP-chip data. Six consensus Dichaete binding sites were found within the region, four of them on the − strand. The two highest scoring binding sites were both on the reverse strand and had weight scores of 8.6 each. The first (coordinates 4,527,919 to 4,527,929) is close to the beginning of the HG4-7 element and overlaps with a Dichaete core binding interval. The second (coordinates 4,528,478 to 4,528,485) is identified as conserved in at least 11 out of 12 Drosophila species (specifically, the 'AACAAT' part of the motif) and is situated next to and partially overlaps a POU motif. It is likely that Dichaete and Pdm physically interact at this site to regulate the expression of *hunchback*.

The other Dichaete binding interval within the HZ4 enhancer fully overlaps with constructs HG4-1 and HG4-3, which are expressed in early born neurons, and very slightly overlaps with HG4-5, which is expressed in late born neurons. This is the Dichaete interval that the accessibility data indicates is only present at stage 10. It perhaps stands to reason that the region required for expression in neuroblasts is bound in earlier stages, whereas the region required for neuron development is bound slightly later in development. Further to that, based on the positioning of predicted binding sites, it seems possible that Dichaete is helping regulate the neuroblast expression enhancer element by creating DNA loops in this region (Figure 6.14).

6.4.3 Motif searching

The binding motif for Dichaete generated from the Berkeley ChIP-chip study was used, and the motifs for Hb, Kr and Nub (Pdm) were downloaded from the JASPAR database (Bryne et al., 2007). At first, motif searching was attempted using nmscan, a part of the NMICA software (Down and Hubbard, 2005). However, it was found

that if the same threshold is used for all motifs, the number of binding sites found varies greatly. As expected, the shorter and less specific motifs were found an order of magnitude more frequently than the longer and more specific ones. Next a threshold optimisation was attempted using experimentally validated binding sites from REDfly (Halfon et al., 2007). Unfortunately, only *hb* had anywhere nearly enough known binding sites (103 in total) to attempt this, so the strategy was abandoned. Therefore I used a different motif search tool, matrix scan, part of the RSAT package (Thomas-Chollier et al., 2008), which normalizes the output scores to take the motif length into account (settings used were Markov Order 1 and the high stringency threshold >= weight score 6). In the cases where two overlapping motifs were found, the higher scoring one was selected.

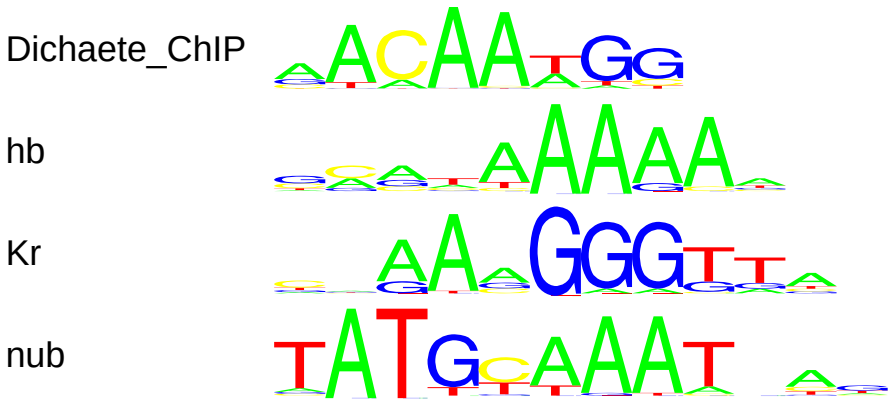

Figure 6.13: TF binding motifs used

To start with, just the *hb* HZ4 enhancer was scanned, as it was expected that there may be some interaction between the four factors here. The HZ4 enhancer was found to have 5 Dichaete, 3 Kr, 9 Hb and 5 Nub binding sites (Figure 6.14). The distance of the nearest other transcription factor sites within a 1 kb range from either side of each Dichaete site was assessed. As expected if this is indeed a cofactor, the Nub binding sites were closest to Dichaete binding sites, with the closest one being 130 bp away. This is actually much further away from Dichaete than the POU motif found by Hirono et al. (2011), and this is because for this analysis, the Nubbin position weight matrix from JASPAR was used. There were also some Hb binding sites relatively nearby, with the closest one being 210 bp away, though this may simply reflect a high number of potential Hb binding sites found due to the low specificity of the position weight matrix for this factor. The Kr binding sites were far away, with the closest being 507 bp away.

The same analysis was repeated for the genes identified as high confidence targets in neuroblasts, to see whether more general patterns emerge. The analysis was performed using the Dichaete high confidence intervals associated with direct neuroblast targets. The 116 genes associated with more than one Dichaete interval and also flagged in at

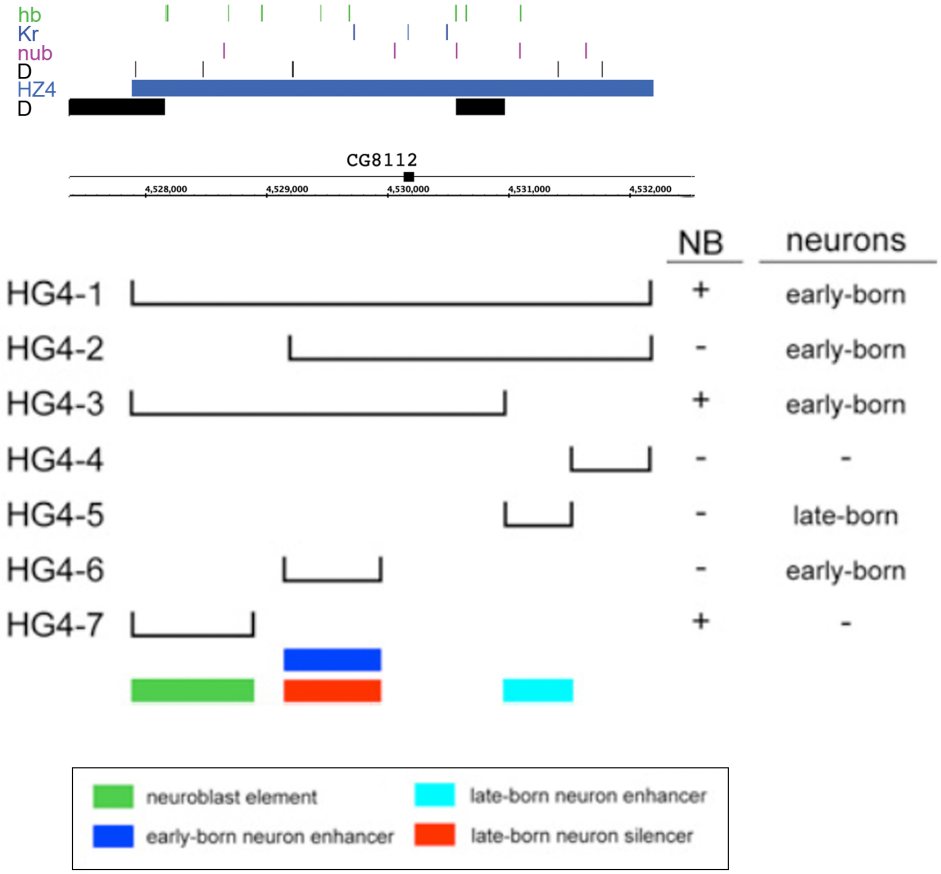

Figure 6.14: A more detailed view of Dichaete action in the HZ4 enhancer region. The top of the figure shows the predicted binding sites for Hb (green), Kr (blue), Nub (pink) and Dichaete (black), as well as the position of the Dichaete core binding intervals (black rectangles) and the HZ4 enhancer (blue rectangle). The different enhancer regions investigated by Hirono et al. (2011) are shown below.

least one neuroblast gene expression study were found to be associated with a total of 428 Dichaete intervals. These intervals were scanned with binding motifs for Hb, Kr, Dichaete and Nub (Figure 6.13) using RSAT, with the settings described above. The search identified 323 binding sites for Dichaete, 204 for Kruppel, 403 for Hunchback and 260 for Nubbin. The distances of the other predicted binding sites from Dichaete were then analysed, to check whether a particular distance tends to occur frequently. For each Dichaete binding site, the closest sites within a 1 kb window on either side were found (Figure 6.15). There appears to be a trend in site distances from Dichaete for all 3 transcription factors, with all of them being most likely to be found roughly 100 bp away from Dichaete. Interestingly, the distance between the Dichaete and Vvl binding sites in the *slit* regulatory region was proposed to be between 50 and 70 bp (Ma et al., 2000), so the distances found here constitute plausible ones for cofactor action.

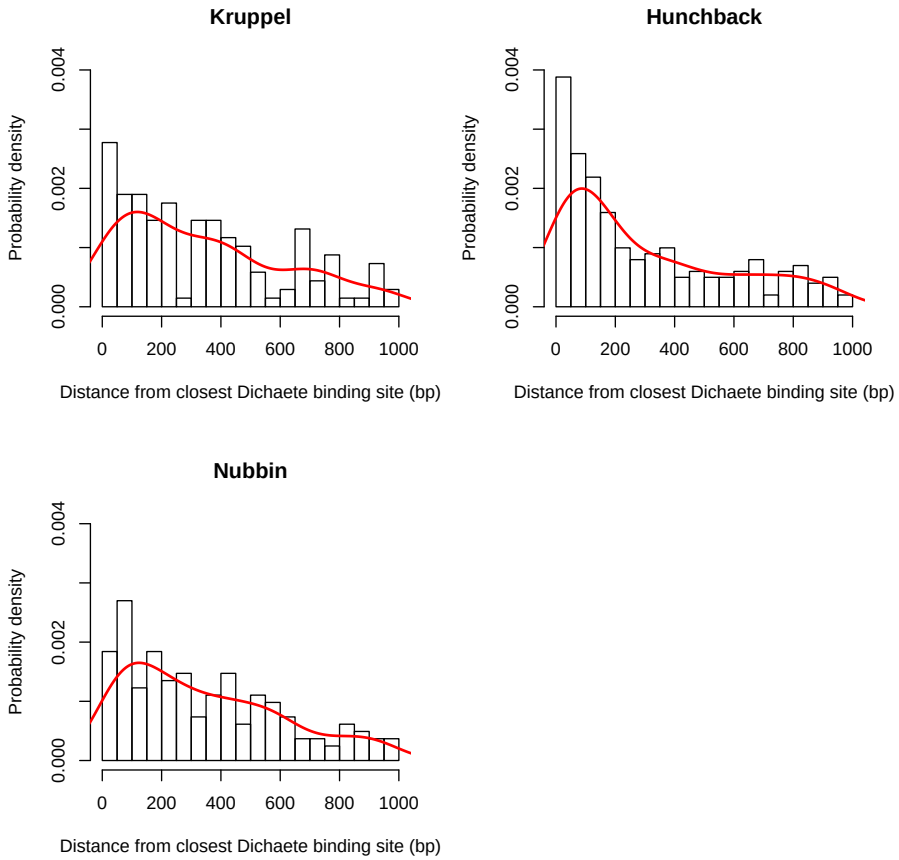

Figure 6.15: Distance of different transcription factor binding sites from Dichaete binding sites. The sites were predicted using RSAT. Only the sites closest to each Dichaete binding site are shown. There appears to be a peak about 100 bp away from Dichaete binding sites for all 3 transcription factors.

6.4.4 Dichaete and Prospero

Prospero is a transcription factor that facilitates the differentiation of neuroblasts into neurons. In dividing neuroblasts, Prospero is kept outside of the nucleus by Miranda, which anchors it to the basal cortex of the cell, resulting in asymmetric division in which Prospero ends up in the ganglion mother cell. At that point, Prospero is released and enters the nucleus, regulating gene expression to bring about differentiation (Choksi et al., 2006). Dichaete and Prospero binding intervals were found to significantly overlap, and the connection between them was thought to be potentially interesting. Work so far indicates that they operate antagonistically: while the presence of Dichaete ensures the continued proliferation of neuroblasts, the presence of Prospero in the nucleus brings about neuron differentiation (Choksi et al., 2006). It is therefore likely that they are targeting the same genes, but with opposite effects. However, direct evidence for the relationship between Dichaete and Pros is currently lacking. For example, it is not known currently known whether both factors are in the nucleus at the same time at any point. It is possible that in the GMC, *Dichaete* expression is quickly switched off, or that Dichaete is outcompeted by Prospero at relevant binding sites. Alternatively, it is possible that Dichaete and Prospero in combination can have the opposite effect from Dichaete paired with other transcription factors. This is an interesting area to further research given the role of Sox2 and mammalian Prox proteins in regulating neurogenesis (Torii et al., 1999). Here I present a preliminary analysis of their relationship.

The connection between Dichaete and Prospero has been previously highlighted by Shen (2006) in her PhD work on Dichaete. The effect of dominant negatives and a *Dichaete* null mutant on *prospero* expression was characterised using immunohistochemistry. It was found that in *Dichaete* null mutants, the expression of *prospero* is much stronger, and the layer formed by neuroblasts and GMCs is much denser compared to wild type, suggesting an increase in neuroblast proliferation. This is unusual because a study of Dichaete expression specifically in neuroblasts found that Dichaete promotes continued proliferation, and needs to be switched off in order for neuroblasts to exit the cell cycle (Maurange et al., 2008). However, the phenotype could be due to downstream effects, for example defects on the Notch signalling pathway in *Dichaete* mutants. At stage 16, glial differentiation defects are present, including loss and disorganisation of the longitudinal glia (Shen, 2006). When looking at dominant negatives, both HMG^{-} and EnRep display a loss of *prospero* expression. In the case of EnRep, the repression makes sense and suggests that *prospero* is a direct target, which could be either activated by Dichaete, or slightly repressed (and further repressed by EnRep). In the case of HMG^{-}, it is less clear why it would lead to a reduction of *prospero* expression, unless it is either normally activated by Dichaete, with HMG^{-} sequestering

away cofactors, or the reduction in expression is due to indirect effects. While the precise effect of Dichaete on *prospero* expression is unclear, this combined with the high confidence binding data strongly suggests that Prospero is a direct target.

In order to find out more about the possible interaction between Dichaete and Prospero, their binding profiles were compared. The Prospero 4-7h embryo DamID binding data from Choksi et al. (2006) was reanalysed using the TiMAT pipeline and remapped to Genome Release 5 using the UCSC LiftOver tool. The FDR 1% dataset had very few intervals (307 in total), so was thought to be too stringent. Instead, the FDR 25% dataset was used, which contained a total of 1478 intervals. This binding profile was compared to the Dichaete core 0-12 hour dataset (6227 intervals). Intersecting the two binding sets resulted in 1225 interval overlaps, associated with 642 genes. The gene list is highly specific, with 292 genes involved in developmental processes (p-value 1.8E−50), 146 in transcriptional regulation (p-value 1.8E−46) and 86 in cell fate commitment (p-value 1.4E−43). 117 of the genes are involved in the generation of neurons and this is likely to be an interesting target list (including *Antp*, *Abd-A*, *hth*, *ey*, *svp*, *hb*, *Kr*, *nub*, *cas*, *pros*, *mira*, *N*, *Dl*, *sim* and *sli*). 27 out of 35 genes defined as being part of a neural stem cell transcriptional network (Southall and Brand, 2009) (Figure 6.16).

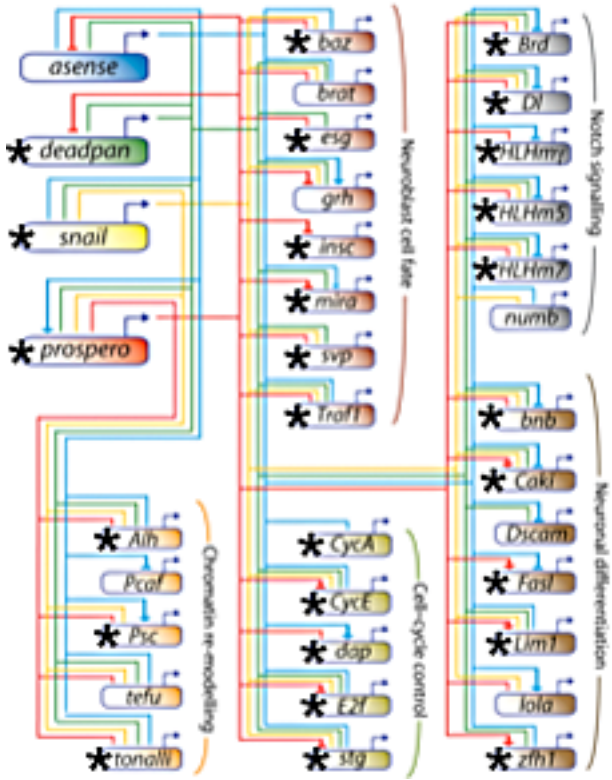

Figure 6.16: The figure, taken from Southall and Brand (2009), shows the neural stem cell transcriptional network as determined from overlapping binding data for Asense, Deadpan, Prospero and Snail. 27 out of the 35 genes are bound by Dichaete and Prospero together in the same genomic locations are labelled with a black asterisk.

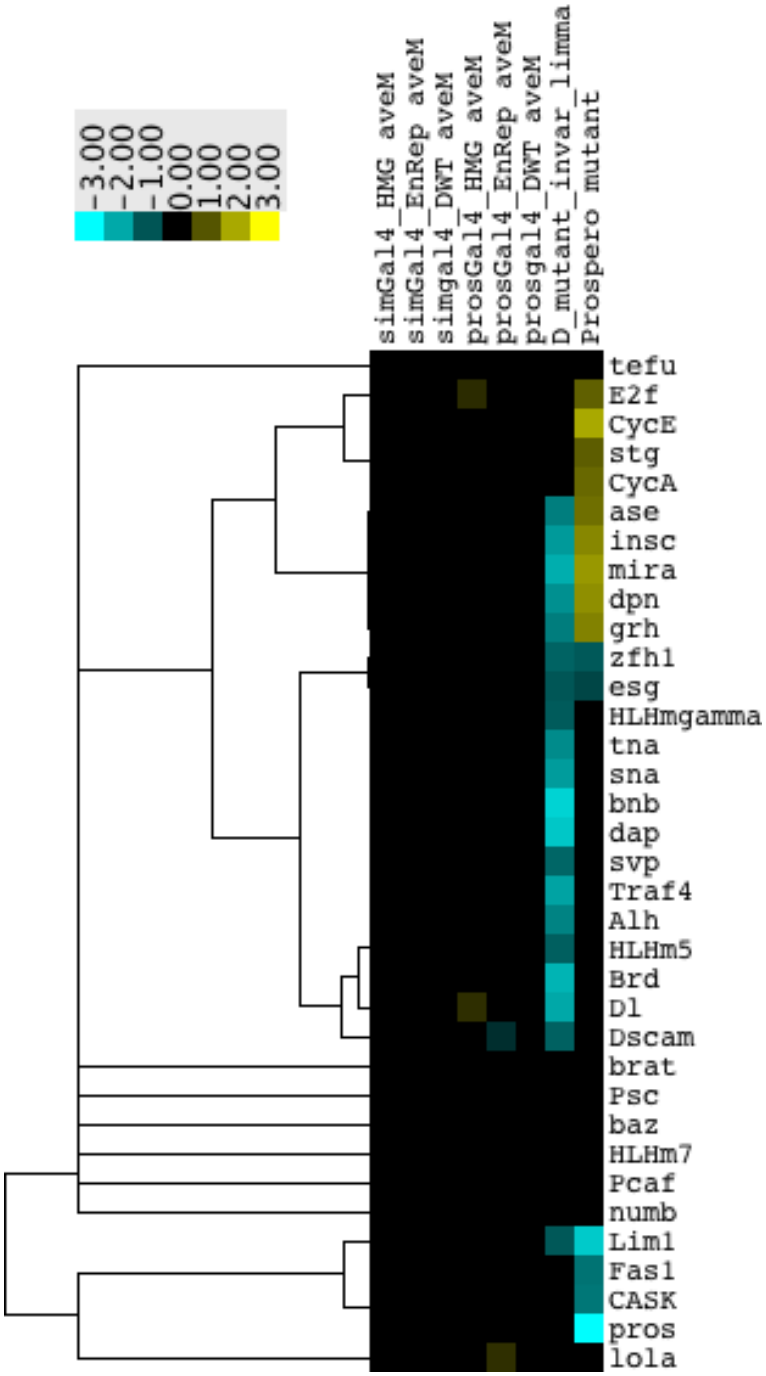

Figure 6.17: The Dichaete gene expression data was compared to the *prospero* mutant expression data from Choksi et al. (2006), looking specifically at genes identified as being part of the neural stem cell transcriptional network by Southall and Brand (2009).

It was thought interesting to further investigate the effects of Dichaete and Prospero on the neural stem cell transcriptional network genes. The log2 fold-change for genes in *prospero* mutants were taken from Choksi et al. (2006) and visualised alongside the Dichaete gene expression experiments (Figure 6.17). While overall not enough genes were differentially expressed to draw definite conclusions, it appears that the majority of genes involved in neuroblast cell fate determination tend to be regulated antagonistically, whereas the two neuronal differentiation genes for which data is available, *zfh1* and *Lim1*, appear to be regulated in unison. This opens up the possibility that Dichaete acts antagonistically to Prospero in neuroblast, maintaining pluripotency and preventing differentiation, but then goes on to act in unison once Prospero enters the nucleus of the GMC, with the two factors together facilitating correct neuronal differentiation.

6.5 Conclusions

This chapter concludes the research part of my PhD. In previous chapters, I presented a high confidence Dichaete binding dataset for the Stage 1-15 embryonic period, along with a series of gene expression studies aimed at functional validation. Furthermore, I developed a number of filtering methods for identifying direct targets and increasing the signal to noise ratio of the data. In this chapter I have made use of these datasets and methods to identify direct Dichaete targets in the hope that this will aid future research on Dichaete and Sox genes

Further to that, I have made use of a large number of available binding datasets to scan for cofactors, and identified links between Dichaete and the neuroblast temporal sequence of transcription factors Hunchback and Kruppel, as well as a link with the differentiation factor Prospero. It appears that Dichaete interacts and cooperates with the first three transcription factors in the temporal sequence, as well as being an integral part of the neural stem cell transcriptional network. This conclusion has a pleasing similarity with the role of its homologue Sox2 in maintaining stem cell pluripotency in vertebrates. The relationship of Dichaete with Prospero appears to be complex, with Dichaete acting antagonistically to Prospero in neuroblasts in terms of preventing cell cycle exit, but potentially acting cooperatively once the differentiation process has begun in order to facilitate correct neuron differentiation.

One of the limitations of this work is that much of it is based on datasets from a range of developmental time periods, so the next step would be to focus on specific stages of development and parts of the nervous system, to more precisely determine the interactions between Dichaete and the various transcription factors involved. It would also be useful to do some validation, which can be performed using immunohistochemistry

and/ or in situs with interesting direct targets to determine the effects of *Dichaete* null mutations and dominant negatives on specific transcription factors. What might be of particular interest is that, while the results presented here have all been performed using whole embryos, it is likely that the effects are non-uniform, so may vary for example in particular neuroblast lineages.

In terms of computational work, further analysis can be performed using existing data that, due to time constraints, I have been unable to complete. I have generated a genome-wide network of binding with the available embryonic binding datasets: this is a useful resource and could be further analysed by using, for example, clustering and principal component analysis to determine clusters of genes that are co-regulated and the extents to which different transcription factors influence gene expression together. While much detailed analysis has already been published by BDTNP and the modENCODE consortium, a similar analysis with a more specific focus on Dichaete has not yet been perfomed. Motif searching could also be used further to predict binding sites and do a more in-depth analysis of their occurrence patterns. However, such analyses should ideally be followed up by in vivo validation, for example by using cloned constructs attached to marker genes in wild type and mutants.

Part of what is interesting about the modENCODE data analysis presented in this chapter is that a non-specific analysis with a large number of different transcription factor binding datasets still points very clearly towards connections specific to CNS development, neuroblasts and neuronal differentiation, processes Dichaete is known to be involved in. The high degree of co-occurrence of Dichaete, Hb and Kr, as well as the presence of their binding motifs, suggest that there are genuine interactions between them. Dichaete and Prospero also appear to have a complex relationship that may be interesting to investigate further. I hope that my work has created an initial roadmap that can guide further Dichaete research.

Chapter 7

Conclusions

The main objective of my PhD project was to provide the first in-depth genome-wide view of the developmental and molecular action of Dichaete during *Drosophila melanogaster* embryogenesis. Specifically, I set out to uncover the locations Dichaete binds to in the genome; to assess the functionality of the binding using results from gene expression studies; to use computational methods to predict potential cofactors and their binding patterns and to gain an overall understanding of how Dichaete and its cofactors act together to modulate gene expression. I believe that I have in part been successful, and that this thesis presents the first rough roadmap of Dichaete genomics.

7.1 Dichaete genome-wide binding during embryogenesis

The Dichaete in vivo binding profile was investigated using two independent genome-wide approaches: ChIP-chip and DamID. While my own ChIP-chip studies did not produce data with a sufficiently clear signal, several other Dichaete ChIP-chip datasets had since been published and these I analysed as part of my thesis. The DamID data I generated, on the other hand, is high quality and appears to specifically highlight Dichaete activity. Since ChIP-chip and DamID are independent assays, they effectively cross-validate each other, and this was used to create a high confidence binding dataset for the 0-12 hour period of embryonic development. Since these high confidence binding intervals are validated by two methods and come from multiple highly concordant datasets, I believe that this dataset provides a reliable general view of Dichaete action during early embryogenesis. Additionally, the high concordance between the two different types of data is encouraging and means that these two techniques can be used to complement each other and generate more robust binding datasets in general.

While looking at accessibility provides an approximation of where Dichaete may be binding at specific developmental times, it was clear from the comparative analysis that, as would be expected, having a tighter time window for a ChIP-chip experiment has advantages. For example, the Dichaete binding motif was only successfully retrieved from D_Berk ChIP data. While the high significance of overlaps between the different Dichaete datasets suggests that they are performing well in terms of identifying broad regions of Dichaete binding, the fact that a subsection of sequences close to the estimated peak centre from D_Berk were enriched for the Dichaete binding motif, while the others did not appear to, suggests that peaks in ChIP-chip data from a tight developmental window may be more closely indicative of actual Dichaete binding sites, rather than just correctly reflecting the general area of binding. Furthermore, there appears to be signal in the quantitative information from this dataset, which may have functional significance (Biggin, 2011). My analysis suggests that this correlation between quantitative information and function is lost in datasets from broader developmental times, probably because the binding picture obtained is averaged out over both time and multiple tissues.

As well as using narrow time windows, tissue specific profiles would be ideal where possible, but this is currently technically challenging. While tissue specific ChIP where dissection is not possible needs to involve, for example, sorting nuclei by fluorescence, the development of a less active Dam methylase would make it possible to do tissue specific DamID by driving the construct using tissue-specific drivers, making tissue specific binding profiles much easier and more widely accessible to perform.

With or without access to these new technologies, the next logical step for the study of Dichaete in neurogenesis would be to obtain a more specific binding profile, for example at late stage 8-stage 9. A temporal series may also be interesting, to test whether specific enhancers are only bound at particular points during development. My analysis uncovered some interesting candidate targets and suggests that these can potentially be identified from the Dichaete binding data deconvolved by using chromatin accessibility, if only specific enhancers were studied. In particular, some of the CRMs documented in REDfly (Halfon et al., 2007) which overlap with Dichaete data could be used.

Genes associated with multiple Dichaete binding intervals also showed increased enrichment for involvement in a range of developmental processes. This is likely due to the fact that complex CRM architecture makes it possible to integrate a range of signals (Hermsen et al., 2006) and precisely control temporal expression. It would be interesting to see whether the number of binding sites associated with a gene correlates with the changes in gene expression in mutants, and also to try and dissect the precise functions of different components of multiply bound CRMs by creating reporter constructs where individual binding sites are mutated.

7.2 Dichaete mechanisms of action

While the data presented in my thesis is not sufficient to precisely determine the mode of action Dichaete uses for transcriptional regulation, there were two hypotheses in particular that are consistent with the evidence presented in this thesis. First, Dichaete is likely to act by binding to enhancers: it frequently binds to long introns and overlaps with a large proportion of known CRMs as characterised in REDfly (Halfon et al., 2007). This is consistent with the limited previous evidence on Dichaete action, where it is known to regulate the expression of *slit* in the midline together with the POU protein Vvl and the PAS domain protein Single minded by binding to a 1 kb intronic enhancer (Ma et al., 2000). The known CRMs Dichaete binds to could also provide a starting point for further analysis of its action.

A second potential mechanism of action consistent with the available data is that Dichaete facilitates Polymerase pausing. The majority of Dichaete binding peak centres land in genic regions, and there is a high degree of overlap with PolII binding intervals, with some Dichaete intervals showing strong correspondence with PolII peaks, for example in the region of the *mod(mdg4)* insulator protein gene. This is unlikely to be the only mode of action of Dichaete, since it also binds to a large number of enhancers and a sizeable portion of the binding peaks falls outside genic regions. However, it is possible that for a subset of its regulatory targets, this is the mechanism Dichaete uses for modifying their transcription rate. The fact that interrupting Polymerase pausing sometimes results in repression and sometimes in derepression (Zeitlinger et al., 2007a; Muse et al., 2007; Gilchrist et al., 2008) is consistent with the reported context dependent effects Dichaete can have on gene expression (Phochanukul and Russell, 2009).

Dichaete, like other Sox genes, is known to bend DNA at a sharp angle (Werner, Bianchi, Gronenborn and Clore, 1995; Ma et al., 1998), and it is possible that this structural property is in some way integral to its function: for example with multiple Dichaete binding sites bending DNA to create chromatin loops. This is consistent with either of the two modes of action described above: Dichaete could be bending DNA at enhancer regions to bring distant enhancer regions together. It is also possible that the presence of bent DNA in the middle of a genic region slows down the progress of PolII, resulting in pausing. This hypothesis is not possible to test from the currently available data, but techniques such as chromosome conformation capture (3C) (Dekker, 2005) could shed further light on the effect of Dichaete on local chromatin geometry and how this may regulate transcription.

7.3 Dichaete direct targets

For the functional analysis of binding intervals, data from a number of gene expression studies were used. The analysis of *Dichaete* mutant embryos at stage 10-11 revealed a very large number of differentially expressed genes, with the majority of the genes in the genome changing expression. Despite the large number of genes flagged, the genes in question nonetheless appear to be specific to known Dichaete functions, including a number of genes involved in nervous system development, segmentation and the regulation of gene expression, among other processes. The specificity of the list appears to increase when filtered for higher fold-change values, which may reflect that genes exhibiting more pronounced changes in gene expression are more likely to be direct targets.

Several tissue specific gene expression profiles were obtained by using dominant negative constructs combined with appropriate Gal4 drivers. The effect of dominant negatives was studied in the midline (using the *sim*Gal4 driver) and in neuroblasts (using the *pros*Gal4 driver) at developmental stages 8-9. The midline results only identified a small number of differentially expressed genes, notably do not include the known midline target *slit*, so this analysis clearly does not represent an extensive catalogue of Dichaete midline targets. It is thought that this is because the gene expression experiments were performed using whole embryos and since the midline is a relatively small tissue, all but the strongest changes in gene expression are lost by being averaged out over the rest of the embryo. However, the experiment still highlights a number of interesting direct targets: for example, the gene *bancal*, which has a known role in regulating apoptosis, and *Rala*, which is involved in the regulation of the Notch signalling pathway and cell fate commitment of photoreceptors. Clearly a future study using newly developed methods for purifying specific cell populations from the embryo (Salmand et al., 2011) will be of use here. Since the dominant negative constructs used in my study are tagged with GFP, they may be of use in this type of study.

The neuroblast results provide a more extensive list of targets, with a total of 325 genes found to be associated with both Dichaete binding at stage 9 and differentially expressed in the gene expression screen. A number of these were genes involved in neurogenesis, including *SoxN*, *Delta*, *lola*, *mid*, *nkd*, *hb*, *ac* and *hkb*. The gene *ac* in particular has already been shown to change in expression in proneural clusters in both*Dichaete* and *SoxN* mutants (Overton et al., 2002), so the conclusion from my analysis that it is a direct target is consistent with previous evidence. Previous research suggests that Dichaete has a context-dependent effect. *ac* is derepressed in some intermediate column proneural clusters in *Dichaete* mutants, suggesting that Dichaete is acting as a repressor there, but it seems to be activated by Dichaete in the medial column (Overton et al., 2002; Zhao and Skeath, 2002; Zhao, Boekhoff-

Falk, Wilson and Skeath, 2007). *SoxN* mutants display a partial loss of *ac* expression in the medial column, whereas a *Dichaete* and *SoxN* double mutant shows an almost complete loss of *ac* expression. Further, more detailed experiments on binding patterns and effects of SoxN around *ac* may shed some light on the partial redundancy between Dichaete and SoxN.

The fact that *SoxN* was found to be a target makes sense: since Dichaete and SoxN are both present in medial and intermediate column neuroblasts and share some functionality, having a feedback loop between them seems logical. This has indeed already been suggested by a previous study - Overton et al. (2002) report that in about half of the *SoxN* null mutant embryos studied, the levels of *Dichaete* expression were reduced in the anterior half of the neuroectoderm. Other genes identified as likely direct targets in the CNS also show binding and expression effects that suggest it is likely that the network of transcription factors in neuroblasts is highly interconnected. If this is the case, it suggests a mechanism to facilitate precise tuning of different cell fate outcomes. In this respect I note that Dichaete, Hb and Kr all bind to each others regulatory regions, further emphasising the potential for a cross-regulatory network (Figure 7.1). It is also interesting to note that the neuroblast differentiation genes *nkd*, *mid*, *hb* and *hkb* are all associated with overlapping binding intervals for Dichaete, Hb and Kr, and it is likely that these are some of the regulatory targets during the transcription factor temporal sequence in neuroblasts.

7.4 The role of Dichaete in neuroblasts

Dichaete has been previously reported to have a role in maintaining the mitotic activity of neuroblasts in the larval neuroectoderm, with downregulation of Dichaete necessary for cell cycle exit (Maurange et al., 2008). The data from my PhD highlights this role and provides some further insight into the details of its role. In particular, Dichaete appears to interact with the first three transcription factors in the series (Hb, Kr and Pdm) and may be preventing cell cycle exit by directly regulating both apoptosis and differentiation genes. The presence of the Hb binding motif in the output from the Dichaete motif searching I performed strengthens the evidence that the factors are cooperating. In addition, Pdm is a POU protein with previously reported genetic interactions with Dichaete (Ma et al., 1998). The occurrence of Hb, Kr and Pdm binding motifs within Dichaete binding intervals suggests that the factors are frequently located approximately 100 bp away from Dichaete binding. Although it is currently not known whether at this distance they act as direct Dichaete cofactors, they at least appear to be acting at the same general regulatory regions. It is also possible that the motif occurence distances are connected to the structural properties of Dichaete and

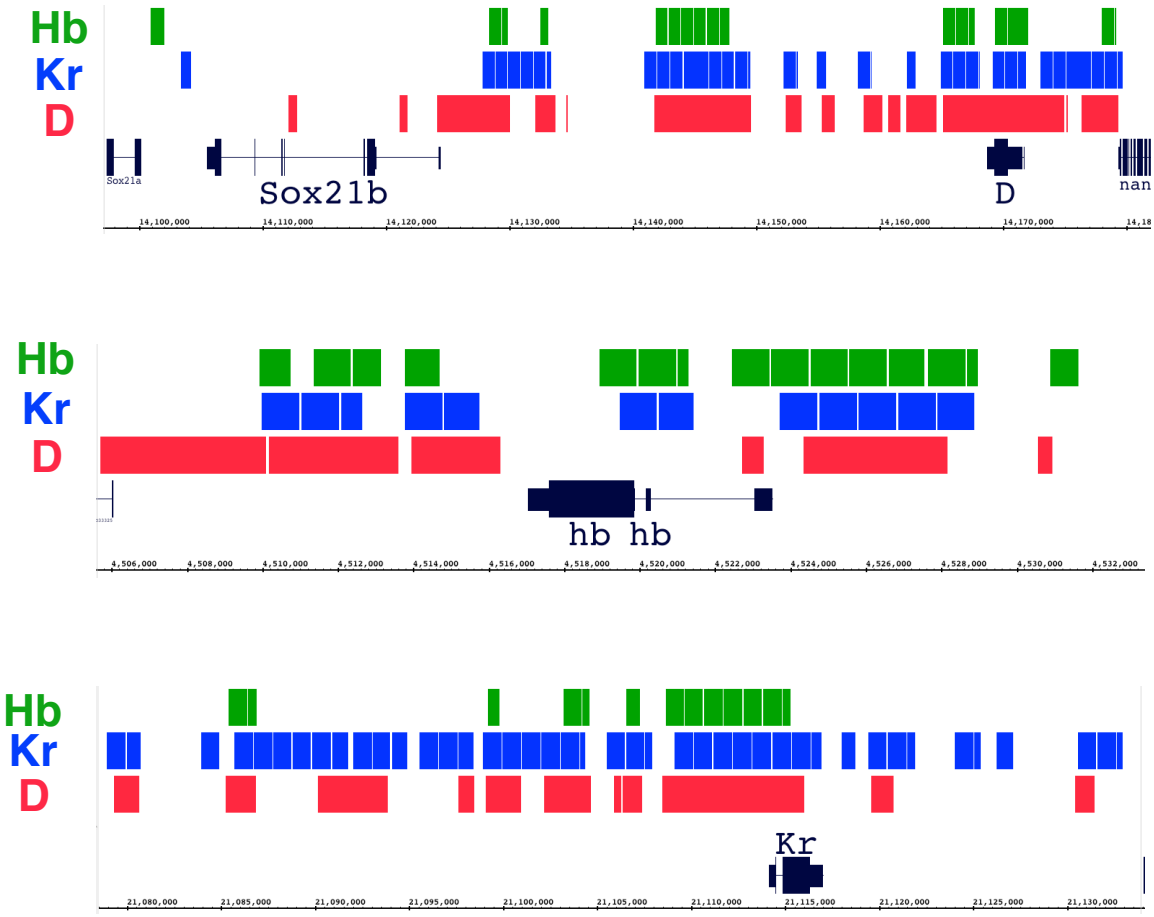

Figure 7.1: The binding in the regulatory regions of *Dichaete*, *hb* and *Kr*. It is notable that all of the transcription factors in question bind multiple times to each other's regulatory regions, suggesting a complex cross-regulatory network.

the angle at which it bends DNA. Establishing the resulting geometry of bound regions will make an interesting further analysis.

Since *hb* has been highlighted in the analysis as a direct target of Dichaete, I performed an analysis of the recently characterised *hb* enhancer HZ4 (Hirono et al., 2011). I found some overlap between Dichaete binding data and the enhancer, in particular in the regions controlling expression in neuroblasts and also in early neurons. Taking into account the chromatin accessibility data for different time periods suggests that Dichaete binds just upstream of (but not overlapping) the neuroblast regulatory region at stages 5 and 9 only, and that it binds the early neuron regulatory region at stage 10 only. However, for this type of detailed analysis the limitations of the binding data I present here becomes apparent and suggest that it is insufficiently detailed. While my analysis provides a reliable overview of Dichaete binding during stages 1-15, it is possible that there are more transient binding events occuring specifically in developing neuroblasts, which this data is not sensitive enough to identify. This will be interesting to investigate further, with particular focus on how the binding patterns change from stage 9 to stage 11, and whether different parts of the enhancer are bound at different times. A more focused study, using ChIP coupled with specific PCR assays, may be helpful in this respect.

7.5 Dichaete and Prospero

Dichaete and Prospero binding profiles were found to significantly overlap, and in particular, joint Dichaete and Prospero binding regions were found to be associated with 27 out of 35 genes from the neuroblast transcriptional network identified by Southall and Brand (2009) (Figure 7.2). This suggests the close involvement of both factors in neuroblast transcriptional control. In particular, it was interesting that the expression results from *Dichaete* and *prospero* mutants show opposite changes for a number of genes involved in neuroblast differentiation, but the same changes for two of the genes involved in neuronal development (Figure 7.2). This opens up the possibility that Dichaete and Prospero act antagonistically when it comes to neuroblast differentiation, but possibly later cooperate to bring about correct neuronal development. This conclusion stands up when we consider that Prospero is a differentiation factor and Dichaete keeps neuroblasts in a pluripotent state. However, the involvement in neuronal differentiation has not been investigated, and while solid conclusions cannot be based on such a small number of genes, this will be an interesting area to further explore. Careful qRT-PCR or *in situ* analysis of relevant target genes in single and double mutants will help clarify the regulatory inputs from Dichaete and Prospero.

My analysis and that of Maurange et al. (2008) suggests that Dichaete activates *mi-*

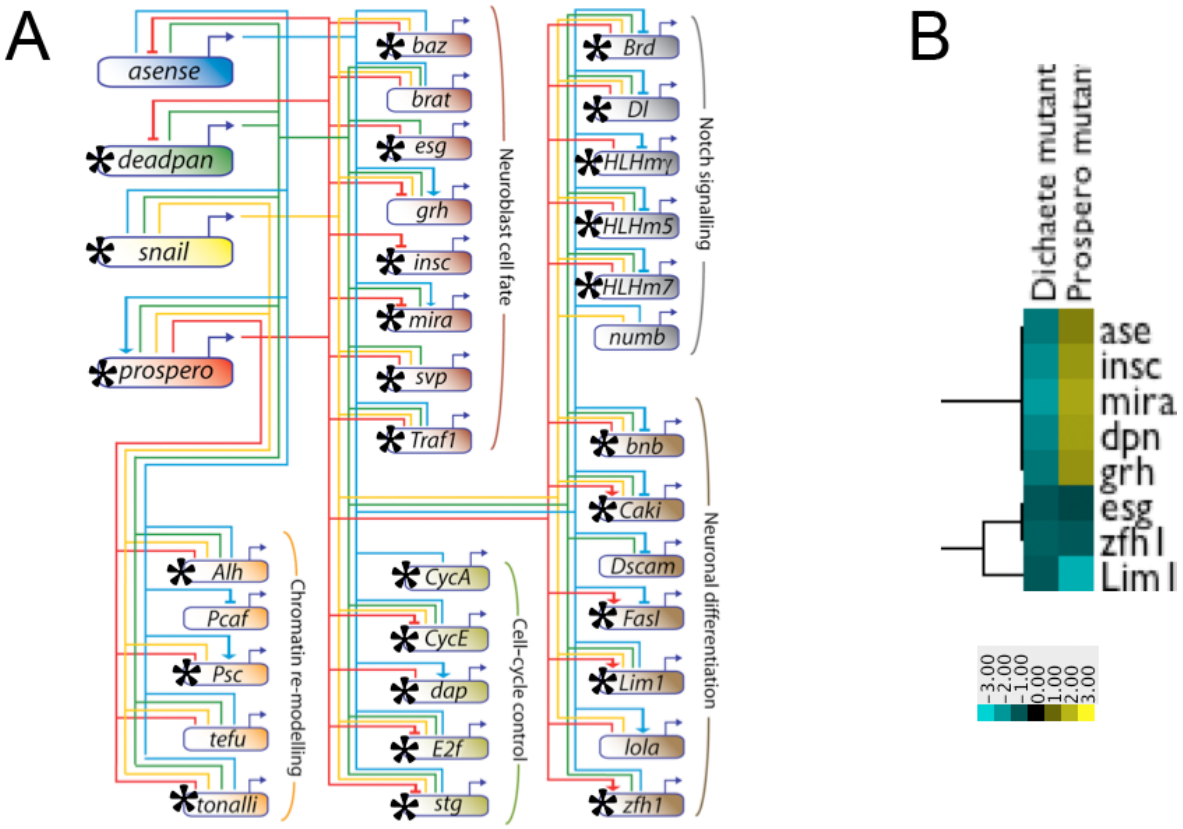

Figure 7.2: (A) The neuroblast transcriptional network, as identified by Southall and Brand (2009), with the genes associated with overlapping Dichaete and Prospero binding labelled with a black asterisk. The majority of genes were found to be associated with overlapping Dichaete and Prospero binding. (B) The clustering of gene expression data is shown for the genes which have been flagged as differentially expressed in both *Dichaete* and *Prospero* mutants. Dichaete and Prospero appear to have the opposite transcriptional effect on the majority of neuroblast cell fate genes for which data is available, but have the same effect on two of the neuronal differentiation genes.

randa in neuroblasts, presumably to keep Prospero out of the nucleus and thus maintain neuroblasts in an undifferentiated state. However, it is unclear what happens from there, including the issue of whether Dichaete and Prospero are ever located in the same nucleus together. It would be interesting to resolve the question of whether Dichaete persists in the nucleus of ganglion mother cell, and subsequently in its daughter cells. If the two colocalise together, it is possible that they cooperate to bring about the later stages of neuron differentiation. If this is the case, it may represent a switch in Dichaete function based on changing co-factors; for example, from the temporal neuroblast TFs to Prospero. An initial analysis of Dichaete and Prospero using immunofluorescence should help clarify the spatio-temporal relationship between these to proteins as a prelude to further analysis.

7.6 Future work

Due to its importance as a developmental transcription factor and high degree of functional conservation with relatively distantly related species such as mouse and chick (Soriano and Russell, 1998; Blanco et al., 2005), Dichaete continues to be an interesting subject of study. Further to that, the relationship between Dichaete and SoxNeuro as yet remains unclear, and may serve as an interesting model for functional redundancy between closely related Sox factors. Here I present some potential directions in which to take the study of Dichaete further.

7.6.1 Further Dichaete *in vivo* binding studies

As discussed, there are a number of ways of improving our knowledge of Dichaete in vivo binding. The most technically challenging would be to perform tissue specific ChIP-chip using FACS, or tissue specific DamID using an improved construct that can be driven in specific tissues. However, even without these methods, the data would be much improved if it was collected from a tighter developmental window and possibly as a time series. My analysis suggests that early developing neuroblasts will be particularly interesting to focus on, and that there is a strongly interconnected network of transcription factors that Dichaete is an integral part of. A different experiment that could also be performed using currently available methods is ChIP-chip and DamID using larval brains, to shed some light on the role of Dichaete in post-embryonic nervous system development. At present such a dataset has not yet been published.

In terms of data analysis, our understanding of DamID data could possibly be improved by studying the patterns of the GATC sequence occurring in the genome. The technique relies on being able to methylate GATC sites in the genome, which are likely

to be non-uniformly distributed, but this has as yet not been taken into account when analysing DamID data. The strategy of using ChIP-chip and DamID data together appears to be a promising one, and is likely to increase the reliability and reduce the number of false positives from any data obtained.

7.6.2 SoxNeuro studies

Some particularly interesting questions regarding functional redundancy include how Dichaete and SoxNeuro find their specific targets when present in the same tissues (such as medial and intermediate column neuroblasts), to what extent they are functionally redundant, whether as suggested they sometimes act antagonistically in a context-dependent manner, and what the mechanism for their differential function is in vivo. There is currently an ongoing PhD project by Enrico Ferrero in the Russell lab, performing ChIP-chip, DamID and gene expression experiments on SoxNeuro similar to the ones I performed on Dichaete, and hopefully with the joint data from our two projects, some of the questions about the precise relationship between Dichaete and SoxNeuro can finally be answered.

7.6.3 Molecular action of Dichaete

Dichaete belongs to the slightly unconventional group of transcription factors which bind to the minor groove of DNA, and in the process of binding it is known to introduce a sharp bend into the DNA. It is possible that the role of Dichaete is a structural one and that the mechanism by which it contributes to gene expression regulation is by forming structures such as chromatin loops. This hypothesis is experimentally testable using chromatin conformation capture (3C) (Dekker, 2005). It will be interesting to apply this method to the *achaete-scute* complex and potentially the regulatory regions surrounding *hunchback*, as the binding patterns there have already been well studied and contains probable *bona fide* Dichaete targets. This data may help shed light on the question of the Dichaete mode of action, by comparing sites of intragenic bending to the presence of corresponding PolII peaks, and also by identifying binding sites through motif searches and studying in more detail how Dichaete-induced bending alters enhancer geometry and binding site distances.

7.6.4 Conservation studies

A number of transcription factors have now been studied in closely related species, including some in mammalian, yeast and Drosophila species (Odom et al., 2007; Borneman et al., 2007; Schmidt et al., 2010; He et al., 2011; Hemberg and Kreiman, 2011).

The results are varied, and suggest that while some transcription factors have binding sites whose locations are strongly conserved by evolution, other ones diverge rapidly between even very closely related species, perhaps providing a tool for evolution to shape and optimise. It is unclear which group Sox genes in general and Dichaete in particular belong to, but an evolutionary study of both Dichaete and SoxNeuro between closely related Drosophila species may prove to be interesting and insightful. Sarah Carl, a PhD student in the Russell lab, is currently beginning this project, which should offer an interesting continuation of Dichaete research.

7.6.5 In-depth studies of Dichaete target expression patterns

The BDGP in situ database provides a major resource for studying the gene expression patterns during embryonic development in *Drosophila*. This resource has not been used very much in this thesis, primarily because Dichaete has been shown to act both as an activator and as a repressor, and therefore the 'expected' patterns of expression of direct Dichaete targets could not be determined with certainty. However, a preliminary analysis using the GEMs tool in Fly Express Kumar et al. (2011) revealed that the predicted Dichaete targets do in fact show consistent patterns of expression in the nervous system. Further to that, now that a large number of potential Dichaete targets have been identified as part of this project, this provides a much more solid starting point for drawing on this data. Future studies could use this to identify and filter for direct targets, as well as doing an in-depth analysis of the expression patterns of known targets during different stages of embryogenesis, to further dissect the potential action of Dichaete at different times and in different tissues.

7.6.6 Dichaete and Sox2 binding conservation

There are now a number of ChIP-chip datasets available for Sox2 in mice, including a fascinating study on the role of Sox2, Sox3 and Sox11 in neuronal lineage development (Bergsland et al., 2011). A comparative analysis of the Sox2 binding in mice and Dichaete binding in *Drosophila* may yield some interesting insights into the conservation of Sox binding between vertebrates and invertebrates. While genome alignments are technically challenging even between mammalian species, and likely close to impossible between mice and fruit flies, the level of homology between the genes associated the binding intervals in each species should be relatively straight-forward to analyse, and may yield some interesting insights into the functional conservation between the two factors. The analysis may also be improved by including gene expression studies from both species to facilitate comparisons between likely direct targets.

7.6.7 The mode of action of Dichaete dominant negative constructs

Dominant negative constructs are a commonly used tool for studying transcription factor function via gene expression studies. However, the mechanisms by which they cause gene expression changes are often unclear, making the data analysis somewhat ambiguous and inconclusive. In Chapter 5, I presented a pilot study of Dichaete dominant negative protein binding in the genome, done in collaboration with Aleksandra Mandic, Louis Chauviere and Bettina Fischer. The initial results were encouraging and this is a promising project that could be taken further. The gene expression data using the constructs is already available, and all the constructs are GFP-tagged, making the project relatively straight-forward from a technical perspective.

7.7 Conclusions

In this thesis I present the first in-depth genome-wide study of Dichaete action. The data I have generated, combined with data from other groups, gives a broad overview of the role of Dichaete in *Drosophila* embryogenesis. While there is already an abundance of available high throughput genomics datasets for *Drosophila*, and their number will carry on increasing in the coming years, what is often lacking is an in-depth systematic analysis of the available data. Here I present such an analysis for Dichaete. The Dichaete core binding dataset, as well as the direct targets identified using binding data, gene expression and computational methods, are carefully constructed and will hopefully form a foundation for future research. Further to that, during the course of my project I believe I have developed several simple but effective analysis and filtering methods, such as looking at clusters of binding to identify direct targets, and using DNaseI sensitivity data to add a dynamic context to the available binding maps. Finally, I identified a list of specific cofactors which link back to the Dichaete role in the nervous system - specifically its interaction with Hb, Kr and Pdm in the neuroblast temporal series. Additionally, I highlighted an interesting connection between Dichaete and Prospero in neuroblasts and GMCs, as well as suggesting a number of potential future paths of enquiry that stem from my analysis. I believe that my work makes a novel contribution to the field, and provides a useful stepping stone in our understanding of Dichaete biology.

Bibliography

Adoutte, A., Balavoine, G., Lartillot, N., Lespinet, O., Prud'Homme, B. and De Rosa, R. (2000). The new animal phylogeny: reliability and implications, *Proceedings of the National Academy of Sciences* **97**(9): 4453.

Adryan, B. and Teichmann, S. (2006). FlyTF: a systematic review of site-specific transcription factors in the fruit fly *Drosophila melanogaster*, *Bioinformatics* **22**(12): 1532.

Adryan, B., Woerfel, G., Birch-Machin, I., Gao, S., Quick, M., Meadows, L., Russell, S. and White, R. (2007). Genomic mapping of Suppressor of Hairy-wing binding sites in *Drosophila*, *Genome Biology* **8**(8): R167.

Akam, M. (1987). The molecular basis for metameric pattern in the *Drosophila* embryo, *Development* **101**(1): 1.

Aleksic, J. and Russell, S. (2009). ChIPing away at the genome: the new frontier travel guide, *Molecular BioSystems* **5**(12): 1421–1428.

Allen, M., Hillier, L., Waterston, R. and Blumenthal, T. (2011). A global analysis of *C. elegans* trans-splicing, *Genome Research* **21**(2): 255.

Altenhein, B., Becker, A., Busold, C., Beckmann, B., Hoheisel, J. and Technau, G. (2006). Expression profiling of glial genes during *Drosophila* embryogenesis, *Developmental Biology* **296**(2): 545–560.

Arendt, D. and Nubler-Jung, K. (1999). Comparison of early nerve cord development in insects and vertebrates, *Development* **126**(11): 2309.

Avilion, A., Nicolis, S., Pevny, L., Perez, L., Vivian, N. and Lovell-Badge, R. (2003). Multipotent cell lineages in early mouse development depend on SOX2 function, *Genes & Development* **17**(1): 126.

Babu, M., Luscombe, N., Aravind, L., Gerstein, M. and Teichmann, S. (2004). Structure and evolution of transcriptional regulatory networks, *Current Opinion in Structural Biology* **14**(3): 283–291.

Baldi, P. and Long, A. (2001). A Bayesian framework for the analysis of microarray expression data: regularized t-test and statistical inferences of gene changes, *Bioinformatics* **17**(6): 509.

Barabasi, A. and Oltvai, Z. (2004). Network biology: understanding the cell's functional organization, *Nature Reviews Genetics* **5**(2): 101–113.

Barski, A., Cuddapah, S., Cui, K., Roh, T., Schones, D., Wang, Z., Wei, G., Chepelev, I. and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome, *Cell* **129**(4): 823–837.

Bauer, S., Grossmann, S., Vingron, M. and Robinson, P. (2008). Ontologizer 2.0 - a multifunctional tool for go term enrichment analysis and data exploration, *Bioinformatics* **24**(14): 1650–1651.

Berezikov, E., Robine, N., Samsonova, A., Westholm, J., Naqvi, A., Hung, J., Okamura, K., Dai, Q., Bortolamiol-Becet, D., Martin, R. et al. (2011). Deep annotation of *Drosophila melanogaster* microRNAs yields insights into their processing, modification, and emergence, *Genome Research* **21**(2): 203.

Bergman, C., Carlson, J. and Celniker, S. (2005). *Drosophila* DNase I footprint database: a systematic genome annotation of transcription factor binding sites in the fruitfly, *Drosophila melanogaster*, *Bioinformatics* **21**(8): 1747.

Bergsland, M., Ramsköld, D., Zaouter, C., Klum, S., Sandberg, R. and Muhr, J. (2011). Sequentially acting Sox transcription factors in neural lineage development, *Genes & Development* **25**: 2453–2464.

Bernstein, B., Humphrey, E., Erlich, R., Schneider, R., Bouman, P., Liu, J., Kouzarides, T. and Schreiber, S. (2002). Methylation of histone H3 Lys 4 in coding regions of active genes, *Proceedings of the National Academy of Sciences of the United States of America* **99**(13): 8695.

Beuchle, D., Struhl, G. and Muller, J. (2001). Polycomb group proteins and heritable silencing of *Drosophila* Hox genes, *Development* **128**(6): 993.

Bhat, K. (1999). Segment polarity genes in neuroblast formation and identity specification during *Drosophila* neurogenesis, *Bioessays* **21**(6): 472–485.

Biggin, M. (2011). Animal transcription networks as highly connected, quantitative continua, *Developmental Cell* **21**(4): 611–626.

Birch-Machin, I., Gao, S., Huen, D., McGirr, R., White, R. and Russell, S. (2005). Genomic analysis of heat-shock factor targets in *Drosophila*, *Genome Biology* **6**(7): R63.

Blanco, J., Girard, F., Kamachi, Y., Kondoh, H. and Gehring, W. (2005). Functional

analysis of the chicken δ1-crystallin enhancer activity in *Drosophila* reveals remarkable evolutionary conservation between chicken and fly, *Development* **132**(8): 1895–1905.

Borneman, A., Gianoulis, T., Zhang, Z., Yu, H., Rozowsky, J., Seringhaus, M., Wang, L., Gerstein, M. and Snyder, M. (2007). Divergence of transcription factor binding sites across related yeast species, *Science* **317**(5839): 815.

Bowles, J., Schepers, G. and Koopman, P. (2000). Phylogeny of the Sox family of developmental transcription factors based on sequence and structural indicators, *Developmental Biology* **227**(2): 239–255.

Boyer, L., Lee, T., Cole, M., Johnstone, S., Levine, S., Zucker, J., Guenther, M., Kumar, R., Murray, H., Jenner, R. et al. (2005). Core transcriptional regulatory circuitry in human embryonic stem cells, *Cell* **122**(6): 947–956.

Boyer, L., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L., Lee, T., Levine, S., Wernig, M., Tajonar, A., Ray, M. et al. (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells, *Nature* **441**(7091): 349–353.

Brand, A. and Perrimon, N. (1993). Targeted gene expression as a means of altering cell fates and generating dominant phenotypes, *Development* **118**(2): 401.

Bridges, C. and Morgan, T. (1923). The third-chromosome group of mutant characters of *Drosophila melanogaster*, *Publs Carnegie Inst* **327**: 1–251.

Brody, T., Stivers, C., Nagle, J. and Odenwald, W. (2002). Identification of novel *Drosophila* neural precursor genes using a differential embryonic head cDNA screen, *Mechanisms of Development* **113**(1): 41–59.

Bryne, J., Valen, E., Tang, M., Marstrand, T., Winther, O., da Piedade, I., Krogh, A., Lenhard, B. and Sandelin, A. (2007). JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update, *Nucleic Acids Research* .

Buck, M. and Lieb, J. (2004). ChIP-chip: considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments, *Genomics* **83**(3): 349–360.

Buescher, M., Hing, F. and Chia, W. (2002). Formation of neuroblasts in the embryonic central nervous system of *Drosophila melanogaster* is controlled by SoxNeuro, *Development* **129**(18): 4193.

Buhler, J., Ideker, T. and Haynor, D. (2000). Dapple: improved techniques for finding

spots on DNA microarrays, *University of Washington CSE Technical Report UWTR* pp. 08–05.

Celniker, S., Dillon, L., Gerstein, M., Gunsalus, K., Henikoff, S., Karpen, G., Kellis, M., Lai, E., Lieb, J., MacAlpine, D. et al. (2009). Unlocking the secrets of the genome, *Nature* **459**(7249): 927–930.

Chalancon, G. and Babu, M. (2010). Structure and evolution of transcriptional regulatory networks.

Chan, E., Quon, G., Chua, G., Babak, T., Trochesset, M., Zirngibl, R., Aubin, J., Ratcliffe, M., Wilde, A., Brudno, M. et al. (2009). Conservation of core gene expression in vertebrate tissues, *J Biol* **8**(3): 33.

Chao, A., Jones, W. and Bejsovec, A. (2007). The HMG-box transcription factor SoxNeuro acts with Tcf to control Wg/Wnt signaling activity, *Development* **134**(5): 989–997.

Charroux, B., Angelats, C., Fasano, L., Kerridge, S. and Vola, C. (1999). The levels of the bancal product, a *Drosophila* homologue of vertebrate hnRNP K protein, affect cell proliferation and apoptosis in imaginal disc cells, *Molecular and Cellular Biology* **19**(11): 7846.

Cherbas, L., Willingham, A., Zhang, D., Yang, L., Zou, Y., Eads, B., Carlson, J., Landolin, J., Kapranov, P., Dumais, J. et al. (2011). The transcriptional diversity of 25 *Drosophila* cell lines, *Genome Research* **21**(2): 301.

Cheung, M., Down, T., Latorre, I. and Ahringer, J. (2011). Systematic bias in high-throughput sequencing data and its correction by BEADS, *Nucleic Acids Research* .

Chi, N. and Epstein, J. (2002). Getting your Pax straight: Pax proteins in development and disease, *Trends in Genetics* **18**(1): 41–47.

Choksi, S., Southall, T., Bossing, T., Edoff, K., de Wit, E., Fischer, B., van Steensel, B., Micklem, G. and Brand, A. (2006). Prospero acts as a binary switch between self-renewal and differentiation in *Drosophila* neural stem cells, *Developmental Cell* **11**(6): 775–789.

Choo, S., White, R. and Russell, S. (2011). Genome-wide analysis of the binding of the Hox protein Ultrabithorax and the Hox cofactor Homothorax in *Drosophila*, *PloS One* **6**(4): e14778.

Chung, W., Agius, P., Westholm, J., Chen, M., Okamura, K., Robine, N., Leslie, C.

and Lai, E. (2011). Computational and experimental identification of mirtrons in *Drosophila melanogaster* and *Caenorhabditis elegans*, *Genome Research* **21**(2): 286.

Clark, A., Eisen, M., Smith, D., Bergman, C., Oliver, B., Markow, T., Kaufman, T., Kellis, M., Gelbart, W., Iyer, V. et al. (2007). Evolution of genes and genomes on the *Drosophila* phylogeny., *Nature* **450**(7167).

Cleary, M. and Doe, C. (2006). Regulation of neuroblast competence: multiple temporal identity factors specify distinct neuronal fates within a single early competence window, *Genes & Development* **20**(4): 429.

Collignon, J., Sockanathan, S., Hacker, A., Cohen-Tannoudji, M., Norris, D., Rastan, S., Stevanovic, M., Goodfellow, P. and Lovell-Badge, R. (1996). A comparison of the properties of Sox-3 with Sry and two related genes, Sox-1 and Sox-2, *Development* **122**(2): 509.

Cox, R., Surette, M. and Elowitz, M. (2007). Programming gene expression with combinatorial promoters, *Molecular Systems Biology* **3**(1).

Cremazy, F., Berta, P. and Girard, F. (2000). Sox Neuro, a new *Drosophila* Sox gene expressed in the developing central nervous system, *Mechanisms of Development* **93**(1): 215–219.

Cremazy, F., Berta, P. and Girard, F. (2001). Genome-wide analysis of Sox genes in *Drosophila melanogaster*, *Mechanisms of Development* **109**(2): 371–375.

Crews, S. (1998). Control of cell lineage-specific development and transcription by bHLH-PAS proteins, *Genes & Development* **12**(5): 607.

Dailey, L. and Basilico, C. (2001). Coevolution of HMG domains and homeodomains and the generation of transcriptional regulation by Sox/POU complexes, *Journal of Cellular Physiology* **186**(3): 315–328.

Davidson, E. (2001). *Genomic regulatory systems*, Academic Press.

de Hoon, M., Imoto, S., Nolan, J. and Miyano, S. (2004). Open source clustering software, *Bioinformatics* **20**(9): 1453–1454.

De Wit, E., Braunschweig, U., Greil, F., Bussemaker, H. and Van Steensel, B. (2008). Global chromatin domain organization of the *Drosophila* genome, *PLoS Genetics* **4**(3): e1000045.

Dekker, J. (2005). The three 'C's of chromosome conformation capture: controls, controls, controls, *Nature Methods* **3**(1): 17–21.

Down, T. and Hubbard, T. (2005). NestedMICA: sensitive inference of over-represented motifs in nucleic acid sequence, *Nucleic Acids Research* **33**(5): 1445.

Driever, W. and Nusslein-Volhard, C. (1988). A gradient of bicoid protein in *Drosophila* embryos, *Cell* **54**(1): 83–93.

Eaton, M., Prinz, J., MacAlpine, H., Tretyakov, G., Kharchenko, P. and MacAlpine, D. (2011). Chromatin signatures of the *Drosophila* replication program, *Genome Research* **21**(2): 164.

Egelhofer, T., Minoda, A., Klugman, S., Lee, K., Kolasinska-Zwierz, P., Alekseyenko, A., Cheung, M., Day, D., Gadel, S., Gorchakov, A. et al. (2010). An assessment of histone-modification antibody quality, *Nature Structural & Molecular Biology* **18**(1): 91–93.

Ercan, S., Lubling, Y., Segal, E. and Lieb, J. (2011). High nucleosome occupancy is encoded at X-linked gene promoters in *C. elegans*, *Genome Research* **21**(2): 237.

Erives, A. and Levine, M. (2004). Coordinate enhancers share common organizational features in the *Drosophila* genome, *Proceedings of the National Academy of Sciences of the United States of America* **101**(11): 3851.

Eun, S., Lea, K., Overstreet, E., Stevens, S., Lee, J. and Fischer, J. (2007). Identification of genes that interact with *Drosophila* liquid facets, *Genetics* **175**(3): 1163.

Farkas, I., Wu, C., Chennubhotla, C., Bahar, I. and Oltvai, Z. (2006). Topological basis of signal integration in the transcriptional-regulatory network of the yeast, *Saccharomyces cerevisiae*, *BMC Bioinformatics* **7**(1): 478.

Feingold, E., Good, P., Guyer, M., Kamholz, S., Liefer, L., Wetterstrand, K., Collins, F., Gingeras, T., Kampa, D., Sekinger, E. et al. (2004). The ENCODE (ENCyclopedia Of DNA Elements) Project, *Science(Washington)* **306**(5696): 636–40.

Filion, G., van Bemmel, J., Braunschweig, U., Talhout, W., Kind, J., Ward, L., Brugman, W., de Castro, I., Kerkhoven, R., Bussemaker, H. et al. (2010). Systematic protein location mapping reveals five principal chromatin types in *Drosophila* cells, *Cell* **143**(2): 212–214.

Fu, A. and Adryan, B. (2009). Scoring overlapping and adjacent signals from genome-wide ChIP and DamID assays, *Molecular BioSystems* **5**(12): 1429–1438.

Fullwood, M., Wei, C., Liu, E. and Ruan, Y. (2009). Next-generation DNA sequencing of paired-end tags (PET) for transcriptome and genome analyses, *Genome Research* **19**(4): 521.

Furlong, E., Andersen, E., Null, B., White, K. and Scott, M. (2001). Patterns of gene expression during *Drosophila* mesoderm development, *Science* **293**(5535): 1629.

Gabilondo, H., Losada-Pérez, M., del Saz, D., Molina, I., León, Y., Canal, I., Torroja,

L. and Benito-Sipos, J. (2011). A targeted genetic screen identifies crucial players in the specification of the *Drosophila* abdominal Capaergic neurons, *Mechanisms of Development* **128**(3-4): 208–221.

Gelbart, W., Crosby, M., Matthews, B., Chillemi, J., Russo Twombly, S., Emmert, D., Bayraktaroglu, L., Smutniak, F., Kossida, S., Ashburner, M. et al. (1999). The FlyBase database of the *Drosophila* genome projects and community literature, *Nucleic Acids Research* **31**(1): 172–175.

Gelbart, W., Crosby, M., Matthews, B., Rindone, W., Chillemi, J., Twombly, S., Emmert, D., Ashburner, M., Drysdale, R., Whitfield, E. et al. (1997). FlyBase: a *Drosophila* database. The FlyBase consortium., *Nucleic Acids Research* **25**(1): 63.

Gerstein, M., Lu, Z., Van Nostrand, E., Cheng, C., Arshinoff, B., Liu, T., Yip, K., Robilotto, R., Rechtsteiner, A., Ikegami, K. et al. (2010). Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE project, *Science* **330**(6012): 1775.

Gilchrist, D., Nechaev, S., Lee, C., Ghosh, S., Collins, J., Li, L., Gilmour, D. and Adelman, K. (2008). NELF-mediated stalling of Pol II can enhance gene expression by blocking promoter-proximal nucleosome assembly, *Genes & Development* **22**(14): 1921.

Gonzalez, A., Chaouiya, C. and Thieffry, D. (2008). Logical modelling of the role of the Hh pathway in the patterning of the *Drosophila* wing disc, *Bioinformatics* **24**(16): i234–i240.

Graveley, B., Brooks, A., Carlson, J., Duff, M., Landolin, J., Yang, L., Artieri, C., van Baren, M., Boley, N., Booth, B. et al. (2010). The developmental transcriptome of *Drosophila melanogaster*, *Nature* **471**: 473–479.

Greenspan, R. (2004). *Fly pushing: the theory and practice of Drosophila genetics*, Cold Spring Harbor Laboratory Pr.

Greil, F., De Wit, E., Bussemaker, H. and Van Steensel, B. (2007). HP1 controls genomic targeting of four novel heterochromatin proteins in *Drosophila*, *The EMBO Journal* **26**(3): 741–751.

Griffiths, A., Gelbart, W., Miller, J. and Lewontin, R. (2002). *Modern Genetic Analysis*, New York: W. H. Freeman.

Grosskortenhaus, R., Pearson, B., Marusich, A. and Doe, C. (2005). Regulation of temporal identity transitions in *Drosophila* neuroblasts, *Developmental Cell* **8**(2): 193–202.

Haddrill, P., Charlesworth, B., Halligan, D. and Andolfatto, P. (2005). Patterns of intron sequence evolution in *Drosophila* are dependent upon length and GC content, *Genome Biology* **6**(8): R67.

Haiminen, N., Mannila, H. and Terzi, E. (2008). Determining significance of pairwise co-occurrences of events in bursty sequences, *BMC Bioinformatics* **9**(1): 336.

Halfon, M., Gallo, S. and Bergman, C. (2007). REDfly 2.0: an integrated database of cis-regulatory modules and transcription factor binding sites in *Drosophila*, *Nucleic Acids Research* **36**: D594–598.

Hare, E., Peterson, B., Iyer, V., Meier, R. and Eisen, M. (2008). Sepsid Even-skipped enhancers are functionally conserved in *Drosophila* despite lack of sequence conservation, *PLoS Genetics* **4**(6): e1000106.

Harley, V., Clarkson, M. and Argentaro, A. (2003). The molecular action and regulation of the testis-determining factors, Sry (sex-determining region on the Y chromosome) and Sox9 [Sry-related high-mobility group (HMG) box 9], *Endocrine Reviews* **24**(4): 466–487.

Harley, V., Lovell-Badge, R. and Goodfellow, P. (1994). Definition of a consensus DNA binding site for Sry., *Nucleic Acids Research* **22**(8): 1500.

Hauck, B., Gehring, W. and Walldorf, U. (1999). Functional analysis of an eye specific enhancer of the eyeless gene in *Drosophila*, *Proceedings of the National Academy of Sciences of the United States of America* **96**(2): 564.

Hays, R., Gibori, G. and Bejsovec, A. (1997). Wingless signaling generates pattern through two distinct mechanisms, *Development* **124**(19): 3727–3736.

He, Q., Bardet, A., Patton, B., Purvis, J., Johnston, J., Paulson, A., Gogol, M., Stark, A. and Zeitlinger, J. (2011). High conservation of transcription factor binding and evidence for combinatorial regulation across six *Drosophila* species, *Nature Genetics* **43**(5): 414–420.

Hemberg, M. and Kreiman, G. (2011). Conservation of transcription factor binding events predicts gene expression across species, *Nucleic Acids Research* **39**(16): 7092–7102.

Hermsen, R., Tans, S. and Ten Wolde, P. (2006). Transcriptional regulation by competing transcription factor modules, *PLoS Computational Biology* **2**(12): e164.

Hiraoka, Y., Ogawa, M., Sakai, Y., Taniguchi, K., Fujii, T., Umezawa, A., Hata, J. and Aiso, S. (1998). Isolation and expression of a human Sry-related cDNA, hSOX201,

Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression **1396**(2): 132–137.

Hirono, K., Margolis, J., Posakony, J. and Doe, C. (2011). Identification of *hunchback* cis-regulatory DNA conferring temporal expression in neuroblasts and neurons, *Gene Expression Patterns* .

Holland, S., Ioannou, D., Haines, S. and Brown, W. (2005). Comparison of Dam tagging and chromatin immunoprecipitation as tools for the identification of the binding sites for *S. pombe* CENP-C, *Chromosome Research* **13**(1): 73–83.

Holohan, E., Kwong, C., Adryan, B., Bartkuhn, M., Herold, M., Renkawitz, R., Russell, S. and White, R. (2007). CTCF genomic binding sites in *Drosophila* and the organisation of the bithorax complex, *PLoS Genetics* **3**(7): e112.

Holt, R. and Jones, S. (2008). The new paradigm of flow cell sequencing, *Genome Research* **18**(6): 839.

Hooper, S., Boué, S., Krause, R., Jensen, L., Mason, C., Ghanim, M., White, K., Furlong, E. and Bork, P. (2007). Identification of tightly regulated groups of genes during *Drosophila melanogaster* embryogenesis, *Molecular Systems Biology* **3**(1).

Hoskins, R., Landolin, J., Brown, J., Sandler, J., Takahashi, H., Lassmann, T., Yu, C., Booth, B., Zhang, D., Wan, K. et al. (2011). Genome-wide analysis of promoter architecture in *Drosophila melanogaster*, *Genome Research* **21**(2): 182.

Huber, W., Von Heydebreck, A., Sültmann, H., Poustka, A. and Vingron, M. (2002). Variance stabilization applied to microarray data calibration and to the quantification of differential expression, *Bioinformatics* **18**(suppl 1): S96–S104.

Hueber, S., Bezdan, D., Henz, S., Blank, M., Wu, H. and Lohmann, I. (2007). Comparative analysis of *Hox* downstream genes in *Drosophila*, *Development* **134**(2): 381.

Huen, D. and Russell, S. (2010). On the use of resampling tests for evaluating statistical significance of binding-site co-occurrence, *BMC Bioinformatics* **11**(1): 359.

Ishii, Y., Weinberg, K., Oda-Ishii, I., Coughlin, L. and Mikawa, T. (2009). Morphogenesis and cytodifferentiation of the avian retinal pigmented epithelium require downregulation of Group B1 Sox genes, *Development* **136**(15): 2579–2589.

Isshiki, T., Pearson, B., Holbrook, S. and Doe, C. (2001). *Drosophila* neuroblasts sequentially express transcription factors which specify the temporal identity of their neuronal progeny, *Cell* **106**(4): 511–521.

Iyer, V., Horak, C., Scafe, C., Botstein, D., Snyder, M. and Brown, P. (2001). Genomic

binding sites of the yeast cell-cycle transcription factors SBF and MBF, *Nature* **409**(6819): 533–538.

Jacob, F. and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins, *Journal of Molecular Biology* **3**(3): 318–356.

John, A., Smith, S. and Jaynes, J. (1995). Inserting the Ftz homeodomain into *engrailed* creates a dominant transcriptional repressor that specifically turns off Ftz target genes in vivo, *Development* **121**(6): 1801.

Johnson, D., Li, W., Gordon, D., Bhattacharjee, A., Curry, B., Ghosh, J., Brizuela, L., Carroll, J., Brown, M., Flicek, P. et al. (2008). Systematic evaluation of variability in ChIP-chip experiments using predefined DNA targets, *Genome Research* **18**(3): 393.

Johnson, D., Mortazavi, A., Myers, R. and Wold, B. (2007). Genome-wide mapping of in vivo protein-DNA interactions, *Science* **316**(5830): 1497.

Jothi, R., Balaji, S., Wuster, A., Grochow, J., Gsponer, J., Przytycka, T., Aravind, L. and Babu, M. (2009). Genomic analysis reveals a tight link between transcription factor dynamics and regulatory network architecture, *Molecular Systems Biology* **5**(1).

Kalodimos, C., Biris, N., Bonvin, A., Levandoski, M., Guennuegues, M., Boelens, R. and Kaptein, R. (2004). Structure and flexibility adaptation in nonspecific and specific protein-DNA complexes, *Science* **305**(5682): 386.

Kamachi, Y., Sockanathan, S., Liu, Q., Breitman, M., Lovell-Badge, R. and Kondoh, H. (1995). Involvement of Sox proteins in lens-specific activation of crystallin genes., *The EMBO Journal* **14**(14): 3510.

Kamachi, Y., Uchikawa, M. and Kondoh, H. (2000). Pairing Sox off: with partners in the regulation of embryonic development, *Trends in Genetics* **16**(4): 182–187.

Kamachi, Y., Uchikawa, M., Tanouchi, A., Sekido, R. and Kondoh, H. (2001). Pax6 and Sox2 form a co-DNA-binding partner complex that regulates initiation of lens development, *Genes & development* **15**(10): 1272.

Kankel, M., Hurlbut, G., Upadhyay, G., Yajnik, V., Yedvobnick, B. and Artavanis-Tsakonas, S. (2007). Investigating the genetic circuitry of *mastermind* in *Drosophila*, a *notch* signal effector, *Genetics* **177**(4): 2493.

Kaplan, T., Li, X., Sabo, P., Thomas, S., Stamatoyannopoulos, J., Biggin, M. and Eisen, M. (2011). Quantitative models of the mechanisms that control genome-wide patterns of transcription factor binding during early *Drosophila* development, *PLoS Genetics* **7**(2): e1001290.

Kearney, J., Wheeler, S., Estes, P., Parente, B. and Crews, S. (2004). Gene expression profiling of the developing *Drosophila* CNS midline cells, *Developmental Biology* **275**(2): 473–492.

Kharchenko, P., Alekseyenko, A., Schwartz, Y., Minoda, A., Riddle, N., Ernst, J., Sabo, P., Larschan, E., Gorchakov, A., Gu, T. et al. (2010). Comprehensive analysis of the chromatin landscape in *Drosophila melanogaster*, *Nature* .

Kishi, M., Mizuseki, K., Sasai, N., Yamazaki, H., Shiota, K., Nakanishi, S. and Sasai, Y. (2000). Requirement of Sox2-mediated signaling for differentiation of early *Xenopus* neuroectoderm, *Development* **127**(4): 791.

Kohler, J., Schafer-Preuss, S. and Buttgereit, D. (1996). Related enhancers in the intron of the beta 1 tubulin gene of *Drosophila melanogaster* are essential for maternal and CNS-specific expression during embryogenesis, *Nucleic Acids Research* **24**(13): 2543–2550.

Kondoh, H. and Kamachi, Y. (2010). SOX-partner code for cell specification: regulatory target selection and underlying molecular mechanisms, *The International Journal of Biochemistry & Cell Biology* **42**(3): 391–399.

Konikoff, C., Karr, T., McCutchan, M., Newfeld, S. and Kumar, S. (2011). Comparison of embryonic expression within multigene families using the FlyExpress discovery platform reveals more spatial than temporal divergence, *Developmental Dynamics* .

Konikoff, C., McCutchan, M., Van Emden, B., Busick, C., Davis, K., Ji, S., Wu, L., Ramos, H., Brody, T., Panchanathan, S. et al. (2009). FlyExpress: Visual mining of spatiotemporal patterns for genes and publications in *Drosophila* embryogenesis., *Tempe, AZ* **85287**.

Koopman, P., Schepers, G., Brenner, S. and Venkatesh, B. (2004). Origin and diversity of the Sox transcription factor gene family: genome-wide analysis in *Fugu rubripes*, *Gene* **328**: 177–186.

Krejci, A., Bernard, F., Housden, B., Collins, S. and Bray, S. (2009). Direct response to Notch activation: Signaling crosstalk and incoherent logic, *Science Signaling* **2**(55): ra1.

Kumar, S., Konikoff, C., Van Emden, B., Busick, C., Davis, K., Ji, S., Wu, L., Ramos, H., Brody, T., Panchanathan, S. et al. (2011). FlyExpress: Visual mining of spatiotemporal patterns for genes and publications in *Drosophila* embryogenesis, *Bioinformatics* .

Kwong, C., Adryan, B., Bell, I., Meadows, L., Russell, S., Manak, J. and White, R.

(2008). Stability and dynamics of polycomb target sites in *Drosophila* development, *PLoS Genetics* **4**(9).

Larroux, C., Luke, G., Koopman, P., Rokhsar, D., Shimeld, S. and Degnan, B. (2008). Genesis and expansion of metazoan transcription factor gene classes, *Molecular Biology and Evolution* **25**(5): 980–996.

Lee, T., Jenner, R., Boyer, L., Guenther, M., Levine, S., Kumar, R., Chevalier, B., Johnstone, S., Cole, M., Isono, K. et al. (2006). Control of developmental regulators by Polycomb in human embryonic stem cells, *Cell* **125**(2): 301–313.

Lewis, M., Chang, G., Horton, N., Kercher, M., Pace, H., Schumacher, M., Brennan, R. and Lu, P. (1996). Crystal structure of the lactose operon repressor and its complexes with DNA and inducer, *Science* **271**(5253): 1247.

Li, X., MacArthur, S., Bourgon, R., Nix, D., Pollard, D., Iyer, V., Hechmer, A., Simirenko, L., Stapleton, M., Hendriks, C., Chu, H., Ogawa, N., Inwood, W., Sementchenko, V., Beaton, A., Weiszmann, R., Celniker, S., Knowles, D., Gingeras, T., Speed, T., Eisen, M. and Biggin, M. (2008). Transcription factors bind thousands of active and inactive regions in the *Drosophila* blastoderm, *PLoS Biology* **6**(2): e27.

Li, X., Thomas, S., Sabo, P., Eisen, M., Stamatoyannopoulos, J. and Biggin, M. (2011). The role of chromatin accessibility in directing the widespread, overlapping patterns of *Drosophila* transcription factor binding, *Genome Biology* **12**(4): R34.

Lieb, J., Liu, X., Botstein, D. and Brown, P. (2001). Promoter-specific binding of Rap1 revealed by genome-wide maps of protein-DNA association, *Nature Genetics* **28**(4): 327–334.

Ligr, M., Siddharthan, R., Cross, F. and Siggia, E. (2006). Gene expression from random libraries of yeast promoters, *Genetics* **172**(4): 2113.

Lin, S. and Riggs, A. (1975). The general affinity of lac repressor for *E. coli* DNA: implications for gene regulation in procaryotes and eucaryotes, *Cell* **4**(2): 107–111.

Liu, T., Rechtsteiner, A., Egelhofer, T., Vielle, A., Latorre, I., Cheung, M., Ercan, S., Ikegami, K., Jensen, M., Kolasinska-Zwierz, P. et al. (2011). Broad chromosomal domains of histone modification patterns in *C. elegans*, *Genome Research* **21**(2): 227.

Liu, X., Huang, J., Chen, T., Wang, Y., Xin, S., Li, J., Pei, G. and Kang, J. (2008). Yamanaka factors critically regulate the developmental signaling network in mouse embryonic stem cells, *Cell Research* **18**(12): 1177–1189.

Lockhart, D., Dong, H., Byrne, M., Follettie, M., Gallo, M., Chee, M., Mittmann, M., Wang, C., Kobayashi, M., Norton, H. et al. (1996). Expression monitor-

ing by hybridization to high-density oligonucleotide arrays, *Nature Biotechnology* **14**(13): 1675–1680.

Loh, Y., Wu, Q., Chew, J., Vega, V., Zhang, W., Chen, X., Bourque, G., George, J., Leong, B., Liu, J. et al. (2006). The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells, *Nature Genetics* **38**(4): 431–440.

Ludwig, M., Bergman, C., Patel, N. and Kreitman, M. (2000). Evidence for stabilizing selection in a eukaryotic enhancer element, *Nature* **403**(6769): 564–567.

Lund, A. and van Lohuizen, M. (2004). Polycomb complexes and silencing mechanisms, *Current Opinion in Cell Biology* **16**(3): 239–246.

Lyko, F. (2001). DNA methylation learns to fly, *Trends in Genetics* **17**(4): 169–172.

Lyne, R., Smith, R., Rutherford, K., Wakeling, M., Varley, A., Guillier, F., Janssens, H., Ji, W., Mclaren, P., North, P. et al. (2007). FlyMine: an integrated database for *Drosophila* and *Anopheles* genomics, *Genome Biology* **8**(7): R129.

Ma, H., Buer, J. and Zeng, A. (2004). Hierarchical structure and modules in the *Escherichia coli* transcriptional regulatory network revealed by a new top-down approach, *BMC Bioinformatics* **5**(1): 199.

Ma, Y., Certel, K., Gao, Y., Niemitz, E., Mosher, J., Mukherjee, A., Mutsuddi, M., Huseinovic, N., Crews, S., Johnson, W. et al. (2000). Functional interactions between *Drosophila* bHLH/PAS, Sox, and POU transcription factors regulate CNS midline expression of the *slit* gene, *Journal of Neuroscience* **20**(12): 4596.

Ma, Y., Niemitz, E., Nambu, P., Shan, X., Sackerson, C., Fujioka, M., Goto, T. and Nambu, J. (1998). Gene regulatory functions of *Drosophila* Fish-hook, a high mobility group domain Sox protein, *Mechanisms of Development* **73**(2): 169–182.

MacArthur, S., Li, X., Li, J., Brown, J., Chu, H., Zeng, L., Grondona, B., Hechmer, A., Simirenko, L., Keränen, S., Knowles, D., Stapleton, M., Bickel, P., Biggin, M. and Eisen, M. (2009). Developmental roles of 21 *Drosophila* transcription factors are determined by quantitative differences in binding to an overlapping set of thousands of genomic regions, *Genome Biology* **10**(7).

Mace, D., Varnado, N., Zhang, W., Frise, E. and Ohler, U. (2010). Extraction and comparison of gene expression patterns from 2D RNA in situ hybridization images, *Bioinformatics* **26**(6): 761.

Mahony, S. and Benos, P. (2007). STAMP: a web tool for exploring DNA-binding motif similarities, *Nucleic Acids Research* **35**(suppl 2): W253.

Mann, R. and Carroll, S. (2002). Molecular mechanisms of selector gene function and evolution, *Current Opinion in Genetics & Development* **12**(5): 592–600.

Margolin, J., Friedman, J., Wolfram, K., Vissing, H., Thiesen, H. and Rauscher III, F. (1994). Kruppel-associated boxes are potent transcriptional repression domains, *Proceedings of the National Academy of Sciences of the United States of America* pp. 4509–4513.

Margolis, J. (1992). *Regulation of the Drosophila gap segmentation gene hunchback*, PhD thesis, University of California, San Diego.

Margolis, J., Borowsky, M., Shim, C. and Posakony, J. (1994). A small region surrounding the distal promoter of the hunchback gene directs maternal expression, *Developmental Biology* **163**(2): 381–388.

Margolis, J., Borowsky, M., Steingrímsson, E., Shim, C., Lengyel, J. and Posakony, J. (1995). Posterior stripe expression of hunchback is driven from two promoters by a common enhancer element, *Development* **121**(9): 3067–3077.

Markstein, M., Zinzen, R., Markstein, P., Yee, K., Erives, A., Stathopoulos, A. and Levine, M. (2004). A regulatory code for neurogenic gene expression in the *Drosophila* embryo, *Development* **131**(10): 2387.

Martinez-Antonio, A., Janga, S. and Thieffry, D. (2008). Functional organisation of *Escherichia coli* transcriptional regulatory network, *Journal of Molecular Biology* **381**(1): 238–247.

Masui, S., Nakatake, Y., Toyooka, Y., Shimosato, D., Yagi, R., Takahashi, K., Okochi, H., Okuda, A., Matoba, R., Sharov, A. et al. (2007). Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells, *Nature Cell Biology* **9**(6): 625–635.

Maurange, C., Cheng, L. and Gould, A. (2008). Temporal transcription factors and their targets schedule the end of neural proliferation in *Drosophila*, *Cell* **133**(5): 891–902.

Maurange, C. and Gould, A. (2005). Brainy but not too brainy: starting and stopping neuroblast divisions in *Drosophila*, *Trends in Neurosciences* **28**(1): 30–36.

McKimmie, C., Woerfel, G. and Russell, S. (2005). Conserved genomic organisation of Group B Sox genes in insects., *BMC Genetics* **6**(1): 26.

McLean, C., Bristor, D., Hiller, M., Clarke, S., Schaar, B., Lowe, C., Wenger, A. and Bejerano, G. (2010). Great improves functional interpretation of cis-regulatory regions, *Nature biotechnology* **28**(5): 495–501.

Meadows, L., Chan, Y., Roote, J. and Russell, S. (2010). Neighbourhood continuity is not required for correct testis gene expression in *Drosophila*, *PLoS Biology* **8**(11): e1000552.

Meredith, J. and Storti, R. (1993). Developmental regulation of the *Drosophila tropomyosin II* gene in different muscles is controlled by muscle-type-specific intron enhancer elements and distal and proximal promoter control elements, *Developmental Biology* **159**(2): 500–512.

Mi, H., Lazareva-Ulitsky, B., Loo, R., Kejariwal, A., Vandergriff, J., Rabkin, S., Guo, N., Muruganujan, A., Doremieux, O., Campbell, M. et al. (2005). The PANTHER database of protein families, subfamilies, functions and pathways, *Nucleic Acids Research* **33**(suppl 1): D284.

Mikkelsen, T., Ku, M., Jaffe, D., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T., Koche, R. et al. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells, *Nature* **448**(7153): 553–560.

Montalta-He, H., Leemans, R., Loop, T., Strahm, M., Certa, U., Primig, M., Acampora, D., Simeone, A. and Reichert, H. (2002). Evolutionary conservation of otd/Otx2 transcription factor action: a genome-wide microarray analysis in *Drosophila*, *Genome Biology* **3**(4): 1–0015.

Moorman, C., Sun, L., Wang, J., De Wit, E., Talhout, W., Ward, L., Greil, F., Lu, X., White, K., Bussemaker, H. et al. (2006). Hotspots of transcription factor colocalization in the genome of *Drosophila melanogaster*, *Proceedings of the National Academy of Sciences of the United States of America* **103**(32): 12027.

Mukherjee, A., Melnattur, K., Zhang, M. and Nambu, J. (2006). Maternal expression and function of the *Drosophila* Sox gene Dichaete during oogenesis, *Developmental Dynamics* **235**(10): 2828–2835.

Mukherjee, A., Shan, X., Mutsuddi, M., Ma, Y. and Nambu, J. (2000). The *Drosophila Sox* gene, *fish-hook*, is required for postembryonic development, *Developmental Biology* **217**(1): 91–106.

Murphy, E., Zhurkin, V., Louis, J., Cornilescu, G. and Clore, G. (2001). Structural basis for Sry-dependent 46-X, Y sex reversal: modulation of DNA bending by a naturally occurring point mutation, *Journal of Molecular Biology* **312**(3): 481–499.

Muse, G., Gilchrist, D., Nechaev, S., Shah, R., Parker, J., Grissom, S., Zeitlinger, J. and Adelman, K. (2007). RNA polymerase is poised for activation across the genome, *Nature Genetics* **39**(12): 1507–1511.

Mutsuddi, M., Mukherjee, A., Shen, B., Manley, J. and Nambu, J. (2010). *Drosophila*

Pelle phosphorylates Dichaete protein and influences its subcellular distribution in developing oocytes, *Int. J. Dev. Biol* **54**: 1309–1315.

Myers, R., Stamatoyannopoulos, J., Snyder, M., Dunham, I., Hardison, R., Bernstein, B., Gingeras, T., Kent, W., Birney, E., Wold, B. et al. (2011). A user's guide to the encyclopedia of DNA elements (encode), *PLoS Biology* **9**(4): e1001046.

Nambu, P. and Nambu, J. (1996). The *Drosophila fish-hook* gene encodes a HMG domain protein essential for segmentation and CNS development., *Development (Cambridge, England)* **122**(11): 3467.

Negre, N., Brown, C., Lijia, M., Bristow, C., Miller, S., Wagner, U., Kheradpour, P., Eaton, M., Loriaux, P., Sealfon, R. et al. (2011). A cis-regulatory map of the *Drosophila* genome, *Nature* **471**(7339): 527–531.

Negre, N., Brown, C., Shah, P., Kheradpour, P., Morrison, C. et al. (2010). A comprehensive map of insulator elements for the *Drosophila* genome, *PLoS Genetics* **6**(1).

Negre, N., Hennetin, J., Sun, L., Lavrov, S., Bellis, M. et al. (2006). Chromosomal distribution of PcG proteins during *Drosophila* development, *PLoS Biology* **4**(6): e170.

Nelson, C., Hersh, B., Carroll, S. et al. (2004). The regulatory content of intergenic dna shapes genome architecture, *Genome Biology* **5**(4).

Ng, H., Robert, F., Young, R. and Struhl, K. (2002). Genome-wide location and regulated recruitment of the RSC nucleosome-remodeling complex, *Genes & Development* **16**(7): 806.

Nicol, J., Helt, G., Blanchard Jr, S., Raja, A. and Loraine, A. (2009). The Integrated Genome Browser: free software for distribution and exploration of genome-scale datasets, *Bioinformatics* **25**(20): 2730–2731.

Nishiguchi, S., Wood, H., Kondoh, H., Lovell-Badge, R. and Episkopou, V. (1998). Sox1 directly regulates the γ-*crystallin* genes and is essential for lens development in mice, *Genes & Development* **12**(6): 776.

Niu, W., Lu, Z., Zhong, M., Sarov, M., Murray, J., Brdlik, C., Janette, J., Chen, C., Alves, P., Preston, E. et al. (2011). Diverse transcription factor binding features revealed by genome-wide ChIP-seq in *C. elegans*, *Genome Research* **21**(2): 245.

Noordermeer, J., Johnston, P., Rijsewijk, F., Nusse, R. and Lawrence, P. (1992). The consequences of ubiquitous expression of the Wingless gene in the *Drosophila* embryo, *Development* **116**(3): 711–719.

Noyes, M., Meng, X., Wakabayashi, A., Sinha, S., Brodsky, M. and Wolfe, S. (2008).

A systematic characterization of factors that regulate drosophila segmentation via a bacterial one-hybrid system, *Nucleic acids research* **36**(8): 2547–2560.

Nusslein-Volhard, C. and Wieschaus, E. (1980). Mutations affecting segment number and polarity in *Drosophila*, *Nature* **287**(5785).

Odom, D., Dowell, R., Jacobsen, E., Gordon, W., Danford, T., MacIsaac, K., Rolfe, P., Conboy, C., Gifford, D. and Fraenkel, E. (2007). Tissue-specific transcriptional regulation has diverged significantly between human and mouse, *Nature Genetics* **39**(6): 730–732.

Oliveros, J. (2007). VENNY. an interactive tool for comparing lists with venn diagrams.

Osaki, E., Nishina, Y., Inazawa, J., Copeland, N., Gilbert, D., Jenkins, N., Ohsugi, M., Tezuka, T., Yoshida, M. and Semba, K. (1999). Identification of a novel Sry-related gene and its germ cell-specific expression, *Nucleic Acids Research* **27**(12): 2503–2510.

Overton, P. (2003). *The role of Sox genes in the development of Drosophila melanogaster*, PhD thesis, University of Cambridge, Cambridge, UK.

Overton, P., Chia, W. and Buescher, M. (2007). The *Drosophila* HMG-domain proteins SoxNeuro and Dichaete direct trichome formation via the activation of Shavenbaby and the restriction of Wingless pathway activity, *Development* **134**(15): 2807–2813.

Overton, P., Meadows, L., Urban, J. and Russell, S. (2002). Evidence for differential and redundant function of the Sox genes Dichaete and SoxN during CNS development in *Drosophila*, *Development* **129**(18): 4219.

Papatsenko, I., Levine, M. and Papatsenko, D. (2010). Temporal waves of coherent gene expression during *Drosophila* embryogenesis, *Bioinformatics* **26**(21): 2731.

Pavlopoulos, A. and Akam, M. (2011). *Hox* gene *Ultrabithorax* regulates distinct sets of target genes at successive stages of *Drosophila* haltere morphogenesis, *Proceedings of the National Academy of Sciences of the United States of America* **108**(7): 2855.

Pevny, L. and Placzek, M. (2005). *Sox* genes and neural progenitor identity, *Current Opinion in Neurobiology* **15**(1): 7–13.

Phelps, C. and Brand, A. (1998). Ectopic gene expression in *Drosophila* using GAL4 system, *Methods* **14**(4): 367–379.

Phochanukul, N. and Russell, S. (2009). No backbone but lots of Sox: Invertebrate Sox genes, *The International Journal of Biochemistry & Cell Biology* **42**(3): 453–464.

Piano, F., Parisi, M., Karess, R. and Kambysellis, M. (1999). Evidence for redundancy but not trans factor–cis element coevolution in the regulation of drosophila yp genes, *Genetics* **152**(2): 605–616.

Piccolo, S., Sasai, Y., Lu, B. and De Robertis, E. (1996). Dorsoventral patterning in *Xenopus*: inhibition of ventral signals by direct binding of Chordin to BMP-4, *Cell* **86**(4): 589–598.

Pindyurin, A., Moorman, C., de Wit, E., Belyakin, S., Belyaeva, E., Christophides, G., Kafatos, F., van Steensel, B. and Zhimulev, I. (2007). SUUR joins separate subsets of PcG, HP1 and B-type lamin targets in *Drosophila*, *Journal of Cell Science* **120**(14): 2344.

Pomraning, K., Smith, K. and Freitag, M. (2009). Genome-wide high throughput analysis of DNA methylation in eukaryotes, *Methods* **47**(3): 142–150.

Pradervand, S., Weber, J., Thomas, J., Bueno, M., Wirapati, P., Lefort, K., Dotto, G. and Harshman, K. (2009). Impact of normalization on miRNA microarray expression profiling, *Rna* **15**(3): 493.

Ptashne, M. (2005). Regulation of transcription: from lambda to eukaryotes., *Trends in Biochemical Sciences* **30**(6): 275.

Putnam, N., Srivastava, M., Hellsten, U., Dirks, B., Chapman, J., Salamov, A., Terry, A., Shapiro, H., Lindquist, E., Kapitonov, V. et al. (2007). Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization, *Science* **317**(5834): 86.

Pym, E., Southall, T., Mee, C., Brand, A. and Baines, R. (2006). The homeobox transcription factor even-skipped regulates acquisition of electrical properties in *Drosophila* neurons, *Neural Development* **1**(1): 1–16.

Remenyi, A., Lins, K., Nissen, L., Reinbold, R., Schöler, H. and Wilmanns, M. (2003). Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers, *Genes & Development* **17**(16): 2048–2059.

Ren, B., Robert, F., Wyrick, J., Aparicio, O., Jennings, E., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E. et al. (2000). Genome-wide location and function of DNA binding proteins, *Science* **290**(5500): 2306.

Rex, M., Orme, A., Uwanogho, D., Tointon, K., Wigmore, P., Sharpe, P. and Scotting, P. (1997). Dynamic expression of chicken Sox2 and Sox3 genes in ectoderm induced to form neural tissue, *Developmental Dynamics* **209**(3): 323–332.

Rhead, B., Karolchik, D., Kuhn, R., Hinrichs, A., Zweig, A., Fujita, P., Diekhans, M., Smith, K., Rosenbloom, K., Raney, B. et al. (2010). The UCSC genome browser database: update 2010, *Nucleic Acids Research* **38**(suppl 1): D613.

Riaz, F. (2009). *The application of mutant DNA adenine methyltransferase enzymes in DAM identification.*, PhD thesis, University of Cambridge, Cambridge, UK.

Riddle, N., Minoda, A., Kharchenko, P., Alekseyenko, A., Schwartz, Y., Tolstorukov, M., Gorchakov, A., Jaffe, J., Kennedy, C., Linder-Basso, D. et al. (2011). Plasticity in patterns of histone modifications and chromosomal proteins in *Drosophila* heterochromatin, *Genome Research* **21**(2): 147.

Robertson, G., Hirst, M., Bainbridge, M., Bilenky, M., Zhao, Y., Zeng, T., Euskirchen, G., Bernier, B., Varhol, R., Delaney, A. et al. (2007). Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing, *Nature Methods* **4**(8): 651–657.

Rohs, R., West, S., Sosinsky, A., Liu, P., Mann, R. and Honig, B. (2009). The role of DNA shape in protein–DNA recognition, *Nature* **461**(7268): 1248–1253.

Roy, S., Ernst, J., Kharchenko, P., Kheradpour, P., Negre, N., Eaton, M., Landolin, J., Bristow, C., Ma, L., Lin, M. et al. (2010). Identification of functional elements and regulatory circuits by *Drosophila* modENCODE, *Science* **330**(6012): 1787.

Russell, S. (2000). The *Drosophila* dominant wing mutation Dichaete results from ectopic expression of a Sox-domain gene, *Molecular and General Genetics MGG* **263**(4): 690–701.

Russell, S., Meadows, L. and Russell, R. (2009). *Microarray technology in practice*, Academic Press, San Diego, USA.

Russell, S., Sanchez-Soriano, N., Wright, C. and Ashburner, M. (1996). The Dichaete gene of *Drosophila melanogaster* encodes a Sox-domain protein required for embryonic segmentation, *Development* **122**(11): 3669.

Saldanha, A. (2004). Java Treeview - extensible visualization of microarray data, *Bioinformatics* **20**(17): 3246–3248.

Salmand, P., Iché, Torres, M. and Perrin, L. (2011). Tissue-specific cell sorting from *Drosophila* embryos: Application to gene expression analysis., *Fly* **5**(3).

Sanchez-Soriano, N. and Russell, S. (2000). Regulatory mutations of the *Drosophila* Sox gene *Dichaete* reveal new functions in embryonic brain and hindgut development, *Developmental Biology* **220**(2): 307–321.

Sandberg, M., Källström, M. and Muhr, J. (2005). Sox21 promotes the progression of vertebrate neurogenesis, *Nature Neuroscience* **8**(8): 995–1001.

Sandmann, T., Jensen, L., Jakobsen, J., Karzynski, M., Eichenlaub, M., Bork, P. and Furlong, E. (2006). A temporal map of transcription factor activity: mef2

directly regulates target genes at all stages of muscle development, *Developmental cell* **10**(6): 797–807.

Sasai, Y. et al. (2001). Roles of Sox factors in neural determination: conserved signaling in evolution?, *International Journal of Developmental Biology* **45**(1; SPI): 321–326.

Sawamoto, K., Winge, P., Koyama, S., Hirota, Y., Yamada, C., Miyao, S., Yoshikawa, S., Jin, M., Kikuchi, A. and Okano, H. (1999). The *Drosophila* Ral GTPase regulates developmental cell shape changes through the Jun NH2-terminal kinase pathway, *The Journal of Cell Biology* **146**(2): 361.

Schena, M., Shalon, D., Davis, R. and Brown, P. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray, *Science* **270**(5235): 467.

Schepers, G., Teasdale, R. and Koopman, P. (2002). Twenty pairs of Sox: extent, homology, and nomenclature of the mouse and human Sox transcription factor gene families., *Developmental Cell* **3**(2): 167.

Schmidt, D., Wilson, M., Ballester, B., Schwalie, P., Brown, G., Marshall, A., Kutter, C., Watt, S., Martinez-Jimenez, C., Mackay, S. et al. (2010). Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding, *Science* **328**(5981): 1036.

Schmiedeberg, L., Skene, P., Deaton, A. and Bird, A. (2009). A temporal threshold for formaldehyde crosslinking and fixation, *PloS One* **4**(2): e4636.

Schwartz, Y., Kahn, T., Nix, D., Li, X., Bourgon, R., Biggin, M. and Pirrotta, V. (2006). Genome-wide analysis of polycomb targets in *Drosophila melanogaster*, *Nature Genetics* **38**(6): 700–705.

Shannon, P., Markiel, A., Ozier, O., Baliga, N., Wang, J., Ramage, D., Amin, N., Schwikowski, B. and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks, *Genome research* **13**(11): 2498–2504.

Shen, S. (2006). *Characterisation of Dichaete functions and targets during Drosophila embryonic development*, PhD thesis, University of Cambridge, Cambridge, UK.

Sinclair, A., Berta, P., Palmer, M., Hawkins, J., Griffiths, B., Smith, M., Foster, J., Frischauf, A., Lovell-Badge, R. and Goodfellow, P. (1990). A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif, *Nature* **346**(6281): 216–217.

Skeath, J. (1999). At the nexus between pattern formation and cell-type specification:

the generation of individual neuroblast fates in the *Drosophila* embryonic central nervous system, *Bioessays* **21**(11): 922–931.

Skeath, J. and Thor, S. (2003). Genetic control of *Drosophila* nerve cord development, *Current Opinion in Neurobiology* **13**(1): 8–15.

Slattery, M., Ma, L., Négre, N., White, K. and Mann, R. (2011). Genome-wide tissue-specific occupancy of the Hox protein Ultrabithorax and Hox cofactor Homothorax in *Drosophila*, *PloS One* **6**(4): e14686.

Smyth, G. (2005). Limma: linear models for microarray data, *Bioinformatics and computational biology solutions using R and Bioconductor* pp. 397–420.

Solomon, M., Larsen, P. and Varshavsky, A. (1988). Mapping protein-DNA interactions in vivo with formaldehyde: evidence that histone H4 is retained on a highly transcribed gene., *Cell* **53**(6): 937.

Soriano, N. and Russell, S. (1998). The *Drosophila* Sox-domain protein Dichaete is required for the development of the central nervous system midline, *Development* **125**(20): 3989.

Southall, T. and Brand, A. (2009). Neural stem cell transcriptional networks highlight genes essential for nervous system development, *The EMBO Journal* **28**(24): 3799–3807.

Spencer, W., Zeller, G., Watson, J., Henz, S., Watkins, K., McWhirter, R., Petersen, S., Sreedharan, V., Widmer, C., Jo, J. et al. (2011). A spatial and temporal map of*C. elegans* gene expression, *Genome Research* **21**(2): 325.

Stanley, S., Bailey, T. and Mattick, J. (2006). Gonome: measuring correlations between go terms and genomic positions, *BMC bioinformatics* **7**(1): 94.

Stathopoulos, A., Van Drenth, M., Erives, A., Markstein, M. and Levine, M. (2002). Whole-genome analysis of dorsal-ventral patterning in the *Drosophila* embryo, *Cell* **111**(5): 687–701.

Stevanovic, M. (2003). Modulation of Sox2 and Sox3 gene expression during differentiation of human neuronal precursor cell line NTERA2, *Molecular Biology Reports* **30**(2): 127–132.

Stormo, G. (2000). DNA binding sites: representation and discovery, *Bioinformatics* **16**(1): 16.

Sun, L., Chen, L., Greil, F., Negre, N., Li, T., Cavalli, G., Zhao, H., Van Steensel, B. and White, K. (2003). Protein–DNA interaction mapping using genomic tiling path

microarrays in *Drosophila*, *Proceedings of the National Academy of Sciences of the United States of America* **100**(16): 9428.

Sykacek, P., Kreil, D., Meadows, L., Auburn, R., Fischer, B., Russell, S. and Micklem, G. (2011). The impact of quantitative microarray optimization on gene expression analysis, *BMC Bioinformatics* **12**(73).

Taher, L. and Ovcharenko, I. (2009). Variable locus length in the human genome leads to ascertainment bias in functional inference for non-coding elements, *Bioinformatics* **25**(5): 578–584.

Takahashi, K. and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors, *Cell* **126**(4): 663–676.

Thomas-Chollier, M., Sand, O., Turatsinze, J., Defrance, M., Vervisch, E., Brohée, S., van Helden, J. et al. (2008). RSAT: regulatory sequence analysis tools, *Nucleic Acids Research* **36**(suppl 2): W119–W127.

Thomas, S., Li, X., Sabo, P., Sandstrom, R., Thurman, R., Canfield, T., Giste, E., Fisher, W., Hammonds, A., Celniker, S. et al. (2011). Dynamic reprogramming of chromatin accessibility during *Drosophila* embryo development, *Genome Biology* **12**(5): R43.

Tolhuis, B., Muijrers, I., De Wit, E., Teunissen, H., Talhout, W., Van Steensel, B. and Van Lohuizen, M. (2006). Genome-wide profiling of PRC1 and PRC2 Polycomb chromatin binding in *Drosophila melanogaster*, *Nature Genetics* **38**(6): 694–699.

Torii, M., Matsuzaki, F., Osumi, N., Kaibuchi, K., Nakamura, S., Casarosa, S., Guillemot, F. and Nakafuku, M. (1999). Transcription factors Mash-1 and Prox-1 delineate early steps in differentiation of neural stem cells in the developing central nervous system, *Development* **126**(3): 443–456.

Tullius, T. (2009). DNA binding shapes up, *Nature* **461**(7268): 1225–1226.

Uchikawa, M., Kamachi, Y. and Kondoh, H. (1999). Two distinct subgroups of Group B Sox genes for transcriptional activators and repressors: their expression during embryonic organogenesis of the chicken, *Mechanisms of Development* **84**(1-2): 103–120.

Urig, S., Gowher, H., Hermann, A., Beck, C., Fatemi, M., Humeny, A. and Jeltsch, A. (2002). The *Escherichia coli* Dam DNA methyltransferase modifies DNA in a highly processive reaction, *Journal of Molecular Biology* **319**(5): 1085–1096.

Uwanogho, D., Rex, M., Cartwright, E., Pearl, G., Healy, C., Scotting, P. and Sharpe, P. (1995). Embryonic expression of the chicken Sox2, Sox3 and Sox11 genes sug-

gests an interactive role in neuronal development, *Mechanisms of Development* **49**(1-2): 23–36.

van Steensel, B., Delrow, J. and Henikoff, S. (2001). Chromatin profiling using targeted DNA adenine methyltransferase, *Nature Genetics* **27**(3): 304–308.

van Steensel, B. and Henikoff, S. (2000). Identification of in vivo DNA targets of chromatin proteins using tethered Dam methyltransferase, *Nature Biotechnology* **18**(4): 424–428.

van Steensel, B. and Henikoff, S. (2003). Epigenomic profiling using microarrays, *Biotechniques* **35**(2): 346–357.

Vogel, M., Peric-Hupkes, D. and van Steensel, B. (2007). Detection of in vivo protein–DNA interactions using DamID in mammalian cells, *Nature Protocols* **2**(6): 1467–1478.

Von Hippel, P. (2004). Completing the view of transcriptional regulation, *Science* **305**(5682): 350.

Von Mering, C., Jensen, L., Snel, B., Hooper, S., Krupp, M., Foglierini, M., Jouffre, N., Huynen, M. and Bork, P. (2005). STRING: known and predicted protein–protein associations, integrated and transferred across organisms, *Nucleic Acids Research* **33**(suppl 1): D433–D437.

Walter, J. and Biggin, M. (1997). Measurement of in vivo DNA binding by sequence-specific transcription factors using UV cross-linking, *Methods* **11**(2): 215–224.

Wang, X., Lee, C., Gilmour, D. and Gergen, J. (2007). Transcription elongation controls cell fate specification in the *Drosophila* embryo, *Genes & Development* **21**(9): 1031.

Wegner, M. (1999). From head to toes: the multiple facets of Sox proteins, *Nucleic Acids Research* **27**(6): 1409.

Wegner, M. and Stolt, C. (2005). From stem cells to neurons and glia: a Soxist's view of neural development, *Trends in Neurosciences* **28**(11): 583–588.

Werner, M., Bianchi, M., Gronenborn, A. and Clore, G. (1995). Nmr spectroscopic analysis of the DNA conformation induced by the human testis determining factor Sry, *Biochemistry* **34**(37): 11998–12004.

Werner, M., Huth, J., Gronenborn, A. and Marius Clore, G. (1995). Molecular basis of human 46X, Y sex reversal revealed from the three-dimensional solution structure of the human Sry-DNA complex, *Cell* **81**(5): 705–714.

Wheeler, S., Pearson, J. and Crews, S. (2011). Time-lapse imaging reveals stereotypical patterns of *Drosophila* midline glial migration, *Developmental Biology* .

Wilczynski, B. and Furlong, E. (2010). Dynamic CRM occupancy reflects a temporal map of developmental progression, *Molecular Systems Biology* **6**(1).

Wilson, M., Barbosa-Morais, N., Schmidt, D., Conboy, C., Vanes, L., Tybulewicz, V., Fisher, E., Tavaré, S. and Odom, D. (2008). Species-specific transcription in mice carrying human chromosome 21, *Science* **322**(5900): 434.

Wilson, M. and Koopman, P. (2002). Matching Sox: partner proteins and co-factors of the Sox family of transcriptional regulators, *Current Opinion in Genetics & Development* **12**(4): 441–446.

Winter, R., Berg, O. and Von Hippel, P. (1981). Diffusion-driven mechanisms of protein translocation on nucleic acids. the *Escherichia coli* lac repressor-operator interaction: kinetic measurements and conclusions, *Biochemistry* **20**(24): 6961–6977.

Wolffe, A. and Leblanc, B. (2000). Creating molecular clues to uncover gene function., *Nature Biotechnology* **18**(4): 379.

Wright, E., Snopek, B. and Koopman, P. (1993). Seven new members of the Sox gene family expressed during mouse development., *Nucleic Acids Research* **21**(3): 744.

Yu, H. and Gerstein, M. (2006). Genomic analysis of the hierarchical structure of regulatory networks, *Proceedings of the National Academy of Sciences of the United States of America* **103**(40): 14724.

Yu, Z., Syu, L. and Mellerick, D. (2005). Contextual interactions determine whether the *Drosophila* homeodomain protein, Vnd, acts as a repressor or activator, *Nucleic Acids Research* **33**(1): 1.

Yuan, H., Corbi, N., Basilico, C. and Dailey, L. (1995). Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3., *Genes & Development* **9**(21): 2635.

Zaslaver, A., Bren, A., Ronen, M., Itzkovitz, S., Kikoin, I., Shavit, S., Liebermeister, W., Surette, M. and Alon, U. (2006). A comprehensive library of fluorescent transcriptional reporters for *Escherichia coli*, *Nature Methods* **3**(8): 623–628.

Zeitlinger, J. et al. (2007a). RNA polymerase stalling at developmental control genes in the *Drosophila melanogaster* embryo, *Nature Genetics* **39**(12): 1512–1516.

Zeitlinger, J. et al. (2007b). Whole-genome ChIP–chip analysis of dorsal, twist, and snail suggests integration of diverse patterning processes in the *Drosophila* embryo, *Genes & Development* **21**(4): 385.

Zeitouni, B., Sénatore, S., Séverac, D., Aknin, C., Sémériva, M. and Perrin, L. (2007). Signalling pathways involved in adult heart formation revealed by gene expression profiling in *Drosophila*, *PLoS Genetics* **3**(10): e174.

Zhang, L., Kasif, S., Cantor, C. and Broude, N. (2004). Gc/at-content spikes as genomic punctuation marks, *Proceedings of the National Academy of Sciences of the United States of America* **101**(48): 16855.

Zhang, Y., Wheatley, R., Fulkerson, E., Tapp, A. and Estes, P. (2011). Mastermind mutations generate a unique constellation of midline cells within the *Drosophila* CNS, *PloS One* **6**(10): e26197.

Zhao, G., Boekhoff-Falk, G., Wilson, B. and Skeath, J. (2007). Linking pattern formation to cell-type specification: Dichaete and Ind directly repress *achaete* gene expression in the *Drosophila* CNS, *Proceedings of the National Academy of Sciences of the United States of America* **104**(10): 3847–3852.

Zhao, G. and Skeath, J. (2002). The Sox-domain containing gene *Dichaete/fish-hook* acts in concert with Vnd and Ind to regulate cell fate in the *Drosophila* neuroectoderm, *Development* **129**(5): 1165–1174.

Zhao, G., Wheeler, S. and Skeath, J. (2007). Genetic control of dorsoventral patterning and neuroblast specification in the *Drosophila* central nervous system, *Int J Dev Biol* **51**(2): 107–115.

Zhao, S., Nichols, J., Smith, A. and Li, M. (2004). SoxB transcription factors specify neuroectodermal lineage choice in ES cells, *Molecular and Cellular Neuroscience* **27**(3): 332–342.

Zhong, L., Wang, D., Gan, X., Yang, T. and He, S. (2011). Parallel expansions of Sox transcription factor Group B predating the diversifications of the arthropods and jawed vertebrates, *PloS One* **6**(1): e16570.

Zubay, G., Schwartz, D. and Beckwith, J. (1970). Mechanism of activation of catabolite-sensitive genes: a positive control system, *Proceedings of the National Academy of Sciences of the United States of America* **66**(1): 104.

Appendix 1

A GO enrichment of Dichaete targets

Table 1: GO enrichments for the four Dichaete binding datasets. Four gene lists were created, for genes associated with between 1 and 4 datasets in total (named '1 dataset' to '4 datasets' according to the number of associations of the genes within it to different datasets). The GO terms shown were found to be significantly enriched for all 4 of the gene lists. The numbers shown are numbers of genes from each of the gene lists which were found to belong to the GO category in question.

GO term	1 dataset	2 datasets	3 datasets	4 datasets
cell differentiation	862	574	236	36
positive regulation of cellular process	320	216	73	14
neuron development	372	252	118	22
negative regulation of transcription, DNA-dependent	119	79	34	9
regulation of cellular component size	109	76	38	11
regulation of cellular component organization	228	138	55	13
negative regulation of biosynthetic process	181	119	47	11
cardiac cell differentiation	27	25	15	6
regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	693	423	152	29
cell division	155	101	40	10
biological regulation	1928	1073	391	58
regulation of multicellular organismal process	288	198	82	15
cellular component organization or biogenesis	1382	751	264	38
eye development	262	188	88	14
peripheral nervous system development	79	65	39	9
regulation of biological quality	513	299	116	21
transcription from RNA polymerase II promoter	208	127	48	13
response to stimulus	1297	752	310	45
transcription, DNA-dependent	511	310	111	22
regulation of cellular macromolecule biosynthetic process	651	401	146	30
oogenesis	347	218	90	17
nervous system development	640	442	204	30
cellular component organization at cellular level	1177	653	233	34
cell part morphogenesis	326	214	103	19
chemotaxis	171	130	63	10
imaginal disc-derived wing morphogenesis	193	140	64	15
segmentation	194	136	60	13
organ morphogenesis	430	321	143	29

Continued on next page

Table 1 – continued from previous page

GO term	1 dataset	2 datasets	3 datasets	4 datasets
post-embryonic morphogenesis	323	241	109	22
cell growth	81	58	33	9
compound eye morphogenesis	200	143	71	11
transcription	641	391	142	29
imaginal disc-derived appendage development	217	162	75	19
organ formation	52	46	27	6
signal transduction	700	406	171	25
cell development	712	464	189	30
developmental process involved in reproduction	379	248	90	17
negative regulation of cellular macromolecule biosynthetic process	180	118	47	11
imaginal disc development	367	279	124	25
open tracheal system development	154	116	57	10
negative regulation of cellular biosynthetic process	181	119	47	11
positive regulation of biological process	349	234	81	16
developmental process	1464	933	366	57
regulation of macromolecule metabolic process	831	495	181	31
cellular macromolecule biosynthetic process	1082	557	171	31
nucleic acid metabolic process	1114	569	191	30
axonogenesis	220	158	75	14
cell morphogenesis involved in differentiation	323	221	108	19
neurogenesis	501	347	160	26
regionalization	353	251	111	22
response to chemical stimulus	403	271	124	18
regulation of cellular process	1641	933	348	54
cell morphogenesis	360	236	119	21
system development	1066	726	314	51
regulation of nitrogen compound metabolic process	694	423	152	29
negative regulation of RNA metabolic process	126	84	36	9
cellular component organization or biogenesis at cellular level	1214	667	233	34
cellular component organization	1343	736	263	38
regulation of signaling	297	179	77	14
localization of cell	177	131	62	12
negative regulation of macromolecule biosynthetic process	180	118	47	11
regulation of cellular biosynthetic process	679	415	148	30
growth	212	138	73	19
cell motility	173	129	61	12
gland development	157	118	59	10
cell fate commitment	227	180	88	15
regulation of gene expression	749	452	169	31
wing disc pattern formation	65	54	31	7
tube development	94	70	38	8
regulation of transcription from RNA polymerase II promoter	162	108	45	12
imaginal disc morphogenesis	254	197	87	18
cellular developmental process	882	580	240	37
wing disc morphogenesis	194	141	64	15
regulation of RNA metabolic process	535	333	119	23
neuron projection development	313	215	103	19

Continued on next page

Table 1 – continued from previous page

GO term	1 dataset	2 datasets	3 datasets	4 datasets
imaginal disc-derived appendage morphogenesis	214	160	73	18
eye morphogenesis	211	153	74	13
wing disc development	249	181	81	18
multicellular organismal process	1641	1006	385	57
anatomical structure development	1318	855	343	55
signaling	825	472	190	28
negative regulation of biological process	508	310	131	21
regulation of growth	124	85	45	15
regulation of developmental growth	48	36	18	7
post-embryonic development	384	279	121	24
cell migration	169	126	61	12
cell fate determination	123	99	50	12
cell projection morphogenesis	326	214	103	19
cell morphogenesis involved in neuron differentiation	309	211	102	19
circulatory system development	67	58	27	8
regulation of biosynthetic process	679	415	148	30
locomotion	350	255	116	20
anatomical structure morphogenesis	943	631	272	47
sensory organ development	326	237	106	20
instar larval or pupal development	366	264	116	24
regulation of metabolic process	918	539	189	32
post-embryonic appendage morphogenesis	206	153	71	18
negative regulation of nitrogen compound metabolic process	160	106	44	10
appendage morphogenesis	220	165	74	18
compound eye development	248	175	84	12
generation of neurons	481	332	154	25
regulation of transcription, DNA-dependent	478	300	109	23
axon guidance	164	124	61	10
organ development	818	577	259	45
RNA biosynthetic process	513	311	111	22
neuron projection morphogenesis	309	211	101	19
regulation of cell size	89	62	36	11
pattern specification process	379	271	122	23
anatomical structure formation involved in morphogenesis	312	210	86	15
cellular component morphogenesis	411	271	133	24
embryonic pattern specification	177	120	54	11
appendage development	223	167	76	19
female gamete generation	349	219	91	17
negative regulation of cellular process	451	282	116	21
cardiovascular system development	67	58	27	8
neuron differentiation	431	294	138	23
negative regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	160	106	44	10
post-embryonic organ morphogenesis	254	197	87	18
post-embryonic organ development	270	211	92	18
regulation of biological process	1753	981	366	54
regulation of cellular metabolic process	826	492	167	30
muscle structure development	192	130	64	13
regulation of multicellular organismal development	217	151	65	13
blastoderm segmentation	162	110	50	10

Continued on next page

Table 1 – continued from previous page

GO term	1 dataset	2 datasets	3 datasets	4 datasets
asymmetric cell division	60	44	24	8
heart development	67	58	27	8
sexual reproduction	478	289	109	18
metamorphosis	324	245	109	23
embryo development	430	300	136	23
regulation of cell communication	172	108	49	11
regulation of transcription	592	369	137	29
regulation of primary metabolic process	811	485	170	30
negative regulation of transcription	145	99	41	10
regulation of anatomical structure size	139	95	46	13
ovarian follicle cell development	150	112	52	11
cellular component movement	187	139	68	14
multicellular organismal development	1320	859	348	56
tissue development	413	296	145	19
respiratory system development	155	116	57	10
dorsal/ventral pattern formation, imaginal disc	50	42	24	6
instar larval or pupal morphogenesis	314	236	104	22
gamete generation	467	283	107	18
behavior	275	188	93	17
developmental growth	130	88	48	12
regulation of developmental process	269	192	80	16
cellular process involved in reproduction	373	234	88	16
cell projection organization	367	243	114	19
regulation of macromolecule biosynthetic process	651	401	146	30
macromolecule biosynthetic process	1087	557	171	31

B modENCODE data used in Chapter 3

Original data used
BEAF32_0-12h_ChIP-chip_rel5.gff3
CP190_0-12h_ChIP-chip_rel5.gff3
CTCF_C-term_0-12h_ChIP-chip_rel5.gff3
CTCF_N-term_0-12h_ChIP-chip_rel5.gff3
H3K27Me3_0-12h_ChIP-chip_rel5.gff3
H3K27Me3_0-4h_ChIP-chip_rel5.gff3
H3K27Me3_0-4h_ChIP-seq_rel5.gff3
H3K36Me3_0-12h_ChIP-chip_rel5.gff3
H3K36Me3_2-4h_ChIP-chip_rel5.gff3
H3K4Me1_0-4h_ChIP-chip_rel5.gff3
H3K4Me1_0-4h_ChIP-seq_rel5.gff3
H3K4Me1_4-8h_ChIP-chip_rel5.gff3
H3K4Me1_4-8h_ChIP-seq_rel5.gff3
H3K4Me1_8-12h_ChIP-chip_rel5.gff3
H3K4Me1_8-12h_ChIP-seq_rel5.gff3
H3K4Me3_0-12h_ChIP-chip_rel5.gff3
H3K4Me3_2-4h_ChIP-chip_rel5.gff3
H3K9Me3_0-12h_ChIP-chip_rel5.gff3
H3K9Me3_0-4h_ChIP-chip2_rel5.gff3
H3K9Me3_0-4h_ChIP-chip_rel5.gff3
H3K9Me3_0-4h_ChIP-seq_rel5.gff3
mod(mdg4)_0-12h_ChIP-chip_rel5.gff3
su(Hw)_0-12h_ChIP-chip2_rel5.gff3
su(Hw)_0-12h_ChIP-chip_rel5.gff3

Table 2: Original modENCODE data files used for analysis in Chapter 3

Data used	Union outcome
H3K4Me1_0-4h_ChIP-chip_rel5.gff3	0-4 h H3K4Me1data
H3K4Me1_0-4h_ChIP-seq_rel5.gff3	0-4 h H3K4Me1 data
H3K4Me1_0-4h_ChIP-chip_rel5.gff3	0-12 h H3K4Me1 data
H3K4Me1_0-4h_ChIP-seq_rel5.gff3	0-12 h H3K4Me1 data
H3K4Me1_4-8h_ChIP-chip_rel5.gff3	0-12 h H3K4Me1 data
H3K4Me1_4-8h_ChIP-seq_rel5.gff3	0-12 h H3K4Me1 data
H3K4Me1_8-12h_ChIP-chip_rel5.gff3	0-12 h H3K4Me1 data
H3K4Me1_8-12h_ChIP-seq_rel5.gff3	0-12 h H3K4Me1 data
H3K9Me3_0-4h_ChIP-chip2_rel5.gff3	0-4 h H3K9Me3 data
H3K9Me3_0-4h_ChIP-chip_rel5.gff3	0-4 h H3K9Me3 data
H3K9Me3_0-4h_ChIP-seq_rel5.gff3	0-4 h H3K9Me3 data
H3K27Me3_0-4h_ChIP-chip_rel5.gff3	0-4 h H3K27Me3 data
H3K27Me3_0-4h_ChIP-seq_rel5.gff3	0-4 h H3K27Me3 data
CTCF_C-term_0-12h_ChIP-chip_rel5.gff3	0-12 h CTCF data
CTCF_N-term_0-12h_ChIP-chip_rel5.gff3	0-12 h CTCF data
su(Hw)_0-12h_ChIP-chip2_rel5.gff3	0-12 h su(Hw) data
su(Hw)_0-12h_ChIP-chip_rel5.gff3	0-12 h su(Hw) data

Table 3: modENCODE data on which unions were performed, with the union outcome shown.

Final datasets used for analysis
BEAF32_0-12h_ChIP-chip_rel5.gff3
CP190_0-12h_ChIP-chip_rel5.gff3
CTCF_0-12h_ChIP_union_rel5.gff3
H3K27Me3_0-12h_ChIP-chip_rel5.gff3
H3K27Me3_0-4h_ChIP_union_rel5.gff3
H3K36Me3_0-12h_ChIP-chip_rel5.gff3
H3K36Me3_2-4h_ChIP-chip_rel5.gff3
H3K4Me1_0-12h_ChIP_union_rel5.gff3
H3K4Me1_0-4h_ChIP_union_rel5.gff3
H3K4Me3_0-12h_ChIP-chip_rel5.gff3
H3K4Me3_2-4h_ChIP-chip_rel5.gff3
H3K9Me3_0-12h_ChIP-chip_rel5.gff3
H3K9Me3_0-4h_ChIP_union_rel5.gff3
mod(mdg4)_0-12h_ChIP-chip_rel5.gff3
suHw_0-12h_ChIP_union_rel5.gff3

Table 4: Final modENCODE data files used for comparison to Dichaete data in Chapter 3, after unions were performed with some of them.

Appendix 2

C Transcription factors with motifs flagged in NMICA searches of Dichaete binding data

Table 5: The motifs were first found using NMICA, then the matches to existing motifs from different databases were performed using Stamp. The matches were then filtered by e-value - only the transcription factors whose motifs matched the ones found with NMICA with an e-value $< 1E-04$ are shown here. The transcription factor names are shown, along with the number of Dichaete datasets that the transcription factor motif in question was found in, out of a maximum of 4.

TF name	Symbol	D datasets with motif
Suppressor of Hairless	SuH	4
Biniou	bin	1
Sine oculis	so	3
Snail	sna	3
Knirps	kni	2
Zeste	z	1
Adh transcription factor 1	Adf1	1
Topoisomerase 2	Top2	3
Ovo	ovo	1
Dorsal	dl	3
Deformed	Dfd	1
Grainyhead	grh	1
Eyeless	ey	2
Hunchback	hb	1
Rough	ro	1
Mothers against dpp	Mad	1
Sloppy paired 1	slp1	1
Chorion factor 2	Cf2	2
Dichaete	D	1
Boundary element associated factor of 32 kD	BEAF-32	3
Tailless	tll	1
Ventral nervous system defective	vnd	1
Bicoid	bcd	2
Achintya	achi	1
CG11085	CG11085	1
Deformed epidermal autoregulatory factor 1	Deaf1	1
Brinker	brk	1
Broad	br	1
Slow border cells	slbo	1
Traintrack	ttk	2
Adult enhancer factor 1	Aef1	1

Continued on next page

Table 5 – continued from previous page

TF name	Symbol	D datasets with motif
Drop	Dr	1
Hormone receptor-like in 46	Hr46	1
Runt	run	1

Appendix 3

D *sim*Gal4 gene expression data

Table 6: Expression data for genes differentially expressed in the *sim*Gal4 expression studies. The cutoffs used were M value $>= 0.5$ or $< - - 0.5$ and p-value < 0.05

		***sim*Gal4 HMG$^-$**		***sim*Gal4 EnRep**		***sim*Gal4 DWT**	
Transcript	**Gene**	**M**	**p**	**M**	**p**	**M**	**p**
CG30029-RA	CR30029	0.72	0.048	0.37	0.131	0.64	0.054
CG12045-RA	Cpr100A	-0.60	0.008	-0.21	0.518	0.01	0.991
CG6478-RA	TwdlB	-2.17	0.000	-0.58	0.481	-0.64	0.479
CG10211-RA	CG10211	-0.50	0.011	-0.08	0.712	0.02	0.924
CG5661-RA	Sema-5c	0.78	0.031	-0.28	0.062	0.47	0.123
CG1919-RA	Cpr62Bc	-1.30	0.002	-0.37	0.554	-0.44	0.432
CG13063-RA	CG13063	-1.15	0.002	-0.38	0.339	-0.24	0.665
CG2976-RA	CG2976	-0.57	0.113	-0.50	0.112	-0.64	0.033
CG7111-RA	Rack1	-0.22	0.499	-0.54	0.018	-0.29	0.121
CG10112-RA	Cpr51A	-1.86	0.001	-0.34	0.748	-0.64	0.309
CG13794-RA	CG13794	-1.15	0.022	-0.76	0.016	-0.72	0.025
CG17697-RB	fz	0.24	0.257	-0.27	0.494	0.61	0.026
CG14643-RA	TwdlG	-0.95	0.005	0.03	0.957	0.04	0.894
CG10475-RA	Jon65Ai	0.67	0.016	0.29	0.700	0.22	0.754
CG17127-RA	CG17127	-1.17	0.005	-0.14	0.826	-0.10	0.802
CG7930-RB	TpnC73F	-0.86	0.049	-0.57	0.468	0.19	0.779
CG4688-RA	CG4688	-0.65	0.030	-0.47	0.000	-0.30	0.025
CG32405-RA	Cpr65Av	-0.53	0.047	-0.11	0.408	0.07	0.837
CG6259-RA	CG6259	-0.24	0.054	-0.55	0.000	-0.06	0.060
CG15345-RA	CG15345	-1.31	0.004	-0.75	0.011	-0.59	0.018
CG11430-RA	olf186-F	-0.21	0.269	-0.56	0.030	-0.41	0.019
CG6447-RA	TwdlL	-2.60	0.000	-1.19	0.288	-0.97	0.467
CG4060-RA	TwdlW	-1.03	0.001	-0.17	0.795	-0.43	0.330
CG17082-RA	CG17082	0.59	0.017	0.27	0.202	0.14	0.413
CG2177-RB	CG2177	-0.89	0.003	-0.15	0.520	-0.48	0.085
CG4670-RA	CG4670	-0.48	0.008	-0.51	0.006	-0.52	0.020
CG15658-RA	CG15658	-0.54	0.036	-0.44	0.058	-0.52	0.007
CG4464-RB	RpS19a	0.53	0.044	-0.02	0.896	0.27	0.389
CG15117-RA	CG15117	-0.57	0.025	-0.08	0.657	-0.13	0.306
CG10541-RA	Tektin-C	0.28	0.182	0.11	0.454	0.64	0.012
CG1304-RA	CG1304	0.81	0.026	0.17	0.585	0.45	0.096
CG32659-RA	Ten-a	-0.57	0.049	-0.10	0.795	0.12	0.695
CG10901-RB	osk	0.73	0.035	0.03	0.799	0.28	0.301
CG7532-RA	l(2)34Fc	-0.59	0.045	-0.25	0.646	0.04	0.958
CG32305-RA	CG32305	-0.44	0.024	-0.50	0.029	-0.20	0.446
CG4147-RA	Hsc70-3	0.17	0.123	0.10	0.573	0.64	0.031

Continued on next page

Table 6 – continued from previous page

		*sim*Gal4 HMG$^-$		*sim*Gal4 EnRep		*sim*Gal4 DWT	
Transcript	**Gene**	**M**	**p**	**M**	**p**	**M**	**p**
CG4835-RA	CG4835	0.29	0.089	0.53	0.013	0.52	0.013
CG1840-RA	CG1840	0.56	0.008	0.24	0.167	0.46	0.001
CG4354-RA	slbo	-0.61	0.011	-0.08	0.852	-0.54	0.024
CG5514-RA	CG5514	-0.45	0.056	-0.70	0.018	-0.67	0.033
CG9337-RA	CR9337	-0.54	0.004	-0.38	0.028	-0.50	0.015
CG10657-RA	CG10657	-0.59	0.010	-0.03	0.943	-0.22	0.513
CG1169-RA	Osi18	-0.64	0.001	-0.19	0.714	-0.04	0.856
CR32205-RA	CR32205	-0.62	0.007	-0.56	0.003	-0.38	0.056
CG13463-RA	CG13463	-1.55	0.001	-0.90	0.023	-1.10	0.000
CG5326-RA	CG5326	-0.88	0.005	-0.10	0.884	-0.32	0.481
CG32692-RA	CG32694	-0.67	0.013	0.04	0.936	-0.09	0.765
CG4466-RA	Hsp27	0.84	0.001	-0.08	0.663	0.30	0.263
CG11242-RA	CG11242	0.63	0.001	0.25	0.099	0.14	0.531
CG14095-RA	CG14095	-0.70	0.002	0.31	0.579	-0.23	0.328
CG30413-RA	CG30413	-1.43	0.006	-0.56	0.355	-1.07	0.001
CG8507-RA	CG8507	0.51	0.013	0.34	0.069	0.02	0.933
CG10119-RA	LamC	-0.63	0.026	-0.25	0.676	-0.18	0.618
CG16886-RA	CG16886	-0.93	0.000	-0.06	0.901	-0.62	0.034
CG1330-RA	Ccp84Ae	-0.67	0.018	0.07	0.856	0.37	0.575
CG2849-RA	Rala	1.45	0.008	0.58	0.020	1.14	0.006
CG8965-RA	CG8965	0.29	0.115	0.26	0.440	0.62	0.040
CG15251-RA	CG43386	-0.88	0.025	0.14	0.803	-0.11	0.839
CG14446-RA	CG14446	0.51	0.001	0.27	0.110	0.45	0.115
CG9614-RA	pip	-0.56	0.029	-0.09	0.444	-0.06	0.331
CG9564-RA	Try29F	-0.57	0.002	-0.13	0.301	-0.44	0.000
CG14147-RA	CG14147	-1.19	0.000	0.15	0.856	-0.42	0.384
CG15212-RA	CG15212	-1.67	0.002	-0.43	0.548	-0.73	0.306
CG7203-RA	CG7203	-0.73	0.007	-0.38	0.162	-0.16	0.535
CG11061-RA	GM130	-0.51	0.029	-0.02	0.833	-0.22	0.097
CG6868-RA	tld	-0.00	0.953	0.00	0.958	0.51	0.002
CG11205-RB	phr	-0.79	0.000	-0.78	0.005	-0.83	0.000
CG10287-RA	Gasp	-0.75	0.010	-0.03	0.962	0.06	0.896
CG1155-RA	Osi14	-0.65	0.013	0.11	0.803	-0.04	0.885
CG13067-RA	CG13067	-0.98	0.022	-0.52	0.354	-0.08	0.938
CG31364-RA	l(3)neo38	0.72	0.000	0.48	0.072	0.57	0.029
CG13049-RA	CG13049	-1.17	0.001	-0.05	0.951	-0.36	0.486
CG31532-RA	CG42574	0.58	0.023	0.31	0.150	0.39	0.028
CG18543-RA	mtrm	0.80	0.042	-0.08	0.538	-0.38	0.335
CG10901-RA	osk	0.83	0.049	0.02	0.860	0.08	0.610
CG17462-RA	Trf4-2	0.61	0.001	0.30	0.137	0.22	0.608
CG10591-RA	CG10591	-0.86	0.027	-0.43	0.293	-0.45	0.397
CG14897-RB	CG42232	1.16	0.001	0.85	0.036	0.71	0.009
CG7577-RA	ppk20	-0.01	0.955	2.18	0.003	0.10	0.212
CR32477-RA	CR42862	-0.53	0.009	-0.37	0.014	-0.55	0.041
CG13059-RA	CG13059	-1.93	0.010	-1.03	0.129	-0.95	0.279
CG13738-RA	CG13738	-0.56	0.004	-0.12	0.412	-0.41	0.000
CG12876-RA	ALiX	-0.09	0.201	-0.39	0.027	-0.53	0.004
CG14993-RA	Faa	0.75	0.014	0.60	0.025	0.63	0.065
CG10953-RA	CG10953	-1.37	0.021	-0.65	0.351	-0.18	0.857
CG32237-RA	CG32237	-1.05	0.001	-0.27	0.335	-0.10	0.903
CG7660-RA	Pxt	0.81	0.013	0.06	0.517	-0.01	0.981
CG3509-RA	CG3509	0.87	0.001	0.23	0.552	0.12	0.714
CG8844-RB	Pdsw	0.60	0.016	0.16	0.018	0.36	0.056

Continued on next page

Table 6 – continued from previous page

		*sim*Gal4 HMG$^-$		*sim*Gal4 EnRep		*sim*Gal4 DWT	
Transcript	**Gene**	**M**	**p**	**M**	**p**	**M**	**p**
CR31808-RA	CG31808	-1.14	0.005	-0.37	0.413	-0.68	0.129
CG18817-RB	Tsp42Ea	-0.77	0.019	-0.34	0.205	-0.46	0.108
CG4998-RA	CG4998	-0.66	0.006	0.15	0.690	-0.13	0.700
CG6081-RA	Cyp28d2	-1.13	0.009	-0.41	0.013	-0.56	0.086
CG15239-RA	CG15239	-0.67	0.024	-0.10	0.748	-0.19	0.346
CG11584-RB	CG11584	-0.88	0.003	0.02	0.973	-0.22	0.476
CG14143-RA	CG14143	-0.68	0.038	0.32	0.333	-0.29	0.069
CG17018-RD	CG17018	0.69	0.018	0.19	0.349	0.00	0.987
CG2342-RA	Ccp84Ag	-1.18	0.002	-0.12	0.847	-0.10	0.890
CG17052-RA	obst-A	-0.85	0.006	-0.16	0.847	-0.11	0.856
CG13991-RA	CG13991	-0.85	0.001	-0.88	0.006	-0.82	0.001
CG8776-RA	nemy	-0.64	0.012	-0.15	0.579	-0.05	0.687
CG13425-RD	bl	-1.02	0.002	-1.04	0.009	-0.92	0.001
CG3759-RA	CG3759	-0.57	0.014	-0.12	0.187	0.02	0.919
CG6146-RA	Top1	0.56	0.004	0.32	0.157	0.30	0.243
CG9155-RC	Myo61F	1.68	0.000	1.61	0.001	0.22	0.553
CG12743-RA	otu	0.77	0.015	-0.00	0.987	0.03	0.917
CG17227-RB	lig3	0.24	0.202	0.61	0.001	0.51	0.005
CG13425-RC	bl	-1.22	0.000	-1.24	0.015	-1.18	0.001
CG3520-RA	CG3520	-0.52	0.010	-0.25	0.072	-0.13	0.698
CG14897-RA	CG42232	0.78	0.033	0.72	0.013	0.67	0.016
CG3396-RA	Ocho	0.60	0.006	0.11	0.477	0.14	0.362
CG1151-RA	Osi6	-1.65	0.031	-0.53	0.410	-0.02	0.965
CG12418-RA	Glut4EF	0.57	0.005	0.59	0.125	0.49	0.159
CG13784-RB	CG13784	0.64	0.004	0.62	0.008	0.40	0.100
CG30141-RA	Obp57a	-0.72	0.026	-0.38	0.020	0.16	0.180
CG13069-RA	CG13069	-1.71	0.002	-0.40	0.544	-0.33	0.769
CG14243-RA	TwdlD	-1.82	0.001	-0.54	0.522	-1.04	0.195
CG8960-RA	CG8960	-0.11	0.089	-0.79	0.004	-0.76	0.015
CG3502-RA	CG3502	0.16	0.287	-0.46	0.002	-0.62	0.001
CG2962-RA	CG2962	-2.82	0.001	-1.27	0.296	-1.54	0.065
CG14453-RA	CG14453	-0.89	0.007	-0.77	0.013	-0.81	0.028
CG8050-RA	Cys	0.63	0.025	-0.18	0.306	0.82	0.013
CG13047-RA	CG13047	-1.08	0.004	0.04	0.948	-0.02	0.979
CR41110-RA	fog	0.88	0.006	0.39	0.030	0.44	0.411
CR32862-RA	snRNA:U1:82Eb	1.40	0.011	0.75	0.039	0.78	0.337
CG18783-RA	Kr-h1	-0.56	0.007	-0.32	0.150	-0.22	0.346
CG7980-RA	RabX5	0.28	0.001	0.54	0.018	0.28	0.347
CG8640-RA	Cpr65Ea	-2.12	0.001	-0.67	0.540	-0.76	0.266
CG6874-RA	HipHop	-0.31	0.061	-0.58	0.010	-0.16	0.479
CG31813-RA	CG31813	-0.66	0.045	-0.17	0.591	-0.11	0.659
CG17534-RA	GstE9	-0.04	0.844	0.30	0.006	0.60	0.014
CG17090-RB	hipk	0.23	0.219	0.38	0.107	0.55	0.012
CG12063-RA	mey	-0.66	0.001	-0.09	0.562	-0.41	0.004
CG14265-RB	CG14265	-0.42	0.305	-0.69	0.214	-0.83	0.016
CG2341-RA	Ccp84Ad	-0.71	0.028	0.06	0.865	0.14	0.842
CG1742-RB	Mgstl	0.51	0.004	0.47	0.002	0.42	0.002
CG15740-RA	CG15740	-0.82	0.002	0.02	0.976	-0.25	0.464
CG31012-RB	cindr	-0.58	0.011	-0.16	0.296	-0.37	0.098
CG9614-RI	pip	-0.51	0.011	-0.09	0.219	-0.17	0.066
CG15022-RA	CG15022	-0.92	0.002	-0.11	0.817	-0.71	0.020
CG10414-RA	Atac2	-0.45	0.006	-0.53	0.001	-0.51	0.014
CG12297-RA	BG4	0.25	0.064	0.88	0.004	0.54	0.004

Continued on next page

Table 6 – continued from previous page

		*sim*Gal4 HMG$^-$		*sim*Gal4 EnRep		*sim*Gal4 DWT	
Transcript	**Gene**	**M**	**p**	**M**	**p**	**M**	**p**
CG32453-RA	CG32453	-0.79	0.012	-0.42	0.184	-0.63	0.057
CG12283-RA	kek1	-0.64	0.013	-0.33	0.048	-0.34	0.006
CG12885-RA	CG12885	-0.76	0.001	-0.39	0.023	-0.40	0.106
CG14685-RD	Cap-H2	0.62	0.007	0.50	0.018	0.50	0.004
CG15321-RA	nej	0.60	0.005	0.38	0.054	0.31	0.236
CG33320-RA	CheB38a	-0.97	0.012	-0.38	0.001	-0.46	0.027
CG3598-RA	CG3598	0.68	0.049	0.29	0.278	0.99	0.269
CG9628-RA	CG9628	-0.82	0.001	-0.40	0.216	-0.37	0.211
CG6460-RA	TwdlK	-1.38	0.002	-0.22	0.785	0.02	0.984
CG13041-RA	CG13041	-0.40	0.038	-0.52	0.043	-0.15	0.403
CG7429-RA	CG7429	0.64	0.003	0.07	0.463	0.12	0.316
CG12449-RB	Gfat1	-0.53	0.044	-0.29	0.572	0.09	0.822
CG5468-RA	TwdlM	-1.36	0.011	-0.37	0.586	-0.24	0.700
CG13066-RA	CG13066	-1.23	0.008	-0.47	0.267	-0.08	0.933
CG31220-RA	CG31220	0.58	0.000	0.44	0.017	0.21	0.037
CG32694-RD	CG32694	-0.69	0.004	0.11	0.834	-0.02	0.972
CG2961-RA	Ipod	-0.60	0.010	-0.07	0.880	-0.40	0.135
CG31362-RA	Jon99Ciii	1.52	0.000	0.75	0.356	0.81	0.288
CG6643-RA	Esyt2	-0.52	0.025	0.11	0.554	0.04	0.778
CG15013-RB	dyl	-0.66	0.027	-0.14	0.696	-0.06	0.803
CG1944-RA	Cyp4p2	-1.45	0.004	-1.02	0.002	-1.17	0.001
CG14756-RA	CG14756	-0.90	0.040	-0.42	0.362	-0.57	0.196
CG1259-RB	Cpr64Ad	-1.67	0.000	-0.23	0.808	0.07	0.938
CG13123-RA	CG13123	-0.19	0.046	-0.61	0.002	-0.09	0.068
CG30422-RA	egg	0.55	0.000	0.23	0.191	0.16	0.307
CG30458-RA	CG30458	-0.79	0.002	-0.19	0.602	0.04	0.937
CG4547-RA	Atx-1	-0.07	0.726	-0.54	0.049	-0.22	0.328
CG32266-RA	CG32266	-1.57	0.004	-0.82	0.114	-0.22	0.783
CG13394-RA	CG42819	-0.68	0.045	-0.13	0.533	0.03	0.938
CG15236-RB	CG15236	-0.90	0.015	-0.63	0.008	-0.47	0.004
CG9057-RA	Lsd-2	0.51	0.016	0.03	0.883	0.11	0.537
CG1555-RA	cn	0.21	0.292	0.31	0.233	0.58	0.022
CG13068-RA	CG13068	-1.05	0.003	-0.46	0.300	0.11	0.894
CG14096-RA	CG14096	-0.79	0.009	0.28	0.499	0.01	0.983
CG12177-RA	CG12177	-0.42	0.043	-0.58	0.000	-0.57	0.008
CG1157-RA	Osi15	-1.48	0.010	-0.41	0.637	-0.30	0.600
CG4396-RA	fne	-0.53	0.045	-0.17	0.549	-0.18	0.268
CG14573-RA	CG14573	-0.69	0.003	-0.32	0.187	0.02	0.973
CG33117-RA	Victoria	0.65	0.000	0.29	0.010	0.40	0.010
CG13362-RA	CG13362	-0.50	0.033	-0.23	0.242	-0.16	0.309
CG4115-RA	CG4115	-1.15	0.007	-0.44	0.581	-0.29	0.480
CG13060-RA	CG13060	-0.73	0.003	-0.22	0.286	-0.21	0.434
CG8628-RA	CG8628	0.82	0.010	0.20	0.505	0.80	0.262
CG31641-RB	stai	0.80	0.027	0.23	0.134	0.05	0.498
CG4038-RA	CG4038	0.50	0.005	0.19	0.232	0.46	0.015
CG9411-RA	CG9411	-0.64	0.011	-0.27	0.173	-0.18	0.177
CG11529-RA	CG11529	0.07	0.454	0.12	0.106	0.65	0.042
CG14639-RA	TwdlF	-1.10	0.027	-0.29	0.603	-0.39	0.038
CG32570-RA	TwdlY	-1.16	0.020	-0.21	0.715	-0.20	0.664
CG14302-RA	CG14302	0.60	0.039	-0.21	0.796	-0.04	0.925
CG5494-RA	Cpr92F	-0.53	0.004	-0.05	0.824	-0.25	0.241
CG13674-RA	CG13674	-1.09	0.039	-0.30	0.532	-0.02	0.979
CG32701-RA	l(1)G0320	0.13	0.641	0.20	0.052	0.50	0.014

Continued on next page

Table 6 – continued from previous page

		***sim*Gal4 HMG$^-$**		***sim*Gal4 EnRep**		***sim*Gal4 DWT**	
Transcript	**Gene**	**M**	**p**	**M**	**p**	**M**	**p**
CG6610-RA	CG6610	-0.28	0.176	-0.57	0.010	-0.29	0.168
CG32694-RA	CG32694	-0.81	0.029	0.06	0.912	-0.25	0.614
CG18124-RA	mTTF	-0.49	0.041	-0.19	0.030	-0.65	0.015
CG10812-RA	dro5	-0.77	0.128	-0.87	0.019	-0.72	0.181
CG13454-RA	CG13454	-0.68	0.045	-1.02	0.001	-0.04	0.829
CG13631-RA	CG13631	-0.70	0.028	-0.35	0.436	-0.40	0.058
CG8598-RA	eco	-0.31	0.142	-0.55	0.023	-0.07	0.705
CG12076-RA	YT521-B	-0.26	0.200	-0.36	0.235	-0.50	0.004
CG12487-RA	BobA	0.03	0.849	-0.73	0.016	-0.46	0.120
CG5476-RA	TwdlN	-1.29	0.009	-1.01	0.154	-0.55	0.530
CG31164-RA	Ir94a	-0.62	0.028	-0.65	0.008	-0.35	0.017
CG10214-RA	CG10214	0.04	0.812	-0.56	0.013	-0.03	0.860
CG8235-RA	CG8235	-0.37	0.242	-0.60	0.003	-0.03	0.882
CG3777-RC	CG3777	-0.68	0.038	-0.06	0.885	0.06	0.874
CG14736-RA	CG14736	0.81	0.011	0.02	0.804	0.03	0.722
CG1740-RA	Ntf-2	-0.39	0.061	-0.69	0.044	-0.57	0.046
CG15213-RA	CG15213	-0.90	0.017	-0.45	0.431	-0.11	0.858
CG11797-RA	Obp56a	0.68	0.028	0.39	0.232	1.02	0.210
CG7361-RA	RFeSP	-0.34	0.205	1.53	0.001	-0.28	0.119
CG4584-RB	dUTPase	0.09	0.687	-0.52	0.044	-0.18	0.613
CG15535-RA	CG15535	-0.24	0.099	-0.28	0.004	-0.50	0.021
CG12822-RA	CG12822	-0.39	0.017	-0.39	0.003	-0.55	0.046
CG15592-RA	Osi9	-0.98	0.023	-0.38	0.524	-0.21	0.514
CG3819-RA	CG3819	0.60	0.008	0.21	0.700	0.50	0.346
CG2615-RB	ik2	-0.29	0.231	-0.50	0.046	-0.44	0.086
CG5025-RA	Sps2	-0.30	0.098	-0.42	0.079	-0.50	0.033
CG15737-RA	wisp	0.61	0.048	0.06	0.765	-0.18	0.457
CG7380-RA	baf	-0.25	0.309	-0.58	0.042	-0.35	0.257
CG14688-RA	CG14688	-0.33	0.183	-0.52	0.020	-0.44	0.098

E *pros*Gal4 gene expression data

Table 7: Expression data for genes differentially expressed in the *pros*Gal4 expression studies. The cutoffs used were M value $>= 0.5$ or $< - - 0.5$ and p-value < 0.05

		***pros*Gal4 HMG$^-$**		***pros*Gal4 EnRep**		***pros*Gal4 DWT**	
Transcript	**Gene**	**M**	**p**	**M**	**p**	**M**	**p**
CG3429-RA	swa	-1.08	0.103	0.81	0.043	0.03	0.970
CG1088-RA	Vha26	-0.80	0.047	0.74	0.097	-0.42	0.390
CG5864-RA	AP-1sigma	-0.85	0.017	0.57	0.202	-0.11	0.862
CG4148-RA	wek	-0.53	0.023	0.36	0.238	-0.15	0.783
CG2943-RA	CG2943	-0.60	0.008	0.34	0.241	-0.08	0.794
CG4759-RA	RpL27	-0.59	0.006	0.11	0.701	-0.41	0.151
CG6420-RA	CG6420	-0.70	0.012	0.41	0.247	-0.36	0.641
CG4165-RA	CG4165	-0.83	0.036	0.48	0.343	-0.05	0.932
CG3362-RA	CG3362	-0.56	0.046	0.21	0.337	-0.38	0.489
CG5946-RB	CG5946	-0.69	0.042	0.55	0.117	-0.15	0.829
CG5661-RA	Sema-5c	0.58	0.007	-0.12	0.122	0.20	0.475
CG33281-RA	CG33281	-0.29	0.117	-0.50	0.013	-0.34	0.232
CG32606-RA	mamo	0.27	0.331	0.65	0.009	0.35	0.430

Continued on next page

Table 7 – continued from previous page

		*pros*Gal4 HMG$^-$		*pros*Gal4 EnRep		*pros*Gal4 DWT	
Transcript	**Gene**	**M**	**p**	**M**	**p**	**M**	**p**
CG32654-RC	Sec16	-0.52	0.007	0.55	0.092	-0.21	0.657
CR33328-RA	CR33328	0.69	0.006	0.06	0.722	0.13	0.775
CG30289-RA	CG30289	-0.13	0.497	-0.50	0.000	-0.51	0.160
CG5263-RA	smg	-1.62	0.042	0.76	0.103	-0.52	0.670
CG12626-RA	CG42339	0.15	0.210	0.03	0.780	0.64	0.006
CG5362-RA	Mdh1	-1.18	0.123	0.78	0.039	-0.24	0.844
CG9926-RA	CG9926	0.28	0.081	0.10	0.463	0.52	0.036
CG30096-RA	CG30096	-0.53	0.007	0.46	0.155	0.12	0.813
CG5053-RA	RASSF8	-0.81	0.040	0.49	0.236	-0.15	0.744
CR31273-RA	bxd	0.75	0.003	0.62	0.016	-0.01	0.952
CG14883-RA	CG14883	-0.69	0.041	0.70	0.026	0.04	0.960
CG17369-RB	Vha55	-0.45	0.229	0.75	0.030	0.27	0.709
CG18361-RA	dsh	-0.57	0.029	0.39	0.262	-0.06	0.934
CG7641-RA	Nca	-0.59	0.012	0.46	0.262	-0.10	0.832
CG4817-RA	Ssrp	0.59	0.046	-0.25	0.221	0.17	0.649
CG6951-RA	CG6951	-0.39	0.072	0.50	0.033	0.04	0.939
CG17883-RB	CG17883	-0.75	0.029	0.61	0.106	0.01	0.988
CG11589-RA	VhaM9.7-c	-0.64	0.028	0.60	0.046	-0.16	0.758
CG6549-RC	fws	-0.57	0.012	0.32	0.368	0.07	0.884
CG7548-RA	CG7548	-0.32	0.029	-0.54	0.015	-0.42	0.051
CG6171-RA	CG6171	-0.88	0.022	0.65	0.078	-0.23	0.750
CG8944-RA	CG8944	-0.68	0.014	0.55	0.147	0.12	0.872
CG13928-RA	CG13928	0.20	0.563	0.74	0.015	0.29	0.277
CG5515-RA	CG5515	-0.68	0.035	0.47	0.165	-0.11	0.880
CG14100-RA	CG14100	-0.62	0.024	0.70	0.038	0.25	0.718
CG8605-RA	CG8605	-0.07	0.757	-0.59	0.014	-0.61	0.359
CG9279-RA	CG9279	-0.37	0.075	-0.72	0.035	-0.92	0.175
CG1815-RA	CG1815	-0.71	0.018	0.49	0.080	-0.05	0.931
CG4433-RA	CG4433	-0.22	0.186	0.59	0.018	0.28	0.543
CG4999-RC	Tsp66E	-1.20	0.027	0.81	0.024	-0.34	0.695
CG12052-RE	lola	-0.59	0.010	0.56	0.288	0.09	0.880
CG8931-RA	CG8931	-0.76	0.009	0.51	0.184	-0.19	0.742
CG4538-RB	CG4538	-0.68	0.022	0.59	0.100	0.21	0.786
CG7299-RA	CG7299	-0.75	0.040	-0.96	0.048	-0.76	0.192
CG13794-RA	CG13794	-1.16	0.041	-1.32	0.004	-1.09	0.159
CG1927-RA	CG1927	-1.11	0.040	0.62	0.128	-0.29	0.761
CG12106-RA	CG12106	-0.17	0.684	0.56	0.021	0.34	0.707
CG12787-RA	hoe1	-1.26	0.051	0.78	0.036	-0.16	0.885
CG32782-RC	tlk	-0.57	0.046	0.51	0.194	-0.28	0.657
CG10240-RA	Cyp6a22	-0.77	0.007	0.49	0.181	0.36	0.649
CG2790-RA	CG2790	-0.57	0.019	0.26	0.439	-0.20	0.707
CG10475-RA	Jon65Ai	-1.62	0.003	-1.42	0.081	-1.82	0.007
CG14850-RA	CG14850	-0.51	0.033	-0.41	0.355	-0.67	0.027
CG10529-RA	Lcp65Ae	-1.90	0.158	-1.28	0.398	-3.00	0.024
CG2204-RA	G-oalpha47A	-0.42	0.324	0.64	0.009	-0.02	0.978
CG6321-RA	CG6321	-0.96	0.032	0.37	0.119	-0.27	0.727
CG3309-RA	CG3309	-0.58	0.018	0.53	0.101	0.00	0.996
CG10939-RA	Sip1	-0.54	0.014	0.40	0.138	-0.23	0.533
CG7184-RA	Mkrn1	-0.94	0.003	0.47	0.186	-0.40	0.473
CG7192-RA	CG7192	-0.57	0.043	0.39	0.094	-0.25	0.624
CG32198-RB	CG32198	-1.33	0.006	-1.42	0.027	-1.63	0.019
CG8223-RA	CG8223	-0.60	0.001	0.32	0.326	-0.27	0.505
CG14011-RA	CG14011	0.21	0.072	0.53	0.014	0.20	0.010

Continued on next page

Table 7 – continued from previous page

		*pros*Gal4 HMG$^-$		*pros*Gal4 EnRep		*pros*Gal4 DWT	
Transcript	**Gene**	**M**	**p**	**M**	**p**	**M**	**p**
CG5880-RA	CG5880	-0.75	0.006	0.47	0.112	0.01	0.991
CG8357-RA	Drep-1	-0.14	0.284	0.53	0.026	0.46	0.365
CG4688-RA	CG4688	-0.42	0.032	-0.77	0.073	-0.51	0.007
CG9238-RB	CG9238	-0.50	0.037	0.68	0.114	0.03	0.950
CG4390-RA	CG4390	-0.56	0.026	0.21	0.389	-0.28	0.618
CG4217-RA	TFAM	-0.58	0.023	0.60	0.265	-0.27	0.609
CG13142-RA	CG13142	-0.57	0.010	0.50	0.228	0.01	0.976
CG11981-RA	Prosbeta3	-0.91	0.021	0.58	0.063	-0.32	0.583
CG7138-RA	r2d2	-0.64	0.008	0.39	0.324	0.03	0.951
CG14103-RA	CG14103	-0.91	0.009	0.77	0.068	0.11	0.869
CG13037-RA	mRpS34	0.16	0.538	-0.68	0.031	-0.40	0.382
CG2924-RA	CG2924	-0.90	0.126	0.83	0.036	-0.19	0.827
CG3407-RA	CG3407	-0.57	0.032	0.92	0.151	-0.09	0.887
CG17490-RC	CG17490	-0.61	0.047	0.48	0.213	-0.02	0.970
CG15345-RA	CG15345	-1.04	0.048	0.25	0.798	-0.85	0.319
CG2060-RA	Cyp4e2	-0.47	0.082	0.66	0.002	0.14	0.802
CG11430-RA	olf186-F	-0.28	0.099	-0.81	0.026	-0.83	0.117
CG4060-RA	TwdlW	-0.56	0.045	-0.67	0.149	-0.32	0.595
CG31517-RA	CG31517	-0.94	0.025	0.32	0.302	-0.30	0.682
CG8997-RA	CG8997	-0.62	0.010	-1.18	0.006	-0.87	0.037
CG16869-RA	Ance-2	-0.17	0.212	-0.78	0.013	-0.21	0.734
CG7700-RA	Gos28	-0.46	0.289	0.51	0.034	0.09	0.911
CG10387-RA	tos	0.11	0.536	-0.50	0.036	-0.21	0.574
CG13323-RA	CG13323	-0.67	0.022	-0.97	0.025	-0.95	0.045
CG2177-RB	CG2177	-1.77	0.023	-0.04	0.833	-1.97	0.150
CG4670-RA	CG4670	-0.70	0.013	0.08	0.541	-0.18	0.620
CG2702-RA	Pbp95	-0.61	0.034	0.51	0.093	-0.02	0.974
CG6897-RA	bora	-0.53	0.039	0.28	0.220	0.08	0.850
CG18734-RG	Fur2	-1.39	0.018	0.63	0.148	-0.37	0.730
CG12244-RA	lic	-0.88	0.123	0.71	0.046	-0.00	0.997
CG8841-RC	CG8841	-0.97	0.021	0.80	0.051	0.13	0.861
CG4898-RE	Tm1	-0.73	0.046	0.36	0.237	-0.05	0.945
CG4620-RB	unk	-0.74	0.048	0.46	0.270	-0.06	0.921
CG33122-RA	cutlet	0.06	0.281	-0.61	0.009	-0.44	0.064
CG5804-RA	CG5804	-0.58	0.022	-0.65	0.066	-0.54	0.154
CG17665-RA	CG17665	-0.53	0.108	0.51	0.014	0.21	0.762
CG31849-RA	CG31849	-0.55	0.049	0.49	0.006	-0.07	0.895
CG1225-RA	RhoGEF3	0.77	0.007	-0.35	0.374	-0.13	0.816
CG2714-RB	crm	-0.77	0.153	0.58	0.009	-0.07	0.944
CG8988-RA	S2P	-1.04	0.038	0.63	0.108	-0.16	0.887
CG8311-RA	CG8311	-0.37	0.201	0.50	0.035	0.33	0.604
CG1692-RA	mal	-0.82	0.009	0.43	0.167	-0.09	0.898
CG11577-RA	CG11577	-0.65	0.047	0.52	0.044	-0.03	0.967
CG3889-RA	CSN1b	-0.57	0.058	0.55	0.031	-0.09	0.887
CG1044-RA	dos	0.13	0.302	-0.51	0.015	-0.10	0.613
CG10541-RA	Tektin-C	1.66	0.000	-0.04	0.847	0.88	0.251
CG8798-RA	CG8798	-0.57	0.025	0.44	0.080	0.24	0.694
CG9786-RA	hb	0.33	0.163	0.13	0.545	0.56	0.036
CG2041-RA	lgs	-0.57	0.028	0.42	0.155	0.07	0.860
CG3348-RA	CG3348	-0.43	0.064	-0.87	0.012	-0.36	0.401
CG8634-RA	Cpr65Ec	0.12	0.678	0.68	0.025	1.30	0.499
CG32189-RA	CG32189	-0.38	0.007	-0.50	0.020	-0.17	0.355
CG3679-RA	CG3679	-1.09	0.030	0.74	0.076	-0.02	0.982

Continued on next page

Table 7 – continued from previous page

		*pros*Gal4 HMG$^-$		*pros*Gal4 EnRep		*pros*Gal4 DWT	
Transcript	**Gene**	**M**	**p**	**M**	**p**	**M**	**p**
CG9773-RA	CG9773	-0.36	0.251	0.52	0.026	0.28	0.696
CG6668-RA	atl	-0.59	0.031	0.46	0.113	0.00	0.994
CG9681-RA	PGRP-SB1	0.76	0.001	-0.10	0.383	-0.13	0.638
CG9834-RA	endoB	-0.52	0.002	0.61	0.161	0.12	0.836
CG17342-RA	Lk6	-0.50	0.005	0.09	0.808	-0.42	0.246
CG32659-RA	Ten-a	-0.10	0.435	-0.53	0.013	-0.76	0.208
CG5690-RA	CG5690	0.53	0.013	0.00	0.983	-0.29	0.685
CG17508-RA	CG17508	-1.13	0.037	0.60	0.151	-0.11	0.908
CG32305-RA	CG32305	-0.60	0.028	-0.65	0.057	-0.46	0.137
CG3319-RA	Cdk7	-0.51	0.005	0.28	0.352	-0.09	0.810
CG4287-RA	CG4287	-0.73	0.000	0.39	0.185	0.14	0.774
CG7044-RA	CG7044	-0.61	0.010	0.39	0.169	0.08	0.871
CG6376-RA	E2f	0.51	0.016	0.15	0.482	-0.01	0.922
CG5704-RA	CG5704	-0.94	0.049	0.59	0.041	-0.11	0.874
CG18495-RB	Prosalpha1	-0.61	0.029	0.20	0.548	-0.46	0.227
CG5295-RA	bmm	-0.70	0.017	0.24	0.457	-0.37	0.365
CG7465-RA	CG7465	-0.68	0.269	-0.25	0.650	-1.07	0.004
CG11523-RA	CG11523	-0.24	0.459	0.60	0.020	0.48	0.558
CG5277-RA	Ip259	-0.43	0.107	-0.05	0.826	-0.56	0.009
CG32030-RA	Fhos	-1.12	0.027	0.67	0.105	-0.19	0.800
CG7324-RA	CG7324	-0.74	0.012	0.45	0.130	-0.32	0.561
CG14025-RA	Bsg25D	-0.96	0.013	0.61	0.147	0.07	0.915
CG4911-RA	CG4911	-1.10	0.030	0.71	0.135	-0.05	0.949
CG8177-RK	CG8177	-1.02	0.030	0.69	0.040	-0.20	0.820
CG3947-RA	pex16	-0.54	0.038	0.26	0.341	-0.41	0.404
CG4867-RA	bc10	-0.57	0.075	0.55	0.028	-0.15	0.800
CG6498-RA	CG6498	-0.60	0.034	0.36	0.331	0.30	0.653
CG8420-RA	CG8420	-0.77	0.019	0.55	0.139	-0.13	0.859
CG18812-RA	CG18812	-1.19	0.025	0.73	0.073	-0.15	0.872
CG7869-RA	SuUR	-0.77	0.019	0.27	0.391	-0.17	0.758
CG32791-RA	CG32791	0.51	0.034	0.15	0.110	0.61	0.244
CG14360-RA	Or88a	0.26	0.130	-0.03	0.740	0.51	0.045
CG9474-RA	Snap24	-0.75	0.011	0.43	0.215	-0.23	0.667
CG3262-RA	CG3262	-0.59	0.024	0.21	0.398	-0.45	0.329
CG5725-RB	fbl	-0.52	0.007	0.36	0.229	-0.09	0.826
CG4069-RA	CG4069	-0.55	0.023	0.36	0.154	0.13	0.805
CG9326-RB	vari	-0.44	0.052	0.88	0.012	0.55	0.338
CG5052-RA	pim	-0.76	0.030	0.48	0.286	-0.09	0.913
CG1275-RB	CG1275	-0.57	0.024	0.34	0.203	0.00	0.996
CG8569-RA	CG8569	-0.80	0.049	0.46	0.036	-0.05	0.949
CG30000-RA	CG30000	-0.89	0.012	0.52	0.194	-0.26	0.733
CG13588-RA	CG42360	-0.57	0.003	0.29	0.179	-0.15	0.795
CG1169-RA	Osi18	-0.58	0.058	-1.17	0.000	0.12	0.864
CG15085-RA	edl	0.50	0.021	-0.54	0.179	-0.35	0.559
CG2682-RA	d4	-0.52	0.043	0.39	0.275	0.10	0.834
CG7220-RB	CG7220	-0.86	0.100	0.61	0.002	-0.28	0.736
CG7498-RA	CG42445	-0.71	0.130	0.58	0.041	0.08	0.929
CG13641-RA	CG13641	0.44	0.063	0.54	0.044	0.07	0.854
CG7597-RB	Cdk12	-0.57	0.045	0.41	0.126	0.12	0.867
CG33336-RA	p53	-0.64	0.007	0.19	0.234	-0.21	0.646
CG16879-RA	CG16879	0.36	0.184	0.49	0.022	0.55	0.010
CG3029-RA	or	-0.30	0.264	0.51	0.035	0.25	0.704
CG8057-RA	alc	-0.57	0.147	0.64	0.041	0.30	0.709

Continued on next page

Table 7 – continued from previous page

		*pros*Gal4 HMG$^-$		*pros*Gal4 EnRep		*pros*Gal4 DWT	
Transcript	**Gene**	**M**	**p**	**M**	**p**	**M**	**p**
CG11242-RA	CG11242	-1.05	0.044	0.83	0.024	0.01	0.989
CG8196-RA	Ance-4	-0.29	0.120	-0.51	0.036	0.38	0.525
CG6547-RA	mRpL37	0.57	0.045	-0.29	0.273	-0.02	0.922
CG8507-RA	CG8507	-1.44	0.023	0.83	0.027	-0.21	0.827
CG5217-RA	CG5217	0.08	0.324	0.21	0.214	0.50	0.047
CG14478-RA	CG14478	-0.59	0.015	0.27	0.330	-0.04	0.954
CG6995-RA	Saf-B	-0.56	0.008	0.19	0.598	-0.07	0.874
CG14290-RB	CG14290	-0.50	0.005	0.30	0.033	0.01	0.982
CG5061-RA	capt	-0.53	0.018	0.48	0.126	0.01	0.993
CG1389-RA	tor	-0.99	0.029	0.81	0.133	0.12	0.902
CG12785-RA	Mat89Ba	-0.58	0.026	0.29	0.321	-0.07	0.874
CG8593-RA	qm	-0.82	0.032	0.69	0.020	0.34	0.763
CG17082-RB	CG17082	-0.80	0.007	0.46	0.244	-0.12	0.773
CG10237-RA	CG10237	-1.07	0.021	0.69	0.022	-0.20	0.796
CG10272-RB	gpp	0.75	0.012	0.05	0.719	0.25	0.607
CG14052-RA	CG14052	0.23	0.465	0.69	0.006	0.46	0.048
CG4676-RA	CG4676	-0.06	0.752	-0.52	0.027	-0.39	0.562
CG1221-RA	miple	0.83	0.031	0.28	0.145	0.51	0.236
CG12290-RA	CG12290	0.32	0.151	0.65	0.010	0.51	0.040
CG3688-RA	l(2)35Bd	-0.61	0.009	0.38	0.256	0.02	0.970
CG9070-RA	Cpr47Eg	-0.61	0.166	-0.35	0.611	-0.85	0.009
CG8460-RA	CG8460	-0.88	0.018	0.60	0.150	-0.24	0.692
CG18348-RA	Cpr67Fb	-1.27	0.006	-0.92	0.040	-0.94	0.078
CG8274-RA	Mtor	-0.53	0.037	0.21	0.514	-0.09	0.807
CG11880-RB	CG11880	-0.67	0.033	0.46	0.090	-0.09	0.904
CG11518-RA	pygo	-0.58	0.022	0.52	0.172	0.04	0.957
CG16785-RA	fz3	0.53	0.008	-0.10	0.569	0.22	0.570
CG15251-RA	CG43386	-0.51	0.156	-0.31	0.644	-0.72	0.046
CG11490-RA	CG11490	-0.79	0.015	0.56	0.077	0.09	0.910
CG32346-RA	E(bx)	0.65	0.002	0.02	0.925	0.28	0.496
CG6298-RA	Jon74E	-1.04	0.004	-1.33	0.000	-1.25	0.002
CG3403-RA	Mob4	-0.06	0.507	-0.69	0.025	-0.64	0.318
CG17490-RB	CG17490	-0.67	0.009	0.42	0.222	0.08	0.841
CG15882-RA	CG15882	0.63	0.016	0.31	0.011	0.51	0.102
CG2054-RA	Cht2	0.63	0.010	0.63	0.013	0.34	0.193
CG9540-RA	Ag5r2	-2.16	0.002	-2.57	0.001	-2.10	0.062
CG31523-RA	CG31523	-0.66	0.017	0.32	0.236	-0.35	0.510
CG1034-RA	bcd	-1.40	0.048	0.67	0.117	-0.23	0.851
CG1966-RA	Acf1	-0.60	0.001	-0.06	0.826	-0.35	0.077
CG11596-RB	CG11596	-0.52	0.007	0.36	0.270	-0.03	0.955
CG9796-RA	CG9796	-0.94	0.033	0.52	0.035	-0.28	0.655
CG40293-RA	Stlk	-0.88	0.010	0.63	0.235	0.01	0.985
CG31673-RA	CG31673	-0.69	0.035	0.45	0.059	-0.06	0.939
CG9617-RA	Sgt1	-0.58	0.041	0.36	0.207	0.15	0.857
CG14215-RA	CG14215	-0.72	0.009	0.13	0.682	-0.42	0.417
CG9233-RA	fu2	-0.59	0.048	0.47	0.129	-0.02	0.969
CG11982-RA	CG11982	-0.84	0.030	0.38	0.099	-0.39	0.578
CG4059-RA	ftz-f1	-1.08	0.039	0.65	0.053	-0.18	0.854
CG14147-RA	CG14147	-1.47	0.049	-1.86	0.022	-1.76	0.073
CG15361-RA	Nplp4	-1.39	0.010	-1.68	0.003	-1.61	0.018
CG6190-RA	Ube3a	-0.52	0.027	0.30	0.220	0.07	0.888
CG8529-RA	Dyb	0.23	0.219	-0.53	0.028	-0.07	0.799
CG8256-RA	Gpo-1	-1.16	0.003	-1.78	0.000	-1.55	0.014

Continued on next page

Table 7 – continued from previous page

		*pros*Gal4 HMG$^-$		*pros*Gal4 EnRep		*pros*Gal4 DWT	
Transcript	**Gene**	**M**	**p**	**M**	**p**	**M**	**p**
CG13316-RB	Mnt	-0.53	0.038	0.37	0.195	-0.02	0.967
CG6859-RA	CG6859	-0.63	0.013	0.33	0.238	0.04	0.937
CG7737-RA	CG7737	-0.92	0.028	0.49	0.069	-0.20	0.800
CG32491-RM	mod(mdg4)	-0.71	0.020	0.37	0.102	-0.13	0.830
CG11061-RA	GM130	-0.54	0.007	0.56	0.089	0.12	0.827
CG31729-RA	CG31729	-0.70	0.018	0.29	0.379	-0.16	0.798
CG14358-RA	CCHa1	0.27	0.037	0.27	0.081	0.50	0.011
CG4583-RA	Ire1	-0.52	0.040	0.18	0.515	0.07	0.918
CG17134-RA	CG17134	-0.22	0.062	-0.58	0.030	-0.30	0.192
CG11205-RB	phr	-0.32	0.055	-1.15	0.022	-0.91	0.161
CG4898-RC	Tm1	-0.55	0.018	0.27	0.397	-0.12	0.867
CG4620-RA	unk	0.50	0.020	0.05	0.842	-0.13	0.740
CG8430-RA	Got1	-0.77	0.040	0.48	0.157	-0.01	0.990
CG4977-RA	kek2	0.51	0.021	0.37	0.009	0.32	0.045
CG7878-RA	CG7878	-0.22	0.008	-0.69	0.022	-0.68	0.057
CG4300-RA	CG4300	-1.12	0.035	0.59	0.079	-0.21	0.837
CG12891-RB	whd	-0.79	0.021	0.56	0.073	0.13	0.886
CG7118-RA	Jon66Ci	-0.72	0.047	-1.28	0.006	-1.32	0.103
CG17168-RA	CG17168	-0.66	0.027	0.32	0.192	-0.24	0.712
CG10390-RA	mia	-0.02	0.763	-0.59	0.030	-0.64	0.253
CG33181-RA	CG33181	-1.22	0.041	0.69	0.034	-0.12	0.910
CG3445-RA	phol	0.14	0.237	-0.73	0.037	-0.20	0.497
CG13067-RA	CG13067	0.07	0.914	0.62	0.600	-0.88	0.006
CG8632-RB	CG8632	-1.00	0.039	0.64	0.172	-0.15	0.884
CG1578-RA	CG1578	-1.06	0.041	0.48	0.077	-0.21	0.852
CG6467-RA	Jon65Aiv	-2.24	0.014	-2.35	0.041	-1.88	0.213
CG31364-RA	l(3)neo38	0.55	0.017	-0.28	0.369	-0.23	0.747
CG9062-RB	CG9062	-0.93	0.031	0.59	0.050	-0.02	0.983
CG13049-RA	CG13049	-0.88	0.083	-1.28	0.011	-0.65	0.486
CG31332-RA	Unc-115b	0.63	0.009	0.32	0.192	0.09	0.871
CG33275-RA	CG33275	-0.50	0.082	0.54	0.026	0.10	0.859
CG2048-RB	dco	-0.59	0.002	0.37	0.271	-0.05	0.928
CG11804-RA	ced-6	-0.90	0.027	0.63	0.184	0.00	0.996
CG17462-RA	Trf4-2	-1.25	0.026	0.64	0.163	-0.11	0.896
CG17260-RA	CG17260	-0.78	0.003	0.38	0.247	0.13	0.806
CG7899-RA	Acph-1	-1.10	0.023	0.78	0.098	-0.01	0.995
CG3532-RA	CG3532	-0.77	0.017	0.53	0.122	-0.13	0.823
CG6425-RA	CG6425	-0.66	0.006	0.49	0.277	-0.25	0.602
CG17985-RA	CG17985	-0.66	0.018	0.50	0.093	0.00	0.999
CG31175-RA	Dys	-0.82	0.012	0.44	0.222	0.16	0.814
CG17818-RA	rdgBbeta	-0.70	0.010	0.50	0.142	-0.26	0.615
CG1832-RA	CG1832	-0.64	0.015	0.57	0.129	-0.23	0.568
CG31992-RD	gw	-0.50	0.019	0.43	0.215	0.13	0.794
CG6391-RA	Aps	-0.67	0.040	0.46	0.091	-0.51	0.176
CG7595-RA	ck	-0.77	0.033	0.52	0.147	0.09	0.872
CG7577-RA	ppk20	0.23	0.066	3.62	0.001	0.26	0.392
CG31322-RA	Aats-met	-0.59	0.001	0.18	0.451	-0.09	0.792
CG5659-RC	ari-1	-0.51	0.001	0.51	0.243	0.18	0.719
CG5385-RA	CG5385	-0.57	0.019	0.45	0.265	-0.23	0.563
CG8948-RC	Graf	-0.50	0.045	0.45	0.257	0.15	0.726
CG7960-RA	Bro	0.78	0.024	0.20	0.418	0.74	0.115
CG10686-RA	tral	-0.74	0.006	0.55	0.208	-0.07	0.905
CG11993-RA	Mst85C	-0.70	0.018	0.39	0.151	-0.20	0.696

Continued on next page

Table 7 – continued from previous page

		*pros*Gal4 HMG$^-$		*pros*Gal4 EnRep		*pros*Gal4 DWT	
Transcript	**Gene**	**M**	**p**	**M**	**p**	**M**	**p**
CG30383-RA	CG30383	0.01	0.945	-0.56	0.007	-0.14	0.663
CG30287-RA	CG30287	-0.18	0.591	-0.50	0.017	-0.44	0.441
CG13778-RA	Mnn1	-0.65	0.031	0.47	0.129	-0.02	0.971
CG8177-RG	CG8177	-1.11	0.036	0.74	0.074	-0.31	0.759
CG13059-RA	CG13059	-0.21	0.784	0.66	0.606	-0.96	0.027
CG7133-RA	CG7133	-0.53	0.014	0.39	0.194	0.21	0.656
CG32564-RA	CG32564	-0.22	0.510	-0.02	0.962	-0.51	0.042
CG3273-RA	sced	-0.85	0.014	0.50	0.158	-0.04	0.962
CG10083-RA	CG10083	-0.77	0.046	0.58	0.099	0.05	0.950
CG5725-RA	fbl	-0.55	0.006	0.34	0.250	-0.07	0.874
CG18030-RA	Jon99Fi	-1.91	0.002	-2.31	0.003	-1.75	0.046
CG30437-RA	CG42345	-0.75	0.011	0.76	0.060	0.33	0.667
CG11254-RC	mael	-0.76	0.047	0.58	0.161	-0.07	0.948
CG30022-RA	CG30022	-0.61	0.035	0.48	0.017	-0.12	0.780
CG10847-RA	enc	-0.70	0.091	0.56	0.002	-0.30	0.687
CG14993-RA	Faa	1.49	0.006	1.36	0.015	1.27	0.111
CG31794-RD	Pax	-0.91	0.026	0.49	0.085	-0.11	0.876
CG7554-RA	comm2	0.32	0.035	-0.52	0.044	-0.33	0.537
CG3167-RA	MAN1	-0.89	0.027	0.44	0.084	-0.33	0.690
CR32162-RA	snRNA:U12:73B	0.36	0.318	0.77	0.022	0.62	0.186
CG6403-RA	CG6403	-0.36	0.063	-0.55	0.011	-0.58	0.027
CG31092-RA	LpR2	-0.71	0.070	0.73	0.048	0.16	0.877
CG5721-RA	CG5721	0.00	0.990	-0.53	0.038	-0.55	0.280
CG7494-RA	mRpL1	-0.53	0.023	0.32	0.094	-0.12	0.826
CG15015-RA	Cip4	-0.56	0.019	0.34	0.255	-0.09	0.894
CG5560-RA	dob	-0.89	0.022	0.48	0.178	-0.23	0.813
CG2221-RA	l(1)G0289	-0.66	0.035	0.41	0.183	-0.10	0.873
CG6580-RA	Jon65Aii	-0.72	0.033	-1.05	0.001	-1.15	0.010
CG14830-RA	CG14830	0.19	0.508	-0.67	0.018	-0.33	0.647
CG6157-RA	dah	-0.88	0.023	0.32	0.454	0.07	0.941
CG6457-RA	yip7	-0.88	0.104	-1.33	0.057	-1.57	0.036
CG3509-RA	CG3509	-0.59	0.335	0.90	0.001	0.34	0.773
CG7614-RA	Mat1	0.09	0.626	-0.54	0.041	-0.50	0.130
CG1242-RA	Hsp83	-0.22	0.489	-0.78	0.124	-0.62	0.020
CG1896-RA	CG1896	-0.79	0.032	0.43	0.344	-0.10	0.894
CG4965-RA	twe	-1.51	0.024	0.76	0.175	-0.36	0.697
CG5026-RB	CG5026	-1.24	0.039	0.68	0.059	-0.21	0.833
CG9258-RA	nrv1	-0.66	0.050	0.68	0.004	0.14	0.826
CG16713-RA	CG16713	-0.34	0.085	-0.79	0.014	-0.64	0.046
CG12141-RB	Aats-lys	-0.85	0.022	0.40	0.298	-0.32	0.724
CG7427-RA	CG7427	-0.59	0.006	0.35	0.396	0.04	0.936
CG32027-RA	CR32027	-0.50	0.045	0.28	0.281	-0.04	0.936
CG8097-RA	CG8097	0.51	0.021	-0.28	0.372	-0.13	0.799
CG12737-RA	Crag	-1.17	0.017	0.53	0.168	-0.13	0.879
CG1058-RA	rpk	-0.54	0.008	0.33	0.211	-0.08	0.881
CG31991-RB	mdy	-0.22	0.028	0.56	0.016	0.24	0.330
CG8108-RA	CG8108	-0.71	0.045	0.30	0.451	-0.21	0.712
CG6932-RA	CSN6	-0.78	0.026	0.42	0.182	-0.47	0.460
CG18177-RA	CG18177	-0.93	0.020	0.61	0.182	-0.44	0.460
CG8986-RA	TwdlBeta	-0.59	0.068	-0.81	0.026	-0.75	0.176
CG5789-RA	CG5789	-1.26	0.029	0.48	0.178	-0.35	0.719
CR33471-RA	Ir48a	-0.02	0.908	-0.55	0.007	-0.26	0.475
CG5629-RA	Ppcs	-0.63	0.026	0.54	0.095	0.32	0.592

Continued on next page

Table 7 – continued from previous page

		*pros*Gal4 HMG$^-$		*pros*Gal4 EnRep		*pros*Gal4 DWT	
Transcript	**Gene**	**M**	**p**	**M**	**p**	**M**	**p**
CR31808-RA	CG31808	-0.31	0.682	-0.80	0.013	-0.99	0.281
CG8144-RC	ps	0.50	0.011	0.17	0.628	-0.08	0.919
CG9320-RA	ns4	-0.68	0.028	0.60	0.121	-0.27	0.568
CG8287-RA	Rab8	-0.80	0.038	0.58	0.187	-0.25	0.708
CG7953-RA	CG7953	-0.80	0.007	-1.12	0.012	-0.92	0.075
CG6889-RA	tara	0.79	0.035	-0.24	0.554	-0.13	0.891
CG2863-RA	Nle	0.11	0.076	-0.55	0.019	-0.47	0.415
CG7146-RA	CG7146	-0.61	0.010	0.51	0.169	-0.06	0.913
CG10859-RA	CG10859	-0.39	0.065	-0.71	0.016	-0.77	0.106
CG4877-RA	CG4877	-0.50	0.002	0.38	0.240	-0.26	0.497
CG8379-RB	CG8379	-0.67	0.031	0.50	0.086	0.06	0.920
CG10360-RA	ref(2)P,ref(2)P	-0.97	0.015	0.45	0.143	-0.27	0.702
CG31776-RA	CG31776	-0.06	0.622	-0.56	0.000	-0.30	0.116
CG2103-RB	pgant6	0.52	0.028	-0.27	0.388	-0.07	0.882
CG14616-RC	l(1)G0196	-0.81	0.139	0.63	0.017	0.14	0.881
CG6081-RA	Cyp28d2	-0.80	0.030	-0.85	0.015	-0.48	0.383
CG32775-RA	GlcAT-I	-0.66	0.021	0.49	0.194	0.03	0.971
CG32077-RA	nol	-0.75	0.016	-1.26	0.003	-0.99	0.020
CG33196-RB	dp	0.60	0.002	0.29	0.171	0.07	0.908
CG15741-RA	CG15741	-0.64	0.076	-0.89	0.016	-1.09	0.143
CG1179-RA	LysB	-0.45	0.089	-0.87	0.035	-0.56	0.311
CG2150-RA	CG2150	-0.26	0.334	-0.36	0.187	-0.67	0.019
CG13460-RA	CG13460	-0.59	0.017	-0.80	0.023	-0.42	0.368
CG9575-RA	Rab35	-0.83	0.018	0.46	0.151	-0.10	0.899
CG33226-RA	CG33226	0.09	0.810	-0.53	0.012	-0.32	0.553
CG15829-RA	CG15829	-0.17	0.329	-0.60	0.046	-0.40	0.323
CG9795-RA	CG9795	-0.72	0.030	0.41	0.241	-0.14	0.835
CG17498-RA	mad2	-0.60	0.009	0.40	0.237	-0.33	0.489
CG14143-RA	CG14143	-0.90	0.046	-1.17	0.006	-1.40	0.023
CG2604-RA	CG2604	-0.52	0.002	0.29	0.447	-0.15	0.774
CG7996-RA	snk	-0.17	0.583	0.50	0.044	0.22	0.729
CG16979-RA	CG16979	0.00	0.964	-0.59	0.049	-0.49	0.327
CG31223-RA	CG31223	-0.18	0.241	-0.21	0.301	-0.52	0.037
CG10492-RA	CG10492	-0.89	0.018	0.40	0.301	-0.15	0.868
CG4404-RA	CG4404	-0.76	0.041	0.43	0.233	-0.49	0.560
CG10241-RA	Cyp6a17	0.49	0.076	0.83	0.014	0.76	0.014
CG2233-RA	CG2233	-0.68	0.085	-1.12	0.000	-1.16	0.065
CG17018-RD	CG17018	-1.23	0.019	0.75	0.085	-0.19	0.850
CG18180-RA	CG18180	-2.20	0.007	-2.85	0.000	-2.48	0.042
CG3766-RA	scat	0.16	0.283	-0.52	0.044	-0.29	0.439
CG12310-RA	CG12310	-1.86	0.002	-2.33	0.002	-2.13	0.023
CG7823-RA	RhoGDI	-0.57	0.015	0.52	0.115	-0.26	0.572
CG9446-RA	coro	-0.67	0.040	0.44	0.163	0.03	0.970
CG10425-RA	CG10425	-0.73	0.047	0.53	0.105	-0.12	0.876
CG2316-RB	CG2316	-0.65	0.025	0.71	0.116	0.18	0.756
CG15172-RA	CG15172	0.36	0.090	0.07	0.849	0.52	0.012
CG13897-RA	CG13897	-0.76	0.044	0.15	0.413	-0.34	0.603
CG13391-RA	Aats-ala	0.12	0.250	-0.52	0.044	-0.36	0.440
CG12877-RA	CG12877	-0.77	0.047	0.38	0.126	-0.16	0.834
CG5029-RA	SamDC	-0.79	0.019	0.53	0.132	-0.24	0.722
CG16926-RA	CG16926	-0.19	0.226	-0.64	0.040	-0.34	0.231
CG15887-RA	CG15887	-0.16	0.390	-0.62	0.004	-0.27	0.497
CG5214-RA	CG5214	-0.38	0.075	0.54	0.021	0.13	0.836

Continued on next page

Table 7 – continued from previous page

		*pros*Gal4 HMG$^-$		*pros*Gal4 EnRep		*pros*Gal4 DWT	
Transcript	**Gene**	**M**	**p**	**M**	**p**	**M**	**p**
CG8790-RB	Dic1	-0.63	0.046	0.31	0.045	-0.38	0.445
CG9734-RA	glob1	-0.95	0.023	0.54	0.209	-0.18	0.816
CG15817-RB	CG15817	-0.55	0.030	0.61	0.259	0.12	0.788
CG14285-RA	CG14285	1.29	0.001	-0.09	0.658	-0.09	0.638
CG11790-RA	CG11790	-0.49	0.154	0.51	0.020	0.02	0.981
CG3107-RB	CG3107	-0.82	0.046	0.61	0.103	0.11	0.905
CG8631-RA	msl-3	-0.62	0.044	0.19	0.375	0.30	0.648
CG13425-RD	bl	-0.98	0.010	-1.05	0.008	-1.42	0.015
CG9388-RA	AP-47	-0.72	0.272	0.75	0.035	0.07	0.952
CG1544-RA	CG1544	0.09	0.391	-0.50	0.014	-0.39	0.310
CG4656-RA	Rassf	-0.76	0.035	0.49	0.209	0.11	0.887
CG31363-RE	Jupiter	-0.36	0.044	0.11	0.607	-0.53	0.034
CG6146-RA	Top1	-0.83	0.046	0.51	0.235	0.06	0.944
CG1417-RC	slgA	-0.27	0.134	0.51	0.013	-0.11	0.779
CG15439-RA	CG15439	-0.53	0.024	0.14	0.362	-0.10	0.789
CG9155-RC	Myo61F	2.26	0.001	2.72	0.000	0.07	0.851
CG12743-RA	otu	-0.16	0.777	0.58	0.015	0.47	0.624
CG17299-RA	SNF4Agamma	-0.92	0.014	0.49	0.285	-0.04	0.948
CG17579-RA	sca	0.60	0.013	-0.26	0.266	-0.55	0.486
CG11140-RB	Aldh-III	-0.79	0.010	0.53	0.231	-0.43	0.406
CG30388-RA	Magi	-0.68	0.033	0.14	0.482	-0.45	0.131
CG5222-RA	CG5222	-0.58	0.007	0.26	0.254	-0.48	0.207
CG1838-RA	myoglianin	-0.83	0.004	0.61	0.245	-0.19	0.722
CG7332-RA	CG7332	-0.58	0.002	0.64	0.203	0.20	0.723
CG1468-RA	CG1468	-0.81	0.007	-1.02	0.004	-0.69	0.065
CG10918-RA	CG10918	-0.78	0.012	-1.12	0.003	-0.90	0.006
CG31633-RA	CG31633	-0.56	0.037	0.46	0.127	0.35	0.474
CG6483-RA	Jon65Aiii	-1.45	0.003	-1.39	0.059	-1.41	0.041
CG32301-RA	CG32301	-0.75	0.054	-0.91	0.006	-0.64	0.066
CG4586-RA	CG4586	-1.07	0.038	0.52	0.181	-0.11	0.885
CG6975-RA	gig	-0.50	0.047	0.44	0.276	-0.06	0.900
CG7041-RA	HP1b	-0.53	0.015	0.36	0.387	-0.27	0.587
CG31173-RA	Gr93c	0.62	0.032	0.25	0.071	0.42	0.132
CG17927-RK	Mhc	-0.44	0.492	-0.11	0.854	-0.97	0.036
CG13425-RC	bl	-1.13	0.025	-1.41	0.008	-1.46	0.063
CG3520-RA	CG3520	-0.75	0.001	-0.15	0.366	-0.51	0.052
CG14897-RA	CG42232	0.61	0.018	0.18	0.502	0.08	0.710
CG7001-RA	Pk17E	-0.60	0.025	0.46	0.173	0.02	0.968
CG14804-RA	Vps26	-0.84	0.030	0.56	0.220	-0.22	0.716
CG5893-RA	D	1.94	0.006	1.10	0.002	1.12	0.274
CG31714-RA	CG31714	0.45	0.038	0.45	0.015	0.59	0.003
CG14352-RA	CG14352	-0.91	0.027	0.67	0.143	-0.42	0.605
CG31893-RA	Peritrophin-15b	-0.38	0.048	-0.61	0.006	-0.29	0.046
CG1483-RB	Map205	-0.50	0.026	0.34	0.201	-0.22	0.592
CG12418-RA	Glut4EF	0.98	0.015	-0.07	0.854	0.07	0.920
CG7456-RA	CG7456	-0.61	0.006	0.46	0.164	-0.03	0.951
CG9195-RA	Scamp	-0.39	0.199	0.63	0.048	0.04	0.957
CG1910-RA	CG1910	0.51	0.001	-0.15	0.404	-0.08	0.865
CG8811-RB	muskelin	-1.16	0.030	0.61	0.170	-0.23	0.798
CG7824-RA	CG7824	-0.56	0.035	0.45	0.230	-0.02	0.979
CG1651-RB	Ank	-0.54	0.031	0.43	0.320	-0.14	0.723
CG1659-RA	unc-119	-0.54	0.030	0.52	0.173	-0.14	0.808
CG8174-RA	SRPK	0.68	0.039	-0.06	0.774	0.05	0.924

Continued on next page

Table 7 – continued from previous page

		*pros*Gal4 HMG$^-$		*pros*Gal4 EnRep		*pros*Gal4 DWT	
Transcript	**Gene**	**M**	**p**	**M**	**p**	**M**	**p**
CG13731-RA	CG13731	-0.36	0.594	0.33	0.801	-1.44	0.021
CG7037-RA	Cbl	-0.89	0.026	0.56	0.132	-0.44	0.542
CG8073-RA	Pmm45A	-0.69	0.045	0.25	0.524	-0.46	0.255
CG5320-RA	Gdh	-0.92	0.019	0.41	0.244	-0.14	0.871
CG6650-RA	CG6650	-0.59	0.001	0.38	0.235	-0.06	0.934
CG13784-RB	CG13784	0.92	0.003	-0.18	0.602	-0.15	0.794
CG7212-RA	cdm	-0.69	0.042	0.34	0.349	-0.01	0.987
CG1903-RA	sno	0.58	0.046	0.07	0.735	0.28	0.408
CG6605-RA	BicD	-0.90	0.002	0.47	0.104	-0.16	0.751
CG14438-RA	CG14438	-0.79	0.026	0.35	0.196	-0.29	0.592
CG4257-RC	Stat92E	-0.54	0.007	0.45	0.185	0.21	0.712
CG30020-RA	CG30020	-0.52	0.026	0.43	0.283	0.03	0.919
CG3502-RA	CG3502	0.43	0.115	-1.27	0.032	-0.96	0.174
CG3508-RB	CG3508	-1.08	0.014	0.54	0.214	-0.11	0.910
CG1091-RB	CG1091	-0.70	0.044	0.40	0.071	-0.14	0.859
CG6509-RB	CG6509	-0.54	0.046	0.38	0.150	0.12	0.840
CG4920-RA	ea	-1.01	0.033	0.57	0.016	-0.23	0.804
CG10808-RA	synaptogyrin	-0.08	0.628	-0.43	0.006	-0.50	0.045
CG1172-RA	CG1172	-0.60	0.026	0.49	0.179	-0.02	0.975
CG6955-RA	Lcp65Ad	-0.34	0.062	-0.60	0.011	-0.60	0.017
CG3726-RA	CG3726	-0.54	0.037	0.59	0.152	0.21	0.812
CG31086-RA	CG31086	-0.59	0.101	-0.57	0.017	-0.59	0.134
CG7156-RA	CG7156	-0.95	0.003	0.44	0.299	-0.19	0.810
CG31373-RA	CG31373	-0.10	0.543	-0.80	0.000	-0.41	0.100
CG30467-RA	CG30467	-0.85	0.002	0.47	0.198	-0.03	0.957
CG10283-RA	CG10283	0.13	0.505	-0.55	0.008	-0.51	0.357
CG1994-RA	l(1)G0020	0.15	0.571	-0.50	0.046	-0.54	0.407
CR30024-RA	tRNA:SeC	0.26	0.489	0.83	0.016	0.65	0.149
CG14829-RA	CG14829	-0.15	0.510	-0.71	0.007	-0.38	0.414
CG31794-RA	Pax	-0.86	0.004	0.52	0.108	0.06	0.925
CG4802-RA	CG4802	-0.55	0.008	0.30	0.241	-0.21	0.657
CG8924-RA	CG8924	-0.93	0.047	0.61	0.082	-0.31	0.753
CG1829-RA	Cyp6v1	-0.64	0.019	0.64	0.108	0.23	0.784
CG9213-RA	CG9213	-0.51	0.012	0.36	0.174	-0.01	0.984
CG8449-RA	CG8449	-0.68	0.017	0.39	0.063	0.02	0.961
CG14072-RA	CG14072	0.39	0.265	0.67	0.007	0.65	0.046
CG10725-RB	CG10725	-0.25	0.268	-0.57	0.028	-0.42	0.125
CG5490-RA	Tl	-0.79	0.038	0.49	0.209	-0.06	0.921
CG4428-RA	Atg4	-0.62	0.049	0.17	0.348	-0.38	0.410
CG12746-RD	CG12746	-0.50	0.039	0.37	0.147	-0.03	0.955
CR31400-RB	Hsromega	0.72	0.012	-0.35	0.289	-0.04	0.928
CG12047-RB	mud	-0.88	0.024	0.54	0.158	-0.10	0.908
CG8839-RC	CG8839	-0.46	0.056	0.53	0.006	0.12	0.725
CG8107-RA	CalpB	-0.75	0.018	0.67	0.085	-0.09	0.889
CG7254-RA	GlyP	-0.89	0.069	0.61	0.038	0.15	0.878
CG11217-RA	CanB2	-0.35	0.264	0.79	0.036	0.31	0.653
CG12129-RA	CG12129	-0.89	0.021	0.53	0.159	-0.10	0.912
CG12559-RD	rl	-1.23	0.012	0.60	0.133	-0.35	0.618
CG6634-RA	mid	-0.24	0.337	-0.78	0.025	-0.20	0.569
CG8983-RA	ERp60	-0.83	0.006	0.59	0.090	-0.04	0.958
CG6258-RA	RfC38	-0.53	0.004	0.25	0.345	-0.00	0.996
CR31696-RA	pncr003:2L	-0.69	0.041	-0.42	0.438	-0.90	0.062
CG10626-RA	Lkr	0.53	0.014	0.10	0.353	0.62	0.020

Continued on next page

Table 7 – continued from previous page

		*pros*Gal4 HMG$^-$		*pros*Gal4 EnRep		*pros*Gal4 DWT	
Transcript	**Gene**	**M**	**p**	**M**	**p**	**M**	**p**
CG8589-RA	tej	-0.78	0.013	0.57	0.188	-0.24	0.782
CG9218-RA	sm	0.81	0.004	0.02	0.944	0.08	0.860
CG7085-RA	l(2)s5379	-0.75	0.014	0.50	0.233	-0.35	0.481
CG9078-RA	ifc	-0.70	0.002	0.50	0.290	-0.22	0.664
CG9272-RA	CG9272	-0.66	0.011	0.70	0.183	-0.03	0.943
CG15104-RA	Topors	-0.67	0.010	0.66	0.166	-0.04	0.935
CG10978-RA	jagn	-0.56	0.016	0.49	0.236	-0.30	0.470
CG7902-RA	bap	0.59	0.137	-0.56	0.018	-0.25	0.426
CG14005-RA	CG14005	-0.71	0.004	0.57	0.297	-0.12	0.787
CG31813-RA	CG31813	-0.55	0.028	-0.11	0.708	-0.13	0.778
CG5876-RA	heix	-0.56	0.029	0.46	0.159	-0.06	0.868
CG8768-RA	CG8768	-0.56	0.021	0.42	0.293	-0.17	0.733
CG17090-RB	hipk	0.10	0.477	0.51	0.030	0.46	0.117
CG3248-RA	Cog3	0.95	0.013	0.18	0.415	0.32	0.446
CG7851-RA	Scgalpha	-0.33	0.134	-0.68	0.018	-0.68	0.191
CG7840-RA	CG7840	-1.04	0.008	0.65	0.131	-0.40	0.578
CG1031-RA	alpha-Est1	-0.94	0.031	0.56	0.049	-0.07	0.928
CG1836-RB	Rad23	-0.60	0.035	0.48	0.195	-0.23	0.690
CG12179-RA	CG12179	-0.33	0.263	0.51	0.015	0.08	0.878
CG8282-RA	Snx6	-0.80	0.038	0.72	0.177	-0.19	0.794
CG7129-RA	l(3)05822	-0.54	0.033	0.40	0.194	-0.10	0.849
CG10255-RA	Lap1	-0.69	0.016	0.50	0.241	-0.22	0.688
CG5037-RA	CG5037	-0.58	0.025	0.48	0.105	-0.30	0.546
CG9755-RC	pum	-0.96	0.027	0.82	0.129	-0.26	0.773
CG11911-RA	CG11911	-0.94	0.036	-1.10	0.034	-1.23	0.027
CG32302-RA	CG32302	-0.51	0.053	-0.69	0.098	-0.92	0.020
CG14265-RB	CG14265	-0.86	0.098	-1.35	0.004	-1.30	0.002
CG6292-RA	CycT	0.55	0.017	-0.24	0.327	-0.24	0.651
CG8609-RA	MED4	-0.57	0.035	0.32	0.175	-0.17	0.783
CG1618-RA	comt	-0.58	0.005	0.30	0.029	-0.26	0.457
CG5991-RA	CG5991	-0.92	0.004	0.30	0.223	-0.45	0.529
CG40300-RA	AGO3	-0.84	0.012	0.63	0.277	0.28	0.679
CG9144-RA	Fbw5	-0.56	0.032	0.33	0.211	-0.06	0.908
CG16704-RA	CG16704	-0.54	0.207	-0.96	0.005	-1.00	0.033
CG15825-RB	fon	-0.84	0.191	-0.67	0.306	-1.29	0.033
CG3082-RC	l(2)k09913	-0.70	0.001	0.18	0.419	-0.31	0.600
CG12182-RA	CG12182	-0.47	0.065	-0.51	0.004	-0.47	0.072
CG2864-RA	Parg	-0.99	0.018	0.55	0.147	-0.15	0.866
CG17494-RA	CG17494	-0.94	0.007	0.54	0.188	-0.22	0.727
CG17331-RA	CG17331	-0.88	0.010	0.42	0.268	-0.38	0.558
CG31510-RA	CG31510	-0.69	0.007	0.38	0.265	0.10	0.848
CG9538-RA	Ag5r	-0.73	0.042	-0.89	0.018	-0.76	0.038
CG10372-RA	Faf	-0.57	0.049	0.40	0.199	-0.00	0.998
CG10295-RA	Pak	-0.58	0.010	0.45	0.181	0.08	0.892
CG4592-RA	CG4592	0.34	0.102	0.56	0.004	0.32	0.008
CG6767-RA	CG6767	-0.83	0.019	0.69	0.050	-0.08	0.917
CG2186-RA	CG2186	-0.69	0.019	0.31	0.315	-0.31	0.616
CG8233-RB	Rcd1	-0.61	0.011	0.44	0.087	-0.03	0.958
CG7730-RC	CG7730	-1.16	0.007	0.75	0.092	-0.19	0.824
CG2229-RA	Jon99Fii	-0.71	0.095	-1.01	0.003	-0.69	0.450
CG8869-RA	Jon25Bii	-1.55	0.002	-2.18	0.000	-2.05	0.026
CG7785-RA	CG7785	-0.44	0.044	0.74	0.044	0.17	0.694
CG6342-RA	Irp-1B	-0.79	0.043	0.61	0.101	-0.25	0.758

Continued on next page

Table 7 – continued from previous page

		*pros*Gal4 HMG$^-$		*pros*Gal4 EnRep		*pros*Gal4 DWT	
Transcript	**Gene**	**M**	**p**	**M**	**p**	**M**	**p**
CG18179-RA	CG18179	-0.50	0.008	-0.72	0.017	-0.59	0.102
CG18681-RA	epsilonTry	-1.65	0.003	-2.12	0.001	-1.94	0.023
CG1619-RB	Hmr	-0.66	0.048	0.33	0.028	-0.03	0.960
CG10328-RA	nonA-l	-0.54	0.026	0.46	0.230	-0.15	0.740
CG11148-RB	CG11148	-0.60	0.029	0.44	0.126	0.15	0.817
CG9127-RA	ade2	-0.51	0.031	0.11	0.374	-0.28	0.535
CG12297-RA	BG4	-0.03	0.745	0.67	0.002	0.66	0.123
CG10343-RA	CG10343	-0.51	0.018	0.32	0.170	0.03	0.957
CG12885-RA	CG12885	-0.74	0.019	-0.95	0.006	-0.89	0.070
CG8301-RA	CG8301	-0.90	0.004	0.51	0.139	-0.07	0.916
CG11567-RA	Cpr	-0.93	0.048	0.50	0.064	-0.31	0.735
CG12140-RA	CG12140	-0.59	0.032	0.27	0.074	-0.31	0.638
CG8726-RA	CG8726	-0.81	0.018	0.57	0.091	0.07	0.928
CG31108-RA	CG31108	-0.91	0.036	0.54	0.082	-0.10	0.901
CG4118-RA	nxf2	-0.68	0.028	0.40	0.151	-0.14	0.816
CG11504-RB	CG11504	-0.53	0.032	0.29	0.233	-0.06	0.921
CG8630-RA	CG8630	-0.42	0.279	-0.22	0.730	-0.84	0.013
CG1636-RA	CG1636	-0.60	0.044	0.41	0.043	-0.11	0.798
CG33320-RA	CheB38a	-1.29	0.038	-1.46	0.007	-1.79	0.145
CG11455-RB	CG11455	0.13	0.437	-0.24	0.226	-0.88	0.025
CG12234-RA	Ranbp21	-0.61	0.012	0.38	0.033	-0.07	0.909
CG9768-RA	hkb	0.35	0.021	-0.84	0.027	-0.29	0.615
CG12428-RC	CG12428	-1.12	0.015	0.52	0.034	-0.23	0.670
CG1981-RA	Thd1	-0.69	0.020	0.66	0.086	0.00	0.998
CG3136-RA	Atf6	-0.58	0.006	0.61	0.153	0.02	0.955
CG1643-RA	Atg5	-0.83	0.010	0.66	0.156	-0.33	0.603
CG41099-RB	CG41099	-0.72	0.010	0.46	0.153	-0.08	0.890
CG2938-RB	CG2938	-1.20	0.014	0.53	0.180	-0.31	0.742
CG6460-RA	TwdlK	-0.95	0.080	-0.71	0.501	-1.26	0.048
CG11614-RA	nkd	-0.04	0.850	-0.81	0.013	-0.89	0.283
CG14222-RA	CG14222	-0.52	0.035	0.36	0.246	-0.16	0.718
CG6099-RA	m4	0.20	0.284	-0.51	0.046	-0.54	0.425
CG10422-RA	bam	-0.57	0.023	0.46	0.177	0.05	0.928
CG9153-RA	CG9153	-0.72	0.049	0.48	0.158	0.03	0.977
CG11771-RA	CG11771	-0.91	0.019	0.66	0.081	-0.14	0.815
CG7433-RA	CG7433	-0.92	0.045	0.80	0.081	-0.04	0.962
CG5608-RA	CG5608	-0.83	0.011	0.57	0.208	0.31	0.565
CG14487-RA	Ir54a	-0.01	0.991	0.52	0.003	0.55	0.167
CG14967-RA	CG14967	-0.65	0.023	0.40	0.180	-0.11	0.822
CG17223-RA	alpha4GT1	-0.65	0.018	0.60	0.131	-0.17	0.756
CG4572-RB	CG4572	-0.56	0.011	0.34	0.158	-0.34	0.344
CG5300-RA	Klp31E	-0.50	0.003	0.35	0.283	-0.06	0.845
CG11305-RA	Sirt7	-0.63	0.031	0.50	0.263	-0.32	0.463
CG5726-RA	CG5726	-0.82	0.021	0.86	0.050	0.01	0.990
CG4749-RA	CG4749	-0.12	0.672	0.62	0.001	0.38	0.094
CG3208-RB	RhoGAP5A	-0.96	0.013	0.57	0.143	-0.09	0.881
CG14629-RA	CG14629	-0.54	0.042	0.28	0.406	-0.75	0.330
CG3268-RA	phtf	-0.61	0.019	0.42	0.215	-0.23	0.618
CG6303-RA	Bruce	0.63	0.002	0.25	0.363	0.48	0.191
CG14997-RB	CG14997	-0.76	0.008	0.43	0.228	-0.26	0.652
CG17633-RA	CG17633	-0.91	0.029	-1.09	0.022	-0.99	0.013
CG6188-RA	CG6188	-0.29	0.319	-0.56	0.002	-0.59	0.261
CG14803-RA	CG14803	-0.56	0.008	0.47	0.200	0.16	0.775

Continued on next page

Table 7 – continued from previous page

		*pros*Gal4 HMG$^-$		*pros*Gal4 EnRep		*pros*Gal4 DWT	
Transcript	**Gene**	**M**	**p**	**M**	**p**	**M**	**p**
CG32656-RA	Muc11A	-0.58	0.037	-0.80	0.009	-0.39	0.003
CG12414-RA	nAcRalpha-80B	-1.00	0.005	0.68	0.196	-0.04	0.959
CG8024-RA	ltd	-1.27	0.038	0.73	0.035	-0.36	0.717
CG7927-RA	CG7927	-0.60	0.039	0.27	0.393	-0.29	0.468
CG40006-RC	CG40006	-0.67	0.019	0.59	0.151	0.12	0.841
CG12467-RA	CG42248	-0.94	0.015	0.60	0.219	0.03	0.964
CG7448-RB	CG7448	-0.08	0.196	-0.53	0.043	-0.08	0.652
CG4665-RA	Dhpr	-0.01	0.969	-0.67	0.045	-0.61	0.154
CG10574-RA	I-2	-0.57	0.020	0.39	0.320	-0.04	0.950
CG5395-RA	nmd	-0.91	0.017	0.51	0.063	-0.28	0.688
CG5084-RA	CG5084	-0.39	0.130	-0.38	0.239	-0.58	0.005
CG7429-RA	CG7429	0.95	0.004	-0.31	0.042	-0.07	0.858
CG30283-RA	CG30283	-0.43	0.094	-0.91	0.005	-0.72	0.166
CG18211-RA	betaTry	-1.71	0.003	-2.11	0.005	-1.92	0.023
CG12116-RA	CG12116	-1.16	0.051	-1.19	0.063	-1.42	0.032
CG9456-RA	Spn1	-0.11	0.161	-0.54	0.012	-0.44	0.028
CG32697-RA	l(1)G0232	-0.95	0.024	0.50	0.204	-0.25	0.775
CG13822-RA	CG13822	0.88	0.056	1.00	0.022	0.19	0.308
CG5439-RA	CG5439	-0.53	0.043	0.35	0.190	-0.15	0.774
CG12690-RA	CHES-1-like	0.63	0.011	0.05	0.858	0.03	0.848
CG6544-RC	fau	-0.05	0.878	-0.19	0.382	-0.55	0.026
CG11306-RA	CG11306	-0.11	0.596	-0.56	0.038	-0.86	0.264
CG3954-RB	csw	-1.10	0.025	0.69	0.143	-0.36	0.661
CG4495-RA	CG4495	-0.71	0.017	0.47	0.116	-0.44	0.376
CG6575-RA	glec	0.60	0.042	0.11	0.672	-0.01	0.984
CG30149-RB	rig	-0.80	0.000	0.45	0.247	-0.02	0.973
CG31220-RA	CG31220	0.61	0.075	0.53	0.003	0.55	0.116
CG12104-RA	CG12104	-0.70	0.100	0.56	0.039	0.05	0.945
CG14879-RA	CG14879	-0.32	0.001	-0.55	0.004	-0.11	0.642
CG10424-RA	CG10424	-0.68	0.040	0.44	0.137	0.16	0.830
CG30005-RA	CG30005	-0.50	0.004	0.53	0.076	0.25	0.721
CG6551-RA	fu	-0.63	0.030	0.39	0.146	-0.19	0.756
CG32267-RA	CG32267	-0.58	0.043	0.67	0.130	-0.23	0.703
CG17888-RC	Pdp1	0.57	0.021	0.43	0.072	0.19	0.003
CG6493-RA	Dcr-2	-0.50	0.011	0.43	0.119	-0.10	0.825
CG10851-RG	B52	0.15	0.336	-0.58	0.018	-0.68	0.399
CG17800-RE	Dscam	0.32	0.107	-0.55	0.030	-0.37	0.498
CG31362-RA	Jon99Ciii	-1.04	0.017	-1.30	0.002	-1.34	0.133
CG1944-RA	Cyp4p2	-1.42	0.009	-1.86	0.005	-1.60	0.044
CG5460-RB	H	-0.59	0.242	0.57	0.048	0.02	0.986
CG17998-RA	Gprk2	-0.54	0.006	0.35	0.237	-0.04	0.946
CG6538-RA	TfIIFbeta	-0.85	0.020	0.44	0.147	-0.23	0.757
CG1444-RA	CG1444	-0.64	0.007	0.45	0.238	0.14	0.859
CG5447-RA	CG5447	-0.53	0.006	0.43	0.294	-0.22	0.654
CG7008-RA	Tudor-SN	-0.50	0.032	0.27	0.140	-0.29	0.492
CG10017-RA	CG34340	-0.33	0.271	-0.57	0.004	-0.60	0.265
CG7537-RB	inx5	-0.12	0.265	-0.59	0.022	-0.45	0.200
CG8448-RC	mrj	-0.78	0.020	0.53	0.078	-0.04	0.964
CG10722-RA	nesd	0.12	0.654	-0.58	0.049	-0.43	0.548
CG5489-RA	Atg7	-0.71	0.027	0.42	0.140	-0.24	0.747
CG8599-RA	Su(var)3-7	-0.77	0.030	0.41	0.068	-0.09	0.912
CG7481-RA	RhoGAP18B	-0.97	0.038	0.53	0.065	-0.37	0.692
CG30325-RA	CG30325	0.22	0.299	0.20	0.314	0.54	0.026

Continued on next page

Table 7 – continued from previous page

		*pros*Gal4 HMG$^-$		*pros*Gal4 EnRep		*pros*Gal4 DWT	
Transcript	**Gene**	**M**	**p**	**M**	**p**	**M**	**p**
CG5978-RA	GAPsec	-0.60	0.008	0.34	0.327	0.15	0.798
CG1667-RA	CG1667	-0.96	0.011	0.63	0.132	0.12	0.838
CG11279-RA	CG11279	0.08	0.752	-0.69	0.015	-0.48	0.339
CG9480-RB	Glycogenin	-0.71	0.008	0.36	0.223	-0.14	0.768
CG5568-RA	CG5568	-1.43	0.025	0.71	0.141	-0.21	0.848
CG6839-RA	CG6839	-0.59	0.057	-0.78	0.019	-0.81	0.078
CG5461-RA	bun	0.51	0.029	-0.15	0.597	-0.03	0.933
CG4362-RA	CG4362	-0.56	0.002	-0.90	0.005	-0.64	0.250
CG6584-RB	SelR	-0.72	0.190	0.52	0.026	0.02	0.987
CG5588-RA	Mtl	-0.79	0.013	0.55	0.021	-0.14	0.797
CG13123-RA	CG13123	0.01	0.940	-0.86	0.028	-0.26	0.334
CG9434-RA	Fst	-0.90	0.037	-1.15	0.036	-1.01	0.104
CG4332-RA	CG4332	-0.83	0.001	0.42	0.235	-0.06	0.937
CG5203-RA	CHIP	-1.17	0.005	0.62	0.230	-0.06	0.938
CG1921-RC	sty	0.73	0.008	-0.05	0.885	0.03	0.966
CG5604-RA	CG5604	-0.77	0.018	0.43	0.135	0.01	0.984
CG32954-RG	Adh	-0.39	0.094	-0.55	0.025	-0.27	0.364
CG8588-RB	pst	-0.58	0.032	0.40	0.080	0.05	0.928
CG9670-RA	fal	-1.27	0.016	0.60	0.114	-0.30	0.720
CG17686-RA	DIP1	-0.16	0.462	-0.54	0.036	-0.58	0.197
CG16762-RA	CG16762	-0.34	0.114	-0.60	0.023	-0.35	0.371
CG9077-RA	Cpr47Ec	-0.75	0.008	-1.12	0.009	-0.99	0.010
CG9242-RA	bur	-0.66	0.018	0.72	0.102	-0.16	0.768
CG17667-RA	CG42709	-0.95	0.014	0.39	0.189	-0.38	0.578
CG8253-RA	tun	-0.86	0.043	0.63	0.083	-0.23	0.772
CG32266-RA	CG32266	-0.31	0.381	0.62	0.547	-0.57	0.029
CG32499-RA	Cda4	-0.70	0.012	0.16	0.750	-0.42	0.403
CG8629-RA	CG8629	-0.60	0.021	-0.53	0.124	-0.48	0.126
CG15236-RB	CG15236	-1.48	0.017	-1.55	0.000	-1.19	0.013
CG6871-RA	Cat	-0.53	0.004	0.17	0.257	-0.10	0.785
CG1803-RB	regucalcin	-0.62	0.107	-0.58	0.009	-0.31	0.475
CG4673-RA	CG4673	0.02	0.879	-0.53	0.037	-0.48	0.325
CG1227-RA	CG1227	-0.58	0.046	0.49	0.146	-0.10	0.864
CG4912-RB	eEF1delta	-0.64	0.033	0.46	0.154	-0.43	0.445
CG6214-RO	MRP	-0.52	0.013	0.22	0.158	-0.08	0.818
CG2257-RA	Ubc-E2H	-0.76	0.136	0.93	0.037	0.05	0.960
CG3796-RA	ac	0.07	0.719	-0.74	0.033	-0.66	0.301
CG9646-RA	CG9646	-0.82	0.015	0.47	0.139	-0.13	0.822
CG5788-RA	UbcD10	-0.75	0.009	0.36	0.297	-0.38	0.536
CG2100-RA	CG2100	-0.57	0.019	0.29	0.480	-0.33	0.429
CG7291-RA	Npc2a	-0.89	0.109	0.74	0.005	-0.27	0.748
CG32684-RB	alpha-Man-I	-0.86	0.016	0.51	0.270	-0.03	0.967
CG32300-RB	oxt	-0.60	0.027	0.60	0.104	0.05	0.950
CG17486-RA	CG17486	-0.53	0.034	0.49	0.092	-0.19	0.696
CG2051-RA	CG2051	-0.76	0.027	0.60	0.089	-0.15	0.803
CG15415-RA	Spindly	0.65	0.009	-0.08	0.817	-0.12	0.771
CG11798-RB	chn	0.51	0.012	-0.32	0.357	-0.19	0.643
CG3045-RA	CG3045	-0.80	0.035	0.74	0.018	0.04	0.962
CG6323-RB	Tsp97E	-0.57	0.039	0.38	0.110	-0.14	0.800
CG1962-RA	CG1962	-1.34	0.015	0.59	0.168	-0.37	0.677
CG3925-RA	CG3925	-0.66	0.022	0.42	0.151	-0.22	0.699
CG12177-RA	CG12177	-0.69	0.014	-1.02	0.006	-1.12	0.016
CG1157-RA	Osi15	-1.27	0.051	-1.49	0.018	-0.97	0.259

Continued on next page

Table 7 – continued from previous page

		*pros*Gal4 HMG$^-$		*pros*Gal4 EnRep		*pros*Gal4 DWT	
Transcript	**Gene**	**M**	**p**	**M**	**p**	**M**	**p**
CG11471-RA	Aats-ile	-0.58	0.003	0.33	0.383	-0.15	0.787
CG12052-RX	lola	-0.84	0.026	0.60	0.134	0.10	0.907
CG31272-RA	CG31272	0.25	0.559	0.65	0.039	0.21	0.583
CG13319-RA	CG13319	-0.53	0.078	0.52	0.031	-0.01	0.987
CG3871-RA	Six4	0.54	0.032	-0.49	0.110	-0.14	0.769
CG32491-RG	mod(mdg4)	-0.12	0.428	0.04	0.680	-0.53	0.025
CG8390-RA	vlc	-0.75	0.034	0.48	0.111	-0.10	0.883
CG32241-RA	CG32241	-0.55	0.028	-0.56	0.054	-0.70	0.102
CG5899-RA	Etl1	-0.57	0.002	-0.59	0.045	-1.00	0.068
CG3707-RA	wapl	-0.80	0.032	0.49	0.152	-0.04	0.962
CG33117-RA	Victoria	0.44	0.011	0.40	0.025	0.57	0.009
CG9096-RC	CycD	-0.78	0.032	0.59	0.051	-0.03	0.964
CG17270-RA	CG17270	-0.96	0.009	0.40	0.181	-0.25	0.773
CG17060-RA	Rab10	-0.66	0.021	0.32	0.413	-0.33	0.621
CG18525-RA	Spn5	-0.83	0.035	0.48	0.046	-0.18	0.846
CG5026-RA	CG5026	-1.15	0.039	0.61	0.126	-0.19	0.854
CG11897-RB	CG11897	-0.89	0.013	0.68	0.150	-0.03	0.966
CG15371-RA	Gr8a	-0.53	0.014	-0.68	0.005	-0.56	0.011
CG8725-RA	CSN4	-0.57	0.012	0.19	0.162	-0.22	0.667
CG2087-RA	PEK	-1.06	0.028	0.50	0.068	-0.17	0.853
CG3359-RD	mfas	0.26	0.182	-0.78	0.016	-0.77	0.237
CG6113-RA	Lip4	0.53	0.004	0.01	0.972	0.39	0.349
CG1594-RA	hop	-0.54	0.022	0.27	0.095	-0.16	0.798
CG8628-RA	CG8628	-0.59	0.001	-0.92	0.014	-0.46	0.415
CG31641-RB	stai	-0.40	0.377	0.64	0.008	0.12	0.883
CG3376-RB	CG3376	-0.85	0.019	0.31	0.097	-0.14	0.866
CG12399-RA	Mad	-0.96	0.023	0.57	0.090	-0.03	0.969
CG4032-RA	Abl	-0.63	0.032	0.48	0.276	0.01	0.992
CG4040-RA	CG42388	0.11	0.376	0.52	0.012	0.56	0.187
CG2083-RA	CG2083	0.70	0.005	-0.25	0.069	0.01	0.990
CG10504-RA	Ilk	-0.63	0.011	0.51	0.098	0.20	0.724
CG2258-RA	CG2258	-1.00	0.045	0.71	0.012	0.19	0.853
CG15168-RA	CG15168	-0.64	0.045	0.34	0.212	-0.06	0.914
CG11984-RB	CG11984	-1.98	0.042	0.66	0.224	-0.44	0.649
CG3632-RC	CG3632	-0.52	0.034	0.52	0.145	0.01	0.987
CG1165-RA	LysS	-1.09	0.007	-1.35	0.009	-1.18	0.100
CG7038-RA	mRpL30	-0.23	0.177	-0.02	0.900	-0.58	0.003
CG3448-RB	CG3448	-0.50	0.029	0.50	0.082	-0.07	0.904
CG14302-RA	CG14302	-0.52	0.047	-0.65	0.038	-0.53	0.162
CG1848-RC	LIMK1	-0.51	0.046	0.48	0.154	-0.28	0.578
CG11350-RB	CG11350	-0.68	0.061	-0.73	0.006	-0.50	0.109
CG10743-RB	Liprin-beta	-0.64	0.026	0.48	0.227	-0.13	0.781
CG6515-RA	Takr86C	0.29	0.321	0.55	0.072	0.62	0.005
CG31876-RA	Cpr30F	-0.47	0.012	-0.65	0.045	-0.48	0.131
CG4780-RA	membrin	0.09	0.706	-0.57	0.042	-0.09	0.320
CG3619-RA	Dl	0.54	0.003	-0.35	0.377	-0.12	0.857
CG8014-RA	Rme-8	-0.73	0.021	0.37	0.253	-0.16	0.804
CG1862-RA	Ephrin	-0.35	0.082	0.65	0.041	0.25	0.596
CG3264-RA	CG3264	-0.59	0.017	-0.68	0.054	-0.36	0.213
CG5072-RA	Cdk4	-0.54	0.010	0.34	0.424	-0.12	0.774
CG8808-RA	Pdk	-1.22	0.048	0.69	0.137	-0.07	0.950
CG12567-RA	CG12567	-1.11	0.020	0.68	0.189	-0.12	0.886
CG12052-RS	lola	0.65	0.003	-0.00	0.984	-0.07	0.889

Continued on next page

Table 7 – continued from previous page

		*pros*Gal4 HMG$^-$		*pros*Gal4 EnRep		*pros*Gal4 DWT	
Transcript	**Gene**	**M**	**p**	**M**	**p**	**M**	**p**
CG32555-RC	RhoGAPp190	0.51	0.017	-0.37	0.177	-0.23	0.635
CG9468-RA	CG9468	-0.40	0.048	-0.65	0.001	-0.20	0.299
CG3008-RA	CG3008	-0.52	0.041	0.42	0.213	-0.00	0.997
CG7197-RA	CG7197	-0.85	0.029	0.46	0.104	-0.17	0.834
CG13947-RA	CG13947	-0.50	0.017	-0.49	0.070	-0.58	0.023
CG16757-RA	Spn	-0.92	0.049	0.58	0.111	-0.05	0.955
CG14230-RA	CG14230	-0.40	0.257	0.54	0.027	0.15	0.844
CG18024-RA	SoxN	0.65	0.002	-0.16	0.696	-0.30	0.695
CG10601-RB	mirr	0.64	0.006	-0.40	0.148	-0.19	0.734
CG14764-RA	CG14764	-1.34	0.039	0.62	0.100	-0.62	0.583
CG14509-RA	CG14509	-0.92	0.002	-0.76	0.023	-0.62	0.107
CG18124-RA	mTTF	-0.92	0.025	0.14	0.226	-0.57	0.396
CG10812-RA	dro5	-0.80	0.045	-0.51	0.265	-1.14	0.087
CG10971-RA	Hip1	-0.65	0.027	0.36	0.007	-0.32	0.527
CG3002-RB	Gga	-1.21	0.102	0.88	0.026	-0.08	0.946
CG17528-RD	CG17528	-0.51	0.033	0.35	0.219	-0.17	0.714
CG3303-RA	CG3303	-0.88	0.005	0.41	0.179	-0.31	0.645
CG31080-RA	TwdlH	-0.88	0.005	-1.14	0.003	-1.15	0.079
CG13454-RA	CG13454	-0.11	0.739	-1.00	0.012	-0.42	0.480
CG8091-RA	Nc	-1.10	0.068	0.58	0.047	-0.24	0.818
CG7487-RA	RecQ4	0.68	0.018	-0.13	0.577	0.08	0.822
CG5760-RA	rtet	-0.89	0.007	0.49	0.127	-0.19	0.801
CG2698-RA	CG2698	-0.53	0.045	0.27	0.319	-0.14	0.777
CG3593-RA	r-l	-0.57	0.043	0.31	0.013	-0.29	0.641
CR33318-RA	CR33318	-0.61	0.014	-0.22	0.377	-0.36	0.011
CG2944-RB	gus	-1.07	0.012	0.52	0.164	-0.31	0.710
CG10630-RA	blanks	0.18	0.148	-0.73	0.032	-0.47	0.344
CG7378-RA	CG7378	0.35	0.357	0.82	0.003	0.65	0.160
CG13461-RA	CG13461	-0.57	0.003	-0.84	0.028	-0.77	0.076
CG31961-RA	CG31961	-0.68	0.044	0.51	0.075	-0.09	0.915
CG12676-RA	ed	0.52	0.043	0.03	0.884	0.31	0.387
CG8598-RA	eco	0.30	0.180	-0.73	0.039	-0.10	0.750
CG12537-RA	rdx	0.53	0.010	-0.30	0.099	-0.03	0.935
CG40129-RA	Gprk1	-0.64	0.005	0.53	0.079	0.19	0.777
CG8090-RA	CG8090	-0.58	0.002	0.29	0.159	0.01	0.986
CG12713-RA	CG12713	-0.75	0.008	0.33	0.211	-0.03	0.973
CG6665-RA	CG6665	-0.62	0.044	0.28	0.190	-0.27	0.618
CG32756-RA	CG32756	0.08	0.762	0.56	0.012	0.91	0.325
CG1877-RC	lin19	-0.73	0.033	0.38	0.251	-0.07	0.929
CG9493-RA	Pez	-0.76	0.022	0.40	0.189	0.02	0.981
CG11999-RA	CG11999	-0.50	0.011	0.18	0.500	-0.13	0.793
CG8871-RA	Jon25Biii	-1.97	0.003	-2.41	0.002	-1.92	0.079
CG6147-RA	Tsc1	-0.78	0.008	0.48	0.251	0.07	0.926
CG9311-RA	mop	-0.93	0.031	0.52	0.134	-0.29	0.663
CG33130-RC	Patronin	-0.70	0.049	0.35	0.345	-0.50	0.049
CG32548-RB	CG32548	-0.77	0.024	-0.76	0.023	-0.41	0.141
CG15101-RA	Jheh1	-0.66	0.037	0.40	0.243	-0.19	0.593
CG8312-RA	CG8312	0.03	0.918	0.75	0.010	-0.45	0.637
CG17149-RA	Su(var)3-3	-0.74	0.043	0.49	0.161	-0.19	0.698
CG4993-RB	PRL-1	-0.67	0.023	0.80	0.178	-0.19	0.748
CG2204-RC	G-oalpha47A	-1.05	0.030	0.66	0.091	0.03	0.933
CG6310-RA	CG6310	-0.53	0.168	0.61	0.044	-0.09	0.728
CG32473-RA	CG32473	-1.38	0.198	0.61	0.030	-0.04	0.960

Continued on next page

Table 7 – continued from previous page

		*pros*Gal4 HMG$^-$		*pros*Gal4 EnRep		*pros*Gal4 DWT	
Transcript	**Gene**	**M**	**p**	**M**	**p**	**M**	**p**
CG11034-RA	CG11034	-0.73	0.038	0.56	0.172	-0.10	0.894
CG12214-RB	CG12214	-0.28	0.035	0.62	0.010	0.36	0.395
CG8470-RA	mRpS30	-0.64	0.019	0.44	0.310	-0.12	0.821
CG3672-RA	Cpr67B	-0.33	0.114	-0.43	0.291	-0.57	0.008
CG8104-RA	nudE	-0.60	0.029	0.57	0.221	-0.18	0.749
CG11811-RA	CG11811	-0.74	0.023	0.44	0.336	-0.19	0.762
CG10413-RA	CG10413	-0.95	0.022	0.63	0.121	0.00	0.997
CG7370-RA	CG7370	0.40	0.117	0.66	0.018	0.59	0.105
CG9705-RA	CG9705	-0.83	0.010	0.54	0.340	-0.35	0.532
CG6072-RA	sra	-0.96	0.025	0.53	0.111	-0.37	0.635
CG11377-RA	CG11377	0.09	0.353	-0.57	0.013	-0.64	0.299
CG11760-RA	CG8116	-0.32	0.234	0.71	0.027	0.08	0.841
CG7361-RA	RFeSP	-0.26	0.321	2.61	0.001	0.00	0.999
CG11370-RA	CG11370	-0.59	0.013	-0.82	0.004	-0.61	0.299
CG17492-RA	mib2	-0.50	0.016	0.27	0.495	0.10	0.848
CG7400-RB	Fatp	0.54	0.009	0.32	0.217	0.49	0.259
CG15524-RA	sas-6	-0.09	0.115	-0.51	0.034	-0.69	0.200
CG12052-RP	lola	-0.40	0.279	0.56	0.027	0.03	0.912
CG9126-RB	Stim	-0.91	0.049	0.59	0.176	-0.31	0.692
CG17715-RE	CG17715	-0.56	0.018	0.19	0.376	-0.44	0.312
CG32365-RA	CG32365	-1.32	0.040	0.59	0.124	-0.36	0.730
CG11055-RC	CG11055	-1.10	0.047	0.63	0.097	-0.18	0.874
CG4698-RA	Wnt4	0.81	0.044	0.18	0.109	0.09	0.633
CG9113-RC	AP-1gamma	-0.51	0.017	0.43	0.139	-0.04	0.941
CG10962-RA	CG10962	-0.84	0.034	0.79	0.066	-0.06	0.947
CG5796-RA	Ppox	-0.79	0.026	0.46	0.193	-0.18	0.798
CG8402-RA	PpD3	-0.66	0.101	0.64	0.025	-0.11	0.886
CG15592-RA	Osi9	-0.82	0.121	-0.91	0.009	-0.39	0.381
CG17163-RA	CG17163	-0.56	0.045	0.26	0.363	0.12	0.837
CG8978-RA	Sop2	-0.15	0.430	-0.39	0.050	-0.58	0.020
CG3819-RA	CG3819	-0.44	0.037	-0.64	0.045	-0.58	0.100
CG5850-RA	CG5850	-0.70	0.029	0.62	0.081	-0.07	0.901
CG33070-RA	Sxl	0.69	0.032	-0.14	0.649	0.37	0.292
CG12142-RA	Tsp42Eg	0.34	0.402	0.59	0.033	0.11	0.695
CG31789-RA	CG31789	-1.29	0.005	-1.34	0.050	-0.77	0.338
CG15296-RA	CG15296	0.24	0.404	0.60	0.010	0.30	0.185
CG15737-RA	wisp	-1.14	0.111	0.79	0.017	0.23	0.855
CG31007-RA	CG31007	0.16	0.157	0.06	0.167	0.50	0.027
CG10383-RA	CG10383	-0.65	0.038	0.35	0.243	-0.03	0.943
CG6105-RA	l(2)06225	-0.52	0.038	0.14	0.620	0.32	0.581
CG15812-RA	pfk	-0.53	0.044	0.18	0.461	0.09	0.875
CG11660-RA	CG11660	-0.87	0.007	0.41	0.248	-0.22	0.703
CG6335-RA	Aats-his	-0.73	0.037	0.28	0.179	0.03	0.959
CG1746-RA	CG1746	-0.58	0.035	0.53	0.525	-0.31	0.299
CG9572-RA	CG9572	-0.11	0.495	0.10	0.109	0.51	0.009
CG6051-RA	CG6051	-0.55	0.017	0.40	0.241	0.06	0.902
CG6542-RA	EDTP	-1.23	0.009	0.67	0.158	-0.37	0.592
CG32140-RA	nuf	-0.34	0.496	0.52	0.040	0.17	0.024
CG3329-RA	Prosbeta2	-0.53	0.095	0.52	0.002	-0.10	0.827
CG6005-RA	CG6005	-0.64	0.047	0.28	0.203	0.03	0.954
CG31320-RA	CG31320	-0.46	0.041	0.58	0.036	0.24	0.593

F *Dichaete* mutant gene expression data

Table 8: Expression data for genes differentially expressed in the *Dichaete* mutant expression study. Invariant normalization was used, and the data was subsequently analysed using limma. The cutoffs used were M value $>= 1$ or $< --1$ and p-value < 0.01

		***Dichaete* mutant**				***Dichaete* mutant**	
Transcript	**Gene**	**M**	**p**	**Transcript**	**Gene**	**M**	**p**
CG4250-RA	CG4250	-1.01	0.001	CG30104-RA	CG30104	-1.27	0.001
CG11928-RA	CG34351	-1.44	0.001	CG5400-RA	Eh	-1.44	0.001
CG10659-RA	CG10659	-1.15	0.001	CG32693-RA	Gr9a	-1.18	0.001
CG4714-RA	CG4714	-1.15	0.001	CG2983-RA	CG2983	-1.05	0.001
CG7452-RA	Syx17	-1.27	0.001	CG17820-RA	fit	-1.11	0.001
CG31773-RA	CG31773	-1.18	0.001	CG14864-RA	CG14864	-1.02	0.001
CG18577-RA	CG18577	-1.23	0.001	CG30029-RA	CR30029	-1.60	0.001
CG14481-RA	CG42561	-1.34	0.001	CG2750-RA	CG2750	-1.01	0.002
CG3040-RA	CG3040	-1.41	0.001	CG3931-RA	Rrp4	-1.29	0.001
CG3713-RA	CG3713	-1.43	0.001	CG7485-RA	Oct-TyrR	-1.14	0.001
CG14316-RA	CG14316	-1.02	0.001	CG14591-RB	CG14591	-1.13	0.001
CG14934-RA	Mal-B1	-1.19	0.002	CG7571-RA	Oatp74D	-1.54	0.001
CG32082-RA	CG32082	-1.33	0.001	CG14499-RA	CR14499	-1.03	0.001
CG10834-RA	CG10834	-1.25	0.001	CG14926-RA	CG14926	-1.62	0.001
CG32238-RA	CG32238	-1.20	0.001	CG17868-RB	Or35a	-1.06	0.001
CG14856-RA	CG14856	-1.62	0.001	CG3644-RB	bic	-1.30	0.001
CG31468-RA	CG31468	-1.44	0.001	CG14973-RA	CG42324	-1.67	0.001
CG6969-RA	CG6969	-1.11	0.001	CG4725-RA	CG4725	-1.41	0.001
CG7399-RA	Hn	-1.40	0.001	CG14691-RB	CG14691	-1.18	0.001
CG14963-RA	CG14963	-1.15	0.001	CG11659-RA	CG11659	-1.58	0.001
CG7533-RC	chrb	-1.25	0.001	CG17525-RA	GstE4	-1.32	0.001
CG12045-RA	Cpr100A	-1.10	0.001	CG3395-RB	RpS9	-1.47	0.001
CG31439-RA	Muc96D	-1.07	0.001	CG10936-RA	CG10936	-1.06	0.001
CG31657-RA	PNUTS	-1.39	0.001	CG1663-RA	CG1663	-1.14	0.001
CG17268-RA	Pros28.1A	-1.23	0.001	CG4836-RC	CG4836	-1.13	0.001
CG13353-RA	CG42287	-1.22	0.001	CG4658-RA	CG4658	-1.18	0.001
CG5388-RA	CG5388	-1.13	0.001	CG5338-RB	RpS19b	-1.00	0.001
CG11337-RB	CG11337	-1.11	0.001	CG17921-RA	HmgZ	-1.95	0.001
CG18317-RA	CG18317	-1.03	0.001	CG33207-RB	pxb	-1.90	0.001
CG1168-RA	7B2	-1.07	0.001	CG31231-RA	CG31231	-1.23	0.001
CG17075-RA	CG17075	-1.17	0.001	CG4753-RA	CG4753	-1.03	0.001
CG2932-RA	Bteb2	-1.28	0.001	CG33114-RA	Gyc32E	-1.19	0.001
CG11293-RA	CG11293	-1.09	0.001	CG14744-RA	CG14744	-1.48	0.001
CG5603-RC	CYLD	-1.00	0.004	CR31356-RA	CR31356	-1.68	0.001
CG10911-RA	CG10911	-1.31	0.001	CG32796-RA	boi	-1.13	0.001
CG13577-RA	CG13577	-1.13	0.001	CG17991-RA	CG17991	-1.30	0.001
CG3221-RA	dgt3	-1.47	0.001	CG32836-RA	NKAIN	-2.09	0.001
CG2102-RA	cas	-1.82	0.001	CG6976-RD	Myo28B1	-1.16	0.001
CG13813-RA	CG13813	-1.01	0.001	CG13434-RA	Nnf1a	-1.42	0.001
CG4314-RA	st	-1.13	0.001	CG17778-RA	CG17778	-1.17	0.001
CG14307-RA	fru	-1.21	0.001	CG1792-RA	CG1792	-1.20	0.001
CG18331-RA	Muc68Ca	-1.31	0.001	CG13437-RA	CG13437	-1.11	0.001
CG5706-RA	CG5706	-1.22	0.001	CG4207-RA	bonsai	-1.37	0.001
CG10211-RA	CG10211	-1.30	0.001	CG17237-RA	CG17237	-1.28	0.001
CG10638-RA	CG10638	-1.11	0.003	CG15000-RA	nab	-1.01	0.001
CG2097-RA	Sym	-1.17	0.001	CG18557-RA	CG18557	-1.26	0.001

Continued on next page

Table 8 – continued from previous page

		Dichaete mutant				*Dichaete* mutant	
Transcript	**Gene**	**M**	**p**	**Transcript**	**Gene**	**M**	**p**
CG11401-RA	Trxr-2	-1.04	0.001	CG10795-RA	CG10795	-1.03	0.001
CG17970-RA	Cyp4ac2	-1.25	0.001	CG3416-RA	Mov34	-1.13	0.001
CG13618-RA	CG13618	-1.09	0.001	CG11527-RA	Tig	-1.46	0.001
CG14731-RA	CG14731	-1.02	0.001	CG31187-RA	CG34384	-1.04	0.001
CG7579-RA	CG7579	-1.32	0.001	CG3253-RA	CG3253	-1.72	0.001
CG4847-RC	CG4847	-1.13	0.001	CG10737-RC	CG10737	-1.14	0.001
CG33300-RA	Muc30E	-1.35	0.001	CG10340-RA	CG10340	-1.61	0.001
CG31266-RA	CG31266	-1.46	0.001	CG14529-RA	CG14529	-1.25	0.001
CG31363-RB	Jupiter	-1.44	0.001	CG32100-RA	CG32100	-1.39	0.001
CG32040-RA	CG32040	-1.28	0.001	CG5661-RA	Sema-5c	-1.26	0.001
CG31516-RA	CG31516	-1.03	0.001	CG14357-RA	CG14357	-1.42	0.001
CG31913-RA	CG31913	-1.17	0.001	CG14843-RA	CG34388	-1.19	0.001
CG7567-RA	CG7567	-1.00	0.008	CG5392-RA	Reck	-1.19	0.001
CG9520-RB	C1GalTA	-1.35	0.003	CG12286-RA	kar	-1.32	0.001
CG8028-RA	CG8028	-1.22	0.001	CG10623-RA	CG10623	-1.91	0.001
CG13031-RA	CG13031	-1.02	0.001	CG4440-RA	CG4440	-1.78	0.001
CG8907-RA	CG8907	-1.28	0.001	CG7992-RA	CG7992	-1.03	0.001
CG5043-RA	CG5043	-1.42	0.001	CG18646-RA	CG42629	-1.26	0.001
CG13870-RA	CG13870	-1.06	0.001	CG32847-RB	CG32847	-1.34	0.001
CG10000-RA	CG10000	-1.11	0.001	CG12486-RA	CG12486	-1.27	0.001
CG7476-RA	mthl7	-1.20	0.001	CG9372-RA	CG9372	-1.31	0.001
CG11997-RA	CG11997	-1.01	0.001	CR33328-RA	CR33328	-1.66	0.001
CG14462-RA	CG34384	-1.00	0.001	CG5423-RA	robo3	-1.10	0.001
CG13455-RA	CG13455	-1.25	0.001	CG8434-RA	lbk	-1.85	0.001
CG5142-RA	CG5142	-1.06	0.001	CG6652-RA	CG6652	-1.08	0.001
CG14031-RA	Cyp4ac3	-1.33	0.001	CG10699-RA	Lim3	-1.50	0.001
CG5122-RA	CG5122	-1.21	0.001	CG1919-RA	Cpr62Bc	-1.17	0.001
CG11147-RB	CG11147	-1.06	0.001	CG8816-RA	Ak6	-1.04	0.001
CG8584-RA	CG8584	-1.15	0.001	CG32922-RA	Skeletor	-1.24	0.001
CG13724-RA	CG13724	-1.14	0.001	CG18110-RA	CG18110	-1.44	0.001
CG7886-RA	CG7886	-1.16	0.001	CG13687-RA	Ptth	-1.20	0.001
CG4324-RA	CG4324	-1.17	0.001	CG16739-RA	CG16739	-1.03	0.001
CG30289-RA	CG30289	-1.23	0.001	CG5925-RA	desat2	-1.17	0.001
CG13784-RA	CG13784	-1.07	0.001	CG13243-RA	CG13243	-1.44	0.001
CG9465-RA	CG9465	-1.35	0.001	CG3948-RA	zetaCOP	-1.64	0.001
CG13063-RA	CG13063	-1.00	0.001	CG13744-RA	CG13744	-1.18	0.001
CG4882-RA	CG4882	-1.16	0.001	CG5557-RA	sqz	-1.21	0.001
CG12768-RA	CG12768	-1.07	0.001	CG32006-RA	CG32006	-1.04	0.003
CG32700-RB	CG32700	-1.66	0.001	CG11301-RA	Mes4	-1.19	0.001
CG8083-RA	CG8083	-1.08	0.003	CG7377-RA	CG7377	-1.62	0.001
CG1265-RB	CG1265	-1.03	0.001	CG8487-RB	garz	-1.59	0.001
CG6789-RA	CG6789	-1.36	0.001	CG10050-RA	CG10050	-1.20	0.001
CG9338-RA	CG9338	-1.21	0.001	CG6656-RA	CG6656	-1.00	0.001
CR30298-RA	CR30298	-1.74	0.001	CG9926-RA	CG9926	-1.03	0.001
CG8850-RB	CG8850	-1.45	0.001	CG12511-RA	CG12511	-1.22	0.001
CG30487-RA	CG30487	-1.17	0.001	CG8421-RA	Asph	-1.56	0.001
CG14430-RA	bou	-1.86	0.001	CG13788-RA	Gr28b	-1.33	0.001
CG18265-RA	CG18265	-1.50	0.002	CG18814-RA	CG18814	-1.04	0.002
CG18765-RA	CG18765	-1.28	0.001	CG9475-RA	Rpt3R	-1.00	0.001
CG31960-RA	CG31960	-1.60	0.001	CG6571-RA	rdgC	-1.12	0.007
CG13661-RA	CG34110	-1.40	0.001	CG5565-RA	CG5565	-1.52	0.001
CG9432-RB	l(2)01289	-1.25	0.001	CG30103-RA	CG30103	-1.07	0.001
CG10130-RA	Sec61beta	-2.16	0.001	CG3680-RA	CG3680	-1.10	0.002

Continued on next page

Table 8 – continued from previous page

		Dichaete mutant				*Dichaete* mutant	
Transcript	**Gene**	**M**	**p**	**Transcript**	**Gene**	**M**	**p**
CG6612-RB	Adk3	-1.02	0.001	CG11312-RA	insc	-1.83	0.001
CG10658-RA	Hf	-1.18	0.001	CG10090-RA	Tim17a1	-1.32	0.001
CG30147-RB	Hil	-1.25	0.001	CG9289-RA	CG9289	-1.24	0.001
CG12643-RA	CG12643	-1.85	0.001	CG11182-RA	PHDP	-1.20	0.001
CG8531-RA	CG8531	-1.35	0.001	CG13953-RA	Camta	-1.18	0.001
CG3959-RA	pelo	-1.15	0.001	CG16781-RA	CG16781	-1.34	0.001
CG7076-RA	Cpr66Cb	-1.21	0.001	CG12030-RA	Gale	-1.77	0.001
CG13672-RA	Dscam4	-1.08	0.001	CG12535-RB	CG12535	-1.11	0.001
CG14249-RB	beat-VII	-1.18	0.001	CG8541-RA	CG8541	-1.41	0.001
CG5747-RA	mfr	-1.31	0.001	CG3996-RA	CG42795	-1.00	0.001
CG9427-RA	CG9427	-1.23	0.001	CG8964-RA	CG8964	-1.01	0.001
CG6527-RA	CG6527	-1.04	0.001	CG4653-RA	CG4653	-1.07	0.001
CG13779-RA	CG13779	-1.51	0.001	CG1322-RB	zfh1	-1.18	0.001
CG5187-RA	Doc2	-1.87	0.001	CG5948-RA	CG5948	-1.36	0.001
CR31273-RA	bxd	-1.13	0.001	CG32944-RB	CG32944	-1.21	0.001
CG6608-RA	Tpc1	-1.02	0.004	CG6582-RA	Aac11	-1.18	0.001
CG12073-RA	5-HT7	-1.04	0.001	CG13942-RA	Shroom	-1.36	0.001
CG8127-RB	Eip75B	-1.63	0.001	CG6698-RA	NtR	-1.33	0.001
CG12763-RA	Dpt	-1.32	0.001	CG12490-RA	CG12490	-1.60	0.001
CG13162-RA	ana3	-1.19	0.001	CG12362-RB	CG12362	-1.24	0.001
CG12892-RA	Caf1-105	-1.12	0.001	CG5269-RA	vib	-1.36	0.001
CG4269-RA	CG4269	-1.32	0.001	CG13477-RA	CG13477	-1.19	0.001
CG5424-RA	f	-1.03	0.002	CG10723-RA	Kua	-1.21	0.001
CG7888-RA	CG7888	-1.12	0.001	CG3032-RA	CG3032	-1.02	0.001
CG14112-RA	SNCF	-1.73	0.002	CG1856-RA	ttk	-1.54	0.001
CR33315-RA	CR33315	-1.12	0.001	CG31869-RA	CG31869	-1.19	0.001
CG4260-RA	alpha-Adaptin	-1.12	0.001	CG5638-RA	Rh7	-1.26	0.001
CG15650-RA	CG15650	-1.06	0.001	CG11608-RA	CG11608	-1.06	0.001
CG30053-RA	CG30053	-1.26	0.001	CG5262-RA	CG5262	-1.13	0.001
CG3629-RA	Dll	-1.31	0.001	CG7514-RA	CG7514	-1.04	0.001
CG14957-RA	CG14957	-1.23	0.001	CG31160-RA	CG34376	-1.21	0.001
CG18445-RA	oys	-1.97	0.001	CG7589-RA	CG7589	-1.28	0.001
CG3671-RC	Mvl	-1.25	0.001	CG10851-RD	B52	-1.24	0.001
CG10823-RA	SIFR	-1.51	0.001	CG7503-RA	Con	-1.40	0.001
CG14982-RB	CG14982	-1.05	0.001	CG11700-RA	CR11700	-1.21	0.001
CG7557-RA	CG7557	-1.07	0.001	CG11000-RA	CG11000	-1.19	0.001
CG31710-RA	CG31710	-1.47	0.001	CG5367-RA	CG5367	-1.16	0.001
CG11437-RA	CG11437	-1.13	0.001	CG7958-RA	tna	-1.64	0.001
CG17941-RA	ds	-1.78	0.001	CG3168-RA	CG3168	-1.09	0.001
CG10914-RA	CG10914	-1.28	0.001	CG5758-RB	CG5758	-1.30	0.001
CG4321-RA	Cyp4d8	-1.35	0.001	CG7068-RA	TepIII	-1.27	0.001
CG33140-RA	CG33140	-1.03	0.001	CG12946-RA	Whamy	-1.09	0.001
CG11619-RA	CG11619	-1.35	0.001	CG5623-RA	CG5623	-1.33	0.001
CG6012-RA	CG6012	-1.22	0.001	CR32525-RA	CR32525	-1.34	0.001
CG33234-RA	CG33234	-1.04	0.001	CG6405-RA	CG6405	-1.24	0.001
CG4817-RA	Ssrp	-1.17	0.001	CG6451-RA	blue	-1.45	0.001
CG32240-RA	CG32240	-1.22	0.001	CG33012-RB	CG33012	-1.35	0.001
CG14905-RA	CG14905	-1.14	0.001	CG2976-RA	CG2976	-1.07	0.001
CG2816-RA	CG2816	-1.58	0.001	CG13702-RB	AlCR2	-1.29	0.001
CG4334-RA	CG4334	-1.11	0.001	CG32946-RA	CG31530	-1.03	0.001
CG17950-RA	HmgD	-1.14	0.001	CG5378-RA	Rpn7	-1.50	0.001
CG17957-RA	Sry-alpha	-1.42	0.002	CG18375-RA	ASPP	-1.14	0.001
CG13466-RA	DCX-EMAP	-1.11	0.001	CG5723-RB	Ten-m	-1.89	0.001

Continued on next page

Table 8 – continued from previous page

		Dichaete mutant				*Dichaete* mutant	
Transcript	**Gene**	**M**	**p**	**Transcript**	**Gene**	**M**	**p**
CG9826-RB	CG9826	-1.26	0.001	CG18538-RA	CG18538	-1.02	0.001
CG4221-RA	CG4221	-1.38	0.001	CG33474-RA	CG33474	-1.61	0.001
CG4779-RA	hgo	-1.14	0.001	CG5772-RA	Sur	-1.21	0.001
CG6706-RB	GABA-B-R2	-1.09	0.001	CG14458-RA	CG33769	-1.14	0.001
CG11404-RA	CG11404	-1.33	0.001	CG15018-RA	CG33514	-1.13	0.001
CG2857-RA	Tpc2	-1.00	0.002	CG10308-RA	CycJ	-1.06	0.001
CG15538-RA	Osi23	-1.00	0.001	CG3548-RA	CG3548	-1.17	0.001
CG32031-RA	Argk	-1.16	0.002	CG17738-RA	CG17738	-1.25	0.001
CG4509-RB	Cad86C	-1.50	0.001	CG14709-RA	Mrp4	-1.54	0.001
CG7994-RA	CG42796	-1.30	0.001	CG33269-RA	CG33269	-1.71	0.001
CG4863-RC	RpL3	-1.51	0.001	CG7548-RA	CG7548	-1.06	0.001
CG10651-RA	CG10651	-1.36	0.001	CG31561-RA	Osi16	-1.21	0.001
CG1946-RA	CG1946	-1.47	0.001	CG10688-RA	CG10688	-1.03	0.001
CG14556-RA	CG14556	-1.09	0.003	CG31805-RA	CG31805	-1.27	0.001
CG32283-RA	dro3	-1.05	0.002	CG5681-RA	CG5681	-1.54	0.001
CG14391-RA	CG14391	-1.09	0.002	CG5155-RA	CG5155	-1.19	0.001
CG4807-RA	ab	-1.90	0.001	CG18469-RA	CG18469	-1.04	0.006
CG11066-RA	scaf	-1.13	0.001	CG14876-RA	CG43318	-1.12	0.001
CG8087-RA	CG8087	-1.40	0.001	CG16890-RA	CG16890	-1.29	0.001
CR7249-RA	Cyp6a16Psi	-1.68	0.001	CG7629-RA	AttD	-1.09	0.001
CG31146-RD	Nlg1	-1.33	0.001	CG5408-RA	trbl	-1.35	0.002
CG8975-RA	RnrS	-1.58	0.002	CG8822-RA	PpD6	-1.22	0.001
CG9547-RA	CG9547	-1.10	0.001	CG12433-RA	CG12433	-1.25	0.001
CG4467-RA	CG4467	-1.14	0.001	CG5060-RA	cic	-1.12	0.001
CG6713-RA	Nos	-1.22	0.001	CG1867-RA	Or98b	-1.47	0.001
CG30270-RA	CG30270	-1.31	0.001	CG3288-RA	CG3288	-1.34	0.001
CG10031-RA	CG10031	-1.11	0.001	CG10718-RA	neb	-1.19	0.001
CG1363-RA	blow	-1.65	0.001	CG12017-RA	CG12017	-1.03	0.001
CR9700-RA	Or85e	-1.44	0.001	CG13490-RA	CG34369	-1.19	0.002
CG4531-RA	aos	-1.13	0.002	CG5528-RA	Toll-9	-1.10	0.001
CG10601-RA	mirr	-1.44	0.001	CG31988-RA	CG31988	-1.03	0.001
CG9224-RA	sog	-1.48	0.001	CG8457-RA	Cyp6t3	-1.10	0.001
CG5169-RA	GckIII	-1.72	0.001	CG6729-RA	CG6729	-1.19	0.001
CG5493-RA	CG5493	-1.32	0.001	CG8605-RA	CG8605	-1.61	0.001
CG9279-RA	CG9279	-1.33	0.001	CG5070-RA	CG5070	-1.32	0.001
CG8100-RA	CG8100	-1.23	0.001	CG9261-RA	nrv2	-1.76	0.001
CG1262-RA	Acp62F	-1.35	0.001	CG4636-RA	SCAR	-1.41	0.001
CG13811-RA	CG42674	-1.05	0.001	CG5546-RA	MED19	-1.58	0.001
CR32665-RB	roX2	-1.07	0.003	CG3973-RA	pigs	-1.08	0.001
CG9460-RA	Spn42De	-1.31	0.001	CG5344-RB	wkd	-1.16	0.001
CG13771-RA	CG13771	-1.13	0.001	CG12841-RA	Tsp42Ek	-1.26	0.001
CG32031-RC	Argk	-1.36	0.001	CG5458-RA	CG5458	-1.06	0.001
CG11321-RB	CG11321	-1.10	0.001	CG8102-RA	CG8102	-1.27	0.001
CG1865-RA	Spn43Ab	-1.00	0.003	CG12724-RA	IP3K2	-1.28	0.001
CG10073-RA	CG10073	-1.33	0.001	CG10120-RB	Men	-1.80	0.001
CG11158-RA	CG11158	-1.19	0.001	CG9361-RA	Task7	-1.21	0.001
CG12496-RA	CG12496	-1.17	0.001	CG9995-RA	htt	-1.25	0.001
CG8167-RA	Ilp2	-1.17	0.001	CR31511-RA	α-Est4aPsi	-1.46	0.001
CG8870-RA	CG8870	-1.02	0.001	CG12581-RA	CG12581	-2.00	0.001
CG5121-RA	MED28	-1.35	0.001	CG8445-RA	calypso	-1.42	0.001
CG14482-RA	CG14482	-1.74	0.001	CG18287-RA	ppk19	-1.07	0.001
CG16825-RA	CG16825	-1.36	0.001	CG32811-RA	CG32811	-1.09	0.002
CG10595-RB	d	-1.54	0.001	CG9626-RA	CG9626	-1.17	0.001

Continued on next page

Table 8 – continued from previous page

		Dichaete mutant				*Dichaete* mutant	
Transcript	**Gene**	**M**	**p**	**Transcript**	**Gene**	**M**	**p**
CG9500-RA	CG9500	-1.13	0.002	CG16876-RA	nimC4	-1.48	0.001
CG32853-RA	mthl12	-1.12	0.003	CG6589-RA	spag4	-1.49	0.001
CG13692-RA	CG13692	-1.30	0.001	CG6898-RA	Zip3	-1.09	0.001
CG5592-RA	CG5592	-1.39	0.001	CG14199-RA	ksh	-1.55	0.001
CG10153-RA	CG10153	-1.24	0.001	CG3999-RA	CG3999	-1.57	0.001
CG10112-RA	Cpr51A	-1.35	0.001	CG10079-RB	Egfr	-1.83	0.001
CG12674-RA	CG12674	-1.06	0.001	CG30118-RA	CG30118	-1.87	0.001
CR14753-RA	Cyp6a15Psi	-1.03	0.001	CG14004-RA	CG34381	-1.43	0.001
CG3984-RA	CG3984	-1.67	0.001	CG16801-RA	Hr51	-1.21	0.002
CG7191-RA	CG7191	-1.01	0.001	CG13713-RA	CG13713	-1.28	0.001
CG14291-RA	CG14291	-1.06	0.008	CG8557-RA	CG8557	-1.60	0.001
CG5837-RA	Hem	-1.49	0.001	CG7299-RA	CG7299	-1.33	0.001
CG8427-RA	SmD3	-1.23	0.001	CG18088-RA	CG18088	-1.24	0.001
CG13290-RA	CG33993	-1.16	0.006	CG6349-RA	DNApol-α180	-1.24	0.001
CG6630-RA	CG42533	-1.13	0.002	CG6601-RA	Rab6	-1.52	0.001
CG8595-RA	Toll-7	-1.16	0.001	CG9770-RA	p	-1.44	0.001
CG7841-RA	CG7841	-1.44	0.001	CG31174-RA	CG31174	-1.19	0.001
CG17697-RB	fz	-1.05	0.001	CG9283-RA	Cpr76Ba	-1.01	0.001
CG17083-RA	CG17083	-1.14	0.001	CG7973-RB	CG33985	-1.41	0.001
CG7670-RA	WRNexo	-1.71	0.001	CG17623-RA	CG42686	-1.00	0.001
CG6174-RA	Arp87C	-1.42	0.001	CG16964-RA	CG16964	-1.00	0.001
CG1675-RA	CG1675	-1.21	0.001	CG9135-RB	CG9135	-1.80	0.001
CG32560-RA	CG42684	-1.04	0.002	CG9347-RA	ninaB	-1.10	0.001
CG7815-RA	ran-like	-1.04	0.001	CG14643-RA	TwdlG	-1.05	0.001
CG11099-RA	CG11099	-1.37	0.001	CG4972-RA	CG4972	-1.46	0.001
CG10475-RA	Jon65Ai	-1.12	0.001	CG18598-RA	CG18598	-1.19	0.001
CG13335-RA	CG42808	-1.47	0.001	CG32985-RA	CG32985	-1.25	0.001
CG15109-RB	CG15109	-1.14	0.001	CG4336-RA	rux	-1.60	0.001
CG13299-RA	CG13299	-1.01	0.001	CG17196-RA	CG17196	-1.06	0.001
CG13310-RA	CG13310	-1.20	0.001	CG11396-RA	CG11396	-1.49	0.001
CG14850-RA	CG14850	-1.28	0.001	CG10529-RA	Lcp65Ae	-1.49	0.001
CG6833-RA	CG6833	-1.09	0.001	CG9894-RA	CG9894	-2.15	0.002
CG15590-RA	Osi5	-1.22	0.001	CG31718-RA	Ir31a	-1.11	0.001
CG17127-RA	CG17127	-1.24	0.001	CG32282-RA	dro4	-1.49	0.001
CG6890-RA	Tollo	-2.06	0.001	CG6800-RA	CG6800	-1.15	0.001
CG3250-RA	Os-C	-1.23	0.001	CG2852-RA	CG2852	-1.80	0.001
CG18788-RA	CG18788	-1.36	0.001	CG13315-RA	CG13315	-1.38	0.001
CG14541-RA	CG42261	-1.21	0.001	CG13260-RA	CG42389	-1.00	0.001
CG8647-RA	mp	-1.06	0.001	CG14015-RA	CG14015	-1.32	0.001
CG4650-RA	CG4650	-1.20	0.001	CG2671-RA	l(2)gl	-1.19	0.001
CG16987-RA	daw	-1.13	0.001	CG8354-RA	m6	-1.19	0.001
CG6883-RA	trh	-1.89	0.001	CG6784-RB	CG42827	-1.13	0.001
CG32483-RA	CG32483	-1.18	0.001	CG7930-RB	TpnC73F	-1.15	0.001
CG2867-RA	Prat	-1.78	0.001	CG17735-RA	CG42574	-1.03	0.005
CG31817-RA	CG31817	-1.07	0.001	CG32198-RB	CG32198	-1.38	0.001
CG4677-RA	lmd	-1.97	0.001	CG12292-RA	spict	-1.30	0.001
CG14011-RA	CG14011	-1.12	0.001	CG5872-RA	CG5872	-1.11	0.001
CG7759-RA	CG7759	-1.31	0.001	CG17562-RA	CG17562	-1.00	0.001
CG6847-RA	CG6847	-1.06	0.002	CG4552-RA	CG4552	-1.06	0.001
CG15904-RA	Ir56d	-1.19	0.001	CG6282-RB	CG6282	-1.24	0.001
CG7148-RA	CG7148	-1.46	0.001	CG1030-RA	Scr	-1.67	0.001
CG3689-RB	CG3689	-1.72	0.001	CG4527-RD	slik	-1.02	0.001
CG2614-RA	CG2614	-1.46	0.001	CG9073-RA	TpnC47D	-1.25	0.001

Continued on next page

Table 8 – continued from previous page

		Dichaete mutant				*Dichaete* mutant	
Transcript	**Gene**	**M**	**p**	**Transcript**	**Gene**	**M**	**p**
CG8967-RA	otk	-1.38	0.001	CG10950-RA	CG10950	-1.12	0.001
CG4956-RA	CG4956	-1.07	0.001	CG11053-RA	CG34345	-1.09	0.001
CG41105-RC	CG34334	-1.10	0.001	CG3306-RA	CG3306	-1.06	0.001
CG8251-RA	Pgi	-1.06	0.001	CG7856-RA	CG7856	-1.08	0.001
CG12523-RA	CG12523	-1.12	0.001	CG3244-RA	Clect27	-1.33	0.001
CG31823-RA	CG31823	-1.40	0.001	CG10369-RA	Irk3	-1.68	0.001
CG11309-RB	CG11309	-1.01	0.001	CG12268-RA	CG12268	-1.25	0.001
CG7890-RA	Hs3st-B	-1.08	0.001	CG8497-RA	Rhp	-1.46	0.001
CG12231-RA	CG12231	-1.18	0.001	CG7847-RA	sr	-1.18	0.001
CG31744-RA	Gr36b	-1.32	0.001	CG3792-RA	CG3792	-1.31	0.001
CG3738-RA	Cks30A	-1.13	0.001	CG18155-RA	CG18155	-1.36	0.001
CG10325-RA	abd-A	-2.01	0.001	CG16916-RA	Rpt3	-1.65	0.001
CG31269-RA	CG31269	-1.16	0.001	CG12919-RA	egr	-1.24	0.001
CG5842-RA	nan	-1.46	0.001	CG2762-RA	ush	-1.47	0.001
CG10177-RA	CG10177	-1.23	0.001	CG33056-RA	CG33056	-1.16	0.001
CG2559-RA	Lsp1alpha	-1.00	0.001	CG6338-RA	Ets97D	-1.30	0.001
CG6738-RA	CG6738	-1.10	0.001	CG16956-RA	CG16956	-1.22	0.001
CG10361-RA	CG10361	-1.26	0.001	CG7656-RA	CG7656	-1.47	0.001
CG13248-RA	CG13248	-1.15	0.001	CG8277-RA	eIF4E-5	-1.43	0.001
CG7251-RA	CG7251	-1.13	0.001	CG9775-RA	CG9775	-1.55	0.001
CG12865-RA	CG12865	-1.12	0.001	CG12857-RA	CG12857	-1.12	0.001
CG8286-RA	P58IPK	-1.67	0.001	CG13970-RA	CG13970	-1.62	0.001
CG15203-RA	CG15203	-1.19	0.001	CG6259-RA	CG6259	-1.10	0.003
CG10993-RA	CG10993	-1.78	0.001	CG31778-RA	CG31778	-1.01	0.001
CG5933-RA	Mta70	-1.61	0.001	CG15874-RA	Pgam5-2	-1.14	0.001
CG4894-RC	Ca-alpha1D	-1.45	0.001	CG11759-RA	Kap3	-1.00	0.003
CG1154-RA	Osi12	-1.25	0.001	CG11882-RA	CG11882	-1.25	0.001
CG3803-RA	CG3803	-1.00	0.003	CG2105-RA	Corin	-1.00	0.002
CG14654-RA	CG14654	-1.18	0.001	CG30084-RA	Zasp52	-1.46	0.003
CG15280-RA	CR15280	-1.16	0.001	CG11866-RA	CG11866	-1.90	0.001
CG10472-RA	CG10472	-1.10	0.001	CG7297-RA	pgant8	-1.22	0.001
CG14536-RA	Herp	-1.38	0.001	CG5064-RA	Srp68	-1.29	0.001
CG13140-RA	dpr19	-1.01	0.001	CG9657-RA	CG9657	-1.28	0.001
CG13076-RB	Notum	-1.71	0.001	CG33351-RA	CheB42b	-1.02	0.001
CG9712-RA	TSG101	-1.36	0.005	CG31845-RA	CR31845	-1.07	0.001
CG10366-RA	CG10366	-1.11	0.002	CG3290-RA	CG3290	-1.16	0.001
CG3349-RA	CG3349	-1.19	0.001	CG11929-RA	CG11929	-1.18	0.001
CG11387-RA	ct	-1.64	0.001	CG14258-RA	CG14258	-1.03	0.001
CG31664-RA	CG31664	-1.13	0.001	CG33203-RC	CG33203	-1.03	0.001
CG32081-RA	CG32081	-1.15	0.001	CG15293-RA	CG15293	-1.28	0.001
CG33478-RA	Or46a	-1.08	0.001	CG1319-RA	CG1319	-1.17	0.001
CG33197-RB	mbl	-1.14	0.001	CG5442-RA	SC35	-1.62	0.001
CG32356-RB	ImpE1	-1.04	0.001	CG14275-RA	CG14275	-2.06	0.001
CG11333-RA	CG11333	-1.03	0.001	CG13037-RA	mRpS34	-1.60	0.001
CG4276-RD	aru	-1.88	0.001	CG6308-RA	CG6308	-1.07	0.001
CG3346-RA	pon	-1.99	0.001	CG31391-RA	CG31391	-1.54	0.001
CG1958-RA	CG1958	-1.26	0.001	CG32096-RB	rols	-1.25	0.001
CG15634-RA	CG15634	-1.89	0.001	CG8613-RA	CG8613	-1.01	0.001
CG5538-RA	CG5538	-1.19	0.001	CG7507-RA	Dhc64C	-1.10	0.001
CG14088-RA	CG14088	-1.13	0.001	CG13196-RA	CG13196	-1.14	0.001
CG7289-RA	CG7289	-1.38	0.001	CG17438-RA	CR17438	-1.05	0.001
CG32052-RA	CG32052	-1.03	0.001	CG9813-RB	CG9813	-1.33	0.001
CG9541-RA	CG9541	-1.11	0.001	CG3573-RA	CG3573	-1.02	0.001

Continued on next page

Table 8 – continued from previous page

		Dichaete mutant				*Dichaete* mutant	
Transcript	**Gene**	**M**	**p**	**Transcript**	**Gene**	**M**	**p**
CG31632-RA	sens-2	-1.12	0.001	CG3474-RA	Cpr35B	-1.48	0.001
CG3129-RA	Rab-RP4	-1.41	0.001	CG15210-RA	CG15210	-1.15	0.001
CG31524-RA	CG31524	-1.06	0.001	CG15427-RE	tutl	-1.22	0.001
CG6447-RA	TwdlL	-1.23	0.001	CG17360-RA	CG17360	-1.14	0.001
CG12172-RA	Spn43Aa	-1.51	0.001	CG6042-RA	Cyp12a4	-1.01	0.001
CG12269-RA	CG12269	-1.11	0.001	CG31406-RA	CG31406	-1.04	0.001
CG7339-RA	CG7339	-1.67	0.001	CG7722-RA	Spn47C	-1.31	0.001
CG3992-RA	srp	-1.51	0.001	CG11419-RA	Apc10	-1.14	0.001
CG13021-RB	CG13021	-1.25	0.001	CG10353-RB	CG10353	-1.27	0.002
CG6333-RA	CG6333	-1.00	0.001	CG1969-RB	CG1969	-1.19	0.001
CG14577-RA	ORMDL	-1.00	0.001	CG15393-RA	CG15393	-1.39	0.001
CG15002-RB	mas	-1.60	0.001	CG1968-RA	CG1968	-1.19	0.001
CG2075-RA	aly	-1.15	0.001	CG6138-RA	CG6138	-1.03	0.001
CG9149-RA	CG9149	-1.29	0.001	CG15366-RA	CG15366	-1.35	0.001
CG9614-RE	pip	-1.54	0.001	CG16978-RA	CG16978	-1.09	0.001
CG10488-RA	eyg	-1.27	0.001	CG6071-RA	CG6071	-1.12	0.001
CG33228-RA	CG33228	-1.15	0.001	CG9376-RA	CG9376	-1.00	0.001
CG8661-RA	CG8661	-1.12	0.001	CG9799-RA	CG9799	-1.75	0.001
CG3090-RA	Sox14	-1.26	0.001	CG9614-RD	pip	-1.24	0.001
CG9097-RA	bab1	-1.03	0.001	CG9876-RA	CG9876	-1.01	0.001
CG17054-RC	Cap-G	-1.20	0.001	CG14216-RA	CG14216	-1.17	0.001
CG3894-RA	CG3894	-1.32	0.001	CG31348-RA	Octbeta3R	-1.08	0.001
CG9565-RA	Nep3	-1.05	0.001	CG32131-RA	CG34400	-1.05	0.001
CG3108-RA	CG3108	-1.08	0.001	CG9254-RA	CG9254	-1.32	0.001
CG33290-RA	CG33290	-1.00	0.001	CG6131-RA	Cpr97Ea	-1.24	0.001
CG30042-RA	Cpr49Ab	-1.04	0.001	CG6140-RA	CG6140	-1.29	0.001
CG4060-RA	TwdlW	-1.03	0.001	CG31021-RA	CG31021	-1.35	0.001
CG15106-RA	Jheh3	-1.14	0.001	CG2666-RB	kkv	-1.20	0.001
CG15405-RA	CG34393	-1.51	0.001	CG11482-RA	Mlh1	-1.41	0.001
CG9134-RA	CG9134	-1.10	0.001	CG8997-RA	CG8997	-1.26	0.001
CG11710-RB	CG11710	-1.07	0.001	CR31930-RA	Gr22d	-1.26	0.001
CG33047-RA	Fuca	-1.20	0.001	CG9633-RA	RpA-70	-1.55	0.002
CG17117-RD	hth	-1.84	0.001	CG11113-RA	CG11113	-1.04	0.001
CG10210-RA	tst	-1.13	0.001	CG16869-RA	Ance-2	-1.34	0.001
CG7123-RA	LanB1	-2.09	0.001	CG10297-RA	Acp65Aa	-1.01	0.001
CG16817-RA	CG16817	-1.61	0.001	CG9451-RA	CG9451	-1.15	0.001
CG9614-RK	pip	-1.48	0.001	CG7833-RA	Orc5	-1.08	0.001
CG17712-RA	CG17712	-1.36	0.001	CG7963-RA	CG7963	-1.33	0.001
CG10300-RA	CG10300	-1.03	0.001	CG12202-RA	Nat1	-1.15	0.003
CG10566-RA	CG10566	-1.00	0.001	CG18673-RA	CG18673	-1.00	0.001
CG13374-RA	pcl	-1.05	0.001	CG15124-RA	CG15124	-1.25	0.001
CG4409-RA	CG4409	-1.24	0.001	CG13317-RA	Ilp7	-1.24	0.001
CG1299-RA	CG1299	-1.03	0.001	CG12846-RA	Tsp42Ed	-1.33	0.001
CG2194-RC	su(r)	-1.19	0.001	CG3725-RH	Ca-P60A	-1.01	0.003
CG11012-RA	Ugt37a1	-1.19	0.001	CG2177-RB	CG2177	-1.40	0.001
CG17183-RA	MED30	-1.52	0.001	CG8333-RA	HLHmγ	-1.08	0.003
CG31738-RB	CG42389	-1.02	0.002	CG13000-RA	CG13000	-1.06	0.001
CG5022-RA	CG5022	-1.15	0.001	CG17193-RA	CG17193	-1.08	0.001
CG6912-RA	CG6912	-1.28	0.001	CG6821-RA	Lsp1γ	-1.22	0.001
CG31202-RA	CG31202	-1.06	0.001	CG15555-RA	CG15555	-1.21	0.001
CG8013-RA	Su(z)12	-1.17	0.001	CG7791-RA	CG7791	-1.25	0.001
CG13287-RA	CG13287	-1.07	0.001	CG8733-RA	Cyp305a1	-1.26	0.001
CG14032-RA	Cyp4ac1	-1.36	0.001	CG33115-RA	nimB4	-1.19	0.002

Continued on next page

Table 8 – continued from previous page

		Dichaete mutant				*Dichaete* mutant	
Transcript	**Gene**	**M**	**p**	**Transcript**	**Gene**	**M**	**p**
CG8404-RA	Sox15	-1.01	0.002	CG14567-RA	CG14567	-1.07	0.001
CG17265-RA	CG17265	-1.53	0.001	CG6231-RA	CG6231	-1.01	0.001
CG7206-RA	CG7206	-1.14	0.001	CG8395-RA	Rrp42	-1.08	0.001
CG32256-RA	Gr64c	-1.26	0.001	CG4697-RA	CSN1a	-1.11	0.001
CG11062-RA	Actbeta	-1.30	0.001	CG12313-RA	ttm2	-1.37	0.001
CG31730-RA	CG31730	-1.43	0.001	CG8246-RA	Poxn	-1.34	0.001
CG16997-RA	Phae2	-1.11	0.001	CG14365-RA	CG34383	-1.18	0.001
CG5911-RB	ETHR	-1.28	0.001	CG7779-RA	Cng	-1.11	0.001
CG8853-RA	CG8853	-1.04	0.001	CG15678-RA	pirk	-1.02	0.001
CG17594-RA	scro	-1.51	0.001	CG1665-RA	CG1665	-1.06	0.001
CG8582-RB	Sh3beta	-1.21	0.001	CG3711-RA	CG3711	-1.08	0.001
CG32531-RA	mRpS14	-1.30	0.001	CG9106-RA	CG9106	-1.04	0.001
CG2663-RA	CG2663	-1.17	0.001	CG9013-RA	Vha16-4	-1.20	0.001
CG10972-RA	ppk12	-1.02	0.001	CG8355-RC	sli	-1.39	0.001
CG4979-RA	sxe2	-1.03	0.001	CG11076-RA	CG11076	-1.55	0.001
CG4143-RB	mbf1	-1.37	0.001	CG8297-RA	CG8297	-1.16	0.002
CG7879-RA	CG7879	-1.24	0.001	CG10436-RA	Pbprp1	-1.08	0.002
CG4661-RA	CG4661	-1.15	0.001	CG11620-RA	CG42399	-1.25	0.001
CG7912-RA	CG7912	-1.00	0.003	CG1225-RA	RhoGEF3	-1.57	0.001
CG15599-RA	CG15599	-1.01	0.001	CG3760-RB	CG3760	-1.68	0.001
CG13415-RA	Cby	-1.15	0.001	CG15262-RA	CG15262	-1.07	0.001
CG8539-RA	CG8539	-1.17	0.001	CG12950-RA	CG12950	-1.22	0.001
CG5860-RA	CG5860	-1.32	0.001	CG18550-RA	yellow-f	-1.25	0.001
CG13922-RA	mRpL46	-1.27	0.001	CG32458-RA	nrm	-1.48	0.001
CG32506-RA	CG32506	-1.02	0.001	CG14064-RA	beat-VI	-1.26	0.001
CG2505-RA	alpha-Est2	-1.26	0.001	CG8230-RA	CG8230	-1.43	0.001
CG12209-RA	CG12209	-1.14	0.001	CG10663-RA	CG10663	-1.03	0.001
CG7678-RA	Vha100-4	-1.20	0.001	CG3242-RA	sob	-1.23	0.001
CG1399-RA	CG1399	-1.32	0.001	CG4629-RA	CG4629	-1.00	0.001
CG15224-RB	CkIIbeta	-1.33	0.001	CG6297-RB	JIL-1	-1.44	0.001
CG11033-RA	Kdm2	-1.09	0.001	CG7921-RA	Mgat2	-1.33	0.001
CG1572-RA	CG1572	-1.00	0.001	CG7163-RA	mkg-p	-1.37	0.001
CG1028-RI	Antp	-1.41	0.001	CG5958-RA	CG5958	-1.91	0.001
CG3868-RA	CG3868	-1.25	0.001	CG11951-RA	CG11951	-1.02	0.001
CG11780-RA	beta4GalT7	-1.09	0.001	CG31622-RC	Gr39a	-1.12	0.001
CG17329-RA	CG17329	-1.23	0.001	CG15311-RA	CG15311	-1.06	0.001
CG1044-RA	dos	-1.27	0.001	CG11895-RA	stan	-1.67	0.001
CG31507-RA	TotZ	-1.05	0.002	CG10498-RB	cdc2c	-1.15	0.001
CG6004-RB	Muc68D	-1.53	0.001	CG15377-RA	Or22c	-1.27	0.001
CG14708-RA	CG14708	-1.25	0.001	CG12156-RA	Rab39	-1.42	0.002
CG15890-RA	CG15890	-1.14	0.001	CG5289-RA	Pros26.4	-1.09	0.001
CG31652-RA	CG34124	-1.07	0.001	CG10541-RA	Tektin-C	-1.30	0.001
CG8732-RA	Acsl	-1.13	0.004	CG13186-RA	CG13186	-1.02	0.001
CG33461-RA	CG33461	-1.35	0.001	CG9726-RA	PH4αMP	-1.29	0.001
CG9739-RB	fz2	-1.78	0.001	CG17230-RA	CG17230	-1.07	0.001
CG10391-RA	Cyp310a1	-1.68	0.001	CG13113-RA	CG13113	-1.27	0.001
CG15564-RA	CG15564	-1.23	0.001	CG3371-RA	ebd1	-1.27	0.001
CG12014-RA	CG12014	-1.15	0.001	CG15819-RA	CG42533	-1.06	0.001
CG14296-RB	endoA	-1.39	0.001	CG11796-RA	CG11796	-1.12	0.001
CG31681-RA	CG31681	-1.43	0.001	CG30285-RA	CG30285	-1.26	0.001
CG15327-RA	Ir7d	-1.11	0.001	CG3446-RA	CG3446	-1.22	0.001
CG3348-RA	CG3348	-1.38	0.001	CG31478-RA	mRpL9	-1.25	0.001
CG3320-RB	Rab1	-1.28	0.001	CG8780-RA	tey	-1.20	0.001

Continued on next page

Table 8 – continued from previous page

		Dichaete mutant				*Dichaete* mutant	
Transcript	**Gene**	**M**	**p**	**Transcript**	**Gene**	**M**	**p**
CG30446-RA	Tdc2	-1.18	0.001	CG4050-RB	CG4050	-1.42	0.001
CG13070-RA	CG13070	-1.26	0.001	CG40198-RA	CG40198	-1.19	0.001
CG2759-RA	w	-1.07	0.001	CG11668-RA	CG11668	-1.13	0.001
CG4550-RA	ninaE	-1.19	0.001	CG33481-RB	dpr7	-1.18	0.001
CG31418-RA	CG31418	-1.00	0.001	CG1978-RA	Or45a	-1.01	0.001
CG15685-RA	Ir92a	-1.24	0.001	CG4058-RA	Nep4	-1.00	0.001
CG32189-RA	CG32189	-1.37	0.001	CG1343-RA	Sp1	-1.28	0.001
CG4133-RA	CG4133	-1.40	0.001	CG32071-RA	CG32071	-1.44	0.001
CG9394-RA	CG9394	-1.00	0.001	CG15792-RA	zip	-1.11	0.001
CG33325-RA	CG33325	-1.24	0.001	CG30416-RA	CG30416	-1.28	0.001
CG6053-RA	CG6053	-1.21	0.001	CG3386-RA	CG3386	-1.08	0.001
CG31002-RA	CG31002	-1.63	0.003	CG12242-RA	GstD5	-1.03	0.001
CG6059-RA	CG6059	-1.05	0.002	CG31431-RA	CG31431	-1.41	0.001
CG2246-RA	CG2246	-1.17	0.001	CG13308-RA	CG13308	-1.10	0.001
CG10379-RA	mbc	-1.79	0.001	CG14837-RA	CG14837	-1.05	0.001
CG32062-RB	A2bp1	-1.64	0.001	CG9606-RA	Rrp45	-1.12	0.001
CG10512-RA	CG10512	-1.82	0.001	CG6104-RA	m2	-1.68	0.001
CG9449-RB	CG9449	-1.03	0.001	CG13211-RA	CG42336	-1.17	0.001
CG2969-RA	Atet	-1.23	0.005	CG9591-RA	omd	-1.40	0.001
CG3214-RA	CG3214	-1.38	0.002	CG31104-RA	CG31104	-1.03	0.003
CG6018-RA	CG6018	-1.06	0.002	CG1212-RA	p130CAS	-1.00	0.002
CG9159-RA	Kr-h2	-1.48	0.001	CG31533-RA	CG31533	-1.00	0.006
CG9938-RA	Ndc80	-1.25	0.001	CG13491-RA	Gr58c	-1.08	0.001
CG5242-RA	mRpL40	-1.21	0.001	CG30092-RD	jbug	-1.57	0.001
CG18548-RA	GstD10	-1.30	0.001	CG5399-RA	CG5399	-1.43	0.002
CG10650-RA	CG10650	-1.00	0.001	CG32659-RA	Ten-a	-1.55	0.001
CG4713-RA	l(2)gd1	-1.00	0.001	CG14369-RA	CG14369	-1.40	0.001
CG7761-RA	pcs	-1.63	0.001	CG13564-RA	CG13564	-1.07	0.001
CG13845-RA	CG34376	-1.53	0.001	CG11100-RA	Mes2	-1.07	0.006
CG12725-RA	CG12725	-1.14	0.001	CG31763-RA	bru-2	-1.23	0.001
CG5611-RA	CG5611	-1.33	0.001	CG14861-RA	CG14861	-1.45	0.001
CG10160-RA	ImpL3	-1.29	0.004	CG14632-RB	CG14632	-1.07	0.001
CG5226-RA	CG43066	-1.23	0.001	CG6536-RB	mthl4	-1.06	0.001
CG14579-RA	CG14579	-1.02	0.001	CG10749-RA	CG10749	-1.08	0.001
CG32281-RA	CG32281	-1.08	0.007	CG5690-RA	CG5690	-1.80	0.001
CG14847-RA	cv-c	-1.15	0.002	CG17754-RD	CG17754	-1.29	0.005
CG14972-RA	CG42324	-1.33	0.001	CG4723-RA	CG4723	-1.10	0.002
CG7367-RA	CG7367	-1.13	0.004	CG11654-RA	Ahcy13	-1.48	0.001
CG30339-RA	CG30339	-1.05	0.001	CG4159-RA	CG4159	-1.48	0.001
CG7532-RA	l(2)34Fc	-1.07	0.001	CG3380-RA	Oatp58Dc	-1.89	0.002
CG14542-RA	vps2	-1.23	0.001	CG17267-RA	CG17267	-1.54	0.001
CG13333-RA	CG13333	-2.18	0.002	CG5369-RA	CG34367	-1.06	0.002
CG11320-RA	CG11320	-1.11	0.001	CG6763-RA	CG6763	-1.22	0.002
CG3234-RC	tim	-1.04	0.001	CG17524-RA	GstE3	-1.80	0.001
CG17054-RB	Cap-G	-1.59	0.001	CG2930-RD	CG2930	-1.22	0.001
CG12481-RA	CG12481	-1.00	0.002	CG3210-RA	Drp1	-1.46	0.001
CG14723-RA	HisCl1	-1.25	0.001	CR31273-RB	bxd	-1.24	0.001
CG4147-RA	Hsc70-3	-1.76	0.001	CG10842-RA	Cyp4p1	-1.09	0.001
CG17988-RA	Ance-3	-1.15	0.004	CG32833-RA	CG32833	-1.03	0.001
CG14889-RA	CG42342	-1.02	0.001	CG2211-RA	CG2211	-1.05	0.001
CG3687-RA	CG3687	-1.08	0.002	CG31045-RC	Mhcl	-1.08	0.002
CG6965-RA	mthl5	-1.20	0.001	CG13766-RA	CG13766	-1.19	0.001
CG17843-RA	CG17843	-1.21	0.001	CG31176-RA	CG31176	-1.17	0.001

Continued on next page

Table 8 – continued from previous page

		Dichaete mutant				*Dichaete* mutant	
Transcript	**Gene**	**M**	**p**	**Transcript**	**Gene**	**M**	**p**
CG17929-RA	CG17929	-1.19	0.001	CG17824-RA	CG17824	-1.19	0.001
CG13432-RA	qsm	-1.31	0.001	CG4206-RA	Mcm3	-1.46	0.001
CG32356-RA	ImpE1	-1.00	0.001	CG33316-RB	sick	-1.04	0.001
CG17839-RA	CG17839	-1.32	0.001	CG11966-RA	CG11966	-1.05	0.001
CG5744-RB	Frq1	-1.09	0.001	CG6401-RA	CG6401	-1.09	0.003
CG1722-RA	CG1722	-1.11	0.001	CG30111-RA	grh	-1.15	0.001
CG11221-RA	CG11221	-1.07	0.001	CG1499-RA	nyo	-1.63	0.001
CG32150-RA	CG32150	-1.64	0.001	CG10122-RA	RpI1	-1.79	0.001
CG15302-RA	Or9a	-1.02	0.001	CG1774-RA	CG1774	-1.20	0.001
CG14680-RA	Cyp12e1	-1.01	0.002	CG7598-RA	CG7598	-1.02	0.001
CG3361-RA	mrt	-1.29	0.004	CG7937-RA	C15	-1.16	0.002
CG30070-RA	CG34443	-1.02	0.001	CG3114-RA	ewg	-1.83	0.001
CG7578-RA	sec71	-1.54	0.001	CG6281-RB	Timp	-1.24	0.001
CG4846-RA	beat-Ia	-1.28	0.001	CG10734-RA	CG10734	-1.00	0.001
CG33299-RA	CG33299	-1.03	0.001	CG17795-RA	mthl2	-1.09	0.001
CG18048-RA	CG18048	-1.22	0.001	CG1480-RA	bnk	-1.47	0.002
CG14528-RA	CG14528	-1.27	0.001	CG5646-RA	CG5646	-1.14	0.001
CG31732-RA	yuri	-1.12	0.001	CG1427-RA	CG1427	-1.15	0.001
CG32091-RB	CG32091	-1.43	0.001	CG15009-RC	ImpL2	-2.30	0.001
CG5660-RA	CG5660	-1.05	0.001	CG14355-RA	CG14355	-1.17	0.001
CG9903-RA	CG9903	-1.04	0.001	CG4835-RA	CG4835	-1.34	0.001
CG4786-RA	Rcd2	-1.17	0.001	CG1840-RA	CG1840	-1.02	0.001
CG14841-RA	CG14841	-1.01	0.001	CR32901-RA	snR38:54Ec	-1.10	0.001
CG7565-RA	CG7565	-1.13	0.001	CG8953-RA	Actn3	-1.06	0.002
CG4386-RA	CG4386	-1.08	0.004	CG9383-RA	asf1	-1.52	0.001
CG5432-RA	CG5432	-1.10	0.001	CG9519-RA	CG9519	-1.14	0.001
CG30354-RA	CG30354	-1.04	0.001	CG4439-RA	S-Lap8	-1.04	0.001
CG13850-RA	CG13850	-1.35	0.001	CG6690-RA	CG6690	-1.21	0.001
CG30190-RA	CG42678	-1.33	0.001	CG6939-RA	Sbf	-1.23	0.004
CG40042-RA	CG40042	-1.96	0.001	CG10689-RA	l(2)37Cb	-1.38	0.001
CG13541-RA	CG13541	-1.12	0.001	CG7462-RB	Ank2	-1.03	0.001
CG5500-RA	CG5500	-1.45	0.001	CG31936-RA	Gr22e	-1.13	0.001
CG9203-RA	CG9203	-1.13	0.001	CG10698-RA	GRHRII	-1.34	0.001
CG5115-RB	mun	-1.19	0.001	CG1916-RA	Wnt2	-1.92	0.001
CG5904-RA	mRpS31	-1.23	0.001	CG9247-RA	CG9247	-1.24	0.001
CG13721-RA	CG13721	-1.08	0.001	CG16738-RA	slp1	-1.38	0.001
CG30288-RA	CG30288	-1.12	0.002	CG1655-RA	sofe	-1.51	0.001
CG18214-RB	trio	-1.49	0.001	CG1873-RA	Ef1α100E	-1.13	0.002
CG9463-RA	CG9463	-1.12	0.001	CG9170-RA	CG9170	-1.45	0.001
CG13737-RB	CG13737	-1.06	0.001	CG12755-RB	l(3)mbn	-1.13	0.001
CG32005-RA	pan	-1.02	0.001	CG5387-RA	Cdk5α	-1.05	0.001
CG7368-RA	CG7368	-1.28	0.001	CG18627-RA	betaggt-II	-1.59	0.001
CG12655-RA	CG12655	-1.04	0.001	CG14068-RA	dpr2	-1.01	0.001
CG13671-RA	CG13671	-1.16	0.001	CG8486-RA	CG8486	-1.15	0.007
CG1004-RA	rho	-1.82	0.002	CG12455-RB	CG42817	-1.31	0.001
CG5147-RA	CG5147	-1.47	0.001	CG6654-RA	CG6654	-1.07	0.001
CG32568-RA	CG32568	-1.40	0.001	CR30087-RA	CG30087	-1.07	0.001
CG30486-RA	CG30486	-1.47	0.001	CG5093-RA	Doc3	-1.85	0.001
CG3917-RA	Grip84	-1.22	0.001	CG32791-RA	CG32791	-1.12	0.001
CG32832-RA	CG32832	-1.14	0.001	CG6570-RA	lbl	-1.31	0.001
CG13659-RA	CG13659	-1.30	0.001	CG6824-RA	ovo	-1.12	0.002
CG5562-RA	gbb	-1.96	0.001	CG9432-RA	l(2)01289	-1.14	0.001
CG8564-RA	CG8564	-1.04	0.001	CG30101-RA	CG30101	-1.17	0.001

Continued on next page

Table 8 – continued from previous page

		Dichaete mutant				*Dichaete* mutant	
Transcript	**Gene**	**M**	**p**	**Transcript**	**Gene**	**M**	**p**
CG32019-RA	bt	-1.52	0.001	CG11310-RA	Cpr78Ca	-1.05	0.001
CG10657-RA	CG10657	-1.26	0.001	CG8858-RA	CG8858	-1.04	0.001
CG12640-RA	CG12640	-1.01	0.001	CG9987-RA	CG9987	-1.20	0.001
CG16777-RA	ps	-1.46	0.001	CG12028-RA	dib	-1.28	0.001
CG13083-RB	CG13083	-1.08	0.001	CG9885-RA	dpp	-1.33	0.001
CG3033-RA	CG3033	-1.06	0.001	CG3897-RB	blot	-1.11	0.005
CG4871-RB	ST6Gal	-1.19	0.001	CG8184-RB	CG8184	-1.38	0.001
CG4646-RA	CG4646	-1.43	0.001	CG10011-RA	CG10011	-1.43	0.001
CG18039-RA	KaiRIA	-1.30	0.001	CG3189-RA	Dpit47	-1.45	0.001
CG5186-RA	slim	-1.68	0.001	CG5586-RB	Tusp	-1.07	0.001
CG31907-RA	CG31907	-1.36	0.001	CR31271-RB	iab-4	-1.08	0.001
CG10612-RA	Or83a	-1.12	0.001	CG6607-RA	CG6607	-1.58	0.001
CG12072-RA	wts	-1.42	0.001	CG8127-RA	Eip75B	-1.67	0.001
CG12479-RA	CG12479	-1.30	0.001	CG11255-RA	CG11255	-1.06	0.002
CG32751-RA	CG32751	-1.21	0.001	CG9509-RA	CG9509	-1.25	0.001
CG4268-RD	Pitslre	-1.10	0.001	CG3222-RA	CG3222	-1.04	0.001
CG14788-RA	ns3	-1.28	0.001	CG3000-RA	rap	-1.37	0.001
CG4733-RA	CG4733	-1.04	0.001	CG7883-RA	eIF2B-α	-1.25	0.001
CG7073-RC	sar1	-1.49	0.001	CG13085-RA	CG13085	-1.07	0.001
CG11453-RA	CG11453	-1.22	0.001	CG18605-RA	CG18605	-1.13	0.001
CG15628-RA	CG15628	-1.30	0.001	CG30048-RA	CG30048	-1.12	0.001
CG30052-RA	Obp49a	-1.03	0.001	CG14659-RA	CG14659	-1.17	0.001
CG5261-RA	CG5261	-1.16	0.001	CG14609-RA	CG14609	-1.15	0.001
CG13627-RA	CG13627	-1.04	0.002	CG14343-RA	CG42329	-1.05	0.001
CG31839-RA	nimB2	-1.08	0.001	CG3505-RA	CG3505	-1.18	0.001
CG31116-RC	ClC-a	-1.65	0.001	CG7588-RA	Eig71Ej	-1.19	0.001
CG15116-RA	CG15116	-1.03	0.001	CG3671-RA	Mvl	-1.14	0.001
CG31543-RA	Hph	-1.05	0.003	CG30078-RA	trpm	-1.07	0.001
CG4743-RA	CG4743	-1.21	0.001	CG10922-RA	La	-1.88	0.001
CG10822-RA	CG10822	-1.28	0.001	CG14981-RB	mge	-1.60	0.001
CG13549-RA	yip3	-1.17	0.001	CG31702-RA	CG31702	-1.29	0.001
CG16993-RA	in	-1.31	0.001	CG4995-RB	CG4995	-1.25	0.001
CG10913-RA	Spn6	-1.43	0.001	CG32376-RA	CG32376	-1.17	0.001
CG33264-RB	Or69a	-1.31	0.001	CG31647-RA	CG31647	-1.27	0.001
CG17352-RC	CG17352	-1.90	0.001	CG31280-RA	Gr94a	-1.12	0.001
CG12926-RA	CG12926	-1.20	0.001	CG6957-RA	Oscillin	-1.41	0.001
CG18507-RB	CG18507	-1.38	0.001	CG10946-RA	dpr14	-1.08	0.001
CG5621-RA	CG5621	-1.29	0.001	CG15085-RA	edl	-2.03	0.002
CG30038-RA	MCPH1	-1.31	0.001	CG10958-RA	CG10958	-1.12	0.001
CG33221-RA	CR33221	-1.16	0.001	CG10861-RA	Atg12	-1.27	0.001
CG6404-RA	CG6404	-1.16	0.002	CG4815-RA	CG4815	-1.16	0.001
CG6450-RC	lva	-1.35	0.001	CG18023-RA	Eip78C	-1.03	0.001
CG3731-RA	CG3731	-1.43	0.002	CG31093-RA	CG31093	-1.19	0.001
CG1391-RB	sol	-1.21	0.002	CG1367-RA	CecA2	-1.04	0.001
CG4330-RA	CG4330	-1.36	0.002	CG14329-RA	CG14329	-1.11	0.001
CG1794-RA	Mmp2	-1.23	0.001	CG3652-RA	CG3652	-1.46	0.001
CG5722-RA	Npc1a	-1.00	0.001	CG9820-RA	Or59a	-1.19	0.001
CG18537-RA	CG18537	-1.25	0.001	CG4220-RC	elB	-1.14	0.001
CG33464-RA	CG33464	-1.38	0.001	CG5326-RA	CG5326	-1.08	0.001
CG4778-RA	obst-B	-1.10	0.001	CG10479-RA	CG10479	-1.94	0.001
CG6705-RA	tsl	-1.31	0.001	CG15017-RA	CG33514	-1.32	0.001
CG2846-RA	CG2846	-1.16	0.001	CG11051-RA	Nplp2	-1.89	0.001
CG1028-RH	Antp	-1.60	0.001	CG32263-RA	CG32263	-1.05	0.001

Continued on next page

Table 8 – continued from previous page

		Dichaete mutant				*Dichaete* mutant	
Transcript	**Gene**	**M**	**p**	**Transcript**	**Gene**	**M**	**p**
CG32023-RA	CG32023	-1.14	0.001	CG17737-RA	CG17737	-1.39	0.001
CG10882-RA	gho	-1.14	0.001	CG7669-RA	CG7669	-1.14	0.001
CG3513-RA	CG3513	-1.02	0.001	CG11249-RA	CG11249	-1.12	0.001
CG5670-RE	Atpalpha	-1.38	0.001	CG32236-RA	CG32236	-1.36	0.001
CG15029-RA	CG15029	-1.05	0.001	CG5680-RB	bsk	-1.71	0.003
CG6459-RA	CG6459	-1.60	0.001	CG1438-RA	Cyp4c3	-1.20	0.001
CG9868-RA	Prosbeta5R	-1.03	0.001	CG1102-RA	MP1	-1.20	0.001
CG32922-RB	Skeletor	-1.30	0.001	CG8085-RB	RN-tre	-1.13	0.001
CG16879-RA	CG16879	-1.27	0.001	CG2955-RA	CG2955	-1.00	0.001
CG5407-RA	Sur-8	-1.02	0.003	CG8821-RB	vis	-1.23	0.001
CG4422-RA	Gdi	-1.67	0.001	CG32692-RA	CG32694	-1.05	0.001
CG1239-RA	CG1239	-1.17	0.001	CG12432-RA	CG12432	-1.03	0.001
CG1897-RA	Dr	-1.50	0.001	CG1066-RB	Shab	-1.25	0.001
CG18302-RA	CG18302	-1.09	0.001	CG5059-RA	CG5059	-1.88	0.001
CG3285-RA	CG3285	-1.24	0.001	CG12502-RA	CG12502	-1.00	0.001
CG10717-RB	ImpL1	-1.10	0.001	CG9397-RA	jing	-1.18	0.001
CG10166-RA	CG10166	-1.58	0.001	CG4527-RA	slik	-1.36	0.002
CG8454-RA	Vps16A	-1.23	0.001	CG5168-RA	CG5168	-1.03	0.001
CG6728-RA	ninaG	-1.18	0.002	CG14095-RA	CG14095	-1.01	0.001
CG1072-RB	Awh	-1.40	0.001	CG1942-RA	CG1942	-1.55	0.001
CG12682-RA	CG12682	-1.02	0.001	CG8603-RA	Shroom	-1.45	0.001
CG32953-RA	CG14929	-1.01	0.001	CG18156-RA	Mis12	-1.67	0.001
CG5028-RA	CG5028	-1.16	0.001	CG9259-RA	CG9259	-1.03	0.001
CG7512-RA	CG7512	-1.19	0.001	CG12624-RA	Ir10a	-1.06	0.001
CG4627-RA	CG4627	-1.05	0.001	CG18250-RC	Dg	-1.03	0.001
CG13810-RA	CG13810	-1.34	0.001	CR32865-RA	CR32865	-1.96	0.001
CG16716-RB	CG16716	-1.12	0.001	CG8196-RA	Ance-4	-1.24	0.002
CG12151-RA	Pdp	-1.18	0.001	CG13801-RA	CG13801	-1.20	0.001
CG5341-RA	sec6	-1.38	0.001	CG7549-RA	mtg	-1.12	0.001
CG6547-RA	mRpL37	-1.66	0.001	CG32029-RA	Cpr66D	-1.09	0.001
CG32717-RC	sdt	-1.87	0.001	CG13868-RA	CG13868	-1.95	0.001
CG13700-RA	CG13700	-1.07	0.001	CG6850-RA	Ugt	-1.47	0.001
CG12493-RA	CG12493	-1.24	0.001	CG18231-RA	CG18231	-1.37	0.001
CG3258-RA	ase	-1.50	0.002	CG9994-RA	Rab9	-1.18	0.001
CG4477-RB	CG4477	-1.25	0.001	CG8166-RA	unc-5	-1.86	0.002
CG8443-RA	clu	-1.00	0.001	CG5107-RA	CG5107	-1.21	0.001
CG1273-RB	CG1273	-1.01	0.001	CG16886-RA	CG16886	-1.23	0.001
CG16820-RA	CG16820	-1.24	0.001	CG14411-RA	CG14411	-1.14	0.003
CG9499-RA	ppk7	-1.04	0.001	CG6618-RB	Patsas	-1.25	0.001
CG6893-RA	CG6893	-1.17	0.001	CG5591-RA	CG5591	-1.35	0.001
CG8114-RA	pbl	-1.79	0.001	CG10109-RA	L	-1.79	0.001
CG30163-RA	Cpr60D	-1.17	0.001	CG9308-RA	CG9308	-1.28	0.001
CG8885-RA	Scox	-1.01	0.004	CG10077-RB	CG10077	-1.65	0.001
CG12673-RA	olf413	-1.29	0.001	CG11219-RA	PIP82	-1.13	0.001
CR12953-RA	Ir51a	-1.19	0.001	CG40053-RA	CG40053	-1.07	0.004
CG4500-RA	CG4500	-2.05	0.001	CG5073-RA	CG5073	-1.56	0.001
CG13279-RA	Cyt-b5-r	-1.15	0.001	CG3295-RA	CG3295	-1.06	0.003
CG6754-RB	nbs	-1.16	0.001	CG12384-RA	CG12384	-1.42	0.003
CG9528-RA	retm	-1.44	0.001	CG9760-RB	CG9760	-1.26	0.001
CG1108-RA	alpha-Est6	-1.45	0.001	CG7973-RA	obst-H	-1.46	0.001
CG33136-RA	CG33136	-1.22	0.001	CG16779-RA	CG16779	-1.04	0.001
CG7100-RE	CadN	-1.43	0.001	CG16963-RA	Cry	-1.13	0.001
CG9125-RA	CG9125	-1.30	0.001	CG9323-RA	CG9323	-1.12	0.002

Continued on next page

Table 8 – continued from previous page

		Dichaete mutant				*Dichaete* mutant	
Transcript	**Gene**	**M**	**p**	**Transcript**	**Gene**	**M**	**p**
CG10327-RB	TBPH	-1.86	0.003	CG7814-RA	CG34133	-1.00	0.003
CG8231-RA	Tcp-1zeta	-1.41	0.002	CG4568-RA	fzo	-1.25	0.001
CG16947-RA	CG16947	-1.13	0.001	CG1333-RA	Ero1L	-1.09	0.001
CG13332-RA	CG13332	-1.05	0.001	CG4329-RA	CG4329	-1.15	0.001
CG6751-RA	CG6751	-1.65	0.001	CG3124-RA	CG3124	-1.18	0.001
CG10078-RB	Prat2	-1.57	0.002	CG13586-RA	itp	-1.41	0.001
CG2157-RA	CG2157	-1.46	0.001	CG7961-RB	alphaCop	-1.56	0.001
CG3851-RA	odd	-1.95	0.001	CG5000-RA	msps	-1.41	0.001
CG6889-RB	tara	-1.80	0.004	CG32508-RB	shakB	-1.23	0.001
CG18768-RD	Ank2	-1.02	0.009	CG13314-RA	CG13314	-1.05	0.001
CG8791-RA	CG8791	-1.09	0.001	CG7768-RA	CG7768	-1.37	0.002
CG14013-RA	CG14013	-1.15	0.001	CG33099-RA	CG33099	-1.82	0.002
CG6878-RA	CG6878	-1.22	0.002	CG10933-RA	CG10933	-1.58	0.004
CG1809-RA	CG1809	-1.15	0.003	CG2865-RA	CG2865	-1.37	0.002
CG17732-RA	CG17732	-1.09	0.001	CG4927-RA	CG4927	-1.11	0.001
CG7189-RA	Gr66a	-1.31	0.001	CG4676-RA	CG4676	-1.33	0.002
CG11043-RA	CG11043	-1.18	0.001	CG1221-RA	miple	-1.54	0.001
CG12290-RA	CG12290	-1.30	0.001	CG14237-RA	CG14237	-1.20	0.006
CG8222-RD	Pvr	-1.53	0.001	CG5871-RA	Oga	-1.10	0.002
CG7758-RA	ppl	-1.27	0.002	CG6845-RA	CG6845	-1.35	0.001
CG10307-RA	CG10307	-1.07	0.003	CG7802-RB	neo	-1.15	0.003
CG15626-RA	CG15626	-1.24	0.003	CG4604-RA	GLaz	-1.11	0.002
CG8965-RA	CG8965	-1.65	0.001	CG9035-RA	Tapdelta	-2.24	0.001
CG11052-RA	CG11052	-1.18	0.001	CG17046-RA	klar	-1.96	0.001
CG12327-RA	Best3	-1.02	0.001	CG8380-RA	DAT	-1.38	0.001
CG4641-RA	CG4641	-1.23	0.001	CG8913-RA	Irc	-1.65	0.001
CG13390-RA	CG13390	-1.15	0.001	CG16998-RA	CG16998	-1.14	0.001
CG17147-RA	CG17147	-1.14	0.001	CG8458-RA	wntD	-1.25	0.002
CG10176-RA	CG10176	-1.29	0.001	CG13654-RA	CG13654	-1.27	0.003
CG12908-RB	Ndg	-1.34	0.001	CG6337-RA	CG6337	-1.03	0.001
CG16954-RB	Hsp60D	-1.18	0.001	CG13139-RA	lft	-1.82	0.001
CG11942-RA	skpE	-1.06	0.001	CG10230-RA	Rpn9	-1.23	0.002
CG7250-RA	Toll-6	-2.14	0.001	CG32404-RA	Cpr65Aw	-1.26	0.001
CG5010-RA	CG5010	-1.33	0.001	CG5793-RA	CG5793	-1.48	0.001
CG10992-RA	CG10992	-1.55	0.001	CG7134-RA	cdc14	-1.17	0.001
CG10261-RA	aPKC	-1.49	0.001	CG31777-RA	CG31777	-1.28	0.001
CG33129-RA	CG33129	-1.22	0.001	CG11909-RA	tobi	-1.35	0.001
CG15006-RA	Cpr64Aa	-1.19	0.001	CG7017-RA	CG7017	-1.17	0.001
CG2104-RA	CG2104	-1.10	0.001	CG14638-RA	TwdlU	-1.12	0.001
CG11865-RA	CG11865	-1.30	0.001	CG7376-RA	CG7376	-1.15	0.002
CG15661-RA	CG15661	-1.04	0.001	CG5063-RA	Trax	-1.05	0.004
CG13124-RB	CG13124	-1.43	0.001	CG1998-RA	CG1998	-1.07	0.001
CG9656-RA	grn	-1.59	0.001	CG31097-RA	CG31097	-1.05	0.001
CG33349-RA	ppk25	-1.17	0.001	CG9631-RA	CG9631	-1.19	0.001
CG16785-RA	fz3	-1.09	0.006	CG6083-RA	CG6083	-1.04	0.001
CG11922-RA	fd96Cb	-1.09	0.001	CG3982-RA	CG3982	-1.00	0.005
CG14254-RA	TwdlC	-1.06	0.001	CG11767-RA	Or24a	-1.14	0.001
CG31662-RA	Gr22a	-1.12	0.001	CG30179-RA	Mlp60A	-1.08	0.001
CG1528-RA	gammaCop	-1.34	0.006	CG13198-RA	CG13198	-1.02	0.001
CG33197-RA	mbl	-1.20	0.001	CG15066-RA	IM23	-1.03	0.001
CG5810-RA	CG5810	-1.21	0.001	CG32346-RA	E(bx)	-1.55	0.001
CG11330-RA	cort	-1.01	0.001	CG30385-RA	CG30385	-1.02	0.002
CG15461-RA	CG15461	-1.21	0.001	CG7330-RA	CG7330	-1.11	0.001

Continued on next page

Table 8 – continued from previous page

		Dichaete mutant				*Dichaete* mutant	
Transcript	**Gene**	**M**	**p**	**Transcript**	**Gene**	**M**	**p**
CG2448-RA	FucT6	-1.40	0.002	CG3403-RA	Mob4	-1.78	0.001
CG32094-RA	CG42255	-1.09	0.001	CG7464-RA	CS-2	-1.38	0.001
CG14087-RA	CG14087	-1.51	0.001	CG17390-RA	Oaz	-1.10	0.001
CG3556-RA	CG3556	-1.01	0.001	CG3460-RA	Nmd3	-1.51	0.003
CG5583-RA	Ets98B	-1.06	0.001	CG3352-RA	ft	-1.86	0.003
CG31014-RA	PH4αSG1	-1.14	0.001	CG3829-RA	CG3829	-1.65	0.001
CG6041-RA	CG6041	-1.21	0.001	CG9746-RA	ird1	-1.21	0.002
CG9701-RA	CG9701	-1.22	0.001	CG32139-RA	Sox21b	-1.07	0.001
CG6027-RA	cdi	-1.22	0.001	CG31405-RA	Gr85a	-1.22	0.001
CG7365-RA	CG7365	-1.15	0.001	CG11418-RA	CG11418	-1.03	0.001
CG6329-RB	CG6329	-1.28	0.001	CG11656-RA	CG11656	-1.03	0.001
CG4035-RE	eIF-4E	-1.23	0.001	CG1968-RB	CG1968	-1.27	0.001
CG1448-RA	inx3	-2.34	0.001	CG15387-RA	CG15387	-1.86	0.001
CG32042-RE	PGRP-LA	-1.22	0.001	CG6356-RA	CG6356	-1.46	0.001
CG9148-RA	scf	-1.73	0.002	CG15438-RA	CG15438	-1.18	0.002
CG9614-RA	pip	-1.32	0.001	CG10481-RA	CG10481	-1.08	0.001
CG6069-RA	CG34129	-1.22	0.001	CG9494-RA	Tsp29Fa	-1.17	0.001
CG33227-RB	CG34370	-1.16	0.001	CG11821-RA	Cyp12a5	-1.19	0.001
CG15427-RD	tutl	-1.33	0.001	CG8660-RC	Fibp	-1.04	0.001
CG13190-RA	cuff	-1.42	0.001	CG9614-RC	pip	-1.21	0.001
CG6421-RA	CG6421	-1.26	0.001	CG8673-RA	CG8673	-1.00	0.001
CG33098-RB	CG33098	-1.16	0.001	CG33339-RA	CG33339	-1.10	0.001
CG9564-RA	Try29F	-1.21	0.001	CG3212-RA	Sr-CIV	-1.27	0.001
CG31089-RA	CG31089	-1.38	0.001	CG33287-RA	CG33287	-1.27	0.001
CG6034-RA	CG6034	-1.16	0.001	CG13025-RA	CG13025	-1.45	0.001
CG14147-RA	CG14147	-1.06	0.001	CG14815-RA	CG14815	-1.04	0.001
CG31020-RA	spdo	-1.15	0.001	CG10405-RA	CG10405	-1.05	0.001
CG15025-RA	CG15025	-1.14	0.001	CG15403-RA	CG15403	-1.32	0.001
CG17334-RA	lin-28	-1.38	0.001	CG15902-RA	Ugt86Dj	-1.16	0.001
CG15361-RA	Nplp4	-1.13	0.001	CG9911-RA	CG9911	-1.40	0.001
CR31273-RD	bxd	-1.03	0.001	CG8141-RA	CG8141	-1.15	0.001
CG33045-RG	Kaz1-ORFB	-1.26	0.001	CG17721-RA	CG17721	-1.59	0.001
CG9614-RJ	pip	-1.14	0.001	CG17117-RA	hth	-2.18	0.001
CG31161-RA	CG31161	-1.29	0.001	CG7122-RA	RhoGAP16F	-1.22	0.003
CG16815-RA	CG16815	-1.06	0.001	CG31225-RA	Ir94f	-1.04	0.001
CG9422-RA	CG9422	-1.01	0.001	CG10382-RA	wrapper	-1.13	0.001
CG31766-RB	Ance-3	-1.37	0.001	CG4596-RA	CG4596	-1.14	0.001
CG12201-RA	CG12201	-1.36	0.001	CG2835-RB	G-sα60A	-1.87	0.001
CG17063-RA	inx6	-1.10	0.001	CG10561-RA	CG10561	-1.01	0.001
CG18672-RA	CG18672	-1.17	0.001	CG15122-RA	Ir56c	-1.07	0.001
CG14077-RA	CG14077	-1.02	0.001	CG17212-RA	rho-6	-1.06	0.001
CG13322-RB	CG13322	-1.45	0.003	CG2191-RA	Smvt	-1.35	0.002
CG15735-RA	CG15735	-1.02	0.004	CG17180-RA	CG17180	-1.07	0.001
CG15605-RA	CG15605	-1.27	0.001	CG3879-RA	Mdr49	-1.18	0.001
CG5021-RA	CG5021	-1.60	0.001	CG17191-RA	CG17191	-1.19	0.001
CG3239-RA	CG3239	-1.21	0.001	CG6203-RA	Fmr1	-1.45	0.001
CG13421-RA	Obp57c	-1.18	0.001	CG2239-RC	jdp	-1.39	0.001
CG15553-RA	CG15553	-1.13	0.001	CG13008-RA	CG13008	-1.29	0.001
CG14618-RA	CG14618	-1.02	0.002	CG13285-RA	CG13285	-1.30	0.001
CG17224-RA	CG17224	-1.19	0.002	CG33113-RG	Rtnl1	-1.07	0.001
CG6806-RA	Lsp2	-1.15	0.001	CG10960-RA	CG10960	-1.00	0.001
CG32425-RB	CG32425	-1.20	0.001	CG7264-RA	CG7264	-1.05	0.001
CG12837-RA	Tsp42Er	-1.05	0.001	CG4947-RA	Tgt	-1.01	0.001

Continued on next page

Table 8 – continued from previous page

		Dichaete mutant				*Dichaete* mutant	
Transcript	**Gene**	**M**	**p**	**Transcript**	**Gene**	**M**	**p**
CG12241-RA	CG12241	-1.42	0.001	CG12311-RA	tw	-1.41	0.001
CG8245-RA	CG8245	-1.01	0.001	CG2522-RA	Gtp-bp	-1.30	0.001
CG32410-RA	CG33523	-1.14	0.001	CG5909-RA	CG5909	-1.21	0.001
CG6868-RA	tld	-1.41	0.003	CG1650-RA	unpg	-1.20	0.001
CG6304-RA	CG6304	-1.30	0.001	CG8577-RA	PGRP-SC1b	-1.28	0.001
CG17134-RA	CG17134	-1.20	0.001	CG10335-RA	Pbgs	-1.40	0.001
CG11205-RB	phr	-1.32	0.001	CG4582-RA	CG4582	-1.29	0.001
CG9449-RA	CG9449	-1.60	0.001	CG4620-RA	unk	-1.40	0.001
CG9160-RA	mtacp1	-1.07	0.009	CG10108-RA	phyl	-1.69	0.003
CG33121-RA	Spn28Db	-1.07	0.001	CG10970-RA	CG10970	-1.30	0.001
CG8353-RA	CG8353	-1.50	0.001	CG4977-RA	kek2	-1.26	0.001
CG5803-RA	Fas3	-1.89	0.001	CG15115-RA	CG15115	-1.10	0.001
CG17662-RA	CG17662	-1.03	0.001	CG33093-RA	CG33093	-1.08	0.001
CG31848-RA	CG31848	-1.27	0.001	CG11498-RA	CG11498	-1.22	0.001
CG7911-RA	CG7911	-1.78	0.002	CG31761-RA	bru-2	-1.37	0.001
CG3815-RA	CG3815	-1.10	0.001	CG3758-RA	esg	-1.02	0.001
CG14811-RA	CG14811	-1.06	0.001	CG2713-RA	ttm50	-1.38	0.001
CG1828-RA	dre4	-1.22	0.009	CG7748-RA	OstStt3	-1.95	0.001
CG15259-RA	nht	-1.16	0.001	CG31287-RA	CG31287	-1.05	0.002
CG12948-RA	CG12948	-1.60	0.001	CG18545-RA	CG18545	-1.15	0.001
CG13920-RA	CG13920	-1.53	0.003	CG12934-RA	CG12934	-1.16	0.001
CG16971-RB	CR42862	-1.55	0.003	CG13400-RA	D12	-1.33	0.001
CG7676-RA	cona	-1.26	0.001	CG7276-RA	CG7276	-1.38	0.001
CG32428-RA	CG32428	-1.07	0.001	CG9981-RA	CG9981	-1.15	0.001
CG5030-RA	CG34139	-1.23	0.001	CG15224-RA	CkIIbeta	-1.38	0.001
CG10287-RA	Gasp	-1.11	0.001	CG5949-RA	DNApol-δ	-1.04	0.001
CG11949-RA	cora	-1.70	0.001	CG11778-RA	CG11778	-1.24	0.001
CG11575-RA	Ir100a	-1.20	0.001	CG7227-RA	CG7227	-1.05	0.001
CG11886-RA	Slbp	-1.35	0.001	CG31622-RB	Gr39a	-1.00	0.001
CG30156-RA	CG30156	-1.03	0.001	CG17304-RA	CG17304	-1.02	0.001
CG9076-RA	Cpr47Ed	-1.06	0.001	CG15778-RA	CG42449	-1.06	0.001
CG31501-RA	nxf4	-1.09	0.001	CG7321-RA	Rim	-1.13	0.001
CG15725-RA	CG15725	-1.16	0.001	CG14704-RA	PGRP-LB	-1.03	0.001
CG2144-RA	CG2144	-1.24	0.004	CG9273-RA	RPA2	-1.06	0.004
CG31651-RA	pgant5	-1.69	0.001	CG10540-RA	cpa	-1.43	0.001
CG1317-RB	CG1317	-1.05	0.002	CG13160-RA	CG13160	-1.04	0.001
CG1155-RA	Osi14	-1.06	0.001	CG2110-RA	Cyp4ad1	-1.30	0.001
CG3345-RA	CG3345	-1.07	0.001	CG9704-RA	Nrt	-1.46	0.001
CG32092-RB	CG42255	-1.15	0.001	CG10390-RA	mia	-1.37	0.001
CG14120-RA	CG14120	-1.39	0.001	CG13110-RA	CG13110	-1.12	0.001
CG9784-RA	CG9784	-1.04	0.001	CG6117-RA	Pka-C3	-1.33	0.001
CG33265-RA	Muc68E	-1.11	0.002	CG4120-RA	Cyp12c1	-1.44	0.001
CG10560-RA	CG10560	-1.10	0.001	CG7454-RA	Or85a	-1.04	0.001
CG15818-RA	CG15818	-1.02	0.001	CG9952-RA	ppa	-1.67	0.001
CG32105-RB	CG32105	-1.22	0.001	CG33489-RA	CG33489	-1.02	0.001
CG13221-RA	Vhl	-1.19	0.001	CG5656-RA	CG5656	-1.60	0.001
CG15478-RA	CG15478	-1.01	0.001	CG6136-RA	CG6136	-1.26	0.001
CG1429-RF	Mef2	-1.03	0.001	CG4017-RA	CG4017	-1.07	0.001
CG30431-RA	CG30431	-1.44	0.001	CG13067-RA	CG13067	-1.04	0.001
CG1600-RC	CG1600	-1.30	0.001	CG15483-RA	CG15483	-1.10	0.001
CG3423-RA	SA	-1.74	0.001	CG33481-RA	dpr7	-1.30	0.001
CG31414-RA	CG31414	-1.41	0.001	CG17195-RA	CG17195	-1.17	0.001
CG7760-RA	cato	-1.30	0.001	CG14116-RA	CG34428	-1.41	0.001

Continued on next page

Table 8 – continued from previous page

		Dichaete mutant				*Dichaete* mutant	
Transcript	**Gene**	**M**	**p**	**Transcript**	**Gene**	**M**	**p**
CG17461-RA	Kif3C	-1.00	0.001	CG7306-RA	obst-F	-1.34	0.001
CG4128-RE	nAcRα-30D	-1.30	0.001	CG2082-RA	CG2082	-1.03	0.001
CG33323-RA	Fer1	-1.05	0.001	CG31645-RA	CG34381	-1.05	0.001
CG31319-RA	cv-c	-1.50	0.001	CG40249-RA	CG40249	-1.41	0.001
CG11459-RA	CG11459	-1.16	0.001	CG3385-RA	nvy	-1.16	0.001
CG31538-RA	CG31538	-1.23	0.001	CG6056-RA	AP-2sigma	-1.61	0.001
CG9722-RA	CG9722	-1.11	0.001	CG32159-RB	dsx-c73A	-1.02	0.001
CG1984-RA	djl	-1.26	0.001	CG14720-RA	CG14720	-1.10	0.001
CG2093-RA	Vps13	-1.72	0.001	CG9603-RA	CG9603	-1.93	0.001
CG11630-RA	CG11630	-1.25	0.001	CG31370-RA	CG31370	-1.32	0.001
CG6103-RD	CrebB-17A	-1.12	0.001	CG11395-RA	CG11395	-1.10	0.001
CG9884-RA	oaf	-1.04	0.002	CG31009-RA	Cad99C	-1.17	0.003
CG8644-RA	Osi22	-1.03	0.001	CG2275-RB	Jra	-1.27	0.001
CG9630-RA	CG9630	-1.36	0.001	CG6467-RA	Jon65Aiv	-1.06	0.001
CG33456-RA	mute	-1.09	0.001	CG31364-RA	l(3)neo38	-1.85	0.001
CG31332-RA	Unc-115b	-1.98	0.001	CG15440-RA	CG15440	-1.05	0.001
CG12770-RA	Vps28	-1.10	0.002	CG3088-RA	CG3088	-1.17	0.001
CG17320-RA	ScpX	-1.30	0.001	CG13487-RA	Fili	-1.39	0.001
CG5237-RA	unc79	-1.18	0.001	CG30067-RB	Obp50a	-1.06	0.001
CG5398-RA	CG5398	-1.02	0.001	CG2989-RA	Cht6	-1.08	0.002
CG32639-RA	CG32639	-1.23	0.002	CG4710-RA	Pino	-1.13	0.002
CG14366-RA	CG34383	-1.33	0.001	CG7756-RA	Hsc70-2	-1.38	0.001
CG13842-RA	CG13842	-1.19	0.001	CG5610-RA	nAcRα-96Aa	-1.00	0.001
CG1856-RD	ttk	-1.92	0.001	CG3132-RA	Ect3	-1.73	0.001
CG1447-RB	Ptx1	-1.25	0.005	CG5225-RA	CG5225	-1.29	0.001
CG12052-RU	lola	-1.03	0.003	CG3702-RA	CG3702	-1.57	0.003
CG32280-RA	CG32280	-1.62	0.002	CG3368-RA	CG3368	-1.28	0.003
CG17754-RA	CG17754	-1.10	0.004	CG4722-RA	bib	-2.03	0.003
CG7335-RA	CG7335	-1.18	0.004	CG4158-RA	wor	-1.72	0.002
CG6484-RA	CG6484	-1.09	0.002	CG18314-RA	DopEcR	-1.01	0.001
CG31177-RA	CG42335	-1.09	0.004	CG2906-RA	CG2906	-1.45	0.001
CG33061-RA	CG33061	-1.01	0.005	CG6931-RA	CG6931	-1.10	0.001
CG14693-RA	CG14693	-1.34	0.001	CG5597-RA	CG5597	-1.07	0.007
CG2937-RA	mRpS2	-1.57	0.002	CG18507-RA	CG18507	-1.13	0.008
CG3456-RA	Mct1	-1.20	0.008	CG3669-RA	CG3669	-1.06	0.002
CG7013-RA	Manf	-1.57	0.001	CG1339-RA	Gr43b	-1.24	0.001
CG11208-RA	CG11208	-1.14	0.004	CG17840-RA	CG17840	-1.20	0.001
CG5348-RA	CG5348	-1.20	0.001	CG9752-RA	CG9752	-1.23	0.001
CG3715-RA	Shc	-1.52	0.002	CG6919-RA	oa2	-1.09	0.001
CG18493-RA	CG18493	-1.10	0.001	CG7463-RA	yellow-k	-1.10	0.001
CG5740-RA	CG42390	-1.14	0.001	CG17190-RA	CG17190	-1.19	0.001
CG1442-RA	eIF4E-6	-1.29	0.001	CG10591-RA	CG10591	-1.27	0.006
CG14999-RA	RfC4	-1.20	0.002	CG10121-RB	SP1173	-1.21	0.003
CG1851-RA	Ady43A	-1.13	0.009	CG17645-RA	Pglym87	-1.19	0.001
CG1076-RA	Vha14-2	-1.02	0.006	CG3399-RA	capu	-1.26	0.007
CG32664-RA	Gr10a	-1.24	0.001	CG18581-RA	CG18581	-1.01	0.006
CG4845-RA	psidin	-1.10	0.005	CG17786-RA	CG17786	-1.45	0.003
CG31251-RA	CG31251	-1.05	0.002	CG3172-RA	twf	-1.17	0.004
CG15009-RB	ImpL2	-1.84	0.002	CG4827-RA	veil	-1.06	0.001
CG31900-RA	CG43394	-1.32	0.001	CG9381-RA	mura	-1.32	0.001
CG32666-RB	Drak	-1.36	0.001	CG8026-RB	CG8026	-1.12	0.001
CG10621-RA	CG10621	-1.74	0.003	CG12404-RA	CG12404	-1.17	0.001
CG4438-RA	CG4438	-1.03	0.001	CG8896-RA	18w	-1.80	0.003

Continued on next page

Table 8 – continued from previous page

		Dichaete mutant				*Dichaete* mutant	
Transcript	**Gene**	**M**	**p**	**Transcript**	**Gene**	**M**	**p**
CG31867-RA	CG31867	-1.11	0.001	CG30447-RA	CG30447	-1.00	0.001
CG11196-RA	Dic3	-1.07	0.001	CG32845-RA	CG32845	-1.54	0.001
CG40041-RB	Gpb5	-1.11	0.001	CR33319-RA	CR33319	-1.19	0.001
CG14455-RA	CG14455	-1.22	0.001	CG3808-RA	CG3808	-1.18	0.002
CG5498-RB	CG5498	-1.43	0.001	CG5137-RA	Cyp312a1	-1.12	0.001
CG6646-RA	DJ-1alpha	-1.17	0.001	CG4793-RB	CG4793	-1.34	0.001
CG10695-RA	Pat1	-1.45	0.001	CG5111-RA	CG5111	-1.10	0.001
CG30383-RA	CG30383	-1.12	0.001	CG5474-RA	SsRbeta	-1.72	0.001
CG11145-RA	CG11145	-1.05	0.001	CG8581-RA	fra	-1.07	0.001
CG13720-RA	CG13720	-1.26	0.001	CG12592-RA	CG12592	-1.30	0.003
CG16736-RA	CG16736	-1.17	0.001	CG30287-RA	CG30287	-1.37	0.001
CR32477-RA	CR42862	-2.01	0.003	CG11262-RA	CG11262	-1.17	0.001
CG18731-RA	CG18731	-1.54	0.001	CG5509-RA	CG5509	-1.21	0.001
CG13059-RA	CG13059	-1.33	0.001	CG9121-RA	CG9121	-1.19	0.001
CG5249-RA	Blimp-1	-1.65	0.001	CG13738-RA	CG13738	-1.25	0.001
CG12754-RA	Or42b	-1.09	0.001	CG8075-RA	Vang	-1.24	0.001
CG13827-RA	CG13827	-1.47	0.001	CG12620-RA	CG12620	-1.02	0.001
CG18619-RD	CG18619	-1.17	0.001	CG10047-RA	Syt4	-1.66	0.001
CG31858-RA	t-cup	-1.09	0.001	CG6653-RA	Ugt86De	-1.13	0.001
CG16726-RA	CG33696	-1.46	0.001	CG12370-RA	Dh44-R2	-1.06	0.001
CG8134-RA	CG8134	-1.00	0.001	CR30068-RA	CR30068	-1.29	0.001
CG8843-RA	sec5	-1.11	0.004	CG8417-RA	CG8417	-1.26	0.001
CG5078-RA	CG5078	-1.07	0.001	CG12023-RA	GV1	-1.14	0.001
CG3916-RA	CG3916	-1.02	0.001	CG16786-RA	CG16786	-1.06	0.001
CG1056-RA	5-HT2	-1.17	0.001	CG9266-RB	CG42238	-1.53	0.001
CG16834-RA	lectin-33A	-1.28	0.001	CG6514-RA	TpnC25D	-1.00	0.001
CG6569-RA	CG6569	-1.06	0.001	CG13658-RA	CG13658	-1.22	0.001
CG6803-RC	Mf	-1.11	0.001	CG3625-RB	CG3625	-1.77	0.001
CG6599-RA	CG42534	-1.07	0.001	CG1130-RA	scrt	-1.81	0.001
CG8856-RA	Sr-CII	-1.22	0.004	CG1263-RB	RpL8	-1.03	0.001
CG7204-RA	Neu2	-1.14	0.001	CG3956-RA	sna	-1.85	0.001
CG13653-RA	CG13653	-1.15	0.001	CG12920-RA	CG12920	-1.02	0.001
CG9792-RA	yellow-e	-1.26	0.001	CG5737-RA	dmrt93B	-1.15	0.001
CG30338-RA	CG30338	-1.13	0.001	CG3896-RA	Nox	-1.06	0.002
CG4842-RA	CG4842	-1.00	0.001	CG6441-RA	CG6441	-1.29	0.001
CG8007-RA	CG42796	-1.25	0.001	CG5050-RA	CG5050	-1.00	0.001
CG12799-RA	Ubc84D	-1.20	0.001	CG5892-RA	CG5892	-1.11	0.001
CG13833-RA	CG13833	-1.00	0.001	CG31906-RA	Ndae1	-1.32	0.001
CR31271-RA	iab-4	-1.27	0.001	CG30151-RA	CG30151	-1.15	0.001
CG12758-RB	sano	-1.63	0.001	CG12478-RA	bru-3	-1.10	0.001
CG13154-RA	CG13154	-1.02	0.001	CG32750-RA	CG32750	-1.22	0.001
CG9508-RA	CG42370	-1.08	0.001	CG13527-RA	CG13527	-1.09	0.001
CG5267-RA	CG5267	-1.15	0.001	CG32228-RA	CG32228	-1.00	0.001
CG13475-RA	HGTX	-1.05	0.001	CG5421-RA	CG5421	-1.32	0.001
CG30008-RA	CG30008	-1.19	0.001	CG4211-RA	nonA	-1.25	0.001
CG32712-RA	CG32712	-1.06	0.001	CG14049-RA	Ilp6	-1.31	0.001
CG18412-RA	ph-p	-1.44	0.002	CG15617-RA	CG15617	-1.13	0.001
CG14502-RB	CG14502	-1.11	0.001	CG3625-RA	CG3625	-1.72	0.007
CG14608-RA	CG14608	-2.15	0.001	CG10772-RA	Fur1	-1.17	0.002
CG13624-RA	CG13624	-1.29	0.001	CG7586-RA	Mcr	-1.46	0.002
CG14940-RA	Pde1c	-1.01	0.001	CG3517-RA	CG3517	-1.20	0.001
CG14993-RA	Faa	-0.81	0.002	CG4742-RA	mRpL22	-1.18	0.001
CG7496-RA	PGRP-SD	-1.21	0.001	CG14981-RA	mge	-1.47	0.001

Continued on next page

Table 8 – continued from previous page

		Dichaete mutant				*Dichaete* mutant	
Transcript	**Gene**	**M**	**p**	**Transcript**	**Gene**	**M**	**p**
CG4185-RA	NC2beta	-1.18	0.001	CG7554-RA	comm2	-1.62	0.001
CG14562-RA	CG14562	-1.08	0.001	CG4853-RB	CG4853	-1.27	0.001
CG4963-RA	mfrn	-1.08	0.001	CG10809-RA	CG10809	-1.37	0.003
CG7079-RA	CG7079	-1.12	0.001	CG6993-RA	ss	-1.25	0.001
CG4319-RA	rpr	-1.84	0.002	CG31624-RA	CG31624	-1.01	0.002
CG33159-RA	CG33159	-1.07	0.001	CG30035-RA	Tret1-1	-1.65	0.001
CG5973-RA	CG5973	-1.36	0.001	CG17564-RB	CG17564	-1.36	0.001
CG4170-RD	vig	-1.45	0.002	CG6403-RA	CG6403	-1.02	0.001
CG2939-RA	slp2	-1.01	0.001	CG11121-RA	so	-1.34	0.001
CG13916-RA	SA-2	-1.09	0.001	CG17776-RA	CG17776	-1.46	0.001
CG14328-RA	CG34281	-1.27	0.001	CG1345-RA	Gfat2	-1.25	0.001
CG5721-RA	CG5721	-1.35	0.001	CG17871-RB	Or71a	-1.29	0.001
CG8095-RB	scb	-1.29	0.001	CG6692-RA	Cp1	-1.38	0.001
CG32396-RA	CG32396	-1.18	0.001	CG10284-RA	CG42564	-1.17	0.001
CG15408-RA	CG15408	-1.11	0.001	CG11470-RA	CG11470	-1.05	0.001
CG14913-RA	CG14913	-1.23	0.001	CG3430-RA	CG3430	-1.10	0.001
CG30084-RB	Zasp52	-1.55	0.003	CG13640-RA	CG13640	-1.21	0.001
CG33206-RA	Gmap	-1.64	0.001	CG5315-RA	CG5315	-1.41	0.001
CG33007-RA	CG33007	-1.03	0.001	CG6545-RA	lbe	-1.01	0.001
CG10639-RA	CG10639	-1.03	0.001	CG17821-RA	CG17821	-1.01	0.002
CG14830-RA	CG14830	-1.37	0.001	CG5670-RD	Atpalpha	-1.89	0.001
CG14497-RA	CG34386	-1.19	0.001	CG32444-RA	CG32444	-1.03	0.001
CG32161-RA	CG32161	-1.20	0.001	CG5114-RA	CG5114	-1.28	0.001
CG4803-RA	Takl2	-1.12	0.001	CG18466-RB	Nmdmc	-1.20	0.001
CG11029-RA	CG11029	-1.26	0.001	CG7614-RA	Mat1	-1.01	0.001
CG4421-RA	GstD8	-1.03	0.001	CG32690-RA	CR32690	-1.09	0.001
CG8936-RA	Arpc3B	-1.18	0.001	CG9539-RA	Sec61alpha	-1.76	0.001
CG30372-RA	CG30372	-1.36	0.001	CG12443-RA	ths	-1.10	0.001
CG18096-RA	TepI	-1.25	0.001	CG8053-RB	eIF-1A	-1.03	0.001
CG13050-RA	CG13050	-1.31	0.001	CG4463-RA	Hsp23	-1.22	0.001
CG8927-RA	CG8927	-1.14	0.001	CG5058-RD	grh	-1.80	0.001
CG32680-RB	spri	-1.08	0.001	CG30268-RA	CG30268	-1.26	0.002
CG31882-RA	CG31882	-1.27	0.001	CG32859-RA	eIF4E-7	-1.15	0.001
CG10021-RD	bowl	-1.82	0.001	CG4006-RA	Akt1	-1.35	0.001
CG12501-RA	Or56a	-1.03	0.001	CG1070-RC	Alh	-1.56	0.001
CG13632-RA	CG42331	-1.16	0.001	CG9392-RA	CG9392	-1.08	0.001
CG12011-RA	CG12011	-1.77	0.002	CG14384-RA	CG14384	-1.08	0.001
CG4784-RA	Cpr72Ec	-1.07	0.001	CG4525-RA	CG4525	-1.04	0.001
CG5526-RA	Dhc36C	-1.20	0.001	CG5507-RA	T48	-1.23	0.001
CG9215-RA	CG9215	-1.13	0.001	CG8051-RA	CG8051	-1.07	0.001
CG12623-RA	Ir10a	-1.17	0.001	CG4589-RC	Letm1	-1.24	0.001
CG16752-RA	SPR	-1.05	0.001	CG30329-RA	Vha100-3	-1.34	0.001
CG12605-RB	CG12605	-1.03	0.001	CG16700-RA	CG16700	-1.12	0.001
CG1380-RB	sut4	-1.25	0.001	CR32863-RA	MeU5-C46	-1.37	0.001
CG11281-RA	snky	-1.18	0.001	CG16713-RA	CG16713	-1.10	0.001
CG8192-RA	CG8192	-1.34	0.001	CG1309-RA	CG1309	-1.10	0.001
CG9396-RA	CG9396	-1.08	0.001	CG13792-RA	CG13792	-1.05	0.001
CG5576-RA	imd	-1.40	0.001	CG5441-RA	tx	-1.28	0.001
CG8097-RA	CG8097	-1.61	0.001	CG13867-RA	MED8	-1.02	0.001
CG18657-RA	NetA	-1.75	0.001	CG1271-RB	CG1271	-1.08	0.001
CG8505-RA	Cpr49Ae	-1.01	0.001	CG31873-RA	CG31873	-1.19	0.001
CG6098-RA	Lrr47	-1.22	0.001	CR31400-RC	Hsromega	-2.29	0.001
CG5118-RA	CG5118	-1.39	0.001	CG16885-RB	CG16885	-1.08	0.001

Continued on next page

Table 8 – continued from previous page

		Dichaete mutant				*Dichaete* mutant	
Transcript	**Gene**	**M**	**p**	**Transcript**	**Gene**	**M**	**p**
CG1681-RA	CG1681	-1.57	0.001	CG3280-RA	CG3280	-1.09	0.001
CG9504-RA	Eo	-1.14	0.001	CG9497-RA	CG9497	-1.03	0.001
CG32850-RA	CG32850	-2.15	0.002	CG5590-RA	CG5590	-1.34	0.001
CG10145-RA	mspo	-1.07	0.001	CG32037-RA	CG32037	-1.24	0.001
CG5154-RA	Idgf5	-1.27	0.001	CG5481-RA	lea	-2.05	0.001
CG8546-RA	CG8546	-1.06	0.001	CG3980-RB	Cep97	-1.02	0.004
CG13129-RA	CG34109	-1.23	0.001	CG8205-RE	fus	-1.50	0.001
CG4720-RB	Pk92B	-1.05	0.001	CG13263-RA	Cyt-c-d	-1.01	0.004
CR31808-RA	CG31808	-1.34	0.001	CG12101-RA	Hsp60	-1.17	0.004
CG8144-RC	ps	-1.69	0.001	CG30460-RC	CG30460	-1.13	0.005
CG11086-RA	Gadd45	-1.26	0.001	CG7972-RA	mus301	-1.09	0.001
CG10185-RA	CG10185	-1.11	0.002	CG7088-RA	bnb	-2.49	0.002
CG13295-RA	CG13295	-1.29	0.001	CG4554-RA	CG4554	-1.22	0.005
CG12817-RA	CG12817	-1.02	0.001	CG17152-RA	Ir68b	-1.07	0.002
CG15581-RA	Or83c	-1.00	0.002	CG7708-RA	CG7708	-1.00	0.009
CG6889-RA	tara	-1.69	0.002	CG13396-RA	fy	-1.00	0.002
CG2201-RB	CG2201	-1.16	0.002	CG18766-RA	CG18766	-1.57	0.001
CG8596-RA	CG8596	-1.50	0.009	CG14022-RA	CG14022	-1.67	0.001
CG4648-RA	CG42399	-1.09	0.003	CG32476-RA	mthl14	-1.05	0.002
CG17239-RA	CG17239	-1.03	0.007	CG7230-RA	rib	-2.25	0.004
CG2863-RA	Nle	-1.25	0.005	CG1772-RA	dap	-2.35	0.001
CG8172-RA	CG8172	-1.09	0.002	CG5869-RA	CG5869	-1.79	0.003
CG17544-RB	CG17544	-1.01	0.004	CG10859-RA	CG10859	-1.05	0.005
CG9220-RB	CG9220	-1.25	0.001	CG8846-RA	Thor	-1.63	0.001
CG31121-RA	CG31121	-1.01	0.001	CG8325-RA	l(2)k14710	-1.05	0.001
CG5781-RA	CG5781	-1.01	0.002	CG4109-RA	Syx8	-1.22	0.009
CG12402-RA	CG12402	-1.17	0.001	CG31821-RA	CG31821	-1.03	0.002
CG12264-RA	CG12264	-1.29	0.003	CG31742-RA	CG31742	-1.49	0.001
CG16992-RA	mthl6	-1.15	0.003	CG12917-RA	CG12917	-1.00	0.001
CG9879-RA	CG9879	-1.00	0.001	CG6331-RA	Orct	-1.34	0.001
CG32443-RA	Pc	-1.11	0.001	CG13125-RA	CG13125	-1.11	0.001
CG2374-RA	lbm	-1.23	0.001	CG11849-RA	dan	-1.03	0.001
CG12068-RA	sro	-1.11	0.001	CG32403-RA	Or65c	-1.14	0.001
CG9737-RA	CG9737	-1.17	0.001	CG7131-RA	CG7131	-1.06	0.001
CG16894-RA	CG16894	-1.28	0.001	CG14983-RA	CG14983	-1.04	0.001
CG11537-RA	CG11537	-1.03	0.003	CG6930-RA	l(3)neo38	-2.05	0.001
CG2103-RB	pgant6	-1.61	0.001	CG15596-RA	Osi11	-1.37	0.001
CG12102-RA	CG34104	-1.16	0.001	CG2114-RA	FR	-1.06	0.001
CG1987-RA	Rbp1-like	-1.26	0.001	CG13074-RA	CG13074	-1.04	0.001
CG33348-RA	CheB42a	-1.13	0.001	CG9624-RA	CG9624	-1.00	0.001
CG16755-RA	Or85f	-1.29	0.001	CG3178-RA	Rrp1	-1.47	0.001
CG33171-RE	mp	-1.16	0.001	CG6081-RA	Cyp28d2	-1.39	0.001
CG2005-RC	Ptp99A	-1.55	0.001	CG33494-RA	CG33494	-1.44	0.001
CG11921-RA	fd96Ca	-1.17	0.001	CG7428-RA	halo	-1.71	0.001
CG14118-RA	CG14118	-1.31	0.001	CG32077-RA	nol	-1.01	0.001
CG15287-RA	ms(2)34Fe	-1.07	0.001	CG3412-RA	slmb	-1.34	0.001
CG33476-RA	CG33476	-1.15	0.001	CG33196-RB	dp	-1.59	0.005
CG1133-RA	opa	-1.48	0.001	CG32335-RA	CG32335	-1.25	0.001
CG15771-RA	CG15771	-1.04	0.001	CG14269-RA	CG14269	-1.40	0.001
CG1132-RA	fd64A	-1.11	0.001	CG2918-RA	CG2918	-1.31	0.005
CG11639-RA	TfIIA-S-2	-1.10	0.001	CG4216-RA	term	-1.06	0.001
CG9084-RB	CG9084	-1.07	0.001	CG3401-RA	betaTub60D	-2.54	0.001
CG14086-RA	CG14086	-1.08	0.001	CG8714-RA	sut1	-1.38	0.001

Continued on next page

Table 8 – continued from previous page

		Dichaete mutant				*Dichaete* mutant	
Transcript	**Gene**	**M**	**p**	**Transcript**	**Gene**	**M**	**p**
CG2053-RA	CG2053	-1.16	0.001	CG3544-RA	CG3544	-1.05	0.001
CG33296-RA	CG33296	-1.00	0.001	CG5545-RA	Oli	-1.23	0.005
CG15655-RA	CG34396	-1.35	0.001	CG15427-RA	tutl	-1.04	0.001
CG1179-RA	LysB	-1.57	0.001	CG3822-RA	CG3822	-1.00	0.001
CG7364-RA	TM9SF4	-1.71	0.001	CG7635-RA	Mec2	-1.13	0.001
CG2069-RA	Oseg4	-1.18	0.001	CG6348-RB	ro	-1.00	0.002
CG15437-RA	morgue	-1.28	0.001	CG9602-RA	CG9602	-1.11	0.001
CG1671-RA	CG1671	-1.38	0.001	CG31352-RA	Unc-115a	-1.36	0.001
CG15423-RA	CG15423	-1.35	0.001	CG15829-RA	CG15829	-1.10	0.001
CG8624-RA	melt	-1.09	0.001	CG12189-RA	Rev1	-1.03	0.002
CG2196-RA	salt	-1.21	0.001	CG14214-RA	Sec61γ	-1.30	0.003
CG4928-RA	CG4928	-1.22	0.001	CG33092-RA	P5CDh2	-1.05	0.001
CG1546-RB	PH4αSG2	-1.04	0.001	CG33338-RA	p38c	-1.15	0.001
CG9036-RA	Cpr56F	-1.35	0.001	CG32120-RA	sens	-1.61	0.002
CG6127-RA	Ser	-1.49	0.001	CG30037-RA	CG30037	-1.14	0.001
CG14735-RA	ssp5	-1.20	0.001	CG31019-RA	CG31019	-1.05	0.001
CG10396-RA	CG10396	-1.20	0.001	CG15094-RB	CG15094	-1.16	0.002
CG4103-RA	l(2)35Bc	-1.18	0.001	CG15400-RA	CG15400	-1.18	0.001
CG9908-RA	disco	-1.47	0.003	CG11699-RA	CG11699	-1.32	0.001
CG8080-RA	CG8080	-1.43	0.001	CG1771-RB	mew	-1.71	0.001
CG7695-RA	CG7695	-1.37	0.001	CG7120-RA	CG7120	-1.29	0.001
CG10289-RA	CG10289	-1.31	0.001	CG16979-RA	CG16979	-1.18	0.002
CG16813-RA	CG16813	-1.34	0.001	CG17707-RA	CG17707	-1.01	0.002
CG8254-RA	exex	-1.30	0.001	CG4986-RA	Mst57Dc	-1.00	0.008
CG12194-RA	CG12194	-1.03	0.001	CG17054-RD	Cap-G	-1.47	0.001
CG13348-RA	Aats-phe	-1.08	0.001	CG33013-RB	CG33013	-1.09	0.001
CG15120-RA	CG15120	-1.05	0.001	CG13321-RA	CG13321	-1.03	0.001
CG8332-RB	RpS15	-1.22	0.001	CG2189-RA	Dfd	-1.42	0.001
CG3723-RA	Dhc93AB	-1.01	0.001	CG11718-RA	CG42286	-1.20	0.001
CG8320-RA	CG8320	-1.09	0.002	CG31735-RA	CG31735	-1.12	0.001
CG5018-RA	CG5018	-1.21	0.001	CG2233-RA	CG2233	-1.35	0.001
CG13494-RA	CG13494	-1.13	0.001	CG17221-RA	CG17221	-1.18	0.001
CG15211-RA	CG15211	-1.13	0.001	CG15270-RA	CG15270	-1.11	0.002
CG7945-RC	CR43159	-1.27	0.001	CG17262-RA	cnir	-1.06	0.002
CG8396-RA	Ssb-c31a	-1.77	0.001	CG8386-RA	CG8386	-1.25	0.001
CG32245-RB	CG42540	-1.07	0.001	CG3766-RA	scat	-1.06	0.001
CG12310-RA	CG12310	-1.18	0.001	CG5885-RA	CG5885	-1.61	0.001
CG8837-RA	CG8837	-1.01	0.001	CG10334-RA	spi	-1.91	0.004
CG4701-RA	CG4701	-1.00	0.001	CG4898-RB	Tm1	-1.29	0.001
CG9138-RA	uif	-1.70	0.002	CG9010-RA	CG9010	-1.09	0.001
CG4937-RA	RhoGAP15B	-1.06	0.001	CG4976-RA	Mes-4	-1.36	0.001
CG17660-RA	CG17660	-1.37	0.001	CG8289-RA	CG8289	-1.49	0.002
CG7910-RA	CG7910	-1.22	0.001	CG8909-RB	CG8909	-1.22	0.001
CG15595-RA	Osi13	-1.24	0.001	CG8981-RA	MCPH1	-1.13	0.007
CG10344-RB	CG10344	-1.04	0.001	CG7745-RA	CG7745	-1.12	0.001
CG1693-RA	tty	-1.21	0.001	CG2621-RC	sgg	-1.51	0.001
CG32454-RA	CG32454	-1.14	0.001	CG13391-RA	Aats-ala	-1.29	0.001
CG8214-RA	CG8214	-1.05	0.001	CG12178-RA	Nhe1	-1.10	0.001
CG10248-RA	Cyp6a8	-1.49	0.001	CG13277-RA	LSm7	-1.00	0.009
CG12287-RA	pdm2	-1.41	0.009	CG8303-RA	CG8303	-1.26	0.001
CG10262-RA	CG10262	-1.10	0.001	CG15929-RA	lin-52	-1.07	0.002
CG4901-RA	CG4901	-1.17	0.001	CG14892-RA	CG14892	-1.00	0.001
CG31622-RA	Gr39a	-1.12	0.001	CG3014-RA	CG3014	-1.05	0.001

Continued on next page

Table 8 – continued from previous page

		Dichaete mutant				*Dichaete* mutant	
Transcript	**Gene**	**M**	**p**	**Transcript**	**Gene**	**M**	**p**
CG17324-RA	CG17324	-1.15	0.001	CG17302-RA	Prosbeta4R2	-1.19	0.001
CG11888-RA	Rpn2	-1.38	0.001	CG7395-RA	sNPF-R	-1.38	0.001
CG5999-RA	CG5999	-1.02	0.001	CG7320-RA	CG7320	-1.18	0.001
CG2140-RB	Cyt-b5	-1.20	0.001	CG10539-RA	S6k	-1.62	0.001
CG8728-RA	CG8728	-1.09	0.001	CG13178-RA	CG13178	-1.09	0.001
CG9674-RB	CG9674	-1.23	0.002	CG32087-RA	CG32087	-1.02	0.001
CG10388-RA	Ubx	-1.33	0.001	CG13109-RA	tai	-1.06	0.001
CG33178-RA	CG33178	-1.20	0.001	CG6262-RB	CG6262	-1.19	0.001
CG14285-RA	CG14285	-1.03	0.001	CG32104-RB	CG32104	-1.04	0.009
CG30272-RA	CG30272	-1.25	0.003	CG8776-RA	nemy	-1.26	0.001
CG1429-RD	Mef2	-1.53	0.001	CG33480-RA	hig	-1.00	0.001
CG31413-RA	CG31413	-1.28	0.001	CG1976-RA	RhoGAP100F	-1.01	0.001
CG32110-RA	CG32110	-1.46	0.001	CG7304-RA	CG7304	-1.12	0.001
CG17477-RA	CG17477	-1.23	0.001	CG4125-RA	rst	-1.64	0.001
CG31317-RA	stumps	-1.98	0.001	CG15308-RA	CG15308	-1.02	0.001
CG3097-RA	CG3097	-1.19	0.001	CG6055-RA	CG6055	-1.23	0.001
CG9596-RA	CG9596	-1.10	0.001	CG15480-RA	CG15480	-1.77	0.002
CG15871-RA	mRpL38	-1.06	0.002	CG9588-RA	CG9588	-1.19	0.007
CG1418-RA	CG1418	-1.18	0.003	CG9015-RA	en	-1.10	0.004
CG4196-RB	CG42542	-1.16	0.007	CG17579-RA	sca	-1.97	0.002
CG31757-RA	Pde1c	-1.13	0.001	CG31323-RA	CG31323	-1.13	0.001
CG30021-RA	metro	-1.25	0.001	CG13604-RA	CG13604	-1.01	0.003
CG1082-RA	alpha-Est4	-1.00	0.008	CG14332-RA	CG14332	-1.54	0.001
CG1079-RA	Fie	-1.01	0.001	CG1597-RA	CG1597	-1.32	0.004
CG17012-RA	CG17012	-1.07	0.001	CG33052-RA	CG33052	-1.50	0.001
CG7619-RA	Pros54	-1.54	0.003	CG31043-RA	gukh	-1.40	0.004
CG17669-RA	CG17669	-1.28	0.001	CG14304-RA	CG14304	-1.16	0.003
CG17927-RK	Mhc	-1.08	0.001	CG17838-RA	CG17838	-1.01	0.004
CG10576-RA	CG10576	-1.52	0.001	CG3194-RA	CG3194	-1.00	0.005
CG6388-RA	CG6388	-1.34	0.001	CG3353-RA	CG3353	-1.51	0.001
CG3396-RA	Ocho	-1.46	0.006	CG13611-RA	CG13611	-1.02	0.002
CG3260-RA	Zfrp8	-1.24	0.002	CG7702-RB	CG7702	-1.47	0.002
CG6520-RA	CG6520	-1.59	0.001	CG10234-RA	Hs2st	-1.21	0.001
CG31313-RA	CG31313	-1.21	0.001	CG5893-RA	D	-1.92	0.004
CG31714-RA	CG31714	-1.24	0.001	CG14219-RA	CG14219	-1.00	0.001
CG5658-RA	Klp98A	-1.06	0.002	CG31493-RA	CG31493	-1.24	0.001
CG4783-RA	CG4783	-1.37	0.001	CG7556-RA	CG7556	-1.60	0.001
CG5384-RA	CG5384	-1.35	0.001	CG4383-RA	CG33960	-1.00	0.001
CG30277-RA	Oatp58Da	-1.07	0.001	CG9517-RB	CG9517	-1.10	0.001
CG12418-RA	Glut4EF	-1.72	0.001	CG10619-RB	tup	-1.33	0.001
CG13018-RA	CG13018	-1.53	0.001	CG12403-RA	Vha68-1	-1.18	0.001
CG4437-RA	PGRP-LF	-1.13	0.001	CG7957-RA	MED17	-1.27	0.001
CG10642-RA	Klp64D	-1.08	0.001	CG18269-RA	CG18269	-1.24	0.002
CG13859-RA	CG34377	-1.15	0.001	CG30189-RA	Gr59a	-1.01	0.001
CG11194-RA	Hey	-1.05	0.002	CG32843-RA	Dh31-R1	-1.06	0.001
CG7311-RA	CG7311	-1.28	0.001	CG9366-RA	RhoL	-1.33	0.002
CG14307-RC	fru	-1.07	0.001	CR33294-RA	CR33294	-1.22	0.001
CG13386-RA	CG13386	-1.33	0.001	CG10134-RA	beat-Va	-1.08	0.001
CG14451-RA	CG14451	-1.14	0.001	CG3775-RA	CG3775	-1.02	0.001
CG10575-RA	Ppat-Dpck	-1.11	0.002	CG5411-RA	Pde8	-1.09	0.001
CG8424-RA	Jhedup	-1.04	0.001	CG6638-RA	CG6638	-1.09	0.001
CG14020-RA	CG14020	-1.36	0.001	CG1910-RA	CG1910	-1.51	0.003
CG8695-RA	Mal-A3	-1.08	0.001	CG5870-RA	beta-Spec	-1.13	0.007

Continued on next page

Table 8 – continued from previous page

		Dichaete mutant				*Dichaete* mutant	
Transcript	**Gene**	**M**	**p**	**Transcript**	**Gene**	**M**	**p**
CG13715-RA	axo	-1.42	0.001	CG16735-RA	CR16735	-1.02	0.001
CG30284-RA	CG30284	-1.03	0.001	CG1257-RA	alpha-Est3	-1.08	0.001
CR32385-RA	CR32385	-1.20	0.001	CG8174-RA	SRPK	-1.43	0.001
CG9458-RA	CG9458	-1.03	0.001	CG13058-RA	CG13058	-1.01	0.001
CG9102-RA	bab2	-1.87	0.002	CG7125-RA	PKD	-1.12	0.001
CG6488-RA	CG6488	-1.21	0.001	CG13737-RA	CG13737	-1.09	0.001
CG3017-RA	Alas	-1.24	0.001	CG5550-RA	CG5550	-1.19	0.001
CG11291-RA	CG11291	-1.02	0.001	CG32798-RA	Vml	-1.67	0.001
CG12653-RA	btd	-1.47	0.001	CG13664-RA	Cad96Cb	-1.24	0.001
CG31857-RA	bun	-1.45	0.001	CG5102-RA	da	-1.05	0.006
CG5703-RA	CG5703	-1.26	0.001	CG1818-RA	Updo	-1.18	0.001
CG4448-RA	wda	-1.19	0.001	CR30009-RA	CR30009	-1.58	0.001
CG8416-RB	Rho1	-1.34	0.001	CG13784-RB	CG13784	-1.05	0.001
CG18258-RA	CG18258	-1.00	0.001	CG18810-RA	CG18810	-1.27	0.001
CG3278-RA	Tif-IA	-1.37	0.001	CG16833-RA	CG16833	-1.39	0.001
CG6567-RA	CG6567	-1.45	0.001	CG11983-RA	CG11983	-1.17	0.001
CG6798-RA	nAcRβ-96A	-1.14	0.001	CG32627-RA	NnaD	-1.00	0.001
CG14141-RA	CG14141	-1.14	0.001	CG8506-RA	Rbsn-5	-1.15	0.001
CG3009-RA	CG3009	-1.13	0.001	CG3624-RA	CG3624	-1.72	0.002
CG6568-RA	CG6568	-1.33	0.001	CG11308-RA	sa	-1.27	0.001
CG5125-RA	ninaC	-1.03	0.001	CG30141-RA	Obp57a	-1.07	0.001
CG9284-RB	CG9284	-1.29	0.001	CG12636-RA	CG42685	-1.06	0.001
CG8468-RB	CG8468	-1.84	0.001	CG9983-RB	Hrb98DE	-1.59	0.002
CG30491-RA	CG30491	-1.38	0.001	CG12026-RB	Tmhs	-1.29	0.001
CG13651-RA	danr	-1.12	0.001	CG18131-RD	CG18131	-1.04	0.001
CG14243-RA	TwdlD	-1.05	0.001	CG8536-RA	β4GalNAcTA	-1.11	0.001
CG5732-RA	Gld2	-1.35	0.001	CG3988-RA	γSnap	-1.41	0.001
CG8960-RA	CG8960	-1.92	0.001	CG6417-RA	Oatp33Eb	-1.21	0.001
CG7955-RC	ABCB7	-1.27	0.001	CG5048-RA	CG5048	-1.03	0.001
CG5863-RA	CG5863	-1.42	0.001	CG8120-RA	HP1e	-1.36	0.001
CG12470-RA	CG12470	-1.01	0.001	CG8566-RD	unc-104	-1.20	0.001
CG17975-RA	sut2	-1.37	0.001	CG32227-RA	gogo	-1.46	0.001
CG18585-RA	CG18585	-1.43	0.001	CG13472-RA	CG13472	-1.29	0.001
CG5418-RA	CG5418	-1.26	0.001	CG10706-RA	SK	-1.03	0.001
CG14401-RA	CG14401	-1.03	0.001	CG7852-RC	CG7852	-1.15	0.001
CG30188-RA	CG34371	-1.58	0.001	CG7072-RA	CG34462	-1.21	0.001
CG18340-RA	Ucp4B	-1.07	0.001	CG5630-RA	CG5630	-1.01	0.001
CG30044-RA	s-cup	-1.11	0.001	CG14606-RA	CG14606	-1.17	0.001
CG32310-RA	zormin	-1.08	0.001	CG3502-RA	CG3502	-1.26	0.001
CG31038-RA	CG31038	-1.46	0.001	CG18420-RA	CG18420	-1.01	0.002
CG14820-RA	CG14820	-1.30	0.001	CG14938-RA	crol	-1.72	0.001
CG3246-RA	CG3246	-1.20	0.002	CG17894-RC	cnc	-1.24	0.001
CG18304-RA	CG18304	-1.19	0.001	CG3666-RA	Tsf3	-1.40	0.001
CG3153-RA	Npc2b	-1.04	0.001	CG11695-RA	CG11695	-1.04	0.001
CG13540-RA	CG13540	-1.06	0.001	CG17599-RA	CG17599	-1.01	0.001
CG16914-RA	Lcp9	-1.01	0.001	CG32320-RA	CG32320	-1.08	0.001
CG6494-RA	h	-1.36	0.001	CG7950-RB	CG7950	-1.43	0.001
CG17932-RB	Ugt36Bc	-1.16	0.001	CG31676-RA	CG31676	-1.08	0.001
CG10909-RA	CG10909	-1.29	0.001	CG11211-RA	CG11211	-1.42	0.001
CG32372-RA	ltl	-1.53	0.001	CG33256-RA	Lmpt	-1.09	0.001
CG4318-RA	CG4318	-1.13	0.001	CG30091-RA	CG30091	-1.16	0.001
CG7059-RC	CG7059	-1.22	0.001	CG33129-RC	CG33129	-1.28	0.001
CG12899-RA	CG42732	-1.08	0.001	CG32156-RC	Mbs	-1.30	0.001

Continued on next page

Table 8 – continued from previous page

		Dichaete mutant				*Dichaete* mutant	
Transcript	**Gene**	**M**	**p**	**Transcript**	**Gene**	**M**	**p**
CG1150-RA	Osi3	-1.00	0.001	CG14947-RA	CG14947	-1.07	0.001
CG5618-RA	CG5618	-1.51	0.001	CR32777-RA	roX1	-2.25	0.003
CG17560-RA	CG17560	-1.31	0.001	CG13597-RA	CG13597	-1.18	0.001
CG32235-RA	CG32235	-1.07	0.001	CG14902-RA	decay	-1.32	0.001
CG13614-RA	CG13614	-1.24	0.001	CG18012-RA	CG18012	-1.18	0.001
CG17758-RA	CG42265	-1.04	0.001	CG31240-RA	repo	-1.16	0.001
CG14324-RA	CG14324	-1.16	0.001	CG5375-RA	CG5375	-1.16	0.001
CG18368-RA	CG18368	-1.18	0.001	CG13459-RA	cp309	-1.06	0.001
CG18528-RA	CG18528	-1.38	0.001	CG32388-RA	CG32388	-1.18	0.001
CG5322-RA	CG5322	-1.14	0.001	CG17870-RB	14-3-3zeta	-1.34	0.001
CG8093-RA	CG8093	-1.20	0.001	CG4772-RA	Ugt86Dh	-1.20	0.001
CG12162-RB	CG12162	-1.02	0.001	CG6666-RA	SdhC	-1.20	0.001
CG17453-RA	Cyp317a1	-1.03	0.001	CG11354-RA	Lim1	-1.05	0.006
CG15014-RA	CG15014	-1.32	0.001	CG10283-RA	CG10283	-1.60	0.001
CG1994-RA	l(1)G0020	-1.61	0.001	CG32261-RA	Gr64a	-1.18	0.001
CG15404-RA	CG15404	-1.02	0.001	CG3541-RC	pio	-1.13	0.001
CG1771-RA	mew	-1.57	0.001	CG14911-RA	CG42747	-1.01	0.001
CG4402-RA	lox2	-1.15	0.001	CG14695-RA	CG14695	-1.44	0.001
CG10864-RA	CG10864	-1.16	0.001	CG5553-RA	DNApol-α60	-1.29	0.001
CG1799-RA	ras	-1.19	0.001	CG3473-RA	CG3473	-1.05	0.001
CG13636-RA	CG13636	-1.22	0.001	CG33150-RA	Gr59f	-1.07	0.002
CG11581-RA	CG11581	-1.11	0.001	CG5313-RA	RfC3	-1.45	0.001
CG6571-RD	rdgC	-1.14	0.001	CG33333-RA	CG33333	-1.15	0.001
CG7542-RA	CG7542	-1.20	0.001	CG17819-RA	CG17819	-1.38	0.001
CG31446-RA	CG31446	-1.10	0.001	CG18808-RA	CG18808	-1.10	0.001
CG10674-RA	CG10674	-1.63	0.001	CG14829-RA	CG14829	-1.03	0.001
CG14548-RA	HLHmbeta	-1.78	0.001	CG5669-RA	Spps	-1.41	0.001
CG31663-RA	CG31663	-1.41	0.001	CG6456-RA	Mip	-1.03	0.001
CG9861-RA	CG9861	-1.15	0.001	CG5103-RA	CG5103	-1.30	0.001
CG18455-RA	Optix	-1.28	0.001	CG11020-RA	nompC	-1.27	0.001
CG1486-RB	CG1486	-1.08	0.001	CG7610-RA	ATPsynγ	-1.00	0.003
CG5405-RA	KrT95D	-1.12	0.001	CG4420-RA	rngo	-1.37	0.001
CG32688-RA	Hk	-1.03	0.001	CG6972-RA	CG6972	-1.20	0.001
CG12439-RA	CG12439	-1.28	0.001	CG13047-RA	CG13047	-1.55	0.001
CG4459-RA	CG4459	-1.16	0.001	CG13872-RA	CG13872	-1.20	0.001
CG6704-RA	CG6704	-1.32	0.001	CG18669-RA	CG42811	-1.24	0.001
CG13919-RA	CG13919	-1.34	0.001	CG7538-RA	Mcm2	-1.04	0.003
CG12009-RA	CG12009	-1.12	0.001	CG14377-RA	CG14377	-1.17	0.001
CG13474-RA	CG13474	-1.25	0.001	CG10164-RA	CG10164	-1.13	0.001
CG14493-RA	dpr13	-1.35	0.001	CG31973-RB	Cda5	-1.05	0.001
CG6718-RC	iPLA2-VIA	-1.26	0.001	CG10725-RB	CG10725	-1.08	0.001
CG9206-RA	Gl	-1.02	0.001	CG8956-RD	Ahcy89E	-1.43	0.001
CG5932-RA	CG5932	-1.11	0.001	CG8827-RA	Ance	-2.07	0.002
CG9265-RA	CG9265	-1.08	0.001	CG9257-RA	CG9257	-1.28	0.001
CG32594-RB	be	-1.47	0.001	CG4580-RA	CG4580	-1.05	0.001
CG5983-RA	ACXC	-1.00	0.001	CG12602-RA	Vha100-5	-1.22	0.001
CG18249-RA	CG18249	-1.26	0.001	CR32862-RA	U1:82Eb	-1.57	0.001
CG11280-RA	trn	-2.21	0.001	CG16712-RA	CG16712	-1.31	0.001
CG8191-RA	CG8191	-1.19	0.001	CG12139-RB	Megalin	-1.92	0.001
CG9483-RA	CG9483	-1.19	0.001	CG3964-RA	CG3964	-1.08	0.002
CG4955-RA	CG4955	-1.10	0.001	CG12832-RA	Tsp42Eq	-1.12	0.001
CG5434-RA	Srp72	-1.64	0.001	CG11317-RA	CG11317	-1.10	0.001
CG8094-RA	Hex-C	-1.29	0.001	CG10068-RA	CG10068	-1.24	0.001

Continued on next page

Table 8 – continued from previous page

		Dichaete mutant				*Dichaete* mutant	
Transcript	**Gene**	**M**	**p**	**Transcript**	**Gene**	**M**	**p**
CG11154-RA	ATPsyn-β	-1.81	0.001	CG5404-RA	CG5404	-1.43	0.001
CG5992-RA	Adgf-A	-1.47	0.001	CG6667-RC	dl	-1.04	0.001
CG18228-RA	CR18228	-1.08	0.001	CG5204-RA	CG5204	-1.06	0.001
CR31400-RB	Hsromega	-1.79	0.001	CG8864-RA	Cyp28a5	-1.12	0.001
CG5461-RC	bun	-1.15	0.001	CG5116-RA	CG5116	-1.23	0.001
CG3935-RA	al	-1.36	0.001	CG12730-RA	CG12730	-1.06	0.001
CG16885-RA	CG16885	-1.03	0.001	CG32806-RB	CG32806	-1.10	0.001
CG10589-RA	CG10589	-1.14	0.001	CG16863-RA	CG16863	-1.16	0.001
CG12002-RE	Pxn	-1.24	0.001	CG13211-RB	CG42336	-1.12	0.001
CG30121-RA	CG30121	-1.04	0.001	CG30159-RA	CG30159	-1.14	0.001
CG30110-RA	CG30110	-1.00	0.001	CG8492-RA	CG8492	-1.00	0.001
CG3978-RA	pnr	-1.76	0.001	CG16793-RA	CG16793	-1.04	0.001
CG7142-RA	CG7142	-1.04	0.001	CG13693-RA	CG13693	-1.04	0.001
CG4535-RA	FKBP59	-1.09	0.001	CG7279-RA	Lip1	-1.24	0.001
CG6724-RA	CG6724	-1.10	0.001	CG8202-RA	CG8202	-1.21	0.001
CG9342-RA	Mtp	-1.13	0.001	CG6719-RA	CG6719	-1.30	0.001
CG3298-RB	JhI-1	-1.11	0.001	CG17611-RA	eIF6	-1.37	0.002
CG6168-RB	CG6168	-1.08	0.001	CG16960-RA	Or33a	-1.13	0.001
CG7936-RA	mex1	-1.23	0.001	CG8228-RA	Vps45	-1.04	0.001
CG16931-RA	Eig71Ea	-1.17	0.001	CG1046-RA	zen	-1.10	0.001
CG13293-RA	CG13293	-1.27	0.001	CG12692-RB	CG12692	-1.01	0.002
CG13307-RA	CG34426	-1.08	0.001	CG3694-RA	Gγ30A	-1.11	0.001
CG9847-RA	Fkbp13	-1.17	0.001	CG8258-RA	CG8258	-1.27	0.001
CG7902-RA	bap	-1.56	0.005	CG2209-RA	CG2209	-1.05	0.001
CG6177-RA	ldlCp	-1.39	0.002	CG6792-RA	CG6792	-1.01	0.002
CG13394-RB	CG42820	-1.05	0.001	CG3117-RA	CG3117	-1.06	0.001
CG7764-RA	mrn	-1.29	0.002	CG31783-RA	ninaD	-1.36	0.001
CG8587-RA	Cyp301a1	-1.20	0.001	CG17199-RA	CG17199	-1.28	0.001
CG8340-RA	128up	-1.41	0.002	CG6874-RA	HipHop	-1.38	0.001
CG17234-RA	CG17234	-1.12	0.001	CG12284-RA	th	-1.75	0.002
CG2412-RA	Rad51C	-1.03	0.001	CG5867-RA	CG5867	-1.35	0.001
CG15636-RA	HP6	-1.44	0.001	CG17534-RA	GstE9	-1.52	0.001
CG6836-RA	CG6836	-1.00	0.001	CG7144-RA	LKR	-1.39	0.001
CG10303-RA	Osi4	-1.03	0.001	CG7798-RA	CG7798	-1.17	0.001
CG15138-RA	beat-IIIc	-1.27	0.001	CG9063-RA	CG9063	-1.09	0.002
CG4914-RA	CG4914	-1.21	0.001	CG33103-RA	Ppn	-1.73	0.004
CG31120-RA	CG31120	-1.08	0.003	CG8323-RA	CG8323	-1.16	0.001
CG4952-RE	dac	-1.26	0.001	CG32160-RA	CG32160	-1.35	0.001
CG1104-RA	CG1104	-1.40	0.001	CG33062-RA	CG33062	-1.26	0.001
CG7851-RA	Scgalpha	-1.15	0.001	CG8376-RA	ap	-1.34	0.001
CG10345-RA	CG10345	-1.42	0.001	CG12263-RA	CG12263	-1.02	0.001
CG7884-RA	CG7884	-1.43	0.003	CG8425-RA	Jhe	-1.26	0.001
CG3781-RA	CG3781	-1.38	0.001	CG3734-RA	CG3734	-1.10	0.001
CG13895-RA	CG13895	-1.31	0.001	CG18130-RA	CG18130	-1.06	0.001
CG15148-RA	btv	-1.24	0.001	CG4154-RC	Gyc88E	-1.00	0.001
CG17145-RA	CG17145	-1.03	0.001	CG2720-RA	Hop	-1.48	0.001
CG6330-RA	CG6330	-1.07	0.001	CG16952-RC	CG16952	-1.51	0.002
CG10318-RA	NC2alpha	-1.14	0.002	CG10226-RA	CG10226	-1.11	0.001
CG13226-RA	CG13226	-1.23	0.001	CG12063-RA	mey	-1.28	0.001
CG7239-RA	CG7239	-1.23	0.001	CG32402-RA	Or65b	-1.36	0.001
CG9663-RA	CG9663	-1.23	0.002	CG5005-RA	HLH54F	-1.17	0.001
CG4608-RB	bnl	-1.09	0.001	CG5790-RA	CG5790	-1.35	0.001
CG14745-RA	PGRP-SC2	-1.23	0.002	CG7893-RB	vav	-1.38	0.001

Continued on next page

Table 8 – continued from previous page

		Dichaete mutant				*Dichaete* mutant	
Transcript	**Gene**	**M**	**p**	**Transcript**	**Gene**	**M**	**p**
CG17031-RA	ref2	-1.09	0.001	CG5927-RA	Her	-1.00	0.001
CG4889-RA	wg	-1.69	0.001	CG33127-RA	CG33127	-1.11	0.001
CG11907-RA	Ent1	-1.37	0.008	CG2101-RA	mRpS35	-1.51	0.001
CG15743-RA	CG15743	-1.04	0.004	CG17279-RA	CG17279	-1.24	0.001
CG9044-RA	CG9044	-1.23	0.001	CG31548-RA	CG31548	-1.09	0.002
CG3048-RA	Traf4	-1.93	0.003	CG6736-RA	Ilp4	-1.82	0.002
CG14496-RA	CG34386	-1.28	0.001	CG2112-RA	CG42233	-1.12	0.001
CG15845-RB	Adf1	-1.56	0.005	CG8704-RA	dpn	-1.70	0.002
CG10362-RA	CG10362	-1.08	0.001	CG14098-RA	Rcd7	-1.11	0.001
CG13081-RA	CG13081	-1.11	0.001	CG1999-RA	CG1999	-1.12	0.001
CG32694-RB	CG32694	-1.22	0.001	CG33493-RA	CG33493	-1.26	0.001
CG11911-RA	CG11911	-1.20	0.001	CG11489-RB	srpk79D	-1.02	0.001
CG3972-RA	Cyp4g1	-1.03	0.001	CG1137-RA	CG1137	-1.23	0.001
CG15749-RA	dmrt11E	-1.25	0.001	CG14252-RA	CG14252	-1.12	0.001
CG11763-RA	RanBPM	-1.54	0.001	CG9919-RA	CG9919	-1.27	0.001
CG14085-RA	CG14085	-1.12	0.001	CG33200-RA	CG33200	-1.73	0.001
CG15286-RA	CG15286	-1.06	0.001	CG33194-RA	CheA29a	-1.02	0.002
CG15036-RA	CG15036	-1.02	0.001	CG8745-RA	CG8745	-1.46	0.006
CG5779-RA	proPO-A1	-1.39	0.001	CG32302-RA	CG32302	-1.01	0.002
CG13029-RB	CG13029	-1.34	0.004	CG31059-RA	Gr98b	-1.25	0.001
CG6292-RA	CycT	-1.14	0.004	CG33467-RA	CG33467	-1.04	0.001
CG31385-RA	Neu3	-1.29	0.001	CG32088-RA	CG32088	-1.00	0.001
CG5506-RA	CG5506	-1.30	0.001	CG7369-RA	CG7369	-1.11	0.001
CG14084-RB	Bet1	-1.64	0.001	CG8713-RA	CG8713	-1.12	0.001
CG1742-RB	Mgstl	-1.55	0.001	CG3536-RA	CG42260	-1.01	0.001
CG15196-RA	CG15196	-1.03	0.001	CG31522-RA	CG31522	-1.59	0.001
CG31012-RB	cindr	-1.16	0.001	CG3818-RA	Cpr30B	-1.29	0.001
CG12159-RA	CG12159	-1.12	0.001	CG6038-RA	crim	-1.91	0.001
CG9742-RA	SmG	-1.62	0.001	CG32137-RA	CG32137	-1.26	0.001
CG6024-RA	CG6024	-1.23	0.001	CG7363-RA	w-cup	-1.23	0.001
CG7336-RA	Eig71Eg	-1.12	0.001	CG6327-RA	CG6327	-1.14	0.005
CG14468-RA	Tsp42A	-1.34	0.001	CG13458-RA	CG13458	-1.07	0.001
CG14975-RA	Rdh	-1.11	0.001	CG11593-RA	CG11593	-1.29	0.001
CG6130-RA	h-cup	-1.35	0.001	CG31697-RA	CG31697	-1.07	0.001
CG16704-RA	CG16704	-1.16	0.001	CG33225-RA	CG33225	-1.35	0.001
CG8652-RA	Ugt37c1	-1.03	0.001	CG3948-RB	zetaCOP	-1.31	0.001
CG15825-RB	fon	-1.08	0.001	CG9001-RA	ste24b	-1.24	0.001
CG3166-RB	aop	-1.09	0.001	CG6416-RB	Zasp66	-1.00	0.001
CG9614-RI	pip	-1.33	0.001	CG9870-RA	CG9870	-1.26	0.001
CG15127-RA	CG15127	-1.13	0.001	CG11768-RA	CG11768	-1.16	0.001
CG11582-RA	srw	-1.11	0.002	CG11915-RA	Lmpt	-1.01	0.001
CG33083-RA	Gr97a	-1.23	0.001	CG31343-RA	CG31343	-1.07	0.001
CG15468-RA	SIP3	-1.19	0.001	CG9033-RB	Tsp47F	-1.05	0.001
CG32117-RA	CG32117	-1.20	0.001	CG31087-RA	CG31087	-1.12	0.001
CG33285-RA	CG33285	-1.20	0.001	CG6123-RA	CG6123	-1.14	0.001
CG30033-RB	CG30033	-1.24	0.001	CG13022-RA	CG13022	-1.42	0.001
CG14718-RA	CG14718	-1.02	0.001	CG4054-RA	Fili	-1.18	0.001
CG7404-RA	ERR	-1.37	0.001	CG33232-RC	CG33232	-1.66	0.001
CG14180-RA	CG14180	-1.09	0.001	CG15022-RA	CG15022	-1.18	0.001
CG1539-RB	tmod	-1.62	0.001	CG2017-RB	CG2017	-1.06	0.004
CG11473-RA	CG42492	-1.07	0.001	CG15897-RA	wuho	-1.16	0.001
CG12289-RA	CG12289	-1.08	0.001	CG17717-RA	CG17717	-1.01	0.001
CG9538-RA	Ag5r	-1.07	0.001	CG11101-RA	pwn	-1.11	0.001

Continued on next page

Table 8 – continued from previous page

		Dichaete mutant				*Dichaete* mutant	
Transcript	**Gene**	**M**	**p**	**Transcript**	**Gene**	**M**	**p**
CG31157-RA	CG31157	-1.20	0.001	CG16854-RA	CG16854	-1.07	0.001
CG7105-RA	Proct	-1.12	0.001	CG10286-RA	CG10286	-1.76	0.001
CG31221-RA	CG31221	-1.10	0.001	CG32543-RA	CG42450	-1.16	0.001
CG7830-RA	CG7830	-1.44	0.001	CG10486-RA	CG10486	-1.21	0.001
CG8511-RA	Cpr49Ag	-1.14	0.001	CG7956-RA	CG7956	-1.58	0.001
CG4592-RA	CG4592	-1.22	0.001	CG2827-RA	Tal	-1.05	0.001
CG17044-RA	yellow-e2	-1.65	0.001	CG12912-RA	CG12912	-1.13	0.001
CG12842-RA	CR12842	-1.29	0.001	CG8331-RA	CG8331	-1.59	0.001
CG11007-RA	CG11007	-1.37	0.001	CG10585-RA	CG10585	-1.10	0.001
CG6854-RA	CTPsyn	-1.12	0.001	CG2174-RA	Myo10A	-1.69	0.001
CG15732-RA	Ir11a	-1.17	0.001	CG15602-RA	CG15602	-1.20	0.004
CG31733-RA	ms(2)35Ci	-1.30	0.001	CG3252-RA	NAAT1	-1.04	0.001
CG13418-RA	RpI12	-1.16	0.001	CG3119-RA	CG3119	-1.28	0.001
CG1893-RA	scramb2	-1.46	0.001	CG13492-RB	CG13492	-1.10	0.001
CG8006-RA	CG8006	-1.25	0.001	CG8667-RA	dimm	-1.14	0.001
CG13283-RA	CG13283	-1.25	0.001	CG15269-RA	CG15269	-1.14	0.001
CG8400-RB	casp	-1.44	0.001	CG10956-RA	Spn53F	-1.13	0.001
CG17753-RA	CCS	-1.03	0.001	CG12825-RA	CG12825	-1.04	0.001
CG6225-RA	CG6225	-1.32	0.001	CG18681-RA	epsilonTry	-1.05	0.001
CG14949-RA	CG14949	-1.13	0.001	CG32244-RB	NT1	-1.04	0.001
CG3764-RA	CG3764	-1.41	0.001	CG12236-RA	CG12236	-1.21	0.001
CG12307-RA	CG12307	-1.21	0.001	CG31721-RA	Trim9	-1.74	0.001
CG8238-RA	Buffy	-1.04	0.001	CG13906-RA	nerfin-1	-1.83	0.001
CG6301-RA	CG34460	-1.07	0.001	CG13978-RA	CG13978	-1.01	0.001
CG17124-RA	CG17124	-1.28	0.001	CG11201-RA	TTLL3B	-1.12	0.001
CG4575-RA	CR4575	-1.64	0.001	CG15155-RA	CG15155	-1.01	0.001
CG2657-RA	Ir21a	-1.00	0.001	CG8345-RA	Cyp6w1	-1.01	0.002
CG4132-RA	pkaap	-1.32	0.001	CG12345-RB	Cha	-1.21	0.001
CG10414-RA	Atac2	-1.59	0.001	CG7908-RA	Tace	-1.31	0.001
CG8881-RA	skpB	-1.06	0.001	CG15594-RA	CG15594	-1.13	0.002
CG17167-RA	CG17167	-1.17	0.001	CG13927-RA	GC	-1.02	0.001
CG8980-RA	NiPp1	-1.28	0.001	CG7742-RA	CG7742	-1.19	0.001
CG16936-RA	CG16936	-1.57	0.001	CG4261-RA	Hel89B	-1.75	0.001
CG17213-RA	Gr33a	-1.35	0.001	CG3823-RA	CG3823	-1.00	0.001
CG31286-RB	CG31286	-1.00	0.001	CG13377-RA	CG13377	-1.08	0.002
CG10581-RA	CG10581	-1.38	0.001	CG10247-RA	Cyp6a21	-1.14	0.001
CG14655-RA	CG14655	-1.00	0.001	CG4624-RA	VhaAC39-2	-1.10	0.001
CG10279-RB	Rm62	-1.99	0.001	CG31792-RA	CG31792	-1.07	0.001
CG15919-RA	CG15919	-1.01	0.001	CG3858-RA	gcm2	-1.23	0.001
CG33141-RA	sns	-1.76	0.001	CG15125-RA	CG15125	-1.18	0.001
CG11742-RA	Or85d	-1.21	0.001	CG11876-RD	CG11876	-1.20	0.001
CG30145-RA	Obp57e	-1.21	0.001	CG17323-RA	CG17323	-1.44	0.001
CG17301-RA	Prosβ4R1	-1.05	0.001	CG11887-RA	Elp2	-1.73	0.001
CG10491-RA	vn	-1.04	0.001	CG5998-RA	Adgf-B	-1.21	0.001
CG7102-RA	CG7102	-1.00	0.001	CG15883-RA	Obp18a	-1.24	0.001
CG3164-RC	CG3164	-1.49	0.001	CG13155-RA	CG13155	-1.17	0.001
CG2098-RA	ferrochelatase	-1.14	0.001	CG9673-RA	CG9673	-1.07	0.001
CG13095-RA	Bace	-1.03	0.002	CG33457-RA	CheB53a	-1.34	0.001
CG9780-RA	CG9780	-1.02	0.001	CG3360-RA	Cyp313a1	-1.04	0.001
CG12004-RA	CG12004	-1.45	0.001	CG7450-RB	CrebA	-1.69	0.001
CG6223-RA	betaCop	-1.50	0.001	CG31467-RA	CG31467	-1.13	0.001
CG33207-RA	pxb	-1.26	0.001	CG5653-RA	CG5653	-1.29	0.001
CG15425-RA	CG15425	-1.08	0.001	CG6116-RA	CG6116	-1.48	0.001

Continued on next page

Table 8 – continued from previous page

		Dichaete mutant				*Dichaete* mutant	
Transcript	**Gene**	**M**	**p**	**Transcript**	**Gene**	**M**	**p**
CG11348-RB	nAcRβ-64B	-1.04	0.002	CG1598-RA	CG1598	-1.61	0.001
CG15479-RA	CG15479	-1.00	0.002	CG31411-RA	CG34409	-1.33	0.001
CG7744-RA	CG7744	-1.01	0.001	CG4124-RA	PNUTS	-1.60	0.001
CG9841-RA	EfSec	-1.00	0.003	CG33320-RA	CheB38a	-1.00	0.003
CG3096-RA	Brd	-2.13	0.001	CG15878-RA	CG15878	-1.03	0.001
CG9768-RA	hkb	-1.43	0.001	CG6057-RA	SMC1	-1.63	0.001
CG1470-RA	Gycβ100B	-1.00	0.001	CG15623-RA	c-cup	-1.24	0.001
CG2086-RA	drpr	-1.18	0.001	CG15475-RA	CG15475	-1.03	0.002
CG31948-RB	CG31948	-1.05	0.003	CG6096-RA	HLHm5	-1.13	0.001
CG15449-RA	CG15449	-1.00	0.002	CG8676-RC	Hr39	-1.16	0.001
CG8639-RA	Cirl	-1.27	0.008	CG9025-RA	Fem-1	-1.11	0.002
CG9628-RA	CG9628	-1.76	0.001	CG14233-RA	meso18E	-1.54	0.001
CG11614-RA	nkd	-1.73	0.001	CG3340-RA	Kr	-1.53	0.001
CG6099-RA	m4	-1.91	0.001	CG2812-RA	CG2812	-1.28	0.001
CG2046-RA	CG2046	-1.11	0.001	CG17258-RA	CG17258	-1.10	0.001
CG32201-RA	CG32201	-1.07	0.001	CG18540-RA	CG18540	-1.29	0.001
CG13441-RA	Gr57a	-1.10	0.001	CG5394-RA	Aats-glupro	-1.83	0.001
CG10620-RA	Tsf2	-1.40	0.004	CG4195-RA	l(3)73Ah	-1.31	0.009
CG4706-RA	CG4706	-1.02	0.001	CG1435-RA	CBP	-1.95	0.002
CG2901-RA	CG2901	-1.24	0.001	CG7047-RA	Vdup1	-1.33	0.006
CG13836-RA	CG13836	-1.06	0.001	CG14413-RA	mRpS25	-1.56	0.001
CR32874-RA	Me18S-G1620	-1.13	0.001	CG12689-RA	CG12689	-1.04	0.001
CG31748-RA	Gr36c	-1.29	0.001	CG14855-RA	CG14855	-1.21	0.002
CG15326-RA	Ir7b	-1.04	0.001	CG31321-RB	CG31321	-1.06	0.001
CG1152-RA	Gld	-1.00	0.001	CG3036-RA	CG3036	-1.05	0.002
CG6531-RA	wgn	-1.43	0.002	CG10746-RA	fok	-1.42	0.001
CG12010-RA	CG12010	-1.50	0.001	CG13603-RA	CG13603	-1.10	0.001
CG3477-RA	Pxd	-1.31	0.001	CG3020-RA	Sytbeta	-1.25	0.001
CG7560-RA	CG7560	-1.21	0.001	CG14321-RA	CG14321	-1.93	0.002
CG14919-RA	Ast-C	-1.25	0.001	CG33329-RB	Sp212	-1.01	0.002
CG32250-RA	CG32250	-1.01	0.001	CG14767-RB	CG14767	-1.74	0.002
CG3635-RA	CG3635	-1.18	0.001	CG10737-RA	CG10737	-1.04	0.002
CG4717-RA	kni	-1.71	0.002	CG10799-RA	CG10799	-1.40	0.001
CG7329-RA	CG7329	-1.00	0.001	CG14687-RA	CG14687	-1.48	0.001
CG1495-RA	CaMKI	-1.37	0.001	CG4153-RA	eIF-2beta	-1.00	0.007
CG31469-RA	CG31469	-1.06	0.001	CG7527-RA	CadN2	-1.25	0.001
CG1744-RA	chp	-1.20	0.001	CG33462-RA	CG33462	-1.22	0.001
CG10915-RA	CG10915	-1.36	0.001	CG14535-RA	CG14535	-1.12	0.001
CG15889-RA	Ravus	-1.38	0.001	CG4821-RC	Tequila	-1.18	0.001
CG13229-RA	CG13229	-1.12	0.001	CG6464-RA	salm	-1.84	0.001
CG5333-RA	trus	-1.08	0.001	CG17907-RA	Ace	-1.14	0.001
CG6723-RA	CG6723	-1.48	0.002	CG32343-RA	Atac3	-1.34	0.001
CG11666-RA	CG11666	-1.04	0.003	CG31105-RA	CG31105	-1.07	0.001
CG17010-RA	CG17010	-1.23	0.001	CG12321-RA	CG12321	-1.28	0.001
CG3528-RA	CG3528	-1.13	0.001	CG11246-RA	Rpb8	-1.13	0.007
CG14629-RA	CG14629	-1.74	0.002	CG30015-RB	CG30015	-1.61	0.001
CG2931-RA	CG2931	-1.57	0.001	CG13567-RA	CG13567	-1.26	0.001
CG4789-RA	CG4789	-1.30	0.001	CG17976-RA	sut3	-1.07	0.001
CG32210-RA	l(3)76BDr	-1.28	0.001	CG14886-RA	Gyc-89Db	-1.47	0.001
CG3658-RA	CDC45L	-1.00	0.001	CG13726-RA	Or74a	-1.24	0.001
CG13385-RA	CG13385	-1.16	0.001	CG1782-RA	Uba1	-1.28	0.002
CG3578-RA	bi	-1.36	0.002	CG5346-RA	CG5346	-1.30	0.001
CG13424-RA	lms	-1.00	0.001	CG31299-RA	cu	-1.04	0.001

Continued on next page

Table 8 – continued from previous page

		Dichaete mutant				*Dichaete* mutant	
Transcript	**Gene**	**M**	**p**	**Transcript**	**Gene**	**M**	**p**
CG9735-RA	Aats-trp	-1.36	0.001	CG3663-RA	CG3663	-1.15	0.001
CG5282-RA	CG5282	-1.18	0.001	CG17836-RA	Xrp1	-1.19	0.001
CG4755-RA	RhoGAP92B	-1.13	0.001	CG17121-RA	CG17121	-1.27	0.001
CG10571-RA	ara	-1.53	0.001	CG10798-RA	dm	-1.41	0.001
CG18437-RA	CG18437	-1.01	0.004	CG3570-RA	CG3570	-1.01	0.001
CG3515-RA	CG3515	-1.26	0.001	CG11356-RA	CG11356	-1.25	0.001
CG17633-RA	CG17633	-1.17	0.001	CG14675-RA	glob3	-1.06	0.001
CG1074-RA	CG1074	-1.00	0.004	CG7924-RA	CG7924	-1.15	0.001
CG3389-RA	Cad88C	-1.00	0.001	CG13610-RA	Orct2	-1.32	0.008
CG14719-RA	I-t	-1.01	0.001	CG31119-RA	HdacX	-1.14	0.001
CG6188-RA	CG6188	-1.03	0.001	CG7698-RA	Cpsf73	-1.15	0.004
CG7000-RA	Snmp1	-1.21	0.001	CG6518-RA	inaC	-1.06	0.001
CG17784-RA	CG17784	-1.05	0.001	CG10587-RA	CG10587	-1.35	0.001
CG5888-RA	CG5888	-1.15	0.001	CG32048-RA	CG42673	-1.06	0.001
CG15007-RA	Cpr64Ab	-1.01	0.002	CG14351-RA	haf	-1.31	0.001
CG9880-RA	Or23a	-1.39	0.001	CG3187-RB	Sirt4	-1.25	0.002
CG4781-RA	CG4781	-1.15	0.001	CG14838-RA	CG14838	-1.16	0.001
CG15666-RA	CG15666	-1.04	0.001	CG2917-RA	Orc4	-1.26	0.002
CG31062-RA	side	-1.09	0.001	CG5382-RA	CG5382	-1.55	0.001
CG8800-RA	CG8800	-1.13	0.001	CG4382-RA	CG4382	-1.25	0.001
CG8910-RA	CG8910	-1.17	0.001	CG5429-RA	Atg6	-1.43	0.001
CG6901-RA	CG6901	-1.40	0.001	CG30275-RB	CG30275	-1.03	0.001
CG30345-RA	CG30345	-1.02	0.001	CG1804-RA	kek6	-1.10	0.001
CG12309-RA	CG12309	-1.22	0.001	CG12273-RA	angel	-1.29	0.001
CG10618-RA	CHKov1	-1.17	0.001	CG8892-RC	CG8892	-1.17	0.002
CG13841-RA	CG13841	-1.42	0.001	CG18268-RA	CG18268	-1.03	0.001
CG6687-RA	Spn88Eb	-1.04	0.001	CG18631-RA	CG43370	-1.09	0.001
CG13858-RA	CG34377	-1.00	0.001	CG30187-RA	CG30187	-1.10	0.002
CG6933-RC	CG6933	-1.12	0.001	CG7305-RA	Rim	-1.11	0.001
CG7448-RB	CG7448	-1.05	0.002	CR33258-RA	CG33258	-1.24	0.001
CG4665-RA	Dhpr	-1.36	0.001	CG10132-RA	CG10132	-1.18	0.004
CG4483-RA	CG4483	-1.22	0.001	CG5495-RA	Txl	-1.14	0.001
CG13427-RA	CG13427	-2.53	0.001	CG8422-RA	Dh44-R1	-1.17	0.001
CG4721-RA	CG4721	-1.23	0.001	CG14014-RA	CG14014	-1.37	0.001
CG10693-RA	slo	-1.15	0.001	CG30378-RA	CG30378	-1.16	0.001
CG12766-RA	CG12766	-1.06	0.001	CG8637-RA	trc	-1.39	0.001
CG5859-RA	CG5859	-1.10	0.001	CG12516-RA	CG12516	-1.26	0.001
CG13707-RA	CG13707	-1.02	0.003	CG8031-RA	CG8031	-1.26	0.003
CG5274-RA	CG5274	-1.09	0.002	CG4886-RA	cyp33	-1.59	0.001
CG13675-RA	CG13675	-1.14	0.001	CG7429-RA	CG7429	-1.17	0.001
CG30283-RA	CG30283	-1.07	0.001	CG5910-RA	CG5910	-1.35	0.001
CG18211-RA	betaTry	-1.45	0.001	CG1374-RA	tsh	-1.59	0.001
CG13232-RA	BBS4	-1.08	0.001	CG12116-RA	CG12116	-1.20	0.001
CR31931-RA	Gr22b	-1.29	0.001	CG9456-RA	Spn1	-1.13	0.001
CG3942-RA	CG3942	-1.16	0.001	CG13728-RA	CG13728	-1.24	0.001
CG4843-RC	Tm2	-1.18	0.001	CG30177-RA	CG30177	-1.00	0.001
CG3199-RA	CG3199	-1.07	0.001	CG5379-RA	CG5379	-1.06	0.001
CG12590-RA	CG12590	-1.29	0.001	CG18609-RA	CG18609	-1.24	0.001
CG14062-RA	CG14062	-1.01	0.001	CG10092-RA	CG10092	-1.24	0.001
CG5285-RA	CG5285	-1.00	0.001	CG1112-RA	alpha-Est7	-1.21	0.001
CG9333-RA	Oseg5	-1.33	0.001	CG5687-RA	CG5687	-1.04	0.001
CG18188-RA	Damm	-1.06	0.001	CG12700-RA	skpD	-1.10	0.001
CG5140-RA	nopo	-1.25	0.001	CG9962-RA	CG9962	-1.13	0.001

Continued on next page

Table 8 – continued from previous page

		Dichaete mutant				*Dichaete* mutant	
Transcript	**Gene**	**M**	**p**	**Transcript**	**Gene**	**M**	**p**
CG12369-RA	Lac	-1.09	0.002	CG5076-RA	elk	-1.13	0.001
CG18747-RA	CG18747	-1.11	0.001	CG12690-RA	CHES-1-like	-1.25	0.001
CG1683-RA	Ant2	-1.06	0.001	CG14333-RA	lute	-1.34	0.001
CG9470-RA	MtnA	-1.12	0.001	CG16827-RA	alphaPS4	-1.07	0.001
CG6508-RA	CG6508	-1.12	0.001	CG4337-RA	mtSSB	-1.50	0.001
CG32823-RB	Sdic3	-1.49	0.001	CG6793-RA	CG6793	-1.14	0.001
CG5558-RB	CG42359	-1.36	0.001	CG14137-RA	CG14137	-1.13	0.001
CG9429-RA	Crc	-2.22	0.002	CG13164-RC	SIP2	-1.12	0.001
CG30099-RA	CG30099	-1.19	0.001	CG3592-RA	CG3592	-1.05	0.001
CG11306-RA	CG11306	-1.26	0.002	CG10081-RA	CG10081	-1.21	0.001
CG8852-RA	CG8852	-1.12	0.001	CG10052-RA	Rx	-1.06	0.001
CG7199-RD	Hr78	-1.12	0.001	CG11175-RA	Rcd6	-1.24	0.001
CG16765-RB	ps	-2.08	0.001	CG13648-RA	tnc	-1.18	0.001
CG9730-RA	mRpL21	-1.04	0.001	CG5731-RA	CG5731	-1.10	0.002
CG3987-RA	CG3987	-1.50	0.001	CG7225-RA	wbl	-1.82	0.001
CG9416-RA	CG9416	-1.37	0.001	CG8958-RA	CG8958	-1.13	0.001
CG5697-RA	CG5697	-1.04	0.001	CG14107-RA	CG14107	-1.06	0.001
CR32886-RA	Uhg1	-1.43	0.001	CG10006-RA	CG10006	-1.01	0.001
CG9325-RC	hts	-1.14	0.004	CG5171-RA	CG5171	-2.15	0.001
CG5524-RA	CG5524	-1.19	0.001	CG9294-RB	CG9294	-1.35	0.001
CG32937-RA	FoxP	-1.12	0.002	CG6575-RA	glec	-1.27	0.001
CG13931-RA	CG13931	-1.38	0.001	CG8118-RA	mam	-1.26	0.001
CG1246-RB	CG1246	-1.34	0.001	CG30432-RA	CG30432	-1.14	0.001
CG9503-RA	CG9503	-1.21	0.001	CG17905-RA	ChLD3	-1.15	0.001
CG13518-RA	Obp58b	-1.17	0.001	CG5265-RA	CG5265	-1.02	0.001
CG14766-RA	CG34431	-1.05	0.001	CG14400-RA	CG14400	-1.09	0.001
CG7665-RB	Fsh	-1.18	0.001	CG13969-RA	bwa	-1.44	0.001
CG14437-RA	COQ7	-1.13	0.001	CG11143-RA	Inos	-1.77	0.001
CG11326-RA	Tsp	-1.11	0.001	CG18599-RA	CG18599	-1.21	0.001
CG15551-RA	Ctr1C	-1.27	0.001	CG11597-RA	CG11597	-1.29	0.001
CG30043-RA	CG30043	-1.29	0.001	CG14605-RB	CG14605	-1.38	0.001
CG1075-RA	CR1075	-1.30	0.001	CG3769-RA	CG3769	-1.23	0.001
CG13622-RA	CG13622	-1.17	0.001	CG32308-RA	CG42355	-1.04	0.001
CG10845-RA	CG10845	-1.09	0.001	CG14546-RA	CG14546	-1.16	0.001
CG14937-RA	CG14937	-1.49	0.001	CG17888-RC	Pdp1	-1.26	0.001
CG3665-RB	Fas2	-1.97	0.001	CG31537-RA	cno	-1.27	0.001
CG30074-RA	Obp50d	-1.39	0.001	CG14989-RB	CG14989	-1.00	0.002
CG6981-RB	CG6981	-1.51	0.001	CG10823-RB	SIFR	-1.32	0.001
CG10917-RA	fj	-1.77	0.001	CG10814-RA	CG10814	-1.02	0.001
CG7484-RB	CG7484	-1.67	0.001	CG14979-RA	Gr63a	-1.06	0.001
CG30421-RA	CG30421	-1.24	0.001	CG7551-RB	CG7551	-1.10	0.001
CG1455-RA	CanA1	-1.00	0.001	CG17343-RA	CG17343	-1.13	0.001
CG4908-RA	CG4908	-1.40	0.001	CG11407-RA	CG11407	-1.28	0.001
CG31668-RA	CG31668	-1.05	0.001	CG10908-RA	Der-1	-1.70	0.001
CG7031-RA	CG7031	-1.14	0.001	CG17916-RA	Or92a	-1.15	0.001
CG5751-RA	TrpA1	-1.18	0.001	CG31482-RA	CG31482	-1.16	0.001
CG17266-RA	CG17266	-1.01	0.001	CG7058-RA	CG7058	-1.18	0.001
CG33125-RA	CG33125	-1.10	0.001	CG32149-RC	RhoGAP71E	-1.05	0.001
CG6954-RA	CG6954	-1.64	0.001	CG1849-RA	run	-1.13	0.001
CG5614-RA	CG5614	-1.18	0.001	CG30030-RA	Gr47b	-1.12	0.001
CG5955-RA	CG5955	-1.14	0.001	CG4163-RA	Cyp303a1	-1.24	0.001
CG10851-RG	B52	-1.74	0.001	CG4812-RA	Ser8	-1.26	0.001
CG32970-RA	wb	-1.03	0.001	CG14901-RA	Gr89a	-1.19	0.001

Continued on next page

Table 8 – continued from previous page

		Dichaete mutant				*Dichaete* mutant	
Transcript	**Gene**	**M**	**p**	**Transcript**	**Gene**	**M**	**p**
CG11109-RA	CG11109	-1.52	0.001	CG17800-RE	Dscam	-1.15	0.001
CG14323-RA	CG14323	-1.12	0.001	CG5373-RA	Pi3K59F	-1.34	0.001
CG31362-RA	Jon99Ciii	-1.30	0.001	CG9806-RA	CG9806	-1.02	0.001
CG3989-RA	ade5	-1.35	0.001	CG32386-RA	corn	-1.63	0.001
CG5321-RA	CG5321	-1.14	0.001	CG5765-RA	Muc55B	-1.20	0.001
CG6643-RA	Esyt2	-1.02	0.001	CG17381-RA	Ir94g	-1.21	0.001
CG10233-RA	rtp	-1.18	0.001	CG1944-RA	Cyp4p2	-1.02	0.001
CG32259-RB	CG32259	-1.09	0.001	CG15399-RA	CG15399	-1.15	0.001
CG3541-RB	pio	-1.26	0.002	CG14694-RA	CG14694	-1.01	0.001
CG3427-RA	Epac	-1.22	0.001	CG30081-RA	Ir51b	-1.50	0.001
CG33148-RB	Mctp	-1.23	0.001	CG2182-RA	CG2182	-1.00	0.003
CG10752-RA	CG10752	-1.00	0.001	CG7794-RA	CG7794	-1.05	0.001
CG33324-RA	gpp	-1.06	0.003	CG7540-RA	M6	-1.42	0.001
CG18660-RC	Nckx30C	-1.18	0.001	CG14825-RA	BBS1	-1.09	0.001
CG31659-RA	CG31659	-1.35	0.001	CG31786-RA	CG42750	-1.17	0.001
CG32145-RA	ome	-1.26	0.001	CG14380-RA	CG14380	-1.25	0.001
CG9969-RA	Or63a	-1.08	0.001	CG9855-RA	CG9855	-1.31	0.002
CG31950-RA	CG31950	-1.46	0.001	CG18449-RA	CG18449	-1.09	0.002
CG32855-RA	CG32855	-1.27	0.001	CG8818-RA	CG8818	-1.58	0.001
CG9406-RA	CG9406	-1.07	0.001	CG8932-RA	CG42235	-1.00	0.001
CG9536-RA	CG9536	-1.08	0.001	CG12438-RA	CG12438	-1.17	0.001
CG12436-RA	CG43373	-1.20	0.001	CG32634-RA	Rbp1-like	-1.22	0.001
CG13046-RA	CG13046	-1.08	0.001	CG12424-RB	CG12424	-1.52	0.001
CG18296-RA	axo	-1.17	0.001	CG18662-RA	CG18662	-1.28	0.001
CG30468-RA	Ir52c	-1.00	0.001	CG32856-RA	CG32856	-1.29	0.001
CG10017-RA	CG34340	-1.21	0.001	CG12499-RA	CG12499	-1.16	0.001
CG1362-RB	cdc2rk	-1.11	0.001	CG12008-RC	kst	-1.42	0.001
CG4767-RA	Tektin-A	-1.42	0.001	CG4521-RA	mthl1	-1.05	0.001
CG5518-RA	sda	-1.08	0.001	CG10598-RA	CG10598	-1.26	0.001
CG13504-RA	CG34205	-1.43	0.001	CG31962-RA	Sr-CIII	-1.16	0.001
CG10722-RA	nesd	-1.42	0.001	CG3039-RA	ogre	-2.17	0.001
CG11162-RA	CG11162	-1.32	0.001	CG5929-RA	Ir40a	-1.10	0.003
CG32950-RB	CG14245	-1.07	0.001	CG9264-RB	tadr	-1.20	0.001
CG4427-RA	cbt	-1.85	0.001	CG12617-RA	CG12617	-1.23	0.001
CG4576-RA	CG4576	-1.26	0.001	CG1674-RB	CG1674	-1.09	0.001
CG11279-RA	CG11279	-1.39	0.001	CG8190-RA	eIF2B-γ	-1.20	0.001
CG30203-RA	CG30203	-1.44	0.001	CG32024-RA	CG32024	-1.38	0.001
CG13856-RA	CG13856	-1.31	0.001	CG32813-RA	CG32813	-1.40	0.001
CG1864-RB	Hr38	-1.20	0.001	CG12715-RA	CG12715	-1.09	0.001
CG14104-RA	CG14104	-1.16	0.001	CG13695-RB	gk	-1.26	0.001
CG8502-RC	Cpr49Ac	-1.00	0.001	CG6839-RA	CG6839	-1.30	0.001
CG10067-RA	Act57B	-2.32	0.001	CG31871-RA	CG31871	-1.12	0.001
CG6675-RA	CG6675	-1.29	0.001	CG18223-RB	CG18223	-1.14	0.001
CG16740-RA	Rh2	-1.22	0.001	CG32584-RB	CG42299	-1.27	0.001
CG5197-RA	CG34457	-1.10	0.001	CG9990-RB	CG9990	-1.06	0.001
CG5461-RA	bun	-1.15	0.001	CG30495-RA	CG30495	-1.01	0.001
CG5099-RA	msi	-1.05	0.001	CG1827-RB	CG1827	-1.18	0.001
CG3934-RA	Npc2c	-1.11	0.001	CG14398-RA	CG33510	-1.18	0.001
CG32643-RA	CG32643	-1.07	0.001	CG8795-RA	CG8795	-1.05	0.001
CG3011-RA	CG3011	-1.05	0.001	CG10143-RA	Adgf-E	-1.17	0.001
CG8874-RA	Fps85D	-1.50	0.001	CG12663-RA	Ir7a	-1.15	0.001
CG30106-RA	CCHa1r	-1.12	0.001	CG11212-RA	Ptr	-1.62	0.001
CG8544-RB	sd	-1.97	0.001	CG3977-RA	Ctr1A	-1.02	0.002

Continued on next page

Table 8 – continued from previous page

		Dichaete mutant				*Dichaete* mutant	
Transcript	**Gene**	**M**	**p**	**Transcript**	**Gene**	**M**	**p**
CG3050-RA	Cyp6d5	-1.93	0.001	CG8550-RA	CG8550	-1.01	0.001
CG8982-RA	Acp26Aa	-1.06	0.001	CG3921-RA	CG3921	-1.11	0.001
CG8197-RA	CG8197	-1.06	0.003	CG5065-RA	CG5065	-1.06	0.002
CG1806-RA	CG1806	-1.32	0.001	CG1921-RC	sty	-1.90	0.002
CG32954-RG	Adh	-1.17	0.001	CG10625-RC	CG10625	-1.15	0.001
CG6746-RA	CG6746	-1.61	0.001	CG32988-RA	CG32988	-1.35	0.001
CG17078-RA	CG17078	-1.48	0.001	CG31139-RA	CG31139	-1.17	0.002
CG10181-RA	Mdr65	-1.04	0.001	CG16710-RA	CG16710	-1.39	0.001
CG10309-RA	pad	-1.12	0.001	CG7807-RA	AP-2	-1.24	0.001
CG31749-RA	Fas3	-1.58	0.001	CG4959-RB	CG34168	-1.27	0.001
CG12161-RA	Prosβ2R2	-1.20	0.001	CG13337-RA	CG13337	-1.60	0.001
CG13325-RA	CG13325	-1.17	0.001	CG6744-RA	CG6744	-1.16	0.003
CG12680-RA	CG12680	-1.29	0.001	CG17186-RA	CG17186	-1.00	0.001
CG13306-RA	CG13306	-1.42	0.001	CG3691-RA	Pof	-1.12	0.001
CG13563-RA	CG13563	-1.02	0.001	CG10484-RA	Rpn3	-1.40	0.001
CG6829-RA	Ark	-1.13	0.001	CG17085-RA	CG43129	-1.06	0.001
CG6885-RA	CG6885	-1.30	0.001	CG6790-RA	CG6790	-1.33	0.001
CG32499-RA	Cda4	-1.38	0.001	CG7978-RA	Ac76E	-1.05	0.001
CG12975-RA	CG12975	-1.69	0.001	CG15506-RA	CG15506	-1.16	0.001
CG8590-RA	Klp3A	-1.57	0.002	CG7763-RA	CG7763	-2.13	0.001
CG14526-RA	CG14526	-1.46	0.001	CG13251-RA	CG13251	-1.18	0.001
CG31782-RB	CG31782	-1.17	0.001	CG15180-RA	glob2	-1.25	0.001
CG14001-RA	bchs	-1.28	0.001	CG4645-RA	CG4645	-1.37	0.003
CG2665-RA	PebII	-1.15	0.001	CG8337-RA	malpha	-2.33	0.001
CG17233-RC	CG17233	-1.51	0.001	CG7228-RA	pes	-1.24	0.001
CG8363-RC	Papss	-1.34	0.001	CG2861-RA	CG2861	-1.02	0.001
CG17725-RA	Pepck	-2.14	0.001	CG7177-RA	Wnk	-1.00	0.004
CG4673-RA	CG4673	-1.12	0.001	CG11037-RA	CG11037	-1.04	0.001
CG14122-RA	Smyd4	-1.24	0.001	CG2411-RA	ptc	-2.08	0.001
CG16874-RA	Vm32E	-1.11	0.001	CG1394-RA	CG1394	-1.05	0.001
CG32392-RA	CG32392	-1.37	0.001	CG5874-RA	Nelf-A	-1.62	0.001
CG15632-RA	Taf12L	-1.12	0.001	CG1387-RA	CG1387	-1.07	0.001
CG15877-RA	CG15877	-1.07	0.001	CG6269-RA	unc-4	-1.07	0.001
CG10302-RA	bsf	-1.61	0.001	CG32486-RD	CG32486	-1.81	0.005
CG1071-RA	E2f2	-1.80	0.001	CG7796-RA	dpr12	-1.03	0.001
CG4679-RA	CG4679	-1.40	0.002	CG32487-RA	CG32487	-1.18	0.001
CG8756-RC	verm	-1.15	0.001	CG1773-RA	CG1773	-1.34	0.001
CG8110-RB	syd	-1.18	0.001	CG2574-RA	CG2574	-1.30	0.001
CG15618-RA	CG15618	-1.28	0.003	CG8365-RA	E(spl)	-1.83	0.004
CG33101-RA	Nsf2	-1.13	0.001	CG10943-RA	CG10943	-1.21	0.001
CG8322-RB	ATPCL	-1.14	0.001	CG33060-RA	CG33060	-1.15	0.001
CG1703-RA	CG1703	-1.45	0.002	CG8374-RA	dmt	-1.26	0.004
CG10339-RA	CG10339	-1.00	0.001	CG4631-RA	CG4631	-1.16	0.001
CG7882-RA	CG7882	-1.05	0.001	CG8364-RA	Drep-3	-1.09	0.001
CG1555-RA	cn	-1.72	0.001	CG18335-RA	CG18335	-1.42	0.001
CG3173-RA	CG3173	-1.34	0.002	CG13893-RA	CG13893	-1.08	0.001
CG14770-RA	CG14770	-1.19	0.001	CG10317-RA	CG10317	-1.19	0.001
CG15227-RA	CG15227	-1.08	0.001	CG3796-RA	ac	-1.57	0.001
CG12911-RA	CG12911	-1.23	0.001	CG5835-RA	CG5835	-1.27	0.001
CG10170-RA	CG10170	-1.59	0.003	CG9836-RA	CG9836	-1.65	0.001
CG32434-RA	siz	-1.76	0.001	CG6733-RA	CG6733	-1.35	0.001
CG4090-RA	Mur89F	-1.13	0.001	CG14040-RA	CG14040	-1.02	0.001
CG10253-RA	CG10253	-1.24	0.001	CG13225-RA	Or47a	-1.01	0.001

Continued on next page

Table 8 – continued from previous page

		Dichaete mutant				*Dichaete* mutant	
Transcript	**Gene**	**M**	**p**	**Transcript**	**Gene**	**M**	**p**
CG12056-RA	CG12056	-1.04	0.001	CG32401-RA	Or65a	-1.25	0.001
CG4608-RA	bnl	-1.21	0.001	CG14740-RA	CG14740	-1.12	0.001
CG15695-RA	CG15695	-1.14	0.001	CG7127-RA	exo70	-1.28	0.001
CG5924-RA	CG5924	-1.01	0.001	CG15864-RA	CG15864	-1.03	0.001
CG4881-RA	salr	-1.20	0.001	CG14946-RA	CG14946	-1.56	0.001
CG31557-RA	Obp83ef	-1.25	0.001	CG1728-RA	Tim8	-1.07	0.001
CG17278-RA	CG17278	-1.77	0.001	CG9042-RC	Gpdh	-1.08	0.001
CG7371-RA	CG7371	-1.01	0.001	CG10463-RA	CG10463	-1.26	0.001
CG15646-RA	CG15646	-1.00	0.001	CG14479-RA	CG43164	-1.39	0.001
CG12096-RA	CG12096	-1.27	0.001	CG5017-RA	CG5017	-1.08	0.001
CG13117-RA	CG13117	-1.12	0.001	CG9653-RA	brk	-1.10	0.002
CG31091-RA	CG31091	-1.28	0.001	CG9706-RA	CG9706	-1.15	0.001
CG16719-RA	CG16719	-1.28	0.001	CG10359-RA	CG10359	-1.08	0.001
CG14096-RA	CG14096	-1.16	0.001	CG13079-RA	CG13079	-1.31	0.001
CG6073-RA	CG6073	-1.11	0.001	CG4110-RA	ppk16	-1.15	0.001
CG33492-RA	Ir41a	-1.15	0.001	CG11905-RF	CG11905	-1.33	0.001
CG7422-RA	Snmp2	-1.12	0.001	CG3969-RB	PR2	-1.47	0.001
CG15747-RA	CG15747	-1.07	0.001	CG14251-RA	ms(3)K81	-1.13	0.001
CG9918-RC	Pk1r	-1.39	0.001	CG33473-RB	luna	-1.74	0.001
CG31437-RA	CG31437	-1.24	0.001	CG13194-RA	pyr	-1.46	0.002
CG33193-RA	sav	-1.30	0.001	CG5342-RA	CG5342	-1.29	0.001
CG5718-RA	CG5718	-1.26	0.001	CG11327-RA	CG11327	-1.20	0.001
CG30380-RA	CG30380	-1.30	0.001	CG13028-RA	CG13028	-1.04	0.001
CG2331-RA	TER94	-1.06	0.001	CG2904-RB	ec	-1.03	0.001
CG31050-RA	CG31050	-1.04	0.001	CG11637-RA	CG11637	-1.19	0.002
CG33466-RA	Fs	-1.29	0.001	CG31380-RA	CG31380	-1.37	0.001
CG32085-RA	CG32085	-1.21	0.001	CG15615-RA	CG15615	-1.22	0.001
CG14080-RA	Mkp3	-1.30	0.001	CG8712-RB	CG8712	-1.34	0.001
CG7283-RB	RpL10Ab	-1.07	0.001	CG33293-RA	CG33293	-1.47	0.001
CG31623-RA	dtr	-1.19	0.001	CG15415-RA	Spindly	-1.42	0.001
CG11798-RB	chn	-1.57	0.001	CG3814-RA	CG3814	-1.49	0.001
CG17348-RA	drl	-2.01	0.001	CG6036-RA	CG6036	-1.12	0.001
CG32135-RA	Nxf3	-1.08	0.001	CG2140-RA	Cyt-b5	-1.48	0.001
CG7584-RA	Obp99c	-1.08	0.001	CG5987-RA	CG5987	-1.08	0.002
CG4029-RA	jumu	-1.93	0.001	CG14964-RA	CG14964	-1.17	0.001
CG2064-RA	CG2064	-1.47	0.001	CG11592-RA	CG11592	-1.26	0.001
CG6126-RA	CG6126	-1.48	0.001	CG9140-RA	CG9140	-1.26	0.008
CG9598-RA	bbg	-1.44	0.001	CG31693-RA	CG31693	-1.35	0.001
CG10477-RA	CG10477	-1.08	0.001	CG9472-RA	CG9472	-1.22	0.001
CG8651-RA	trx	-1.60	0.001	CG9852-RA	140up	-1.16	0.001
CG13188-RB	CG13188	-1.18	0.001	CG9610-RA	Poxm	-1.20	0.001
CG31666-RD	chinmo	-1.14	0.001	CG6410-RA	Snx16	-1.33	0.001
CG9614-RG	pip	-1.14	0.001	CG1698-RA	CG1698	-1.17	0.001
CG4829-RB	CG4829	-1.41	0.001	CG11761-RA	trsn	-1.65	0.002
CG8665-RA	CG8665	-1.43	0.001	CG1157-RA	Osi15	-1.24	0.001
CG11913-RA	CG11913	-1.29	0.001	CG32115-RA	CG32115	-1.11	0.001
CG31085-RA	CG33970	-1.15	0.001	CG32136-RB	Tsp68C	-1.10	0.001
CG33283-RA	CR33283	-1.34	0.001	CG5976-RA	isoQC	-1.14	0.001
CG13016-RA	CG13016	-1.18	0.001	CG14717-RA	CG14717	-1.08	0.001
CG4053-RA	CG4053	-1.13	0.001	CG7402-RA	CG7402	-1.38	0.001
CG14153-RA	CG14153	-1.43	0.001	CG3322-RA	LanB2	-2.39	0.001
CG31017-RA	PH4αNE3	-1.11	0.001	CG10393-RA	amos	-1.15	0.001
CG2540-RA	CG2540	-1.07	0.001	CG4099-RA	Sr-CI	-1.56	0.001

Continued on next page

Table 8 – continued from previous page

		Dichaete mutant				*Dichaete* mutant	
Transcript	**Gene**	**M**	**p**	**Transcript**	**Gene**	**M**	**p**
CG2016-RB	CG2016	-1.15	0.001	CG15526-RA	CG15526	-1.35	0.001
CG17330-RA	jhamt	-1.13	0.001	CG31509-RA	TotA	-1.10	0.001
CG31189-RA	CG31189	-1.11	0.001	CG17716-RA	fas	-1.12	0.001
CG17108-RA	CG17108	-1.00	0.001	CG11100-RB	Mes2	-1.36	0.001
CG7990-RB	CG7990	-1.32	0.001	CG7112-RA	CG7112	-1.07	0.001
CG17637-RA	CG17637	-1.08	0.001	CG7104-RA	spz3	-1.20	0.001
CG10278-RA	GATAe	-1.58	0.001	CG16791-RA	CG16791	-1.79	0.001
CG31219-RA	CG31219	-1.07	0.001	CG32542-RA	CG34401	-1.80	0.001
CG9480-RA	Glycogenin	-1.02	0.001	CG10371-RB	Plip	-1.17	0.001
CG8476-RA	CG8476	-1.10	0.001	CG14684-RA	CG14684	-1.25	0.001
CG10293-RA	how	-2.08	0.002	CG31764-RB	vir-1	-1.23	0.001
CG4590-RA	inx2	-1.18	0.003	CG2826-RA	lectin-21Ca	-1.35	0.001
CG18649-RA	CG18649	-1.00	0.001	CG13366-RA	CG13366	-1.13	0.002
CG1915-RC	sls	-1.16	0.001	CG33006-RD	CG43164	-1.75	0.001
CG4396-RA	fne	-1.19	0.001	CG14070-RA	CG14070	-1.16	0.001
CG13312-RA	CG13312	-1.27	0.001	CG4593-RA	CG4593	-1.18	0.001
CG12906-RA	Gr47a	-1.02	0.001	CG2184-RA	Mlc2	-1.36	0.001
CG13701-RA	skl	-1.49	0.001	CG14890-RA	CG42342	-1.15	0.001
CG10582-RA	Sin	-1.40	0.001	CG10236-RA	LanA	-2.27	0.001
CR33326-RA	mir-34	-1.24	0.001	CG15731-RA	CG15731	-1.01	0.001
CG17176-RA	ACXA	-1.13	0.001	CG8306-RA	CG8306	-1.83	0.001
CG3871-RA	Six4	-1.39	0.001	CG13417-RA	Gr93a	-1.13	0.001
CG2872-RB	AlstR	-1.12	0.001	CG13488-RA	CG13488	-1.07	0.001
CG8654-RA	CG8654	-1.28	0.001	CG7781-RA	CG7781	-1.32	0.001
CG31804-RA	CG31804	-1.30	0.001	CG33110-RA	CG33110	-1.05	0.001
CG2684-RA	lds	-1.34	0.001	CG17005-RA	CG17005	-1.35	0.001
CG8398-RA	CG8398	-1.13	0.001	CG6802-RA	Cyp313a4	-1.00	0.001
CG32415-RA	CG42272	-1.36	0.001	CG18174-RA	Rpn11	-1.66	0.001
CG7257-RA	Rpt4R	-1.06	0.001	CG2909-RA	CG2909	-1.08	0.001
CG17752-RA	CG17752	-1.06	0.001	CG12824-RA	CG12824	-1.01	0.001
CG3182-RB	sei	-1.09	0.001	CG7214-RA	CG7214	-1.04	0.001
CG11059-RA	cals	-1.94	0.001	CG8236-RA	CG8236	-1.02	0.001
CG16986-RA	CG16986	-1.01	0.001	CG14317-RA	CG14317	-1.60	0.002
CG32407-RA	CG32407	-1.04	0.001	CG5899-RA	Etl1	-1.48	0.001
CG5883-RA	CG5883	-1.25	0.001	CG13904-RA	CG34057	-1.09	0.001
CG31233-RA	CG31233	-1.04	0.001	CG6300-RA	CG6300	-1.35	0.001
CG8560-RA	CG8560	-1.14	0.001	CG17122-RA	CG17122	-1.17	0.001
CG10327-RC	TBPH	-1.92	0.004	CG11200-RA	CG11200	-1.26	0.001
CG9398-RA	king-tubby	-1.59	0.001	CG15154-RA	Socs36E	-1.90	0.001
CG9115-RA	mtm	-1.91	0.001	CG4933-RA	CG4933	-1.27	0.001
CG8344-RA	RpIII128	-1.14	0.001	CG4974-RA	dally	-1.60	0.001
CG11064-RA	Rfabg	-2.06	0.001	CG17657-RA	frtz	-1.01	0.004
CG4123-RA	Mipp1	-1.73	0.002	CG6234-RA	CG6234	-2.18	0.001
CG12784-RA	CG12784	-1.23	0.001	CG3245-RA	PpN58A	-1.18	0.001
CG12296-RA	klu	-1.45	0.001	CG7906-RA	CG7906	-1.01	0.002
CG12250-RA	ymp	-1.44	0.001	CG15593-RA	Osi10	-1.21	0.001
CG1844-RA	SelG	-1.38	0.001	CG2708-RA	unc-45	-1.44	0.001
CG15169-RA	CG15169	-1.15	0.001	CG13405-RA	CG13405	-1.17	0.001
CG4225-RA	Hmt-1	-1.31	0.001	CG15256-RA	CG15256	-1.11	0.001
CG17024-RA	CG17024	-1.40	0.001	CG17204-RA	mun	-1.16	0.001
CG8510-RA	Cpr49Af	-1.07	0.001	CG31284-RC	wtrw	-1.11	0.001
CG5853-RA	CG5853	-1.46	0.001	CG2794-RA	CG2794	-1.02	0.001
CG10197-RA	kn	-1.01	0.003	CG32451-RB	SPoCk	-1.48	0.001

Continued on next page

Table 8 – continued from previous page

		Dichaete mutant				*Dichaete* mutant	
Transcript	**Gene**	**M**	**p**	**Transcript**	**Gene**	**M**	**p**
CG6354-RB	Rb97D	-1.23	0.001	CG16969-RA	dgt2	-1.54	0.001
CG13362-RA	CG13362	-1.38	0.001	CG12143-RA	Tsp42Ej	-1.30	0.001
CG10246-RA	Cyp6a9	-1.05	0.001	CG13271-RA	Ugt36Bb	-1.04	0.001
CG12254-RA	MED25	-1.10	0.001	CG7266-RC	Eip71CD	-1.65	0.001
CG2331-RB	TER94	-1.43	0.001	CG8300-RA	CG8300	-1.42	0.001
CG13990-RA	Muc26B	-1.00	0.001	CG10251-RA	CG10251	-1.08	0.001
CG14787-RA	CG14787	-1.01	0.001	CG31791-RA	CG43339	-1.25	0.001
CG1691-RA	Imp	-1.55	0.001	CG4899-RA	Pdh	-1.01	0.001
CG15095-RA	l(2)08717	-1.51	0.002	CG2125-RA	ci	-2.10	0.001
CG14866-RA	CG14866	-1.11	0.001	CG17321-RA	CG17321	-1.38	0.001
CG6414-RA	CG6414	-1.04	0.001	CG17300-RA	CG17300	-1.40	0.001
CG9071-RB	NaCP60E	-1.01	0.001	CG31601-RA	CG31601	-1.42	0.001
CG1575-RA	CG1575	-1.20	0.001	CG1048-RA	zen2	-1.19	0.001
CG15371-RA	Gr8a	-1.00	0.001	CG7062-RA	Rab-RP3	-1.25	0.001
CG1213-RB	CG1213	-1.28	0.001	CG15881-RA	CG15881	-1.10	0.001
CG1316-RA	CG1316	-1.68	0.001	CG8785-RA	CG8785	-1.02	0.001
CG11552-RA	Rpn12R	-1.17	0.001	CG13094-RC	Dh31	-1.24	0.001
CG9731-RA	pyd	-1.24	0.001	CG32079-RA	CG32079	-1.54	0.001
CG1411-RB	CRMP	-1.70	0.001	CG15412-RA	CG15412	-1.11	0.001
CG9778-RA	Syt14	-1.06	0.001	CG4115-RA	CG4115	-1.34	0.001
CG3359-RD	mfas	-1.80	0.001	CG7449-RB	hbs	-1.81	0.001
CG11502-RB	svp	-1.21	0.001	CG4007-RA	Nrk	-1.07	0.001
CG6217-RA	knk	-1.17	0.001	CG9945-RA	CG9945	-1.33	0.001
CG31677-RA	CG31677	-1.39	0.001	CG32102-RA	CG32102	-1.29	0.002
CG3437-RA	CG3437	-1.11	0.001	CG31464-RA	CG31464	-1.24	0.001
CG13217-RA	CG13217	-1.50	0.001	CG15422-RA	CG15422	-1.29	0.001
CG8772-RB	CG42708	-1.06	0.001	CG6113-RA	Lip4	-1.74	0.001
CG15828-RB	CG15828	-1.72	0.001	CG14298-RA	CG14298	-1.01	0.001
CG3041-RA	Orc2	-1.44	0.001	CG40188-RC	CG40188	-1.12	0.001
CG2692-RA	gsb-n	-1.43	0.001	CG2977-RA	inx7	-1.25	0.001
CG4446-RA	CG34456	-1.33	0.001	CG9098-RA	CG9098	-1.16	0.001
CG33479-RA	clos	-1.08	0.001	CG3140-RA	Adk2	-1.20	0.001
CG1973-RA	yata	-1.59	0.001	CG5733-RA	Nup75	-1.63	0.001
CG14111-RA	CG14111	-1.26	0.001	CG1634-RB	Nrg	-1.35	0.004
CG15369-RA	CG15369	-1.08	0.001	CG32182-RA	CG32182	-1.68	0.001
CG17470-RA	CG17470	-1.08	0.001	CG32064-RA	S-Lap4	-1.46	0.001
CG9577-RA	CG9577	-1.13	0.001	CG11451-RA	Spc105R	-1.10	0.001
CG5820-RC	Gp150	-2.15	0.001	CG31028-RB	CG31028	-1.18	0.001
CG6476-RB	Su(var)3-9	-1.55	0.001	CG6052-RA	CG6052	-1.17	0.001
CG9766-RB	CG9766	-1.25	0.001	CG6043-RD	CG6043	-1.00	0.001
CG31421-RA	Takl1	-1.07	0.001	CG6050-RA	EfTuM	-1.18	0.001
CG1980-RB	dj	-1.16	0.001	CG14696-RA	CG14696	-1.11	0.001
CG32054-RA	CG32054	-1.39	0.001	CG2083-RA	CG2083	-2.02	0.001
CG9593-RA	CG9593	-1.15	0.001	CG11624-RB	Ubi-p63E	-1.61	0.001
CG6197-RA	CG6197	-1.29	0.001	CG15382-RA	CG15382	-1.49	0.001
CG9650-RA	CG9650	-1.72	0.001	CG31369-RA	CG31369	-1.18	0.001
CG1641-RA	sisA	-1.86	0.001	CG9550-RA	CG9550	-1.22	0.001
CG15546-RA	CG15546	-1.16	0.001	CG33253-RA	CG33253	-1.23	0.001
CG11851-RA	CG11851	-1.06	0.001	CG1138-RA	CG1138	-1.00	0.001
CG8676-RB	Hr39	-1.26	0.001	CG8638-RA	Cpr65Eb	-1.32	0.001
CG12252-RA	Fcp1	-1.09	0.001	CG13202-RA	CG13202	-1.02	0.001
CG9024-RB	Acp26Ab	-1.00	0.004	CG9623-RB	if	-1.13	0.001
CG9119-RA	CG9119	-1.04	0.001	CG31684-RA	CG34447	-1.31	0.001

Continued on next page

Table 8 – continued from previous page

		Dichaete mutant				*Dichaete* mutant	
Transcript	**Gene**	**M**	**p**	**Transcript**	**Gene**	**M**	**p**
CG6454-RA	CG6454	-1.20	0.001	CG9887-RA	VGlut	-1.35	0.001
CG5097-RA	MtnC	-1.51	0.001	CG11841-RA	CG11841	-1.42	0.001
CG11607-RA	H2.0	-1.20	0.001	CG11961-RA	CG11961	-1.36	0.001
CG1421-RA	CG1421	-1.07	0.001	CG9054-RA	Ddx1	-1.11	0.001
CG32141-RA	CG32141	-1.01	0.002	CG31100-RA	CG31100	-1.17	0.001
CG32382-RA	sphinx2	-1.41	0.001	CG6143-RA	Pep	-1.21	0.001
CG6006-RA	CG6006	-1.40	0.001	CG14893-RA	CG14893	-1.01	0.001
CG14173-RA	Ilp1	-1.08	0.001	CG10421-RA	Cad96Cb	-1.10	0.001
CG15152-RA	CG15152	-1.35	0.001	CG2808-RA	CG34340	-1.03	0.001
CG17364-RA	CG17364	-1.06	0.001	CG12120-RA	t	-1.04	0.001
CG15922-RA	CG15922	-1.17	0.001	CG11726-RA	CG11726	-1.26	0.001
CG3054-RA	l(2)k05819	-1.22	0.001	CG13478-RA	shd	-1.07	0.001
CG3219-RA	Klp59C	-1.02	0.001	CG13438-RA	CG13438	-1.25	0.002
CG4194-RA	CG4194	-1.73	0.001	CG4704-RA	CG4704	-1.25	0.001
CG7046-RA	CG7046	-1.52	0.001	CG11136-RA	Lrt	-1.42	0.001
CG15316-RA	Megalin	-1.52	0.001	CG31318-RB	Rpb4	-1.03	0.001
CG10002-RA	fkh	-2.04	0.001	CG11529-RA	CG11529	-1.09	0.001
CG30364-RA	hubl	-1.12	0.001	CG3790-RA	CG3790	-1.29	0.001
CG3526-RA	CG3526	-1.19	0.001	CG14570-RA	CG14570	-1.31	0.001
CG3699-RA	CG3699	-1.20	0.004	CG13602-RA	CG13602	-1.09	0.001
CG14585-RA	Ir75a	-1.15	0.001	CG3469-RA	betaggt-I	-1.09	0.001
CG32057-RC	dpr10	-1.02	0.001	CG14312-RA	CG14312	-1.04	0.001
CG14916-RA	Gr32a	-1.02	0.001	CG32177-RA	CG32177	-1.63	0.001
CG31454-RA	CG31454	-1.01	0.001	CG10794-RA	DptB	-1.01	0.001
CG7295-RA	CG7295	-1.29	0.001	CG14686-RA	Desi	-1.13	0.001
CG14959-RC	ckd	-1.30	0.001	CG33344-RA	CcapR	-1.20	0.001
CG10912-RA	CG10912	-1.15	0.001	CG14531-RA	CheB98a	-1.21	0.001
CG6479-RA	CG6479	-1.05	0.003	CG7889-RA	CG7889	-1.24	0.001
CG17906-RA	CG17906	-1.18	0.001	CG31606-RB	CG31606	-1.32	0.001
CG1077-RA	CG1077	-1.13	0.001	CG6971-RA	CG6971	-1.06	0.001
CG11069-RA	CG11069	-1.46	0.001	CG17280-RA	levy	-1.01	0.004
CG5724-RA	CG5724	-1.27	0.001	CG1165-RA	LysS	-1.55	0.001
CG4271-RA	CG4271	-1.01	0.002	CG31204-RA	CG31204	-1.14	0.001
CG33017-RB	CG33017	-1.05	0.001	CG18472-RA	CG18472	-1.11	0.001
CG7615-RA	fig	-1.06	0.001	CG14598-RA	CG14598	-1.13	0.001
CG5776-RA	CG5776	-1.23	0.001	CG10888-RA	Rh3	-1.20	0.001
CG2930-RB	CG2930	-1.46	0.001	CG13566-RA	CG13566	-1.07	0.002
CG6416-RJ	Zasp66	-1.04	0.001	CG17974-RA	CG17974	-1.31	0.001
CG32209-RB	serp	-1.44	0.002	CG1967-RA	p24-1	-1.80	0.001
CG13716-RA	CG13716	-1.13	0.001	CG17828-RB	mod(r)	-1.17	0.001
CG31148-RA	CG31148	-1.31	0.001	CG3567-RA	Cyp6u1	-1.20	0.001
CG18321-RA	miple2	-1.81	0.001	CG13422-RA	CG13422	-1.18	0.001
CG5693-RA	CG5693	-1.15	0.001	CG32351-RA	S-Lap2	-1.09	0.001
CG5281-RA	CG5281	-1.29	0.003	CG17835-RB	inv	-1.59	0.001
CG7459-RA	Poxm	-1.63	0.001	CG5734-RA	CG5734	-1.33	0.002
CG6284-RA	Sirt6	-1.69	0.001	CG2979-RA	Yp2	-1.22	0.001
CG10772-RB	Fur1	-1.34	0.001	CG15179-RA	sunz	-1.02	0.001
CG6385-RA	CG6385	-1.07	0.001	CG11350-RB	CG11350	-1.05	0.001
CG17549-RA	CG17549	-1.49	0.001	CG14673-RA	CG14673	-1.21	0.001
CG7570-RA	hale	-1.14	0.001	CG5185-RA	Tom	-2.56	0.002
CG7916-RA	CG7916	-1.32	0.001	CG30060-RA	CG30060	-1.11	0.001
CG3388-RA	gsb	-1.69	0.006	CG18568-RA	CG18568	-1.15	0.001
CG31118-RA	RabX4	-1.37	0.002	CG11154-RB	ATPsyn-β	-1.86	0.002

Continued on next page

Table 8 – continued from previous page

		Dichaete mutant				*Dichaete* mutant	
Transcript	**Gene**	**M**	**p**	**Transcript**	**Gene**	**M**	**p**
CG10730-RA	CG10730	-1.06	0.001	CG7691-RA	CG7691	-1.21	0.001
CG6515-RA	Takr86C	-1.05	0.003	CG10186-RC	CG10186	-1.06	0.002
CG31226-RA	CG31226	-1.24	0.001	CG1787-RA	Hexo2	-1.22	0.001
CG10586-RA	CG10586	-1.17	0.001	CG14518-RA	CG14518	-1.27	0.001
CG5639-RA	CG5639	-1.08	0.004	CG33209-RA	comm3	-1.10	0.001
CG14127-RA	CG14127	-1.43	0.001	CG3184-RA	CG3184	-1.18	0.001
CG15003-RA	Teh4	-1.35	0.001	CG6429-RA	CG6429	-1.19	0.001
CG31473-RA	CG31473	-1.22	0.001	CG9873-RA	RpL37b	-1.18	0.001
CG31876-RA	Cpr30F	-1.13	0.001	CG4780-RA	membrin	-1.21	0.001
CG8003-RA	CG8003	-1.71	0.001	CG15657-RA	CG15657	-1.23	0.001
CG7539-RA	Edg91	-1.04	0.001	CG3104-RA	CG3104	-1.07	0.001
CG3619-RA	Dl	-1.99	0.001	CG5381-RA	CG5381	-1.03	0.001
CG32655-RA	CG32655	-1.32	0.001	CG5427-RA	Oatp33Ea	-1.21	0.001
CG30271-RC	CG30271	-1.04	0.001	CG9512-RA	CG9512	-1.18	0.001
CG12410-RA	cv	-1.45	0.001	CG13010-RA	CG13010	-1.02	0.001
CG4434-RA	CG4434	-1.21	0.001	CG7907-RA	CG7907	-1.05	0.001
CG18815-RA	CG18815	-1.79	0.001	CG18266-RA	CG18266	-1.18	0.001
CG6917-RA	Est-6	-1.43	0.001	CG13521-RB	robo	-1.82	0.001
CG7442-RA	CG7442	-1.08	0.001	CG14294-RA	CG14294	-1.20	0.001
CG10573-RA	ko	-1.59	0.001	CG31929-RA	Gr22c	-1.07	0.001
CG8421-RE	Asph	-1.31	0.001	CG5133-RA	Doc1	-1.74	0.001
CG4685-RA	Ssadh	-1.40	0.001	CG1908-RA	CG1908	-1.20	0.001
CG9175-RA	CG9175	-1.00	0.003	CG12765-RA	fsd	-1.26	0.001
CG1264-RA	lab	-1.31	0.001	CG8616-RA	BHD	-1.26	0.002
CG5841-RA	mib1	-1.15	0.001	CG18103-RB	fos28F	-1.20	0.001
CG9232-RA	Galt	-1.37	0.001	CG4394-RB	Traf-like	-1.14	0.001
CG5907-RC	Frq2	-1.15	0.001	CG13749-RA	CG13749	-1.22	0.001
CG13231-RA	CG13231	-1.25	0.001	CG18682-RA	CG18682	-1.25	0.001
CG12052-RS	lola	-1.96	0.001	CR31927-RA	snRNA:U3:22A	-1.02	0.001
CG9454-RA	Spn42Db	-1.07	0.001	CG3940-RA	CG3940	-1.38	0.001
CG9059-RA	CG9059	-1.14	0.001	CG4830-RA	CG4830	-1.07	0.001
CG7094-RA	CG7094	-1.11	0.001	CG13725-RA	CG13725	-1.04	0.001
CG7026-RA	VhaPPA1-2	-1.04	0.001	CG7348-RA	obst-J	-1.21	0.001
CG12589-RA	CG12589	-1.03	0.001	CG18607-RA	CG18607	-1.14	0.001
CG13662-RA	CG34110	-1.04	0.001	CG10042-RB	MBD-R2	-1.70	0.001
CG5215-RA	Zn72D	-1.41	0.001	CG9331-RC	CG9331	-1.65	0.001
CG5069-RA	croc	-1.14	0.001	CG5682-RA	Edem2	-1.01	0.001
CG32555-RC	RhoGAPp190	-1.49	0.001	CR15821-RA	CR15821	-1.22	0.001
CG8834-RA	CG8834	-1.24	0.001	CG30480-RA	CG30480	-1.14	0.001
CG5085-RA	Sirt2	-1.10	0.001	CG5075-RA	Vha68-3	-1.17	0.001
CG7786-RA	CG7786	-1.00	0.002	CG1268-RA	VhaM9.7-a	-1.35	0.002
CG16782-RA	CG16782	-1.21	0.001	CG10562-RA	CG10562	-1.26	0.001
CG9468-RA	CG9468	-1.03	0.001	CG16826-RA	CG16826	-1.05	0.001
CG31949-RA	CG31949	-1.24	0.001	CG11977-RA	CG11977	-1.31	0.001
CG9428-RA	ZIP1	-1.27	0.001	CG8493-RA	Den1	-1.22	0.001
CG8084-RA	ana	-1.25	0.001	CG3579-RA	CG42361	-1.29	0.001
CG11303-RA	TM4SF	-1.13	0.001	CG5123-RA	W	-1.42	0.001
CG1007-RA	emc	-1.87	0.001	CG30128-RB	Obp56c	-1.26	0.001
CG9280-RC	Glt	-1.10	0.001	CG10051-RA	CG10051	-1.40	0.001
CG5467-RA	scrib	-1.44	0.001	CG3949-RA	hoip	-2.15	0.001
CG30488-RA	CG30488	-1.12	0.001	CG18125-RA	CG18125	-1.11	0.001
CG12526-RA	Or67a	-1.32	0.001	CG9635-RF	RhoGEF2	-1.35	0.001
CG4491-RA	noc	-2.00	0.001	CG7223-RB	htl	-1.43	0.001

Continued on next page

Table 8 – continued from previous page

		Dichaete mutant				*Dichaete* mutant	
Transcript	**Gene**	**M**	**p**	**Transcript**	**Gene**	**M**	**p**
CG9415-RA	Xbp1	-1.79	0.001	CG8957-RA	CG42235	-1.23	0.001
CG4821-RA	Tequila	-1.39	0.001	CG4605-RA	Acp32CD	-1.08	0.001
CG12972-RA	ebd2	-1.24	0.001	CG10005-RA	CG10005	-1.59	0.001
CG5170-RE	Dp1	-1.55	0.001	CG5044-RB	CG5044	-1.60	0.001
CG18024-RA	SoxN	-1.74	0.002	CG30389-RA	CG30389	-1.41	0.001
CG5177-RA	CG5177	-1.16	0.001	CG12769-RA	CG12769	-1.10	0.001
CG31902-RA	Spn28Da	-1.01	0.001	CG3292-RA	CG3292	-1.13	0.001
CG10601-RB	mirr	-1.35	0.001	CG6669-RA	klg	-1.31	0.001
CG12462-RA	tlk	-1.08	0.001	CG11247-RB	CG11247	-1.44	0.001
CG12869-RA	CG12869	-1.27	0.001	CG32744-RA	Ubi-p5E	-1.64	0.001
CG31195-RA	CG31195	-1.25	0.002	CG17856-RA	CG17856	-1.23	0.001
CG5264-RA	btn	-1.43	0.001	CG4262-RA	elav	-2.07	0.002
CG30371-RA	CG30371	-1.02	0.001	CG12052-RA	lola	-1.48	0.001
CG30002-RA	CG30002	-1.16	0.001	CG32755-RA	CG32755	-1.22	0.001
CG32694-RA	CG32694	-1.00	0.003	CG4726-RA	CG4726	-1.81	0.001
CG14396-RD	Ret	-1.06	0.001	CG7549-RB	mtg	-1.15	0.001
CG13956-RA	kat80	-1.32	0.001	CG11139-RA	p47	-1.66	0.001
CG12976-RA	CG42337	-1.24	0.001	CG1129-RB	CG1129	-1.76	0.001
CG14877-RA	CG14877	-1.06	0.001	CG18596-RA	CG18596	-1.33	0.001
CG15547-RA	CG15547	-1.11	0.001	CG10352-RA	CG10352	-1.09	0.001
CG1449-RA	zfh2	-1.76	0.001	CG2956-RA	twi	-1.47	0.001
CG14651-RA	CG14651	-1.25	0.001	CG5246-RA	CG5246	-1.08	0.001
CG3611-RA	CG3611	-1.15	0.001	CG3756-RA	CG3756	-1.58	0.001
CG32307-RA	zormin	-1.17	0.001	CG14948-RA	dpr18	-1.02	0.001
CG18418-RA	CG18418	-1.32	0.001	CG14509-RA	CG14509	-1.35	0.001
CG10844-RA	Rya-r44F	-1.51	0.001	CG14813-RA	deltaCOP	-1.43	0.001
CG14534-RA	TwdlE	-1.22	0.001	CG14936-RA	Tsp33B	-1.06	0.001
CG32265-RA	Eip63F-2	-1.00	0.002	CG17885-RA	Or1a	-1.05	0.001
CG3665-RA	Fas2	-2.20	0.001	CG3152-RA	Trap1	-1.17	0.001
CG6980-RA	CG6980	-1.14	0.001	CG10800-RA	Rca1	-1.65	0.001
CG31746-RA	CG42750	-1.15	0.001	CG14559-RA	CG33970	-1.43	0.001
CG6503-RA	CG6503	-1.05	0.002	CG4848-RA	CG4848	-1.22	0.001
CG13445-RA	CG13445	-1.08	0.001	CG17930-RA	CG17930	-1.58	0.001
CG31661-RA	CG31661	-1.36	0.001	CG18336-RA	CG18336	-1.09	0.001
CG17914-RA	yellow-b	-1.30	0.001	CG31481-RA	pb	-1.20	0.001
CG30062-RA	CG30062	-1.06	0.001	CG7057-RB	AP-50	-1.71	0.001
CG33152-RA	hbn	-1.21	0.001	CG6953-RA	fat-spondin	-1.34	0.001
CG3558-RA	CG3558	-1.13	0.001	CG18497-RA	spen	-1.44	0.001
CG14909-RA	VhaM9.7-d	-1.23	0.001	CG5613-RA	CG5613	-1.09	0.001
CG15073-RA	CG15073	-1.22	0.001	CG5952-RA	Fer2	-1.11	0.001
CG2958-RA	lectin-24Db	-1.29	0.001	CG32232-RA	CG32232	-1.12	0.001
CG14900-RA	Cad89D	-1.16	0.001	CG3491-RA	CG3491	-1.06	0.001
CG2297-RA	Obp44a	-1.04	0.001	CG13582-RA	Ir60d	-1.08	0.001
CG4190-RA	Hsp67Bc	-1.46	0.001	CG4322-RA	moody	-1.32	0.001
CG17799-RA	lectin-29Ca	-1.11	0.001	CG17962-RA	Z600	-1.71	0.004
CG14322-RA	CG14322	-1.04	0.001	CG17943-RA	comm	-2.29	0.001
CG17917-RA	CG17917	-1.05	0.001	CG18363-RA	Dic4	-1.18	0.001
CG13454-RA	CG13454	-1.71	0.001	CG5715-RA	CG5715	-1.04	0.001
CG3134-RA	ord	-1.11	0.001	CG6944-RA	Lam	-1.75	0.001
CG18518-RA	CG18518	-1.18	0.001	CG32384-RA	form3	-1.11	0.001
CG33350-RA	CheB42c	-1.03	0.001	CG1786-RA	Cyp318a1	-1.03	0.001
CG4770-RA	CG4770	-1.25	0.001	CG10337-RA	CG10337	-1.04	0.001
CG6614-RA	CG6614	-1.28	0.001	CG14447-RA	Grip	-1.08	0.001

Continued on next page

Table 8 – continued from previous page

		Dichaete mutant				*Dichaete* mutant	
Transcript	**Gene**	**M**	**p**	**Transcript**	**Gene**	**M**	**p**
CG11318-RA	CG11318	-1.39	0.001	CG10850-RA	ida	-1.12	0.001
CG1939-RA	Dpck	-1.10	0.001	CG32257-RA	Gr64b	-1.27	0.001
CG15396-RA	Gr23a	-1.40	0.001	CG18109-RA	CG18109	-1.25	0.001
CG4345-RA	grim	-1.18	0.001	CG14692-RA	CG14692	-1.08	0.001
CG10853-RA	CG10853	-1.63	0.001	CG7631-RA	CG7631	-1.27	0.001
CG3424-RA	path	-1.17	0.001	CG5409-RA	Arp53D	-1.05	0.001
CG7968-RA	CG7968	-1.12	0.001	CG17994-RA	CG17994	-1.19	0.001
CG3610-RA	CG3610	-1.66	0.001	CG33109-RA	CG33109	-1.17	0.001
CG9985-RA	sktl	-1.21	0.001	CG4858-RA	CG4858	-1.58	0.001
CG33322-RA	CG33322	-1.21	0.001	CG6541-RA	Mst33A	-1.14	0.001
CG31368-RA	CG31368	-1.71	0.001	CG10662-RA	sick	-1.47	0.001
CG14544-RA	CG14544	-1.19	0.001	CG31774-RA	fred	-1.38	0.001
CG14436-RA	C3G	-1.38	0.001	CG32132-RA	CG32132	-1.05	0.001
CG15019-RA	CG15019	-1.83	0.001	CG6453-RA	CG6453	-1.31	0.001
CG14379-RA	CheA87a	-1.14	0.001	CG5080-RA	CG5080	-1.68	0.002
CG31941-RA	Obp22a	-1.08	0.001	CG18446-RA	CG18446	-1.34	0.001
CG1099-RA	Dap160	-1.02	0.001	CG3440-RA	Pcp	-1.19	0.001
CG32838-RA	CG42345	-1.04	0.001	CG16717-RA	CG16717	-1.02	0.001
CG8966-RA	CG42235	-1.20	0.001	CG4414-RA	Ugt58Fa	-1.06	0.001
CG32686-RB	CG32686	-1.03	0.001	CG4372-RA	CG4372	-1.33	0.001
CG9535-RA	mmy	-1.08	0.001	CG12437-RA	raw	-1.27	0.001
CG18090-RA	Dsk	-1.26	0.001	CG12375-RA	CG12375	-1.06	0.001
CG10630-RA	blanks	-1.29	0.001	CG8918-RA	CG8918	-1.03	0.001
CG8110-RA	syd	-1.15	0.001	CG1890-RA	CG1890	-1.19	0.001
CG13912-RA	CG13912	-1.46	0.001	CG32676-RA	CG32676	-1.36	0.001
CG7005-RA	Esp	-1.89	0.001	CG10016-RA	drm	-1.83	0.001
CG4005-RA	yki	-1.28	0.001	CG7535-RA	GluClα	-1.04	0.001
CG12007-RA	CG12007	-1.26	0.001	CG14367-RA	CG14367	-1.33	0.001
CR40465-RA	CR40465	-1.12	0.001	CG13461-RA	CG13461	-1.53	0.001
CG4761-RA	knrl	-1.99	0.001	CG14490-RA	CG14490	-1.24	0.001
CG4008-RA	und	-1.26	0.001	CG5517-RA	Ide	-1.55	0.001
CG10596-RA	Msr-110	-1.16	0.001	CG13500-RA	CG13500	-1.03	0.002
CG9211-RA	ihog	-1.44	0.001	CG5161-RA	CG5161	-1.03	0.001
CG6699-RA	beta'Cop	-1.53	0.001	CG4893-RA	CG4893	-1.07	0.001
CG1406-RA	U2A	-1.39	0.002	CG30396-RA	Gr58a	-1.26	0.001
CG1934-RA	ImpE2	-1.74	0.001	CG9204-RC	Ate1	-1.15	0.001
CG12789-RA	santa-maria	-1.20	0.001	CG5488-RA	B-H2	-1.09	0.001
CG8934-RA	CG42235	-1.36	0.001	CG5928-RA	CG5928	-1.22	0.001
CG8598-RA	eco	-1.15	0.001	CG32945-RA	CG32945	-1.14	0.001
CG5012-RA	mRpL12	-1.23	0.001	CG4426-RA	ast	-1.02	0.001
CG32592-RA	hiw	-1.27	0.001	CG16749-RA	CG16749	-1.61	0.001
CG30323-RA	CG30323	-1.02	0.001	CG18241-RA	Toll-4	-1.07	0.001
CG1331-RA	Ccp84Af	-1.38	0.001	CR32646-RB	fne	-1.45	0.001
CG3960-RD	CG34417	-1.27	0.001	CG13088-RA	CG13088	-1.14	0.001
CG9345-RA	Adgf-C	-1.07	0.001	CG30195-RA	CG30195	-1.16	0.001
CG7077-RA	CG7077	-1.02	0.001	CG12819-RB	sle	-1.30	0.001
CG11313-RA	CG11313	-1.13	0.001	CG7409-RA	CG7409	-1.01	0.001
CG18647-RA	bin	-1.36	0.001	CG13694-RA	CG13694	-1.06	0.002
CG8502-RA	Cpr49Ac	-1.20	0.001	CG10113-RA	wa-cup	-1.15	0.001
CG11149-RA	CG11149	-1.08	0.001	CG31870-RC	CG31870	-1.00	0.001
CG12487-RA	BobA	-2.18	0.004	CG5961-RA	CG5961	-1.47	0.001
CG18217-RA	CR18217	-1.08	0.001	CG5196-RA	CG5196	-1.58	0.001
CG9989-RA	CG9989	-1.00	0.001	CG12387-RA	zetaTry	-1.19	0.001

Continued on next page

Table 8 – continued from previous page

		Dichaete mutant				*Dichaete* mutant	
Transcript	**Gene**	**M**	**p**	**Transcript**	**Gene**	**M**	**p**
CR31273-RC	bxd	-1.33	0.001	CG5460-RC	H	-1.16	0.001
CG30494-RA	CG43340	-1.10	0.001	CG8072-RA	CG8072	-1.37	0.001
CG9461-RA	FBX011	-1.61	0.001	CG14395-RA	CG14395	-1.29	0.001
CG6863-RA	tok	-1.74	0.001	CG9444-RA	CG9444	-1.39	0.001
CG10140-RA	CG10140	-1.16	0.001	CG10671-RB	CG10671	-1.09	0.001
CG10101-RA	Ir84a	-1.07	0.001	CG8871-RA	Jon25Biii	-1.09	0.002
CG7245-RA	eys	-1.18	0.001	CG11209-RA	ppk6	-1.07	0.002
CG8542-RA	Hsc70-5	-1.75	0.001	CG8489-RA	soti	-1.27	0.001
CG9996-RA	CG9996	-1.37	0.001	CG13989-RA	CG13989	-1.26	0.002
CG3975-RA	CG3975	-1.17	0.001	CG16789-RA	CG16789	-1.05	0.001
CG12052-RF	lola	-1.36	0.001	CG1424-RA	mst	-1.17	0.001
CG12557-RA	dpr13	-1.16	0.001	CG40045-RA	CG40045	-1.38	0.001
CG30356-RA	CG30356	-1.13	0.001	CG5006-RA	Or33c	-1.10	0.001
CG5762-RA	CG5762	-1.08	0.001	CG13035-RB	CG13035	-1.12	0.001
CG14160-RA	CG14160	-1.31	0.001	CG13788-RB	Gr28b	-1.19	0.001
CG4329-RB	CG4329	-1.18	0.001	CG9339-RB	sky	-1.43	0.001
CG5062-RA	CG5062	-1.09	0.001	CG13247-RA	CG13247	-1.24	0.001
CG5201-RA	Dad	-1.91	0.001	CG12847-RA	Tsp42Ec	-1.28	0.001
CG13873-RB	Obp56g	-1.18	0.001	CG1921-RB	sty	-2.02	0.001
CG5648-RA	Prosα6T	-1.20	0.002	CG5602-RA	DNA-ligI	-1.30	0.001
CG6622-RA	Pkc53E	-1.24	0.004	CG6592-RA	CG6592	-1.41	0.001
CG12091-RA	CG12091	-1.88	0.001	CG13954-RA	CG13954	-1.04	0.001
CG8137-RA	Spn2	-1.05	0.001	CG30456-RA	CG30456	-1.12	0.002
CG12374-RA	CG12374	-1.28	0.002	CG8586-RA	CG8586	-1.15	0.002
CG9522-RA	CG9522	-1.19	0.001	CG17077-RB	pnt	-1.46	0.005
CG7966-RA	CG7966	-1.18	0.003	CG16957-RA	CG16957	-1.08	0.001
CG9020-RA	Aats-arg	-1.34	0.001	CG10035-RA	CG10035	-2.21	0.001
CG8224-RB	babo	-1.13	0.001	CG31747-RA	Gr36a	-1.31	0.001
CG12154-RA	oc	-1.21	0.001	CG18802-RA	α-Man-II	-1.16	0.002
CG32972-RA	CG43332	-1.38	0.001	CG6739-RA	CG6739	-1.09	0.001
CG2160-RA	Socs44A	-1.63	0.001	CG14826-RA	CG14826	-1.37	0.001
CG10483-RA	CG10483	-1.14	0.001	CG6827-RB	Nrx-IV	-1.71	0.001
CG2145-RA	CG2145	-1.30	0.002	CG15576-RB	CG15576	-1.03	0.003
CG3843-RA	RpL10Aa	-1.23	0.001	CG17084-RA	mthl9	-1.26	0.001
CG2198-RA	Ama	-2.32	0.008	CG2839-RA	CG2839	-1.07	0.004
CG13206-RA	Or47b	-1.00	0.001	CG7975-RA	Grx-1	-1.29	0.001
CG12974-RA	CG12974	-1.41	0.001	CG7762-RA	Rpn1	-1.63	0.001
CG13250-RA	CG13250	-1.05	0.001	CG8561-RA	conv	-1.17	0.003
CG8335-RA	CG8335	-1.29	0.001	CG14495-RA	CG14495	-1.09	0.001
CG6870-RA	CG6870	-1.08	0.001	CG1279-RA	Rtnl2	-1.11	0.001
CG3739-RA	CG3739	-1.03	0.002	CG31695-RA	scw	-1.15	0.001
CG8209-RA	CG8209	-1.46	0.001	CG16870-RA	Acyp	-1.00	0.001
CG13930-RA	CG13930	-1.05	0.004	CG32391-RA	CG32391	-1.11	0.001
CG5873-RA	CG5873	-1.16	0.001	CG31210-RA	CG4836	-1.12	0.001
CG13937-RA	CG13937	-1.33	0.001	CG17086-RA	CG43129	-1.09	0.001
CG10301-RA	CG10301	-1.16	0.001	CG10947-RB	CG10947	-1.18	0.001
CG8668-RA	CG8668	-1.00	0.007	CG15128-RA	CG15128	-1.08	0.001
CG17736-RA	schuy	-1.21	0.001	CG2560-RA	Cpr11A	-1.22	0.001
CG15616-RA	Acp53C14b	-1.08	0.001	CG8925-RA	CG8925	-1.10	0.001
CG31116-RA	ClC-a	-1.44	0.001	CG8321-RA	CG8321	-1.55	0.001
CG5773-RA	CG5773	-1.32	0.001	CG15086-RB	CG15086	-1.13	0.001
CG17625-RA	CG17625	-1.01	0.003	CG7848-RA	CG7848	-1.00	0.002
CG12316-RA	CG12316	-1.16	0.001	CG12245-RA	gcm	-1.91	0.001

Continued on next page

Table 8 – continued from previous page

		Dichaete mutant				*Dichaete* mutant	
Transcript	**Gene**	**M**	**p**	**Transcript**	**Gene**	**M**	**p**
CG11210-RA	CG11210	-1.04	0.001	CG17136-RD	Rbp1	-1.07	0.001
CG3777-RC	CG3777	-1.31	0.001	CG14759-RA	CG14759	-1.11	0.001
CG8888-RA	CG8888	-1.06	0.001	CG18111-RA	Obp99a	-1.26	0.002
CG15144-RA	CG15144	-1.05	0.002	CG17131-RA	tyn	-1.51	0.001
CG8442-RA	Glu-RI	-1.07	0.001	CG4169-RA	CG4169	-1.30	0.001
CG3793-RA	CG3793	-1.12	0.001	CG12910-RA	CG12910	-1.25	0.001
CG5830-RA	CG5830	-1.45	0.001	CG33048-RA	Mocs1	-1.07	0.001
CG6730-RA	Cyp4d21	-1.00	0.001	CG32432-RA	CG32432	-1.53	0.001
CG12858-RA	CG12858	-1.31	0.001	CG5002-RA	CG5002	-1.32	0.009
CG6246-RA	nub	-2.03	0.001	CG10250-RA	nau	-1.50	0.001
CG5923-RA	DNApol-α73	-1.21	0.001	CG11905-RE	CG11905	-1.08	0.002
CG6844-RA	nAcRα-96Ab	-1.06	0.001	CG3773-RA	CG3773	-1.00	0.001
CG15733-RA	Ten-a	-1.24	0.001	CG11711-RA	Mob2	-1.03	0.001
CG11801-RA	Elo68beta	-1.15	0.001	CG31556-RA	CG31556	-1.24	0.001
CG15275-RA	CG42313	-1.10	0.001	CG31545-RA	MTA1-like	-1.66	0.001
CG7370-RA	CG7370	-1.03	0.003	CG5959-RA	MCO3	-1.20	0.001
CG7290-RA	CG7290	-1.23	0.001	CG4061-RA	CG4061	-1.16	0.001
CG13526-RA	CG13526	-1.08	0.001	CG13127-RA	CG13127	-1.41	0.001
CG11512-RA	GstD4	-1.13	0.001	CG31090-RA	CG42235	-1.30	0.001
CG9614-RH	pip	-1.04	0.001	CG16708-RA	Cerk	-1.18	0.001
CG31753-RA	ham	-1.25	0.001	CG10357-RA	CG10357	-1.00	0.001
CG14089-RA	CG14089	-1.10	0.001	CG33169-RA	CG33169	-1.08	0.001
CG1988-RB	CG1988	-1.35	0.001	CG33490-RA	CG33490	-1.24	0.001
CG11377-RA	CG11377	-1.45	0.001	CG6153-RA	CG6153	-1.29	0.001
CG15745-RA	CG15745	-1.44	0.001	CG14242-RA	TwdlS	-1.10	0.001
CG14075-RA	CG14075	-1.24	0.001	CG31653-RA	CG34126	-1.05	0.001
CG33197-RC	mbl	-1.12	0.001	CG32072-RA	Elo68alpha	-1.38	0.001
CG30159-RB	CG30159	-1.26	0.001	CG3408-RA	CG3408	-1.06	0.001
CG33472-RA	qvr	-1.06	0.001	CG31436-RA	CG31436	-1.16	0.001
CG33192-RA	MtnD	-1.31	0.001	CG1501-RA	unc	-1.25	0.001
CG5677-RA	Spase22-23	-1.16	0.002	CG14259-RA	CG14259	-1.37	0.001
CG11326-RE	Tsp	-1.58	0.001	CG31057-RA	tau	-1.00	0.001
CG15909-RA	CG15909	-1.06	0.001	CG2898-RA	CG2898	-1.05	0.001
CG9080-RA	Listericin	-1.25	0.001	CG15614-RA	CG15614	-1.59	0.001
CG5479-RA	mRpL43	-1.51	0.001	CG7343-RA	btsz	-1.65	0.001
CG13032-RA	CG13032	-1.08	0.001	CG9801-RA	CG9801	-1.13	0.001
CG15736-RA	Chrac-16	-1.19	0.004	CG14372-RA	CG14372	-1.45	0.001
CG33292-RA	CG33292	-1.09	0.001	CG31622-RD	Gr39a	-1.05	0.001
CG31291-RB	CG31291	-1.20	0.001	CG30349-RA	CG30349	-1.31	0.001
CG4019-RC	CG4019	-1.32	0.001	CG3065-RA	CG3065	-1.40	0.001
CG31519-RA	Or82a	-1.16	0.001	CG15627-RA	Ir25a	-1.12	0.001
CG15406-RA	CG15406	-1.17	0.001	CG6435-RA	CG6435	-1.04	0.001
CG32134-RA	btl	-1.58	0.002	CG6019-RA	mus308	-1.74	0.001
CG30450-RA	Obp56f	-1.15	0.001	CG12212-RA	peb	-1.94	0.001
CG7333-RA	CG7333	-1.14	0.001	CG2121-RA	CG2121	-1.09	0.001
CG11642-RA	TRAM	-1.04	0.003	CG15288-RB	wb	-1.47	0.001
CG13442-RA	CG13442	-1.18	0.001	CG14958-RA	CG14958	-1.03	0.001
CG2062-RA	Cyp4e1	-1.52	0.001	CG9139-RA	Rabex-5	-1.31	0.001
CG31691-RA	TotF	-1.19	0.001	CG1630-RB	IP3K2	-1.24	0.001
CG6061-RA	mip120	-1.06	0.001	CG33223-RA	CG33223	-1.09	0.001
CG8648-RA	Fen1	-1.18	0.001	CG3922-RB	RpS17	-1.26	0.001
CG15822-RC	CG15822	-1.18	0.001	CG4104-RA	Tps1	-2.12	0.001
CG12173-RA	CG12173	-1.02	0.001	CG13188-RA	CG13188	-1.00	0.001

Continued on next page

Table 8 – continued from previous page

		Dichaete mutant				*Dichaete* mutant	
Transcript	**Gene**	**M**	**p**	**Transcript**	**Gene**	**M**	**p**
CG30191-RA	Gr59b	-1.29	0.001	CG9609-RA	CG9609	-1.55	0.001
CG9614-RF	pip	-1.17	0.001	CG16899-RA	FoxP	-1.08	0.001
CG13011-RA	sing	-1.23	0.001	CG31300-RA	CG31300	-1.24	0.001
CG7400-RB	Fatp	-1.45	0.001	CG14151-RA	CG14151	-1.15	0.001
CG15008-RA	Cpr64Ac	-1.22	0.001	CG2507-RB	sas	-1.41	0.001
CG15396-RB	Gr23a	-1.06	0.001	CG11466-RA	Cyp9f2	-1.02	0.001
CG1732-RA	CG1732	-1.22	0.001	CG15350-RA	Cp7Fb	-1.02	0.001
CG12272-RA	CG12272	-1.41	0.001	CG9900-RB	mit(1)15	-1.56	0.001
CG9455-RA	Spn42Dc	-1.06	0.001	CG17107-RA	CG17107	-1.12	0.001
CG7988-RA	CG7988	-1.38	0.001	CG16850-RA	CG42784	-1.13	0.001
CG6185-RA	Ir68a	-1.21	0.001	CG9286-RA	CG9286	-1.55	0.001
CG1028-RL	Antp	-1.16	0.001	CG17011-RA	lectin-30A	-1.29	0.001
CG15118-RA	CG15118	-1.42	0.001	CG14066-RB	larp	-1.22	0.001
CG12348-RB	Sh	-1.00	0.001	CG3126-RA	C3G	-1.21	0.001
CG12905-RA	Obp46a	-1.06	0.001	CG10146-RA	AttA	-1.55	0.001
CG12839-RA	Tsp42En	-1.14	0.001	CG13318-RA	CG13318	-1.10	0.001
CG8328-RA	HLHmδ	-2.34	0.001	CG11648-RA	Abd-B	-1.96	0.001
CG10996-RA	CG10996	-1.27	0.001	CG17174-RA	ACXB	-1.21	0.001
CG1559-RA	Upf1	-1.23	0.001	CG6813-RA	CG6813	-1.06	0.001
CG2219-RA	CG2219	-1.05	0.001	CG1887-RA	CG1887	-1.10	0.001
CG13423-RA	CG13423	-1.29	0.001	CG13001-RA	CG13001	-1.04	0.001
CG7780-RA	DNaseII	-2.11	0.001	CG14572-RA	CG14572	-1.14	0.001
CG31803-RA	CG31803	-1.23	0.001	CG17217-RA	CG17217	-1.12	0.001
CG15208-RA	CG15208	-1.12	0.001	CG14027-RA	TotM	-1.29	0.001
CG33108-RA	CG33108	-1.00	0.001	CG6891-RA	CG6891	-2.05	0.001
CG12822-RA	CG12822	-1.15	0.001	CG18676-RA	Teh3	-1.21	0.001
CG12232-RA	Gα73B	-1.38	0.001	CG12304-RA	CG12304	-1.07	0.001
CG3171-RA	Tre1	-1.17	0.002	CG8234-RA	Tret1-2	-1.42	0.001
CG16959-RA	CG16959	-1.48	0.001	CG32406-RA	CG32406	-1.00	0.001
CG5882-RA	CG5882	-1.07	0.001	CG7771-RA	sim	-1.12	0.001
CG8823-RA	Lip3	-1.08	0.001	CG17577-RA	Cyp9h1	-1.14	0.001
CG6860-RB	Lrch	-1.23	0.001	CG1605-RA	az2	-1.12	0.001
CG13976-RA	Gr98a	-1.15	0.001	CG8526-RA	CG8526	-1.05	0.001
CG7164-RA	CG7164	-1.30	0.001	CG10327-RA	TBPH	-1.55	0.001
CG11186-RA	toy	-1.39	0.001	CG4559-RA	Idgf3	-1.30	0.001
CG4896-RB	CG4896	-1.54	0.001	CG8408-RA	CG8408	-1.40	0.001
CG33116-RA	CG33116	-1.68	0.001	CG8342-RA	m1	-1.65	0.001
CG15109-RC	CG15109	-1.13	0.001	CG32199-RA	CG32199	-1.15	0.001
CG11063-RB	jub	-1.47	0.001	CG17650-RA	CG17650	-1.34	0.001
CG17269-RA	Fancd2	-1.16	0.001	CG8271-RA	Sln	-1.21	0.001
CG6233-RA	Ufd1-like	-1.45	0.001	CG12344-RA	CG12344	-1.31	0.001
CG32459-RA	CG32459	-1.26	0.001	CG2302-RA	nAcRα-7E	-1.35	0.001
CG12249-RA	mira	-2.06	0.002	CG15592-RA	Osi9	-1.11	0.001
CG7857-RA	CG7857	-1.01	0.002	CG13923-RA	CG42676	-1.11	0.001
CG18234-RB	CG18234	-1.04	0.001	CG16932-RA	Eps-15	-1.08	0.001
CG15255-RA	CG15255	-1.48	0.001	CG4279-RA	LSm1	-1.64	0.001
CG3819-RA	CG3819	-1.26	0.001	CG12942-RA	CG12942	-1.15	0.001
CG10195-RA	CG10195	-1.11	0.001	CG33070-RA	Sxl	-2.26	0.001
CG13857-RA	CG13857	-1.02	0.001	CG32450-RA	CG32450	-1.40	0.001
CG16965-RA	CG16965	-1.21	0.001	CG32491-RO	mod(mdg4)	-1.00	0.001
CG13349-RA	CG13349	-1.27	0.001	CG12205-RA	Bsg25A	-1.17	0.001
CG8205-RA	fus	-1.24	0.001	CG10514-RA	CG10514	-1.37	0.001
CG7671-RA	Nup43	-1.37	0.001	CG10244-RA	Cad96Ca	-1.28	0.001

Continued on next page

Table 8 – continued from previous page

		Dichaete mutant				*Dichaete* mutant	
Transcript	**Gene**	**M**	**p**	**Transcript**	**Gene**	**M**	**p**
CG13270-RA	Ugt36Ba	-1.14	0.001	CG8288-RA	mRpL3	-1.28	0.001
CG7266-RA	Eip71CD	-1.66	0.001	CG1287-RA	CG1287	-1.55	0.001
CG12866-RA	CG12866	-1.52	0.001	CG15216-RA	CG42748	-1.13	0.001
CG11019-RA	CG11019	-1.22	0.001	CG10275-RA	kon	-1.65	0.001
CG17049-RA	CG17049	-1.35	0.001	CG15905-RA	CG15905	-1.14	0.001
CG33139-RA	Ranbp11	-1.15	0.001	CG7166-RA	CG7166	-1.13	0.001
CG3841-RA	CG3841	-1.10	0.001	CG14854-RA	CG14854	-1.03	0.001
CG11735-RA	Or85b	-1.13	0.001	CG11875-RA	CG11875	-1.04	0.001
CG6379-RA	CG6379	-1.43	0.001	CG17298-RA	CG17298	-1.12	0.001
CG10440-RA	CG10440	-1.01	0.001	CG11883-RB	CG11883	-1.19	0.001
CG7387-RA	CG7387	-1.14	0.001	CG10489-RA	Pole2	-1.01	0.001
CG5993-RA	os	-1.40	0.001	CG7313-RA	CheA75a	-1.11	0.001
CG12131-RA	Adam	-1.64	0.001	CG13894-RA	CG13894	-1.07	0.001
CG9256-RA	Nhe2	-1.33	0.001	CG5182-RA	Pk34A	-1.22	0.001
CG31648-RA	CG31648	-1.06	0.001	CG8722-RC	Nup44A	-1.01	0.001
CG13148-RA	CG13148	-1.26	0.001	CG9668-RA	Rh4	-1.02	0.001
CG13091-RA	CG13091	-1.13	0.001	CG14119-RA	CG14119	-1.16	0.001
CG13102-RA	CG13102	-1.28	0.001	CG4114-RA	ex	-1.48	0.001
CG3356-RA	CG3356	-1.27	0.001	CG12001-RA	CG12001	-1.16	0.001
CG14232-RA	CG14232	-1.32	0.001	CG31675-RA	CG31675	-1.10	0.001
CG3434-RA	CG3434	-1.10	0.001	CG31463-RA	CG31463	-1.25	0.001
CG14297-RA	CG14297	-1.28	0.001	CG11347-RA	DOR	-1.33	0.001
CG3106-RA	CG3106	-1.20	0.001	CG31064-RA	CG31064	-1.05	0.003
CG33478-RB	Or46a	-1.36	0.001	CG31407-RA	CG31407	-1.34	0.001
CG40485-RB	CG40485	-1.02	0.002	CG1750-RA	CG1750	-1.22	0.001
CG14110-RA	CG14110	-1.48	0.001	CG1743-RB	Gs2	-1.73	0.001
CG7298-RA	CG7298	-1.24	0.001	CG9837-RA	CG9837	-1.23	0.001
CG9377-RA	CG9377	-1.50	0.001	CG31640-RA	CG34380	-1.42	0.001
CG3581-RA	CG3581	-1.20	0.001	CG15816-RA	CG42684	-1.37	0.001
CG31022-RA	PH4αEFB	-1.83	0.001	CG11833-RA	Ssl2	-1.17	0.001
CG3894-RB	CG3894	-2.20	0.001	CG6043-RC	CG6043	-1.04	0.001
CG12398-RA	CG12398	-1.15	0.001	CG31419-RA	CG31419	-1.11	0.001
CG7350-RA	Eig71Ed	-1.21	0.001	CG13213-RA	fbl6	-1.59	0.001
CG10370-RA	Tbp-1	-1.33	0.001	CG1979-RA	CG1979	-1.08	0.001
CG15591-RA	Osi8	-1.23	0.001	CG1502-RA	tsg	-1.01	0.002
CG32053-RA	CG32053	-1.29	0.001	CG2082-RD	CG2082	-1.07	0.001
CG11622-RA	Rlip	-1.08	0.001	CG9162-RA	CG9162	-1.16	0.001
CG31367-RA	CG42724	-1.40	0.001	CG6091-RA	CG6091	-1.07	0.001
CG1553-RC	CG1553	-1.21	0.001	CG11387-RB	ct	-1.06	0.001
CG8676-RA	Hr39	-1.77	0.001	CG15860-RA	pain	-1.13	0.005
CG31004-RA	CG31004	-1.00	0.003	CG12250-RB	ymp	-1.13	0.001
CG15167-RA	CG15167	-1.09	0.001	CG5079-RA	CG5079	-1.12	0.001
CG11958-RB	Cnx99A	-1.90	0.001	CG33307-RA	CG33307	-1.02	0.001
CG15488-RA	nub	-1.07	0.001	CG9582-RA	CG9582	-1.29	0.001
CG6094-RA	CG6094	-1.09	0.001	CG6142-RA	CG6142	-1.10	0.002
CG13039-RA	CG13039	-1.13	0.001	CG14853-RA	CG14853	-1.25	0.001
CG17520-RB	CkIIα	-1.15	0.002	CG14990-RA	CG14990	-1.49	0.001
CG12609-RB	CG12609	-1.14	0.001				

Appendix 4

G All binding datasets used for analysis

Table 9: All binding datasets used for analysis in Chapter 6

No. of peaks	Dataset
7827	BEAF32_0-12h_ChIP-chip_rel5.gff3
2232	CBP_0-4h_ChIP-chip2_rel5.gff3
926	CBP_0-4h_ChIP-chip_rel5.gff3
3296	CBP_0-4h_ChIP-seq_rel5.gff3
7736	CBP_4-8h_ChIP-chip_rel5.gff3
2348	CBP_4-8h_ChIP-seq_rel5.gff3
329	CBP_8-12h_ChIP-chip_rel5.gff3
2407	CBP_8-12h_ChIP-seq_rel5.gff3
10456	CP190_0-12h_ChIP-chip_rel5.gff3
4432	CTCF_C-term_0-12h_ChIP-chip_rel5.gff3
3833	CTCF_N-term_0-12h_ChIP-chip_rel5.gff3
4947	CtBP_0-12h_ChIP-chip_rel5.gff3
75	Dll_0-12h_ChIP-chip_rel5.gff3
1927	E(z)_8-16h_ChIP-chip_rel5.gff3
3011	GATAe_0-8h_ChIP-chip_rel5.gff3
8279	H3K18Ac_2-4h_ChIP-chip_rel5.gff3
5089	H3K27Ac_0-4h_ChIP-chip2_rel5.gff3
5431	H3K27Ac_0-4h_ChIP-chip_rel5.gff3
3658	H3K27Ac_0-4h_ChIP-seq_rel5.gff3
6944	H3K27Ac_4-8h_ChIP-chip_rel5.gff3
1119	H3K27Ac_4-8h_ChIP-seq_rel5.gff3
8101	H3K27Ac_8-12h_ChIP-chip_rel5.gff3
4430	H3K27Ac_8-12h_ChIP-seq_rel5.gff3
1519	H3K27Me3_0-12h_ChIP-chip_rel5.gff3
4429	H3K27Me3_0-4h_ChIP-chip_rel5.gff3
438	H3K27Me3_0-4h_ChIP-seq_rel5.gff3
944	H3K27Me3_4-8h_ChIP-chip_rel5.gff3
245	H3K27Me3_4-8h_ChIP-seq_rel5.gff3
5830	H3K27Me3_8-12h_ChIP-chip_rel5.gff3
230	H3K27Me3_8-12h_ChIP-seq_rel5.gff3
6814	H3K36Me3_0-12h_ChIP-chip_rel5.gff3
5952	H3K36Me3_2-4h_ChIP-chip_rel5.gff3
5238	H3K4Me1_0-4h_ChIP-chip_rel5.gff3
6162	H3K4Me1_0-4h_ChIP-seq_rel5.gff3
6149	H3K4Me1_4-8h_ChIP-chip_rel5.gff3
62	H3K4Me1_4-8h_ChIP-seq_rel5.gff3
9792	H3K4Me1_8-12h_ChIP-chip_rel5.gff3
5326	H3K4Me1_8-12h_ChIP-seq_rel5.gff3
4236	H3K4Me3_0-12h_ChIP-chip_rel5.gff3

Continued on next page

Table 9 – continued from previous page

No. of peaks	Dataset
12968	H3K4Me3_0-4h_ChIP-chip2_rel5.gff3
6095	H3K4Me3_0-4h_ChIP-chip_rel5.gff3
3840	H3K4Me3_0-4h_ChIP-seq_rel5.gff3
3495	H3K4Me3_2-4h_ChIP-chip_rel5.gff3
3504	H3K4Me3_4-8h_ChIP-chip_rel5.gff3
1967	H3K4Me3_4-8h_ChIP-seq_rel5.gff3
6867	H3K4Me3_8-12h_ChIP-chip_rel5.gff3
4508	H3K4Me3_8-12h_ChIP-seq_rel5.gff3
5133	H3K9Ac_0-4h_ChIP-chip_rel5.gff3
5513	H3K9Ac_0-4h_ChIP-seq_rel5.gff3
5844	H3K9Ac_4-8h_ChIP-chip_rel5.gff3
755	H3K9Ac_4-8h_ChIP-seq_rel5.gff3
6432	H3K9Ac_8-12h_ChIP-chip_rel5.gff3
4869	H3K9Ac_8-12h_ChIP-seq_rel5.gff3
3625	H3K9Me2_2-4h_ChIP-chip_rel5.gff3
3511	H3K9Me3_0-12h_ChIP-chip_rel5.gff3
454	H3K9Me3_0-4h_ChIP-chip2_rel5.gff3
2994	H3K9Me3_0-4h_ChIP-chip_rel5.gff3
409	H3K9Me3_0-4h_ChIP-seq_rel5.gff3
805	H3K9Me3_4-8h_ChIP-chip_rel5.gff3
427	H3K9Me3_4-8h_ChIP-seq_rel5.gff3
5646	H3K9Me3_8-12h_ChIP-chip_rel5.gff3
438	H3K9Me3_8-12h_ChIP-seq_rel5.gff3
2091	HDAC11_0-12h_ChIP-chip2_rel5.gff3
7215	HDAC11_0-12h_ChIP-chip_rel5.gff3
1513	HDAC11_0-12h_ChIP-seq2_rel5.gff3
989	HDAC11_0-12h_ChIP-seq_rel5.gff3
9377	HDAC1_0-12h_ChIP-chip2_rel5.gff3
9940	HDAC1_0-12h_ChIP-chip_rel5.gff3
4123	HDAC1_0-12h_ChIP-seq2_rel5.gff3
3011	HDAC1_0-12h_ChIP-seq_rel5.gff3
12119	HDAC3_0-12h_ChIP-chip2_rel5.gff3
12111	HDAC3_0-12h_ChIP-chip_rel5.gff3
2565	HDAC3_0-12h_ChIP-seq2_rel5.gff3
2620	HDAC3_0-12h_ChIP-seq_rel5.gff3
7309	HDAC4a_0-12h_ChIP-chip2_rel5.gff3
8965	HDAC4a_0-12h_ChIP-chip_rel5.gff3
4033	HDAC4a_0-12h_ChIP-seq2_rel5.gff3
2732	HDAC4a_0-12h_ChIP-seq_rel5.gff3
5385	HDAC6_0-12h_ChIP-chip2_rel5.gff3
5709	HDAC6_0-12h_ChIP-chip_rel5.gff3
3485	HDAC6_0-12h_ChIP-seq2_rel5.gff3
2989	HDAC6_0-12h_ChIP-seq_rel5.gff3
1688	HP1a_2-4h_ChIP-chip_rel5.gff3
869	Kr_0-8h_ChIP-chip_rel5.gff3
4485	PolII_0-12h_ChIP-chip_rel5.gff3
5235	PolII_4-8h_ChIP-seq_rel5.gff3
1333	PolII_8-12h_ChIP-seq_rel5.gff3
3220	PolII_unphosp_0-4h_ChIP-chip_rel5.gff3
5040	PolII_unphosp_4-8h_ChIP-chip2_rel5.gff3
5040	PolII_unphosp_4-8h_ChIP-chip_rel5.gff3
105	Stat92E_0-12h_ChIP-chip_rel5.gff3
6438	Trl_0-12h_ChIP-chip_rel5.gff3
1300	Ubx_0-12h_ChIP-chip_rel5.gff3
729	Ubx_3-8h_ChIP-chip2_rel5.gff3

Continued on next page

Table 9 – continued from previous page

No. of peaks	Dataset
161	Ubx_3-8h_ChIP-chip_rel5.gff3
1227	bab1_0-12h_ChIP-chip_rel5.gff3
2697	cad_0-4h_ChIP-chip_rel5.gff3
1700	cad_0-4h_T7GFP_ChIP-chip_rel5.gff3
2207	cad_0-4h_T7GFP_ChIP-seq_rel5.gff3
5626	cad_4-8h_T7GFP_ChIP-seq_rel5.gff3
7054	chinmo_0-12h_ChIP-chip_rel5.gff3
699	cnc_0-12h_ChIP-chip_rel5.gff3
2687	disco_0-8h_ChIP-chip_rel5.gff3
26	en_0-12h_ChIP-chip_rel5.gff3
286	en_7-24h_ChIP-chip_rel5.gff3
1738	eve_1-6h_GFP_ChIP-chip_rel5.gff3
28	ftz-f1_0-12h_ChIP-chip_rel5.gff3
1338	gro_0-12h_ChIP-chip1_rel5.gff3
626	gro_0-12h_ChIP-chip2_rel5.gff3
765	gsbn_7-24h_ChIP-chip_rel5.gff3
2745	h_0-8h_ChIP-chip_rel5.gff3
2995	hkb_0-8h_ChIP-chip_rel5.gff3
3222	inv_0-12h_ChIP-chip_rel5.gff3
2690	jumu_0-8h_ChIP-chip_rel5.gff3
792	kn_0-12h_ChIP-chip_rel5.gff3
3975	mod(mdg4)_0-12h_ChIP-chip_rel5.gff3
2174	mod(mdg4)_8-16h_ChIP-chip_rel5.gff3
333	run_0-12h_ChIP-chip_rel5.gff3
2884	sbb_0-4h_ChIP-chip_rel5.gff3
11773	sens_4-8h_ChIP-chip2_rel5.gff3
2640	sens_4-8h_ChIP-chip_rel5.gff3
16070	sens_4-8h_GFP_ChIP-chip_rel5.gff3
4779	su(Hw)_0-12h_ChIP-chip2_rel5.gff3
3947	su(Hw)_0-12h_ChIP-chip_rel5.gff3
384	ttk_0-12h_ChIP-chip_rel5.gff3
895	zfh1_0-12h_ChIP-chip_rel5.gff3

H TF datasets used for initial analysis

Table 10: Transcription factor datasets used for comparison with Dichaete in Chapter 6

No. of intervals	Dataset
1227	bab1_0-12h_ChIP-chip_rel5.gff3
802	bcd_2-3h_ChIP_Berk_fdr1_union_rel5.gff3
9708	cad_0-8h_ChIP_timeunion_rel5.gff3
7054	chinmo_0-12h_ChIP-chip_rel5.gff3
699	cnc_0-12h_ChIP-chip_rel5.gff3
5534	da_2-3h_ChIP-chip_Berk_fdr1_intervals_sym_rel5.gff
6227	Dichaete_core_uberset.gff3
2687	disco_0-8h_ChIP-chip_rel5.gff3
9358	dl_2-3h_ChIP-chip_Berk_fdr1_intervals_sym_rel5.gff
75	Dll_0-12h_ChIP-chip_rel5.gff3
26	en_0-12h_ChIP-chip_rel5.gff3
1738	eve_1-6h_GFP_ChIP-chip_rel5.gff3
403	ftz_2-3h_ChIP-chip_Berk_fdr1_intervals_sym_rel5.gff
28	ftz-f1_0-12h_ChIP-chip_rel5.gff3

Continued on next page

Table 10 – continued from previous page

No. of intervals	Dataset
3011	GATAe_0-8h_ChIP-chip_rel5.gff3
1412	gro_0-12h_ChIP_union_rel5.gff3
765	gsbn_7-24h_ChIP-chip_rel5.gff3
1070	gt_2-3h_ChIP-chip_Berk_fdr1_intervals_sym_rel5.gff
5145	h_0-8h_ChIP_timeunion_rel5.gff3
2180	hb_2-3h_ChIP_Berk_fdr1_union_rel5.gff3
3840	hkb_0-8h_ChIP_timeunion_rel5.gff3
3222	inv_0-12h_ChIP-chip_rel5.gff3
2690	jumu_0-8h_ChIP-chip_rel5.gff3
792	kn_0-12h_ChIP-chip_rel5.gff3
197	kni_2-3h_ChIP_Berk_fdr1_union_rel5.gff3
4553	Kr_0-8h_ChIP_timeunion_rel5.gff3
204	mad_2-3h_ChIP-chip_Berk_fdr1_intervals_sym_rel5.gff
5458	med_2-3h_ChIP-chip_Berk_fdr1_intervals_sym_rel5.gff
2098	prd_2-3h_ChIP_Berk_fdr1_union_rel5.gff3
1478	prospero_4-7h_FDR25_DamID_Brandlab_rel5.gff3
1244	run_0-12h_ChIP_timeunion_rel5.gff3
2884	sbb_0-4h_ChIP-chip_rel5.gff3
19307	sens_4-8h_ChIP_union_rel5.gff3
368	shn_2-3h_ChIP_Berk_fdr1_union_rel5.gff3
1171	slp1_2-3h_ChIP-chip_Berk_fdr1_intervals_sym_rel5.gff
2814	sna_2-3h_ChIP_Berk_fdr1_union_rel5.gff3
105	Stat92E_0-12h_ChIP-chip_rel5.gff3
429	tll_2-3h_ChIP-chip_Berk_fdr1_intervals_sym_rel5.gff
384	ttk_0-12h_ChIP-chip_rel5.gff3
7674	twi_2-3h_ChIP_Berk_fdr1_union_rel5.gff3
1512	Ubx_0-12h_ChIP_timeunion_rel5.gff3
895	zfh1_0-12h_ChIP-chip_rel5.gff3

Appendix 5

I Midline targets

Table 11: All potential midline targets. The genes shown were differentially expressed in at least one of the *sim*Gal4 studies, as well as being associated with Dichaete in vivo binding

Gene identifier	Gene name
CG10214	CG10214
CG10287	Gasp
CG10812	dro5
CG11430	olf186-F
CG12045	Cpr100A
CG12177	CG12177
CG12297	BG4
CG12487	BobA
CG13425	bl
CG13454	CG13454
CG13784	CG13784
CG14147	CG14147
CG14685	Cap-H2
CG14688	CG14688
CG15117	CG15117
CG15319	nej
CG15345	CG15345
CG15658	CG15658
CG17082	CG17082
CG17090	hipk
CG1740	Ntf-2
CG17697	fz
CG18783	Kr-h1
CG18817	Tsp42Ea
CG2177	CG2177
CG2849	Rala
CG31012	cindr
CG31220	CG31220
CG31808	CG31808
CG3396	Ocho
CG34360	Glut4EF
CG3759	CG3759
CG3777	CG3777
CG4147	Hsc70-3
CG42232	CG42232
CG42574	CG42574

Continued on next page

Table 11 – continued from previous page

Gene identifier	Gene name
CG4396	fne
CG4464	RpS19a
CG4547	Atx-1
CG4584	dUTPase
CG4688	CG4688
CG5025	Sps2
CG5514	CG5514
CG5661	Sema-5c
CG6146	Top1
CG6610	CG6610
CG6643	Esyt2
CG6868	tld
CG6874	HipHop
CG6930	l(3)neo38
CG7111	Rack1
CG7380	baf
CG7577	ppk20
CG8050	Cys
CG8235	CG8235
CG8965	CG8965
CG9559	fog
CG9614	pip
CG9628	CG9628
CR32862	snRNA:U1:82Eb
CR42862	CR42862

Table 12: Potential midline targets at stage 9. The genes shown were differentially expressed in at least one of the *sim*Gal4 studies, as well as being associated with Dichaete binding at stage 9, as estimated from DNase I accessibility data.

Gene identifier	Gene name
CG10214	CG10214
CG10657	CG10657
CG10812	dro5
CG11430	olf186-F
CG12045	Cpr100A
CG12177	CG12177
CG12297	BG4
CG12487	BobA
CG13425	bl
CG13454	CG13454
CG13784	CG13784
CG14685	Cap-H2
CG14688	CG14688
CG15319	nej
CG15345	CG15345
CG15658	CG15658
CG17090	hipk
CG1740	Ntf-2
CG17697	fz
CG18783	Kr-h1
CG18817	Tsp42Ea
CG2177	CG2177
CG2849	Rala

Continued on next page

Table 12 – continued from previous page

Gene identifier	Gene name
CG31012	cindr
CG31220	CG31220
CG3396	Ocho
CG34360	Glut4EF
CG3759	CG3759
CG3777	CG3777
CG4147	Hsc70-3
CG42232	CG42232
CG42574	CG42574
CG4396	fne
CG4547	Atx-1
CG4584	dUTPase
CG4688	CG4688
CG5025	Sps2
CG5514	CG5514
CG5661	Sema-5c
CG6146	Top1
CG6610	CG6610
CG6643	Esyt2
CG6868	tld
CG6874	HipHop
CG6930	l(3)neo38
CG7111	Rack1
CG7380	baf
CG7577	ppk20
CG8050	Cys
CG8235	CG8235
CG8965	CG8965
CG9559	fog
CG9614	pip
CG9628	CG9628
CR32862	snRNA:U1:82Eb
CR42862	CR42862

J Neuroblast targets

Table 13: Potential neuroblast targets at stage 9. The genes shown were differentially expressed in at least one of the *pros*Gal4 studies, as well as being associated with Dichaete binding at stage 9, as estimated from DNase I accessibility data.

Gene identifier	Gene name
CG10083	CG10083
CG10283	CG10283
CG10295	Pak
CG10328	nonA-l
CG1034	bcd
CG10343	CG10343
CG10360	ref(2)P
CG10372	Faf
CG1044	dos
CG10492	CG10492
CG1058	rpk
CG10601	mirr

Continued on next page

Table 13 – continued from previous page

Gene identifier	Gene name
CG10686	tral
CG10812	dro5
CG10847	enc
CG10851	B52
CG10859	CG10859
CG1088	Vha26
CG1091	CG1091
CG10939	Sip1
CG10962	CG10962
CG11034	CG11034
CG11140	Aldh-III
CG11148	CG11148
CG11254	mael
CG11279	CG11279
CG1129	CG1129
CG11306	CG11306
CG11430	olf186-F
CG11455	CG11455
CG11504	CG11504
CG11518	pygo
CG11523	CG11523
CG11567	Cpr
CG11614	nkd
CG1172	CG1172
CG11790	CG11790
CG11798	chn
CG11804	ced-6
CG11880	CG11880
CG11897	CG11897
CG11981	Prosbeta3
CG11982	CG11982
CG11993	Mst85C
CG11999	CG11999
CG12047	mud
CG12052	lola
CG12104	CG12104
CG12140	CG12140
CG12177	CG12177
CG12182	CG12182
CG1221	miple
CG12244	lic
CG12297	BG4
CG12399	Mad
CG12414	nAcRalpha-80B
CG12676	ed
CG12690	CHES-1-like
CG12737	Crag
CG12785	Mat89Ba
CG12891	whd
CG13316	Mnt
CG13323	CG13323
CG13391	Aats-ala
CG13425	bl
CG13454	CG13454
CG13778	Mnn1

Continued on next page

Table 13 – continued from previous page

Gene identifier	Gene name
CG13784	CG13784
CG13822	CG13822
CG13897	CG13897
CG14005	CG14005
CG14025	Bsg25D
CG1417	slgA
CG14222	CG14222
CG14230	CG14230
CG14438	CG14438
CG1444	CG1444
CG14478	CG14478
CG14616	l(1)G0196
CG14879	CG14879
CG14997	CG14997
CG15015	Cip4
CG15085	edl
CG15104	Topors
CG15345	CG15345
CG15524	sas-6
CG1594	hop
CG1636	CG1636
CG1643	Atg5
CG1651	Ank
CG1659	unc-119
CG1667	CG1667
CG16757	Spn
CG16785	fz3
CG1692	mal
CG17060	Rab10
CG17090	hipk
CG17149	Su(var)3-3
CG17223	alpha4GT1
CG17299	SNF4Agamma
CG17342	Lk6
CG17369	Vha55
CG1746	CG1746
CG17579	sca
CG17686	DIP1
CG17800	Dscam
CG17818	rdgBbeta
CG17883	CG17883
CG17998	Gprk2
CG18024	SoxN
CG1829	Cyp6v1
CG18350	Sxl
CG1836	Rad23
CG1838	myoglianin
CG1848	LIMK1
CG1862	Ephrin
CG1877	lin19
CG1910	CG1910
CG1921	sty
CG1966	Acf1
CG1981	Thd1
CG2041	lgs

Continued on next page

Table 13 – continued from previous page

Gene identifier	Gene name
CG2048	dco
CG2087	PEK
CG2177	CG2177
CG2186	CG2186
CG2221	l(1)G0289
CG2257	Ubc-E2H
CG2316	CG2316
CG2682	d4
CG2924	CG2924
CG2938	CG2938
CG2944	gus
CG3001	Hex-A
CG3002	Gga
CG3008	CG3008
CG30283	CG30283
CG30388	Magi
CG3082	l(2)k09913
CG3107	CG3107
CG31220	CG31220
CG31223	CG31223
CG3136	Atf6
CG31363	Jupiter
CG31523	CG31523
CG31729	CG31729
CG31794	Pax
CG31991	mdy
CG31992	gw
CG32300	oxt
CG32346	E(bx)
CG32365	CG32365
CG3248	Cog3
CG32491	mod(mdg4)
CG32499	Cda4
CG32555	RhoGAPp190
CG3262	CG3262
CG32654	Sec16
CG3268	phtf
CG32697	l(1)G0232
CG32756	CG32756
CG33196	dp
CG33226	CG33226
CG3329	Prosbeta2
CG3359	mfas
CG33979	capt
CG33991	nuf
CG3403	Mob4
CG3407	CG3407
CG34157	Dys
CG34360	Glut4EF
CG34412	tlk
CG3445	phol
CG3532	CG3532
CG3619	Dl
CG3632	CG3632
CG3688	l(2)35Bd

Continued on next page

Table 13 – continued from previous page

Gene identifier	Gene name
CG3707	wapl
CG3796	ac
CG3871	Six4
CG3925	CG3925
CG3947	pex16
CG3954	csw
CG40293	Stlk
CG4032	Abl
CG4059	ftz-f1
CG4165	CG4165
CG4217	TFAM
CG42232	CG42232
CG42248	CG42248
CG42274	RhoGAP18B
CG42275	alpha-Man-I
CG42281	bun
CG42345	CG42345
CG42378	RhoGEF3
CG42388	CG42388
CG4257	Stat92E
CG42632	mRpL37
CG42803	gpp
CG4300	CG4300
CG4404	CG4404
CG4538	CG4538
CG4572	CG4572
CG4583	Ire1
CG4656	Rassf
CG4673	CG4673
CG4688	CG4688
CG4698	Wnt4
CG4802	CG4802
CG4898	Tm1
CG4912	eEF1delta
CG4993	PRL-1
CG5029	SamDC
CG5052	pim
CG5053	RASSF8
CG5072	Cdk4
CG5263	smg
CG5277	Ip259
CG5295	bmm
CG5362	Mdh1
CG5439	CG5439
CG5490	Tl
CG5588	Mtl
CG5594	kcc
CG5659	ari-1
CG5661	Sema-5c
CG5725	fbl
CG5726	CG5726
CG5788	UbcD10
CG5876	heix
CG5893	D
CG5946	CG5946

Continued on next page

Table 13 – continued from previous page

Gene identifier	Gene name
CG6051	CG6051
CG6099	m4
CG6105	l(2)06225
CG6146	Top1
CG6157	dah
CG6190	Ube3a
CG6214	MRP
CG6292	CycT
CG6335	Aats-his
CG6376	E2f
CG6391	Aps
CG6498	CG6498
CG6509	CG6509
CG6542	EDTP
CG6551	fu
CG6575	glec
CG6584	SelR
CG6634	mid
CG6650	CG6650
CG6668	atl
CG6859	CG6859
CG6889	tara
CG6930	l(3)neo38
CG6932	CSN6
CG6995	Saf-B
CG7001	Pk17E
CG7003	Msh6
CG7005	Esp
CG7041	HP1b
CG7044	CG7044
CG7085	l(2)s5379
CG7129	l(3)05822
CG7133	CG7133
CG7138	r2d2
CG7156	CG7156
CG7184	Mkrn1
CG7192	CG7192
CG7324	CG7324
CG7378	CG7378
CG7400	Fatp
CG7494	mRpL1
CG7554	comm2
CG7577	ppk20
CG7595	ck
CG7597	Cdk12
CG7785	CG7785
CG7869	SuUR
CG7902	bap
CG8014	Rme-8
CG8057	alc
CG8090	CG8090
CG8097	CG8097
CG8104	nudE
CG8108	CG8108
CG8174	SRPK

Continued on next page

Table 13 – continued from previous page

Gene identifier	Gene name
CG8177	CG8177
CG8223	CG8223
CG8233	Rcd1
CG8274	Mtor
CG8287	Rab8
CG8312	CG8312
CG8390	vlc
CG8402	PpD3
CG8420	CG8420
CG8448	mrj
CG8529	Dyb
CG8589	tej
CG8593	qm
CG8631	msl-3
CG8790	Dic1
CG8839	CG8839
CG8924	CG8924
CG8931	CG8931
CG8948	Graf
CG9078	ifc
CG9096	CycD
CG9113	AP-1gamma
CG9126	Stim
CG9153	CG9153
CG9218	sm
CG9238	CG9238
CG9242	bur
CG9446	coro
CG9474	Snap24
CG9493	Pez
CG9670	fal
CG9755	pum
CG9768	hkb
CG9786	hb
CR31273	bxd
CR31400	Hsromega
CR32027	CR32027

Table 14: Potential neuroblast targets at stage 9, with annotated roles in the generation of neurons. The genes shown were differentially expressed in at least one of the *pros*Gal4 studies, as well as being associated with Dichaete binding at stage 9, as estimated from DNase I accessibility data. The selection was further based on which genes are associated with the 'generation of neurons' GO term.

Gene identifier	Gene name
CG17579	sca
CG5661	Sema-5c
CG1848	LIMK1
CG12047	mud
CG11804	ced-6
CG2682	d4
CG17800	Dscam
CG4898	Tm1
CG9218	sm

Continued on next page

Table 14 – continued from previous page

Gene identifier	Gene name
CG3359	mfas
CG11518	pygo
CG9755	pum
CG3954	csw
CG15085	edl
CG4032	Abl
CG12052	lola
CG3796	ac
CG10295	Pak
CG6376	E2f
CG1966	Acf1
CG18024	SoxN
CG9242	bur
CG1044	dos
CG32555	RhoGAPp190
CG6634	mid
CG5490	Tl
CG6190	Ube3a
CG1594	hop
CG9768	hkb
CG4698	Wnt4
CG32346	E(bx)
CG3619	Dl
CG1921	sty
CG1862	Ephrin
CG3403	Mob4
CG42281	bun
CG11614	nkd
CG9786	hb
CG12676	ed

Table 15: Potential neuroblast targets at multiple Dichaete binding at stage 9. The genes shown were differentially expressed in at least one of the *pros*Gal4 studies, as well as being associated with 4 or more intervals of Dichaete binding at stage 9, as estimated from DNase I accessibility data.

Gene identifier	Gene name
CG11140	Aldh-III
CG11148	CG11148
CG11430	olf186-F
CG11614	nkd
CG11798	chn
CG12052	lola
CG12414	nAcRalpha-80B
CG12676	ed
CG12690	CHES-1-like
CG13316	Mnt
CG13897	CG13897
CG15085	edl
CG16785	fz3
CG18024	SoxN
CG1921	sty
CG2048	dco
CG31363	Jupiter

Continued on next page

Table 15 – continued from previous page

Gene identifier	Gene name
CG31523	CG31523
CG32491	mod(mdg4)
CG33196	dp
CG33991	nuf
CG34360	Glut4EF
CG34412	tlk
CG3619	Dl
CG42281	bun
CG4257	Stat92E
CG5295	bmm
CG5490	Tl
CG5661	Sema-5c
CG5893	D
CG6376	E2f
CG6575	glec
CG6634	mid
CG6889	tara
CG7001	Pk17E
CG7554	comm2
CG8312	CG8312
CG8448	mrj
CG9238	CG9238
CG9755	pum
CG9768	hkb
CG9786	hb
CR31273	bxd

Table 16: GO enrichments of multiply bound neuroblast targets

GO term	p-value	No. of genes
organ development [GO:0048513]	4.9152E-7	21
positive regulation of biological process [GO:0048518]	1.2E-6	15
cell fate commitment [GO:0045165]	1.7E-6	12
regulation of cell fate specification [GO:0042659]	2.7E-6	6
embryonic pattern specification [GO:0009880]	3.2E-6	11
regulation of cell fate commitment [GO:0010453]	3.6E-6	6
positive regulation of cellular process [GO:0048522]	4.7E-6	14
embryo development [GO:0009790]	4.8E-6	15
cell fate specification [GO:0001708]	6.4E-6	8
pattern specification process [GO:0007389]	9.9E-6	14
generation of neurons [GO:0048699]	2.3E-5	15
regionalization [GO:0003002]	4.7E-5	13
system development [GO:0048731]	7.2E-5	24
multicellular organismal development [GO:0007275]	1.3E-4	26
segmentation [GO:0035282]	1.4E-4	10
nervous system development [GO:0007399]	1.5E-4	20
anatomical structure morphogenesis [GO:0009653]	1.7E-4	20
regulation of cell differentiation [GO:0045595]	2.1E-4	10
regulation of developmental process [GO:0050793]	2.7E-4	12
blastoderm segmentation [GO:0007350]	4.2E-4	9
locomotion [GO:0040011]	5.8E-4	12
central nervous system development [GO:0007417]	6.2E-4	9
anatomical structure development [GO:0048856]	1.1E-3	25
developmental process [GO:0032502]	1.6E-3	26

Continued on next page

Table 16 – continued from previous page

GO term	p-value	No. of genes
transcription, DNA-dependent [GO:0006351]	1.7E-3	15
RNA biosynthetic process [GO:0032774]	1.8E-3	15
regulation of biological process [GO:0050789]	2.1E-3	26
multicellular organismal process [GO:0032501]	2.1E-3	28
positive regulation of cellular biosynthetic process [GO:0031328]	2.8E-3	8
positive regulation of biosynthetic process [GO:0009891]	2.9E-3	8
regulation of cellular process [GO:0050794]	2.9E-3	25
regulation of transcription, DNA-dependent [GO:0006355]	3.7E-3	14
regulation of RNA biosynthetic process [GO:2001141]	3.7E-3	14
regulation of cellular biosynthetic process [GO:0031326]	3.9E-3	15
regulation of biosynthetic process [GO:0009889]	3.9E-3	15
organ morphogenesis [GO:0009887]	4.2E-3	12
negative regulation of cell fate specification [GO:0009996]	4.2E-3	4
negative regulation of cell fate commitment [GO:0010454]	4.2E-3	4
cell surface receptor linked signaling pathway [GO:0007166]	5.2E-3	13
cell fate determination [GO:0001709]	5.8E-3	7
positive regulation of cellular metabolic process [GO:0031325]	6.6E-3	8
positive regulation of metabolic process [GO:0009893]	8.6E-3	8
regulation of RNA metabolic process [GO:0051252]	0.01	14
positive regulation of macromolecule biosynthetic process [GO:0010557]	0.01	7
biological regulation [GO:0065007]	0.01	26
regulation of macromolecule biosynthetic process [GO:0010556]	0.01	14
regulation of cellular macromolecule biosynthetic process [GO:2000112]	0.01	14
cellular developmental process [GO:0048869]	0.01	20
negative regulation of cell differentiation [GO:0045596]	0.01	6
regulation of multicellular organismal process [GO:0051239]	0.02	10
negative regulation of biological process [GO:0048519]	0.02	13
sensory organ development [GO:0007423]	0.03	10
positive regulation of cell proliferation [GO:0008284]	0.03	4
neuroblast fate commitment [GO:0014017]	0.03	4
regulation of nucleobase-containing compound metabolic process [GO:0019219]	0.03	14
regulation of nitrogen compound metabolic process [GO:0051171]	0.03	14
cell differentiation [GO:0030154]	0.03	19
positive regulation of macromolecule metabolic process [GO:0010604]	0.03	7
neuron differentiation [GO:0030182]	0.03	11
axis specification [GO:0009798]	0.05	7
negative regulation of cellular process [GO:0048523]	0.05	12

K Global targets

Table 17: High probability direct Dichaete targets during early embryogenesis. The genes shown were differentially expressed in a *Dichaete* mutant, as well as being associated with 4 or more Dichaete high confidence binding intervals

Gene identifier	Gene name
CG10016	drm
CG10021	bowl
CG10079	Egfr
CG10109	L
CG1028	Antp
CG1030	Scr
CG10325	abd-A

Continued on next page

Table 17 – continued from previous page

Gene identifier	Gene name
CG10388	Ubx
CG10479	CG10479
CG10488	eyg
CG10601	mirr
CG10619	tup
CG10772	Fur1
CG10798	dm
CG11100	Mes2
CG11140	Aldh-III
CG11186	toy
CG11280	trn
CG1133	opa
CG11502	svp
CG11614	nkd
CG11648	Abd-B
CG12154	oc
CG12212	peb
CG12287	pdm2
CG12487	BobA
CG12673	olf413
CG13076	Notum
CG13194	pyr
CG1343	Sp1
CG1374	tsh
CG1429	Mef2
CG1449	zfh2
CG14926	CG14926
CG15138	beat-IIIc
CG15154	Socs36E
CG1634	Nrg
CG1691	Imp
CG17046	klar
CG17077	pnt
CG17117	hth
CG17348	drl
CG1771	mew
CG17941	ds
CG18024	SoxN
CG18350	Sxl
CG18657	NetA
CG1897	Dr
CG1921	sty
CG1958	CG1958
CG2125	ci
CG2621	sgg
CG2692	gsb-n
CG2969	Atet
CG2977	inx7
CG30015	CG30015
CG30084	Zasp52
CG3048	Traf4
CG31043	gukh
CG31363	Jupiter
CG31481	pb
CG32019	bt

Continued on next page

Table 17 – continued from previous page

Gene identifier	Gene name
CG32062	A2bp1
CG3242	sob
CG32434	siz
CG3258	ase
CG32592	hiw
CG32676	CG32676
CG33196	dp
CG33232	CG33232
CG3340	Kr
CG34110	lobo
CG34360	Glut4EF
CG34368	Fili
CG34371	CG34371
CG34379	Shroom
CG34380	CG34380
CG34389	cv-c
CG34395	nub
CG3474	Cpr35B
CG3578	bi
CG3619	Dl
CG3796	ac
CG42230	bbg
CG42238	CG42238
CG42281	bun
CG42311	grh
CG42333	Sytbeta
CG42589	sick
CG42611	Megalin
CG42684	CG42684
CG42865	trh
CG43140	pyd
CG4319	rpr
CG4491	noc
CG4717	kni
CG4722	bib
CG4889	wg
CG4974	dally
CG5123	W
CG5187	Doc2
CG5201	Dad
CG5481	lea
CG5661	Sema-5c
CG5723	Ten-m
CG5803	Fas3
CG5893	D
CG6027	cdi
CG6531	wgn
CG6570	lbl
CG6889	tara
CG7100	CadN
CG7230	rib
CG7250	Toll-6
CG7450	CrebA
CG7649	Neu3
CG7807	AP-2

Continued on next page

Table 17 – continued from previous page

Gene identifier	Gene name
CG7958	tna
CG8118	mam
CG8127	Eip75B
CG8355	sli
CG8376	ap
CG8676	Hr39
CG8896	18w
CG9102	bab2
CG9224	sog
CG9381	mura
CG9650	CG9650
CG9656	grn
CG9704	Nrt
CG9739	fz2
CG9755	pum
CR42862	CR42862