The translational landscape of fission yeast meiosis and sporulation

Caia D. S. Duncan and Juan Mata

Department of Biochemistry, University of Cambridge, Cambridge CB2 1QW, United Kingdom

Correspondence to Juan Mata, jm593@cam.ac.uk

Abstract

Sexual development in the fission yeast *Schizosaccharomyces pombe* culminates in meiosis and sporulation. We used ribosome profiling to investigate the translational landscape of this process. We show that the translation efficiency of hundreds of genes is regulated in complex patterns, often correlating with changes in RNA levels. Ribosome-protected fragments show a three-nucleotide periodicity that identifies translated sequences and their reading frame. Using this property, we identified 46 novel translated genes and found that 24% of non-coding RNAs are actively translated. We also detected 19 nested antisense genes, in which both DNA strands encode translated mRNAs, and discovered trends that shape the evolution of these genes. Finally, we identified 1,735 translated upstream ORFs in leader sequences. In contrast with *Saccharomyces cerevisiae*, sexual development in *S. pombe* is not accompanied by large increases in upstream ORF use, suggesting that this is an organism-specific adaptation and not a general feature of developmental processes.

Introduction

Fission yeast diploid cells undergo sexual differentiation (meiosis and sporulation) under conditions of nitrogen starvation[1]. This developmental process is accompanied by a complex gene expression program, in which more than 50% of the genome is regulated. Early microarray studies showed that changes in RNA levels occur in successive expression waves that coincide with major biological events: starvation and pheromone-induced genes (response to environmental changes), early genes (pre-meiotic S phase and recombination), middle genes (meiotic divisions and early steps of spore formation) and late genes (spore maturation)[2]. The induction of most of these groups is mediated by meiosis-specific transcription factors[2-5], although additional regulation of mRNA decay by RNA-binding proteins contributes to shaping the kinetics of the changes[6]. In addition, a subset of early genes appears to be induced solely by changes in RNA stability[7]. More recently, RNA-seq experiments revealed that *S. pombe* cells express hundreds of potentially non-coding RNAs (ncRNAs), many of which are meiosis-specific[8-10]. Although gene expression during sexual differentiation has been extensively studied, nothing is known about the contribution of translational control to this process.

Ribosome profiling can provide a genome-wide view of translation with single-nucleotide resolution[11]. The approach is based on the isolation and sequencing of ribosome-protected mRNA fragments (RPFs), and can be used to identify translated regions and to estimate mRNA translational efficiency (reviewed in 12). This approach was recently applied to the meiotic program of the budding yeast *S. cerevisiae,* revealing extensive translational control and a meiosis-specific increase in the use of open reading frames upstream of annotated coding sequences (uORFs)[13]. However, it is unclear if these are general features of differentiation processes or organism-specific properties. Ribosome profiling of budding yeast and higher eukaryotes also exposed that many genes annotated as ncRNAs appear in the RPF fractions[11,14]. However, whether these ncRNAs are actively translated has been contentious[15,16]. A key advantage of ribosome profiling is that RPFs show a characteristic

triplet periodicity when aligned to mRNA sequences that can be used to determine the frame in which a given region is being translated[11]. This property has been used to identify dually decoded regions[17], but has not been employed systematically to analyse every aspect of a translational landscape. We use it here to discover newly translated genes, to detect translated uORFs and to address the question of whether genes annotated as long ncRNAs are actively translated.

We have used ribosome-profiling to investigate the translational landscape of *S. pombe* cells during vegetative growth and sexual differentiation. We discovered dozens of new genes (including nested antisense transcripts), identified translated uORFs in the 5' leader sequences of 25% of *S. pombe* coding genes, and found that 24% of genes annotated as ncRNAs are actively translated. These results reveal pervasive translation of the fission yeast genome. We also show that sexual differentiation in *S. pombe* is accompanied by a complex translational program, in which the translation efficiency of hundreds of genes is regulated in complex patterns.

Results and discussion

Ribosome profiling in S. pombe

We carried out ribosome profiling with *S. pombe* diploid cells undergoing meiosis and sporulation. To achieve good synchrony we used thermo-sensitive mutants in the Pat1 meiotic inhibitor[18,19]. Diploid cells were blocked in G1 by nitrogen starvation, and entered meiosis synchronously upon inactivation of the Pat1 kinase at increased temperature (Figure S1). In addition, we performed ribosome profiling in wild type haploid cells growing vegetatively. Ribosome profiling involves the purification of RPFs, which are used to generate a library containing adaptors for Illumina sequencing. In parallel, a second library is produced from fragmented mRNAs. The first library allows the identification and quantification of translated regions, while the second one is used for the estimation of mRNA levels. In all current protocols, the production of both libraries involves

PCR-based amplification. As part of our library generation strategy, we used Unique Molecular Identifiers (UMIs)[20,21]. Here, the primer employed for reverse transcription contains a random barcode sequence, which uniquely tags each cDNA fragment. Reads that contain identical sequences with the same barcode are very likely to have originated from the same RNA fragment, and only one of them is retained. Therefore, the reads we used for analysis constituted a truly non-redundant dataset. This approach helps avoid PCR amplification and sampling artefacts, and therefore can also be used to estimate the complexity of the original library[21]. 83 % of reads that did not map to rRNAs were unique, indicating that the complexity of the original libraries was high (Table S1). Overall, 707 million reads were reduced to 250 million unique ones (Table S1). Reads were processed and mapped to the *S. pombe* genome as described in Methods. As expected, RPFs showed a strong bias in their distribution along mRNAs: After accounting for length, 92.4 % of them mapped to annotated coding sequences, 6.9 % were located in 5' leader sequences (also known as 5' Untranslated Regions, or 5'-UTRs) and only 0.7 % in 3'-UTRs. By contrast, mRNA reads showed a much more equal distribution (Figure S2). The results were highly reproducible between independent biological repeats (Figure S3 and Methods).

We used the densities of mRNA fragments and RPFs to estimate mRNA levels and protein synthesis rates, respectively. Both features correlated well with each other over the whole meiotic time course (Table S2, average R= 0.79). Total translation rates (measured as RPF densities) are expected to be better predictors of protein abundance than mRNA levels. Consistently, protein levels estimated from a mass spectrometry study of vegetative cells[22] showed higher correlation with RPF density of cells in similar growth conditions (Figure 1A, R = 0.82) than with RNA levels (Figure 1B, R = 0.68). Direct comparison of RPFs and RNA levels allowed us to estimate gene-specific translation efficiencies (TEs), which were calculated by dividing the normalised number of RPFs (in reads per kilobase per million reads or RPKM) by that of mRNA fragments across coding sequences. In vegetative cells, TEs varied over a range of over 100-fold, and did not correlate with mRNA half-lives (R = -0.12, Figure S4) or mRNA levels (average R over the time course = -0.08, Table S2). As

expected, TEs displayed a positive correlation with the mean number of bound ribosomes determined in a microarray-based polysome profiling study (R = 0.43, Figure S4)[23]. Poor TE in vegetative cells was associated with lowly expressed genes, including all major groups of meiotic genes. Moreover, targets of the nonsense-mediated decay (NMD) pathway[24] were very poorly translated. This phenomenon has been previously observed in *S. cerevisiae*, although it is unclear whether the NMD pathway directly represses translation or whether these mRNAs contain features that make their translation inefficient[25,26].

The translational program of meiosis

We used hierarchical clustering to investigate the changes in RPF levels across the meiotic time course. Translation patterns were highly dynamic and generally correlated with changes in RNA expression (Figure S5). Changes in RNA levels were similar to those reported in a previous microarray study, thus providing independent validation of our data (Figure S5)[2]. As in the case of vegetative cells, TEs varied over a wide range in all time points of the meiotic time course, and were dynamically regulated. 25.8% of all coding genes showed changes of 5-fold of more in TE, and 6.7% of more than 10-fold. We looked for patterns in TE regulation by clustering the 318 genes that showed the strongest variation (Figure 1C). Several groups of clustered genes were enriched in co-expressed genes. For example, a subset of late genes (cluster 1) is poorly translated until 7 hours into meiosis, when its TE increases (Figure 1D). Interestingly, this group is enriched in genes regulated by the Atf21/Atf31 transcription factors, but not in other late genes under the control of other transcription factors. Similarly, a group of middle genes is only translated efficiently at 5 hours (cluster 2), and a cluster enriched in genes expressed in response to nitrogen starvation (cluster 4) shows enhanced TE after nitrogen removal (Figure 1D). In all three cases the peak of TE coincides with that of mRNA levels. Increases in TE coupled to those in RNA levels have been observed before during responses to stress, a phenomenon that has been called potentiation[27-29]. A cluster that

showed a strong drop in TE at 5 hours despite high mRNA levels (cluster 3) was enriched in targets of the Meu5 RNA-binding protein (Figure 1D). Meu5 binds to a subset of the middle genes (so called 'late-decay') and stabilises their transcripts, thus allowing them to be expressed during a longer time window[6]. Although the RNA levels of many of these genes peak at 5 hours, the accumulation of RPFs was delayed compared to that of mRNA fragments. This suggested that translation of some Meu5 targets genes is repressed to delay the production of the corresponding protein. We checked this hypothesis by looking at published protein levels during a meiotic time course for Meu5 targets that did or did not show the decrease in TE (see Figure S11 in reference [6]). In all three cases, the peak of protein expression corresponded to that of RPF accumulation. Finally, we noticed that targets of the Mmi1 protein, an RNA-binding protein that promotes the degradation of a group of meiotic RNAs in vegetative cells [7], showed strong changes in TE, with high efficiency correlated with the expression of their mRNAs and a strong repression in vegetative cells, when expression of these genes is toxic for the cell (Figure 1D). These results suggest that the regulation of TE and the control of mRNA levels are coordinated during the meiotic process. However, this relationship appears to be complex and subtle: in some cases it caused increased translation at peaks of RNA levels (potentiation), whereas in others (some Meu5 targets) it led to delays in protein production with respect to RNA accumulation. Moreover, the majority of groups that showed coordinated changes in TE were not enriched in any known group of genes, suggesting that additional layers of translational control remain to be discovered.

Systematic identification of novel translated regions

In a 28-nucleotide long RPF, nucleotide 13 will typically correspond to the first nucleotide of the codon located at the P site (Figure 2A)[11]. As expected, we found that nucleotide 13 mapped to the first nucleotide within a codon in 75 % of the 28-nucleotide RPFs, while mRNA fragments were equally distributed across all three nucleotides (Table S3). When the data are aggregated for all

annotated coding sequences, this effect leads to a 3-nucleotide periodic pattern, which is observed at all positions (Figure 2B). As has been previously noted, there is an accumulation of RPFs at the initiation codon (Figure 2B) and at the last codon (data not shown)[11,14]. By contrast, mRNA fragments did not show any periodic behaviour (Figure 2B). As long as the RPF coverage is high enough, this unique feature allows the detection of any genomic region that is being translated, as well as the identification of its reading frame. Although translated sequences can in principle be identified simply by the accumulation of RPFs, the use of periodicity allows the distinction between translated regions and contaminants in the RPF sample. For example, mRNA fragments could be protected by the binding of non-ribosomal RNA-binding proteins. We therefore decided to use read periodicity as the basis to define translated regions in parts of the genome with no annotated features ('intergenic regions'), in genes annotated as non-coding RNAs, and in 5' leader sequences. We note that the experimental approach and all the analyses presented below are strand-specific, so regions of the genome containing a feature on one strand were still analysed for the other one. For all three groups of regions we followed the same experimental strategy (see Methods for details): First, we defined all possible ORFs starting with AUG as well as rarer initiation codons. Second, each ORF was screened for the presence of periodic signal from RPFs. Third, a false discovery rate (FDR) was estimated based on a randomization test, and was used to fine-tune the threshold values used to call an ORF as translated.

To assess the specificity of the approach we calculated periodicity scores, which measures the fraction of codons translated in each reading frame (see Methods). This was done for all annotated *S. pombe* coding sequences, using both ribosome footprint and mRNA data. As shown in Figure S6, the periodicity score discriminated clearly between both datasets. Using this approach we could validate the translation in the predicted frame of 4,923 out of 5,102 annotated high-confidence coding sequences (96.5%)[30]. We manually examined genes that displayed periodicity in unexpected frames. This led to the discovery of several mis-annotated genes (Table S4A), in which translation takes place in a frame different from the annotated one (Figure S7) or in which the

annotation of one or more exons was incorrect (Figure S8A and S8B). We also found a clear case of alternative intron retention, which is very rare in *S. pombe* (Figure S8B). A total of 71 coding genes in *S. pombe* are annotated as dubious[30], indicating that the evidence for their existence is poor. Inspection of their behaviour revealed that only 11 appeared to be clearly translated as predicted (15%), and 3 were translated in frames different from those annotated. In addition, there were 25 cases (35%) in which the mRNA was well expressed but no translation was observed, strongly suggesting that these predicted genes may not be translated (Table S4B). In the remaining genes, expression levels were not high enough to allow the evaluation of translation. These results validate, refine and improve the annotation of the fission yeast genome.

This analysis also revealed instances of overlapping coding sequences in the same strand (dually decoded regions). For example, *SPAC3C7.15c* is translated from a long mRNA in the predicted frame during early meiosis. In late meiosis, a short transcript appears that is completely enclosed within the long form of the mRNA but is translated in a different frame (Figure 2C and 2D). We epitope-tagged the predicted short form of the protein and observed a meiosis-specific polypeptide of the expected size, confirming that translation of the short ORF leads to the production of a stable polypeptide (Figure 2E). The sensitivity of the approach is also exemplified by the fact that it allowed the discovery of changes in translation caused by single-nucleotide polymorphisms in our strains compared to the reference strain. We detected translation of the C terminal part of the Nup184 protein, which was not predicted to be translated in *S. pombe* due to an in-frame stop codon created by a single nucleotide deletion[31]. Translation of this region in the strains we used was explained by the presence of single nucleotide insertion that reverts the effect of the deletion in the reference sequence (Figure S9).

We then scanned intergenic regions systematically, and identified 715 translated ORFs with an estimated FDR of 10.3 % (Table S5). The majority of these regions were short (Figure S10) and had a strong tendency to use AUG as initiation codon (90.6%, Table S6). We manually inspected the

46 translated ORFs of 45 codons or more. Examination of the corresponding mRNA data suggested that 39 of them (85%) were transcribed as independent units (see Figure S11 for examples). Of the remaining 6, 2 appeared to be extensions to annotated 5' leaders, 3 were present in the 3'-UTRs of highly translated genes (suggesting that they may reflect leaky termination of translation) and one was located downstream of the *nup184* gene (discussed above). 15 translated ORFs did not overlap with annotated coding sequences (in either strand). Their sequences were generally not conserved, but one of them displayed homology to the N-terminal part of a protein present in multiple copies in *S. cryophilus* and *S. octosporus*. Sequences immediately downstream of the termination codon of the ORF were homologous to the C-terminal part of the proteins, but were not translated in *S. pombe*. Surprisingly, we found 14 cases in which the newly-discovered translated ORF were antisense to an annotated coding sequence (completely overlapping in 10 cases, and more than 80% in the other 4) (Figure 2F and G, Figure S12). In addition, we found a similar situation (with complete overlap) in 5 annotated antisense ncRNAs (see below). These exonic Nested Antisense Genes (eNAGs) are extremely rare in eukaryotic cells, with a single case described in *S. cerevisiae* (the *NAG1* gene)[32]. Simultaneous coding on both strands imposes very strict constraints on the evolution of both proteins. The exact nature of these limitations depends on the relative frame of the sense and antisense coding sequences, which can adopt three different arrangements (Figure S13). We examined this configuration in all 15 *S. pombe* eNAGs that are completely nested within the major coding sequence, and found that in 11 cases arrangement 3 was preferred. In the case of *S. cerevisiae NAG1*, configuration 3 was also used (Figure S13). This organization causes very stringent and specific dependencies between the sequences encoded in both strands, as nucleotides 1 and 2 of every codon on both strands are encoded by the same DNA sequence. Because codons 1 and 2 possess most information content, the encoding of a particular amino acid on one strand can in some cases determine the nature of the amino acid on the antisense strand. For example, proline is encoded by CCN, and glycine by GGN; in arrangement 3, the presence of a CCN codon necessarily implies a GGN codon in the antisense gene (and vice versa). Our results suggest that a specific

arrangement of reading frames is evolutionary favoured when both strands of DNA are translated. In almost every case the new eNAG transcript was meiosis–specific, suggesting that there may be undiscovered eNAG genes in the *S. pombe* genome that are only expressed in other conditions.

Translation of annotated non-coding RNAs

The *S. pombe* genome contains 1,571 annotated long non-coding RNAs (ncRNAs) [9,10,30]. Visual inspection of our data revealed the presence of numerous meiosis-specific novel genes (data not shown). ncRNAs tended to be expressed at low levels, with a median read density of 3.9 RPKM (compared to 39.4 for coding sequences). In vegetative cells and during most of meiosis, reads mapped to ncRNAs accounted for less than 3.5% of all reads. However, in mid-meiosis (3 and 5 hour time points) this number increased to 11%, suggesting that ncRNA function may be especially important during cellular differentiation. Recent work in several eukaryotes has revealed that ncRNAs are often present in ribosomal fractions, although it is unclear to which extent this association represents active translation[11,14-16]. We used triplet periodicity as described above to address this question. We identified 499 translated regions in 375 genes, with a FDR of 7.5% (Table S7). These regions had a strong tendency to start with the canonical AUG (96.8%, Table S6). Their median length was 21 codons, and 37 were longer than 45 (Figure 3A). 28% of translated ncRNAs contained more than one translated ORF (Figure 3B). In some cases, these translated regions were reminiscent of upstream ORFs, with a longer coding sequence preceded by several upstream translated ORFs of small length. ORFs in ncRNAs have the potential to produce short polypeptides. To investigate if this is the case, we epitope-tagged two predicted peptides from the *prl3* (*SPNCRNA.03*) and *prl46* (*SPNCRNA.46*) ncRNA genes. The *prl3* gene is well expressed in vegetative cells and contains an ORF of 49 codons that appears to be highly translated (Figure 3C and 3D). Tagging of the corresponding peptide allowed the detection of a protein of the predicted molecular weight, confirming that the *prl3* gene encodes an expressed polypeptide (Figure 3G). The *prl46* gene

is specifically expressed during meiosis (3-5 hours). We detected a translated sequence of 86 codons that partially overlapped with the annotated *prl46* gene, but contained an additional extension at the 5' side (Figure 3E and 3F). Epitope tagging of the predicted peptide revealed the presence of a protein of the expected molecular weight, which was specifically expressed during mid-meiosis (Figure 3H). The majority of the translated sequences do not have homologs in other organisms, although there are some exceptions: The Prl46 protein and that encoded by *SPNCRNA.557* are conserved in *S. cryophilus* and *S. octosporus*, and the peptide encoded by *SPNCRNA.1597* is conserved in *S. cryophilus*, *S. octosporus* and *Pneumocystis*.

These results show that 24% of genes that have been annotated as ncRNAs are translated to produce short peptides. However, the majority of ncRNAs associate with ribosomes to translate ORFs of very few codons, supporting the idea that the distinction between coding and non-coding RNAs is not clear-cut. The functional importance of these observations is still unclear, although several not mutually exclusive explanations are possible. First, short polypeptides encoded by ncRNAs may have biological activity. For example, peptides of 11 to 32 amino acids from the *Drosophila tarsal-less (tal)* gene (also called *polished rice*) regulate the development of embryonic ectodermal structures and the leg tarsus[33,34], peptides shorter than 30 amino acids modulate calcium uptake and cardial function in *Drosophila*[35], the *S. pombe* ribosomal protein Rpl41 is only 25 amino acid long[30], and the Mat-Mi protein, consisting of only 42 amino acids, is a key regulator of fission yeast meiosis that functions as a transcription factor[36]. A recent peptidomic study aimed at identifying short peptides in K562 cells (human leukaemia) identified 90 polypeptides encoded in 5' leader regions and long intergenic non-coding RNAs (lincRNAs), suggesting that the production of stable small peptides may be widespread[37]. Second, translation could be used to target ncRNAs to polysomes, where they might be degraded (through NMD) or perform functions in translational control. Finally, translation of short ORFs in ncRNAs may reflect pervasive translation in which capped and polyadenylated sequences that reach the cytoplasm would be translated to some degree, even if their translation was not required for their role.

Translation of 5' leader sequences

6.9 % of all reads mapped to 5' leader sequences. Translated ORFs in 5' leaders (upstream ORFs, or uORFs) have the potential to regulate translation, although their effect can be neutral, positive or negative[38]. Although translation of a uORF by a ribosome may down-regulate translation by preventing it from reaching a downstream ORF, ribosome small subunits have the potential to perform scanning after termination and may recognize downstream initiation codons (reinitiation)[39]. In addition, uORFs can encode short peptides that are stably expressed[37]. We examined all predicted ORFs in annotated 5' leaders for periodic footprint patterns, and identified 1,735 translated uORFs in 1,272 genes, with an estimated FDR of 10.0% (Table S8). 26% of the genes contained more than a single uORF, and 7% had 3 or more (Figure 4A). The latter group was enriched in genes encoding transcription factors, periodically expressed genes, middle meiotic genes and genes induced in response to nitrogen starvation. Interestingly, mRNAs encoding four key regulators of meiosis contain four or more uORFs: *pat1* and *mei2,* which encode a kinase and an RNA-binding protein, respectively, that control entry into meiosis[18,19,40], *ste11*, encoding a transcription factor that mediates the response to nitrogen starvation and pheromone communication[5], and *atf21*, which codes for a transcription factor responsible for the induction of a large subset of the late genes[2]. uORFs had a mean length of 13.1 codons, and a median of 10 (Figure 4B). Some were extremely short, with 6.8% containing a single AUG codon (Figure 4C). Similar to ORFs, uORFs displayed enrichment of reads in the last codon. 5' leader sequences contained less ORFs starting with AUG compared to other genomic regions (Table S6), and translated uORFs used AUG as a start codon less frequently (73.9%, Table S6).

We also identified 175 uORFs that overlapped partially with the main ORF (but in a different reading frame) (Table S9). These may be particularly important as repressor elements, as they would usually not allow reinitiation of translation in the frame of the main ORF[39]. Among them,

there were three cases in which a strongly translated uORF overlapped substantially with a downstream ORF that was actively but poorly translated (Figure 4D and Figure S14). In the most extreme case (*SPCC1235.01*) the uORF was 347-codon long, starting 68 nucleotides upstream of the ORF (Figure S14A). Interestingly, the amino acid sequence encoded at the 5' end of the uORF (that does not overlap with the ORF) is conserved in other *Schizosaccharomyces* species, suggesting that both reading frames may be used to produce functional proteins. *SPAC11D3.13* (Figure 4D) and *SPAC6G9.05* (Figure S14B) presented a similar structure, although the length of the uORFs was shorter (56 and 51 codons, respectively). *SPAC11D3.13* had an additional upstream uORF, in an arrangement reminiscent of the regulatory uORFs of the mammalian *ATF4* and *ATF5* mRNAs, which are only translated in stress situations (Figure 4D) (reviewed in 39).

In *S. cerevisiae*, there is an increase in the use of uORFs during meiosis[13]. We examined if a similar phenomenon takes place in *S. pombe* by comparing the ratio between the TEs of every uORF and its corresponding ORF for every time point (Figure S15). Although we observed increases towards the end of meiosis, the changes were generally small, indicating that in *S. pombe* neither nitrogen starvation nor the meiotic program cause a large rise in the use of uORFs.

To investigate global effects of uORFs on translation, we compared the TE for every uORF to that of its downstream ORF. In general, the TEs of uORFs and main coding sequences did not correlate (average R for all time points = 0.20), consistent with the complex and varied function of uORFs[38]. We then examined whether changes in TE of uORFs during the meiotic time course correlated with variations in TE of downstream coding regions (see examples in Figure S15). To do that, we calculated the correlation between the TE of every uORF and their downstream ORFs across all time points. The majority of genes showed strong positive correlations (Figure 4E), although a very small subset displayed a negative relationship. An example of the latter group is shown in Figure 4F. In *SPBC1773.04*, a meiosis-specific 5' extension to the mRNA contains two uORFs whose translation is negatively correlated with that of the main ORF, suggesting that the uORFs compete

for translation with and down-regulate translation of the downstream ORF. A very similar example is displayed in Figure 2F, in this case concerning an antisense transcript. Although temporal regulation of TE through changes in the use of uORFs is relatively common in *S. cerevisiae* meiosis[13], it appears to be rare in *S. pombe*.

Conclusions

Our results reveal pervasive translation of the *S. pombe*, including dually decoded regions, exonic nested antisense genes and frequent translation of annotated ncRNAs. Overall, we found 917 translated ORFs of 20 codons or longer (in 5' leader sequences, annotated ncRNAs, and novel translated regions), suggesting the existence of a huge repertoire of small peptides with potential biological functions. We have experimentally validated instances of several of these types of features, demonstrating that this complexity results in real changes to the proteome. In addition, we have observed substantial use of TE regulation during meiosis, including homodirectional changes in RNA levels and TE (potentiation) and the use of TE changes to delay protein accumulation. The existence of a previous dataset from *S. cerevisiae* allowed us to address if properties of the translational programs of cellular differentiation processes are general or organism-specific. Both *S. pombe* and *S. cerevisiae* use extensive translational control, as demonstrated by the widespread and dynamic changes in TE. In contrast to *S. cerevisiae*, we do not detect a switch to the use of uORFs and unconventional initiation mechanisms, suggesting that this is not a general feature of meiosis or cellular differentiation processes.

Methods

General methods: Standard methods and media were used for fission yeast growth. Wild type and *pat1*-driven meiosis were induced as described[2]. Vegetative cells were grown in rich medium at 32°C. Proteins were tagged with TAP. Tagged proteins were detected by Western blot using peroxidase–anti-peroxidase-soluble complexes (Sigma) and tubulin with mouse monoclonal antibodies (Sigma). For protein detection of Prl3 and Prl46, 50 ml of cells (3-8 106 cells/ml) were treated with 1 mM PMSF for 5 min before centrifugation at 4 °C and freezing. Cells were later thawed, washed with cold lysis buffer (20 mM Tris-HCl pH 8.0, 140 mM KCl, 1.8 mM MgCl2 and 0.1% NP-40), resuspended in 200 µl of lysis buffer containing 1 mM PMSF and 1:100 protease inhibitor cocktail (Sigma P8340), and lysed using a bead beater (Fastprep, MP Biomedicals) at level 6 for 13 seconds. Extracts were cleared by centrifugation at 7,600 g for 5 minutes at 4°C and used for Western blotting. The short peptide overlapping SPAC3C.15c was not visible under these conditions. In order to detect it, cells were washed with cold RIPA buffer (10 mM Tris-HCl pH 7.4, 150 mM NaCl, 2 mM EDTA, 1% Triton X100, 0.1% SDS), resuspended in 200 µl of RIPA buffer, boiled for 5 minutes and frozen. Samples were then processed as described above except that RIPA buffer was used and the extracts were not precleared by centrifugation.

Ribosome profiling, library preparation and sequencing: $3x10^8$ exponentially growing wild type haploid cells, or between $3x10^8$ and $12x10^8$ meiotic cells, were incubated for 5 minutes with 100 µg/ml cycloheximide, pelleted at 4°C and frozen in liquid N_2. Each culture was split for footprint isolation and mRNA fragmentation. We generally followed a published protocol[11] with the modifications stated below. Cells were resuspended in 100 µl of lysis buffer (20 mM Tris-HCl pH 8.0, 140 mM KCl, 5 mM Mg_2Cl, 1 % Triton X-100) with 1 g of chilled glass beads (Biospec) and lysed using a Fastprep 24 bead-beater at level 6 for 13 seconds. The extract was diluted with 400 µl of lysis buffer and cleared by centrifugation in two steps at 4°C at 16,000 g (5 minutes followed by 15 minutes). For footprint isolation, 600 A_{260} units of wild type vegetative cell extract were digested

with 750 Units of RNase I (Life Technologies) for 30 minutes, or 800 units of *pat1* diploid extract were treated for 10 min with 1,500 units of RNase I. Reactions were quenched with 600 units of SUPERaseIn (Life Technologies). Digested extracts in 500 µl were loaded onto an 14 ml linear 10-50% (w/v) sucrose gradient prepared with a Gradient Master (Biocomp), and separated by centrifugation for 160 min at 35,000 rpm in a SW 40Ti rotor (Beckman). The gradients were then fractionated by upward displacement with 55% (w/v) sucrose, and fractions containing monosomes selected for further processing. RNAs were then purified by phenol extraction, passed through a YM-100 column (Millipore), and run on 15% TBE-urea gels (Life Technologies). Fragments of around 28 nucleotides were extracted from the gel. For the preparation of mRNA fragments, total RNA was prepared by phenol extraction, and polyadenylated RNA was purified from 150 µg of total RNA (vegetative cells), or 300-450 µg (*pat1* cells) using oligo-dT_{25} magnetic beads (Life Technologies) following the manufacturer's instructions. Purified mRNA was fragmented by mixing 20 µl of mRNA with 20 µl of 2X alkaline fragmentation buffer (2 mM EDTA, 100 mM Na_2CO_3) followed by incubation for 15 minutes at 95°C. Samples were run on 15% TBE-urea gels (Life Technologies), and fragments of around 28 nucleotides were extracted from the gel. From this point, both mRNA and ribosomal footprint samples were processed identically. RNA samples were purified using Purelink RNA microcolumns (Life Technologies) as described by the manufacturer, except that the samples were initially passed through the column in the presence of 70% ethanol (to favour binding of small RNAs). Samples were then treated with poly nucleotide kinase (PNK, Fermentas) as described[11] and polyadenylated using 12 units of poly-(A) polymerase (NEB) at 37°C for 45 minutes. Reverse transcription reactions were performed using custom primers containing an anchored oligo(dT), 4 nucleotides of known sequence used for multiplexing, and 5 random nucleotides that serve as unique molecular identifiers (see below)[21]. All primers were synthesised by Integrated DNA Technologies (IDT) and are listed in Table S10. Reverse transcription products were gel-purified and circularised using CircLigase II (Epicentre), and amplified by PCR with custom library primers (P3 and P5)[21] for 12 or 15 cycles. Libraries were sequenced on an Illumina Genome Analyzer II (a subset of

the first biological repeat of vegetative haploids), on a Next Seq 500 sequencer (total RNA), or on a HiSeq 2000 platform (all other samples) using standard Illumina primers. We performed two biological replicates for vegetatively growing cells and for key time points of the *pat1* time course (3, 5 and 7 hours). Note that these are completely independent repeats, with cultures grown on different days, and library preparation and sequencing performed separately. The results were highly reproducible (Figure S3), with average correlation coefficients between replicates of 0.97 (mRNA), 0.98 (RPFs) and 0.90 (TEs). To evaluate the effect of mRNA purification on the protocol we sequenced total RNA from vegetative cells and compared the results with those using oligo(dT) purified RNA. The correlation between both samples (read densities across ORFs) was 0.88, indicating that oligo(dT)-purified RNA provides a reasonable estimate of mRNA amounts.

Bioinformatic analyses: All data processing was performed with custom scripts written in Perl (www.perl.org) and all downstream statistical analysis used R (www.r-project.org/). For all analyses, *S. pombe* annotations and sequences available from GeneDB (http://old.genedb.org/), now PomBase (http://www.pombase.org/), on May 9, 2011 were used[30].

The RT primers (Table S10) include a 4 nucleotide barcode that allows multiplexing, which was used to allocate reads from different samples to separate files. The RT primers also contain a 5 nucleotide random sequence that serves as a Unique Molecular Identifier (UMI). Reads that contain the same UMI followed by an identical sequence are very likely to have arisen from the same RNA molecule, and only one of them is retained. This step creates a non-redundant dataset, thus avoiding sampling biases and PCR amplification artefacts[20,21]. Non-redundant reads are then processed to remove adenosine residues at their 3' ends. Reads are mapped to the *S. pombe* rDNA genome[31] with Tophat 2[41] and the following parameters: --min-intron-length 29 --max-intron-length 819 --zpacker 0 --splice-mismatches 0 --max-multihits 1. Unmapped reads were recovered and aligned to the full *S. pombe* genome[31] with Tophat 2 and the following settings: --min-intron-length 29 --max-intron-length 819 --zpacker 0 --splice-mismatches 0. A GFF file containing annotation of the *S. pombe*

genome[30] was provided as a source of exon-exon junction data for Tophat. Aligned data were visualised using the Integrated Genome Viewer[42]. RPF libraries are expected to contain a higher fraction of contaminating rRNA than mRNA libraries, as the latter are oligo(dT)-selected. Consistently, 81.7 % of reads from RPF-derived and 26.5 % from mRNA-derived libraries mapped to the rDNA genome (median from all experiments, Table S1). UMIs were used to calculate the fraction of unique reads for those reads mapping to rRNA and for the remaining set. rRNA reads in the RPF samples typically showed low complexity (median unique 10.1 %, Table S1). This may be caused by the fact that the majority of these reads originate from a small pool of sequences (presumably due to the sequence preferences of RNase I). As the UMI length is 5 nucleotides, there are 1,024 different UMIs. If the number of fragments derived from the same sequence largely exceeds this figure, the likelihood that independent fragments with the same sequence share the same UMI increases and artefactually decreases the observed number of unique sequences. This is unlikely to be a problem for rRNA reads in the poly(A)-purified sample (median unique 73%, Table S1) or for non-rRNA reads (median unique 83.4%, Table S1), which are evenly distributed along the RNAs.

For clustering of RPFs (Figure S5), the 1,719 genes showing strongest changes in gene expression were selected. All the data were normalized to expression levels in vegetative cells of the corresponding time course, and RPF values were used for clustering. For clustering of TEs (Figure 1), the 418 genes that showed the largest variations in TE across the time course were chosen. Clustering was performed with Cluster 3.0[43,44], filtering out data with more than 20% values missing, log-transforming the data, using Pearson correlation and creating and average-linked tree. Clusters were visualized with Treeview[45].

To calculate overall periodicity and to calibrate reads, we counted the number of reads in which position 13 of a read maps to the first, second, or third position of every codon for all annotated coding sequences of the *S. pombe* genome (Figure S16A). Note that the choice of nucleotide 13 is arbitrary, and that similar results would be obtained with a different nucleotide.

The advantage of nucleotide 13 is that for a ribosome starting translation at the initiation codon, it corresponds to the position of the AUG on the mRNA. This was done for mRNA and RPFs, and analysed separately for reads of lengths between 25 and 32 nucleotides (Figure S16B). No bias was observed in any position within a codon for the mRNA fragments. By contrast, enrichments in position 1 were observed for RPFs of 28, 29 and 30 nucleotides (Figure S16B). The fact that fragments longer than 28 nucleotides show periodicity is unexpected, and suggests that some feature of the ribosome allows more precise cutting of RNase I at the 5' than at the 3' of the protected fragments. Reads between 28 and 30 nucleotides were selected for the discovery of translated features (such as uORFs and ncRNAs), while all reads were used to calculate translational efficiencies.

For the identification of novel translated regions, annotated coding sequences, non-coding RNAs, uORFs and intergenic regions were analysed separately. Open reading frames were defined as follows. First, selected regions were scanned in each frame until an AUG was encountered. This was defined as the start of the ORF, which was elongated until a stop codon was found. Second, the process was repeated as above for UUG, except than when an ORF overlapped with a previously defined one in the same frame it was discarded. The process was then performed as above for CUG and GUG. A total of 384,032 ORFs were defined in intergenic regions, 50,574 ORFs in ncRNAs, and 31,996 in 5' leader sequences. Of these, 49.2%, 49.2% and 27.2% started with an AUG codon in intergenic regions, ncRNAs, and 5' leaders, respectively (Table S6). This indicates that AUG ORFs are depleted in 5' leader sequences compared to other parts of the genome. The resulting ORFs were analysed for read periodicity by quantifying the fraction of codons showing enrichment of reads in the first nucleotide. For every codon, the enriched nucleotide (if existing) was defined as that having at least 60% of all the reads that mapped to the codon. In this way, all codons have equal contributions regardless of the total number of reads that map to each of them, thus avoiding biases created by small number of codons with very high number of reads. The fraction of codons enriched in nucleotide 1 within an ORF was defined as the ORF periodicity score. An ORF was defined as

translated when its periodicity score was >= 0.6. To avoid noise from lowly expressed genes, a total number of 10 reads was required for a region to be considered as translated. False Discovery Rates (FDRs) were calculated by randomising the position of the reads within each codon as follows: if a codon contains a_1, a_2, a_3 reads in positions 1, 2 and 3, respectively, the numbers a_1, a_2, and a_3 were randomly assigned to a position within a codon (this creates 3!=6 possibilities per codon). This was followed by the analysis of periodicity as described above. A p-value was also calculated for each feature by assuming a binomial distribution for the fraction of codons enriched at position 1. Note that under this very conservative assumption the smallest possible p-value for an ORF of a single codon is 0.33. However, given that our threshold for calling a feature as translated requires at least 10 reads and that more than 60% of the reads map to the first nucleotide, the probability of passing the threshold for such an ORF is lower than 0.0035 (based on the less conservative assumption that reads are randomly distributed within the codon).

Data deposition: All sequencing raw data have been deposited in ArrayExpress with the following accession numbers: E-MTAB-2176 (vegetative haploid cells), E-MTAB-2179 (*pat1* meiosis, replicate 1),E-MTAB-2265 (*pat1* meiosis, replicate 2) and E-MTAB-2470 (total RNA, vegetative cells).

For reviewers' access to the data please go to https://www.ebi.ac.uk/arrayexpress/ , login using the details below and then search for the experiment number.

Username: Reviewer_E-MTAB-2176, password: 7UxikuNp

Username: Reviewer_E-MTAB-2179, password: nwgswEtn

Username: Reviewer_E-MTAB-2265, password: geeehe9x

Username: Reviewer_E-MTAB-2470, password: k1iqhfce

Acknowledgements

We thank Jürg Bähler and Richard Jackson for comments on the manuscript. This work was funded by a Biotechnology and Biological Sciences Research Council (BBSRC) research grant to Juan Mata (BB/J007153/1).

Author contributions

CD and JM conceived the study and designed the experiments. CD performed all the experiments and JM did the bioinformatics analyses. CD and JM wrote the manuscript. All authors have critically read and approved of the manuscript.

References

1. Yamamoto, M., Imai, I. & Watanabe, Y. *S. pombe* mating and sporulation. in *The Molecular and Cellular Biology of the Yeast Saccharomyces: Life Cycle and Cell Biology* (eds. Pringle, J.R., Broach, J.R. & Jones, E.W.) 1035-1106 (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1997).
2. Mata, J., Lyne, R., Burns, G. & Bähler, J. The transcriptional program of meiosis and sporulation in fission yeast. *Nat Genet* **32**, 143-147 (2002).
3. Horie, S. et al. The *Schizosaccharomyces pombe mei4+* gene encodes a meiosis-specific transcription factor containing a forkhead DNA-binding domain. *Mol Cell Biol* **18**, 2118-29 (1998).
4. Mata, J., Wilbrey, A. & Bähler, J. Transcriptional regulatory network for sexual differentiation in fission yeast. *Genome Biol* **8**, R217 (2007).
5. Sugimoto, A., Iino, Y., Maeda, T., Watanabe, Y. & Yamamoto, M. *Schizosaccharomyces pombe ste11+* encodes a transcription factor with an HMG motif that is a critical regulator of sexual development. *Genes Dev* **5**, 1990-9 (1991).
6. Amorim, M.J., Cotobal, C., Duncan, C. & Mata, J. Global coordination of transcriptional control and mRNA decay during cellular differentiation. *Mol Syst Biol* **6**, 380 (2010).
7. Harigaya, Y. et al. Selective elimination of messenger RNA prevents an incidence of untimely meiosis. *Nature* **442**, 45-50 (2006).
8. Bitton, D.A. et al. Programmed fluctuations in sense/antisense transcript ratios drive sexual differentiation in *S. pombe*. *Mol Syst Biol* **7**, 559 (2011).
9. Rhind, N. et al. Comparative Functional Genomics of the Fission Yeasts. *Science* (2011).
10. Wilhelm, B.T. et al. Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature* **453**, 1239-43 (2008).
11. Ingolia, N.T., Ghaemmaghami, S., Newman, J.R. & Weissman, J.S. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324**, 218-23 (2009).
12. Michel, A.M. & Baranov, P.V. Ribosome profiling: a Hi-Def monitor for protein synthesis at the genome-wide scale. *Wiley Interdiscip Rev RNA* **4**, 473-90 (2013).
13. Brar, G.A. et al. High-resolution view of the yeast meiotic program revealed by ribosome profiling. *Science* **335**, 552-7 (2012).
14. Ingolia, N.T., Lareau, L.F. & Weissman, J.S. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* **147**, 789-802 (2011).
15. Chew, G.L. et al. Ribosome profiling reveals resemblance between long non-coding RNAs and 5' leaders of coding RNAs. *Development* **140**, 2828-34 (2013).
16. Guttman, M., Russell, P., Ingolia, N.T., Weissman, J.S. & Lander, E.S. Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. *Cell* **154**, 240-51 (2013).
17. Michel, A.M. et al. Observation of dually decoded regions of the human genome using ribosome profiling data. *Genome Res* **22**, 2219-29 (2012).
18. Iino, Y. & Yamamoto, M. Mutants of *Schizosaccharomyces pombe* which sporulate in the haploid state. *Mol Gen Genet* **198**, 416-421 (1985).
19. Nurse, P. Mutants of the fission yeast *Schizosacharomyces pombe* which alter the shift between cell proliferation and sporulation. *Mol Gen Genet* **198**, 497 (1985).
20. Konig, J. et al. iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat Struct Mol Biol* **17**, 909-15 (2010).
21. Mata, J. Genome-wide mapping of polyadenylation sites in fission yeast reveals widespread alternative polyadenylation. *RNA Biol* **10**, 1407-14 (2013).
22. Marguerat, S. et al. Quantitative analysis of fission yeast transcriptomes and proteomes in proliferating and quiescent cells. *Cell* **151**, 671-83 (2012).

23. Lackner, D.H. et al. A network of multiple regulatory layers shapes gene expression in fission yeast. *Mol Cell* **26**, 145-55 (2007).
24. Matia-Gonzalez, A.M., Hasan, A., Moe, G.H., Mata, J. & Rodriguez-Gabriel, M.A. Functional characterization of Upf1 targets in *Schizosaccharomyces pombe*. *RNA Biol* **10**, 1057-65 (2013).
25. Isken, O. et al. Upf1 phosphorylation triggers translational repression during nonsense-mediated mRNA decay. *Cell* **133**, 314-27 (2008).
26. Zhang, Z. et al. Nonsense-mediated decay targets have multiple sequence-related features that can inhibit translation. *Mol Syst Biol* **6**, 442 (2010).
27. Lackner, D.H., Schmidt, M.W., Wu, S., Wolf, D.A. & Bahler, J. Regulation of transcriptome, translation, and proteome in response to environmental stress in fission yeast. *Genome Biol* **13**, R25 (2012).
28. MacKay, V.L. et al. Gene expression analyzed by high-resolution state array analysis and quantitative proteomics: response of yeast to mating pheromone. *Mol Cell Proteomics* **3**, 478-89 (2004).
29. Preiss, T., Baron-Benhamou, J., Ansorge, W. & Hentze, M.W. Homodirectional changes in transcriptome composition and mRNA translation induced by rapamycin and heat shock. *Nat Struct Biol* **10**, 1039-47 (2003).
30. Wood, V. et al. PomBase: a comprehensive online resource for fission yeast. *Nucleic Acids Res* **40**, D695-9 (2012).
31. Wood, V. et al. The genome sequence of *Schizosaccharomyces pombe*. *Nature* **415**, 871-880 (2002).
32. Ma, J., Dobry, C.J., Krysan, D.J. & Kumar, A. Unconventional genomic architecture in the budding yeast *Saccharomyces cerevisiae* masks the nested antisense gene *NAG1*. *Eukaryot Cell* **7**, 1289-98 (2008).
33. Galindo, M.I., Pueyo, J.I., Fouix, S., Bishop, S.A. & Couso, J.P. Peptides encoded by short ORFs control development and define a new eukaryotic gene family. *PLoS Biol* **5**, e106 (2007).
34. Kondo, T. et al. Small peptide regulators of actin-based cell morphogenesis encoded by a polycistronic mRNA. *Nat Cell Biol* **9**, 660-5 (2007).
35. Magny, E.G. et al. Conserved regulation of cardiac calcium uptake by peptides encoded in small open reading frames. *Science* **341**, 1116-20 (2013).
36. Kelly, M., Burke, J., Smith, M., Klar, A. & Beach, D. Four mating-type genes control sexual differentiation in the fission yeast. *Embo J* **7**, 1537-47 (1988).
37. Slavoff, S.A. et al. Peptidomic discovery of short open reading frame-encoded peptides in human cells. *Nat Chem Biol* **9**, 59-64 (2013).
38. Hood, H.M., Neafsey, D.E., Galagan, J. & Sachs, M.S. Evolutionary roles of upstream open reading frames in mediating gene regulation in fungi. *Annu Rev Microbiol* **63**, 385-409 (2009).
39. Jackson, R.J., Hellen, C.U. & Pestova, T.V. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat Rev Mol Cell Biol* **11**, 113-27 (2010).
40. Shimoda, C. et al. Cloning and analysis of transcription of the *mei2* gene responsible for initiation of meiosis in the fission yeast *Schizosaccharomyces pombe*. *J Bacteriol* **169**, 93-6 (1987).
41. Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14**, R36 (2013).
42. Thorvaldsdottir, H., Robinson, J.T. & Mesirov, J.P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* **14**, 178-92 (2013).
43. de Hoon, M.J., Imoto, S., Nolan, J. & Miyano, S. Open source clustering software. *Bioinformatics* **20**, 1453-4 (2004).
44. Eisen, M.B., Spellman, P.T., Brown, P.O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* **95**, 14863-8 (1998).

45. Saldanha, A.J. Java Treeview--extensible visualization of microarray data. *Bioinformatics* **20**, 3246-8 (2004).

Figure legends

Figure 1. Quantification of translation using ribosome profiling.

(A) Comparison of RPF levels (this work) with protein levels (Marguerat el al. 2012). RPF levels are presented in reads per kilobase per million reads (RPKM), and protein levels are measured in copies per cell (cpc). The Pearson correlation is indicated inside the box. (B) Comparison of RNA levels to protein amounts. Labelling as in A. (C) Cluster analysis of the 318 genes that showed changes in TE of at least 10-fold across meiotic time points. Columns represent experimental points (V corresponds to vegetative cells and the numbers represent hours after meiotic induction), and rows represent genes. The colours show the TE (log_2 scale, see bar at the top). The position of selected clusters discussed in the text is shown on the left. (D) Changes in TE over meiosis for selected clusters. The plots on the left (green) represent median mRNA levels for the indicated clusters. The graphs on the right show the median TE as a function of time for genes in the indicated cluster (blue) or for all genes (red).

Figure 2. Periodic triplet behaviour of mRNA reads reveals novel genes.

(A) Scheme showing the expected position of RPFs with respect to the P and A sites of the ribosome (based on Ingolia et al. 2009). (B) Average number of reads against nucleotide position for all ORFs. RPFs are shown in black and mRNA fragments in red. The graph on the left corresponds to the region around the start codon and that on the right to codons 151 to 180. (C) Overlapping transcripts in the *SPAC3C7.15c* locus. The heat maps show mRNA (top) and RPF levels (bottom) across the meiotic time course, starting with vegetative cells at the top and progressing downwards. The brightness of the red colour is proportional to the number of reads. The scheme shows the position of the UTRs (arrows) and the ORFs (boxes). L corresponds to the annotated gene and S to the novel translated gene. RPF signal after the S ORF may represent leaky termination. (D) Triplet

periodicity for the L and S forms of the *SPAC3C7.15c* gene. The graphs show the fraction of codons in which a majority of the reads map to nucleotides 1, 2 or 3 for the S and L ORFs. In both cases the reading frame is calculated with respect to that of L. (E) Expression of the S form of *SPAC3C7.15c.* The translated S ORF was tagged with a TAP epitope and detected by Western blot (S) during a meiotic time course. A second Western blot was performed with the same samples and probed with antibodies against tubulin (tub). A single band of the predicted molecular mass was detected at 7 hours, corresponding to a strong RPF signal. (F) Nested antisense gene in the *SPAPYUG7.05* gene. As in C, but data are displayed for both the sense and antisense strands. LS corresponds to the annotated ORF and AS to the newly discovered antisense transcript. The arrows indicate the direction of transcription. Note the presence of a 5' extension (star on the heat map and dotted line on the scheme) that appears at the 5 hour time point and whose translation correlates inversely with that of the downstream ORF. (G) Triplet periodicity for the sense and antisense ORFs of *SPAPYUG7.05*, calculated and displayed as indicated for D.

Figure 3. Identification of translated regions in annotated ncRNAs.

(A) Histogram showing the length distribution of all translated uORFs (measured in nucleotides). (B) Number of translated ORFs per ncRNA. (C) Expression and translation of the *prl3* gene. The heat maps show mRNA (top) and RPF levels (bottom) across the meiotic time course, starting with vegetative cells at the top and progressing downwards. The brightness of the red colour is proportional to the number of reads. The arrow shows the position of the annotated ncRNA, and the box that of the translated ORF. (D) Triplet periodicity for the translated ORF in *prl3*. The graph shows the fraction of codons in which a majority of the reads map to nucleotides 1, 2 or 3 for the ORF. (E) Expression and translation of the *prl46* gene. Labelling as in C. (F) Triplet periodicity for the translated ORF in *prl46*. Labelling as in D. (G) Expression of the Prl3 protein. The translated ORF in *prl3* was tagged with a TAP epitope and detected by Western blot in vegetative cells (V). (H)

Expression of the Prl46 protein. The translated ORF in *prl46* was tagged with TAP and detected by Western blot. A single band of the expected molecular weight was observed at 5 hours into meiosis, coinciding with the peak of translation. The blot was also probed with antibodies against tubulin (tub).

Figure 4. Identification and characterization of uORFs.

(A) Number of translated uORFs per mRNA. (B) Histogram showing the length distribution of all translated uORFs (measured in nucleotides). (C) Examples of very short highly translated uORFs. The x axes show the DNA and protein sequences of a uORF consisting of a single codon (left) or two codons (right). The y axis represents the number of reads at each nucleotide. (D) Translation of the *SPAC11D3.13* locus. The top graphs shows the triplet periodicity across the gene calculated for a running window of 35 codons, and defined as the fraction of codons with a majority of reads in nucleotide 1, 2 or 3 within a codon. The heat maps show mRNA (top) and RPF levels (bottom) across the meiotic time course, starting with vegetative cells at the top and progressing downwards. The brightness of the red colour is proportional to the number of reads. The scheme shows the position of the UTRs (arrows) and the ORFs (boxes). L corresponds to the annotated gene, S1 and S2 to predicted uORFs. Note the change in reading frame as the window moves from one ORF to the next one. S2 overlaps with L and is more heavily translated, and is thus the dominant frame. (E) Histogram displaying the correlation coefficient between uORFs and their corresponding ORFs (see Methods for details). (F) Translation of the *SPBC1773.04* locus. Labelling as in D. L corresponds to the annotated gene. Note the presence of a longer form of the mRNA (star on the heat map and dotted line on the scheme) that appears at the 5 hour time point and whose translation (box S1) correlates inversely with that of the downstream ORF.