The Impact of Clinically Relevant Culture on Human Haematopoietic Stem Cells: A Kinetic Analysis at Single Cell Resolution

UNIVERSITY OF CAMBRIDGE

Carys Sarah Johnson

Homerton College

January 2022

This thesis is submitted for the degree of Doctor of Philosophy

DECLARATION

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the preface and specified in the text.

It is not substantially the same as any work that has already been submitted before for any degree or other qualification except as declared in the preface and specified in the text.

It does not exceed the prescribed word limit for the School of Clinical Medicine Degree Committee.

ABSTRACT

Haematopoietic stem cell (HSC) *ex vivo* gene therapy is now successfully used to treat an increasing number of monogenic disorders affecting the blood system. The ultimate efficacy of this therapy depends upon successful targeting of long-term haematopoietic stem cells (LT-HSCs), which are characterised by a predominantly quiescent status and sustained self-renewal capacity. Clinical *ex vivo* gene therapy protocols target the heterogenous mix of stem and progenitor cells encompassed in the $CD34^+$ fraction with a limited understanding of how *ex vivo* manipulation impacts a purified LT-HSC subset. Importantly, long-term repopulation capacity is lost over culture, although the kinetics and molecular drivers of this process remain unclear. To address these questions, I perform single cell RNA-Seq, *in vitro* functional assays and *in vivo* transplantation in a time resolved manner following cord blood (CB) culture in differentiation facilitating conditions and mobilized peripheral blood (mPB) culture during a lentiviral *ex vivo* gene therapy protocol.

First, I characterise the molecular impact of a 62 hr lentiviral *ex vivo* gene therapy protocol on mPB LT-HSC, short-term HSC (ST-HSC) and $CD34^+$ cells. I reveal that the *ex vivo* protocol dramatically rewires the quiescent LT-HSC transcriptome and demonstrate that the majority of transcriptional changes are attributed to the culture conditions rather than lentiviral transduction.

Secondly, I refine the kinetics associated with HSC functional attrition over the first complete cell cycle *ex vivo*. Long-term repopulation capacity is maintained for the first 6 hr irrespective of HSC source or tested culture conditions. I identify the 6 hr time-point as encompassing an adaptation period where LT-HSCs are rapidly responding to the instructive signals of culture and is molecularly underpinned by transient upregulation of cell stress response signalling. Following the 6 hr time-point, long-term repopulation capacity drops dramatically by 24 hr in CB culture and 62 hr in mPB culture. Loss of HSC function is correlated with reduced survival, cell cycle progression (late G_1-M), sharp upregulation of *MYC* and a reduced ability to resolve proteostatic stress. Taken together, my results demonstrate that the 6 to 24 hr transition point is instrumental for the determination of LT-HSC cell fate *ex vivo*.

Finally, using an *ex vivo* culture system of reversible early G_1 arrest, I formally test whether cell cycle progression drives loss of self-renewal capacity in culture. Long-term *in vivo* transplantation approaches conclusively establish that cell cycle progression is not responsible for the loss of long-term repopulation capacity associated with clinically relevant HSC culture.

ACKNOWLEDGEMENTS

I would firstly like to express sincere thanks to my primary supervisor Dr. Elisa Laurenti for giving me the opportunity to pursue a PhD within her lab. Elisa has been a great mentor in every sense of the word throughout my studies and it has been a pleasure to learn, think creatively and grow as a scientist within her group. I have learnt from Elisa far more than stem cell biology for which I will always be truly grateful. Elisa, thank you for your generosity in time and knowledge throughout my PhD.

I would also like to thank my secondary supervisor Prof. Bertie Göttgens for the continued support and always asking thought provoking questions at the end of my talks. In addition, I would like to acknowledge Xiaonan Wang and Winnie Lau formerly from the Göttgens group for the bioinformatic assistance and teaching me all the tips and tricks of scRNA-Seq library generation.

I have been very fortunate to have the support of another great mentor, Dr. Natalie Francis. I would like to thank Natalie and the Cell & Gene therapy team at GSK for offering a great insight into the world of T cells with particular mention to Steve Howe, Martijn Brugman and Christina Basford.

I would like to express sincere gratitude to Kenny Sham for his bioinformatic efforts in this project. In addition, I would like to extend my thanks to the entire Laurenti group with particular mention to Emily Calderbank and Nicole Mende for their help with the *in vivo* work and Serena Belluschi, a previous PhD student for her preliminary work on this project.

I am also very thankful to Kerstin Kaufmann and Gabriela Krivdova from John Dick's lab in Toronto for performing secondary transplants when our lab was in the process of moving animal facility.

The past year has been very tough at times and I have felt incredibly lucky to have the support of close family and friends. First and foremost, I would like to thank my wonderful mother Sarah Johnson, my father Rodney Johnson, my forever best friend and little sister Elera Johnson, my grandparents John & Jenny Johnson, Jordan So and last but certainly not least my housemate and fellow PhD student Elisabeth O'Flaherty for their unwavering support throughout my studies.

I would like to dedicate this thesis to my father, Rodney Johnson who in turn dedicated his thesis to me when I was a child. Your curious nature and commitment to the progression of good science has always shined through and shaped me immeasurably. You have selflessly provided me with every opportunity possible and I owe this success to you. Thank you for teaching me the rewards associated with hard work.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

ADA	Adenosine Deaminase
ADA-SCID	Adenosine Deaminase – Severe Combined Immunodeficiency
ALD	Adrenoleukodystrophy
AML	Acute Myeloid Leukaemia
ARSA	Arylsulfatase A
ASC	Activated Satellite Cell
BI	Bimodal Index
BM	Bone Marrow
°C	Celsius
CB	Cord Blood
CBSB	Cambridge Blood and Stem Cell Biobank
CDK	Cyclin-Dependent Kinases
CFU-C	Colony Forming Unit-Cell
CFU-S	Colony Forming Unit-Spleen
CKI	Cyclin-Dependent Kinases Inhibitor
CLP	Common Lymphoid Progenitors
CMP	Common Myeloid Progenitor
CP	Canonical Pathways
CRISPR	Clustered Regulatory Interspaced Short Palindromic Repeats
CRU	Competitive Repopulating Unit
DC	Dendritic Cell
DMEM	Dulbecco's Modified Eagle Medium
EHT	Endothelial to Haematopoietic Transition
ELDA	Extreme Limiting Dilution Analysis
EPCR	Endothelial C Protein Receptor
ER	Endoplasmic Reticulum
ERAD	Endoplasmic Reticulum Associated Degradation complex
ERT	Enzyme Replacement Therapy
EPO	Erythropoietin
Ery	Erythroid/Erythrocyte
ESC	Embryonic Stem Cells
FACS	Fluorescence Activated Cell Sorting
FBS	Foetal Bovine Serum
FDG	Force Directed Graph
Flt-3L	Fms like tyrosine kinase receptor 3 Ligand
G1 phase	Gap1 Phase
G2 phase	Gap2 Phase
G-CSF	Granulocyte colony stimulating factor

GOF	Gain Of Function
GMP	Granulocyte-Monocyte Progenitors
G.M.P	Good Manufacturing Practise
GPRC5C	G Protein Coupled Receptor
GSEA	Gene Set Enrichment Analysis
GSVA	Gene Set Variation Analysis
GvHD	Graft versus Host Disease
HIF-1α	Hypoxia Inducible Factor-1 Alpha
HDR	Homology Directed Repair
hr	Hour/hours
HSA	Human Serum Albumin
HSC	Haematopoietic Stem Cell
HSPC	Haematopoietic Stem and Progenitor Cells
HVG	Highly Variable Gene
IB	Injected Bone
IF	Intrafemoral
IV	Intravenous
IMDM	Iscove's Modified Dulbecco's Medium
Lin	Lineage
LOF	Loss Of Function
LMPP	Lymphoid-Primed Multipotent Progenitor
LSK	Lineage- Sca-1+ C-Kit+
LT-HSC	Long-Term Haematopoietic Stem Cell
LT-IC	Long-Term Initiating Culture
LTR	Long-terminal Repeat
Lym	Lymphoid
M phase	Mitotic Phase
Meg	Megakaryocyte
MEP	Megakaryocyte-Erythrocyte Progenitor
MLD	Metachromatic Leukodystrophy
MFI	Median Fluorescence Intensity
mPB	Mobilized Peripheral Blood
MPP	Multipotent Progenitor
My	Myeloid
My-bi	Myeloid-Biased
NHEJ	Non-Homologous End Joining
NK	Natural Killer
NSC	Neural Stem Cells
NSG	NOD scid gamma
NT	Non Transduced
pRB	Phosphorylated Retinoblastoma

PD	Palbociclib treated
PB	Peripheral Blood
PBS	Phosphate-Buffered Saline
PCA	Principal Component Analysis
Pen/Strep	Penicillin/Streptomycin
PD	Palbociclib (PD0332991)
PVA	Polyvinyl Alcohol
QC	Quality Control
RB	Retinoblastoma
RBC	Red Blood Cell
RT	Room Temperature
S	Synthesis Phase
Sca-1	Stem Cell Antigen-1
SCD	Sickle Cell Disease
SCF	Stem Cell Factor
SCID-X1	X linked Severe Combined Immunodeficiency
scRNA-Seq	Single Cell RNA-Sequencing
SCGM	Stem Cell Growth Media
SIN	Self Inactivating
ST-HSC	Short-Term Haematopoietic Stem Cell
TCA	Tricarboxylic acid Cycle
TDT	Transfusion Dependant β-thalassemia
TGF-β	Transforming Growth Factor B
TMRM	Tetramethyl Rhodamine Methyl Ester
TPO	Thrombopoietin
UNTR	Untreated
UMAP	Uniform Manifold Approximation and Projection
UPR	Unfolded Protein Response
WAS	Wiskott Aldrich Syndrome
X-CGD	X linked Chronic Granulomatous Disease

1. INTRODUCTION

A trillion blood cells are produced every day. This impressive output is ensured by a hierarchy of haematopoietic cells carefully organised to minimize exhaustion and facilitate unperturbed, lifelong blood production. At the top of this hierarchy exists a rare, predominantly quiescent population of haematopoietic stem cells (HSCs) capable of sustained self-renewal. Upon asymmetric division, HSCs produce a series of highly proliferative, committed progenitors with limited self-renewal capacity which replenish a diverse repertoire of functional blood cells. HSCs are arguably the best studied adult stem cell with decades of research underpinning our understanding of the blood system.

1.1 A brief history of Haematopoietic Stem Cell research

1.1.1 Early functional studies of the haematopoietic hierarchy and an early model of haematopoiesis

The first attempt to study cells with haematopoietic repopulation ability came from Jacobson et al. who made the observation that spleen cells could rescue lethally irradiated mice with only transient anaemia and leukopenia observed (Jacobson et al., 1951). Shortly after, Ford et al. were able to conclusively determine that the rescue of the blood system was mediated by transplanted cells through tracking donor spleen cells with cytogenetic abnormalities (Ford et al., 1956). This work set the stage for perhaps the earliest known stem cell assay – the Colony Forming Unit-Spleen (CFU-S) assay which showed serial repopulation capacity *in vitro* from *in vivo* transplanted spleen cells (Becker et al., 1963; Till and McCulloch, 1961). Conclusions from this work established the founding principle that colony generation is clonal; colonies originate from a population of rare multipotent cells.

The concept of heterogeneity within the HSC pool was developed when cells of different density gradients were found to repopulate recipients to different extents (Worton et al., 1969). It was then found that 5-FU treatment leads to a delayed, transient expansion in CFU-S numbers (Hodgson and Bradley, 1979). This observation formalized the concept that there exists a primitive cell in the bone marrow (BM) which precedes, and can give rise to cells responsible for repopulation in CFU-S assays (Hodgson and Bradley, 1979). Herein, the hierarchical structure of the HSC pool and downstream blood system was developed.

Progression of conceptual ideas was significantly aided by a landmark publication demonstrating that populations of serially transplantable HSCs can be isolated by cell surface marker expression through Fluorescence Activated Cell Sorting (FACS) (Spangrude et al., 1988). After this seminal development,

HSC research expanded to identify cell surface markers for the enrichment of HSC subsets. CD34 was identified to mark a primitive population of human haematopoietic cells *in vitro* (Andrews et al., 1990; Sutherland et al., 1989) in combination with negative selection of CD45RA (Lansdorp et al., 1990; Mayani et al., 1993) and CD38 (Bhatia et al., 1998; Hao et al., 1995) with positive selection of Thy1/CD90 (Baum et al., 1992; Craig et al., 1993). In the mouse system, HSCs were found to express high levels of the c-kit receptor (Orlic et al., 1993) combined with negative/low CD34 expression and high Sca-1 (Osawa et al., 1996; Zhao et al., 2000). These early cell surface markers identified by a range of groups were used to pioneer the characterisation of HSCs *in vitro* and *in vivo* functional assays.

Prospective FACS isolation of phenotypic populations by cell surface markers allowed for characterisation of HSC and progenitor populations leading to a model whereby differentiation occurred over successive, step-wise, lineage restriction events (Akashi et al., 2000; Kondo et al., 1997). In this model, HSCs are represented as a homogenous pool with equipotent differentiation and self-renewal capacity.

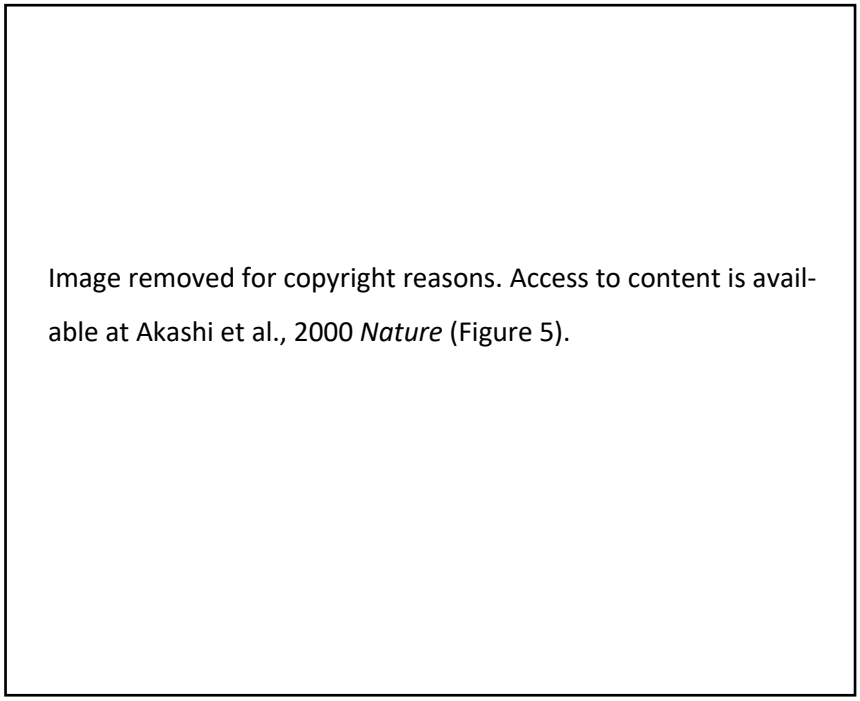

Figure 1.1: An early model of haematopoiesis (Taken from Akashi et al. 2000). LT-HSCs exist at the top of the haematopoietic hierarchy and directly give rise to ST-HSCs only. ST-HSCs give rise to CLPs which produce all mature lymphoid cells and CMPs which produce all mature myeloid (granulocyte, monocyte), megakaryocyte and erythroid cells. Transcription factors found to be expressed in populations by RT-qPCR approaches are highlighted in boxes underneath population and receptors distinguishing populations highlighted at bifurcation points.

1.1.2 Functionally defining the HSC

Following the development of this model and using a variety of experimental approaches, the field formed a functional definition for a bona-fide HSC which is still adhered to. HSCs are functionally defined by their ability for long-term reconstitution upon transplantation into immunocompromised mouse models. More specifically, long-term HSCs (LT-HSC) are exclusively defined by sustained repopulation activity over 16 weeks in a primary transplant and the ability to serially transplant into secondary recipients in both human (Notta et al., 2011) and mouse (Oguro et al., 2013) systems. Both short-term HSCs (ST-HSCs) and multipotent progenitor cells (MPPs) are functionally classified by transient multilineage engraftment in primary animals and thus exhibit limited self-renewal capacity.

1.2 Classical and current approaches to study HSCs

A substantial and growing number of approaches are used to improve our understanding of HSC biology. Over the past decade these approaches have shifted to a reliance on HSC characterization at the single cell level. Ultimately, self-renewal and lineage commitment decisions are made at the individual HSC level and therefore study should reflect this. As such, in the context of *in vitro* and transcriptomic assays, this section will focus on single cell approaches. This section (Introduction 1.2) aims to provide an introduction to the methodology commonly used for study in the HSC biology field whilst subsequent sections will better explore the biological conclusions gained from these studies where relevant to work presented in this thesis.

1.2.1 *In vitro* assays

Semi-solid and liquid colony forming assays have been widely used to study haematopoietic output *in vitro*. However, before the 1990's survival of primitive HSC populations *in vitro* was limited. A key advancement involved the incorporation of an adherent layer of mouse stromal cells (Hogge et al., 1996; Yoshikawa et al., 1999) which increased the efficacy of these assays by providing additional autocrine support. This permitted the discovery that the first lineage commitment event is not a strict myeloid/lymphoid bifurcation point in either mouse (Adolfsson et al., 2005) or human systems (Doulatov et al., 2010) as suggested by earlier studies (Fig.1.1) and in fact, the first lineage restriction event separates erythroid and megakaryocytic fates from lympho-myeloid differentiation (Adolfsson et al., 2005; Doulatov et al., 2010). Another advancement in human HSC culture came through the development of culture conditions which were able to support Myeloid-Erythroid-Megakaryocyte colony generation from single human $CD34^+$ cells (Notta et al., 2016) which included a range of growth factors combined with supportive stromal cell culture. This *in vitro* culture system revealed that the

structure of the hierarchy shifts throughout ontogeny with foetal haematopoiesis characterised by a bias in oligopotent progenitors in contrast to predominantly multipotent and unipotent progenitors found in adult bone marrow (BM) (Notta et al., 2016). Paired daughter cell assays are another single cell *in vitro* approach used to predominantly study mouse haematopoiesis and involve monitoring the first division *ex vivo* before splitting daughter cells for colony formation assays (Ema et al., 2000; Nakajima et al., 2010) or *in vivo* transplantation (Ito et al., 2016; Seita et al., 2007). This assay has provided valuable insights into *ex vivo* study of cytokines favourable for maintenance of stem cell capacity (Ema et al., 2000).

Long-term culture initiating cell assays (LTC-IC) from single cells has also been historically important to study sustained proliferation capacity (≥5 weeks) *in vitro* in the human (Petzer et al., 1996; Prosper et al., 1997) and mouse systems (Dexter et al., 1977; Lemieux et al., 1995). More recently these cultures have provided functional insights into proliferative heterogeneity within primitive mouse HSCs (Kent et al., 2009) and the umbilical cord blood (CB) LT-HSC fraction (Knapp et al., 2018). Importantly, these assays do not read out self-renewal capacity (Table 1.1) which can only be assessed using *in vivo* approaches.

1.2.2 *In vivo* models

Almost everything we know about HSC behaviour is attributed to the use of *in vivo* functional assays utilizing transplantation or transgenic approaches. Knockout technology has enabled the study of individual genes which contribute to HSC function with cell cycle regulators, transcription factors and downstream signalling pathways being extensively characterised through this approach (reviewed in Rossi et al., 2012). Furthermore, conditional knockouts have allowed for timed induction of gene perturbation, refining understanding from earlier more crude knockout approaches and permitting study where gene knockout leads to embryonic lethality, such as in the context of GATA-2 (Menendez-Gonzalez et al., 2019).

Clonal tracking can be achieved from single cell transplantation and is often coupled with the use of fluorescent reporters to effectively monitor progeny. Single cell transplantation of fluorescent reporter HSCs has identified megakaryocytic bias HSCs (Carrelha et al., 2018; Sanjuan-Pla et al., 2013). In addition, native haematopoiesis has recently been studied through inducible transposon tagging systems (Rodriguez-Fraticelli et al., 2018; Sun et al., 2014). Recent advances combine transplantation with lineage tracing and scRNA-Seq for simultaneous molecular and functional analysis (Pei et al., 2020; Rodriguez-Fraticelli et al., 2020).

Elegant pulse-chase systems have used clonal tracking to understand the relationship between divisional history and HSC function (Bernitz et al., 2016; Foudi et al., 2009; Qiu et al., 2014; Wilson et al., 2008). These studies use mouse models with inducible histone H2B-green fluorescent protein (GFP) under the control of a tetracycline response element (TRE). This technique permits initial incorporation of histone H2B-GFP fusion proteins into nucleosomes which is controllable by tetracycline. Initial labelling of the HSC compartment by GFP expression is followed by a period where labelling is inhibited. As such, GFP expression is halved by each cell division during the chase period, allowing for accurate determination of divisional history. This controllable system has conclusively demonstrated that dormant HSCs which divide very infrequently contain the majority of the repopulation capacity encompassed within the HSC pool, thus intimately linking divisional history to HSC function (further explored in Introduction 1.5.2).

Recently, unprecedented access into human haematopoiesis has been achieved through exploiting unique vector integration sites from gene therapy treated patients and subsequent clonal tracking (Aiuti et al., 2013; Biasco et al., 2016; Scala et al., 2018; Wang et al., 2010). This approach has revealed that following gene therapy two "phases" of reconstitution are observed, the first mediated by ST-HSC/MPP populations and the second, later phase mediated by LT-HSCs to reach steady state haematopoiesis (Biasco et al., 2016). A recent clonal tracking study in gene therapy treated patients with Wiskott Aldrich Syndrome (WAS) also highlighted a subset of LT-HSCs with characteristic restrained or latent reconstitution kinetics (Scala et al., 2018).

For the study of human haematopoiesis, immunodeficient mouse strains have been developed to allow xenograft of human cells without immunological rejection, termed "humanized mice" (reviewed in Mian et al., 2021). Key advancements in the progress of humanized mice include the development of *scid* mouse strain ($Prkdc^{scid}$) which recapitulates severe combined immunodeficiency (SCID) through lack of mature lymphoid cell production (Bosma et al., 1983) and subsequently the generation of immunodeficient non-obese diabetic mice (NOD-scid) mice, which show higher levels of human engraftment (Hesselton et al., 1995). Another important milestone involved the incorporation of mutations within the IL-2 receptor common gamma gene (*IL2rγ*) within the NOD-SCID strain (Ito et al., 2002) which leads to further defects in lymphoid cell development (Ohbo et al., 1996). These mice, NOD-*Prkdc* ^{scid} *IL2rg* ^{Tm1Wjl} (NSG) are now considered the gold standard and can recapitulate the human immune system to a reasonable degree (Shultz et al., 2005). To note, there are significant limitations with humanized mouse models including a differentiation bias towards immature lymphoid cells

(Table 1.1). Injection of human erythropoietin (EPO) can improve erythrocyte development (Chen et al., 2009) and the generation of the NOD-SCID-SGM3 (NSG-SGM3) mouse strain which expresses human GM-CSF and IL3 has allowed for improved myeloid reconstitution (Nicolini et al., 2004). Recently, human M-CSF, GM-CSF, IL-3 and TPO expression from an endogenous mouse locus (MISTRG model) has driven improvements in the development of monocytes, macrophages and natural killer (NK) cells (Rongvaux et al., 2014).

The gold standard approach for testing self-renewal capacity is serial transplantation of donor cells at limiting cell concentrations in a limiting dilution assay (LDA). Donor HSCs are first isolated and divided into a minimum of three doses (dose spacing ideally > 3 fold between each respective group) before transplantation (Fig.1.2). Following long-term engraftment, the number of engrafted mice (determined by an arbitrary threshold) is considered against the total number of mice to calculate the HSC frequency of the transplanted population. Reconstitution of the blood system has been achieved by single cell transplantation of mouse HSCs (Osawa et al., 1996) and later of human HSCs (Notta et al., 2011). Single cell transplants of mouse HSCs particularly have also been powerful in advancing our understanding of differentiation biases (Carrelha et al., 2018; Dykstra et al., 2007; Muller-Sieburg et al., 2004; Rodriguez-Fraticelli et al., 2018; Sanjuan-Pla et al., 2013).

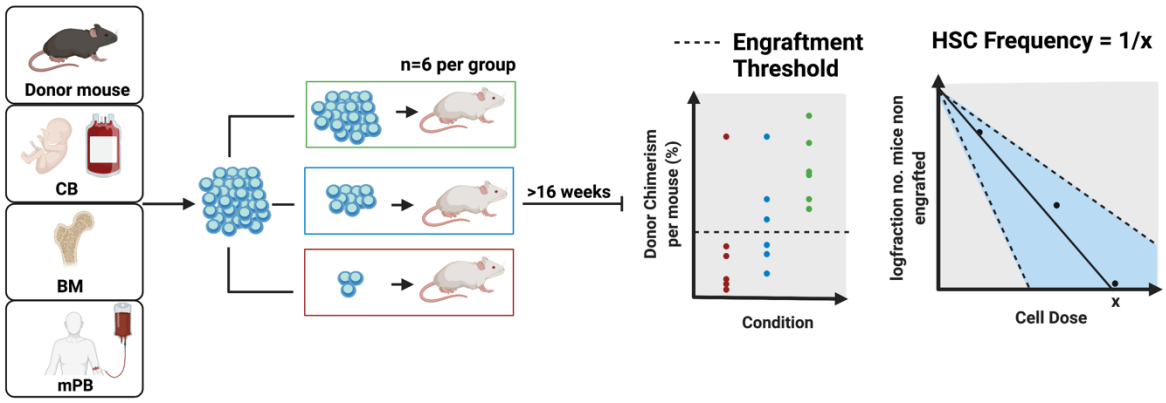

Figure 1.2: Limiting dilution assay schematic to assay HSC frequency in a population. Mouse HSCs can be isolated from a donor mouse or human HSCs can be isolated from a range of tissues including cord blood (CB), bone marrow (BM) and mobilized peripheral blood (mPB). Cells are divided into a minimum of three doses and injected into groups of mice (n=6 mice per group illustrated) for long-term (>16 week) transplantation: low dose (red), medium dose (blue), high dose (green). Engraftment is then assessed by the % donor chimerism (one value per mouse) set by an arbitrary threshold value set (dotted line). The HSC frequency of the transplanted population can then be calculated taking into account the number of tested mice and the number of mice showing successful human engraftment. Created with Biorender.

1.2.3 Single cell transcriptomics

Single cell quantification of gene expression was first performed by qPCR (Eberwine et al., 1992; Lambolez et al., 1992) although was greatly advanced by the development of microfluidics to separate cDNA into discrete reactions. Within haematopoiesis, this approach first revealed heterogenous expression of the myeloid determination factor PU.1 amongst HSCs (Warren et al., 2006) and subsequent scaling up of cell numbers and gene targets motivated progress in our understanding of transcription factor regulatory networks (Moignard et al., 2013) and identified novel progenitor populations (Psaila et al., 2016). A key advance came from the improvement of PCR methods which eventually allowed for untargeted single cell gene expression quantification (reviewed in Svensson et al., 2018). Following this, single cell cDNA amplification was refined and combined with high throughput DNA sequencing, culminating in the first protocol for single cell RNA-Seq (cRNA-Seq) (Tang et al., 2009).

Extensive characterization of large numbers of single blood cells ensued, offering new insights into transcriptional heterogeneity of progenitor populations and a more holistic view of the hierarchy (Paul et al., 2015). Importantly, the advent of single cell sequencing technologies has provided increased resolution where rare LT-HSCs would otherwise be obscured from multipotent progenitors. Nestorawa et al. published one of the earliest single cell landscapes of adult human haematopoiesis, showing for the first time that despite a reduced mRNA within the HSC compartment, a subset of genes showed higher expression in HSCs (Nestorowa et al., 2016).

A substantial body of scRNA-Seq work has now been performed across the haematopoietic hierarchy and taken together, conclusions support the gradual acquisition of lineage specification programs in a continuous manner until uni-lineage progenitor generation rather than a model of “stepwise” transitions between multi and bipotent cells (Macaulay et al., 2016; Notta et al., 2016; Velten et al., 2017). There now exists two main classes of scRNA-Seq platforms, the first consisting of high throughput fluidics-based methods (e.g. 10X genomics, DropSeq) which measure high ($>10^4$) cell numbers with 1000-3000 genes detected per cell and low throughput methods (e.g. Smart-Seq2, CellSeq2) which typically measure much lower cell numbers but detect >5000 genes per cell (reviewed in Watcham et al., 2019).

1.2.4 Limitations and advantages of approaches to study haematopoiesis

Importantly, all experimental approaches outlined above even at the single cell level have caveats and a range of approaches is preferred to draw conclusions. Table 1.1 details a list of key advantages and limitations for techniques used in the present study.

Technique	Advantages	Disadvantages
***In vivo* xenograft transplantation**	Supports long-term human engraftment	NSG mice show a differentiation bias towards immature B cells at expense of myeloid, erythroid and megakaryocyte output
	Potential to assay self-renewal capacity	Lack of cross reactivity of mouse to human cytokines
		Differential dependence on cytokines for differentiation (e.g. IL-7 for B cell development)
		Cell stress may bias/promote uni-lineage output
		Human haematopoiesis may not be fully recapitulated due to differences in life-span and metabolic rate
***In vitro* single cell assays**	Easily accessible for genetic manipulation	Lack of *in vivo* physiological support
	Controllable extrinsic environment	Supraphysiological levels of cytokines may bias differentiation
		HSCs cannot be maintained in quiescence
		Cannot assay self-renewal capacity
Single cell transcriptomics	Unbiased analysis of entire transcriptome	Cell destroyed for molecular analysis
	Inference of differentiation trajectories	Poor detection of low-level expression transcripts/ transcripts in a limited number of cells, "drop-out"
	Analysis of molecular heterogeneity within a population	Low RNA content in HSCs leads to high technical noise
		Aims to fit all cells into a trajectory therefore discrete state identification is minimized

Table 1.1: Advantages and limitations of current approaches in the HSC biology field which have been performed in the present study

1.3 An evolving view of the haematopoietic hierarchy

Our understanding of the haematopoietic hierarchy has progressed through utilizing a combination of techniques, notably aided in recent years by scRNA-Seq. The HSC pool which sits atop the haematopoietic hierarchy is now recognised as molecularly and functionally heterogeneous (Fig.1.3) (reviewed in Haas et al., 2018). Lineage restriction events can still be acquired at discrete decision points although transitionary positions of differentiation exist, surmounting in many potential cell states. This gradual progression allows flexibility to meet the high turnover demands of the blood system.

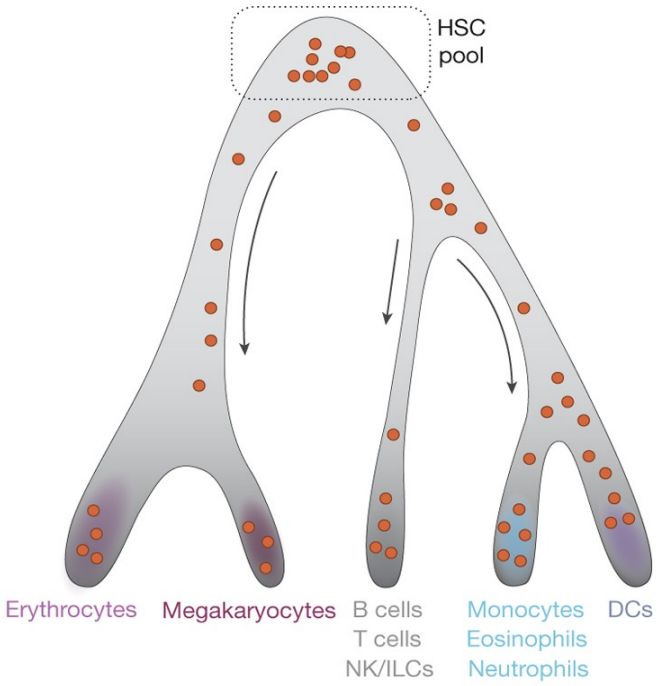

Figure 1.3: Continuous model of haematopoesis. (Taken from Laurenti and Gottgens, 2018). The HSC pool is depicted as molecularly and functionally heterogeneous and downstream lineage specific progenitors exist along a continuum of transcriptional states. The first bifurcation point separates the generation of megakaryocytes and erythroid cells from other lineages. Cells are represented as orange circles. Natural Killer (NK), innate lymphoid cells (ILCs), long-term (LT) short-term (ST) multipotent progenitor (MPP).

1.4 HSC Heterogeneity

Over five decades of research now shows that a simplified perspective of individual HSCs acting as an equipotent, self-renewing pool is perhaps convenient but biologically incorrect. In fact, the phenotypic HSC compartment exhibits a striking degree of molecular and functional heterogeneity. HSCs are heterogeneous in both distinguishing stem cell properties of differentiation capacity and self-renewal. Of note, the field has historically focussed on characterising heterogeneity within the mouse system,

partly due to the relative ease of study, although an increasing body of work is beginning to show significant heterogeneity within the human HSC pool.

1.4.1 Lineage biases

It is a defining property of HSCs to retain the potential for multi-lineage output, yet "balanced" HSCs are rare. Intrinsic myeloid and lymphoid biases have been extensively characterised within the mouse system and predominate the HSC pool (Dykstra et al., 2007; Kent et al., 2009; Morita et al., 2010; Müller-Sieburg et al., 2002; Muller-Sieburg et al., 2004). Myeloid bias HSCs and lymphoid biased HSCs can be isolated by cell surface expression of CD150 within the $CD34^{-}$ $cKit^{+}Sca\text{-}1^{+}Lin^{-}$ (LSK) population (Challen et al., 2010; Kent et al., 2009; Morita et al., 2010).

More recent work has been able to precisely define lineage biases within the branches of the haematopoietic tree. Megakaryocytic biased mouse HSCs have been identified which reside within a primitive subset of the HSC pool (Carrelha et al., 2018; Rodriguez-Fraticelli et al., 2018; Sanjuan-Pla et al., 2013; Shin et al., 2014). Single cell transplantation of platelet bias HSCs showed lineage restriction over impressive serial transplantation, however cells show multi-lineage capacity *in vitro* (Carrelha et al., 2018). Another approach using transposon tagging mediated clonal tracking estimated that up to half of the megakaryocyte lineage is produced by an independent pool of LT-HSCs (Rodriguez-Fraticelli et al., 2018). Platelet restricted LT-HSCs in this study were able to generate multilineage grafts upon subsequent transplantation *in vivo* (Rodriguez-Fraticelli et al., 2018). It is unclear whether a subset of platelet biased LT-HSCs are in fact platelet restricted *in vivo*, which presents challenges in defining cells identified as true HSCs or progenitors. Indeed, platelet biased HSCs and platelet restricted HSCs may both exist, fulfilling distinct roles under steady state and stress respectively. Taken together, findings underscore the degree of heterogeneity within the LT-HSC population. Collectively, this work is important in challenging our ever-changing view of the haematopoietic hierarchy.

1.4.2 Self-renewal capacity

The differentiation biases discussed in mouse HSCs are a distinct phenotype yet intimately linked to differences in self renewal capacity. Myeloid bias mouse HSCs show higher self-renewal capacity compared to lymphoid bias HSCs in single cell transplants (Dykstra et al., 2007; Morita et al., 2010; Muller-Sieburg et al., 2004). It is important to note that differentiation biases are intrinsic and not driven by stochastic processes as they are observed upon serial transplantation (reviewed in Laurenti and Gottgens, 2018).

The human HSC compartment is also heterogenous in self-renewal capacity. A seminal study by Notta et al. showed that a population of CB LT-HSCs can be enriched to a purity of 1 in 10.5 by high cell surface expression of the integrin CD49f combined with positive selection of Thy1/CD90 (Notta et al., 2011). In contrast, $CD90^-/CD49f^-$ counterparts are predominantly comprised of HSCs with transient engraftment kinetics and are denoted as ST-HSCs.

Studies of mouse and human haematopoiesis are caveated by the use of phenotypic cell surface markers to isolate HSCs for which there are now exist many different cell sorting strategies (Table 1.2). It is unclear whether further phenotypic restriction will eventually produce a "pure" population of HSCs or whether the unique combination of intrinsic and extrinsic regulatory factors influencing HSC behaviour will always produce heterogenous self-renewal and differentiation output at the single cell level.

Species	Name	Cell-surface phenotype	Self-renewal	Differentiation
Mouse	LT-HSC	$Lin^-Sca1^+cKit^+CD34^-CD150^+CD135^-CD48^-$ ± EPCR	High	
	IT-HSC	$Lin^-Sca1^+cKit^+CD34^{lo}CD135^-Rho^{lo}CD49b^{hi}$	Intermediate	
	ST-HSC/MPP1	$Lin^-Sca1^+cKit^+CD135^-CD150^-CD48^-$	Low	
	MPP2	$Lin^-Sca1^+cKit^+CD135^-CD150^+CD48^+$	Low	Lym-deficient
	MPP3	$Lin^-Sca1^+cKit^+CD135^-CD150^-CD48^+$	Low	Balanced
	MPP4	$Lin^-Sca1^+cKit^+CD135^+CD150^-CD48^+$	Low	Lym-biased
Human	LT-HSC	$Lin^-CD34^+CD38^-CD45RA^-CD49f^+CD90^+$	High	
	ST-HSC/MPP	$Lin^-CD34^+CD38^-CD45RA^-CD49f^-CD90^-$	Low	

Table 1.2: Commonly used cell surface markers for phenotypic enrichment of HSC subsets in mouse and human (Adapted from Laurenti & Gottgens 2018): Determined capacity for self-renewal and differentiation determined from *in vivo* studies. Long-term HSC (LT-HSC), intermediate term-HSC (IT-HSC), short-term HSC (ST-HSC), multipotent progenitor (MPP), Lymphoid (Lym).

Across both mouse and human studies, the HSC pool is heterogenous in cell cycle status. HSCs with superior self-renewal capacity reside in quiescence for extended periods of time, dividing very infrequently. The importance of quiescence in regulating the HSC state will be discussed in the following section (1.5).

1.5 The regulation and importance of quiescence

1.5.1 Introduction to the cell cycle

The mitotic cell cycle is split into two key phases, the Synthesis (S) phase where DNA is replicated, and the Mitotic (M) phase which generates two identical daughter cells. These are separated by two "Gap" phases and include the pre-replicative Gap 1 (G_1) phase and post-replicative Gap 2 (G_2) phase (Fig 1.4; left). There also exists a state outside of the cell cycle, termed quiescence/(G_0) defined as the reversible absence of cell cycling where cells can re-enter the cell cycle upon given stimulation. Initially, quiescence was thought to represent a passive state of inactivity but is now understood to be an actively maintained collection of states where cells are poised for cell cycle re-entry and activity (reviewed in Velthoven and Rando, 2019).

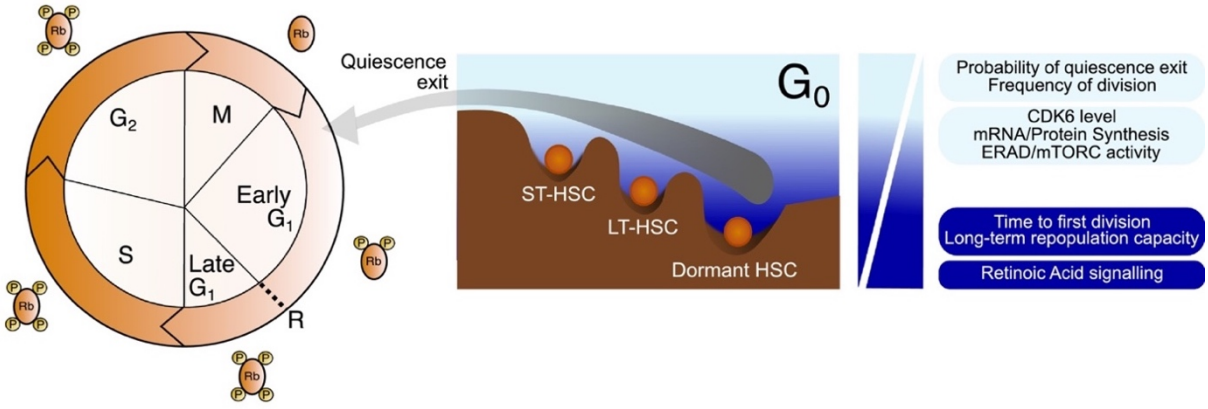

Figure 1.4: Heterogeneity in HSC quiescence status (Taken from Johnson, Belluschi & Laurenti 2020). The cell cycle is divided into four distinct phases Mitosis (M), Gap 1 (G_1), DNA synthesis phase (S), and Gap 2 (G_2). Quiescence (G0) is a continuum of stable molecular states characterised by unphosphorylated retinoblastoma protein (Rb) and low levels of Cyclin dependant Kinases (CDK's)/ Cyclins. Quiescence exit (G_0 to early G_1) is indicated by the arrow and is governed by CDK6 protein levels in HSCs. Rb is progressively phosphorylated throughout quiescence exit until hyperphosphorylation at the restriction point (denoted by R) whereby cells transition into late G_1 and commit to cell division. The strength of the signal required for quiescence exit is determined by the arrow shading and is inversely correlated with CDK6 protein level. ST-HSCs, LT-HSCs and dormant HSCs exist in respectively increasing depths of quiescence. Dormant HSCs divide very infrequently. Molecular and functional features of populations in distinct depths of quiescence indicated in text boxes.

1.5.2 The relationship between quiescence and self-renewal capacity

In many ways quiescent HSCs underpin the hierarchical structure of blood. The unique quality of sustained self-renewal is mediated by quiescence which protects HSCs from exhaustion. The consequence of the longevity of the LT-HSC population is the accumulation of proteostatic, oxidative and DNA damage which increases the likelihood that a deleterious mutation is fixed in the HSC pool.

Increased cell cycling itself is associated with higher DNA replication rates and increased bioenergetic demands. Therefore quiescent HSCs are protected from detrimental mutations and oxidative damage arising from cell cycle progression (reviewed in Wang and Dick, 2005). However, all HSC/MPPs reside predominantly in G_0 and divide much less frequently than progenitors (Foudi et al., 2009). Heterogeneity exists within quiescent populations and this is tightly linked to stem cell function (reviewed in Johnson et al., 2020).

Prolonged periods of quiescence, termed cell dormancy, are observed in a rare proportion of HSCs and this is associated with an increased depth of quiescence (Fig.1.4; right). HSC dormancy has been studied by inducible label retention assays in mice such as the H2BGFP reporter mouse allowing for careful analysis of divisional history *in vivo* (Bernitz et al., 2016; Foudi et al., 2009; Qiu et al., 2014; Wilson et al., 2008). Dormant HSCs cycle extremely infrequently, on average once every 120 days (Wilson et al., 2008) and the great majority of self-renewal capacity in the HSC pool is encompassed within these rare cells with a very low division history (Bernitz et al., 2016; Foudi et al., 2009; Qiu et al., 2014; Wilson et al., 2008). Some studies suggest only the dormant fraction engrafts in secondary transplantation (Qiu et al., 2014; Wilson et al., 2008).

Dormant mouse HSCs show low biosynthetic activity characterised by low transcription, low MYC target gene expression and high retinoic acid signalling (Cabezas-Wallscheid et al., 2017). The dormant fraction is only recruited to enter the cell cycle upon transplantation or stress (Cabezas-Wallscheid et al., 2017; Wilson et al., 2008). Treatment with G-CSF (Wilson et al., 2008), 5-FU (Wilson et al., 2008), lipopolysaccharide (LPS) (Takizawa et al., 2011) and polyinosinic:polycytidylic acid (pIC) (Cabezas-Wallscheid et al., 2017) causes dormant HSC activation and exit from quiescence. Upon cell cycle re-entry following dormancy, homeostatic HSCs become less likely to return to quiescence (Qiu et al., 2014), repopulation potential becomes restricted and self-renewal capacity declines proportionally to each successive division (Bernitz et al., 2016; Qiu et al., 2014). Importantly, injury activated HSCs can later return to dormancy (Wilson et al., 2008).

There are also distinct depths of quiescence within the human HSC pool as determined by heterogenous times of quiescence exit. ST-HSCs show a relatively faster time of quiescence exit *in vitro* compared to LT-HSCs, governed by higher CDK6 protein expression (Laurenti et al., 2015). LT-HSCs lack CDK6 protein expression, resulting in a delayed time to first division and increased time of quiescence exit (Laurenti et al., 2015). It can be speculated that the difference in division kinetics between LT-HSCs and ST-HSCs forms the basis for the distinct functional roles these subsets exhibit within blood

formation. ST-HSCs require a lower level of mitogenic stimulation for quiescence exit and therefore can act as a "first response" to stress or injury whilst limiting the number of divisions required from the LT-HSC population.

An increasing body of work is beginning to show that distinct depths of quiescence also exist within the human LT-HSC compartment. LT-HSCs defined by high expression of the lectin receptor CLEC9A and low CD34 expression ($CD34^{lo}CLEC9A^{hi}$) exhibit a slower time to first division *in vitro* with increased multilineage repopulation capacity *in vivo* (Belluschi et al., 2018). Furthermore, a recent paper by Kaufmann et al. has isolated a population of $CD112^{lo}$ LT-HSCs which show lower *CDK6* expression with transiently reduced, latent repopulation kinetics allowing for more durable long-term repopulation capacity (Kaufmann et al., 2021).

1.5.3 Extrinsic regulation of quiescence: The HSC niche

HSCs reside within a BM niche comprised of haematopoietic and stromal cells connected by an innervated network and surrounding vasculature (Fig.1.5) (reviewed in Pinho and Frenette, 2019). Within the BM space HSCs reside in a relatively hypoxic environment (Nombela-Arrieta et al., 2013; Shima et al., 2010; Takubo et al., 2010) supported through high HIF1α expression (Simsek et al., 2010; Takubo et al., 2010). Hypoxic status is tightly linked to quiescence (reviewed in Suda et al., 2011) and live cell imaging shows that HSCs continuously move away from vascular regions with an associated increase in the proportion of quiescent cells (Lo Celso et al., 2009).

Correct HSC niche maintenance is regulated through secretion of key cytokines, notably CXCL12 and SCF which bind to CXCR4 and KIT on HSCs respectively. A variety of differentiated haematopoietic cells regulate HSC maintenance in the niche, notably megakaryocytes (Bruns et al., 2014), macrophages (Hur et al., 2016), monocytes (Ludin et al., 2012) and regulatory T cells (Hirata et al., 2018). Non haematopoietic cells also play key roles in HSC niche regulation. Osteoblasts produce thrombopoietin (TPO) to regulate HSC quiescence (Qian et al., 2007) and mesenchymal stem cells (MSCs) express genes regulating HSC maintenance (Méndez-Ferrer et al., 2010). Endothelial cells provide a structural component to the niche, produce the majority of SCF (Naveiras et al., 2009) and also produce CXCL12 (Greenbaum et al., 2013) which regulate HSC activity. The relative permeability of endothelial cells also regulates the oxygen gradient in the niche (Itkin et al., 2016). Sympathetic nerves regulate the synthesis of CXCL12 and the normal release of HSCs from the bone marrow in tune with the circadian rhythm (Lucas et al., 2008; Méndez-Ferrer et al., 2008).

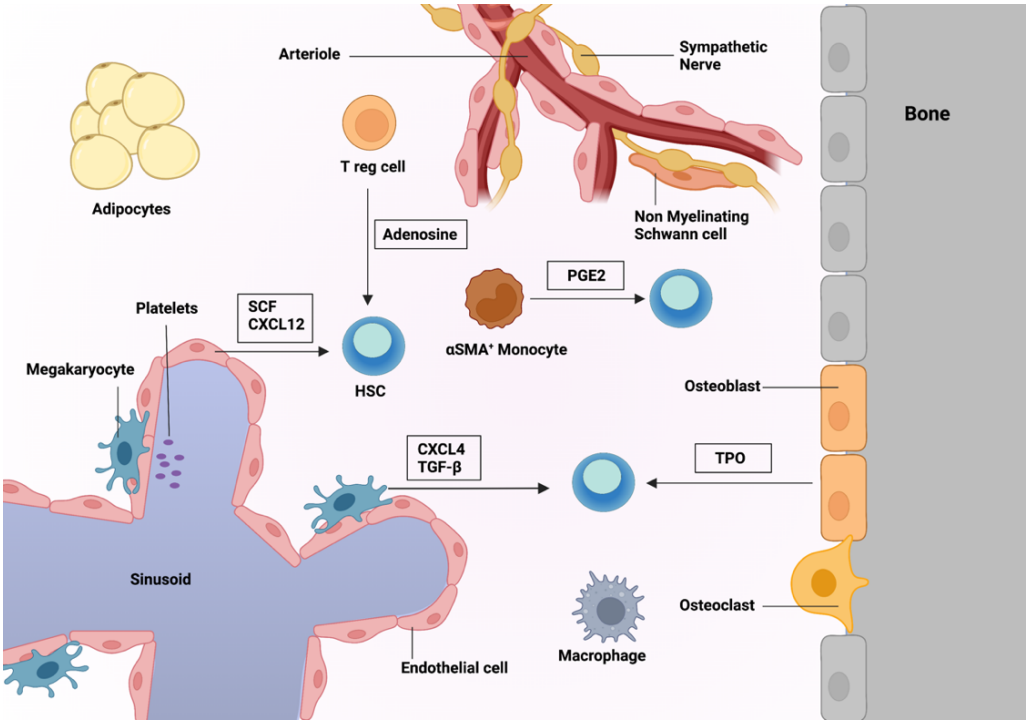

Figure 1.5: Overview of the adult HSC niche and interactions which maintain quiescence under steady state conditions (Adapted from Pinho & Frennete 2019). HSC subsets may reside in distinct niches and regional localization is shown with respect to arterioles and sinusoids. Haematopoeitic and non-haematopoietic cell types shown either directly or indirectly influence HSC behaviour. Secretion of factors which promote or maintain HSC quiescence is indicated. The sympathetic nervous system and associated nonmyelinating Schwann cells also regulate HSC quiescence. Vasculature ensures delivery of oxygen and other nutrients whilst simultaneously removing waste products. Created with Biorender.

Granulocyte colony stimulating factor (G-CSF) induces HSC mobilization by disruption of the CXCR4/CXCL12 signalling axis and the gradient of CXCL12 controls the degree of HSC egress (Shen et al., 2001). G-CSF mediated mobilization of HSCs is not a cell autonomous process and is regulated by the niche. G-CSF signalling in monocytes alone is even sufficient to trigger HSC mobilization (Christopher et al., 2011). Furthermore, perturbation of osteocytes (Asada et al., 2013) and the sympathetic nervous system (Katayama et al., 2006) leads to failed HSC egress upon G-CSF treatment. HSC egress from the BM by G-CSF can be clinically exploited for autologous or allogenic HSC transplantation approaches, the treatment of chemotherapy induced neutropenia and a range of other clinical uses (reviewed in Bendall and Bradstock, 2014)

1.5.4 Transcriptional regulation of the HSC quiescent state

Maintenance of HSC quiescence is also regulated by a multitude of intrinsic factors leading to a distinct epigenetic, transcriptional and post-translational state in quiescent cells. Intrinsic regulation of quiescence is mediated by an impressive array of transcription factors, miRNA's, Cyclins, CDK's and CDK inhibitors (reviewed in Hao, Chen & Cheng. 2016).

The three retinoblastoma (Rb) family members of transcriptional repressors: RB, p107 and p130 inhibit cell cycle progression by blocking E2F family mediated gene transcription. Rb family triple knock out leads to loss of quiescence, HSC pool expansion and impaired reconstitution potential (Viatour et al., 2008). CDK inhibitors block CDK-Cyclin complex activity to maintain HSCs in quiescence such as p21 (Cheng et al., 2000) and p57 (Matsumoto et al., 2011). In addition the INK4 family of CDK inhibitors such as $p18^{INK4c}$ (Yuan et al., 2004) and $p16^{INK4a}$ (Stepanova and Sorrentino, 2005) also maintain HSC quiescence. A number of transcription factors also directly maintain the HSC quiescent state. PU.1 is a master regulator of the cell cycle controlling transcription of *Cdk1*, *Ccnd1* and *E2f1* with loss leading to increased cell cycling and reduced self-renewal capacity (Staber et al., 2013). The AP-1 superfamily also have instrumental roles regulating HSC quiescence such as cFos (Okada et al., 1999), and JunB (Santaguida et al., 2009) with ablation leading to quiescence loss and decreased long-term repopulation capacity. In addition, Gfi-1 (Hock et al., 2004; Zeng et al., 2004), BMI1 (Ikonomi et al., 2020), SATB1 (Will et al., 2013), FOXO3a (Miyamoto et al., 2007) and GATA-2 (Menendez-Gonzalez et al., 2019) amongst many others directly facilitate the HSC quiescent state.

HSC quiescence is also maintained by layers of epigenetic regulation dictated by histone configurations which affect chromatin accessibility and resulting gene expression. MicroRNA's (miRNA's) can promote gene silencing such as miRNA-126 which represses PI3K/AKT signalling (Lechman et al., 2012) and miR-29 with represses *Dnmt3a* (Hu et al., 2015) to maintain HSC quiescence. Using single cell ATAC-Seq, quiescent LT-HSCs were recently found to have a distinct chromatin accessibility signature (Takayama et al., 2021). This state is governed by CTCF mediated silencing of differentiation programs with loss leading to chromatin interactions observed in ST-HSCs (Takayama et al., 2021).

1.5.5 Mechanistic regulation of quiescence exit and the restriction point

Upon receipt of extrinsic signals, HSCs can exit quiescence and re-enter the cell cycle. Cell cycle re-entry is dictated by a shift in the balance of inhibitory and activating regulatory complexes (reviewed in Nakamura-Ishizu et al., 2014). In quiescence, Rb is physically associated with E2F transcription

factors repressing E2F mediated transcription of cell cycle activators (Passegué et al., 2005). Rb is able to recruit chromatin remodelling complexes, histone and DNA methyltransferases and Polycomb repressors to inhibit transcription of cell cycle progression genes (Dahiya et al., 2001; Robertson et al., 2000; Zhang et al., 2000). Upon receipt of mitogenic signals, CDK4/6 complexed with Cyclin D is upregulated to initiate Rb phosphorylation (reviewed in Nakamura-Ishizu et al., 2014). Phosphorylation of Rb causes a conformational change disrupting the E2F interaction (Burke et al., 2010) with Rb allowing transcription of cell cycle progression genes including Cyclin E and CDK2. Rb is then hyperphosphorylated by CDK2/Cyclin E complexes which allows the cell to pass the "restriction point", permitting the transition into late G_1 (Yao et al., 2008).

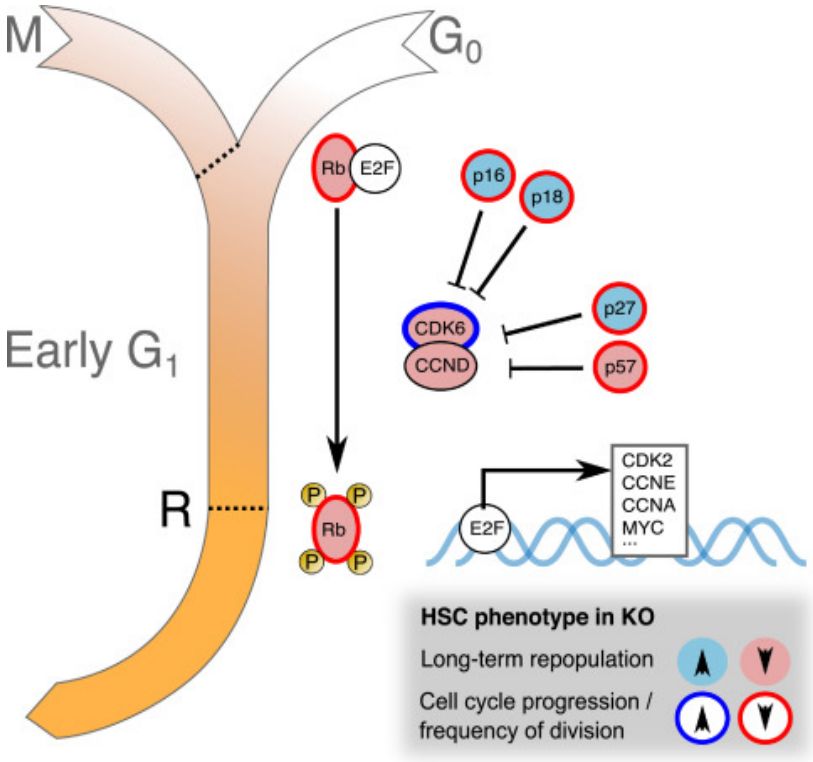

Figure 1.6: HSC Quiescence exit is governed by CDK/Cyclin complex dependant phosphorylation of Rb (taken from Johnson, Belluschi & Laurenti 2020). In quiescence (G_0), Retinoblastoma (Rb) remains unphosphorylated and physically associated to E2F transcription factors, blocking E2F mediated transcription of *CDK2*, *Cyclin E* (CCNE), *Cyclin A* (CCNA) and *MYC* amongst others. Upon receiving mitogenic signals Rb becomes phosphorylated by CDK6/Cyclin D and subsequently Cyclin E/CDK2. Cyclin/CDK complex activity is inhibited by the CDK inhibitors p16, p18, p27 and p57. E2F is released from Rb permitting associated transcription of cell cycle and nucleotide biosynthesis genes. This stage in G_1 is known as the restriction point (R) where cells enter late G_1 and lose dependency on extracellular mitogenic signals for cell cycle progression.

The restriction point was first described by Pardee and colleagues who found that mitogen withdrawal at different points of the cell cycle leads to a ubiquitous block in G_1 (Pardee, 1974). This seminal study revealed that after the restriction point a given cell becomes committed to undergo cell cycle progression independently of mitogenic signals. Recent work has shown that CDK4/6 activity is necessary throughout G_1 for cells to pass the restriction point whereas Cyclin E/CDK activity is in fact dispensable (Chung et al., 2019). To note, the restriction point model has recently been challenged and refined. Commitment to division is still thought to occur in G_1 although the exact point of commitment remains unclear and may involve two bistable switches separated by a period of reversibility (reviewed Pennycook and Barr, 2020).

1.5.7 Features of quiescent and activated HSCs

Quiescent HSCs share many characteristics with other quiescent adult stem cell populations including decreased cell size, reduced RNA content, and reduced metabolic requirements (Fig.1.7). More specifically, quiescent HSCs show high numbers of inactive mitochondria (de Almeida et al., 2017) which maintain a glycolytic metabolism with low ROS accumulation (Ito et al., 2006; Simsek et al., 2010; Vannini et al., 2016). In addition, quiescent HSCs and are reliant on autophagy for recycling of damaged organelles in absence of cell division (Ho et al., 2017; Warr et al., 2013). Translation and protein synthesis rates in quiescent HSCs are also lower with respect to activated populations (Signer et al., 2014, 2016).

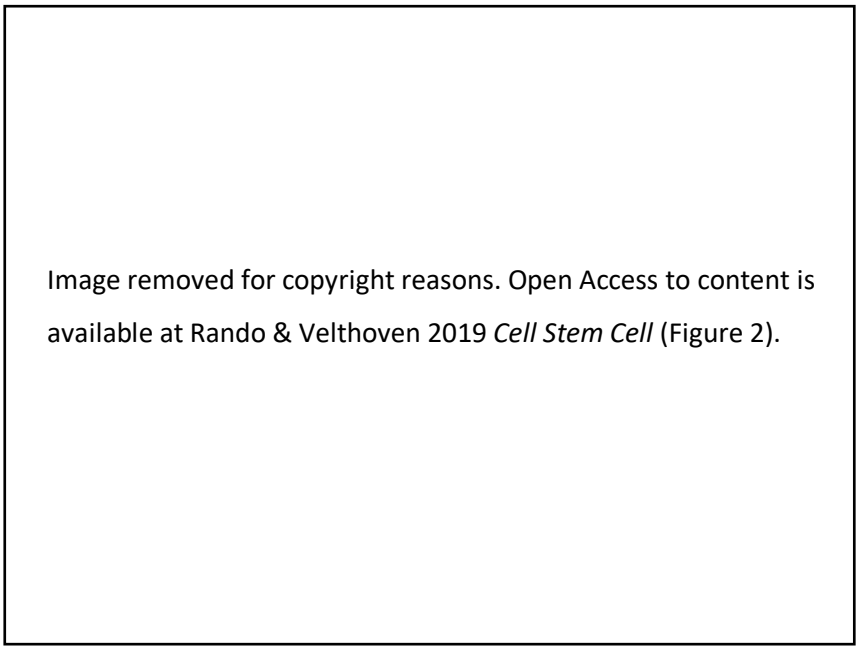

Figure 1.7: Characteristics of quiescent and activated adult stem cell populations (Taken from Rando & Velthoven 2019). Quiescent adult stem cells show shared features in epigenetic, transcriptional and metabolic regulation. Arrows indicate direction of change from quiescent populations compared to respective activated counterparts.

Despite mitigated accumulation of damage, quiescent HSCs show high basal expression of stress response pathways such as the unfolded protein response (Galen et al., 2014) and the integrated stress response pathway (Galen et al., 2018) facilitated through high ER associated degradation complex (ERAD) expression which acts to recognise and target misfolded proteins for ubiquitination (Liu et al., 2020b; Xu et al., 2020). This leads to high "quality control" thresholds in HSCs that promote preferential apoptosis over repair in response to damage induced by unfolded proteins, reactive oxygen species (ROS) generation or DNA damage (Galen et al., 2014, 2018; Milyavsky et al., 2010; Yahata et al., 2011). These features are underpinned by distinct DNA damage repair mechanisms. Quiescent HSCs preferentially undergo error prone non homologous end joining (NHEJ) upon genotoxic stress to promote apoptosis, in contrast to cycling cells which preferentially undergo homology directed repair (HDR) (Milyavsky et al., 2010; Mohrin et al., 2010). Together, this protects the HSC pool by inhibiting the generation of deleterious mutations and damaged HSC proliferation.

1.5.8 Definitions of quiescence and activation for this study

Quiescence exit itself (G_0 to early G_1) is poorly defined and understudied, partially attributed to a lack of experimental control over quiescence status and the fact that no markers currently exist to distinguish G_0 from G_1. For the purpose of this study, I define quiescence exit as the duration of time taken from loss of quiescence status to the hyperphosphorylation of Rb demarcating early and late G_1. Subsequently, late G_1 until cell division (M phase) is defined as cell cycle progression. Finally, "HSC activation" is defined as the entire cell cycle process following loss of quiescence status.

1.6 Clinically relevant HSC culture

1.6.1 Loss of self-renewal capacity *ex vivo*

Whilst HSCs can return to quiescence *in vivo* following activation, return to quiescence *ex vivo* has never been achieved. Prolonged HSC activation and cell cycling *ex vivo* is associated with loss of self-renewal capacity in culture which reduces the efficacy of clinical approaches where culture is unavoidable, such as in *ex vivo* gene therapy protocols. Loss of repopulation capacity was observed following stimulatory *ex vivo* cytokine conditions in early xenograft models of human cord blood (Gan et al., 1997) and mobilized peripheral Blood (mPB) (Gothot et al., 1998) amongst many other studies since.

The mechanisms which drive loss of self-renewal capacity *ex vivo* are poorly studied and much more is understood about loss of self-renewal in the context of *in vivo* models. HSC exit from quiescence *ex*

vivo is due to environmental cues from stimulatory cytokines and the aerobic environment of culture which in turn drive increased mitochondrial activation (Hinge et al., 2020; Vannini et al., 2016), oxidative phosphorylation and ROS production (Manesia et al., 2015). *Ex vivo* culture of LT-HSCs also rapidly increases protein synthesis rates (Kruta et al., 2021) and is associated with increased levels of MYC which directly inhibits lysosomal recycling (García-Prat et al., 2021). This changing external stimuli remodels the mitochondria to an activated state (Hinge et al., 2020), disrupts proteostasis programs (Kruta et al., 2021; Xie et al., 2019) and reduces the dependency on autophagic recycling (García-Prat et al., 2021). These changes are all associated with a loss of self-renewal capacity. Importantly, it remains unclear what causes HSC functional loss rather than what is simply observed alongside HSC functional loss.

1.6.2 *Ex vivo* expansion conditions

Over the past several years dramatic progress has been made in optimising conditions to maintain and even expand functional HSCs in culture. *Ex vivo* HSC expansion approaches which have proven the most successful recapitulate *in vivo* niche interactions. The BM niche is considered hypoxic and an interesting early technique recreating this hypoxic environment showed significant *ex vivo* expansion of mouse HSCs (Danet et al., 2003). The *in vivo* niche also provides significant structural support offering biophysical cues, however, HSCs have traditionally been cultured on hard hydrophobic plastics. To circumvent this, CB and BM LT-HSCs have recently been cultured on hydrophilic zwitterionic hydrogels and demonstrated convincing *ex vivo* expansion (Bai et al., 2019). Furthermore, *ex vivo* expansion of HSCs has been achieved with co-culture of engineered MSCs which overexpress key transcription factors involved in maintenance of stemness (Nakahara et al., 2019).

HSCs are also best maintained *ex vivo* when cultured with cytokines known to be secreted by surrounding niche cells such as stem cell factor (SCF), thrombopoietin (TPO) and FMS like tyrosine kinase 3 ligand (Flt-3L). *In vivo* these soluble factors would be continuously replenished, however *ex vivo* culture leads to an accumulation of inhibitory derived factors such as CCL3, CCL4 and TGFβ-2 promoting unwanted differentiation (Kirouac et al., 2010). Therefore, many studies now implement a "fed-batch" approach, supplementing cultures with fresh media to reduce inhibitory feedback systems (Bai et al., 2019; Csaszar et al., 2012; Xie et al., 2019). Foetal bovine serum (FBS)/ serum albumin is also often added to media which contains growth factors and buffers cells from rapid changes in pH (Ieyasu et al., 2017). Although, FBS shows inconsistency between batches which reduces the reproducibility of HSC culture (Ieyasu et al., 2017). Wilkinson et al. recently replaced serum with the synthetic polymer polyvinyl alcohol (PVA), introduced supportive fibronectin to the culture and optimised high TPO

and low SCF cytokine concentrations to develop a serum free culture system with impressive (236-899 fold) expansion in mouse HSCs over one month (Wilkinson et al., 2019).

A different approach involves the addition of small molecules to culture which maintain the self-renewal capacity of HSCs. Incorporation of the aryl hydrocarbon receptor antagonist StemRegenin1 (SR1) has shown a significantly increased proportion of engraftable mPB and CB $CD34^+$ cells following seven day culture (Boitano et al., 2010). In addition, significant progress was made when the pyrimidoindole derivative UM171 was demonstrated to expand human CB $CD34^+$ cells by ≈13 fold capable of long-term transplantation after a 12 day fed-batch culture protocol (Fares et al., 2014). UM171 is proposed to act by maintaining a more favourable epigenetic state for the transcription of stem cell maintenance genes *ex vivo* (Chagraoui et al., 2021). Inhibition of DEGS1, the final enzyme in sphingolipid biosynthesis by 4HPR has also recently been shown to maintain HSC self-renewal capacity *ex vivo* through coordinated activation of cell stress response programmes and restoring proteostasis (Xie et al., 2019).

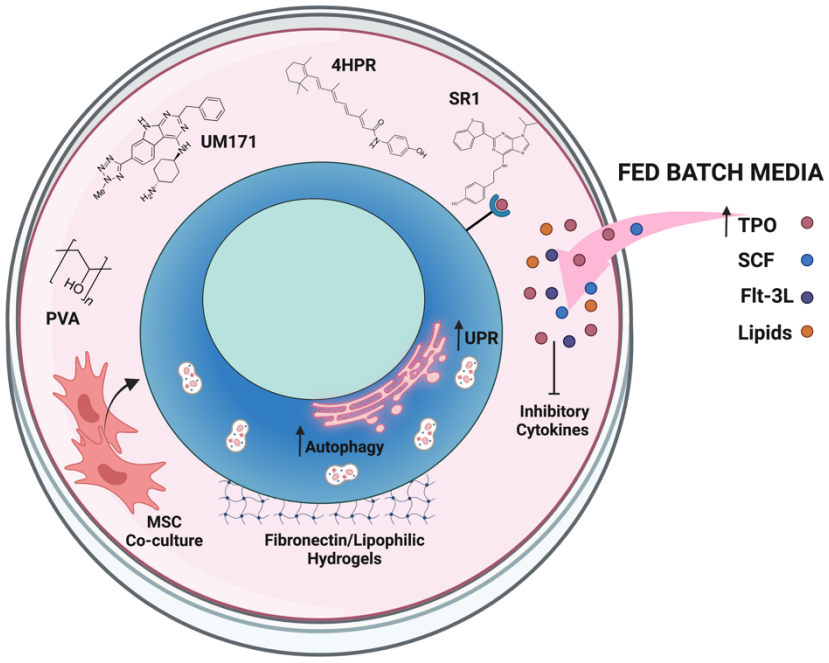

Figure 1.8: A self-renewing *ex vivo* HSC (adapted from Wilkinson et al. 2020). Strategies to promote HSC expansion *ex vivo* include the replacement of serum albumin with Polyvinyl alcohol (PVA), culture with transgenic mesenchymal stem cells (MSCs) and culture using fibronectin or lipophilic hydrogels. Small molecule inhibitors such as SR1, UM171 and 4HPR also expand the self-renewing LT-HSC pool *ex vivo*. A "fed-batch" approach where media including high concentrations of thrombopoietin (TPO), stem cell factor (SCF) in addition to FMS like tyrosine kinase 3 ligand (Flt-3L) and lipids is recommended to dilute inhibitory factors promoting differentiation. Created with Biorender.

1.6.3 Clinical applications of *ex vivo* expanded HSCs

Importantly, the longstanding goal of these studies is to translate findings for clinical use. In addition to resolving HSC functional attrition *ex vivo*, HSC expansion would allow for more effective use of low HSC numbers obtained from paedeatric gene therapy approaches and low starting cell numbers acquired from individual CB units (reviewed in Wilkinson et al., 2020). For allogenic transplantation approaches specifically, ex *vivo* expanded CB HSCs would also often be obtained in a shorter time-frame than donor samples and require a reduced degree of HLA matching due to the vast majority of T cells being naïve (reviewed in Giralt and Bishop, 2009). Expanded HSCs may even require a reduced conditioning regimen, as demonstrated through transplantation of PVA expanded mouse HSCs into non-conditioned recipients (Wilkinson et al., 2019).

A significant challenge remains in developing culture protocols which are scalable for clinical use and adhere to G.M.P standards. SR-1 expanded CB HSCs have now been successfully used to treat haematological malignancies in a phase I/II clinical trial (Wagner et al., 2016) and a recent clinical trial using UM171 expanded CB HSCs for the treatment of haematological malignancies published extremely promising results where no graft failure or rejection occurred (Cohen et al., 2020). It is likely that in the near future, small molecule and biophysical methods to expand HSCs *ex vivo* will develop a combinatorial approach for HSC expansion and this will become an established method of treatment for a range of diseases. Importantly, key questions remain about the mechanism of action for small molecule inhibitors and understanding the mechanisms which drive loss of self-renewal capacity *ex vivo* is crucial in progressing this work.

1.7 HSC *ex vivo* gene therapy for rare monogenic disorders

The concept of gene therapy has fascinated scientists for decades; genetic correction of disease propagating mutations in one durable treatment would allow a continued source of modified progeny, effectively curing the disease. The possibilities and challenges of gene therapy were first speculated by Friedmann and Roblin who concluded at the time of writing that gene therapy would be premature due to a basic lack of understanding on gene regulation and the mechanisms causing disease (Friedmann and Roblin, 1972). Since, our understanding of HSC molecular biology and disease etiology has significantly improved and HSC gene therapy is used to treat a variety of disorders (reviewed in Tucci et al., 2021).

HSCs are ideal targets for *ex vivo* gene therapy, not only due to their unique capacity for both self-renewal and multipotency but features such as the relative ease of accessibility, intrinsic ability to

home to the niche upon transplantation and the unfortunate wealth of monogenic disorders which affect the blood system. Most importantly, there exists an unmet clinical need delivered by HSC gene therapy. Clinical approaches for inherited haematological disease often involve allogenic BM transplantation from a HLA matched donor where availability is scarce (<25%) (reviewed in Aiuti et al., 2017). Imperfect donors dramatically increase the risk of post-transplant complications (Boelens et al., 2013; Walters, 2015) such as graft versus host disease (GvHD) which requires urgent restoration of haematopoiesis to avoid complications from pancytopenia (such as bleeding, infections and anemia) and can be fatal (reviewed in Morgan et al., 2017). Gene therapy directly circumvents this risk by using an autologous approach of cell harvest, correction *ex vivo* and re-infusion of cells back into the same patient (Fig.1.9).

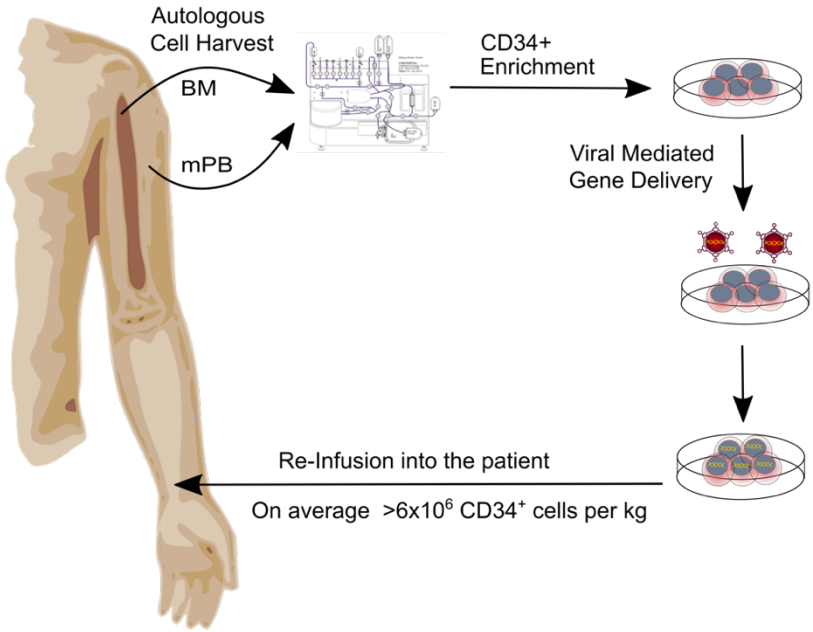

Figure 1.9: HSC *ex vivo* gene therapy workflow: The patient's cells are first harvested and $CD34^+$ enrichment performed. $CD34^+$ cells are cultured and exposed to a viral vector (usually through two rounds of transduction) containing the therapeutic transgene. Corrected $CD34^+$ cells are re-infused back into the patient at high cell doses on average >6 x 10^6 cells per kg of the patient's weight (Aiuti et al., 2009; Biffi et al., 2013). Created with Inkscape.

1.7.1 Promises and challenges in early gene therapy trials

The first disease investigated for HSC gene therapy was the rare, debilitating condition of Adenosine Deaminase - Severe Combined Immunodeficiency (ADA-SCID). Lack of adenosine deaminase (ADA) causes accumulation of toxic metabolic substrates in developing lymphocytes leading to a severely compromised immune response, skeletal dysplasia and cognitive defects (reviewed in Gaspar et al.,

2009). ADA-SCID was considered an ideal candidate for gene therapy for several reasons including the monogenic cause of the disorder and the ubiquitous expression of ADA (Ferrari et al., 1991). In addition, corrected lymphocytes would have a survival advantage due to the correct utility of cytokines and other growth signals (Kume et al., 2002). Importantly, there existed an unmet clinical need; over 75% of ADA-SCID cases have no suitable HLA matched donor for allogenic HSC transplantation leaving treatment options limited to weekly enzymatic replacement therapy (ERT) (reviewed in Aiuti et al., 2009).

The first *ex vivo* gene therapy attempts were to treat ADA-SCID and involved γ-retroviral transduction of CB and BM $CD34^+$ cells (Bordignon et al., 1995; Kohn et al., 1995), although the frequency of genetically modified HSCs was limited with little clinical effect. Soon after, a clinical trial was initiated for X-linked SCID (SCID-X1) to restore the IL2 receptor gamma (IL2RG) (Cavazzana-Calvo et al., 2000) and reported normal numbers of functional T cells with limited numbers of transduced B cells shown two years after therapy (Hacein-Bey-Abina et al., 2002). Following this, a significant breakthrough came through introduction of non-myeloablative chemotherapy with busulfan prior to transplantation which allowed gene modified HSPCs increased access to the niche (Aiuti et al., 2002).

Early clinical trials highlighted the remarkable potential of gene therapy and in the following years, trials for other monogenic disorders using gamma retroviral vectors were conducted including X-linked chronic granulomatous disease (X-CGD) (Ott et al., 2006) and Wiskott-Aldrich Syndrome (WAS) (Boztug et al., 2010). Importantly, some of these trials experienced a major shortcoming caused by the use of γ-retroviral vectors. Gammaretroviral vectors have a tendency to cause insertional mutagenesis due to the long-terminal repeat (LTR) sequences integrating near promoters of proto-oncogenes to ectopically activate expression (Howe et al., 2008; Montini et al., 2009). Indeed, 4 out of 9 children treated with gene therapy for X linked SCID later developed T cell leukaemia (Hacein-Bey-Abina et al., 2008) with similar complications were observed in the trials for X-CGD (Ott et al., 2006) and WAS (Boztug et al., 2010).

Despite setbacks, an extended Phase I/II trial was conducted using a γ-retroviral vector in ten individuals with ADA-SCID and in this seminal study long-term, multi-lineage reconstitution was observed with no insertional mutagenesis (Aiuti et al., 2009). Positive clinical trial data in long-term follow up studies culminated in the decision for European market approval of Strimvelis in 2016, the first *ex vivo* HSC gene therapy product for the treatment of ADA-SCID (reviewed in Aiuti et al., 2017). Shortly after,

a later update of this trial demonstrated robust immune reconstitution up to 13 years post therapy (Cicalese et al., 2016).

1.7.2 Early improvements in viral vectors for gene therapy

Since early setbacks, the field has focussed on the development of safe lentiviral vectors which exploit the core machinery of HIV for efficient nuclear translocation and genomic integration (Naldini et al., 1996). It is now standard practise to use third generation self-inactivating (SIN) lentiviral vectors which have a markedly improved safety profile. All dispensable HIV-1 genes have been removed to reduce virulence (Dull et al., 1998) and deletions are present within the 3' LTR U3 region making the vector replication incompetent (Yu et al., 1986). Upon integration, the 3' UTR is transferred to the 5' LTR U3, "self-inactivating" the 5'LTR promoter/enhancer function. Finally, packaging functions are also separated across three plasmids to reduce the risk of recombination during vector production (reviewed in Gándara et al., 2018). Modified clinical lentiviral vectors have several additional benefits. The native HIV-1 envelope has been replaced by a VSV-G coat protein which allows for increased cell tropism and concentration of vectors to a high titre by ultracentrifugation (Burns et al., 1993). In addition, lentiviral vectors do not rely on dissociation of the nuclear membrane for integration, allowing for transduction in non-dividing/quiescent cells (reviewed in Matreyek and Engelman, 2013).

1.7.3 Current lentiviral vector based clinical trials

Since the seminal investigation studies and the European market approval of Strimvelis in 2016, integrative lentiviral vector approaches are underway for a growing range of immunodeficiency disorders including ADA-SCID, SCID-X1, WAS and X-CGD with the majority of treated patients showing polyclonal engraftment and effective immune reconstitution (reviewed in Tucci et al., 2021).

HSCs have now also been used to treat inherited lysosomal/peroxisomal storage disorders which severely impair the nervous system. This is based on the premise that gene corrected HSCs expressing normal to supraphysiological levels of the target gene give rise to corrected macrophages and microglia which can infiltrate the bone, muscle and brain (reviewed in Sagoo and Gaspar, 2021). One successful application of this approach is within the lysosomal storage disorder Metachromatic leukodystrophy (MLD), where arylsulfatase A (ARSA) deficiency leads to toxic accumulation of sulfatide causing widespread demyelination and neurodegeneration. In a pivotal study, Biffi et al. showed for the first time that HSC gene therapy provided ARSA reconstitution to above normal levels resulting in arrested MLD disease progression (Biffi et al., 2013). Following this, a clinical trial for 33 patients with

MLD was initiated and show reconstituted ARSA expression and with a good safety profile (Fumagalli et al., 2020). HSC gene therapy has also recently been successfully used to treat Hurler Syndrome (Mucopolysaccharidosis I; MPS I) where loss of functional α-L-iduronidase (IDUA) causes glycosaminoglycan (GAG) accumulation in peripheral organs and the nervous system leading to a wide variety of complex symptoms including hepatosplenomegaly, brain atrophy and spinal cord compression (Gentner et al., 2021). Approximately two years following lentiviral gene therapy of mobilized peripheral blood $CD34^+$ cells, 100% of patients (8/8) exhibited sustained supraphysiological blood IDUA activity with detectable levels of IDUA in cerebrospinal fluid (Gentner et al., 2021). Similarly, HSC gene therapy is now being performed for a range of other lysosomal storage disorders such as adrenoleukodystrophy (ALD), MPS-IIIA, Gaucher disease and Fabry disease (reviewed in Sagoo and Gaspar, 2021).

Transfusion dependent β-thalassemia (TDT) and sickle cell disease (SCD) are caused by mutations in the haemoglobin β subunit gene (HBB) and are the most common monogenic diseases worldwide. The gene therapy field has therefore been eager to establish a durable treatment option for these large patient cohorts. However, the design of vectors has been challenging due to the complex regulation of the β-globin gene by introns, the promoter and the β-globin Locus Control Region (LCR) (May et al., 2000). There are now multiple gene therapy clinical trials underway for TDT and SCD using mPB $CD34^+$ cells and a range of lentiviral based approaches which show varying degrees of success (reviewed in Tucci et al., 2021). One recently reported trial used a vector containing the nucleic acid substitution (HbA^{T87Q}) which strongly inhibits polymerization of sickle haemoglobin (Thompson et al., 2018). Variable corrected haemoglobin expression was observed and six of the nine patients with the most severe β^0/β^0 genotype continuing to receive transfusions around 1 year after therapy, albeit with a reduced transfusion volume (Thompson et al., 2018). Another approach used a lentiviral vector containing the β-globin gene under control of the LCR and targeted cells by intrabone injection (Marktel et al., 2019). Here, transfusion was discontinued in 75% of pediatric patients although with a lower success observed in in adults.

1.7.4 Lentiviral vector integration site analysis and safety profile

Lentiviral based HSC gene therapy clinical trials have shown preferential integration within transcriptionally active regions with a semi-random profile and no preferential insertion near proto-oncogene sequences (Aiuti et al., 2013; Biffi et al., 2011, 2013) easing concerns surrounding the safety of gene therapy. In Feburary 2021 bluebird bio halted SCD trials after one incidence of acute myeloid leukaemia (AML) and one incidence of myelodysplastic syndrome (MDS), a blood cell disorder which can

precede leukemia, was reported (reviewed in Sagoo and Gaspar, 2021). Furthermore, in August 2021 a case of MDS, a blood cell disorder which can precede leukemia, was reported in a bluebird bio clinical trial for ALD (bluebird bio, 2021a). Of note, independent analyses of vector integration sites have now shown that in the SCD trial the case of AML was very unlikely attributed to the gene therapy (bluebird bio, 2021b) and similarly the causality of the MDS in the SCD patient was determined as due to underlying inflammation in the SCD patients bone marrow (Sagoo and Gaspar, 2021). However, evidence suggests that in the ALD trial, the case of MDS was likely mediated by lentiviral insertion (bluebird bio, 2021a) and it can be speculated this may be attributed to the broad, constitutive promoter (*MND)* used in this study (Eichler et al., 2016). Overall, recent lentiviral vector safety concerns highlight the importance of careful vector design to mitigate insertional oncogenesis and longitudinal follow up studies to determine the causality of diseases unfortunately arising after therapy.

1.7.5 Genome editing approaches to gene therapy

Genome editing approaches are recently showing exciting promise within gene therapy. Unlike semi-randomly integrating vectors, genome editing involves the precise modification of DNA through site specific deletion, insertion and substitution. Genome editing can be performed by a range of methods, notably through the widely used Clustered Regulatory Interspaced Short Palindromic Repeats (CRISPR)- Cas systems (reviewed in Ferrari et al., 2021). The distinctive repeat sequences of the CRISPR locus were first described in *E.coli* by Ishino et al. (Ishino et al., 1987) and later in a halophilic archaeon by Mojica et al. (Mojica et al., 1993), who subsequently identified the repeating sequences as foreign genetic elements from bacteriophage DNA (Mojica et al., 2005). Subsequent research unveiled a powerful system involving enzymatic cleavage (Bolotin et al., 2005) which acts as an adaptive immune response (Barrangou et al., 2007), with precise DNA targeting (Deltcheva et al., 2011), and can be exploited for a wide range of applications. More specifically, using only a single stranded guide RNA associated with a Cas-9 enzyme, a targeted double strand break can be induced at a gene of interest (reviewed in Doudna, 2020). Upon Cas-9 cleavage the target site undergoes endogenous repair, typically through either error prone non-homologous end joining (NHEJ) or the high fidelity homology directed repair (HDR) (reviewed in Adli, 2018). When the CRISPR-Cas system is combined with a DNA repair template in actively dividing cells, HDR allows for integration of the template sequence, effectively permitting precise genome editing (Doudna, 2020).

Both NHEJ and HDR mediated applications have been investigated therapeutically for the treatment of monogenic blood disorders. Recently, a neat, indirect method to treat TDT and SCD was achieved in a clinical trial by electroporation mediated CRISPR-Cas9 targeting of the transcription factor *BCL11A*

in mPB CD34$^+$ cells (Frangoul et al., 2021). NHEJ restored γ-globin synthesis and recapitulated the phenotype of foetal haemoglobin with transfusion independence observed over one year following therapy (Frangoul et al., 2021). HDR mediated approaches have also been investigated, although are still at pre-clinical stage of development for the treatment of SCD (Dever et al., 2016; Romero et al., 2019) and X-CGD (De Ravin et al., 2017) amongst many others (reviewed in Ferrari et al., 2021).

1.7.6 Challenges posed by recent success and possible solutions

Importantly, despite great progress there are still many improvements to be made. It remains a significant challenge to scale up the investigational studies from rare diseases to manufacturing personalised medicine at a commercial scale. Within the umbrella of "scalability issues" exists remaining challenges related to logistics, cost, vector production and the composition of the target cell population.

One significant drawback of HSC gene therapy is that cells need to be infused within days of correction, resulting in chronically unwell patients travelling to specialised centres. To overcome this, cryopreserved cells are currently being tested in a lentiviral based clinical trial to treat ADA-SCID and MLD (Fumagalli et al., 2020). This permits increased flexibility with the added benefits of full characterisation of the product before infusion and adjustment of the conditioning regimen where necessary (reviewed in Kohn et al., 2019).

Gene therapy is a high cost treatment and Strimvelis is marketed at €594,000 per patient (reviewed in Kohn et al., 2019). Of note, this cost is approximately double that of annual ERT and combined with the reduced cost of intensive care over the patient's lifetime, a one-time curative gene therapy is both more cost effective and provides a higher quality of life than alternative treatments (reviewed in Kohn et al., 2019). Nonetheless, the cost of gene therapy is a major concern. The treatment cost is attributed to the vector development costs, the high dose of corrected CD34$^+$ cells required for re-infusion and G.M.P standards which must be adhered to throughout the therapeutic process. LT-HSCs which hold the potential for long-term reconstitution are rare (<1%) within the CD34$^+$ compartment (Notta et al., 2011) and therefore the majority of viral vector is used to transduce short-term progenitor populations. Both HSC and progenitor populations are key to facilitating combined short and long-term recovery following therapy (Biasco et al., 2016; Scala et al., 2018) however, it is critical that a sufficient LT-HSC population is corrected. Future therapies may aim to re-infuse pre-selected phenotypic LT-HSC, ST-HSC and progenitor populations in order to deliver a better understood cell composition (Zonari et al., 2017).

Large quantities of purified G.M.P viral vectors are required to transduce high numbers of $CD34^+$ patient cells, and vector production is the most expensive component of gene therapy (reviewed in Sagoo and Gaspar, 2021). Recent advances in viral vector research aim to improve transduction efficiency and production scale. Previously, manufacturing of viral vectors was performed by transient transfection into producer cells which is laborious and poorly reproducible. Stable producer cell lines have now been generated to consistently produce high titre vectors and if implemented in clinical trials will reduce costs (reviewed in Sagoo and Gaspar, 2021). Various small molecules have also been introduced to improve transduction such as poloxamers and PGE2 which promote membrane integrity following viral infection (Hauber et al., 2018; Zonari et al., 2017).

A recurring challenge in the field, fuelled by the recent emergence of leukaemia in some treated patients (bluebird bio, 2021b; Orchard therapeutics, 2020) is how best to deliver the corrected gene. Integrating viral vectors remain the method of choice for gene addition although, electroporation is now the preferred delivery method for programmable nucleases and DNA templates (Genovese et al., 2014) and it is likely that an increasing number of clinical trials will implement this approach (Frangoul et al., 2021). Importantly, HSCs are notoriously difficult to edit, limiting the progress of this avenue of gene therapy. This is partially attributed to their quiescent status as gene correction using an exogenous DNA template requires HDR which only active in S-G_2 cell cycle phases (Genovese et al., 2014). Recently, double stranded breaks have been shown to activate the p53 response causing a proliferation delay (Schiroli et al., 2019) and limiting the number of edited HSCs capable of repopulation (Ferrari et al., 2020). Transient p53 inhibition and forced cell cycle progression through upregulating components of the HDR machinery markedly improved HDR efficiency and improved polyclonal graft contribution (Ferrari et al., 2020) which may be implemented in future clinical trials.

Finally, targeting of LT-HSCs *ex vivo* is still a major challenge. Despite the sustained presence of corrected T cells in ADA-SCID treated patients up to 13 years since therapy (Cicalese et al., 2016) limited myeloid cell correction is observed with gene marking in the $CD34^+$ compartment only at 1-2% (Aiuti et al., 2017). Moreover in a recent clinical trial for X-CGD, one out of seven patients showed a transient response, highlighting that the LT-HSC population had not been sufficiently targeted (Kohn et al., 2020). Further, considerable variation in is observed within trials at matched time-points as exemplified by recent TDT trials (Marktel et al., 2019; Thompson et al., 2018) which may suggest a lack of understanding of the cell composition being re-infused into the patient. Exacerbating this issue further is that HSC numbers are limited to what is collected. This is due to loss of LT-HSC self-renewal capacity

ex vivo and cell loss during the *ex vivo* engineering process. *Ex vivo* expansion conditions (reviewed in Wilkinson et al., 2020) may provide an effective solution to this challenge, although the mechanisms which drive loss of self-renewal capacity *ex vivo* in current clinically approved conditions are poorly understood.

1.8 An introduction to the Laurenti laboratory expertise and the motivations for the present study

Within Dr. Elisa Laurenti's laboratory, we are interested in applying single cell biology approaches to understand the dynamics of human blood production over an entire lifetime and across anatomical space. This work is primarily carried out through integrated single cell transcriptomics, *in vitro* single cell functional assays and *in vivo* xenograft transplantation approaches. Most notably, the Laurenti laboratory discoveries have included:

1. **Dissecting human foetal development:** HSC/MPP numbers and differentiation potential are, for the first time, shown to change throughout human gestation, with an early erythroid predominance shifting to a greater lymphoid contribution over gestational age (Popescu et al., 2019). Collectively this work provides the first single cell atlas of human foetal development.

2. **Characterising human HSCs at distinct anatomical sites:** Human HSPC subsets across mPB, peripheral blood (PB), spleen and BM are discovered to have distinct transcriptional signatures and functional characteristics based on anatomical location (Mende et al., 2022). This work shows a functional bias in PB HSPCs towards erythroid-megakaryocytic differentiation and identifies resident splenic progenitors which are only activated to proliferate in response to stress. This study represents the first integrated transcriptional and functional comparison of extramedullary HSPCs.

3. **Characterising functional heterogeneity in human LT-HSC pool**: The human LT-HSC fraction is revealed to be polarised by cell surface expression of CLEC9A and CD34, allowing the prospective isolation of novel functional subsets (Belluschi et al., 2019). CLEC9A^{hi}CD34^{lo} LT-HSCs exhibit slower quiescence exit kinetics and multilineage repopulating capacity as opposed to CLEC9A^{lo}CD34^{hi} LT-HSCs which show faster quiescence exit kinetics and myelo-lymphoid restricted differentiation capacity (Belluschi et al., 2019). This work provides

the first evidence that not all human LT-HSCs are multipotent and identifies the earliest point of lymphoid commitment understood to date.

4. **Understanding human aging:** Impressive clonal tracking of human HSCs through somatic mutation burden was performed across a human lifetime (0 – 81 years) and identified rapid, drastic remodelling of the HSC pool after the age of 70 (Mitchell et al., 2022). The shift in clonal dynamics upon ageing is explained by positive selection of mutations and is characterised by a profoundly reduced clonal diversity. This study provides the first explanation as to why optimal, functional haematopoiesis abruptly declines in the elderly.

The basis of the present project was established from Dr. Laurenti's prior interest in studying the impact of HSC quiescence exit on resulting stem cell function. Before establishing their own group, Dr. Laurenti's work pertaining to this interest first demonstrated that dormant HSCs which cycle extremely infrequently capture the vast majority of long-term repopulation potential in the HSC pool (Wilson et al., 2008). This paper, amongst subsequent others around the time (Bernitz et al., 2016; Foudi et al., 2009; Qiu et al., 2014), established the dogma that cell cycle progression *in vivo* drives loss of long-term repopulation capacity in HSCs. Following this, Dr. Laurenti discovered that CDK6 protein levels differentially regulate quiescence exit length within the human HSC pool, with ST-HSCs showing higher CDK6 protein levels being primed for rapid cell cycle re-entry in comparison to LT-HSCs (Laurenti et al., 2019). Overall, this paper speculates that intrinsic differences in cell cycle kinetics may form the basis for the different functional roles of ST-HSCs and LT-HSCs in blood production. Taken together, and given that LT-HSC culture is associated with both quiescence exit and loss of self-renewal capacity, these studies warranted asking the provocative question of whether cell cycle progression drives loss of self-renewal capacity *ex vivo,* and this formed the motivation for the present study.

2. AIMS

The field of gene therapy is building on remarkable progress but is hampered by outstanding issues. This thesis is divided into three chapters to contribute to addressing current challenges.

Chapter 1 Aim: Characterise the effect of a lentiviral *ex vivo* gene therapy protocol on the transcriptome of LT-HSCs with respect to ST-HSCs and progenitor populations

Targeting of long-term HSCs *ex vivo* is still a major challenge as exemplified by the transient success observed in some clinical trials and high variation between patients treated with the same protocol. It is therefore paramount to understand the effect of the *ex vivo* modification protocol on the LT-HSC compartment. To address this question, I perform scRNA-Seq of LT-HSC, ST-HSC and $CD34^+$ compartments in 4 mPB healthy donors before and after a 62 hr *ex vivo* lentiviral gene therapy protocol.

Chapter 2 Aim: Define the kinetics associated with loss of stem cell capacity and identify potential drivers of self-renewal capacity loss in culture

Secondly, it is well appreciated that HSC repopulation potential reduces following *ex vivo* manipulation. Long-term reconstitution is observed up to 13 years following therapy (Cicalese et al., 2016) but the extent to which long-term repopulation capacity is lost and the mechanisms driving such attrition in HSC function are unclear. To tackle this question, I perform long-term transplantation of CB and mPB HSPC populations at successive time-points in clinically relevant culture and dissect transcriptional changes in a time resolved manner using scRNA-Seq.

Chapter 3 Aim: Is cell cycle progression responsible for the loss of self-renewal capacity associated with HSC culture?

Finally, loss of repopulation capacity *ex vivo* is correlated with increased cell cycling. However, the dependency of cell cycle progression to this process is unclear and often over simplified. The presumption from current literature is that cell cycle progression itself drives the loss of self-renewal capacity associated with HSC culture. To formally test this question, I reversibly arrest cell cycle progression in early G_1 and assess resulting HSC behaviour through scRNA-Seq combined with *in vitro* and *in vivo* functional stem cell assays.

3. METHODS

All methods were performed at Cambridge Stem Cell Institute (Cambridge, UK) unless otherwise indicated.

3.1 Sample Preparation

3.1.1 Umbilical cord blood

CB samples were obtained with informed consent from healthy donors by Cambridge Blood and Stem Cell Biobank (CBSB) in accordance with regulated procedures approved by the relevant research and ethics committees (07/MRE05/44 and 18/EE/0199 research studies). CB samples were pooled independently of sex and processed as a single sample, with the exception of scRNAseq experiments where only single sex CB samples were used.

CB Mononuclear cells (MNCs) were isolated from whole blood (diluted 1:1 in PBS) by Pancoll (PAN-biotech) density gradient centrifugation at 500g for 25 minutes with the brake off. MNCs were harvested and incubated with Red Blood Cell Lysis Buffer for 15 minutes at 4°C (BioLegend). MNCs were then incubated in PBS + 3% foetal calf serum (FCS) (90μl/10^8 cells) containing $CD34^+$ selection beads (Miltenyi Biotech) (30μl/10^8 cells) and FcR blocking reagent (30μl/10^8 cells) for 30 minutes at 4°C. Cells were washed, resuspended in an appropriate volume of AutoMACS buffer (2-4ml) and $CD34^+$ selection carried out using the AutoMACS cell separation machine (Miltenyi Biotech) selecting programme "Posseld2". $CD34^+$ cells were counted manually and stored at -150°C until further use.

3.1.2 Mobilised peripheral blood

Mobilized peripheral blood (mPB) was obtained from healthy male donors aged 25 to 28 by administration of daily Filgrastim (Neupogen) (10μg/kg/day) for 5 days. Apheresis was performed on day 5 and 6 using the Optia Spectra (Terumo BCT). mPB samples were shipped from the US to the GSK research facility (Stevenage, UK). The human biological samples were sources ethically and their research use was in accord with the terms of the informed consents under an IRB/RC approved protocol.

mPB $CD34^+$ cells were enriched by the CliniMacs Prodigy system (Miltenyi Biotech) using the LP-34 programme for 5 hours (hr) and 45 minutes. Briefly, the LP-34 programme consists of successive platelet washes, CD34 AB-nanobead conjugate incubation, immunomagnetic column selection, cell

cultivation, volume reduction and finally target cell elution. Following $CD34^+$ enrichment, cells were cryopreserved in 10% DMSO (Sigma) in FCS (RBMI). Experimental work on mPB $CD34^+$ cells was also covered under 07/MRE05/44 and 18/EE/0199 REC approved research studies. mPB sample preparation was performed at the GSK research facility (Stevenage, UK).

3.2 Lentiviral vector generation

3.2.1 Lentiviral vector production and purification

Third generation self-inactivating (SIN) lentiviral vectors expressing GFP under control of an EIF1α promoter were produced by transient transfection in HEK293T cells. HEK293T cells were cultured at 37 °C/ 5% CO_2 in an orbital shaker and passaged until the optimum density ($1.5x10^6$ cells/ml) achieved. Cells were centrifuged at 300g for 10 minutes and resuspended in G.M.P Stem Cell Growth Medium (SCGM) (CellGenix) containing Polyethylenimine PEIPro (Polyplus) and plasmids generated at GSK, Stevenage: Gag-Pol (pG3-SYNGP), Rev (pG3-REV), VSVG (pG3-VSVG) and the transfer plasmid (pG3T). Cells were incubated for 24 hr at 37 °C/ 5% CO_2. 5mM Sodium Butyrate (Sigma) was added to enhance transfection and cells incubated for a further 48 hr. HEK293T cells were then centrifuged at 1000g for 20 minutes at 4 °C and the supernatant clarified using 0.8μm and 0.45μm vacuum filters. The clarified supernatant was centrifuged at 5000g for 20 hr at 4 °C, the supernatant discarded and the pellet air dried. Pellets were resuspended in SCGM and stored at -80°C. Lentiviral vector production and purification was performed at the GSK research facility (Stevenage, UK).

3.2.2 Vector Titration

CEM A3-04 cells were cultured in RPMI until optimal density ($1.5x10^6$ cells/ml) obtained. CEM cells were centrifuged at 300g for 5 minutes and resuspended in RPMI containing 8μg/ml protamine (Sigma) and plated in a 24 well plate at a density of $1.5x10^6$ cells/ml. The density of cells was kept constant between wells. Serial dilutions (4^{-1} to 4^{-10}) of purified lentiviral vector were added to wells and cells incubated at 37°C/5% CO_2 for 2 hr. 1.5ml of RPMI 1640 (supplemented with 1% Glutamine, 10% FCS and 1% Penicillin/Streptomycin (Pen/Strep) (ThermoFisher Scientific) was added to each well and plates were again incubated at 37°C/ 5% CO_2 for 5 hr. GFP positivity was determined by flow cytometry following 4 days of culture. CEM cells were centrifuged at 500g for 5 minutes, resuspended in PBS + 5% Human Serum Albumin (HSA) (Irvine Scientific) containing 1% 7-AAD Viability dye (Biolegend). GFP positivity was analysed on the MACSQuant Analyzer 10 (Miltenyi Biotech). Vector titre (TU/ml) was calculated as = (Number of cells x $\%GFP^+$ x Dilution Factor) / Transduction volume (0.5

mL). Appropriate measures to control for psuedotransduction were taken. Vector titration was performed at the GSK research facility (Stevenage, UK).

3.3 Flow Cytometry

3.3.1 Fluorescence-Activated Cell Sorting (FACS)

mPB and CB $CD34^+$ cells were thawed by dropwise addition of pre-warmed Iscove's Modified Dulbecco's Medium (IMDM) (Thermo Fischer Scientific) supplemented with 0.1mg/ml DNAse (Lorne Laboratories) and 50% FCS. Cells were centrifuged at 500g for 5 minutes and incubated in PBS + 3% FCS containing an antibody mix (Table 3.1; Panel A) for 20 minutes. All staining was performed at room temperature unless indicated. Cells were washed and resuspended in PBS + 3% FCS for cell sorting on the BD FACS Aria Fusion (BD Biosciences) at the NIHR Cambridge BRC Cell Phenotyping hub. Populations were isolated (Fig.3.1) in either a bulk or single cell manner depending on experimental purpose. Single cells were sorted with the single cell purity setting and index data recorded for all surface markers to allow for retrospective correlation of cell surface phenotype and function. To assess cell surface expression of phenotypic markers at 0 hr and 62 hr following the gene therapy protocol, voltages were not altered for any channels (n=3 mPB scRNA-Seq experiments analysed). Purity for all sorts was >95%.

	Antibody Panel				
Fluorochrome	**A**	**B**	**C**	**D**	**E**
FITC				CD45	
PE		GlyA*	GlyA*		GlyA*
PeCy5	CD49f*	CD45	CD45	CD45**	CD14
PeCy7	CD38	CD14	CD14	CD38*	
APC	CD90	CD56	CD33		CD33*
AF700	CD19		CD19		CD19*
APCCy7	CD34	CD11b	CD3	CD34*	CD3
BV421	CD45RA	CD15			
BV450					CD45
BV510	Zombie	CD41a	CD45	Sytox Blue***	
Vio Bright B515					CD45**

Table 3.1: Antibody panels for cell sorting and flow cytometry analysis. All antibodies/reagents purchased from Biolegend or otherwise BD Biosciences (where indicated *), Beckman Coulter (**) or ThermoFisher (***) and were titrated before experimental use using mononuclear cells (MNCs) and appropriate controls. The CD34-APCCy7 antibody (clone: 581) was custom made from BD Biosciences. The FITC channel was compensated for in experiments involving GFP. The BV510 channel was used to distinguish Live/Dead cells using Zombie Aqua (Biolegend). Panel A: Isolation of phenotypic compartments of LT-HSC, ST-HSC, $CD34^+$, $CD34^+/CD38^-$ and $CD34^+/CD38^+$ populations. Panel B: Determination of colony size and lineage output derived from single sorted LT-HSCs. Panel C: Determination of graft size and lineage output from primary transplants into NSG animals and secondary transplants from mPB in NSG-SGM3 animals. Panel D: Isolation of Viable (Sytox $Blue^-$) $CD45^{++}$ populations from primary mouse BM transplanted with cultured human CB. Panel E: Determination of graft size and composition from secondary NSG-SGM3 animals transplanted with CB.

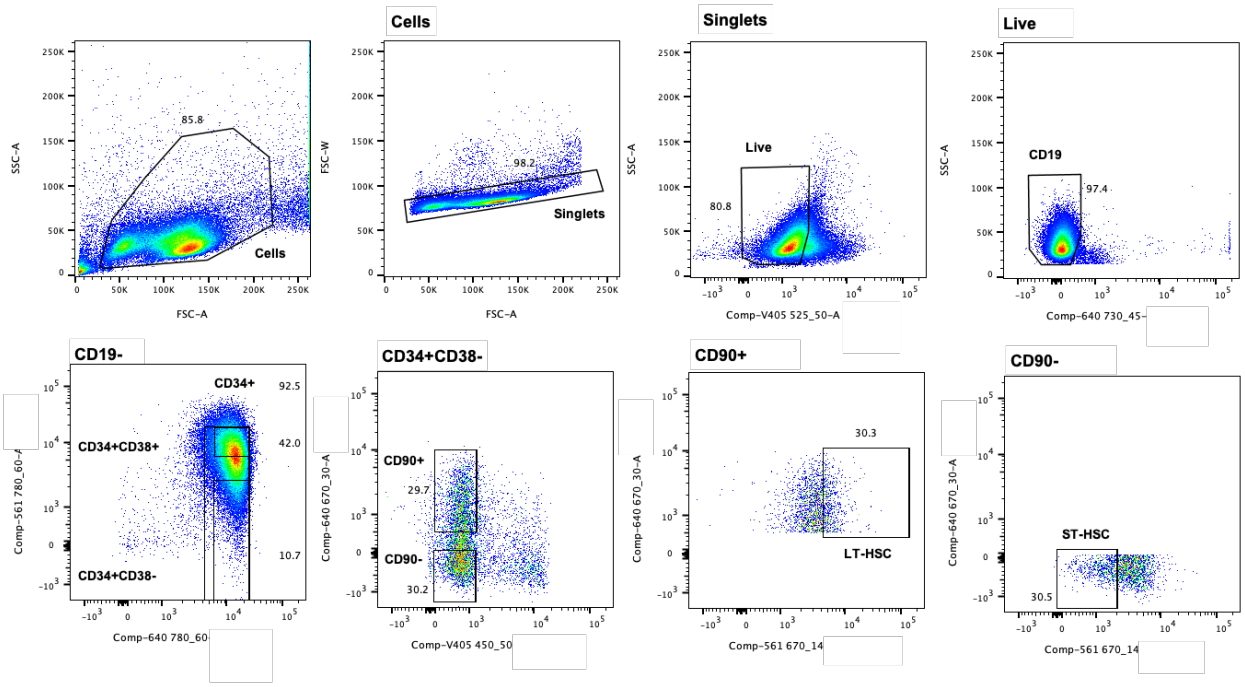

Figure 3.1: Example gating strategy for isolation of phenotypic populations. Representative 0 hr mPB CD34$^+$ enriched donor sample. LT-HSC: Zombie$^-$/CD19$^-$/CD34$^+$/CD38$^-$/CD90$^+$/CD49f$^+$ (top 30% of CD90/CD49f expression). ST-HSC: Zombie$^-$/CD19$^-$/CD34$^+$/CD38$^-$/CD90$^-$/CD49f$^-$ (bottom 30% of CD90/CD49f expression).

For scRNA-Seq experiments performed before and after the gene therapy protocol the following populations were sorted at 0 hr: CD34$^+$ (Zombie$^-$/CD19$^-$/CD34$^+$), LT-HSC: Zombie$^-$/CD19$^-$/CD34$^+$/CD38$^-$/CD90$^+$/CD49f$^+$ (top 30% of CD90/CD49f expression) and ST-HSC: Zombie$^-$/CD19$^-$/CD34$^+$/CD38$^-$/CD90$^-$/CD49f$^-$ (bottom 30% of CD90/CD49f expression). After single cell sorting for scRNA-Seq, populations were bulk sorted and exposed to the lentiviral gene therapy protocol (Methods 3.4.1). At 62hr: Zombie$^-$ (Live) fractions from each population were re-sorted for scRNA-Seq.

3.4 Cell Culture conditions

Cells were cultured under three distinct conditions at 37°C/ 5% CO_2 across all *in vitro* experiments. The CDK6 inhibitor Palbiciclib (PD0332991; denoted as PD) (Sigma) was added at 200nM to *ex vivo* gene therapy conditions, differentiation facilitating conditions and single cell My-Meg-Ery-NK differentiation assay conditions where indicated.

3.4.1 *Ex vivo* lentiviral gene therapy protocol

Ex vivo gene therapy conditions comprise of G.M.P SCGM (CellGenix) supplemented with L-Glutamine (1%), Pen/Strep (1%) in addition to the following cytokines: SCF (300ng/ml), Flt-3L (300ng/ml), IL-3

(60ng/ml) and TPO (100ng/ml) as advised by internal GSK cell culture protocols at the beginning of my PhD and published in Biffi et al., for the first time (Biffi et al., 2013). Cells were cultured in *ex vivo* gene therapy media in a 96 well flat bottom plate coated with 33.3μg/ml of Retronectin (Takara). The 62 hr clinical protocol consists of pre-stimulation in gene therapy media (24 hr) followed by transduction with a lentiviral vector containing GFP (14 hr), an interim incubation in gene therapy media without the vector (10 hr) and a second hit of transduction (14 hr). Vectors of ≥1.19E+08 TU/ml were used for all experiments at multiplicity of infection (MOI) of 100-300. Of note, for scRNA-Seq experiments vectors of ≥4.44E+08 TU/mL were used. Transduction efficiency is estimated by GFP^+ (% of Singlets) at the 62 hr time-point relative to non-transduced.

3.4.2 Differentiation facilitating conditions

Differentiation facilitating conditions comprise of StemPro media (Stem Cell Technologies) supplemented with Nutrients (0.028%), Pen/Strep (1%), L-Glu (1%), human LDL (50ng/ml) in addition to the following cytokines: SCF (100ng/ml), Flt-3L (20ng/ml), TPO (100ng/ml), EPO (3 units/ml), IL-6 (50ng/ml), IL-3 (10ng/ml) and GM-CSF (20ng/ml).

3.4.3 Single cell My-Meg-Ery-NK differentiation assay conditions

Single cell differentiation conditions facilitating the growth of myeloid (My), megakaryocyte (Meg), erythroid (Ery) and lymphoid (Lym; only natural killer (NK) cells supported) colonies comprise of StemPro base media (Stem Cell Technologies) supplemented with Nutrients (0.035%), Pen/Strep (1%), L-Glu (1%), human LDL (50ng/ml) in addition to the following cytokines: SCF (100ng/ml), Flt-3L (20ng/ml), TPO (100ng/ml), EPO (3 units/ml), IL-6 (50ng/ml), IL-3 (10ng/ml), GM-CSF (20ng/ml), IL-11 (50ng/ml), IL-2 (10ng/ml) and IL-7 (40ng/ml).

3.5 *In vitro* functional assays

3.5.1 Time to first division

Cells were single cell sorted into 96 well u-bottom plates containing 100μl of indicated media per well, centrifuged at 500g for 5 minutes and manually counted every 12 hr for 96 hr using a light inverted microscope.

3.5.2 Single cell My-Meg-Ery-NK assay

Cells were single cell sorted into a 96 well flat bottom plate containing 100µl of My-Ery-Meg-NK differentiation assay media (Methods 3.4.3). Cells were harvested after 21 days in culture into 96 well u-bottom plates and incubated for 20 minutes in the dark with an antibody mix (Table 3.1; Panel B). Cells were washed with PBS + 3% FCS and resuspended in 100µl of PBS + 3% FCS for analysis on the LSR Fortessa X-20 (BD Biosciences) using the HTS plate reader. Colony size and lineage output were analysed on colonies with ≥30 cells in ($CD45^+$ & $GlyA^+$ & $CD41a^+$) gates (Figure 3.2).

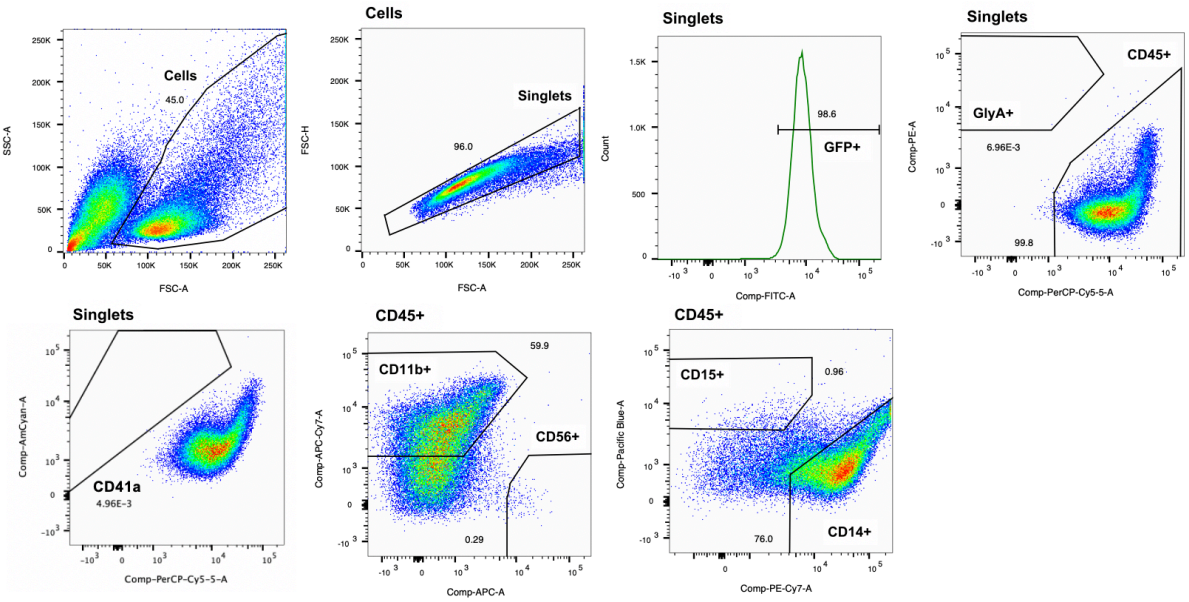

Figure 3.2: Example gating strategy for determination of colony size and lineage output from single cell My-Meg-Ery-NK assay. Representative gating of a colony harvested from a single mPB LT-HSC after 21 days culture in My-Meg-Ery-NK media. A true colony was determined if ≥30 cells observed in ($CD45^+$ & $GlyA^+$ & CD41a) gates. Myeloid: $CD45^+/CD11b^+$. Monocyte: $CD45^+/CD14^+$. Granulocyte: $CD45^+/CD15^+$. Lymphoid (NK only): $CD45^+/CD56^+$. Megakaryocyte: $CD41a^+$. Undifferentiated colonies were determined as having ≥30 cells in ($CD45^+$ & $GlyA^+$ & CD41a) gates however no true lineage.

3.5.3 Cell Cycle Assay

Cells were harvested at indicated time-points and incubated with Cytofix/Cytoperm (BD Biosciences) for 10 minutes to fix cells. Cells were washed with 400µl of 1x Permwash (BD Biosciences) and centrifuged at 500g for 5 minutes. Cells were stained overnight in Anti-Rb (phospho S807/S811) (Cell Signalling Technologies). Cells were then washed with 1x Permwash, centrifuged at 500g for 5 minutes and

resuspended in 1x Permwash containing 0.5μg/ml DAPI (ThermoFisher Scientific) for 15 minutes. Cells were again washed with 1x Permwash and resuspended in PBS + 3% FCS for analysis on LSR Fortessa X-20 (BD Biosciences). DAPI was recorded on a linear scale and samples analysed at ≤ 30 events/second.

3.5.4 Cell Size measurement

100 cells/well were sorted into a 384 well plate containing My-Meg-Ery-NK differentiation assay media either untreated (UNTR) or treated with PD. Cell images were taken every 24 hr. Cell size was analysed using ImageJ and expressed as cell diameter (μm). Cell size measurement experiments and analysis were performed by Serena Belluschi.

3.5.5 Mitochondrial Activity assay

CB populations were cultured in differentiation facilitating conditions, harvested at indicated timepoints and stained with 100nM Tetramethylrhodamine (TMRM) (Life Technologies) for 40 minutes at 37°C/ 5% CO_2. Cells were washed with PBS + 3% FCS, centrifuged at 500g for 5 minutes and resuspended in an appropriate volume of PBS + 3% FCS for analysis on the LSR Fortessa X-20 (BD Biosciences).

3.5.6 Tunicamycin proteostatic stress treatment and apoptosis assay

mPB populations were cultured for 6 hr and 24 hr before addition of 3μg/ml Tunicamycin (Sigma). An equivalent concentration of DMSO was added to control wells. Cells were cultured with Tunicamycin/DMSO for a further 24hr. After culture, cells were washed and resuspended in Annexin-V staining mix comprised of: 50μl of fridge cold PBS containing 2.5μl Annexin-V (PE) (BD Biosciences) and 2.5ul 7-AAD (BD Biosciences) per sample. Cells were incubated for 15 minutes and 200μl of 1x Annexin-V binding buffer (BD Biosciences) added for immediate analysis on the LSR Fortessa (BD Biosciences).

3.6 single cell RNA-Seq library preparation

Sequencing libraries were prepared using the Smart-Seq2 protocol (Picelli et al., 2014) although adapted in the following sections. Single cells were first sorted into 96 well PCR plates containing 4μl of cell lysis buffer per well (0.2% Triton X-100 (Sigma), 2U/μl RNAse inhibitor (Clontech), 5mM DTT (Invitrogen) and 1mM dNTPs (Invitrogen)) and stored at -80°C. An annealing mix was then added containing Extrernal RNA Consortium Controls (ERCC) (Ambion, Life Technologies) diluted to 1:3,000000,

5μM Oligo-dT (Invitrogen), 10mM dNTPs (Invitrogen) and nuclease free H_20 (ThermoFisher Scientific). 23 PCR cycles were performed to account for the low RNA content of HSCs. The DNA: Ampure XP beads ratio used was 1:0.8 for PCR purification. The quality of cDNA produced was detected using the high sensitivity DNA chip (Agilent) and analysed on the 2100 Bioanalyzer (Agilent). Amplified cDNA was quantified by the Picogreen dsDNA Assay kit (Invitrogen) and samples diluted to 0.1-0.15ng/μl. Library preparation was then performed using the Nextera XT DNA sample preparation protocol (Illumina). Tagmentation reaction volumes were altered from the Illumina Nextera NT DNA sample preparation kit as follows: tagmentation DNA buffer (2.5μl), amplicon tagment mix (1.25μl) and using 1.25μl of sample. Tagmentation was performed by incubation of the samples to 55°C for 10 minutes. 1.25μl of Nextera NT buffer was added per sample to strip the Tn5 transposase from tagmented DNA. The size distribution of the amplified pooled library was quantified on the 2100 Bioanalyzer (Agilent). Sequencing was then performed at the CRUK-CI genomics facility by Illumina HiSeq4000. The library preparation was performed by Serena Belluschi for experiments involving time-course culture of CB over differentiation facilitating culture and myself for experiments involving culture of mPB in gene therapy conditions.

3.7 *In vivo* functional assays

3.7.1 Mice

NOD.Cg-PrkdcscidIl2rgtm1Wjl/SzJ (NSG) mice or NOD.Cg-Prkdcscid Il2rgtm1Wjl Tg(CMV-IL3,CSF2,KITLG)1Eav/ MloySzJ (NSG-SGM3) mice were bred in-house or obtained from Charles River. All experimental cohorts were >11 weeks of age. Only female cohorts were used for experiments involving NSG animals. Both male and female animals were used for experiments involving NSG-SGM3 animals. All animals were housed in a specific pathogen free animal facility and experiments conducted under UK Home Office regulations. This research has been regulated under the Animals (Scientific Procedures) Act 1986 Amendment Regulations 2012 following ethical review by the University of Cambridge Animal Welfare and Ethical Review Body (AWERB).

3.7.2 Primary xenograft transplantation experiments

NSG mice were sub-lethally irradiated (2.4 Gy) 24 hr prior to transplantation. For intrafemoral injections, mice were anesthetised with isoflurane and transplanted with the indicated cell dose in PBS + 0.1% Pen/Strep (ThermoFisher) (25μl). Following transplantation, mice were injected subcutaneously with the analgesic buprenorphine (Animalcare) at 0.1mg/kg. For intravenous injection, mice were transplanted with a cell suspension (max 150μl volume) in PBS + 0.1% Pen/Strep by tail vein injection.

For all xenograft experiments involving cell culture, injected cell doses at time-points are representative of the cell count at the time of sort (0 hr). Mice were culled and bone marrow harvested 18-20 weeks post injection. The femur and tibia bones from the two hind legs were taken. For intrafemorally injected mice the injected femur was analyzed separately whereas for intravenously injected mice the BM from the four bones was analysed together. Bone marrow was stained in an antibody panel (Table 3.1; Panel C) for 20 minutes before washing and resuspending in PBS + 3% FCS for analysis on the LSR Fortessa X-20 (BD Biosciences). Mice were considered engrafted if human cells (hCD45^{++} &GlyA^{+}) ≥0.01% of Singlets and if ≥30 cells were present in any lineage determination gate (Fig.3.3). Intrafemoral injections for CB cultured in differentiation facilitating conditions were performed by myself, Serena Belluschi, Emily Calderbank or Nicole Mende. Intravenous injections were performed by myself.

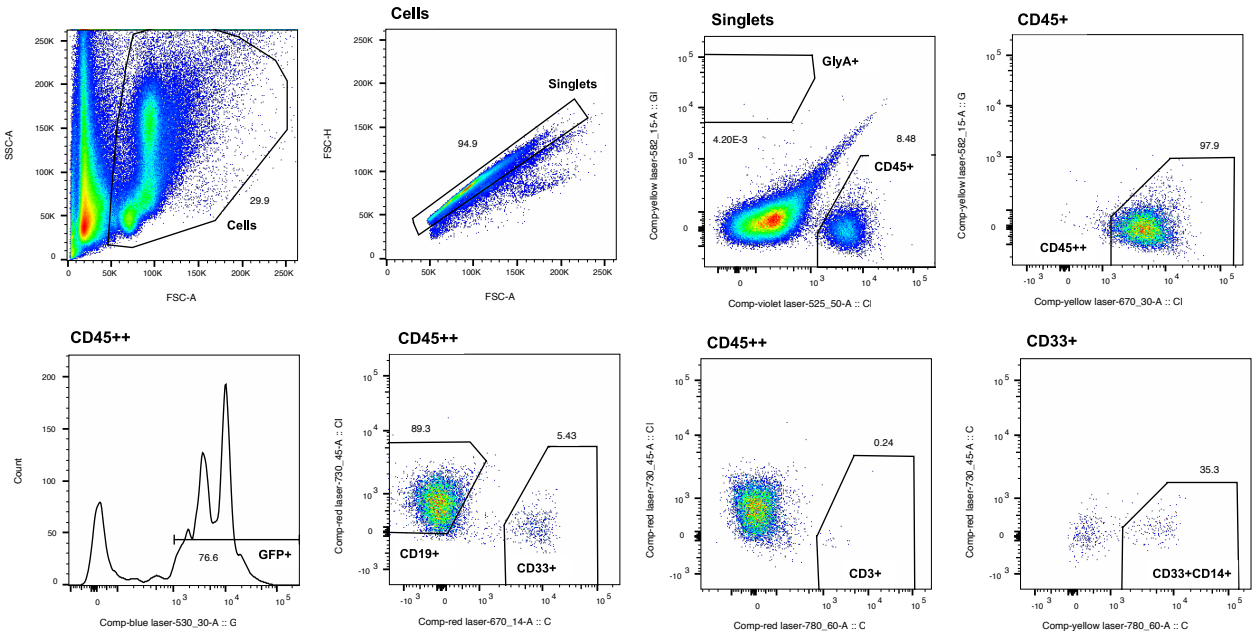

Figure 3.3: Example gating strategy for analysis of mouse BM after primary transplantation (18-20 weeks). Representative example from a mouse injected with mPB CD34^{+}/CD38^{-} cells cultured in gene therapy conditions (62 hr) and transduced with a lentiviral vector containing GFP. Erythroid (Ery) cells were identified as CD45^{-}/GlyA^{+}. Myeloid (My) cells were identified as CD45^{++}/CD33^{+}. Lymphoid (Lym) cells were identified as CD45^{++}/CD19^{+}. Mice were considered engrafted if (hCD45^{++} &GlyA^{+}) ≥0.01% of Singlets and a lineage detected by containing ≥30 cells in any respective lineage gate. Same gating strategy was used for secondary transplantation experiments.

3.7.3 Secondary xenograft transplantation experiments

For secondary transplantation experiments of cultured CB, NSG-SGM3 mice were irradiated using 2.25Gy 24 hr prior to transplantation. Primary BM samples were shipped to Toronto on dry ice and cells were thawed in X-VIVO 10 media (Lonza) + 50% fetal bovine serum (Wisent) supplemented with DNase (100 μg ml^{-1}, Roche). Viable (SytoxBlue^{-}) (ThermoFisher Scientific) human CD45^{++} cells were sorted (Table 3.1; Panel D) on the Aria Fusion (BD Biosciences). Cells were pooled based on condition and intrafemorally injected in three doses (4 fold decrease between each dose). At 8 weeks post transplant, the injected femur was taken alongside the alternate femur of the hind legs. Bone marrow was stained in an antibody panel (Table 3.1; Panel E) and analysed for human engraftment on the FACSCelesta (BD Biosciences). The injected femur was analysed separately. Mice were considered engrafted if human cells (hCD45^{++} &GlyA^{+}) ≥0.01% of Singlets and if ≥30 cells were present in any lineage determination gate. Secondary intrafemoral transplantation experiments for cultured CB were kindly performed at the Princess Margaret Cancer Centre, Toronto by Kerstin Kaufmann, Gabriela Krivdova, and Jessica McLeod (John Dick group) and analysis performed by myself.

For secondary transplants of cultured mPB, primary mouse BM was thawed by dropwise addition of pre-warmed Iscove's Modified Dulbecco's Medium (IMDM) (Thermo Fisher Scientific) supplemented with 0.1mg/ml DNAse (Lorne Laboratories) + 50% FCS. Cells were counted three times using a haemocytometer, an average taken and divided into three doses based on cell count. Cells were intravenously transplanted by tail vein injection using a cell suspension (max 150μl volume) in PBS + 0.1% Pen/Strep. At 8 weeks post-transplant the femur and tibia bones from the two hind legs were taken. Bone marrow was stained in an antibody mix (Table 3.1; Panel C) and analysed on the LSR Fortessa X-20 (BD Biosciences). Mice were considered engrafted if human cells (hCD45^{++} &GlyA^{+}) ≥0.01% of Singlets and if ≥30 cells were present in any lineage determination gate (Fig.3.3).

3.8 single cell RNA-Seq analysis

3.8.1 scRNA-Seq dataset of CB culture in differentiation facilitating conditions

Two repeats of the CB time-course experiment were completed (two batches). Read alignment was performed using GSNAP (Wu et al., 2016) and initial quality control (QC) was performed by FastQC (Andrews, 2010). The count matrix was generated by HTSeq (Anders et al., 2015) and genes were filtered out if expressed in < 3 cells. Additional QC was then performed in the *bglab* package (Jawaid, 2017) using the determined thresholds (Table 3.2).

QC thresholds	Batch 1	Batch 2
Reads mapped to gene	> 2 x 10^5	> 2 x 10^5
Ratio of genes to total number of reads	> 0.3	> 0.2
Ratio of mitochondrial reads to nuclear genes	< 0.15	< 0.2
Ratio of nuclear genes to total mapped reads	> 0.75	No threshold
Number of genes with 10 reads per million	> 2000	No threshold
Ratio of ERCC spike-in to mapped reads	< 0.2	No threshold

Table 3.2: Quality control thresholds for analysis of the CB time-course scRNA-Seq dataset. Values generated by Kenny Sham.

Batch Number	Time-point	Before QC	After QC
Batch 1	0 hr	72	29
	6 hr	72	49
	24 hr	72	28
	72 hr UNTR	72	42
	72 hr PD	72	27
Batch 2	0 hr	95	56
	6 hr	95	85
	24 hr UNTR	95	58
	24 hr PD	95	80
	72 hr UNTR	95	42

Table 3.3: Number of sequenced CB LT-HSCs cultured in differentiation facilitating conditions and number of cells retained for analysis following QC from each batch. Values generated by Kenny Sham.

A natural log transformation was performed with a pseudocount of 1 added to mitigate the mean-variance relationship, reduce skewing of data and account for drop-outs. ComBat (Zhang et al., 2020) was then used for batch correction specifying time-points as covariates. Highly variable genes (HVG)

were selected in scanpy (parameters: minimum_mean = 0.05, maximum_mean = 13, minimum_dispersion = 0.1, maximum_dispersion = 3). 16276 HVG were retained for further analysis. A PCA space was calculated and BBKNN performed to integrate the datasets from two experiments (batches). 15 nearest neighbours were calculated using the K-nearest neighbour (KNN) algorithm and a force directed graph (FDG) or uniform manifold approximation and projection (UMAP) were generated. A diffusion map was also generated to calculate Dpt pseudotime from a starting 0 hr cell. Differential expression analysis between time-points and conditions (including PD treated cells) was performed using the raw count matrix by DESeq2 (Love et al., 2014), taking into account the batch and time-point in the design matrix. The size factor and dispersion were estimated and a Wald test performed for statistical analysis. The list was filtered for *FDR < 0.05*. 10010 genes differentially expressed were found between any pairwise comparison over the UNTR time-course of CB LT-HSC culture. Cell cycle phase assignment was performed in scanpy using the Satija lab cell cycle scoring and regression (Satija lab, 2017) with a list of 97 genes identified as key for cell cycle progression (Tirosh et al., 2016). Cell cycle regression was then performed using the function "regress_out", which regresses the G_2M and S score from the dataset. Analysis performed by Kenny Sham.

3.8.2 scRNA-Seq dataset of mPB culture in gene therapy conditions

mPB LT-HSC, ST-HSC and $CD34^+$ populations were sequenced before and after the *ex vivo* gene therapy protocol (Methods 3.4.1) from repeat experiments using individual mPB donors (four batches total). Read alignment was performed using GSNAP (Wu et al., 2016) and initial QC by FastQC (Andrews, 2010). The count matrix was generated by HTSeq (Anders et al., 2015). Additional QC was performed independently for each batch using the Smart-Seq2 pre-processing package smqpp (Wang, 2020).

QC thresholds	Batch 1	Batch 2	Batch 3	Batch 4
Reads mapped to nuclear genome (log10)	$> 10^{5.2}$	$> 10^{5.2}$	> 15000	> 100000
Ratio of genes to total number of reads	> 0.3	> 0.3	> 0.1	> 0.2
Number of genes with 10 reads per million	No threshold set	> 1000	> 1000	> 1000
Ratio of mitochondrial reads to nuclear genes	< 0.2	< 0.2	< 0.2	< 0.2
Ratio of ERCC spike-ins to mapped reads	< 0.4	< 0.4	< 0.2	< 0.2

Table 3.4: Quality control thresholds for scRNA-Seq analysis of mPB cultured in the *ex vivo* gene therapy protocol. Values generated by Xiaonan Wang/Kenny Sham.

Batch Number	Cell Type	Before QC	After QC	After outlier removal
Batch 1	LT-HSC UNTR	210	189	188
	ST-HSC	141	130	130
	$CD34^+$	190	175	175
Batch 2	LT-HSC UNTR	162	131	131
	LT-HSC PD	96	75	75
	ST-HSC	155	105	104
	$CD34^+$	153	127	126
Batch 3	LT-HSC UNTR	134	80	80
	LT-HSC PD	24	12	12
	ST-HSC	112	40	40
	$CD34^+$	110	80	77
Batch 4	LT-HSC UNTR	240	117	117
	LT-HSC PD	44	7	7
	ST-HSC	230	120	120
	$CD34^+$	232	132	130

Table 3.5: Number of mPB cells cultured in gene therapy conditions sequenced and number of cells retained for analysis following QC from each batch/repeat experiment. Values generated by Xiaonan Wang/Kenny Sham.

The total number of cells after quality control and outliers filtering was 1512. The scanpy function 'filter_genes' was then used to filter low quality genes with parameters: min_cells = 1, and default values for others. The counts were normalised using the function 'normalise_data' from the package smqpp (Wang, 2020) with default parameters used and a pseudo-count of 1 added. 1599 HVG were found and retained using a smqpp function 'tech_var' (parameters useERCC = True, meanForFit = 10, and default values for others). A PCA space was calculated and 10 nearest neighbours for each cell were found. A UMAP was then created in scanpy using default parameters. Differential expression analysis between time-points and conditions (including PD treated cells) was performed using the raw count matrix in DESeq2, taking into account the number of genes and the batch in the design matrix (~ n_genes + batch). The size factor and dispersion were estimated and the Wald test performed for statistical analysis. The list was filtered for *FDR < 0.05*. Cell cycle phase assignment was performed using the Satija lab package (Satija lab, 2017) in the same manner as the CB scRNA-Seq dataset in differentiation facilitating conditions. Analysis was performed initially by Xiaonan Wang and subsequently by Kenny Sham.

3.8.3 Integration of mPB scRNA-Seq dataset to CB scRNA-Seq time-course dataset

Raw counts were natural log normalised by the package smqpp (Wang, 2020) using default parameters and a pseudocount of 1 added. ComBat was performed (Zhang et al., 2020) in scanpy for batch correction with default parameters (no covariate specified). Highly variable genes were found using the scanpy function 'highly_variable_genes' (parameters: min_mean = 1, max_mean = 8, min_disp = 0.5), and default for others. 3255 HVGs were found using the parameters (min_mean = 1, max_mean = 8, min_disp = 0.5; default for others) for the integration of the mPB and CB UNTR scRNA-Seq datasets whereas 3643 HVG were found for the integration of the mPB and CB scRNA-Seq datasets with PD treated conditions. The PCA was calculated with scanpy and 15 nearest neighbours determined for each cell. A UMAP was created for visualisation purposes and a diffusion map was generated to calculate Dpt pseudotime from a chosen 0 hr cell. Integration was performed by Kenny Sham.

3.8.4 Integration of CB or mPB scRNA-Seq dataset to multi-tissue landscape of differentiation and prediction of cell cluster labels

The CB and mPB scRNA-Seq datasets were combined with the multi-tissue landscape of differentiation (Fig.3.4) in the Seurat package (Stuart et al., 2019). Integration was performed separately for CB and mPB datasets. Seurat objects were created from the raw count matrix, filtering for genes expressed in > 3 cells with default values for other parameters. The SCTransform function was used to normalise the datasets and regress the percentage of mitochondrial genes from all counts. The CB/mPB datasets were first prepared for integration using 'PrepSCTIntegration' function which determines the features to use in integration and 'SelectIntegrationFeatures' was used to find 3000 HVG. The dataset was then subsetted accordingly. The 'IntegrateData' function was used to combine the datasets with the parameters: normalization.method = 'SCT' and default values used for other parameters. A PCA was then generated for the combined dataset using default parameters and a UMAP was created for visualisation purposes. The seurat object was used for label projection on the multi-tissue landscape. Leiden clusters 1,3,18,14,24,25,26 were excluded from label projection because clusters represent cells which are unsupported by *in vitro* culture conditions. 'MapQuery' was used to predict the leiden cluster label on the the CB/mPB dataset using the functions reference.reduction = PCA and reduction.model = UMAP. Integration of CB/mPB scRNA-Seq datasets and label transfer were performed by Kenny Sham.

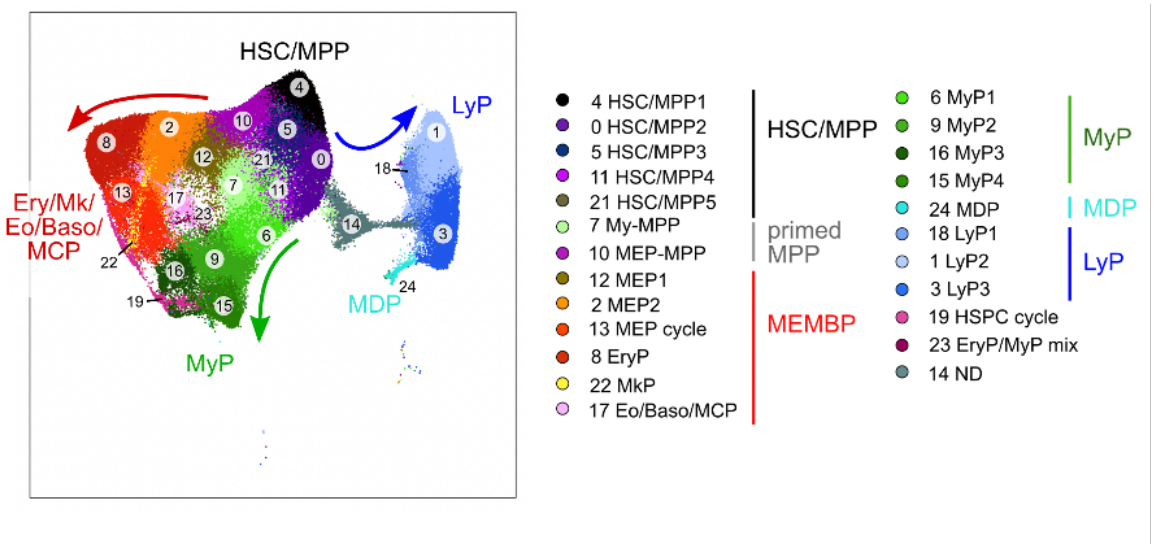

Figure 3.4: Multi-tissue landscape of haematopoeitic differentiation (Mende et al., 2022). Generated from 118,783 $CD19^{-}CD34^{+}$ HSPCs isolated from BM, peripheral blood (PB) and spleen from 10 adult donors. (Number of donors per organ: n=4 BM, n=3 spleen, n=8 PB). Generated by Hugo Bastos.

3.8.5 Downstream analysis of scRNA-Seq datasets

degPatterns (Patano, 2017) was used to group genes based on expression pattern, inputting the count matrix for 10010 genes representing the union of all differentially expressed genes over the CB scRNA-Seq dataset in differentiation facilitating conditions. 8966 genes remained after filtering. A distance matrix was generated from all pairwise comparisons between time-points and hierarchical clustering was performed. 15 clusters were manually found to fit the dataset with patterns showing no redundancy.

To assess gene expression variability within the CB scRNA-Seq dataset Bayesian Analysis of Single cell Sequencing data (BASiCS) was used on the filtered count matrix (Vallejos et al., 2015) and a correction was performed to mitigate the mean-variance dependency (Eling et al., 2019). Differentially variable genes were extracted for analysis. Systematic Identification of Bimodally Distributed genes (SIBER) was also performed (Tong et al., 2013) on the filtered count matrix for the 10010 genes differentially expressed between any pairwise time-point comparison. Within the SIBER package, log-normal distributions were fitted to each gene at each time-point and the distance between two peaks of expression and the expression level quantified. Genes which showed a zero count for for > 20% of all cells in each time-point were considered as zero-inflated and values excluded. Genes with a bimodal index (BI) $\geq$ 1.9 were considered bimodally distributed. SIBER was performed on both CB scRNA-Seq batches separately and an average taken.

The Ratio of Global Unshifted Energy (ROGUE) package was employed (Liu et al., 2020a) to assess the uniformity of expression within clusters of the CB scRNA-Seq dataset by an entropy based statistic. The ROGUE value was calculated for each time-point using default parameters. Of note, there is no ROGUE threshold value advised and this package was used to assess uniformity of a given cluster relatively within the dataset.

3.8.6 Analysis of scRNA-Seq data against curated biological networks/pathways

Gene Set Enrichment Analysis (GSEA) (v4.2) was performed against the c2 curated pathway database using a pre-ranked list by the stat value (value of the Wald test statistic) from the DESeq2 output specifying 10000 permutations. Gene-Set Variation Analysis (GSVA) was also performed (Hänzelmann et al., 2013) against c2 curated pathways in R using default parameters. GSVA calculates an enrichment score per cell, providing information on the variability between pathway expression within a given sample. Cluego (Bindea et al., 2009), a Cytoscape plug in was used for enrichment of gene lists to KEGG/Reactome pathways and Gene Ontology terms with a significance threshold of $p < 0.05$ imposed. The Cytoscape plug in iRegulon (Janky et al., 2014) was used for analysis of transcription factor regulatory networks from inputted gene lists. The 10k motif collection and the 1120 ChIP-seq track collection was selected for enrichment specifying the $FDR < 0.05$.

3.8.7 Analysis of scRNA-Seq data against custom signatures

All gene signatures were created using the top 100 differentially expressed genes contrasting previously reported populations. Signatures were created from $CD34^{lo}CLEC9A^{hi}$ (Subset1) and $CD34^{hi}CLEC9A^{lo}$ (Subset 2) obtained from Belluschi et al. 2018, human dormant LT-HSC populations from the Cabezas-Wallscheid group contrasting activated (cycling) human HSCs (Zhang et al. Nature Cell Biology*, in revision*) and clusters from a multi-tissue landscape of haematopoeisis (Fig 3.4) (Mende et al., 2022). Gene signatures were also used from a landscape of human BM differentiation (Velten et al., 2017). Gene signatures were also generated from Affymetrix microarray data in contralateral (CSC), activated (ASC) and quiescent muscle satellite cell (QSC) populations in mouse (Rodgers et al., 2014). First gene lists were converted from mouse to human using bioMaRt (Durinck et al., 2005) and the top 100 differentially expressed genes contrasting each subset were identified. To calculate enrichment of gene signatures GSEA was performed against a pre-ranked list by the stat value (Wald test statistic) from the DESeq2 output. Significant enrichment to signatures was considered specifying the $FDR < 0.05$.

For analysis of enrichment to custom signatures whilst mitigating the effects of cell cycle, a compiled manually curated list of 613 genes from “Reactome: Cell_Cycle_Checkpoints” and “Reactome: Cell_cycle_mitotic” gene-sets was removed from the count matrix where indicated. Pre-ranked GSEA was then performed (ranked by stat value) specifying 10000 permutations.

3.9 Plotting and Statistical Analysis

Analysis of the HSC frequency from transplanted populations was performed by using Extreme Limiting Dilution Analysis (ELDA) software (Hall, 2014) taking into account the number of engrafted mice as a proportion of the total mice used and the cell dose transplanted. For analysis of a statistical difference in the HSC frequency within two groups, a Chi-Squared test was performed within the ELDA software. Flow cytometry data was analysed using FlowJo software (v10). For analysis of colony data derived from single sorted LT-HSCs, FlowJo v.10 gating statistics were exported and data further analysed in R Studio (v.1.2). Graphpad Prism (v9.3), python (v3.8.6) and R Studio (v1.2) were used for the creation of all plots. Statistical analysis between multiple groups was performed in Graphpad Prism or R Studio. Normality of data was deduced from the Shapiro-Wilks normality test. For statistical analysis between two groups a parametric test (Students t-test) or a non-parametric test (Mann-Whintey U-test) was performed. For analysis between multiple groups, an analysis of variance (ANOVA) test was performed. For categorical analysis, a Fishers Exact test was performed. All statistical tests were performed with a confidence interval of 95%.

4. RESULTS

4.1 Chapter 1: *Ex vivo* gene therapy culture conditions significantly rewire the transcriptome of LT-HSC, ST-HSC and CD34$^+$ cells

4.1.1 An *ex vivo* gene therapy protocol modulates predominantly shared changes across LT-HSC, ST-HSC and CD34$^+$ populations

The vast majority of CD34$^+$ cells (>99%) do not retain the potential for long-term blood reconstitution and instead are important for short-term, steady-state blood formation and tissue repair (Notta et al., 2011). It is also well recognised that progressive *ex vivo* culture almost always leads to a loss in HSC repopulation capacity. As the longstanding goal of HSC gene therapy is a one-time, lifelong treatment option, it is therefore critical to study the impact of this *ex vivo* modification protocol on underlying HSC regulation and function. However, to our knowledge no scRNA-Seq study to date has focussed analysis on a purified HSC subset cultured in gene therapy conditions. To this end, scRNA-Seq of CD34$^+$ (CD19$^-$/CD34$^+$), ST-HSC (CD19$^-$/CD34$^+$/CD38$^-$/CD45RA$^-$/CD90$^-$/CD49f$^-$) and LT-HSC (CD19$^-$/CD34$^+$/CD38$^-$/CD45RA$^-$/CD90$^+$/CD49f$^+$) populations was performed on mPB from 4 independent healthy donors at 0 hr (immediately following single cell sorting) and after a 62 hr *ex vivo* gene therapy protocol. Cells were cultured in *ex vivo* gene therapy media (Biffi et al., 2013) and exposed to two hits of lentiviral transduction (Fig.4.1a). At the 62 hr time-point populations were harvested and live cells were sorted for scRNA-Seq. Lentiviral vectors were generated at GSK using clinical constructs in which the therapeutic transgene was replaced with *pGFP*, allowing for convenient single cell sorting of GFP$^+$ transduced populations. I then performed library preparation of the four donors using the Smart-Seq2 protocol (Picelli et al., 2014).

2069 mPB cells were sequenced and 68% of cells passed QC (508 CD34$^+$, 394 ST-HSCs and 516 LT-HSCs). UMAP visualisation shows two discrete cell clusters based on culture time-point, suggesting that the impact of culture duration is more prominent than the phenotypic separation of populations (Fig.4.1b). Next, differential gene expression was performed comparing 62 hr GFP$^+$ to 0 hr from each population by DESeq2 and demonstrated changes in over 800 genes within each subset (Fig.4.1c). Interestingly, many of these protocol-induced changes are shared between subsets with only 12% (255 genes) uniquely changed within the LT-HSC compartment (Fig.4.1d). Only one anti-correlated gene was observed going up in LT-HSCs and down in the other populations (*BRPF1, FDR = 0.0369*; data not shown).

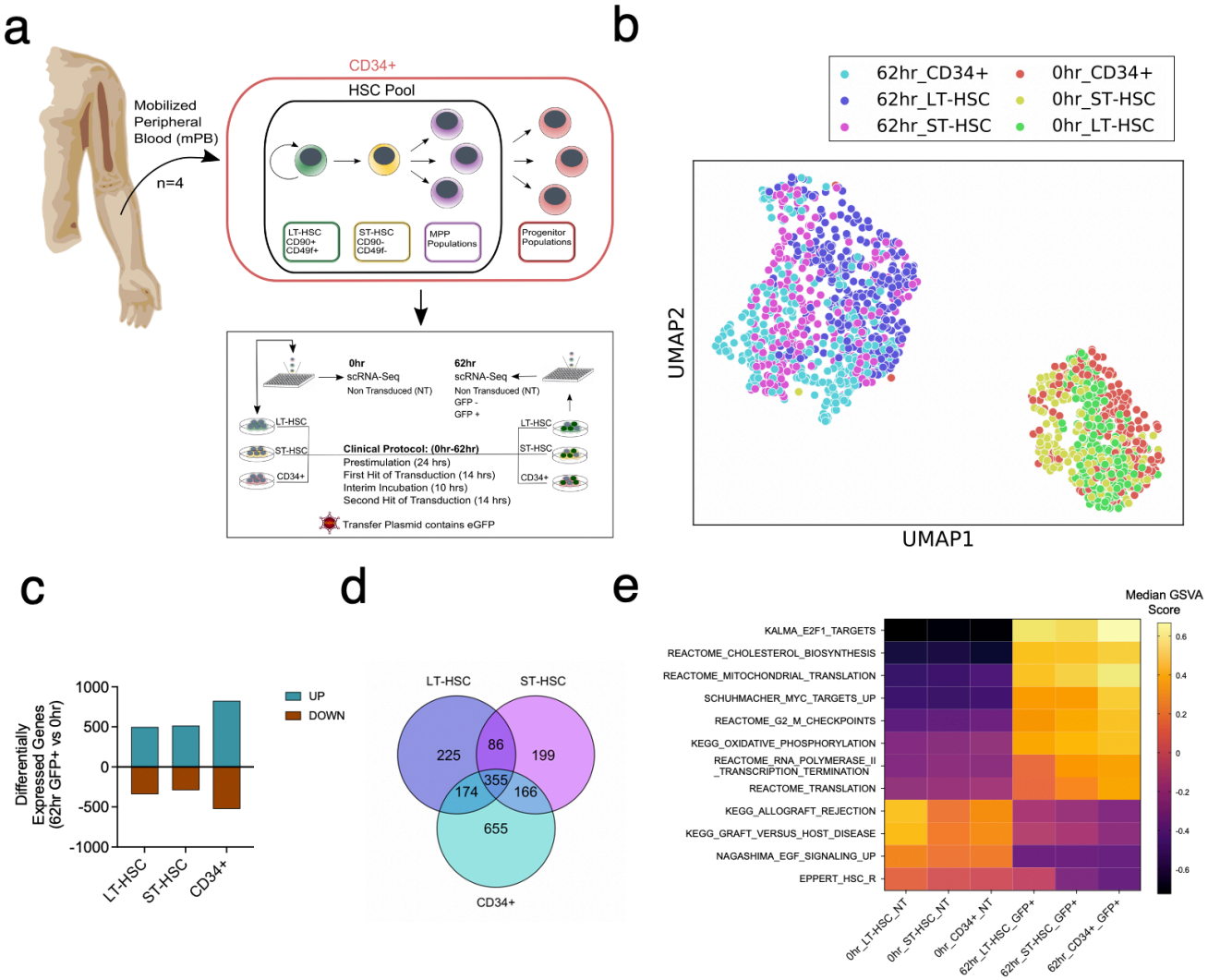

Figure 4.1: The 62 hr *ex vivo* gene therapy protocol causes shared changes to the transcriptome of HSC subsets and CD34⁺ cells.

a. scRNA-Seq workflow from mobilized peripheral blood donors before and after a gene therapy protocol consisting of 62 hr culture and transduction with a pGFP lentiviral vector. Live cells (Zombie⁻) were sorted at the 62 hr time-point. 2069 cells were sequenced and 67% of cells passed quality control.
b. UMAP visualization of 508 CD34⁺, 394 ST-HSC and 516 LT-HSCs at 0 hr and at 62 hr time-points. UMAP constructed using 1599 HVG.
c. Differentially expressed genes after transduction and culture in populations (62 hr GFP+ vs 0hr; *FDR<0.05*) using DESeq analysis.
d. Venn Diagram of common differentially expressed genes (62 hr GFP⁺ vs 0hr) across populations.
e. Heatmap of top differentially expressed c2 pathways by GSVA (62 hr GFP⁺ vs 0hr) in indicated subsets. Median GSVA score across population indicated.

Gene Set Variation Analysis (GSVA) (Hänzelmann et al., 2013) was then performed to better understand the biological pathways associated with differentially expressed genes following culture. GSVA calculates an enrichment score per pathway per cell. This allows for improved interrogation of pathway variation within single cells of a given time-point when compared to GSEA. Mirroring the results

at the gene level, over 95% of significantly enriched GSVA pathways *(FDR<0.01)* were common to all populations. The most significant commonly upregulated pathways include "Reactome: G_2M checkpoints", "Reactome: Mitochondrial translation" and "KEGG: Oxidative phosphorylation" indicating that cells are upregulating metabolic biosynthesis programs and entering mitosis (Fig.4.1e). Gene-sets commonly downregulated after the clinical protocol include "Nagashima: EGF signalling", which regulates HSC regeneration following myelosuppressive injury (Doan et al., 2013) and the "Eppert: HSC" signature generated from upregulated genes in CB HSCs compared to downstream progenitors (Eppert et al., 2011). This is accompanied by downregulation of the KEGG database gene-sets "Allograft rejection" and "Graft versus host disease" which, upon further inspection is driven by downregulation of MHC Class I and Class II genes/cofactors. Slight but significant shifts in the median GSVA score are observed between $CD34^+$, ST-HSC and LT-HSC populations respectively within time-points (Fig.4.1e). Compared to other cell types, LT-HSCs maintain higher expression of the "Eppert: HSC" gene-set and lower expression of "Schuhmacher: Myc targets", "KEGG: Oxidative phosphorylation" and "KEGG: Allograft rejection" at both time-points.

In summary, these data demonstrate that significant transcriptional change is acquired within the LT-HSC compartment during an *ex vivo* gene therapy protocol aimed to preserve long-term repopulation capacity. LT-HSCs change very similar biological pathways when compared to ST-HSCs and $CD34^+$ cells over this protocol, although the degree to which these pathways are changed varies between subsets. This broadly suggests that the overwhelming change from the extrinsic environment of *ex vivo* culture causes the induction of common biological processes.

4.1.2 LT-HSCs show a reduced amplitude of gene expression change over culture and a distinct regulatory network

I next sought to analyse differences in the regulatory framework and the amplitude of gene expression change in $CD34^+$, ST-HSC and LT-HSC populations *ex vivo*. LT-HSCs show a reduced fold change in both upregulated and downregulated genes compared to ST-HSCs and $CD34^+$ cells following the e*x vivo* gene therapy protocol (62 hr GFP^+ vs 0 hr NT) (Fig.4.2a). Whilst no interesting biological pathways could be observed from extraction of genes uniquely changed in each of the subsets (Fig.4.1d), iRegulon transcription factor target analysis demonstrates that subsets regulate different transcriptional networks over culture (Fig.4.2b).

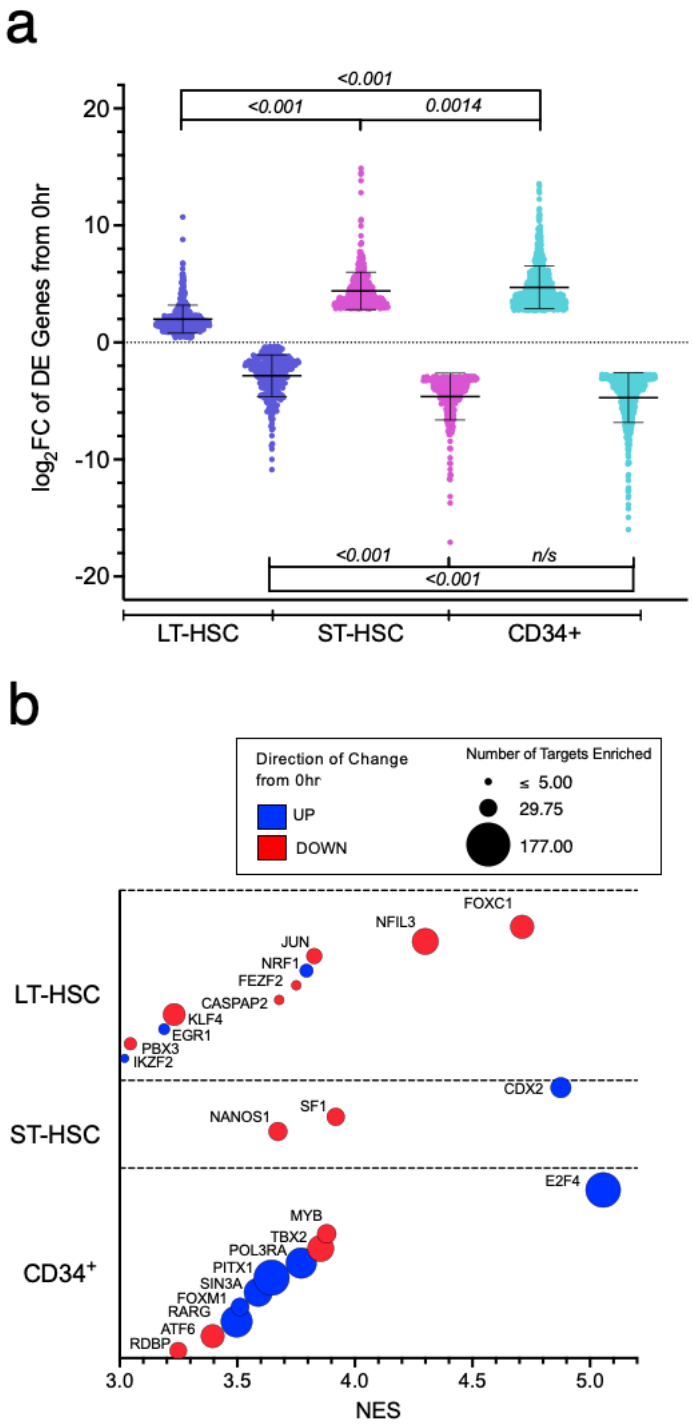

Figure 4.2: LT-HSCs show a reduced amplitude of gene expression change and a unique transcription factor regulatory network over culture.

a. $\log_2$ fold change of gene expression in populations exposed to gene therapy culture conditions and lentiviral transduction (62 hr GFP+ vs 0 hr NT). Line at mean and S.D. shown. Unpaired t-test performed and significant values indicated *(p<0.05).*

b. iRegulon analysis of enriched transcription factor motifs from genes uniquely changed in respective subsets over culture. Only enriched transcription factor networks which are unique to each subset are-shown *(FDR<0.01).* Direction of change over culture indicated by colour of datapoints. Normalised Enrichment Score (NES) calculated by iRegulon. Size of datapoints indicate the number of enriched genes involved in the transcription factor regulatory network.

255 genes uniquely changed in LT-HSCs over the gene therapy protocol (126 downregulated and 99 upregulated; *FDR<0.05*) showed enrichment to 10 transcription factor motifs which were not observed in other subsets (Fig.4.2b). The master regulators of stem cell function, JUN (Li et al., 2019) and KLF4 (Guo et al., 2009) transcription factor networks are exclusively downregulated in LT-HSCs following culture and transduction. The NRF1 network, which acts as a checkpoint regulator of the mitochondrial UPR upon metabolic activation (Mohrin et al., 2015) is also exclusively upregulated in LT-HSCs, suggesting unique mechanisms of mitochondrial activation upon quiescence exit. ST-HSCs show unique upregulation of the caudal homeobox transcription factor network CDX2 which has a critical role in developmental haematopoiesis (El Omar et al., 2021). $CD34^+$ cells show unique upregulation of POL3RA and E2F4 networks which are involved in transcription regulation and cell cycle progression respectively. These data imply that LT-HSCs have a differential regulatory network in cell cycle and metabolic processes over culture when compared to other HSPC subsets.

4.1.3 LT-HSCs show differential regulation of antigen presentation and processing markers after culture

To further disentangle differences in LT-HSC behaviour after the *ex vivo* gene therapy protocol, LT-HSCs were compared to ST-HSCs and $CD34^+$ cells at 0 hr and at 62 hr (GFP^+ populations) respectively. Despite clear functional differences as previously reported in CB (Laurenti et al., 2013; Notta et al., 2011), uncultured mPB LT-HSCs are strikingly transcriptionally similar to ST-HSCs and $CD34^+$ cells (Fig.4.3a). At 62 hr, LT-HSCs become transcriptionally more distinct from ST-HSCs and $CD34^+$ cells (Fig.4.3b). Cluego biological pathway analysis (Bindea et al., 2009) was then performed on genes upregulated in LT-HSCs with respect to both ST-HSCs and $CD34^+$ cells and this was performed at both time-points. Analysis showed significant enrichment at 0 hr to "Reactome: MHC class II antigen presentation" *(padj<0.01)* (including enriched *CD74* and *HLA-DRA*) and analysis at 62 hr showed enrichment to "Reactome: MHC Class I antigen presentation" (including the Beta-2 microglobulin component of MHC Class I molecules *B2M*, in addition to *HLA-A, HLA-E* and *HLA-G*) (Fig.4.3c).

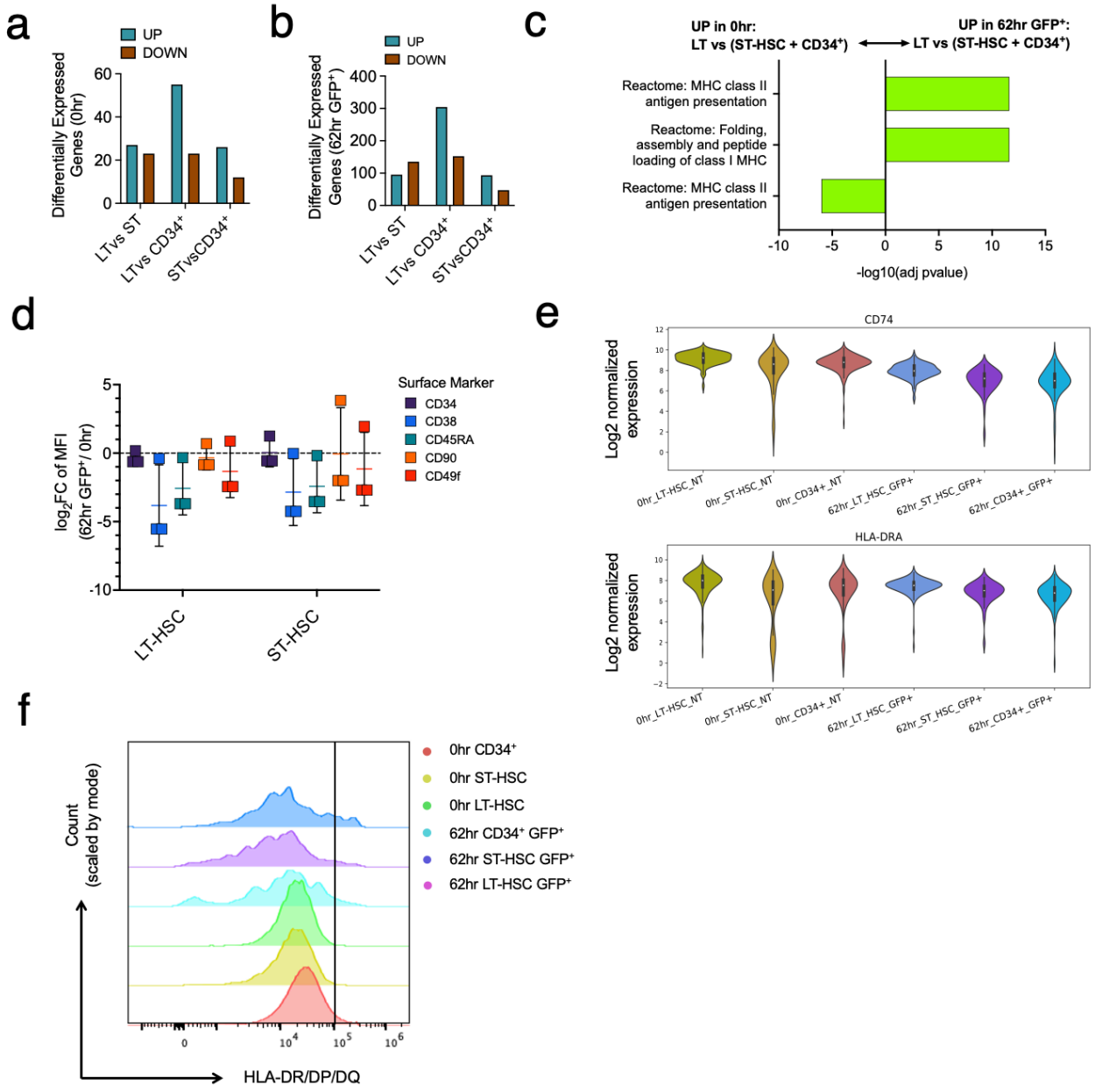

Figure 4.3: LT-HSCs maintain higher expression of antigen processing and presentation markers over culture whilst changing expression of phenotypic LT-HSC cell surface markers.

a. -b. Number of differentially expressed genes *(FDR < 0.05)* within 0 hr (a) 62 hr (b) subsets.

c. Cluego pathway analysis of significantly enriched Reactome pathways *(padj < 0.01)* at indicated time-points. Cluego pathway analysis performed separately at both time-points.

d. log_2fold change (log_2fc) of cell surface marker median fluorescence intensity (MFI) after culture (62 hr GFP$^+$/ 0hr) in LT-HSCs and ST-HSCs.

e. Normalised expression of *CD74* (upper) and *HLA-DRA* (lower) in subsets before (0 hr) and after the 62 hr gene therapy protocol (62 hr GFP$^+$ populations shown).

f. HLA-DR/DP/DQ cell surface expression in subsets before (0 hr) and after the 62 hr gene therapy protocol (62 hr GFP$^+$ populations shown; representative example of n=3 biological repeats).

To assess how phenotypic markers used to isolate HSC populations change following culture, I analysed index data acquired from four mPB scRNA-Seq experiments upon cell sorting at both time-points (0 hr and 62 hr GFP^+). Cell surface expression of phenotypic markers used to isolate HSCs are altered during this protocol (Fig.4.3d), and the field is lacking reliable LT-HSC cell surface markers after culture. I therefore hypothesized that candidate cell surface markers may be found from MHC regulators which show higher expression in LT-HSCs compared to ST-HSCs and $CD34^+$ cells at 62 hr. 8 genes were determined to be significantly higher in LT-HSCs compared to ST-HSCs and $CD34^+$ cells at both 0 hr and 62 hr. Further inspection showed seven out of eight genes were involved in MHC Class II antigen processing and presentation, including the alpha chain of HLA-DR (*HLA-DRA)* and the MHC class II regulator *CD74* (Fig.4.3e). These data demonstrate that LT-HSCs show differential expression of genes involved in antigen presentation and processing which they modify throughout culture.

I next sought to investigate whether these genes are expressed at the cell surface level through antibody staining of: CD74, HLA-DRA and a broad MHC Class II antibody (encompassing HLA-DR/DP/DQ). CD74 cell surface expression in cultured LT-HSCs proved unreliable over repeat experiments (data not shown), perhaps attributed to the multifaceted role of this protein as both an intracellular MHC Class II chaperone protein and a cell surface receptor (reviewed in Schroder, 2016). Despite this shortcoming, a proportion of 62 hr GFP^+ LT-HSCs exhibited a higher median fluorescence intensity (MFI) in HLA-DR/DP/DQ cell surface expression (Fig.4.3f) with respect to ST-HSCs and $CD34^+$ cells (representative example shown; n=3 biological repeat experiments). Although these results proved interesting it was not deemed feasible to perform long-term transplants of the MHC II hi/lo fractions for several reasons. Firstly, the HLA locus is the most polymorphic region of the human genome with variations across ethnicities (Lim et al., 2019), therefore variability in MHC Class II expression is expected within and across populations. Secondly, MHC Class II expression variability has been shown to impact susceptibility to many immune disorders (Kim et al., 2017; Wang et al., 2009). Therefore, HLA protein expression may not represent a reliable cell surface in the context of clinical research. Finally, the HLA-DR/DP/DQ antibody encompasses all MHC Class II expression which may produce a heterogenous response given the diversity of potential HLA-DR/DP/DQ protein combinations. In addition, isolation of the required cell dose from cultured mPB LT-HSC MHC II hi/lo fractions to produce long-term stable grafts *in vivo* would involve initial cell sorting (0 hr) of a very high number of LT-HSCs. Taken together, this approach would significantly deplete sample availability with the added complexity of a potentially heterogenous response.

4.1.4 Transduction alone exerts a minimal effect on the transcriptome of CD34$^+$, ST-HSC and LT-HSC populations.

Results generated show that the *ex vivo* manipulation protocol has a considerable effect on the HSC transcriptome. To discern the effect of transduction alone, GFP$^+$ populations were compared to non-transduced (NT) populations at the 62 hr time-point. The GFP$^-$ population was not compared after it was found that a significant proportion of GFP$^-$ cells at 62 hr later upregulate GFP to become GFP$^+$ in the context of single cell differentiation assays (data not shown). This is perhaps attributed to the second hit of transduction occurring 14 hr prior to cell harvest. Lentiviral vectors of ≥ 4.44E+08 TU/ml were used for transduction with a trend for higher transduction efficiency observed in LT-HSCs when compared to CD34$^+$ cells (Fig.4.4a). Transduction alone caused a minimal effect on the transcriptome and cells do not cluster based on GFP status at the 62 hr time-point (Fig.4.4b). The CD34$^+$ compartment showed the highest number of differentially expressed genes acquired from lentiviral transduction (Fig.4.4c). To note, this number is ~20 fold reduced when compared to the number of differentially expressed genes acquired from the combined effect of culture and transduction. The only significantly enriched biological pathway determined from differentially expressed genes (62 hr GFP$^+$ vs 62 hr NT) is the GO term "Intracellular transport of virus" (*padj = 0.001*) observed in CD34$^+$ GFP$^+$ cells (Fig.4.4d). Enriched genes include Exportin *XPO1*, and the viral sensor *BST2*.

One gene that is commonly upregulated in all GFP$^+$ populations relative to NT is *S100A7A* which encodes the anti-microbial peptide koebnerisin. *S100A7A* shows successively higher expression levels in NT, GFP$^-$ and GFP$^+$ subsets respectively across all populations (Fig.4.4e). Koebnerisin was originally identified from psoriatic skin (Wolf et al., 2003) and is overexpressed in a variety of bacterial and viral infections (Gläser et al., 2009; Jansen et al., 2013; Wolf et al., 2008). Reliably increased expression in all subsets may suggest the potential to use of *S100A7A* as a biomarker of transduction, although lack of cell surface expression does not allow for convenient flow cytometry isolation of populations. Overall, these data broadly suggest that lentiviral transduction causes a minor increase in cell stress and inflammatory related genes across all subsets with no significant disruption to homeostatic function. Importantly, it should be considered that changes acquired as a result of the lentiviral vector in this study are reduced from that expected if a therapeutic transgene were incorporated, where corrected target gene expression would modulate downstream signalling pathways and have pleiotropic effects on mature blood cell production (Aiuti et al., 2013; Biffi et al., 2013; Mamcarz et al., 2019).

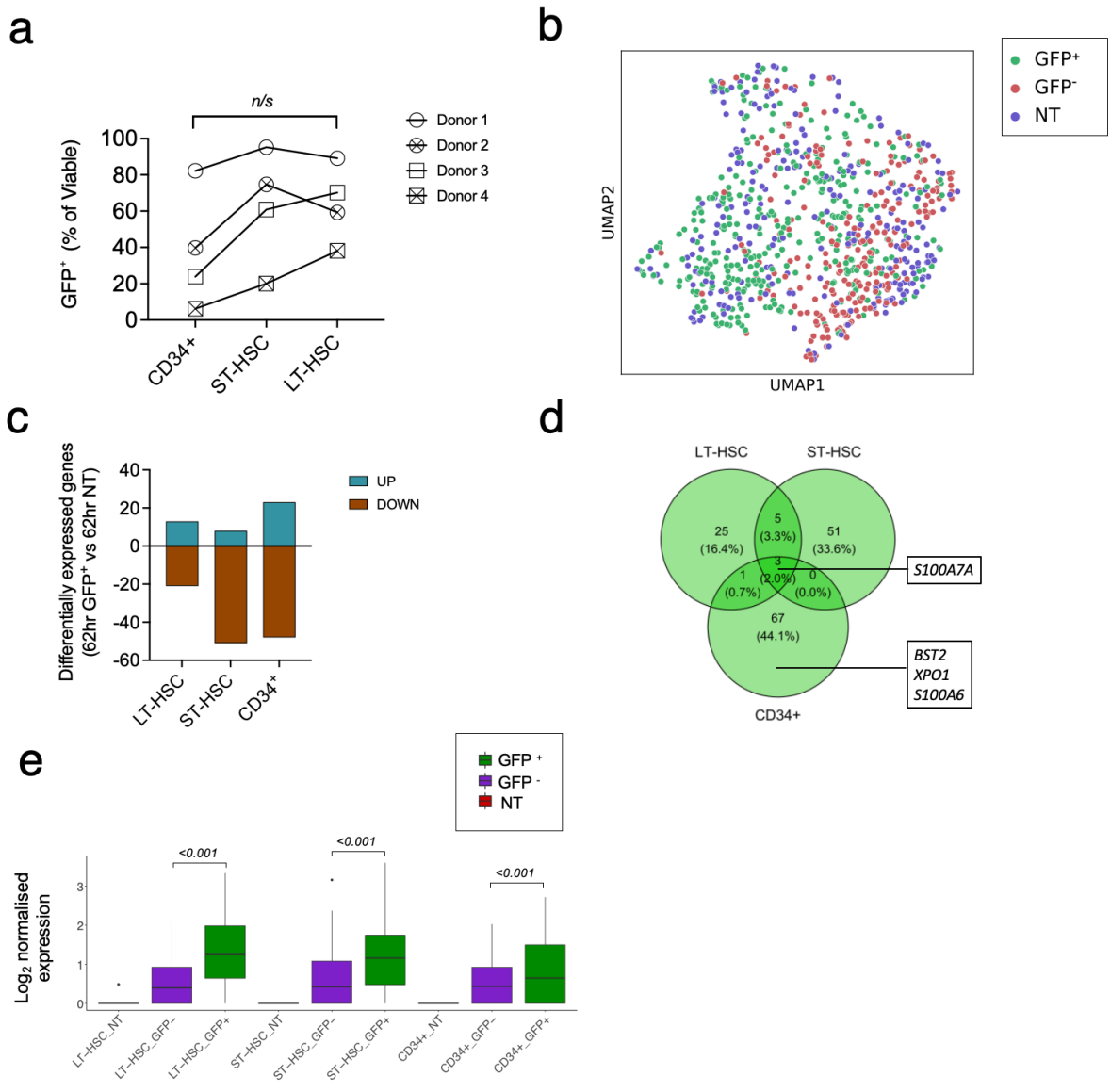

Figure 4.4: Transduction alone has a minimal effect on the transcriptome.

a. Lentiviral transduction efficiency of populations at the 62 hr time-point (GFP⁺% of Viable cells). Paired t-test performed (no significant values generated; CD34⁺ cells vs LT-HSCs $p = 0.053$).

b. UMAP visualization of 62 hr NT, GFP⁻ and GFP⁺ and cells generated based on 820 cells. UMAP constructed using 1599 HVG.

c. Number of differentially expressed genes ($FDR < 0.05$) acquired from lentiviral transduction (62 hr GFP⁺ vs 62 hr NT).

d. Venn Diagram illustrating shared differentially expressed genes between subsets (62 hr GFP⁺ vs 62 hr NT).

e. Barplot of $\log_2$ normalised expression of *S100A7A* in 62 hr CD34⁺, ST-HSC and LT-HSCs. Line at median. Upper and lower whiskers indicate the 25th and 75th percentile of expression. ($FDR < 0.01$ for all 62 hr GFP⁺ vs 62 hr GFP⁻ populations).

4.1.5 Cultured LT-HSCs remaining in the G_1 phase of the cell cycle show enrichment of HSC signatures in contrast to cycling S-G_2M cells

One of the most striking transcriptional changes observed across all subsets is the upregulation of cell cycle progression pathways. In fact, the top differentially expressed GSVA pathway between 62 hr GFP^+ cells relative to 0 hr cells across all populations is "KALMA_E2F1 targets" (Fig.4.1e). It is well appreciated that progressive HSC culture leads to a loss of long-term HSC repopulation capacity, however a subset of LT-HSCs remain after the *ex vivo* gene therapy protocol, as evidenced by successful long-term haematopoietic reconstitution in gene therapy treated patients (reviewed in Sagoo and Gaspar, 202 1). Given these findings, I next wanted to establish whether 62 hr LT-HSCs in different stages of the cell cycle display transcriptional differences which may indicate that cell cycle status drives functional heterogeneity within the cultured LT-HSC population. Transcriptome-based cell cycle allocation was thus determined for all cells in the dataset using a panel of 97 marker genes (Tirosh et al., 2016) within the Satija group package (Methods 3.8.2). As expected, $CD34^+$, ST-HSC and LT-HSC populations are assigned to later phases of the cell cycle at 62 hr (Fig.4.5a+b).

Recently, the Cabezas-Wallscheid group (Frieburg, Germany) identified a cell surface marker that enriches for dormant human LT-HSCs and I helped contribute to this work by performing single cell differentiation assays using this population (Zhang et al., Nature Cell Biology*, in revision*). The team kindly shared an RNA-Seq signature generated from this population with our group. I was next interested in determining how the *ex vivo* culture process impacts this signature because dormant *in vivo* HSCs have been conclusively shown to retain the majority of the self-renewal capacity within the HSC pool (Bernitz et al., 2016; Foudi et al., 2009; Wilson et al., 2008). We have determined that cells are allocated to later phases of the cell cycle following 62 hr culture (Fig.4.5a+b) and therefore, this may lead to lower enrichment of 62 hr mPB cells to the dormant HSC signature driven solely by cell cycle related genes. To circumvent this issue, two comprehensive cell cycle gene-sets ("Reactome: Cell_Cycle_Checkpoints" combined with "Reactome: Mitotic_Cell_Cycle"; 613 non overlapping genes total) were removed from the dormant HSC signature and from the mPB scRNA-Seq dataset. In absence of cell cycle related genes, LT-HSCs show a drastic decrease in expression of this signature (Fig.4.5c) after the 62 hr protocol. This indicates that following culture LT-HSCs are becoming transcriptionally more distinct to a population of highly enriched LT-HSCs and that this is not due to cell cycle related changes.

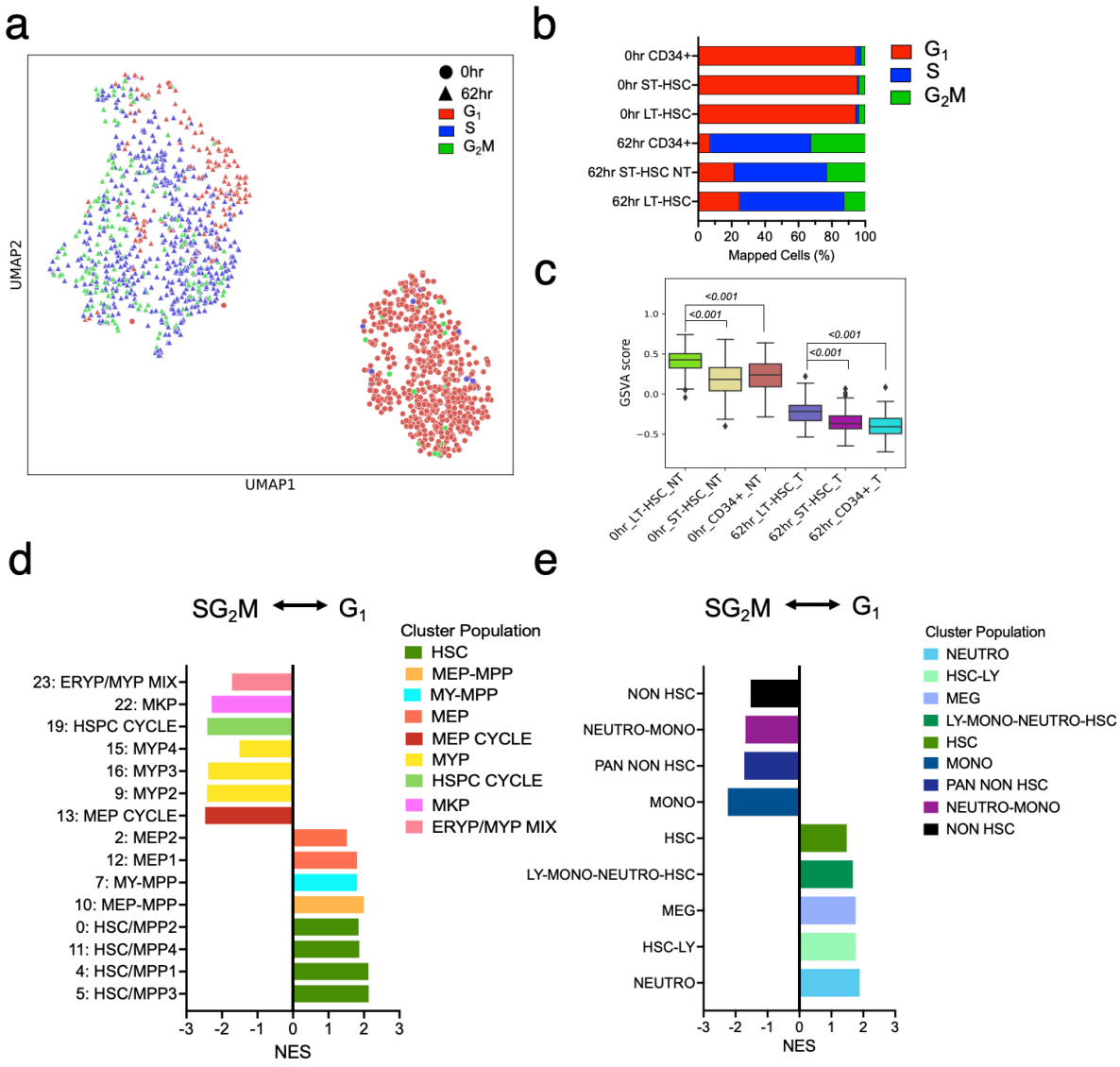

Figure 4.5: 62 hr mPB LT-HSCs in the G_1 phase of the cell cycle enrich to HSC signatures in contrast to cycling (S-G_2M phase) 62 hr LT-HSCs

a. UMAP visualization of 508 CD34$^+$, 394 ST-HSC and 516 LT-HSCs at 0 hr and following a 62 hr *ex vivo* lentiviral gene therapy protocol. UMAP constructed using 1599 HVG. Cells coloured by transcriptional cell cycle phase assignment.

b. Barplot quantification of cell cycle scores in (a)

c. GSVA score of human HSC dormancy signature (generated from Cabezas-Wallscheid group) for mPB LT-HSC, ST-HSC and CD34$^+$ populations at 0 hr and 62 hr. Manually curated cell cycle gene-set (613 genes) removed from mPB scRNA-Seq dataset and from HSC dormancy signature. Line at median. Upper and lower whiskers indicate the 25th and 75th percentile of expression.

d. GSEA enrichment of G_1 and S-G_2M phase 62 hr mPB LT-HSCs (only values *FDR<0.05* shown) to multi-tissue landscape of differentiation (Mende et al., 2022). Manually curated cell cycle gene-set (613 genes) removed from mPB scRNA-Seq dataset and differentiation signatures.

e. GSEA enrichment of G_1 and S-G_2M phase 62 hr LT-HSCs (only values *FDR<0.05* shown) to landscape of BM differentiation using signatures for HSCs and progenitor populations. Velten et al., 2017; supplementary table 4). Manually curated cell cycle gene-set (613 genes) removed from mPB scRNA-Seq dataset and differentiation signatures.

Next, I sought to determine whether LT-HSCs in the G_1 phase of cell cycle at 62 hr showed transcriptional differences in signatures of HSC differentiation when compared to 62 hr LT-HSCs in S-G_2M phases. Signatures were first generated from a multi-tissue landscape of haematopoietic differentiation produced by our lab (methods 3.8.4; Fig.3.4) combining BM, spleen and non-mobilised PB (10 adult donors, 118783 $CD19^-CD34^+$ single cells) (Mende et al., 2022) and enrichment was performed using GSEA. It is known that progenitors cycle more frequently than HSCs, therefore to mitigate potential enrichment being driven by cell cycle genes, the manually curated cell cycle gene-set (613 genes) was removed from both the mPB scRNA-Seq dataset and signatures generated from Mende et al., 2022.

G_1 phase mPB LT-HSCs showed significant enrichment *(FDR<0.05)* to HSC/MPP clusters in addition to erythroid-primed multipotent progenitors and megakaryocytic-erythroid progenitors (MEP-MPP and MEP) and my-biased multipotent progenitors (MY-MPP) (Fig.4.5d). Cycling (S-G_2M phase) mPB LT-HSCs showed enrichment to myeloid progenitors (MYP) and cycling erythroid progenitors amongst others (Fig.4.5d). Results were confirmed by repeating this methodology using a single cell landscape of BM differentiation (Velten et al., 2017). Similarly, cell cycle related genes (613 genes) were removed from Velten et al. signatures and the mPB scRNA-Seq dataset. GSEA was then performed to compare 62 hr LT-HSCs in the G_1 and S-G_2-M phases of the cell cycle. G_1 LT-HSCs again showed enrichment to BM HSC signatures whereas S-G_2M cells showed enrichment to downstream signatures for monocyte and neutrophil progenitors *(FDR<0.05)* (Fig.4.5e). These results collectively suggest that cultured G_1 phase LT-HSCs exhibit a transcriptional profile more similar to primitive cells with a higher stem cell capacity. In contrast, cycling cells acquire myeloid progenitor transcriptomic traits.

4.1.6 Chapter 1 Summary:

Overall, data obtained in this chapter shows that a clinically approved gene therapy protocol which aims to maintain HSC function *ex vivo* causes marked transcriptional change in mPB $CD34^+$, ST-HSC and LT-HSC populations. Transcriptional changes acquired are predominantly attributed to the culture process rather than lentiviral transduction. Culture induced transcriptional changes are largely conserved between LT-HSCs, ST-HSCs and $CD34^+$ cells. However, the amplitude of gene expression change in LT-HSCs is attenuated compared to other populations and transcriptional change occurs via a distinct transcription factor regulatory framework.

After culture, LT-HSCs show downregulation of a transcriptional signature generated from an enriched population of human dormant LT-HSCs (Zhang et al., Nature cell biology*, in revision*) and culture results in allocation to later phases of the cell cycle. Finally, LT-HSCs remaining in the G_1 phase of the cell cycle at 62 hr show enrichment of HSC signatures as opposed to cycling S-G_2M phase cells which instead enrich to downstream progenitor signatures.

Taken together, the loss of LT-HSC transcriptional identity over culture may be linked to culture duration and progression through the cell cycle. It is apparent from this dataset that LT-HSCs show unique regulatory mechanisms in culture and LT-HSCs should be studied in further detail to better understand loss of long-term repopulation capacity *ex vivo*. Also, given the degree of transcriptional change acquired and the importance of cell cycle status in clustering, it would be useful to dissect changes in a time-course nature based on time-points relevant to cell cycle progression stages. Therefore, in subsequent chapters I will focus on understanding the impact of culture in a more precise, temporal manner at single cell resolution.

4.2 Chapter 2: Defining the transcriptional and functional dynamics of single HSC activation over an entire cell cycle

4.2.1 Quiescence exit occurs at approximately 24 hr in CB and mPB LT-HSC culture independently of culture conditions

Results obtained thus far demonstrate that there may be value in dissecting HSC activation during culture at time-points meaningful to cell cycle progression. The cell cycle kinetics of mPB LT-HSCs and CB LT-HSCs were first characterised in contrasting media compositions. mPB LT-HSCs were cultured in *ex vivo* gene therapy media (no lentiviral transduction performed) aimed to maintain HSC repopulation capacity (Biffi et al., 2013), whereas CB LT-HSCs were cultured in differentiation facilitating conditions aimed to promote cell cycle progression (Laurenti et al., 2015). mPB LT-HSCs show a gradual increase in the proportion of cells having completed quiescence exit over the 6-62 hr culture window, by which time 89% of cells have hyperphosphorylated Rb (Fig.4.6a) marking commitment to cell division. Concurrently, the proportion of cycling cells (S-G_2M phases) as determined by DAPI/pRb staining is significantly increased within the 24-62 hr window (Fig.4.6b) (*p=0.0357*). This is supported by an average time to first division in single sorted mPB LT-HSCs of 67.09 hr (Fig.4.6c+h).

CB LT-HSCs cultured in differentiation facilitating conditions also show an accumulation of hyperphosphorylated Rb over the 6-48 hr window (*p=0.0276*) (Fig.4.6d). At 48 hr of culture, 97% of LT-HSCs have passed the restriction point and 38% of these cells are actively cycling (S-G_2M phase) (Fig.4.6e). The average time to first division of single sorted CB LT-HSCs in these conditions is 60.96 hr (Fig.4.6f+h). Of interest, at matched time-points of 0, 6 and 24 hr of culture the proportion of mPB and CB LT-HSCs having hyperphosphorylated Rb is very similar (Fig.6a+d) and interpolation of this data (Sigmoidal regression analysis R^2>0.97) indicates that both mPB and CB LT-HSCs exit quiescence at approximately 24 hr (Fig.4.6g). However, CB LT-HSCs show a trend for a lower time to first division (*p=0.0589*) (Fig.4.6c+h). These data may suggest that the difference in division kinetics is attributed to a faster cycling time through the late G_1-M cell cycle phases. From these experiments we cannot conclude whether the trend for a longer time to first division in mPB LT-HSCs compared to CB LT-HSCs is attributed to the difference in media composition, stem cell source or a combination of these factors.

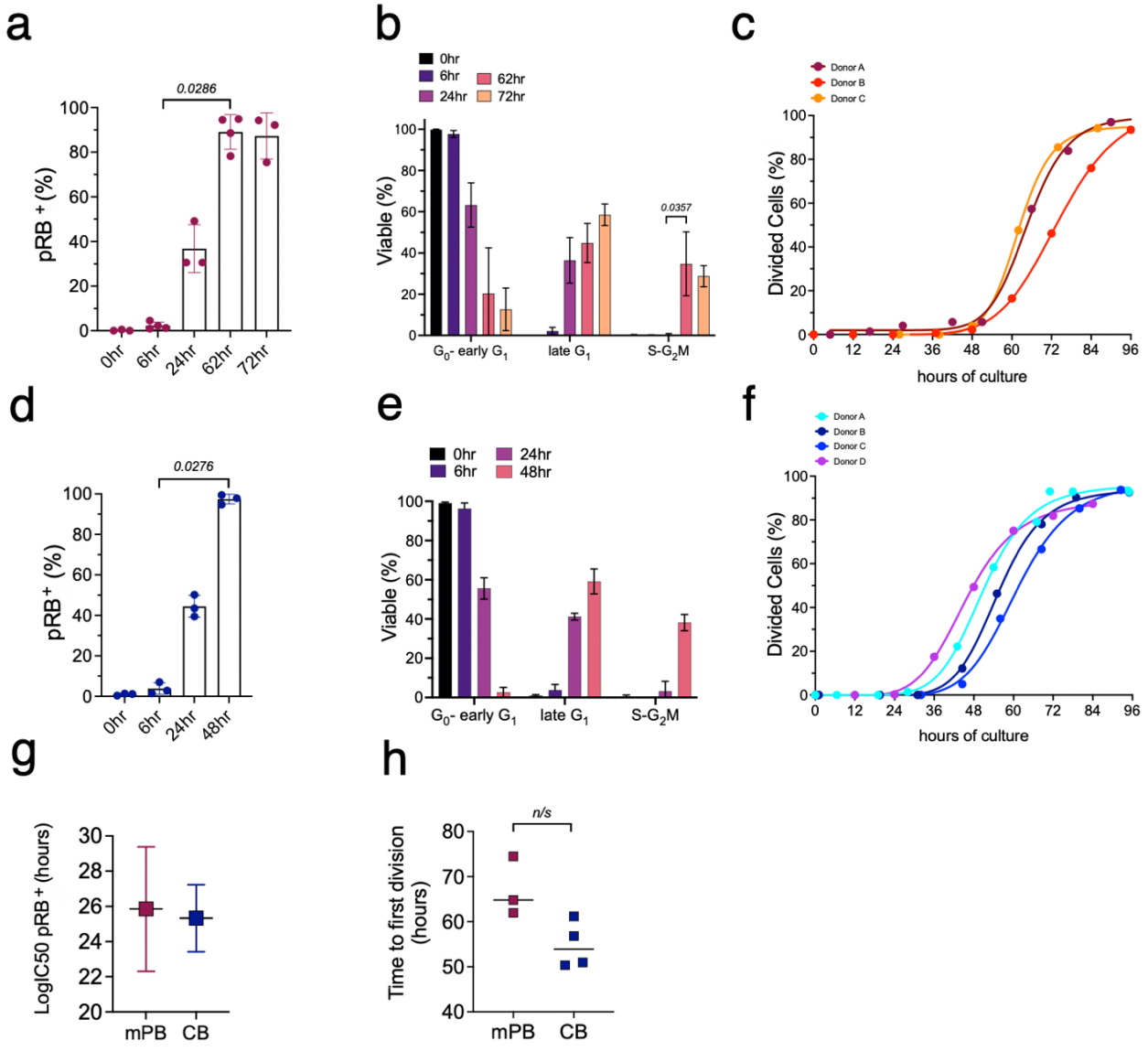

Figure 4.6: Kinetics of mPB and CB LT-HSC quiescence exit and cell cycle progression.

a. pRb in bulk mPB LT-HSC populations cultured in gene therapy conditions (Mann Whitney U-test performed and significant values indicated), n=3 or n=4 independent mPB samples.
b. Cell cycle status as determined by pRb/DAPI staining in bulk mPB LT-HSCs.
c. Cumulative time to first division kinetics of single sorted mPB LT-HSCs (n=3 independent mPB samples). Curve shows sigmoidal regression analysis (R^2 >0.99).
d. pRb staining in CB LT-HSCs cultured in differentiation facilitating conditions (Mann Whitney U-test performed and significant values indicated).
e. Cell cycle status as determined by pRb/DAPI staining in bulk CB LT-HSCs.
f. Cumulative time to first division kinetics of single sorted CB LT-HSCs (n=4 independent CB samples). Curve shows sigmoidal regression analysis (R^2 >0.99).
g. $LogIC_{50}$ of quiescence exit time interpolated from Sigmoidal regression analysis in mPB (a) and CB (d).(mPB R^2 =0.977; CB R^2 =0.995). Upper and lower confidence limits indicated (95%).
h. Mean time to first division of CB and mPB LT-HSCs (CB = 60.96 hr; mPB = 67.09 hr) Unpaired t-test performed (*p=0.0587*).

CB LT-HSCs have a delayed time to first division in culture with respect to ST-HSCs which is attributed to a longer duration of quiescence exit (Laurenti et al., 2015). mPB donor samples have been treated with the mobilisation agent G-CSF which has been shown to induce proliferation in the HSC compartment (Morrison et al., 1997; Wright et al., 2001), although recent data in mouse suggests that G-CSF mobilizes dormant LT-HSCs without proliferation (Bernitz et al., 2017). I was therefore interested in determining whether G-CSF alters the delayed kinetics of LT-HSCs with respect to ST-HSCs in culture.

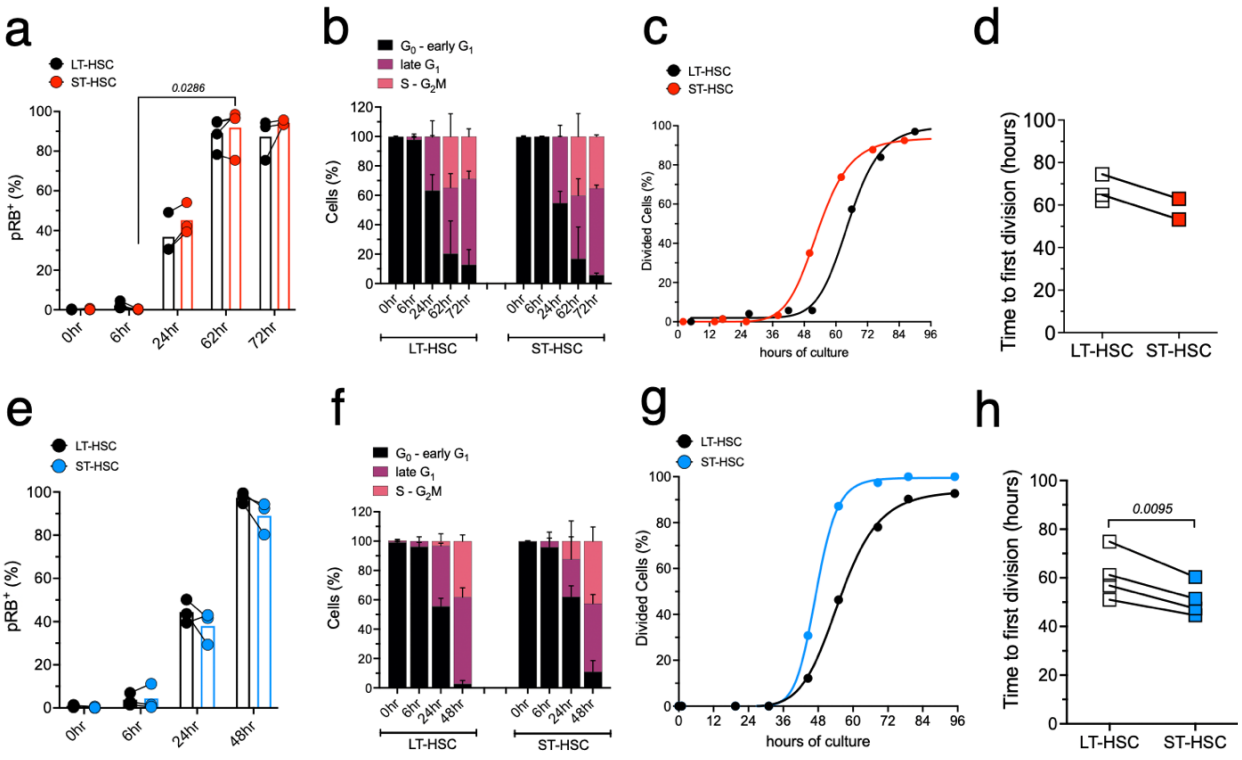

Figure 4.7: Delayed kinetics of LT-HSCs compared to ST-HSCs are comparable in CB and mPB irrespective of media.

a. pRb staining of matched mPB LT-HSC and ST-HSC bulk populations (n=3 matched donor samples). Mann Whitney U-test performed on ST-HSCs and significant values indicated.
b. Cell cycle status of mPB LT-HSC and ST-HSC bulk populations as determined by pRb/DAPI staining (n=3 matched donor samples).
c. Cumulative time to first division kinetics of single sorted mPB LT-HSCs and ST-HSCs (n=2 matched donor samples). Curve shows representative example. Sigmoidal regression analysis (R^2 >0.99).
d. Mean time to first division in single sorted mPB LT-HSCs and ST-HSCs (n=2 matched donor samples; n=1 unmatched LT-HSC repeat).
e. pRb staining of CB LT-HSC and ST-HSC bulk populations (n=3 matched donor samples). Mann Whitney U-test performed on ST-HSCs at time-point comparisons, no significant values determined.
f. Cell cycle status of CB LT-HSC and ST-HSC bulk populations as determined by pRb/DAPI staining (n=3 matched donor samples).
g. Cumulative time to first division kinetics of single sorted CB LT-HSC and ST-HSCs. Curve shows representative example. Sigmoidal regression analysis (R^2 >0.99).
h. Mean time to first division in single sorted CB LT-HSCs and ST-HSCs (n=4 matched donor samples). Paired t-test performed and significant values indicated.

Of note, experiments involving LT-HSCs were performed at the same time as those in Fig.4.6. mPB ST-HSCs showed similar levels of hyperphosphorylated Rb and a higher proportion of cells in later phases of the cell cycle compared to LT-HSCs at matched time-points (Fig.4.7a+b). mPB ST-HSCs showed a reduced time to first division of approximately 9 hr (n=2 matched biological repeats) compared to LT-HSCs (Fig.4.7c+d). The increased proportion of ST-HSCs in later phases of the cell cycle compared to LT-HSCs was also observed in CB (Fig.4.7e+f) with a similar reduction in time to first division of 10 hr (n=4 matched repeat experiments) (Fig.4.7g+h). These data demonstrate that G-CSF treatment is not able to accelerate the difference in quiescence exit time between the LT-HSC and ST-HSC compartments over *ex vivo* culture, perhaps suggesting that this difference is intrinsic or that LT-HSCs are indeed more "resistant" to G-CSF induced cell cycling.

Collectively, these results delineate the quiescence exit and cell division kinetics of mPB and CB HSC subsets cultured in contrasting conditions. LT-HSC quiescence exit in CB and mPB cultured across both conditions occurs at approximately 24 hr of culture. Cell division occurs at approximately 61 hr for CB LT-HSCs cultured in differentiation facilitating conditions and 67 hr for mPB LT-HSCs cultured in gene therapy media. Finally, ST-HSCs show a similar reduction in time to first division compared to LT-HSCs in both CB and mPB HSC sources.

4.2.2 A sharp decline in CB LT-HSC long-term repopulation capacity between 6 and 24 hr of culture

Understanding the kinetics of self-renewal loss in culture is important to drive improvements to HSC *ex vivo* protocols for therapeutic use. Few studies have transplanted $CD34^+$ enriched populations after successive time-points in culture (Ferrari et al., 2020; Gan et al., 1997; Kennedy et al., 2009; Zonari et al., 2017) however, to our knowledge no study to date has investigated this question using a purified population of LT-HSCs, particularly at early time-points $\leq$ 24 hr of culture.

To this end, we set out to better refine the kinetics of long-term repopulation capacity loss over clinically relevant culture. First, we performed limited dilution analysis (LDA) transplantation experiments using CB LT-HSCs cultured in differentiation facilitating conditions for 0, 6, 24 and 72 hr (Fig.4.8a). These time-points were chosen to represent distinct cell cycle phases over these culture conditions (Fig.4.6). At indicated time-points, three doses of CB LT-HSCs were transplanted by intrafemoral injection into NSG animals. Of note, the cell dose is representative of the 0 hr time-point (cells were not re-counted before injection) (Table 4.1; Fig.4.8a).

Hours of culture	Cell Dose	Number of tested mice	Number of engrafted mice
0hr	100	4	4
0hr	25	6	6
0hr	5	10	4
6hr	100	4	4
6hr	25	3	2
6hr	5	4	3
24hr	300	12	12
24hr	50	12	4
24hr	10	8	0
72hr	700	14	13
72hr	300	12	9
72hr	50	13	4

Table 4.1: Human CB LT-HSC cell dose and mouse information for primary *in vivo* transplantation studies in differentiation facilitating media. Human CB LT-HSC cell dose (representative of 0 hr at the time of sort), the number of mice tested and the number of mice showing human engraftment 18 weeks post transplantation are indicated. Mice were considered engrafted if human cells ($hCD45^{++}$ &$GlyA^{+}$) $\geq$ 0.01% of Singlets and $\geq$ 30 cells present in any lineage determination gate. 0hr: 20 mice, n=2 experiments; 6hr: 11 mice, n=2 experiments; 24 hr 32 mice n=3 experiments; 72 hr 39 mice, n=5 experiments).

Human engraftment was read out by flow-cytometry at 18 weeks post transplantation. Mice were considered engrafted if human cells ($hCD45^{++}$ & $GlyA^{+}$) $\geq$ 0.01% of Singlets and if $\geq$ 30 cells were present in any lineage determination gate. Limiting dilution analysis showed that long-term repopulation capacity is unchanged (*p=0.652)* for the first 6 hr of culture such that the HSC frequency is determined as ≈ 1 in 8.1 at 0 hr and ≈1 in 10.57 at 6 hr of culture. Strikingly, between 6 to 24 hr in culture the HSC frequency of the transplanted population drops drastically by 10 fold (≈1 in 98.3 at 24 hr; Fig4.8b), corresponding to the time cells have completed quiescence exit. An additional decrease in HSC frequency is observed between 24 - 72 hr (≈ 1 in 217 at 72 hr), when cells are progressing through the cell cycle and entering mitosis, although to a much lower degree (2 fold reduction in %LTRC from 24 – 72 hr) than at earlier time-points.

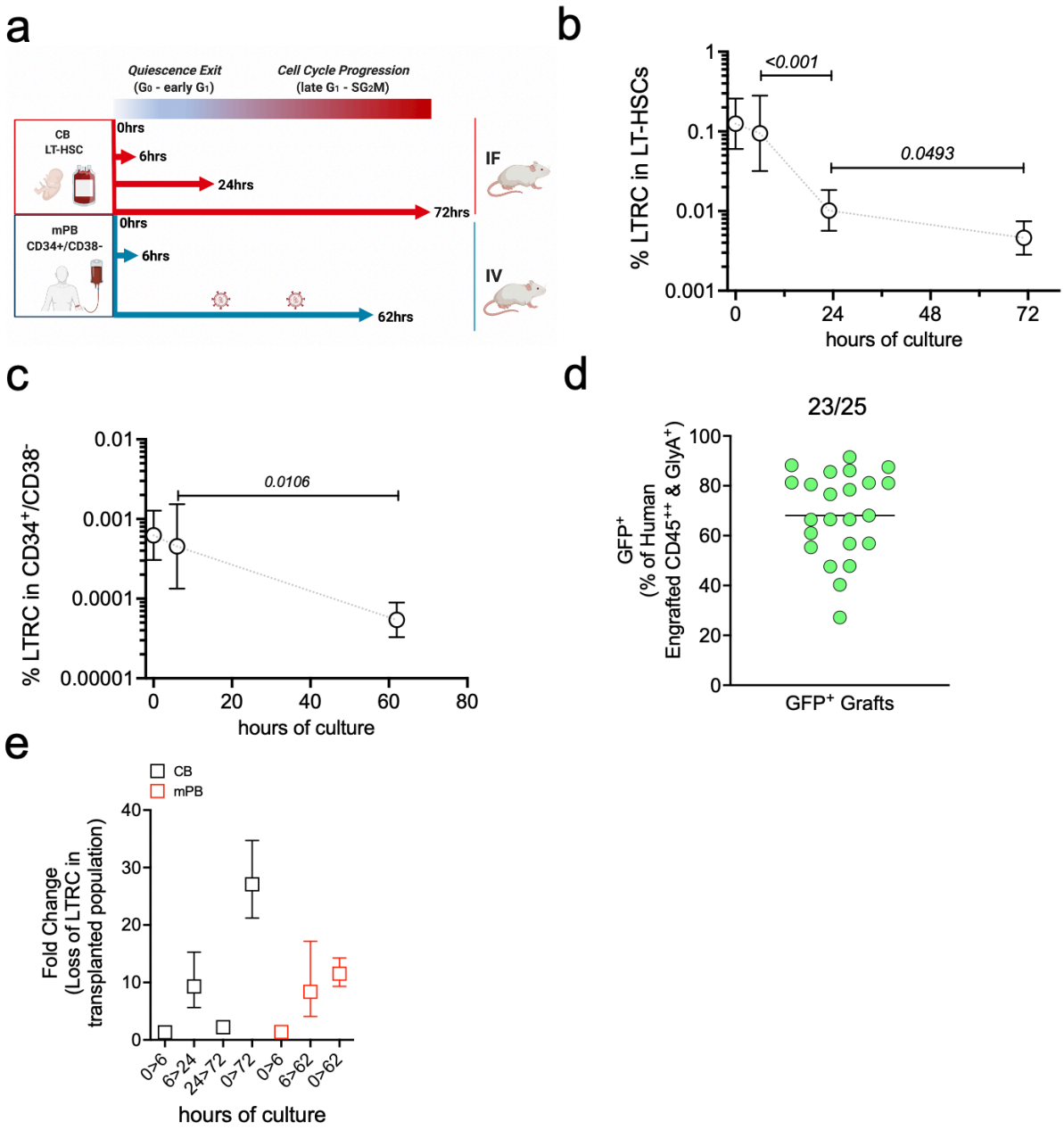

Figure 4.8: HSC frequency is maintained for the first 6 hr of culture and is significantly reduced by 72 hr.

a. Workflow of *in vivo* transplantation. CB LT-HSCs were cultured in differentiation facilitating conditions and transplanted at indicated doses and time-points by intrafemoral (IF) injection. mPB $CD34^+/CD38^-$ cells were cultured in *ex vivo* gene therapy conditions (including exposure to a lentiviral vector containing GFP) and transplanted at indicated doses and time-points by intravenous (IV) injection. Mice were harvested at 18 weeks post transplantation. Created with Biorender.

b. Extreme limiting dilution analysis (ELDA) of the HSC frequency / % long-term repopulating cell (%LTRC) within the transplanted CB LT-HSC population. Chi-squared test performed and significance indicated.

c. ELDA analysis of HSC frequency within the transplanted mPB $CD34^+/CD38^-$ population. Chi-squared test performed and significance indicated.

d. GFP^+ proportion of human $CD45^{++}$ cells from mice engrafted with mPB cultured for 62 hr in gene therapy conditions. Number of engrafted mice with GFP^+ grafts (determined by ≥ 30 cells in GFP^+ gate) as a proportion of total engrafted mice indicated above graph.

e. Fold change of HSC frequency (determined by ELDA analysis) from time-point comparisons for CB and mPB transplantations.

I next sought to measure the kinetics of long-term repopulating HSC loss in the gene therapy protocol conditions. mPB $CD34^+/CD38^-$ cells were cultured in these conditions for 0, 6 and 62 hr, this final time-point marking completion of an *ex vivo* gene therapy protocol provided by GSK at the beginning of my PhD studies. At specified time-points cells were transplanted by intravenous (IV) injection in three doses (Table 4.2). To note, $CD34^+/CD38^-$ cells were used instead of LT-HSCs to better represent clinical protocols of *ex vivo* gene therapy which use the $CD34^+$ fraction (Aiuti et al., 2013; Biffi et al., 2013; Hacein-Bey-Abina et al., 2002), reading out LT-HSC function at 18 weeks post transplantation. The $CD34^+$ fraction itself could not be chosen due to the high dose of cells required per mouse to achieve long-term engraftment (LT-HSCs are <1% of $CD34^+$) (Notta et al., 2011).

Hours of culture	Cell Dose	Number of tested mice	Number of engrafted mice
0hr	5000	5	4
0hr	1250	9	6
0hr	50	6	1
6hr	50	6	0
6hr	1250	3	1
6hr	5000	2	2
62hr	55000	12	9
62hr	25000	13	12
62hr	8000	4	4

Table 4.2: Human mPB $CD34^+/CD38^-$ cell dose and mouse information for primary *in vivo* transplantation studies in gene therapy protocol conditions. Human mPB $CD34^+/CD38^-$ cell dose (representative of 0 hr at the time of sort), the number of mice tested and the number of mice showing human engraftment 18 weeks post transplantation are indicated. Mice were considered engrafted if human cells ($hCD45^{++}$ & $GlyA^+$) $\geq$ 0.01% of Singlets and $\geq$ 30 cells present in any lineage determination gate. 0 hr: 20 mice, n=2 experiments; 6 hr: 11 mice n=1 experiment; 62 hr: 29 mice n=3 experiments.

Integrating the data from the LDAs at each time-point suggested that mPB cultured in gene therapy conditions follows similar kinetics of repopulation capacity loss to the studies in CB. Long-term repopulation capacity is maintained for the first 6 hr (Fig.4.8c; 0 hr vs 6 hr *p=0.672*) with a HSC frequency of ≈ 1 in 1601 at 0 hr and ≈1 in 2208 at 6hr. Subsequently the HSC frequency declines by approximately 10-fold over 62 hr in culture (≈ 1 in 18466 at 62 hr) as cells are undergoing cell cycle progression. Of note, LDA experiments in mPB were performed at 24 hr but there were unanticipated health issues with this cohort and the number of mice remaining did not yield statistical power to perform an LDA. Mice with GFP^+ grafts ($\geq$ 30 GFP^+ cells) after exposure to the full 62 hr *ex vivo* gene therapy protocol and transplantation accounted for the majority of engrafted mice (23/25 mice; Fig.4.8d) and GFP^-

grafts were only observed in mice with <1% human engraftment (data not shown). This demonstrates that long-term repopulating HSCs were targeted by the transduction protocol and that as a whole, expression of the vector was not silenced over the 18-week experiment.

Due to the difference in population transplanted between studies performed with CB and mPB, the loss of long-term repopulation capacity over culture was not directly compared between the two sources. Instead, the fold change relative to uncultured mPB/CB was calculated. The early dynamics of long-term repopulation capacity in culture over 6 hr are similar between mPB and CB despite differences in the medium and stem cell source (Fig.4.8e). The results obtained suggest that mPB cultured in gene therapy conditions show a reduced loss in long-term repopulation capacity compared to CB cultured in differentiation facilitating conditions, although it should be considered that the final time-point in these studies differs by 10 hr (Fig.4.8e). Importantly, the cell dose chosen in 62 hr mPB transplantation experiments is significantly higher when compared to other time-points because it was expected that %LTRC would significantly decline between 24 – 62 hr. However, results demonstrate that there is a high proportion of engrafted mice at the 62 hr time-point using current cell doses (only 4/29 mice are not engrafted). Caution should therefore be taken when drawing conclusions of the % LTRC at 62 hr using currently obtained results. Due to the high proportion of engrafted mice at 62 hr, this experiment should be repeated using the same cell doses as the 0, 6 and 24 hr time-points.

4.2.3 *Ex vivo* culture causes extensive transcriptional rewiring of uncultured LT-HSCs within the first 6 hr

We next sought to investigate transcriptional changes associated with HSC activation over culture in an effort to better understand the kinetics of long-term repopulation capacity loss *ex vivo*. CB LT-HSCs were sorted by flow cytometry and cultured in differentiation facilitating conditions over a time-course for 0, 6, 24 and 72 hr before scRNA-Seq. These culture conditions were chosen due to the significant reduction in long-term repopulation capacity observed over 72 hr (Fig.4.8b) and aimed to encompass the full breadth of HSC activation over the first complete cell cycle *ex vivo*. Libraries were prepared by the Smart-Seq2 protocol (Methods 3.6) (Picelli et al., 2014). 668 total cells were sequenced, pre-processed and 64% of cells passed quality control (Methods 3.8.1). Following selection of highly variable genes, cell clustering was performed by the K-Nearest Neighbour (KNN) approach for downstream analysis. Force directed graph (FDG) visualisation demonstrates that *ex vivo* cultured CB LT-HSCs cluster by time in culture (Fig.4.9a) and pseudotime ordering largely followed time in culture (Fig.9b).

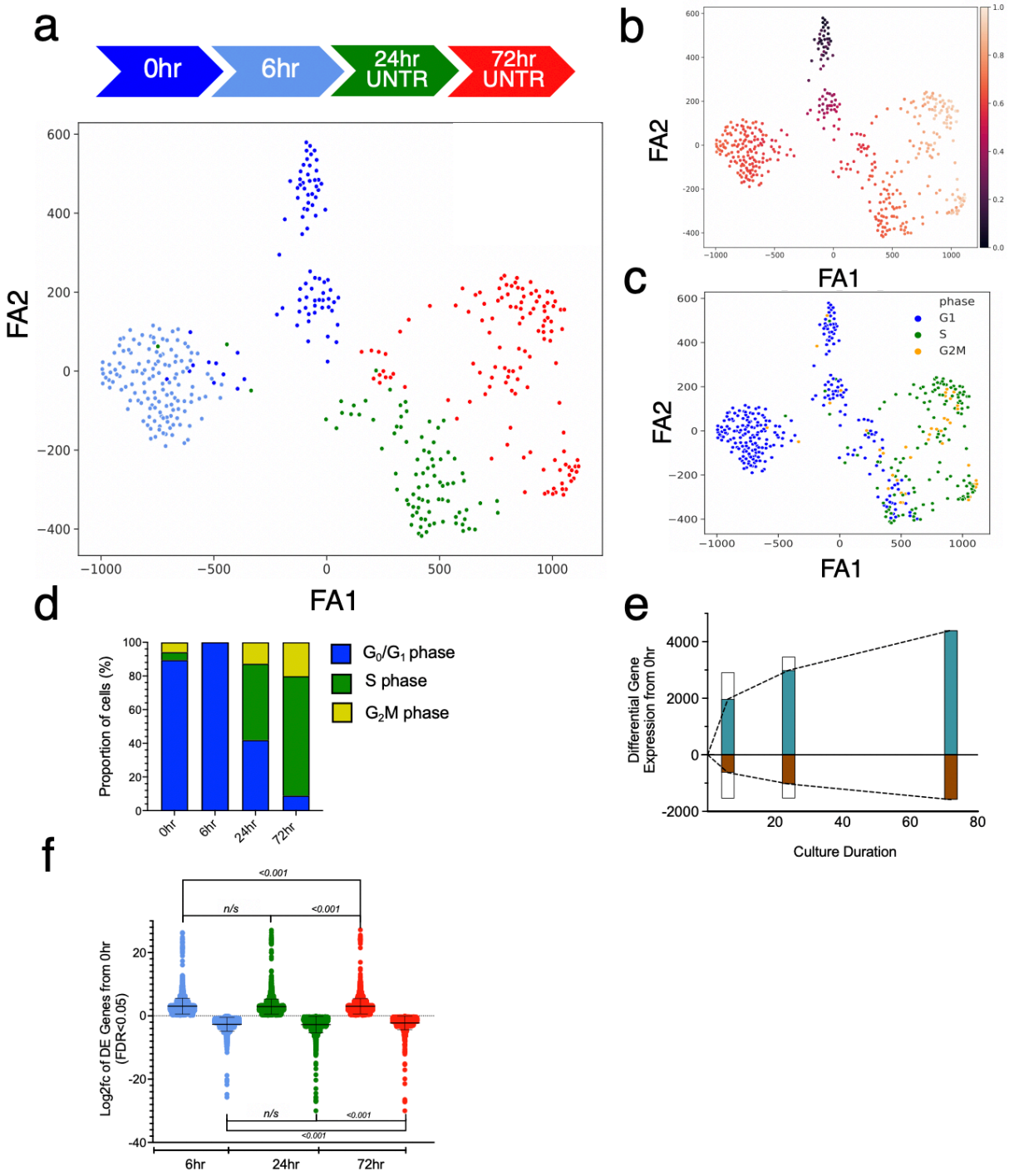

Figure 4.9: *Ex vivo* culture over 72 hours causes significant transcriptional rewiring from the quiescent CB LT-HSC state.

a. – c. FDG visualization of scRNA-Seq from 389 CB LT-HSCs cultured in differentiation facilitating conditions coloured by time-point (a) pseudotime (b) transcriptionally allocated cell cycle status (c). FDG constructed from 16256 HVG.

d. Barplot quantification of c.

e. Number of differentially expressed genes from uncultured CB LT-HSCs. Upregulated genes indicated by a positive change and downregulated genes indicated by a negative change. Dashed lines indicate changes between each successive time-point. Coloured boxes indicate differentially expressed genes in time-point comparison also differentially expressed in 0 hr vs 72hr. Unshaded boxes indicate differential expression from 0 hr not observed in 0 hr vs 72 hr comparison.

f. log2FC of differentially expressed genes from 0 hr (Unpaired t-test performed and significant values indicated (6 hr vs 24 hr upregulated *p=0.173*; 6 hr vs 24 hr downregulated *p=0.0583*).

Transcriptional assignment of cell cycle status (Fig.4.9c+d) showed an increasing proportion of LT-HSCs allocated to S-G_2M phases as culture progresses (58% at 24 hr and 91% at 72 hr) in agreement with functional analysis (Fig.4.6b). Differential gene expression was then performed for all pairwise comparisons in the dataset using DeSeq2. *Ex vivo* culture has a rapid effect on the transcriptome with 40% of all gene expression change observed between 0 - 72 hr of culture happening within the first 6 hr (Fig.4.9e), before the vast majority of cells (>95%) have undergone quiescence exit (Fig.4.8a). The union of all differentially expressed genes between any pairwise comparison over the time-course was then selected (10010 genes). The amplitude of gene expression change (log_2FC) from 0 hr to any other time-point was then calculated for each of these genes. The median log_2FC is significantly higher for changes occurring between 24 – 72 hr for both upregulated and downregulated genes when compared to 6 – 24 hr (Fig.4.9f). Results may suggest that the amplitude of gene change accelerates only once cells have surpassed quiescence exit.

4.2.4 A subset of uncultured CB LT-HSCs are primed for aerobic metabolism

Ex vivo culture significantly rewires the quiescent CB LT-HSC transcriptome with marked differential expression observed over successive time-points. Until now, differential expression has only been considered between time-points, not informing on heterogeneity within clusters. To better quantify transcriptional heterogeneity of CB LT-HSCs within each time-point the ROGUE package was used with the aim to determine the uniformity of the single cell populations at each time-point (Liu et al., 2020a). ROGUE captures the degree of randomness or entropy expected within a population of single cells. A heterogenous cell population will show increased fluctuations in expression levels, constraining the degree of randomness observed and decreasing the determined ROGUE value (Liu et al., 2020a). Surprisingly, the time-point with the lowest uniformity is 0 hr (Fig.4.10a) whereby a subset of uncultured HSCs locate closer to cultured cells by separation along the FA2 axis (Fig.4.10a). To investigate this observation, 0 hr cells were manually categorised into “0 hr-EARLY” and “0 hr-LATE” (0 hr-EARLY $\geq$ 350 FA2) denoted by their respective pseudotime score (Fig.10b).

2207 genes were significantly upregulated and 829 genes were significantly downregulated in 0 hr-LATE compared to 0 hr-EARLY cells (data not shown). The degree of transcriptional heterogeneity within the 0 hr time-point is therefore larger than the transcriptional change induced by culture for 6 hr, but not for 24 hr. GSEA was then used to assess the potential similarity of 0 hr-EARLY and 0 hr-LATE cells to previously characterised populations. Signatures were generated from a subset of CB LT-HSCs previously discovered by our lab defined by $CD34^{lo}CLEC9A^{hi}$ cell surface expression, which exhibit

slower quiescence exit kinetics and multipotent potential (Belluschi et al., 2018) and a population of dormant human LT-HSCs (Zhang et al., Nature Cell Biology, *in revision*). 0 hr-EARLY LT-HSCs show significant enrichment (*FDR< 0.01*) to both the human dormant LT-HSC signature and multi-potent CD34^{lo}CLEC9A^{hi} LT-HSCs compared to 0 hr-LATE cells (Fig.4.10c). Conversely, signatures of CB LT-HSCs with myelo-lymphoid restricted potential (CD34^{hi}CLEC9A^{lo}) which have faster quiescence exit kinetics (Belluschi et al., 2018) were enriched in 0 hr-LATE cells (Fig.4.10c). In line with this observation, *CDK6* expression is lower in 0 hr-EARLY cells (Fig.4.10d; *FDR<0.001*) which may suggest a degree of cell cycle priming within the LT-HSC compartment.

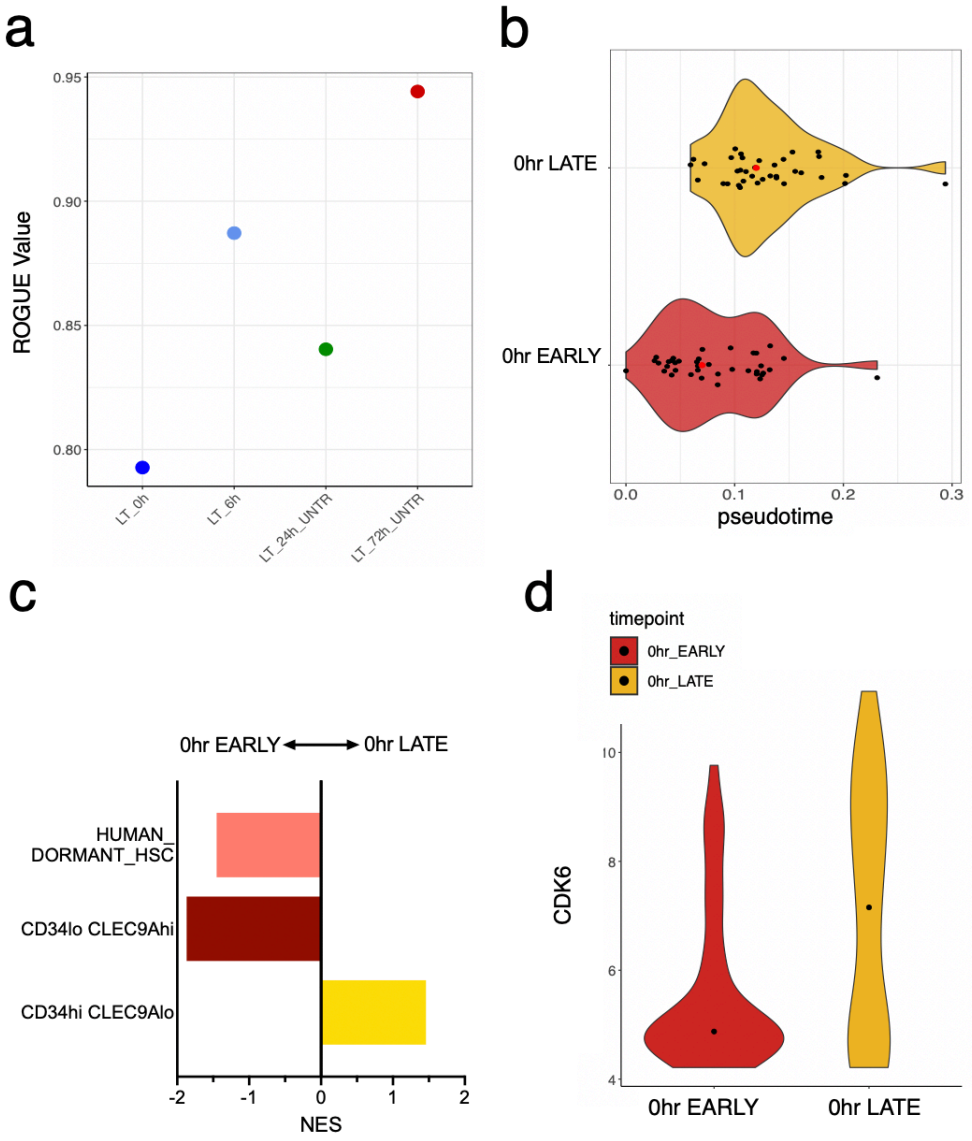

Figure 4.10: A subset of uncultured LT-HSCs with a higher pseudotime score (0hr-LATE) show increased *CDK6* expression.

a. ROGUE score indicating population uniformity at each time-point of culture.
b. Pseudotime score of 0 hr-EARLY and 0 hr-LATE.
c. GSEA enrichment to signatures generated from Belluschi *et al*. 2018 and a human dormant LT-HSC signature generated from the Cabezas-Wallscheid lab (*Nature Cell Biology, in revision*). GSEA was performed from a pre-ranked list generated from the De-Seq2 output between 0 hr EARLY compared to 0 hr-LATE cells.
d. CDK6 transcript expression in 0 hr-EARLY and 0 hr -LATE cells.

0 hr-LATE LT-HSCs are also determined to show upregulation of genes involved in maintaining chromosome organisation such as the centromeric stability proteins *CENPM, CENPO*, *BUB1, AUNIP* (Fig.4.11a) (> 16 log_2FC; *FDR < 0.01*). Quiescent cells have a tightly condensed chromosome structure, therefore it can be suggested that in 0 hr-LATE cells centromeric and associated structural re-organisation genes are increased to prepare for rapid changes in the epigenetic and transcriptional landscape upon HSC activation in culture (Takayama et al., 2021).

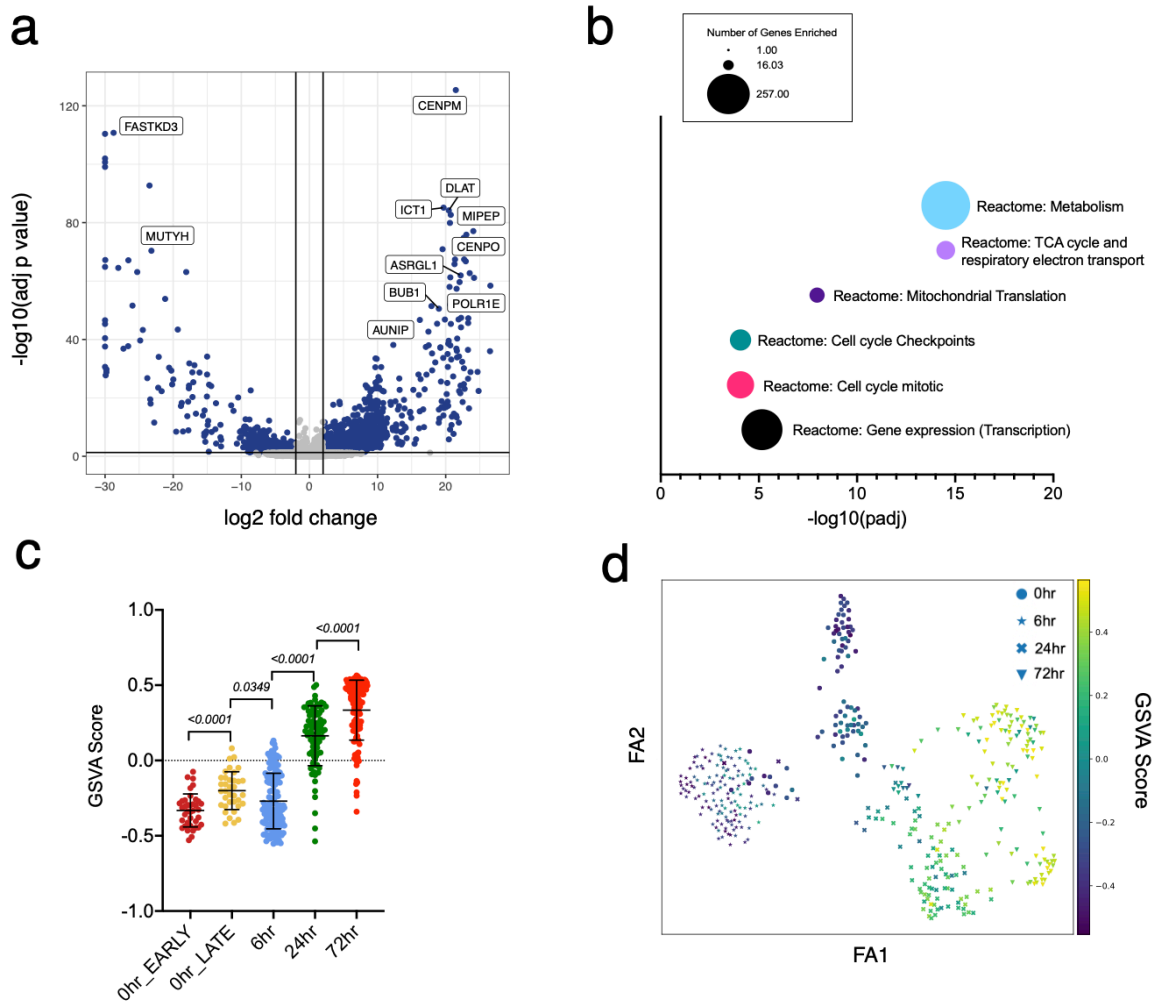

Figure 4.11: 0hr-LATE LT-HSCs show metabolic priming at the transcriptional level.

a. Volcano plot of normalised reads from 0 hr-EARLY and 0 hr-LATE uncultured LT-HSCs. Blue data points indicate significant differentially expressed genes of ≥ 2 log_2fold change (*FDR < 0.05*).
b. Balloon plot of significantly enriched Cluego pathways (*FDR < 0.001*) (from KEGG and Reactome databases). Size of data-point indicates number of enriched target genes.
c. GSVA scores per cell of Reactome TCA Cycle and respiratory electron transport pathways. Unpaired t-test performed and significance indicated.
d. Force directed graph coloured by GSVA score per cell of Reactome TCA Cycle and respiratory electron transport pathways in (c). Representative of 389 single CB LT-HSCs. Shape indicates culture time-point.

Another notable feature of 0 hr-LATE cells is the strikingly higher (> 15 log_2FC) basal levels of key genes involved in oxidative phosphorylation such as *MIPEP*, a peptidase involved in metabolic protein maturation*, DLAT* a component of the pyruvate dehydrogenase complex and the mitochondrial tRNA hydrolase *ICT1* (Fig.4.11a). Cluego biological pathway analysis was performed to analyse significantly enriched Reactome, KEGG and Gene Ontology (GO) pathways (Bindea et al., 2009). Enriched Reactome pathways include "Gene expression (Transcription)" "TCA Cycle and Respiratory Electron Transport", "Metabolism" and "Mitochondrial Translation" (Fig.4.11b). GSVA scores calculated per cell demonstrate that 0 hr-LATE cells have significantly higher pathway expression of "Reactome: TCA Cycle and Respiratory Electron Transport" on average within the population compared to both 0 hr-EARLY cells and also 6 hr cultured LT-HSCs, although it should be noted that the 6 hr time-point shows heterogenous expression of this pathway (Fig.4.11c + d) Taken together, this may suggest that on average, 0 hr-LATE cells show a higher degree of metabolic activation than LT-HSCs cultured for 6 hr. Overall, these data show that despite their quiescent status, uncultured LT-HSCs show a gradient of *CDK6* transcript expression and transcriptional metabolic priming.

4.2.5 Dynamic patterns of gene activation and repression exist within the first cell cycle *ex vivo*

To better understand the molecular changes over the time-course, the Bioconductor package degPatterns (Patano, 2017) was implemented to group differentially expressed genes over the time-course by expression pattern. 15 groups of expression were selected (accounting for 8966 genes) that did not show redundancy in pattern (Fig.4.12). The majority (>70%) of genes were involved in dynamic patterns of expression whereby a change in direction or amplitude was observed over 72 hr culture, rather than a continuous trend. To unpick key trends exhibited over the time-course, patterns of upregulation and downregulation were classified into Early, Continuous, Late and Transient based on direction and amplitude of change (Fig.4.12).

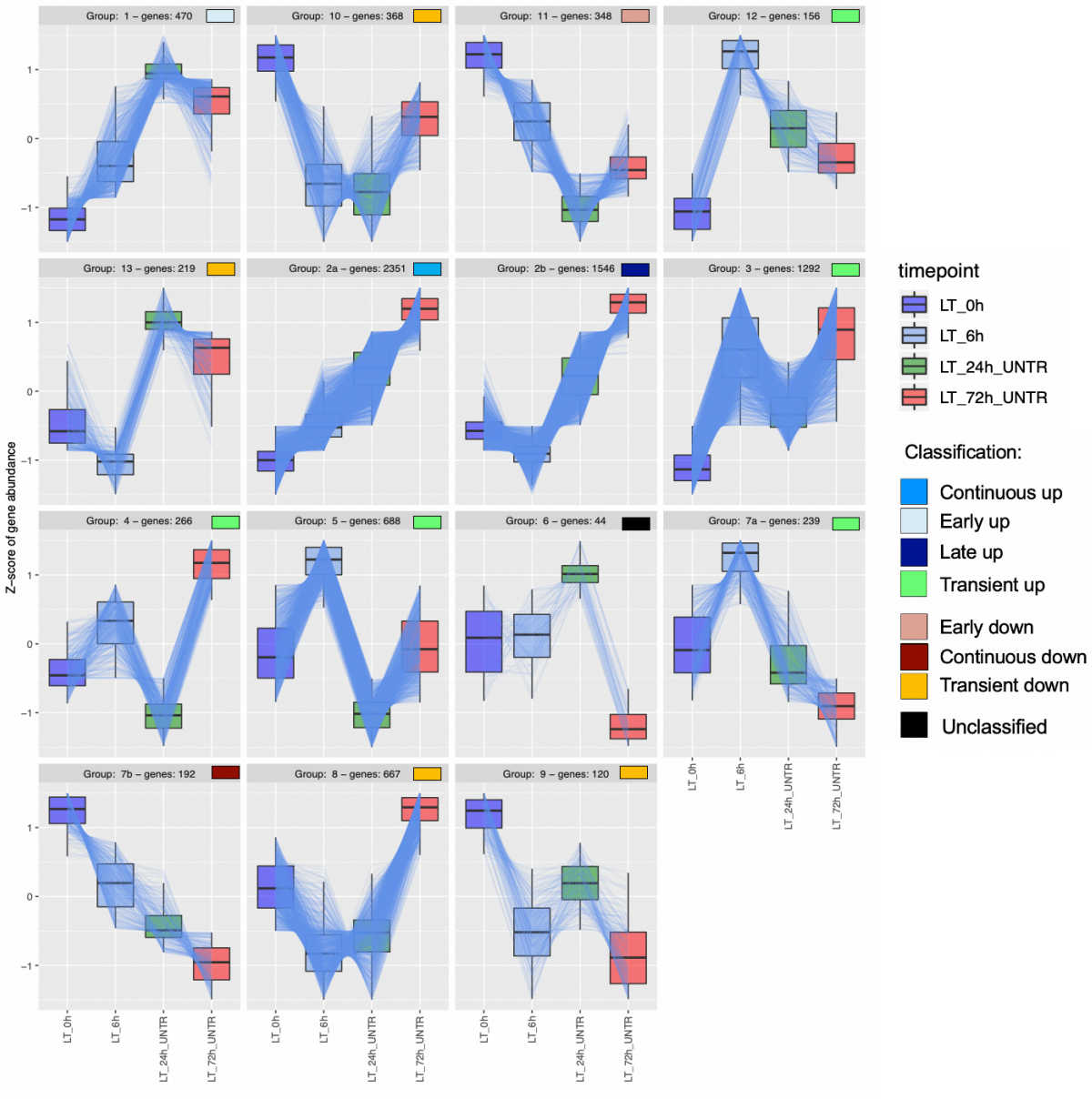

Figure 4.12: Distinct patterns of gene expression activation and repression over 72 hr in CB LT-HSC culture
Raw output of boxplots generated by degPatterns package to group genes based on expression pattern over the 72 hr CB LT-HSC time-course (representative of 8966 filtered genes; min group size > 15). Connecting lines represent individual genes. Line at median. Upper and lower box whiskers indicate the 25th and 75th percentile of expression. Group number and the number of genes fitting this pattern are indicated. Coloured tiles indicate the designated classification.

Upregulated genes show a variety of expression trends over the time-course (Fig.4.13a).The continuous up grouping (2351 genes) depicts a gradual increase of expression over 72 hr and includes genes involved in pathways previously associated with HSC activation from quiescence (Cabezas-Wallscheid et al., 2017; Kruta et al., 2021; Simsek et al., 2010) including “KEGG: RNA Polymerase”, “Reactome: Protein Folding” and “Reactome: TCA Cycle and Respiratory Electron Transport” (Fig.4.13b+e).

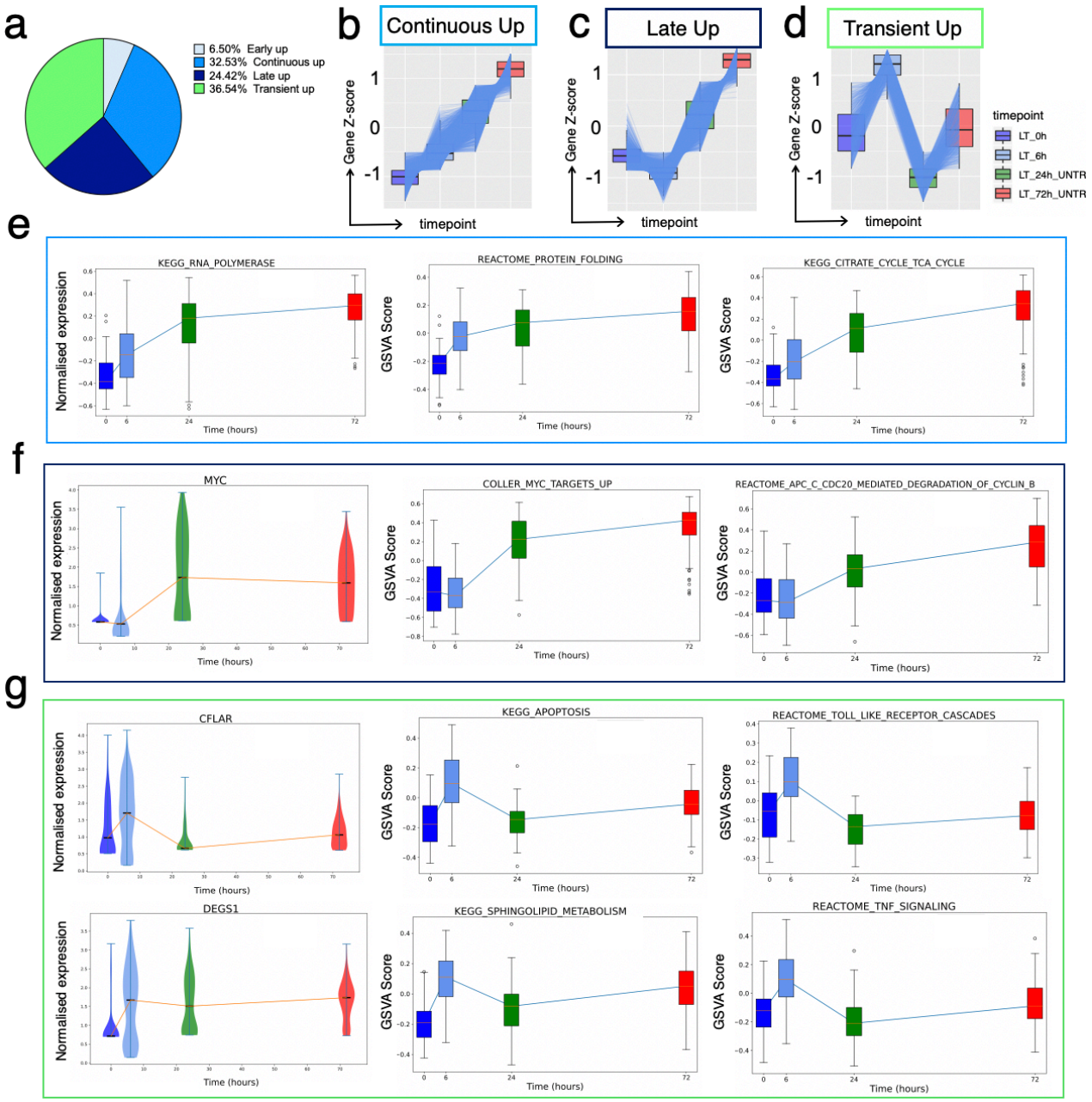

Figure 4.13: Dynamics of gene expression activation during 72 hr CB LT-HSCs culture

a. Pie chart denoting groupings of upregulated gene expression dynamics from genes differentially expressed over 72 hr fit into an expression pattern by degPatterns (representative of 7227 genes).

b. -d. Examples of genes following continuous up, late up and transient up patterns. Boxes represent time-points. Line at median. Upper and lower whiskers of box indicate the 25th and 75th percentile of expression.

e. GSVA scores per cell of c2 curated pathways (GSEA analysis) clustered by DEGpatterns following a continuous up trend scaled over time.

f. Violin plot of *MYC* natural log normalised expression (left) GSVA scores per cell of c2 curated pathways (GSEA analysis) following a late up trend scaled over time (middle-right).

g. Violin plot of *CFLAR* and *DEGS1* natural log normalised expression (Left: upper+lower). GSVA scores per cell of c2 curated pathways (GSEA analysis) following a transient up trend scaled over time (Middle-Left: upper + lower).

Data indicates that transcription, translation and mitochondrial activity are continuously increasing over 72 hr in culture.

Late activation (1765 genes) is characterised by upregulation only after the 6 hr time-point (Fig.4.13c+f). Interestingly, the master regulator of cell growth, proliferation and differentiation *MYC* follows this pattern in addition to various MYC target pathways "Coller_MYC_Targets_Up" (Fig.4.13f), "Schuhmacher _Myc_Targets_Up" and "YU_Myc_Targets_Up" (data not shown). Other genes following this trend are predominantly involved in cell cycle progression (87% of enriched KEGG/Reactome pathways directly relate to mitotic related processes; data not shown) as supported by late upregulation of "Reactome: APC-C CDC20 mediated degradation of Cyclin B" (Fig.4.13f).

The transient up classification consists of an initial increase in expression between 0 hr to 6 hr followed by a sharp change in the direction of expression (2641 genes) (Fig.13d+g). The pro-survival gene *CFLAR* follows this pattern and is accompanied by GSVA pathways "KEGG: Apoptosis", "Reactome: Toll Like Receptor Cascades" and "Reactome: TNF signalling" (Fig.4.13g). In addition, *DEGS1*, the final enzyme in *de novo* sphingolipid biosynthesis follows this trend as supported by transient upregulation of "KEGG: Sphingolipid metabolism" (Fig.4.13g). When pharmacologically inhibited *ex vivo*, *DEGS1* has recently been shown to promote maintenance of HSC self-renewal capacity by restoring optimal proteostasis (Xie et al., 2019). Taken together, our data broadly indicate that cell stress and apoptotic pathways are transiently activated in culture.

Downregulated genes predominantly follow a dynamic, transient trend over 72 hr in culture (68% of filtered downregulated genes) (Fig.4.14a). The metabolic pathways "KEGG: Primary bile acid biosynthesis", "KEGG: Retinol metabolism" (Fig.4.14b,d) and similarly "Reactome: Retinoic Acid Biosynthesis" (data not shown) are transiently downregulated during 72 hr of culture. Bile acids have recently been shown to be crucial in expanding foetal liver HSCs to regulate proteostatic stress (Sigurdsson et al., 2016) and retinoic acid signalling is crucial for the maintenance of HSC dormancy and self-renewal capacity (Cabezas-Wallscheid et al., 2017; Schönberger et al., 2021). Our data suggest transient suppression of pathways may regulate HSC function in culture.

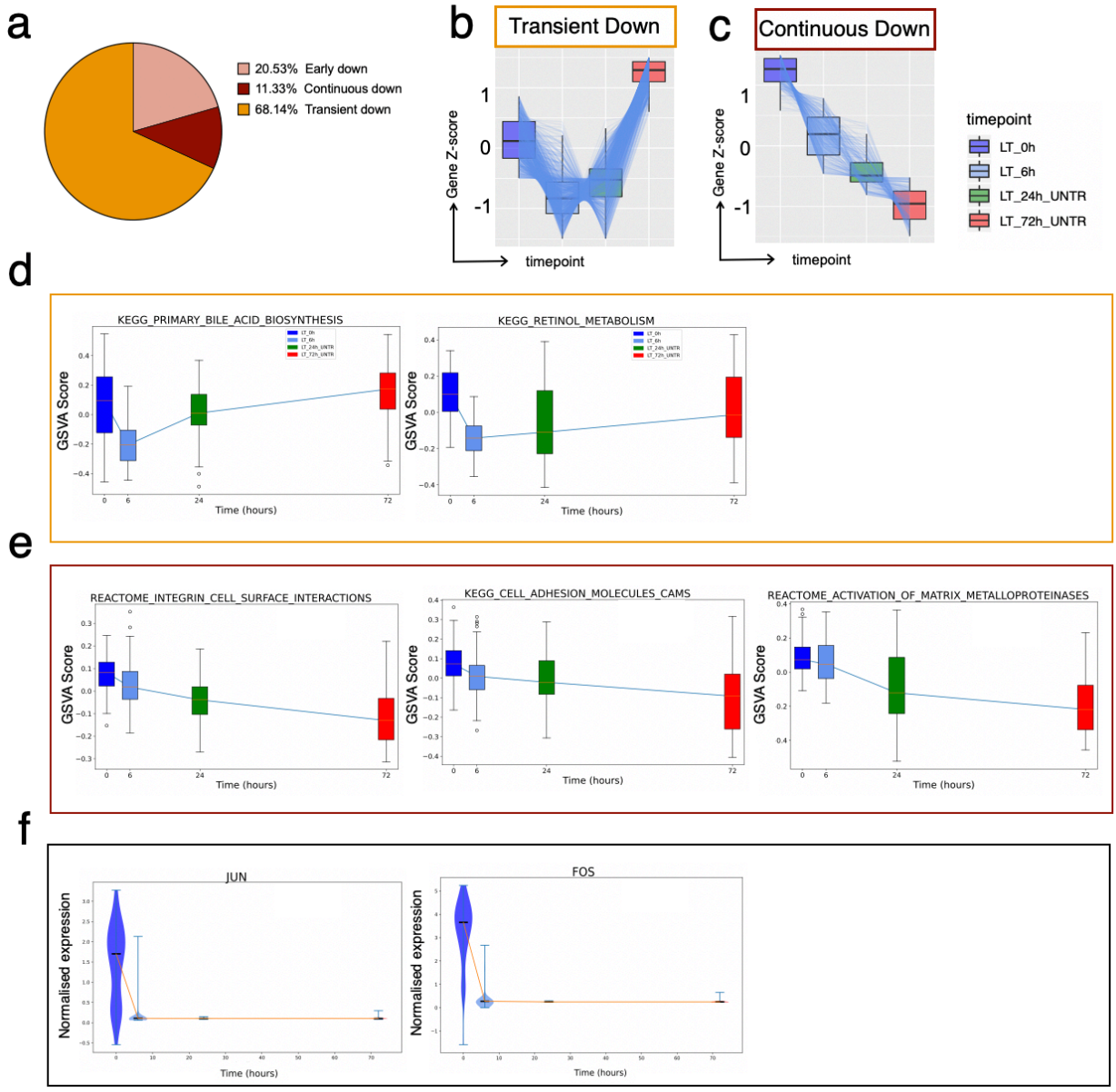

Figure 4.14: Dynamics of gene expression downregulation during 72 hr CB LT-HSC culture.

a. Pie chart denoting groupings of downregulated gene expression dynamics from genes differentially expressed over 72 hr fit into an expression pattern by degPatterns (representative of 1695 genes).
b. -c. Examples of genes following transient down and continuous down patterns. Boxes represent time-points. Line at median. Upper and lower whiskers of box indicate the 25th and 75th percentile of expression.
d. GSVA scores per cell of c2 curated pathways (GSEA analysis) scaled by time following a transient down expression pattern.
e. GSVA scores per cell of c2 curated pathways (GSEA analysis) scaled by time following a continuous down expression pattern.
f. Violin plot of selected unclassified natural log normalised expression of genes.

Continuously downregulated GSVA pathways include “Reactome: Integrin Cell Surface Interactions”, “KEGG: Adhesion Molecules CAMS” and “Reactome: Activation of Matrix Metalloproteinases” (Fig.4.14c+e). Integrin signalling and adhesion molecules such as cadherins and matrix metalloproteinases play important roles in HSC niche maintenance (Theodore et al., 2017; Wagers et al., 2002; Wein et al., 2010). Therefore, these data suggest that that ECM and niche interactions expressed *in vivo* are downregulated in a continuously decreasing manner.

10.4% of genes differentially expressed over the time-course could not be classified into specific grouping suggesting their expression over culture exists as a rare pattern not found in > 15 genes. The AP-1 transcription factors involved in cell cycle regulation, *JUN* and *FOS* could not be classified by degPatterns (Fig.4.14f) and show high expression in 0 hr LT-HSCs before sharply dropping by 6 hr, suggesting that their function is only necessary in quiescent uncultured LT-HSCs.

Taken together, data presented show that the first 72 hr of CB LT-HSC culture is associated with dynamic differential gene expression. Continuously increased expression over time of genes involved in transcription, translation and oxidative phosphorylation is accompanied by serially decreasing expression of genes involved in maintenance of the *in vivo* HSC niche. Interestingly, many differentially expressed genes show transient patterns in culture, such as transient upregulation of cell stress and apoptotic pathways alongside transient downregulation of retinoic acid signalling and bile acid metabolic processes.

4.2.6 Genes bimodally distributed at 6 hr show transient expression dynamics during 72 hr of culture

Our data demonstrate that the average expression value of key genes and biological pathways modulating HSC behaviour changes dynamically in LT-HSCs between culture time-points. Next, in an effort to better understand the molecular heterogeneity that exists within the LT-HSC fraction, I sought to investigate gene expression variability and bimodality in culture by using methods to assess the distribution of gene expression within time-points.

To identify bimodally distributed genes, SIBER (Systematic Identification of Bimodally Expressed genes) was performed using the filtered count matrix for 10010 genes differentially expressed between any time-point comparison over 72 hr in culture (Tong et al., 2013). Briefly, SIBER fits a lognormal distribution to a given gene’s expression distribution and quantifies the distance between two

components of expression to yield a bimodal index (BI) per gene (Methods 3.8.5). Genes with a BI ≥ 1.9 were classified as bimodally distributed.

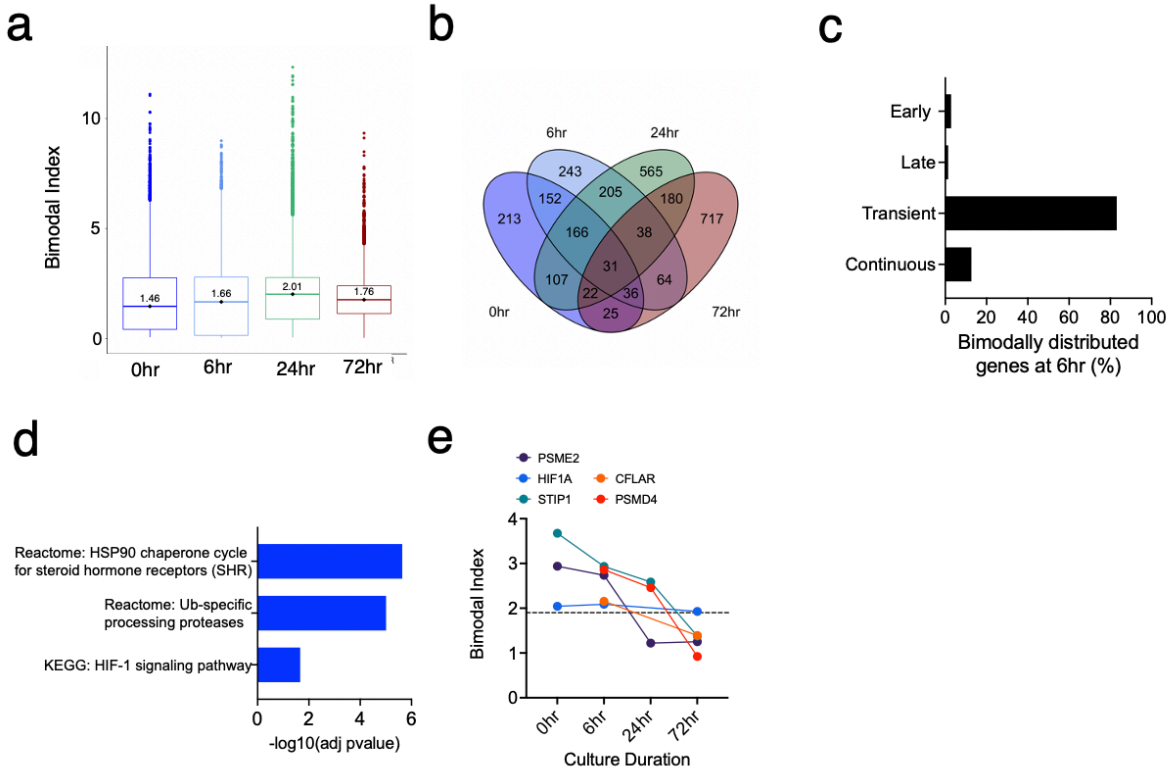

Figure 4.15: Gene expression variability is dynamic over the time-course and 6 hr bimodally distributed genes are transiently expressed over the 72 hr time-course.

a. Distribution of the calculated bimodal index over time-course per gene. Calculated using the filtered count matrix of 10010 genes differentially expressed between any pairwise comparison over 72 hr in culture. Line at median. Upper and lower box indicated at 25th and 75th percentile.
b. Venn Diagram of common bimodally distributed genes between pairwise comparisons over the 72 hr time-course.
c. Classification of expression pattern for the combined list (116 genes total) of genes showing increased variability and bi-modality at 6 hr (including upregulated and downregulated expression dynamics).
d. Enriched KEGG and Reactome pathways (selected pathways shown; *FDR<0.05*) from genes showing increased variability and bimodality at 6 hr.
e. Mean bimodal index (calculated for n=2 scRNA-Seq experiments separately) of selected genes from pathway analysis in (d). Dotted line indicates determined bimodal index threshold of ≥ 1.9. Missing data-points indicate excluded values due to high zero inflated count (not considered bimodal due to technical effects).

The median BI increases sequentially from 0 hr to 24 hr with the highest stepwise change in BI occurring between 6 hr to 24 hr (Fig.4.15a). The number of bimodal genes is dynamic over the time-course with only 1.1% (31 genes) being consistently bimodal over all four time-points (Fig.4.15b). To complement this approach and explore gene expression variability in our dataset from a different perspective, we used the package BASiCS (Bayesian Analysis of Single-Cell Sequencing data). BASiCS gives a distinct

read out of molecular heterogeneity by identifying increased variability (calculated by residual over-dispersion) from that expected for the mean expression value of a given gene (Eling et al., 2019; Vallejos et al., 2015). We reasoned that combining the lists of differentially variable and bimodal genes allowed for increased confidence of genes which were bimodal as a result of biological meaning rather than attributed to technical effects.

Variability in transcript expression within populations has recently been shown to predict later cell fate choices between populations (Grün, 2020) and given that a sharp drop in long-term repopulation capacity is observed between the 6 -24 hr transition, we focussed our analysis on variable genes at the 6 hr time-point. Genes at the 6 hr time-point which showed both increased variability using the BASiCS package and a bimodal distribution of expression using the SIBER package (combined list of 116 genes; hereafter referred to as "bimodal genes") were compared to other time-points. It was identified that 6 hr bimodal genes predominantly show a transient classification of expression over the time-course (Fig.4.15c). Cluego pathway analysis was performed on bimodal genes at 6 hr and significant enrichment to "Reactome: HSP90 chaperone cycle for steroid hormone receptors", "Reactome: Ub-specific processing proteases" and "KEGG: HIF1a signalling pathway" was observed (Fig.4.15d). Key enriched genes from these pathways show a reduction in bimodality between 6 – 24 hr / 6 - 72 hr and include the hypoxia-Inducible factor *HIF1a*, the anti-apoptotic regulator *CFLAR*, the stress induced phosphoprotein 1 *STIP1*, and the proteosome subunits *PSME2* and *PSDM3* (Fig.4.15e).

Overall, these data suggest that genes which are bimodally distributed and show increased variability at the 6 hr time-point largely show transient dynamics of expression and are involved in cell stress related pathways.

4.2.7 Common molecular mechanisms of activation from quiescence between adult stem cell populations

Data so far builds a picture that quiescence exit *ex vivo* is associated with dynamic transcriptional changes impacting cell stress and long-term repopulating capacity. Previous work by Rodgers *et al*. has defined an injury induced state of "Galert" in muscle satellite cells contralateral to a site of injury (CSCs) which represents an intermediary state of transcriptional activation when compared to quiescent (QSC) and activated (ASC) populations (Rodgers et al., 2014). CSCs show the highest transcriptional similarity to QSCs as opposed to ASCs (Fig.4.16a) with the Galert state characterised by increased cell size, mitochondrial activity, faster cell cycle kinetics and enhanced tissue regenerative capacity (Rodgers et al., 2014). To better understand how our time-course data of activation over

culture corresponds to *in vivo* stem cell activation in response to injury, transcriptional signatures from ASC, CSC and QSC mouse populations were generated and genes were converted to human orthologs. GSEA from the 72 hr CB LT-HSC time-course to these signatures (only significant values shown; *FDR < 0.05*) demonstrates that muscle stem cell activation *in vivo* is fairly concordant with CB LT-HSC activation *ex vivo*. The strongest enrichment of the CSC signature was at 24 hr compared to 6 hr (Fig.4.16b) suggesting that this transition may be transcriptionally similar to the characterised state of Galert.

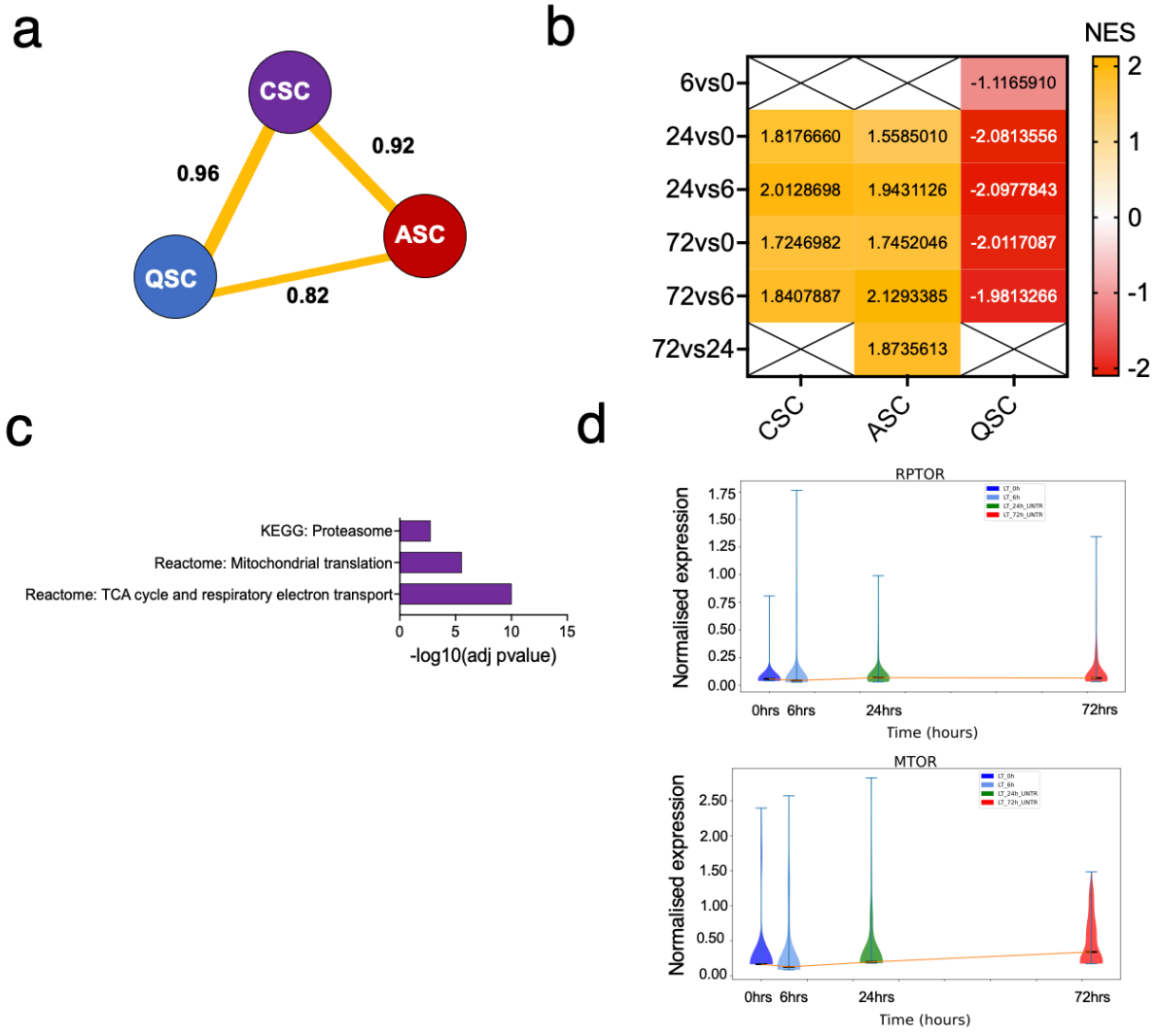

Figure 4.16: LT-HSC exit from quiescence and activation *ex vivo* show common molecular features to muscle stem cell activation *in vivo*.

a. Illustration of transcriptional association between muscle stem cell populations (adapted from Rodgers et al. 2014; Figure 2). Measured by Pearsons correlation coefficient. Satellite cells in muscle contralateral to site of injury (CSCs) exhibit an intermediary state of transcriptional activation and distinct cell cycle state, from quiescent (QSC) and activated (ASC) populations.
b. GSEA analysis of ASC, QSC, and CSC signatures (only significantly enriched values shown; *FDR < 0.05)*. Positive enrichment indicates enrichment to leading term.
c. Cluego biological pathway analysis from genes differentially expressed between 6 to 24 hr and enriched in CSC signature.
d. Natural log normalised expression of key regulators involved in MTOR signaling scaled by time (upper - lower).

No significant enrichment of ASC or CSC signatures was observed at 6 hr relative to 0 hr indicating that both time-points are primitive in activation. Removal of cell cycle associated genes from the signatures (Methods 3.8.7) has a very minimal effect on the enrichment score (average difference in NES = 0.023; data not shown) indicating that cell cycle genes are not the main driver of the enrichment observed.

Biological pathway analysis (Cluego; *FDR < 0.01*) was performed on differentially expressed genes from 6 – 24 hr which were also found in the CSC signature. Enriched biological pathways included "Reactome: Mitochondrial Translation" (including *MRPL/MRPS* genes), "Reactome: TCA Cycle and Respiratory Electron Transport" (including *COX7B*, *NDUFA11*, *SDHB*) and "KEGG: Proteasome" (including *PSMB3*, *PSMB3*, *PSMB5*) (Fig.4.16c). Activation of mTORC1 is required for the injury induced state of Galert such that genetic ablation of *Rptr* leaves CSCs unresponsive to injury (Rodgers et al., 2014). However, we do not see an increase in notable regulators of MTOR signalling (Fig.4.16d) or downstream signal transduction pathways (no change in expression of *TSC1/2*, *VEGFA*, *PIK3A*, *AKT1*; data not shown). Importantly, mTORC signalling may be differentially regulated over culture post-transcriptionally. Collectively, my data suggest that there are similarities in the molecular mechanisms of activation from quiescence between *ex vivo* and *in vivo* populations and more broadly across adult stem cell sources.

4.2.8 LT-HSCs cultured for 6 hr maintain survival in response to proteostatic stress

The 6 to 24 hr transition shows a dramatic drop in CB long-term repopulation capacity and transiently expressed genes over this time window show enrichment to cell stress related pathways. Given these observations, I next wanted to perform Annexin-V/7-AAD staining to investigate whether transcriptional changes in cell stress response pathways have functional implications on cell viability over the time-course. Of note, the following experiments would have ideally been performed with CB HSC populations, however the COVID-19 pandemic caused a complete halt to CB donations and therefore, I instead used mPB samples to address this question. Annexin-V/7-AAD staining was performed at each time-point in mPB LT-HSCs and ST-HSCs in three matched mPB donor samples across two experiments. The long-term repopulation capacity of ST-HSCs is drastically reduced (1 in approximately 500 in uncultured CB) (Notta et al., 2011). This population was included in these experiments to identify changes differential within LT-HSCs which may inform on the mechanism driving long-term repopulation capacity loss over culture. A significant drop in viability is observed between the 6 to 24 hr timepoint in LT-HSCs (average 24% reduction) and ST-HSCs (average 31% reduction) with a subsequent slight increase by 62 hr of culture (Fig.4.17a). LT-HSCs show a reduction in the proportion of cells in

the early apoptotic phase (Annexin-V$^+$/7-AAD$^-$) in matched samples at 0 hr and a trend for this reduction at other time-points (Fig.4.17b). This may suggest that LT-HSCs have a delayed apoptotic response to culture induced stress relative to ST-HSCs. The reduction in viability does not directly result in a significant drop in absolute cell counts, highlighting that only a proportion of the population are undergoing apoptosis (Fig.4.17c) and therefore the reduction in stem cell frequency at 24 hr (Fig.4.8b+d) is not primarily explained by a lower number of starting cells injected than at earlier time-points.

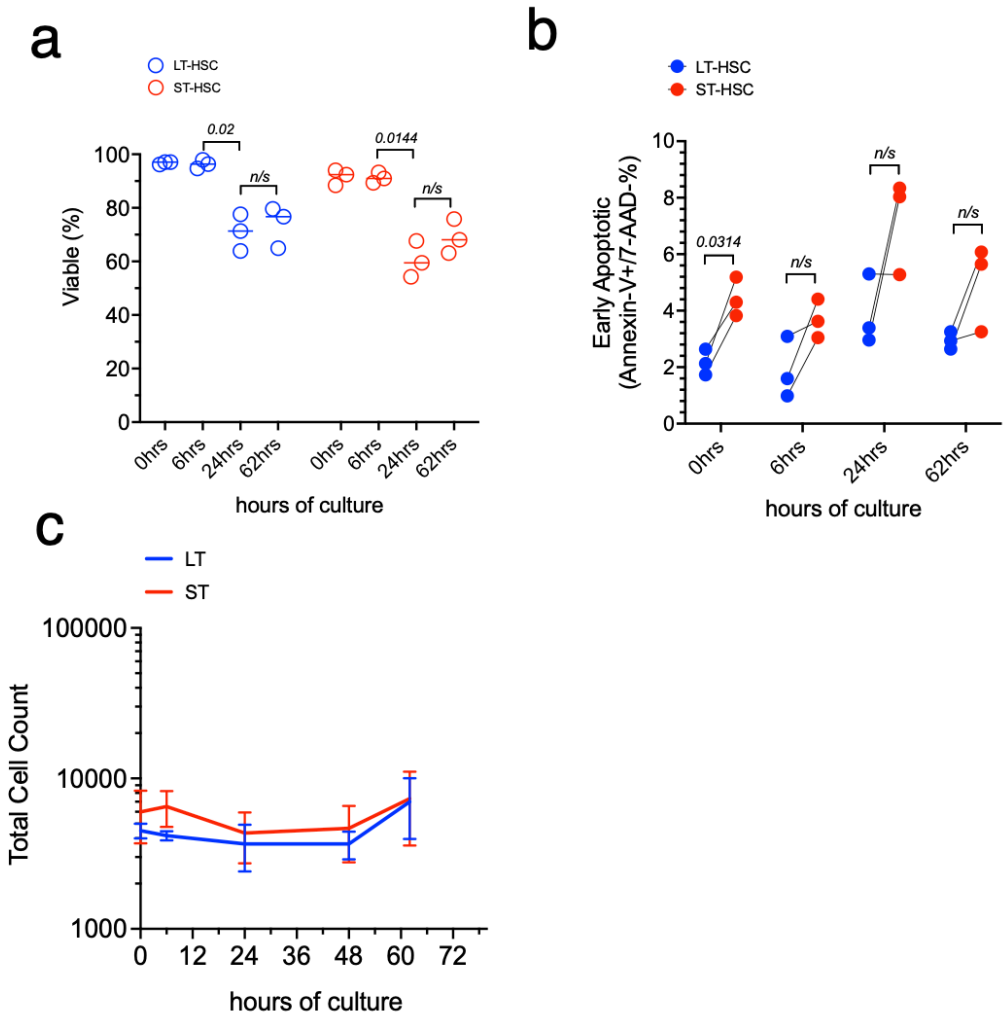

Figure 4.17: A significant drop in viability is observed between 6 to 24 hr in mPB LT-HSC and ST-HSC populations cultured in gene therapy media.

a. Proportion of viable (Annexin-V$^-$ / 7-AAD$^-$) mPB LT-HSCs and ST-HSCs from matched donor samples at indicated time-points of culture in *ex vivo* gene therapy conditions (n=3 matched mPB donor samples performed over n=2 repeat experiments, mean values from mPB donor samples plotted; paired t-test performed and significant values indicated).

b. Proportion of early apoptotic (Annexin-V$^+$ / 7-AAD$^-$) mPB LT-HSCs and ST-HSCs from matched donor samples at indicated time-points (n=3 matched mPB donor samples performed over n=2 repeat experiments, mean values from mPB donor samples plotted; paired t-test performed with significant values indicated) (6 hr LT-HSC vs 6 hr ST-HSC *p=0.1151*; 24 hr LT-HSC vs 24 hr ST-HSC *p=0.187*; 62 hr LT-HSC vs 62 hr ST-HSC *p=0.1652*).

c. Manual cell counts (performed by Trypan blue exclusion) at indicated time-points of culture (n=3 matched mPB donor samples over n=1 experiment).

LT-HSCs maintain high expression of cell stress response pathways (Galen et al., 2014, 2018) which cooperate to promote HSC survival during low stress conditions such as amino acid deprivation (Galen et al., 2018) and preferential apoptosis under high stress conditions such as disrupted proteostasis (Galen et al., 2014). Transient dynamics of cell stress related pathways suggest that cells cultured for different durations may show distinct responses to proteostatic stress induction. To investigate this hypothesis, 6 hr and 24 hr cultured mPB LT-HSCs and ST-HSCs were treated with tunicamycin, which blocks N-linked glycoprotein synthesis and causes misfolded protein accumulation in the ER with subsequent UPR activation (reviewed in DuRose et al., 2006). Following tunicamycin addition to media, cells were cultured for an additional 24 hr (Fig.4.18a). LT-HSCs cultured for 6 hr maintain higher viability upon proteostatic stress induction than 6 hr cultured ST-HSCs (Fig.4.18b; *p=0.0414*). Furthermore, 6 hr cultured LT-HSCs show a trend for a higher survival in response to proteostatic stress when compared to 24 hr cultured LT-HSCs (Fig.4.18b; 6 to 24 hr *p=0.199*). These data may suggest that during the first hours of culture LT-HSCs show an increased ability to respond to stress which is later abolished in tested culture durations.

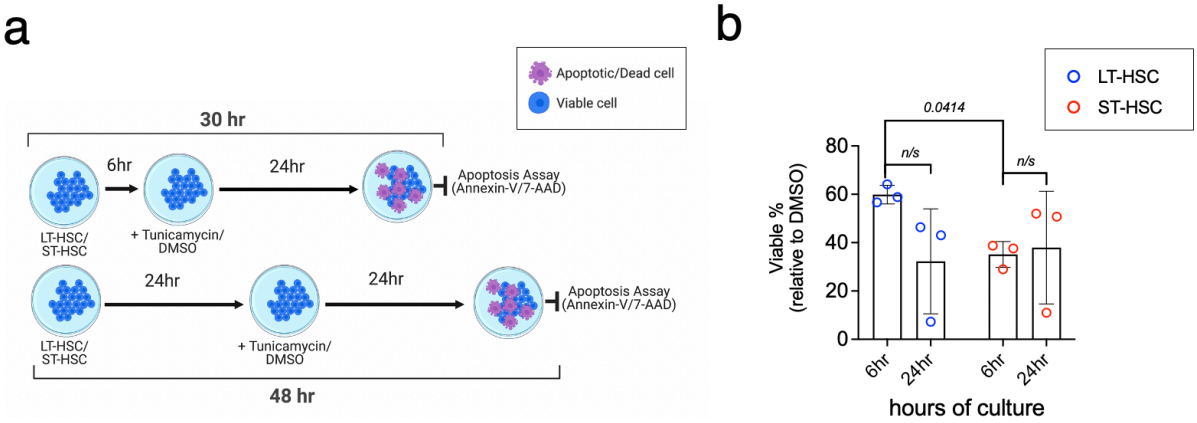

Figure 4.18: LT-HSCs exposed to proteostatic stress at 6 hr show a trend for higher viability when compared to 24 hr of culture.

a. Workflow of culture with proteostatic stress inducing agent tunicamycin. Tunicamycin (3µg/ml) was added to sorted bulk mPB LT-HSC and ST-HSCs at 6 or 24 hr of culture in gene therapy media. An equivalent concentration of DMSO was added to control wells. Cells were harvested for apoptosis assay (Annexin-V/7-AAD staining) following a further 24 hr in culture. Created with Biorender.

b. Viable cells (Annexin-V^- / 7-AAD^-) in response to proteostatic stress induction by tunicamycin treatment in matched LT-HSCs and ST-HSCs from n=3 mPB donor samples (Paired t-test performed and significant values shown; LT-HSC 6 hr vs 24 hr *p=0.199*).

4.2.9 Investigating the application of our datasets as a resource for further study into LT-HSC activation *ex vivo*

My data thus far conclusively demonstrate that during clinically relevant culture, long-term repopulation capacity within the LT-HSC fraction is dramatically reduced. However, a subset of LT-HSCs remain following *ex vivo* gene therapy culture, as shown from longitudinal data acquired from HSC *ex vivo* gene therapy clinical trials(reviewed in Sagoo and Gaspar, 2021). Although importantly, this fraction is not easily identifiable. We hypothesised that the time-course scRNA-Seq dataset may be a useful reference dataset to mine for prospective markers of cultured LT-HSCs or molecular pathways which maintain the LT-HSC state *ex vivo*.

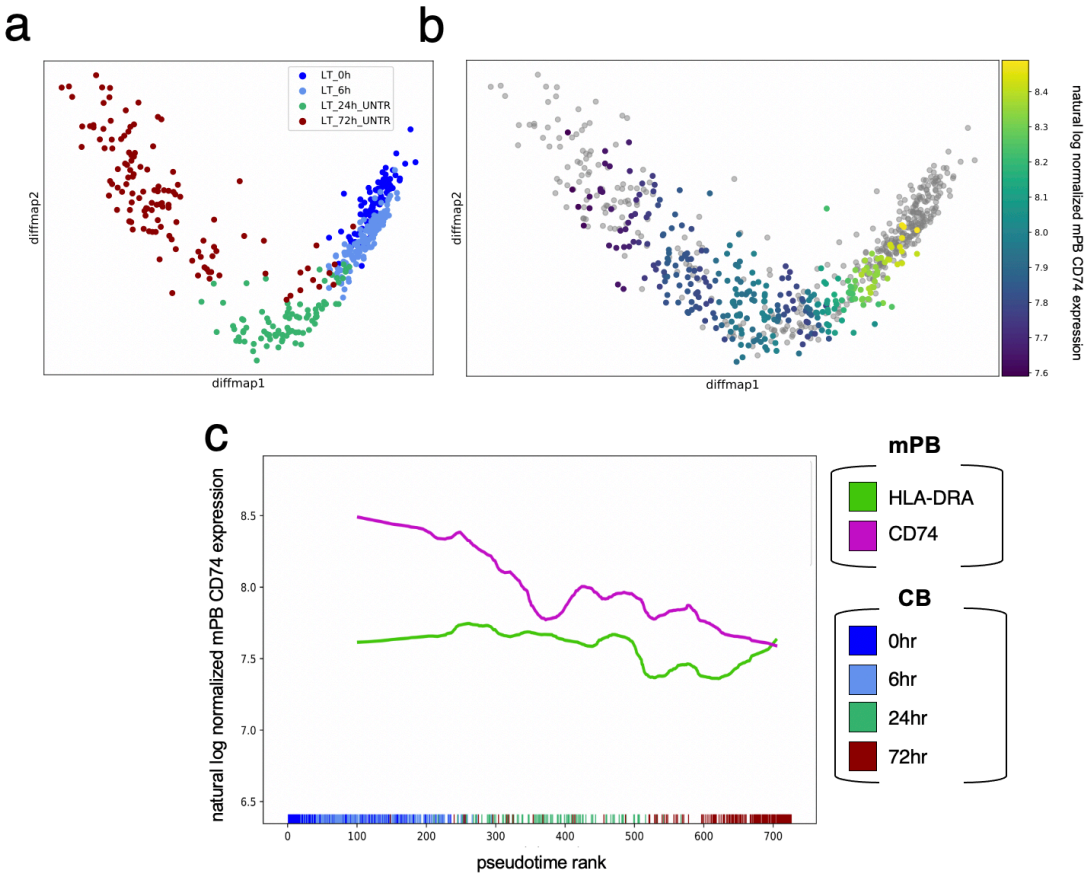

Figure 4.19: 62 hr cultured mPB LT-HSCs with relatively higher CD74 expression show transcriptional similarity to earlier time-points of CB LT-HSC activation *ex vivo.*

a. Diffusion map of CB LT-HSCs cultured in differentiation facilitating conditions for 0, 6, 24 and 72 hr (n=389 single CB LT-HSCs, experimental data equivalent to Fig.4.9; Diffusion map constructed from 3643 HVG).
b. Diffusion map of 62 hr mPB LT-HSCs (GFP^+ and GFP^-) projected on to the CB reference dataset (n=223 single 62 hr GFP^+ and GFP^- mPB LT-HSCs, experimental data equivalent to Fig.4.1; Diffusion map constructed from 3643 HVG). CB dataset coloured in grey. mPB LT-HSCs coloured by smoothed natural log normalised *CD74* expression.
c. Smoothed natural log normalised expression of *CD74* and *HLA-DRA* in mPB LT-HSCs over calculated pseudotime rank from CB time-course scRNA-Seq dataset. (Rug plot indicative of n=389 CB LT-HSCs and n=223 mPB LT-HSCs).

mPB LT-HSCs cultured in gene therapy protocol conditions and sequenced from Chapter 1 were integrated with the CB scRNA-Seq 72 hr time-course dataset. The datasets were then corrected for batch effects using ComBat and highly variable genes were selected in scanpy (Methods 3.8.2). The integration was then visualised by diffusion map and a starting cell picked from the CB scRNA-Seq dataset to calculate a pseudotime score (Fig.4.19a). A subset of LT-HSCs with high *CD74* expression locate towards earlier time-points (notably 0 hr and 6 hr) on the CB scRNA-Seq dataset (Fig.4.19b+c). However, this is not observed for another prospective marker identified in Chapter 1, *HLA-DR,* where expression shows fairly consistent levels over the calculated pseudotime rank (Fig.4.19c). Results obtained suggest that 62 hr mPB LT-HSCs with relatively higher *CD74* expression exhibit transcriptional similarity to CB LT-HSCs cultured for 6 hr. This result is interesting when taken in context that mPB and CB cultured for 6 hr show preserved long-term repopulation capacity (Fig.4.8b+c). We hypothesize that *CD74* may be further investigated as a transcriptional marker to demarcate a population of cultured LT-HSCs. More broadly, this result suggests merit in utilising our data-set as a resource to better understand the molecular consequences of HSC activation *ex vivo*.

4.2.10 Chapter 2 Summary:

Collectively, data presented in this chapter demonstrate that the first cell cycle *ex vivo* is associated with dynamic transcriptional patterns and a significant reduction in long-term repopulation capacity. Curiously, we find that repopulation capacity is maintained for the first 6 hr of culture before significantly dropping within the timeframe that cells are completing quiescence exit. At the 6 to 24 hr transition, we observe an increased proportion of HSCs undergoing apoptosis and a trend for reduced viability in response to proteostatic stress. This is framed at the molecular level by dynamic regulation of cell stress and apoptotic pathways. Taken together, our results suggest that the 6 to 24 hr transition may be key for stem cell fate decisions *ex vivo*.

4.3 Chapter 3: Correlation or Causation: Cell cycle progression and loss of self renewal capacity *ex vivo*

Data presented so far in this thesis demonstrate that *ex vivo* culture of HSCs drives significant transcriptional rewiring from the quiescent state. LT-HSCs must adapt to the oxidative environment of culture and our data suggests that the 6 to 24 hr transition may be crucial in determining the fate of LT-HSCs *ex vivo*. CB and mPB LT-HSCs undergo cell cycle progression over the culture conditions tested and it can be interpreted from our data, alongside many other studies that cell cycle progression *ex vivo* is correlated with a loss of long-term repopulation capacity. Previous literature has inferred that cell cycle progression itself is a major driver of stem cell functional loss in culture. To formally test this, work in this chapter will use the CDK6 inhibitor Palbociclib (PD), to reversibly block cell cycle progression in culture before *in vitro* and *in vivo* functional assays to test stem cell capacity.

4.3.1 Treatment with the CDK6 inhibitor Palbociclib reversibly arrests cell cycle progression *in vitro* with no impact on differentiation output

We first evaluated whether PD treatment could arrest LT-HSCs in early G_1, preventing cell cycle progression over clinically relevant durations of culture. These experiments were performed in the differentiation facilitating conditions and *ex vivo* gene therapy conditions characterised in Chapter 2 and UNTR cells are equivalent to those presented in Chapter 2. A 200nM concentration of PD was chosen for all studies herein based on dose dependent response studies carried out (data not shown) and because this inhibitor concentration has previously been effective within the context of an *in vitro* HSC culture system as shown by Laurenti et al. 2015. Treatment with PD significantly reduces the proportion of cells with Rb phosphorylation at the Ser807/811 site in mPB LT-HSCs (Fig.4.20a-c; *p=0.0024*) and mPB ST-HSCs (Fig4.20c; *p=0.020*) cultured in gene therapy media for 62 hr. Treatment with PD also caused a dramatic reduction in the proportion of divided mPB LT-HSCs cultured in gene therapy conditions (Fig.4.20d + e) and CB LT-HSCs cultured in differentiation facilitating conditions (Fig.4.20f + g). Data from both test culture conditions demonstrates that > 90% of mPB and CB LT-HSCs treated with PD become pharmacologically arrested in early G_1 and cannot pass the restriction point or undergo mitosis during 72 hr of culture. Of interest, CB ST-HSCs cultured in differentiation facilitating conditions show a higher proportion of divided cells when cultured with PD treatment when compared to mPB ST-HSCs cultured in gene therapy media with PD treatment (Fig.4.20e+g). Therefore, it can be concluded that only the LT-HSC fraction are in a complete early G_1 arrest induced by CDK6 inhibition over these culture conditions.

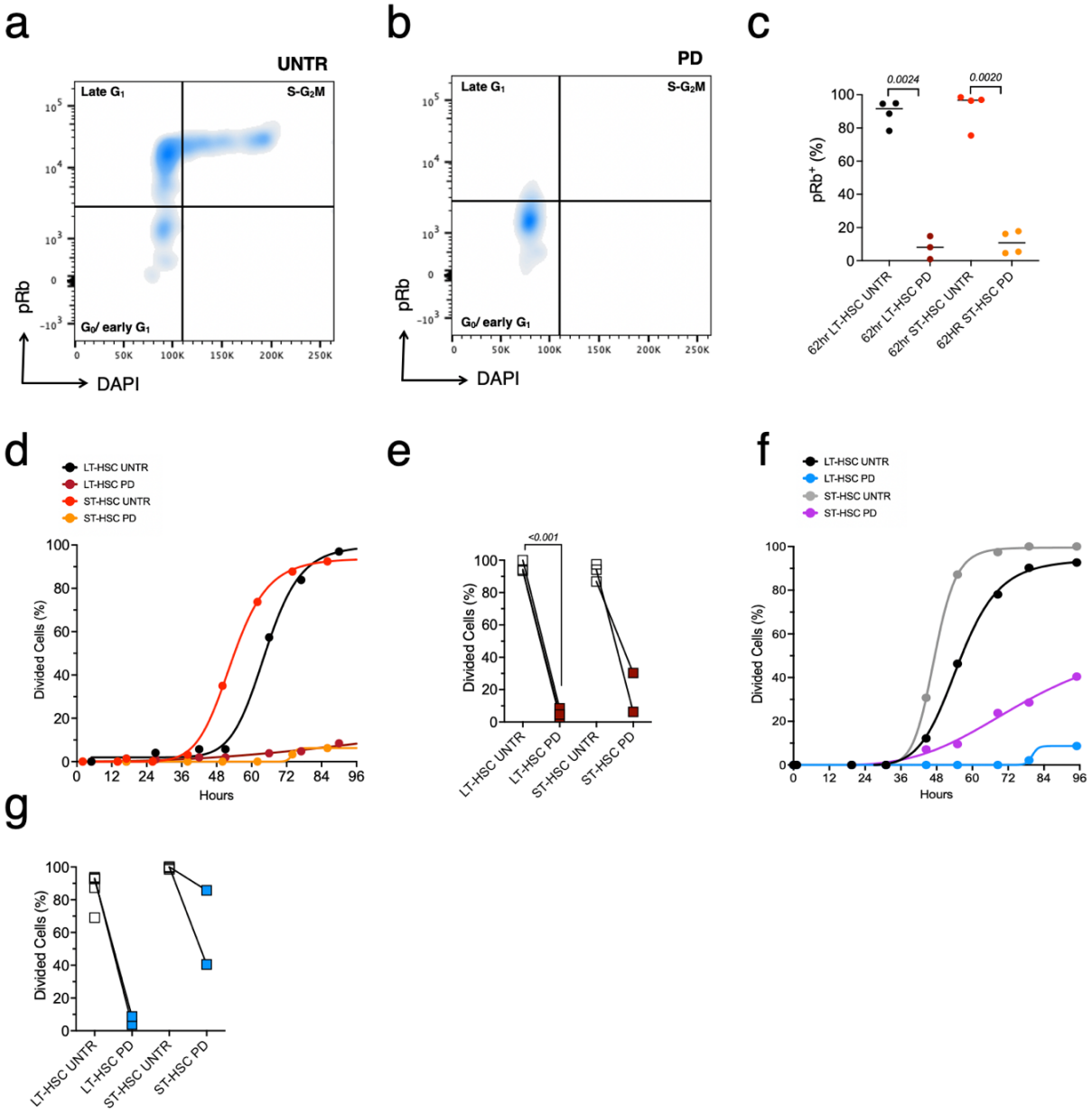

Figure 4.20: CDK6 inhibition by Palbociclib effectively arrests mPB and CB LT-HSCs in early G_1 *ex vivo.*

a. -b Representative example of pRb/DAPI flow cytometry staining in gene therapy media at 62 hr in UNTR mPB LT-HSCs (a) 200nM Palbociclib (PD) treated mPB LT-HSC (b) (gate determined from unstained cells and positive control of actively cycling $CD34^+/CD38^+$ progenitors).

c. Proportion of viable cells showing Rb phosphorylation (Ser807/811) in UNTR/PD treated mPB LT-HSCs and ST-HSCs cultured in *ex vivo* gene therapy conditions (no lentiviral transduction). Paired t-test performed and significant values indicated (n=3 matched independent mPB samples in LT-HSCs, n=1 unmatched untreated LT-HSC; n=4 matched independent mPB samples in ST-HSCs).

d. Cumulative time to first division kinetics of single sorted UNTR/PD treated mPB LT-HSCs and ST-HSCs cultured in gene therapy media. Representative example shown (n=2 matched repeats; n=4 repeat experiments total). Curve shows sigmoidal regression analysis (R^2 >0.99).

e. Quantification of divided single cells from (d) after 96 hr in culture. Connecting lines indicate matched donor samples.

f. Cumulative time to first division kinetics of single UNTR/PD treated CB LT-HSC/ST-HSCs cultured in differentiation facilitating media. Representative example shown (n=2 matched repeats; n=4 repeat experiments total). Curve shows sigmoidal regression analysis (R^2 >0.99).

g. Quantification of divided single cells from (f) after 96 hr in culture. Connecting lines indicate matched donor samples.

We next wanted to determine whether early G_1 arrest during LT-HSC culture could be reversed and whether pharmacological early G_1 arrest impacts *in vitro* differentiation output. mPB LT-HSCs were UNTR or treated with PD and cultured in the 62 hr gene therapy protocol before sorting into single cell My-Meg-Ery-NK differentiation assay conditions (Methods 3.4.3; no PD treatment after 62 hr time-point) (Fig.4.21a). Of note, GFP^+ colonies (≥ 30 cells in GFP^+ gate) consisted of 36.7% of total UNTR colonies and 70.1% of total PD treated colonies, no differences in colony size or lineage output were determined between GFP^+ and GFP^- colonies (data not shown).

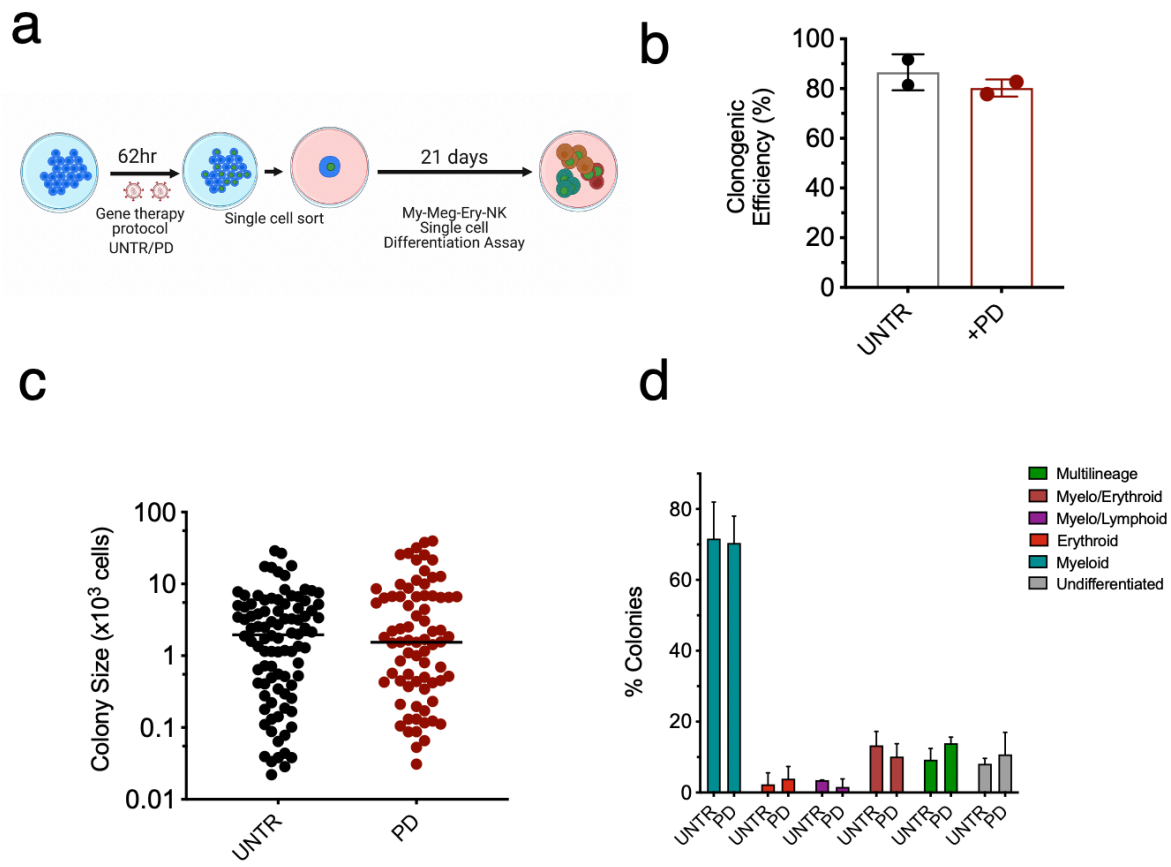

Figure 4.21: mPB LT-HSCs treated with Palbociclib over the *ex vivo* gene therapy protocol generate colonies of equal size and lineage.

a. Experimental workflow. mPB LT-HSCs were exposed to the 62 hr gene therapy protocol (UNTR/ 200nM PD treated) including two hits of lentiviral transduction before single cell sorting into media promoting generation of My-Meg-Ery-NK colonies. Colonies were harvested 21 days post single cell sort (n=2 independent mPB donor samples).
b. Clonogenic efficiency (number of single cells that produced colonies of ≥30 cells) as a proportion of total cells plated (n=2 biological repeats; total n=164 colonies UNTR n=156 colonies +PD). True colonies defined as ≥30 cells.
c. Colony size ($x10^3$ cells) in true colonies.
d. Lineage output from true colonies generated 21 days post single cell sort into single cell differentiation assay media. Multi-lineage is determined by tri-lineage colonies of My/Ery/Lym (NK only) lineage output.

The clonogenic efficiency (the number of single cells generating colonies of size ≥30 cells as a proportion of total single cells plated) is comparable for UNTR and PD treated LT-HSCs in both experiments (Fig.4.21b). Colonies generated are also of approximately equivalent size (Fig.4.21c), demonstrating that the pharmacological early G_1 arrest caused by PD is reversible and cells can later resume normal proliferation. Colonies generated from LT-HSCs treated with PD showed equivalent lineage output compared to UNTR cells with no lineage biases observed (Fig.4.21d). Data presented is caveated by only two biological repeats being performed due to timing restrictions although experiments generated a high number of colonies from single cells (n=164 total UNTR colonies, n=156 total PD colonies).

To the best of our knowledge, this is the only system in which complete and reversible early G_1 arrest can be achieved over 72 hr of *ex vivo* culture in LT-HSCs. The CDK2 inhibitor, Roscovitine, was used in an attempt to arrest LT-HSC cell cycle progression by a previous PhD student, Serena Belluschi, however LT-HSCs only demonstrated slowed time to first division (data not shown). Taken together, these data demonstrate that treatment with the CDK6 inhibitor PD causes a reversible G_1 arrest with cells capable of subsequently unaltered proliferation capacity and differentiation output.

4.3.2 The majority of transcriptional changes observed during *ex vivo* HSC activation are independent of cell cycle progression

To test whether the transcriptional rewiring observed during *ex vivo* HSC activation is dependent on cell cycle progression, PD treated CB LT-HSCs were subject to scRNA-Seq at 24 hr and 72 hr of culture in differentiation facilitating conditions. scRNA-Seq experiments involving PD were performed at the same time as scRNA-Seq using UNTR CB LT-HSCs. A total of 167 PD treated LT-HSCs were sequenced across both time-points. Pre-processing and quality control was performed using the same methodology as used for UNTR cells and 64% of PD treated LT-HSCs passed QC (Methods 3.8.1). LT-HSCs cultured in presence of PD for both 24 hr (representative of of quiescence exit completion) and 72 hr (representative of entering mitosis completion) cluster with 24 hr UNTR cells (Fig.4.22a). PD treated LT-HSCs are transcriptionally allocated to the G_1 phase of the cell cycle (Fig.4.22b) supporting functional data (Fig.4.20f + g) that these cells are effectively arrested in early G_1.

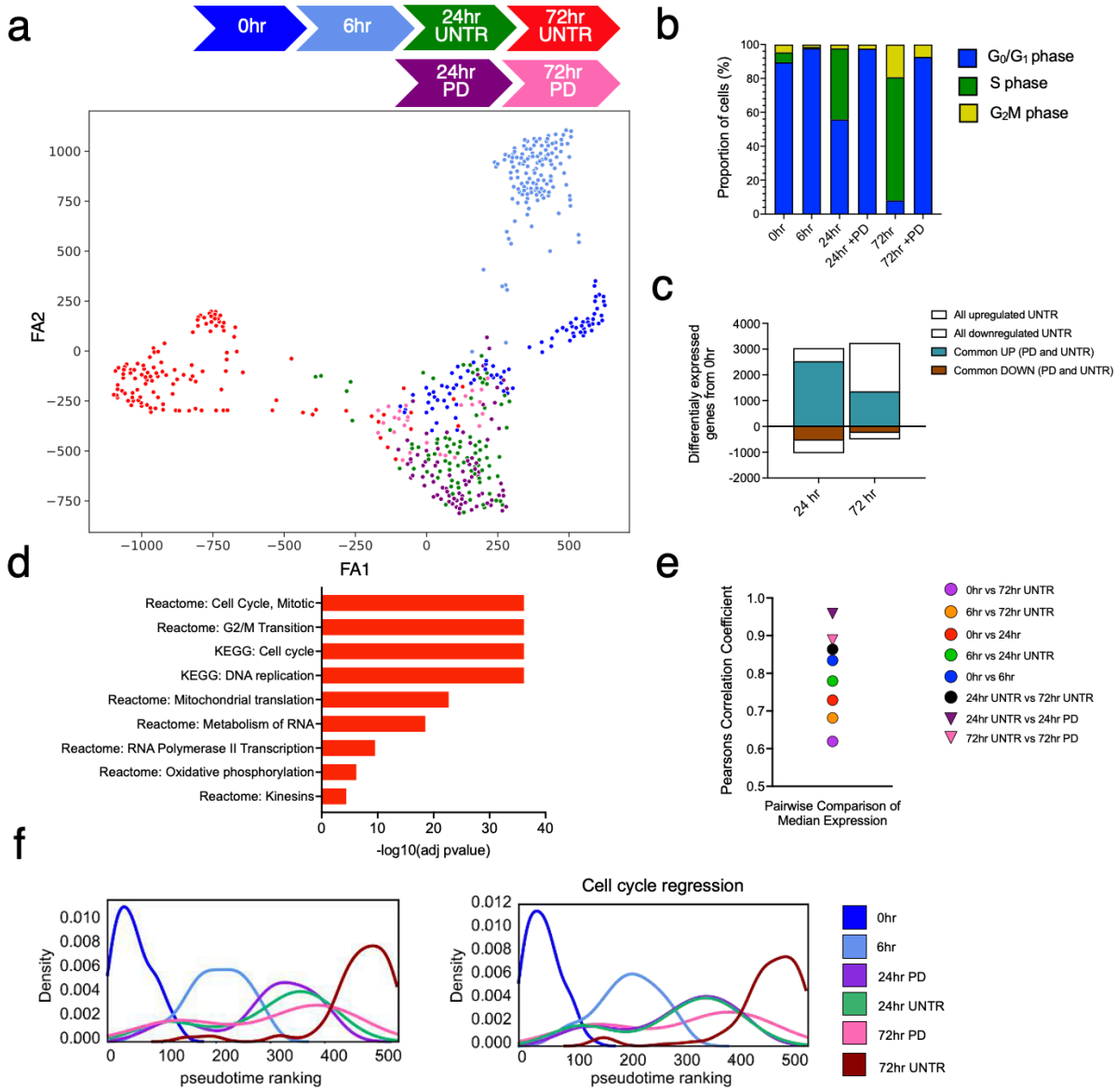

Figure 4.22: The majority of transcriptional change is independent of cell cycle progression during 72 hr CB LT-HSC culture.

a. FDG visualisation of 496 single LT-HSCs at indicated time-points of culture in differentiation facilitating conditions (UNTR) or in the presence of 200nM Palbociclib (PD). FDG calculated based on 14759 HVGs.
b. Proportion of cells allocated to respective cell cycle stages as determined by transcriptional cell cycle allocation.
c. Shared (shaded) and unique (unshaded) differentially expressed genes (*FDR<0.05*) acquired by culture compared to 0 hr CB LT-HSCs. Upregulated genes indicated by positive change and downregulated genes indicated by negative change.
d. Cluego pathway analysis of uniquely changed genes (*FDG < 0.05*) from 0 hr to 72 hr UNTR cells with respect to PD treated cells.
e. Pearsons correlation coefficient comparing median expression value for differentially expressed genes from indicated time-points/condition comparisons.
f. Density plot of cells calculated over DPT pseudotime ranking with cell cycle genes (left) after cell cycle regression (right).

We next sought to determine the transcriptional similarity of HSC *ex vivo* activation between LT-HSCs pharmacologically arrested in G_1 (by PD treatment) compared to UNTR LT-HSCs. At 24 hr of culture, 75% of all differentially expressed genes from uncultured HSCs are common between PD treated and UNTR cells whereas at 72 hr this degree of similarity drops to 43.14% (Fig.4.22c). Pathway analysis of genes changed over 72 hr only in UNTR cells when compared to PD treated LT-HSCs demonstrated the most significant enrichment to mitotic related processes such as "Reactome: Cell Cycle, Mitotic" and "Reactome: G2/M Transition" (Fig.4.22d). To determine whether PD treated LT-HSCs show a similar amplitude of gene expression change to UNTR LT-HSCs, the median gene expression value of differentially expressed genes was plotted for all pairwise time-point comparisons including conditions involving PD treatment and a Pearsons correlation coefficient calculated. As expected, the lowest degree of similarity in the median expression value of differentially expressed genes was observed between uncultured LT-HSCs when compared to 72 hr UNTR LT-HSCs (Fig.4.22e). Of interest, the highest degree of similarity is observed for PD treated cells compared to UNTR cells cultured for matched durations. Pseudotime analysis was then performed to build a temporal trajectory of cell activation over culture. 24 hr and 72 hr PD treated LT-HSCs show a similar density distribution of cells along the pseudotime rank when compared to 24 hr UNTR cells (Fig.4.22f), mirroring the dimensionality reduction analysis. To determine whether the respective ordering of conditions along the pseudotime trajectory was driven by cell cycle genes, cell cycle scoring and regression was performed using the Satija lab package (Methods 3.8.1) and the pseudotime calculation was repeated. The density distribution of cells along the pseudotime axis did not significantly change, indicating that the main driver of transcriptional changes acquired over culture are not cell cycle genes.

Taken together, our data show that cultured LT-HSCs pharmacologically arrested in early G_1 change expression of many shared genes over culture and to a similar amplitude of change when compared to LT-HSCs undergoing cell cycle progression. The highest divergence in differential expression arises at the 72 hr time-point when cells arrested in cell cycle progression do not show upregulation of mitotic related genes. Further, removal of cell cycle genes does not significantly impact the pseudotime ordering of cultured LT-HSCs. It can be determined from this work that the majority of transcriptional changes acquired during human LT-HSC activation in 72 hr culture are therefore independent of cell cycle progression.

4.3.3 Treatment with PD during mPB and CB LT-HSC culture predominantly affects cell cycle related genes

Given our understanding of how PD treatment affects the CB LT-HSC transcriptome during culture, we next sought to investigate the effect of PD treatment on mPB, with the aim of determining whether pharmacological G_1 arrest leads to changes in similar biological pathways. To this end, scRNA-Seq was performed at 62 hr after culture with PD in the gene therapy protocol. scRNA-Seq experiments involving PD were performed at the same time as scRNA-Seq using UNTR mPB LT-HSCs. 164 PD treated mPB LT-HSCs were sequenced and pre-processed. 57% of cells passed QC thresholds (same QC thresholds as for UNTR cells; Methods 3.8.2).

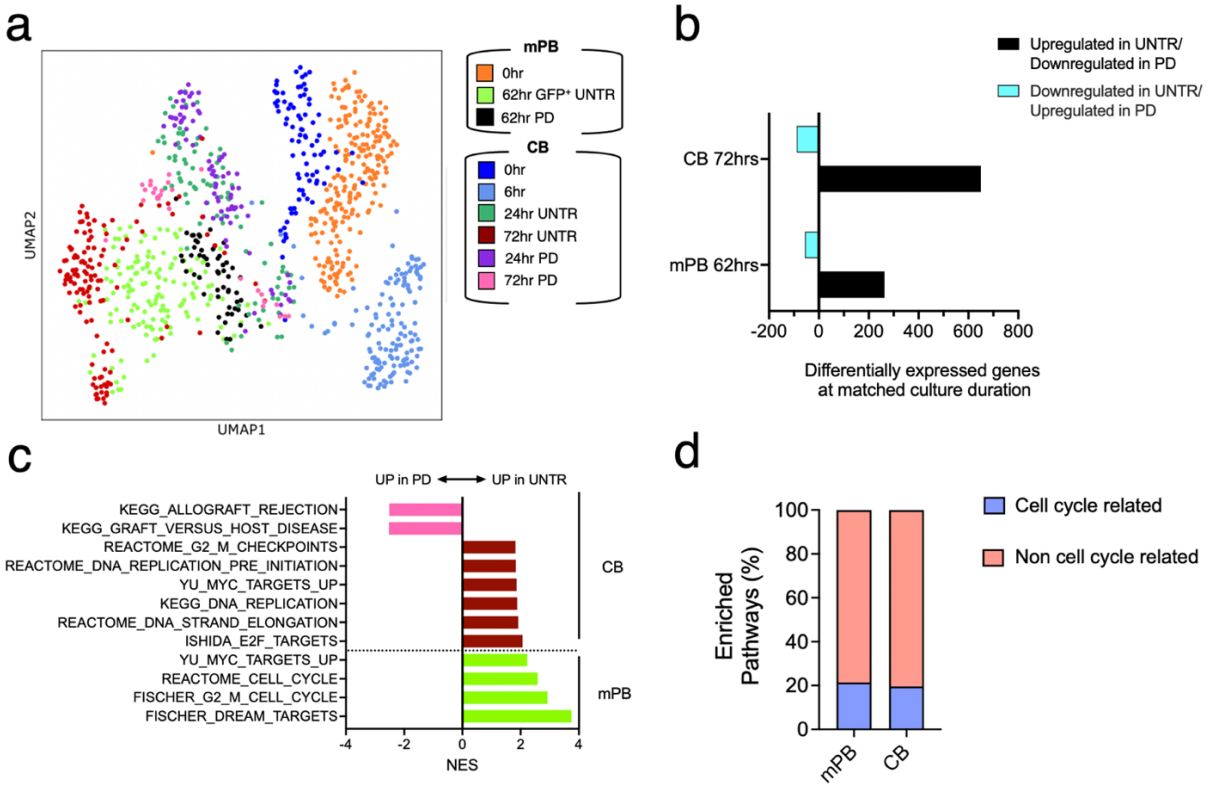

Figure 4.23: mPB and CB LT-HSCs cultured with Palbociclib show shared modulation of predominantly mitotic related processes

a. UMAP visualisation created from 3643 HVG using 920 single cells. UNTR/PD treated mPB LT-HSCs were cultured in 62 hr gene therapy protocol including lentiviral transduction. PD treated CB LT-HSCs were cultured in differentiation facilitating conditions for either 24 hr or 72hr. UNTR CB LT-HSCs were cultured in differentiation facilitating conditions for 0, 6, 24 and 72 hr. UNTR mPB LT-HSCs equivalent to Fig.4.1. UNTR/PD treated CB LT-HSCs equivalent to Fig.4.22.

b. Number of genes differentially expressed (*FDR < 0.05*) as a result of PD treatment at matched culture durations within CB LT-HSCs cultured for 72 hr and mPB LT-HSCs cultured for 62 hr.

c. Pre-ranked GSEA analysis of enriched c2 curated pathways in UNTR/PD treated CB LT-HSCs cultured for 72 hr and UNTR/PD treated mPB LT-HSCs cultured for 62hr. All enriched (*FDR < 0.01*) pathways shown.

d. Proportion of all enriched (*FDR < 0.01*) KEGG and Reactome pathways directly involved in cell cycle related processes (including DNA synthesis, DNA repair and mitotic related processes) from Cluego analysis (97 total enriched pathways in CB; 56 total enriched pathways in mPB).

mPB LT-HSCs treated with PD in gene therapy conditions were integrated to CB LT-HSCs treated with PD in differentiation facilitating conditions. Integration of the two datasets (mPB and CB) was performed using ComBat in scanpy and following integration the PCA space was calculated based on selected HVGs. 15 nearest neighbours were found and a UMAP was created for visualisation purposes (Methods 3.8.3). As expected, mPB and CB LT-HSCs broadly cluster based on culture time-point and 62 hr PD treated mPB LT-HSCs cluster with 24 hr PD treated CB LT-HSCs (Fig.4.23a). This broadly indicates that PD has a similar effect on the transcriptome irrespective of culture conditions and cell source. Differential gene expression analysis was then performed whereby LT-HSCs treated with PD were compared to UNTR LT-HSCs at matched time-points (for CB only 72 hr culture duration tested). PD treatment causes differential expression of 319 genes in mPB LT-HSCs and 739 genes in CB LT-HSCs and in both HSC sources results in predominantly downregulated gene expression with respect to UNTR cells (Fig.4.23b). Pre-ranked GSEA analysis (c2 curated pathway enrichment) was then performed to analyse differences in biological pathway activation as a result of pharmacological G_1 arrest in culture. As expected, results demonstrate significant upregulation of biological pathways predominantly involved in cell cycle progression and also enrichment to MYC targets pathways (*FDR < 0.01*) when compared to PD treated cells (Fig.4.23c). Cluego biological pathway analysis was then performed using differentially expressed genes between PD and UNTR cells to corroborate findings. Analysis revealed that > 75% of all enriched (*FDR < 0.01*) KEGG and Reactome pathways were directly related to cell cycle processes (including DNA synthesis, repair and mitosis programs) in both CB and mPB (Fig.4.23d). Non cell cycle related pathways which were differentially expressed between 62 hr PD treated mPB and 62 hr UNTR mPB cells (12/56) included pathways related to RHO GTPase related signalling, RNA splicing/metabolism and HSF1 activation (data not shown). Non cell cycle related pathways differentially expressed between 72 hr CB LT-HSCs treated with PD and 72 hr UNTR CB cells (19/97) also include RNA splicing/metabolism, RHO GTPase activation and unlike observed in mPB, differential expression of kinesins and oxidative phosphorylation pathways (data not shown).

These data suggest that changes acquired from reversible CDK6 inhibition by PD are predominantly shared between mPB and CB LT-HSCs in culture despite contrasting media conditions, and as expected are largely related to cell cycle progression related processes.

4.3.4 Cell size and mitochondrial activity increase in a cell cycle independent manner over culture

Data presented so far throughout this chapter demonstrate that the majority of the transcriptional changes associated with LT-HSC activation in culture are independent of cell cycle progression and are instead dependent on the duration of the mitogenic signal. We next wanted to tackle whether hallmarks of HSC activation are independent of progression past early G_1. Cell size measurements were performed using live cell microscopy and changes in mitochondrial membrane potential were determined using the cationic dye TetraRhodamine Methyl Ester (TMRM) which accumulates in active mitochondria. Experiments were performed in CB LT-HSCs cultured in differentiation facilitating conditions and ideally, in absence of timing limitations would have been repeated in mPB LT-HSCs cultured in gene therapy media.

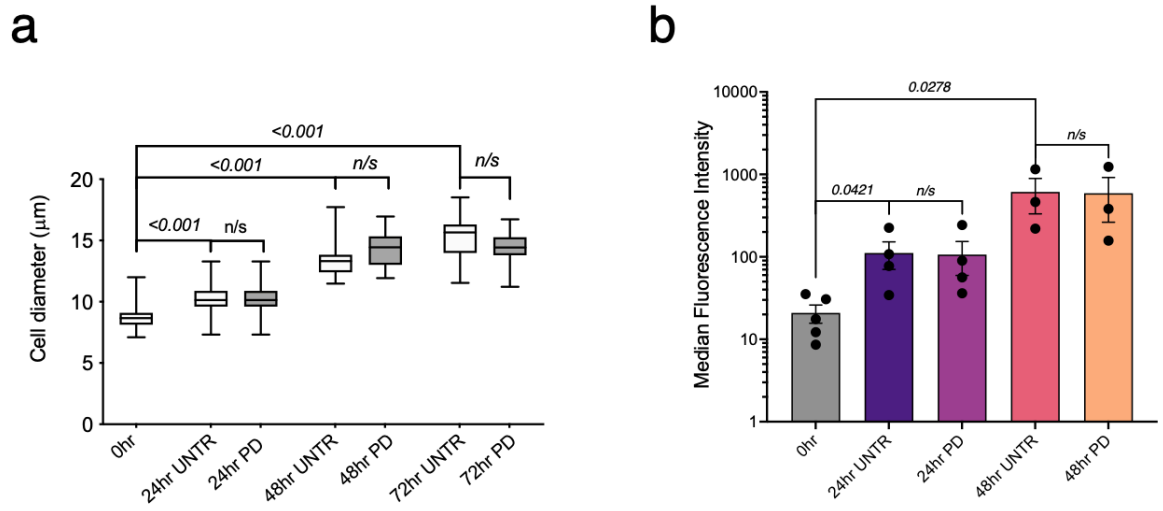

Figure 4.24: LT-HSCs arrested in cell cycle progression show similar cell size and mitochondrial activation at matched culture time-points when compared to UNTR cells.

a. Cell size as measured by cell diameter (μm) over time and in response to PD treatment (n=2 independent CB experiments; 25 cells per condition). Two way ANOVA for multiple comparisons and significant p-values indicated.
b. TMRM staining in CB LT-HSCs at time-points indicated and normalised to unstained control (representative of n=5 independent CB donor samples). Paired t-test performed and significant values indicated.

PD treated CB LT-HSCs increase in cell size over culture similarly to UNTR LT-HSCs, as quantified by cell diameter (μm) measurements (Fig.4.24a). Mitochondrial activity also significantly increased over CB LT-HSC culture and similarly to cell size and this is again independent of progression past early G_1 (Fig.4.24b). These data indicate that LT-HSCs are increasing in size and becoming more metabolically active over culture irrespective of cell cycle progression.

4.3.5 Cell cycle progression does not drive loss of long-term repopulation capacity in culture

To formally test if progression past early G_1 contributes to loss of long-term repopulation capacity in tested culture conditions, CB LT-HSCs (Table 4.3) and mPB $CD34^+/CD38^-$ cells (Table 4.4) were cultured for indicated durations in either UNTR conditions or in the presence of PD and were subsequently transplanted into immunocompromised mice at matched cell doses (Fig.4.25a). Of note, experiments involving UNTR cells cultured for 24 hr and 72 hr time-points were performed at the same time as those presented in Fig.4.8.

Hours of culture	Cell Dose	Number of tested mice	Number of engrafted mice	Condition
24hr	300	12	12	UNTR
24hr	300	13	13	PD
24hr	50	12	4	UNTR
24hr	50	11	5	PD
24hr	10	8	0	UNTR
24hr	10	7	1	PD
72hr	700	14	13	UNTR
72hr	700	15	12	PD
72hr	300	12	9	UNTR
72hr	300	15	14	PD
72hr	50	13	4	UNTR
72hr	50	12	6	PD

Table 4.3: Human CB LT-HSC cell dose and mouse information for primary *in vivo* transplantation studies in differentiation facilitating media including PD treatment. CB LT-HSC cell dose (representative of 0 hr at the time of sort) after culture in differentiation facilitating conditions, the number of mice tested and the number of mice showing human engraftment 18 weeks post transplantation are indicated. Mice were considered engrafted if human cells (hCD45++ &GlyA+) ≥0.01% of Singlets and ≥30 cells present in any lineage determination gate. 24 hr: 32 mice UNTR 31 mice PD, n=3 experiments; 72 hr: 39 mice UNTR, 42 mice PD n=5 experiments. UNTR mice equivalent to Table 4.1.

Hours of culture	Cell Dose	Number of tested mice	Number of engrafted mice	Condition
62hr	55000	12	9	UNTR
62hr	55000	13	12	PD
62hr	25000	13	12	UNTR
62hr	25000	11	11	PD
62hr	8000	4	4	UNTR
62hr	8000	4	3	PD

Table 4.4: Human mPB $CD34^+/CD38^-$ cell dose and mouse information for primary *in vivo* transplantation studies in gene therapy protocol conditions including PD treatment. Human mPB CD34+/CD38- cell dose after culture in gene therapy conditions (representative of 0 hr at the time of sort), the number of mice tested and the number of mice showing human engraftment 18 weeks post transplantation are indicated. 62hr: 29 mice UNTR 28 mice PD n=3 experiments. UNTR mice equivalent to Table 4.2.

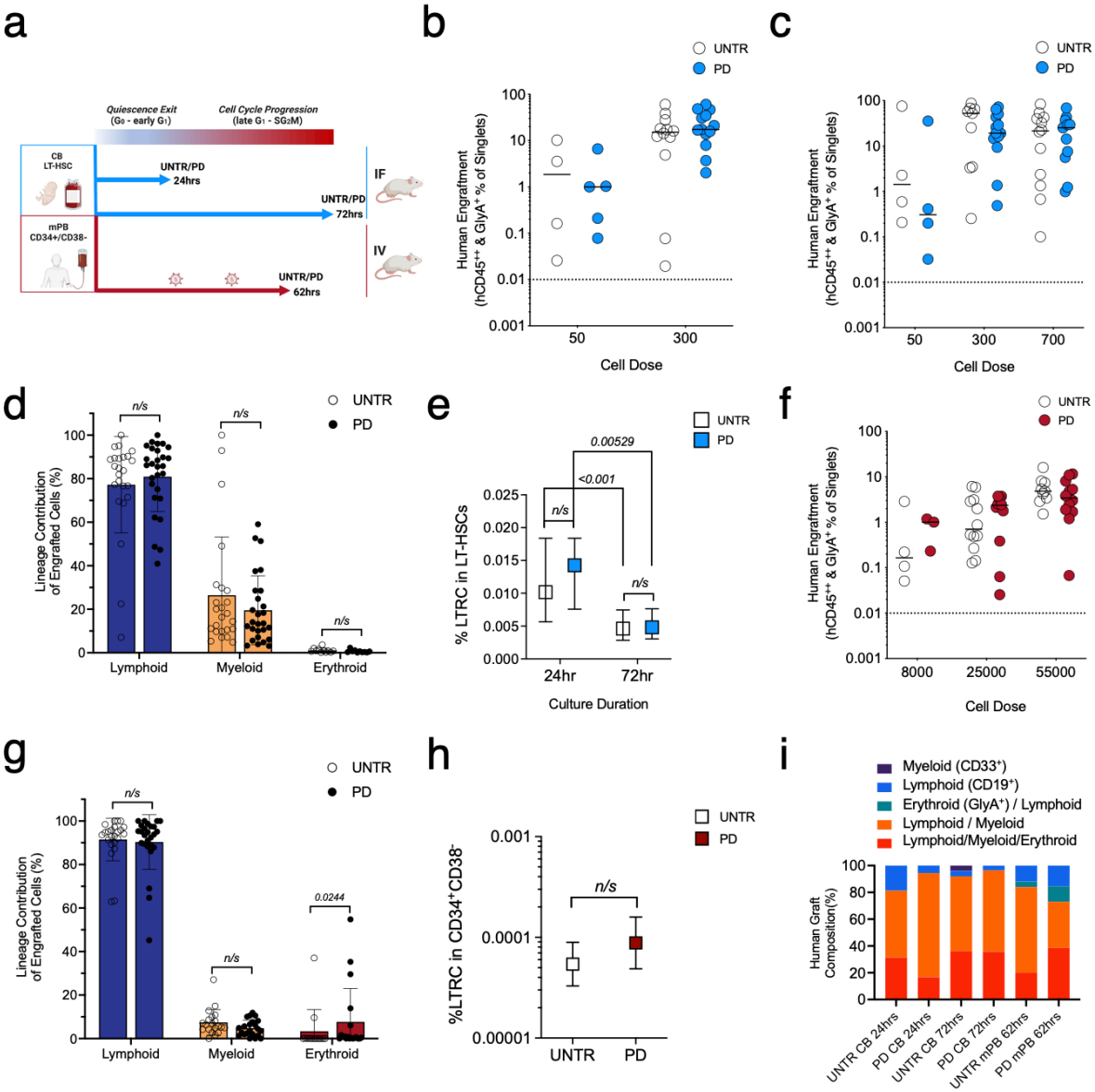

Figure 4.25: Cell cycle progression does not drive long-term repopulation capacity loss in cultured mPB and CB HSCs.

a. Experimental workflow of CB LT-HSCs and mPB $CD34^+/CD38^-$ culture and transplantation. mPB $CD34^+/CD38^-$ cells were exposed to the 62 hr gene therapy protocol including two hits of lentiviral transduction. CB LT-HSCs were cultured in differentiation facilitating conditions. CB LT-HSCs were transplanted by intrafemoral (IF) injection. mPB $CD34^+/CD38^-$ cells were transplanted by intravenous (IV) injection. Mice sacrificed at 18 weeks.

b. Human graft size ($CD45^{++}$ & $GlyA^+$ % of Singlets) from injected bone of mice engrafted with CB LT-HSCs after UNTR/PD culture for 24 hr. Engrafted mice shown. Line at median. Dotted line indicates engraftment threshold (≥30 cells in $CD45^{++}$ & $GlyA^+$ gate and graft size ≥ 0.01% of singlets).

c. Human graft size from injected bone of mice engrafted with CB LT-HSCs after UNTR/PD culture for 72 hr. Engrafted mice shown.

d. Lineage contribution of mice engrafted with UNTR/PD treated CB LT-HSCs cultured for 72 hr. Each datapoint represents % contribution of respective lineage from individual mouse. Colony defined by ≥30 cells in ($CD45^{++}$ & $GlyA^+$ gate) + ≥30 cells in a determined lineage gate. Mann Whitney U-test performed at significance indicated.

e. Human HSC frequency calculated from mice injected with UNTR/PD treated CB LT-HSCs cultured for 24 hr and 72hr. Chi-squared test performed to test for difference between groups and significance indicated. % long-term repopulating cell in the transplanted population (%LTRC).

f. Human graft size from bone marrow of mice engrafted with UNTR/PD treated mPB $CD34^+/CD38^-$ cells cultured for 62 hr in gene therapy conditions. Engrafted mice shown.
g. Lineage contribution of mice engrafted with UNTR/PD treated mPB $CD34^+/CD38^-$ cells cultured for 62 hr. Mann Whitney U-test performed and significant values indicated.
h. Human HSC frequency calculated from mice engrafted with mPB $CD34^+/CD38^-$ cells after 62 hr culture. Chi-squared test performed and significance indicated (% LTRC in 62 hr PD vs 62 hr UNTR *p = 0.180*).
i. Lineage output as proportion of total engrafted mice across conditions and HSC sources. (Fishers Exact test performed and no significant values found; mPB My/Lym PD vs UNTR *p = 0.0571*, mPB Ery/My/Lym PD vs UNTR *p = 0.2203*).

CB LT-HSCs cultured for 24 hr (n=32 mice UNTR, n= 31 mice PD) and 72 hr (n=39 mice UNTR, n=42 mice PD) in the PD condition show no difference in human graft size (Fig.4.25b+c) and lineage output (data not shown for 24hr; Fig.4.25d for 72hr) at 18 weeks post transplantation when compared to the UNTR condition. An LDA was performed to calculate the HSC frequency of the transplanted population, taking into account the number of engrafted mice and the total number of mice used. This yielded no difference in the estimated HSC frequency between cells pharmacologically arrested in G_1 and UNTR cultured cells at 24 hr (*p=0.462*) and 72 hr (*p=0.893*) (Fig.4.25e). Data suggests that cell cycle progression does not drive loss of long-term repopulation capacity in CB LT-HSCs cultured in differentiation facilitating conditions. In fact, the HSC frequency of PD treated cells at 72 hr of culture is significantly lower (approximately 3 fold; *p=0.00529*) than that of cells also arrested in cell cycle progression at 24 hr of culture (Fig.4.25e) further supporting this conclusion.

Similarly to experiments obtained in CB, mPB $CD34^+/CD38^-$ cells in the PD condition cultured in a 62 hr gene therapy protocol (including two hits of lentiviral transduction) showed no significant difference in graft size (Fig. 4.25f) at 18 weeks post transplantation when compared to the UNTR condition (n= 29 mice UNTR, n=28 mice PD). Although, a significantly higher proportion of erythroid cells were observed in grafts from mPB cells cultured with PD with respect to the UNTR condition (Fig.4.25g; *p=0.0244*). An LDA was performed to assess the HSC frequency of the transplanted population and no significant difference was observed between the PD and UNTR condition at 62 hr (Fig.4.25i). In mice transplanted with CB LT-HSCs tri/bilineage grafts (containing myeloid, lymphoid and erythroid cells) were similarly distributed across animals in both the PD and UNTR condition (Fig.4.25i). In mice transplanted with mPB $CD34^+/CD38^-$ cells a trend for increased tri-lineage grafts was observed in the PD condition at the expense of My/Lym grafts (My/Lym PD vs UNTR *p=0.0571*) when compared to the UNTR condition (Fig.4.25i).

To confirm our results in CB are not attributed solely to conditions which facilitate differentiation, CB LT-HSCs were either UNTR or PD treated and cultured in gene therapy media (in the absence of a

lentiviral vector) for 62 hr and injected into immunocompromised mouse models (n=10 mice UNTR, n=10 mice PD) at indicated doses (Table 4.5).

Hours of culture	Cell Dose	Number of tested mice	Number of engrafted mice	Condition
62hr	700	2	2	UNTR
62hr	700	3	3	PD
62hr	300	4	4	UNTR
62hr	300	3	3	PD
62hr	50	4	4	UNTR
62hr	50	4	1	PD

Table 4.5 : Human CB LT-HSC cell dose and mouse information for primary *in vivo* transplantation studies in gene therapy protocol conditions including PD treatment. Human CB LT-HSC cell dose (representative of 0 hr at the time of sort) after culture in gene therapy conditions (in absence of lentiviral transduction), number of mice tested and number of mice showing human engraftment 18 weeks post transplantation is indicated. 62 hr: 10 mice UNTR, 10 mice PD, n=1 experiment.

No differences in the graft size (Fig.4.26a) or lineage composition (Fig.4.26b) were observed in transplanted LT-HSCs cultured with PD compared to UNTR cultured LT-HSCs at matched doses. An LDA could not be performed in this study because an insufficient number of mice were un-engrafted.

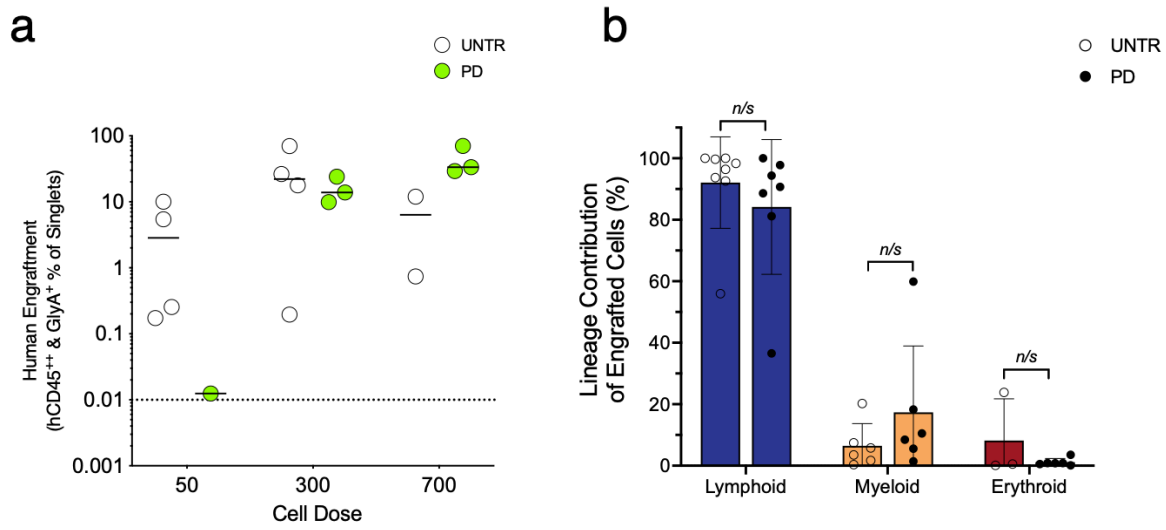

Figure 4.26: CB LT-HSCs arrested in cell cycle progression during 72 hr culture show comparable graft size and lineage output to cells progressing through the cell cycle.

a. Human graft size ($CD45^{++}$ & $GlyA^{+}$ %of Singlets) from bone marrow of UNTR or PD treated CB LT-HSCs cultured for 62 hr in gene therapy conditions in the absence of lentiviral transduction. Mice were transplanted by intravenous injection. Only engrafted mice shown. Line at median.
b. Lineage contribution of mice engrafted with UNTR or PD treated CB LT-HSCs cultured for 62hrs. Colony defined by ≥30 cells in ($CD45^{++}$ & $GlyA^{+}$ gate) + ≥30 cells in a determined lineage gate. Erythroid (GlyA+), Myeloid (CD33), Lymphoid (CD19). Mann Whitney U-test performed and no significant values determined.

To investigate the impact of inhibiting cell cycle progression on self-renewal capacity specifically, secondary transplants were performed in NSG-SGM3 animals. Human $CD45^{++}$ cells from primary mouse bone marrow originally engrafted with 72hr cultured CB LT-HSCs were first sorted and then transplanted into NSG-SGM3 recipients by IF injection. Of note, CB secondary transplantations were only performed for the 72 hr time-point of CB LT-HSC culture (UNTR/PD; differentiation facilitating conditions only). To minimise loss of HSCs during cell sorting for mPB secondary transplants, high doses of primary mouse BM were counted and IV injected with no cell sorting. For both cell sources, matched doses were transplanted between the PD treated condition and UNTR condition (Table 4.6).

Cell Source from Primary Recipi-ents	Cell Dose	Number of tested mice	Number of en-grafted mice	Condition
mPB	2250000	2	2	UNTR
mPB	5000000	2	2	UNTR
mPB	10000000	1	1	UNTR
mPB	100000	4	1	UNTR
mPB	200000	3	2	UNTR
mPB	1000000	4	2	UNTR
mPB	100000	4	3	PD
mPB	200000	3	0	PD
mPB	1000000	3	1	PD
mPB	2250000	1	1	PD
CB	42500	4	0	UNTR
CB	170000	3	1	UNTR
CB	680000	3	1	UNTR
CB	42500	4	1	PD
CB	170000	3	0	PD
CB	680000	3	3	PD

Table 4.6: Human mPB and CB cell dose and mouse information for secondary *in vivo* transplantation studies in gene therapy protocol conditions including PD treatment. Cell dose for secondary transplantation, number of mice engrafted and number of mice tested is indicated. For mPB secondary transplantation experiments, BM from engrafted primary recipients was thawed, counted and injected (IV) in indicated doses based on cell count. For CB secondary transplantation experiments, human $CD45^{++}$ cells were sorted from engrafted primary mice and injected (IF) in three doses. Of note, secondary transplantation for CB experiments was only performed using primary mice transplanted with cells cultured for the 72 hr time-point (UNTR/PD; differentiation facilitating conditions only). For mPB experiments: 16 mice UNTR, 11 mice PD; n=2 experiments. For CB experiments: 10 UNTR and 10 PD; n=1 experiment.

Human graft size was reduced in secondary mice of the PD treated condition (Fig.4.27a) compared to the UNTR condition. However, the frequency of serially-transplantable HSCs in the injected

population was comparable (Fig.4.27b; *p=0.170*). Similar to results in CB, human graft size (Fig.4.27c) and HSC frequency were unchanged (Fig.4.27d; *p=0.809*) between PD and UNTR conditions.

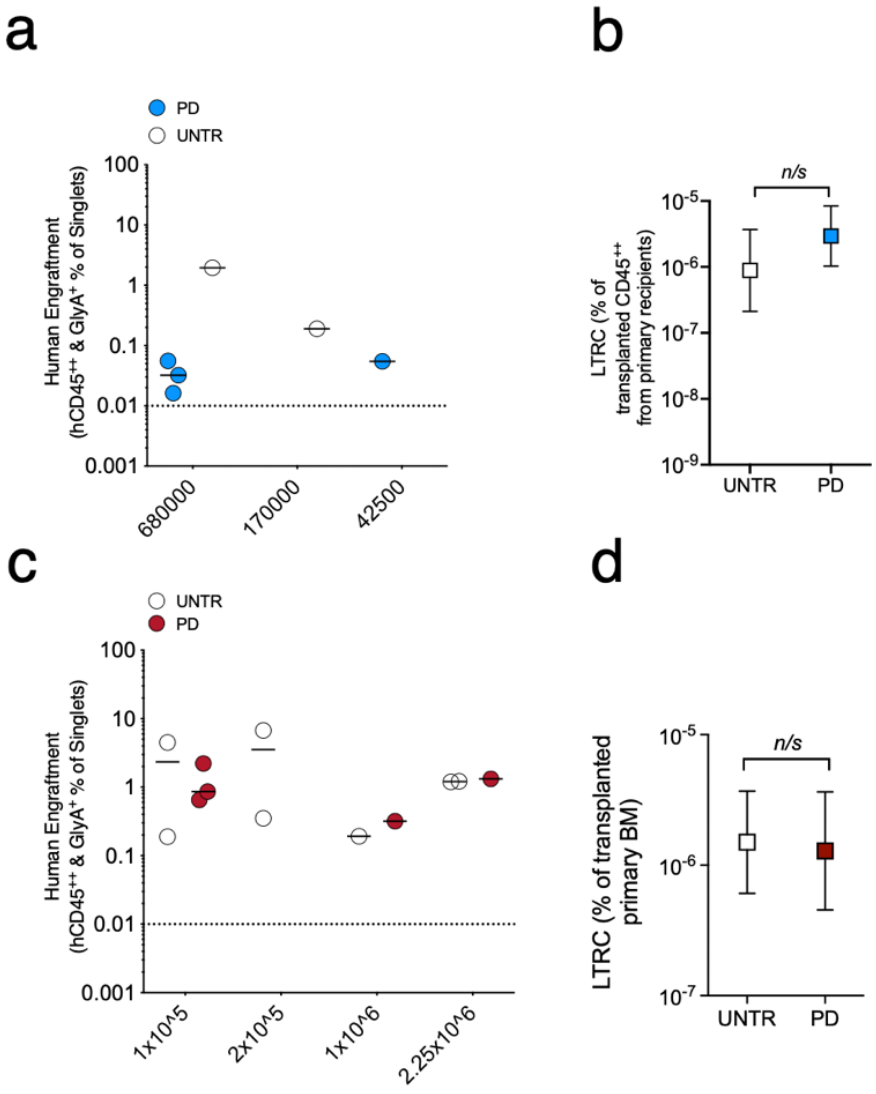

Figure 4.27: Cell cycle progression during the first cell cycle *ex vivo* does not impact the self-renewal capacity of cultured mPB and CB LT-HSCs.

a. Human graft size ($CD45^{++}$ & $GlyA^{+}$ % of Singlets) from secondary mice injected with three doses of human $CD45^{++}$ cells sorted from CB primary recipients. Only engrafted mice shown. Line at median. Dotted line indicates engraftment threshold (≥30 cells in $CD45^{++}$ & $GlyA^{+}$ gate and ≥ 0.01% of Singlets).
b. LDA analysis of the calculated HSC frequency in the transplanted population. Chi-squared test performed and no significant values determined (PD vs UNTR *p=0.170*).
c. Human graft size ($CD45^{++}$ & $GlyA^{+}$ %of Singlets) from secondary mice injected with three doses of unsorted cells from mPB primary recipients. Only engrafted mice shown.
d. LDA analysis of the calculated HSC frequency in the transplanted population. Chi-squared test performed and no significant values determined.

Taken together, this body of work demonstrates that long-term repopulation capacity significantly drops over clinically relevant culture, and that cell cycle progression past early G_1 is not the main driver of such loss in human LT-HSCs (Fig.4.28a + b). In addition, we show that PD treatment over the first cell cycle in culture overall does not drastically alter the lineage output of cells after transplantation. Finally, through secondary transplantation experiments we conclusively demonstrate that cell cycle progression does not impact the self-renewal capacity of cultured LT-HSCs.

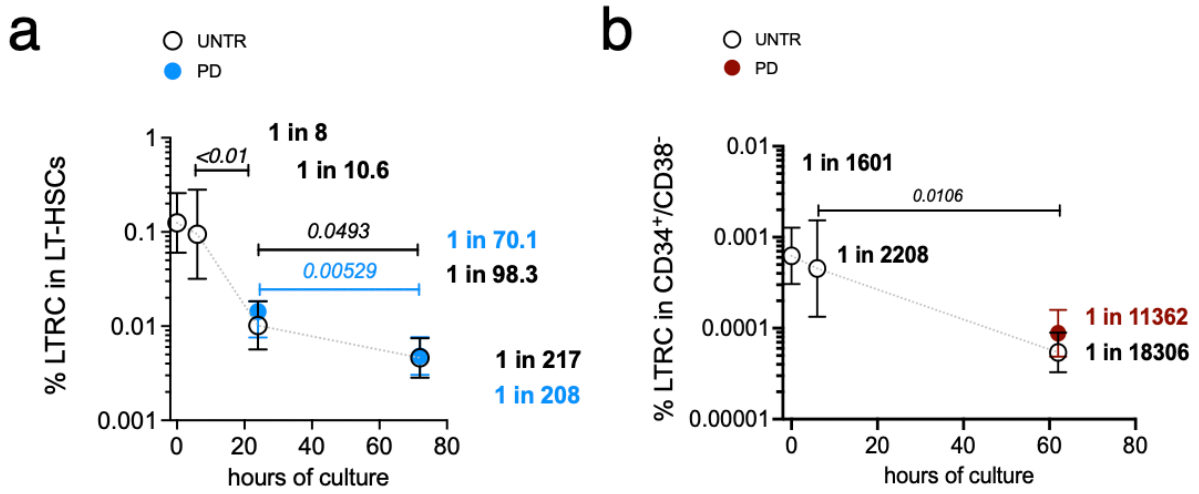

Figure 4.28: Overall dynamics of CB and mPB long-term repopulation capacity loss in clinically relevant culture in either the presence or absence of cell cycle progression.

a. Frequency of long-term repopulation cell (% LTRC) of UNTR/PD treated CB cultured in differentiation facilitating conditions for indicated durations. Calculated by extreme limiting dilution analysis (ELDA). HSC frequency indicated in bold (0 hr UNTR 1 in 8.1, 6 hr UNTR 1 in 10.6, 24 hr UNTR 1 in 90.3, 24 hr PD 1 in 70.1, 72 hr UNTR 1 in 217, 72 hr PD 1 in 208). Chi-squared test performed to test for difference between groups and significance indicated.

b. % LTRC of UNTR/PD treated mPB cultured for indicated durations. Calculated by ELDA. HSC frequency indicated in bold. Chi-squared test performed between groups and significance indicated.

4.3.6 Investigating the effects of cell cycle progression on HSC differentiation

The link between the length of the G_1 phase of the cell cycle and stem cell differentiation has long fascinated scientists and has been extensively studied (reviewed in Dalton and Coverdell, 2015; Orford and Scadden, 2008). A short G_1 phase is an inherent characteristic of pluripotent ESCs and lineage commitment is associated with an increased length of G_1 (reviewed in Boward et al., 2016). Given the link across other stem cell systems between cell cycle kinetics and cell fate I lastly wanted to put our

results into context, building on these ideas and investigating the dependency of cell cycle progression for HSC differentiation specifically. To answer this question to a reasonable degree available from our current datasets, GSVA scores were generated from common-myeloid progenitor (CMP), granulocyte-monocyte progenitor (GMP) and myeloid-erythroid progenitors (MEP) populations of human HSC differentiation (Laurenti et al., 2013). The GSVA scores of cultured mPB and CB LT-HSC populations from scRNA-Seq experiments presented in this thesis were then calculated with respect to the differentiation signatures and visualized (Fig.4.29a). Culture conditions promoting HSC differentiation *in vitro* are biased towards the generation of mostly mature myeloid cell lineages although also support erythroid colonies. Very little supported lymphoid progenitors/mature lymphoid cells are produced and therefore Myeloid-Lymphoid Progenitor (MLP) signatures were not analysed.

Differentiation signatures increase during 72 hr of CB LT-HSC culture with a dynamic similar to the Late Up pattern observed previously (Fig.4.29a, Fig. 4.11). At 24 hr of culture the GMP score is unchanged between UNTR and PD treated CB LT-HSCs while CMP and MEP signatures show significantly increased activation in UNTR cells compared to early G_1 arrested cells (Fig.4.29a). At 72 hr of culture, UNTR cells show significantly higher expression of differentiation signatures compared to PD treated cells for MEP, CMP and GMP signatures (Fig.4.29a). It should be noted that in the presence of PD, differentiation signatures still increase relative to the 0 hr time-points. It is interesting that across CMP, MEP and GMP signatures the degree of upregulation between PD treated and UNTR CB LT-HSCs is increased only after the 24 hr time-point (corresponding to quiescence exit). These data demonstrate that overall, lineage specification signatures are upregulated to a lower degree for cells arrested in early G_1 suggesting that lineage specification is somewhat diminished or delayed in absence cell cycle progression.

Inspection of genes significantly upregulated by progression through the cell cycle in LT-HSCs cultured for 72 hr with respect to PD treated cells showed convincing GMP lineage specification genes including Proteinase 3 *PRTN3*, the elastase *ELANE* and Cathepsin G *CTSG* (data not shown), all of which are expressed by classical monocytes with neutrophil-like characteristics and mature neutrophils (Yáñez et al., 2017). Triphosphate isomerase (*TPI1*) is an enzyme involved in glycolysis and an MEP signature gene. Of interest, *TPI1* is also only upregulated upon progression through the cell cycle by 72 hr CB LT-HSCs. Interestingly, *TPI1* is indispensable for erythropoiesis with triphosphate isomerase deficiency characterised by chronic haemolytic anaemia (Conway et al., 2018).

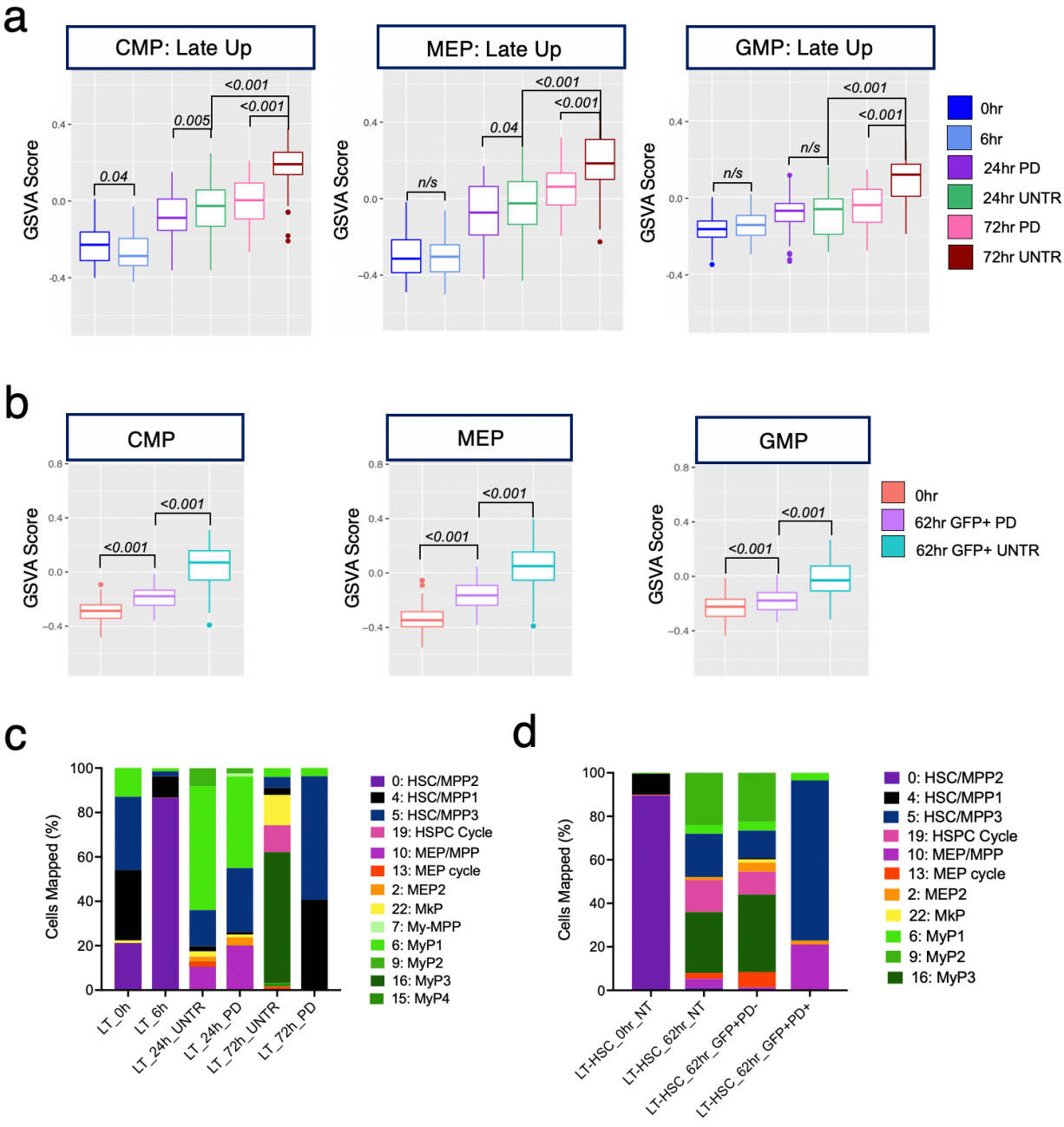

Figure 4.29: Lineage specification programs are upregulated during *ex vivo* culture and show significant downregulation in the absence of cell cycle progression.

a. GSVA score of CMP (left), MEP (middle) and GMP (right) signatures in UNTR/PD treated CB LT-HSCs cultured in differentiation facilitating conditions for 72 hr. Paired t-test performed and significant values indicated (boxplots representative of GSVA score from single cells. 0 hr: n=85, 6 hr: n=134, 24 hr UNTR: n= 58, 24 hr PD: = 80, 72 hr UNTR: n=84, 72 hr PD: n=27).

b. GSVA score of CMP (left), MEP (middle) and GMP (right) signatures in UNTR/PD treated mPB LT-HSCs cultured in gene therapy conditions (including lentiviral transduction) for 62 hr. Paired t-test performed and significant values indicated (boxplots representative of GSVA score from single cells 0 hr: n=218, 62 hr UNTR: n=143, 62 hr PD: n=57).

c. Proportion of single CB LT-HSCs cultured in differentiation facilitating conditions which map to indicated cluster on multi-tissue landscape of differentiation from 118,783 single $CD19^{-}CD34^{+}$ cells (Mende et al., 2022). Clusters 1,3,14, 23,24 excluded from analysis.

d. Proportion of single mPB LT-HSCs cultured in gene therapy conditions which map to indicated cluster on multi-tissue landscape of differentiation from 118,783 single $CD19^{-}CD34^{+}$ cells (Mende et al., 2022). Clusters 1,3,14, 23,24 excluded from analysis.

To corroborate our results, we also calculated the GSVA scores of differentiation signatures in mPB LT-HSCs UNTR or cultured with PD in gene therapy protocol conditions. Similarly to results obtained in CB LT-HSCs, CMP, MEP and GMP signatures all display significantly reduced GSVA scores in PD treated LT-HSCs when compared to UNTR 62 hr GFP^+ LT-HSCs and signatures in PD treated cells were significantly increased from the 0 hr time-point (Fig.4.29b). Therefore, it can be deduced that in both CB and mPB a proportion of lineage determination genes are only upregulated upon cell cycle progression.

Next, we used the multi-tissue landscape of differentiation from Mende et al., 2022. Clusters involved in lymphoid cell specification (1,3,18,24 and the non-descript cluster 14) were first removed from the pipeline because this branch of haematopoiesis is not well supported by our culture conditions. Following this, the CB and mPB scRNA-Seq datasets were integrated to the multi-tissue landscape using the seurat package (Stuart et al., 2019). Integration was performed separately for CB and mPB datasets. The leiden cluster labels from the multi-tissue landscape were then predicted on CB/mPB LT-HSCs through nearest neighbour analysis (Methods 3.8.4). CB LT-HSCs cultured with PD for 24 hr show a similar distribution of cells with nearest neighbours in tested leiden clusters compared to UNTR cells progressing through the cell cycle (Fig.4.29c). This data indicates that despite pharmacological G_1 arrest, cells express lineage determination genes to a similar extent. However, 72 hr LT-HSCs treated with PD maintain transcriptional similarity to HSC/MPP clusters 1 and 3 (Clusters 4+5) whilst the majority of cells progressing through the cell cycle at 72 hr of culture show the highest transcriptional similarity to downstream Myeloid Progenitors (MyP3) (Fig.4.29c).

The majority of uncultured mPB single cells show transcriptional similarity to HSC/MPP2 (Cluster 4) with a similar distribution of projected cells observed in 6 hr cultured CB LT-HSCs (Fig.29c+d). 62 hr UNTR mPB LT-HSCs project to several clusters with the highest proportion of mPB LT-HSCs showing transcriptional similarity to MyP3 (Cluster 4) whereas upon PD treatment 62hr mPB LT-HSCs predominantly project to HSC/MPP3 cluster (Fig.29d). Collectively, these analyses suggest that cell cycle progression contributes to the establishment of lineage specification programmes.

Ideally, to better investigate this question I would like to study the effect of cell cycle inhibition by PD treatment on differentiation over extended culture durations (≥7 days). I had begun optimising the required dose and frequency of PD addition to these cultures to maintain reversible G_1 arrest over a 7 day culture period although this work did not come to fruition within the timing limitations of this PhD.

Overall, both CB and mPB LT-HSCs in pharmacological G_1 arrest show reduced lineage specification signatures, project to a higher proportion of HSC/MPP clusters on a landscape HSC differentiation (Mende et al., 2022) and show reduced expression of key lineage specification genes. These data suggest that cell cycle progression facilitates *in vitro* differentiation.

4.3.7 Chapter 3 Summary:

Data obtained in this chapter use reversible CDK6 inhibition to arrest cell cycle progression in LT-HSCs over clinically relevant culture. We demonstrate that arresting progression past early G_1 does not impair the ability of LT-HSCs to increase cell size and mitochondrial activation over culture. scRNA-Seq of LT-HSCs arrested in cell cycle progression demonstrates that the majority of transcriptional change is acquired independently of progression past early G_1. In addition, we show that cell cycle progression does not impact the self-renewal capacity of LT-HSCs and therefore alternate mechanisms drive the loss in long-term repopulation capacity associated with LT-HSC culture. Finally, we examine the role of cell cycle progression on HSC differentiation *in vitro* and show that upregulation of key lineage specification genes in culture is amplified by cell cycle progression.

5. DISCUSSION

Remarkable progress has been made in recent years within *ex vivo* gene therapy to treat a variety of disorders affecting the haematopoietic system. This progress is in part, attributed to our increasing understanding of the haematopoietic hierarchy through the unparalleled resolution offered by single cell sequencing technologies. Current gene therapy protocols target the $CD34^+$ fraction (a heterogenous mix of stem and progenitor cells) with limited understanding of how *ex vivo* modification impacts the underlying biology of a purified HSC population. Furthermore, it is a continuous challenge in the field that despite *ex vivo* amplification of phenotypic HSC numbers, there is a net decline in long-term repopulation capacity. As such, *ex vivo* gene therapy protocols are limited to re-infusion of a reduced number of LT-HSCs than collected at apheresis. Importantly, the extent to which self-renewal capacity is lost in culture is poorly understood and a lottery exists of how many cells still capable of true long-term reconstitution are re-infused back into the patient. Finally, the mechanisms driving such attrition in HSC function *ex vivo* remain unclear.

This thesis is separated into three chapters although the overarching aims of this work are interlinked; to provide an in-depth analysis of human HSC activation over the first cell cycle *ex vivo* and to investigate changes which may drive loss of self-renewal capacity in culture. To deliver on these aims I used two contrasting culture systems of mPB cultured in *ex vivo* gene therapy conditions (Biffi et al., 2013) and CB cultured in differentiation facilitating conditions (Laurenti et al., 2015). Subsequently, scRNA-Seq, *in vitro* and *in vivo* functional assays were performed to assess changes in HSC behaviour during culture.

By using single cell RNA-Seq of 1512 cells from four healthy mobilized peripheral blood donors I first demonstrate that the *ex vivo* modification protocol has a significant impact on the transcriptome of LT-HSC, ST-HSC and $CD34^+$ cells. We reveal that the majority of this transcriptional change is due to the culture composition rather than lentiviral transduction itself and unpick differential molecular regulation occurring within the LT-HSC compartment over culture.

Next, by using single cell RNA-Seq and *in vivo* functional assays in a time resolved manner, I disentangle the kinetics and molecular rewiring associated with loss of stem cell function *ex vivo*. I demonstrate that long-term repopulation capacity is maintained for the first 6 hr of culture, followed by a marked reduction in HSC function corresponding to the timings of cell cycle progression past early G_1. Time-course scRNA-Seq analysis at matched time-points reveals 10010 differentially expressed genes over

the time-course and such loss in HSC function is framed by dynamic transcriptional patterns of activation and repression.

Finally, I test the dependency of cell cycle progression on the two unique features which underpin HSCs; self-renewal and differentiation capacity. Using a LT-HSC specific *in vitro* model system to reversibly arrest progression past early G_1, my data conclusively demonstrates that loss of self-renewal capacity *ex vivo* is independent of cell cycle progression.

Collectively there are interesting findings throughout all results chapters and this discussion will be broken into key results sections for further analysis.

5.1 A subset of uncultured CB LT-HSCs are primed for oxidative phosphorylation

Curiously, prior to *ex vivo* activation and within 0 hr CB LT-HSCs, we identify two subsets separated by pseudotime score (denoted 0hr-EARLY and 0hr-LATE respectively). 0hr-EARLY LT-HSCs enrich to a population of dormant human LT-HSCs (Zhang et al. Nature cell biology, *in revision*) and a subset of multipotent LT-HSCs with slower quiescence exit kinetics identified by our lab (Belluschi et al., 2018). In contrast, 0hr-LATE cells enrich to a signature generated from a population of lineage restricted LT-HSCs (Belluschi et al., 2018) and express significantly increased levels (> 15 log_2FC from 0hr-EARLY) of *CDK6* and key oxidative phosphorylation genes. Elevated levels of CDK6 protein in ST-HSCs permit faster cell cycle re-entry following quiescence (Laurenti et al., 2015), however it has only very recently been shown that CDK6 transcript expression can subset functionally distinct populations within the phenotypic LT-HSC compartment (Kaufmann et al., 2021). More specifically, $CD112^{lo}$ cell surface expression demarcates a population of LT-HSCs with low C*DK6* expression which exhibit transiently restrained or latent repopulation kinetics, permitting superior and more durable long-term repopulation capacity (Kaufmann et al., 2021).

0hr-EARLY and 0hr-LATE LT-HSCs are roughly equally as prevalent within uncultured cells and therefore cannot be equivalent to those identified in Belluschi et al. whereby lineage restricted LT-HSCs represent a rare subset (Belluschi et al., 2018). Furthermore, previous characterisations of distinct subsets within the LT-HSC compartment have not identified any degree of oxidative phosphorylation priming (Belluschi et al., 2018; Kaufmann et al., 2021; Knapp et al., 2018) which is a preferred mechanism of metabolism upon HSC activation (Manesia et al., 2015). Cell cycle status and metabolism are tightly linked such that *in vivo* quiescence is associated with glycolytic metabolism (Simsek et al., 2010;

Takubo et al., 2013) and maintained by the low ROS levels in the hypoxic BM niche (Itkin et al., 2016). Loss of hypoxia is correlated with loss of quiescence (Lo Celso et al., 2009), increased aerobic metabolism and ROS generation (Hinge et al., 2020; Simsek et al., 2010; Takubo et al., 2013), therefore it is compatible to observe a degree of cell cycle priming within cells transcriptionally primed for aerobic metabolism. Taken together, I identify a population of cells which are more efficiently equipped to respond to the increased bioenergetic demands of culture through an abundance of oxidative phosphorylation genes and a higher degree of cell cycle priming. Our results support the increasing body of work demonstrating intrinsic differences within the LT-HSC fraction (Belluschi et al., 2018; Kaufmann et al., 2021; Knapp et al., 2018) which contribute to heterogeneity of the human HSC pool.

5.2 Global changes of HSPC activation in culture and across stem cell sources

The *ex vivo* gene therapy protocol has a predominantly shared impact on the transcriptome of mPB HSPC subsets with over 95% of changed biological pathways shared between $CD34^+$, ST-HSC and LT-HSCs. Highly upregulated pathways are involved in cell cycle progression, cholesterol biosynthesis, oxidative phosphorylation, translation and MYC target activation. This suggests that both HSCs and progenitors are becoming more metabolically active over culture, as supported by increased cholesterol biosynthesis (Xie et al., 2019), translation and protein synthesis (Kruta et al., 2021) in LT-HSCs over limited culture durations ($\leq$ 48hrs). It should be noted that my results from differentiation facilitating culture of CB LT-HSCs also show upregulated transcription, translation, MYC target pathway activation and oxidative phosphorylation, therefore this observation suggests that these pathways are activated in culture irrespective of media composition.

It can be suggested that the commonality of these pathways is due to the shared oxidative environment of culture. However, increased oxidative phosphorylation and mitochondrial translation pathways are also observed during the transition *in vivo* to an intermediary activation state of “Galert” where muscle stem cells are primed to respond to stress (Rodgers et al., 2014). This therefore suggests that upregulated aerobic metabolism is a “hard-wired” feature of cell activation and is in fact not directly associated with the population studied, the oxidative culture environment or the cell source. Whilst the purpose of our study is not to look specifically for common pathways across cell sources or media compositions, it becomes evident that there is an overlap and we speculate that many biological processes changed over activation are not unique to HSCs. Given the impressive range of transcriptional dynamics we observe in the culture of CB LT-HSCs perhaps it is the amplitude of pathway change

which may distinguish cell activation across different HSC subsets and cell sources, and this concept is explored in the next section.

5.3 LT-HSC specific changes during activation in culture

Despite modulating predominantly shared biological pathways, mPB LT-HSCs cultured in gene therapy conditions show slight but significant reductions in the median GSVA score of upregulated pathways involved in MYC target regulation, mitochondrial translation and E2F1 target regulation at 62 hr when compared to ST-HSCs and $CD34^+$ cells. Results obtained also demonstrate that LT-HSCs show a significantly reduced amplitude of gene expression change over culture. LT-HSCs exist in a deeper state of quiescence such that they have a longer kinetics of quiescence exit than ST-HSCs upon *ex vivo* culture (Laurenti et al., 2015). It can therefore be suggested that the increased depth of the LT-HSC quiescent state drives a lower amplitude of gene expression change as a result of culture, translating into reduced biological pathway activation. Therefore, LT-HSCs may initially be more "resistant" to *ex vivo* activation and exhibit a delayed response to gene expression changes as a direct result of slower cell cycle kinetics.

In contrast to ST-HSCs and $CD34^+$ cells, mPB LT-HSCs also regulate a distinct array of transcription factors over culture including JUN, KLF4, and NRF1 networks. The AP-1 family of transcription factors such as c-Jun (*JUN*) and JunB (*JUNB*) are highly expressed in quiescent HSCs and are quickly downregulated upon proliferation (Min et al., 2008). c-Jun controls G_1-S phase cell cycle progression (Schreiber et al., 1999) through proposed mechanisms regulating CDK and Cyclin D1 expression (reviewed in Shaulian and Karin, 2001). The NRF1 network also shows unique upregulation in LT-HSC culture. NRF1 is a master regulator of mitochondrial biogenesis which directly regulates expression COX subunits critical for oxidative phosphorylation (Dhar et al., 2008). It is interesting that these transcription factor networks are only altered in LT-HSCs when exit from quiescence is common in ST-HSCs and increased cell cycling and aerobic metabolism is common to all subsets.

This may suggest that these regulatory networks have stem cell specific functions in the context of HSC activation. JunB controls proliferation in LT-HSCs (Passegué et al., 2004) and maintains correct notch and TGF-β signalling to regulate myeloid differentiation with *JunB* loss leading to aberrant myeloid expansion (Santaguida et al., 2009). Although the roles of c-Jun (*JUN*) and KLF4 networks in regulating HSC self-renewal and differentiation are poorly understood, these regulatory networks have instrumental roles in maintenance of self-renewal capacity in other stem cell systems. c-Jun signalling

impedes differentiation in human embryonic stem cells (ESCs) by occupying enhancers (Li et al., 2019). KLF4 is a central transcriptional regulator of the pluripotent state (Guo et al., 2009; Takahashi and Yamanaka, 2006) with KLF4 downregulation and relocalization to the cytoplasm initiating exit from pluripotency (Dhaliwal et al., 2018). Taken together, we suggest that KLF4 and JUN networks may be downregulated over only LT-HSC culture exclusively as a direct result of the loss of self-renewal capacity observed. In HSCs specifically, NRF1 regulates protein folding stress incurred from the increased mitochondrial biogenesis required upon the transition from quiescence to activation (Mohrin et al., 2015). The exclusive upregulation of the NRF1 network in LT-HSCs may suggest distinct mechanisms regulating careful metabolic activation from the deep glycolytic state of LT-HSC quiescence (Simsek et al., 2010; Takubo et al., 2013).

5.4 Early preservation of self-renewal capacity in culture

By precise kinetic analysis using *in vivo* transplantation approaches, my findings reveal that the loss of long-term repopulation capacity *ex vivo* is not linear with time. Interestingly, in both mPB cultured in *ex vivo* gene therapy conditions and CB cultured in differentiation facilitating conditions, I demonstrate that the HSC frequency of the transplanted population is maintained for the first 6 hr of culture.

The first 6 hr of CB LT-HSC culture in differentiation facilitating conditions is associated with rapid transcriptional rewiring of the quiescent LT-HSC state; 40% of the transcriptional changes accumulated over 72 hr in culture are acquired within the first 6 hr. We observe fast upregulation of transcription, translation and oxidative phosphorylation related pathways with simultaneous downregulation of cell adhesion molecule pathways and integrin signalling. Cell adhesion molecules such as N-cadherins (Hosokawa et al., 2010), E-selectins (Winkler et al., 2012) and integrin signalling (Khurana et al., 2016) regulate HSC maintenance in the BM niche and the continuous downregulation of these pathways from uncultured LT-HSCs suggests that these niche regulatory interactions are progressively destroyed over culture. We also observe sharp downregulation of AP-1 transcription factors associated with maintenance of a quiescent LT-HSC state such as *JUN* (Min et al., 2008) and *FOS* (Okada et al., 1999). Downregulation of the JUN network is consistent with my data in mPB LT-HSCs cultured in the gene therapy protocol and here we pinpoint that this change is likely occurring rapidly during culture.

Strikingly, we also observe transient upregulation of pathways dynamically regulated at the 6 hr timepoint with genes following this expression pattern involved in cell stress, sphingolipid biosynthesis and apoptotic signalling. In comparison to progenitors, HSCs show high expression of stress response pathways such as the UPR (Galen et al., 2014) and the downstream ISR (Galen et al., 2018). This results in

a carefully maintained proteostatic environment balanced through autophagic recycling (reviewed in García-Prat et al., 2017). High basal levels of stress response pathway expression sets superior "quality control" thresholds for LT-HSCs, resulting in preferential apoptosis in response to oxidative, genotoxic and proteostatic damage (Galen et al., 2014; Milyavsky et al., 2010; Xie et al., 2019; Yahata et al., 2011). This "culling" of unfit HSCs ultimately maintains the integrity and longevity of the HSC pool. Transient upregulation of cell stress related pathways such as TNF signalling and apoptotic pathways may therefore serve as a component of this stress response regulatory axis, proving beneficial to maintain long-term repopulation capacity at 6 hr in culture. Indeed TNF-α treatment both *in vitro* and *in vivo* promotes survival exclusively in HSCs whilst driving apoptosis in progenitors (Yamashita and Passegué, 2019). TNF signalling also directly activates the anti-apoptotic cFLIP (*CFLAR*), which shows transient upregulation at the 6 hr time-point. cFLIP mediates the balance between cell survival and apoptosis in many cells (reviewed in Tsuchiya et al., 2015), therefore upregulated *CFLAR* may promote functional LT-HSC survival at 6 hr of culture. Transient downregulation of bile acid synthesis is also observed at 6 hr of culture. Bile acids are an important component of lipid metabolism and when added to *ex vivo* culture have been shown to enhance protein folding and reduce ER stress to improve HSC engraftment in the mouse system (Miharada et al., 2014) with the same mechanism also demonstrated to suppress ER stress in expanding mouse fetal liver HSCs (Sigurdsson et al., 2016). Taken together, transient dynamics permitting increased TNF-signalling coupled with favourable proteostatic stress response signalling may promote survival and protect the LT-HSC population from functional loss at the 6 hr time-point.

However, transient dynamics of expression exist at the 6 hr time-point which may drive subsequent HSC functional loss *ex vivo*. Transient upregulation of the ceramide biosynthesis enzyme, *DEGS1* likely contributes to later loss of HSC function in culture. Ceramide is the dominant component of sphingolipids which constitute cellular membranes such as the ER, mitochondria and lysosomes (reviewed in Hannun and Obeid, 2018) and therefore has a central role in regulating the interconnected metabolic network within LT-HSCs. Upregulation of *DEGS1* in culture directly mediates loss of LT-HSC self-renewal capacity through disruption of lipid homeostasis and underlying proteostatic programs (Xie et al., 2019). Pharmacological inhibition of *DEGS1* by 4HPR acts as a lipostatic stress stimulant, restoring high expression of coordinated cell stress programs and autophagic recycling to preserve LT-HSC capacity *ex vivo* (Xie et al., 2019). Furthermore, we observe transient downregulation of retinoic acid signalling (retinol metabolism), which is highly expressed within long-lived and highly potent dormant HSCs *in vivo* (Cabezas-Wallscheid et al., 2017; Schönberger et al., 2021). Therefore, differentially

expressed genes with transient dynamics of expression at 6 hr of culture may be detrimental to HSC fitness *ex vivo*.

The 6 hr time-point is also associated with bimodal expression of genes which are predominantly involved in transient dynamics over 72 hr culture. Gene expression variability is inversely correlated with our ability to predict the subsequent transcriptional state of a cell (reviewed in Chalancon et al., 2012), such that stably expressed genes can dictate a transcriptional program to maintain pluripotency (Mason et al., 2014) and conversely, expression noise in transcriptional regulators identifies genes which predict imminent cell fate decisions in embryonic development (Hasegawa et al., 2015) and in erythroid progenitor populations (Grün, 2020). Taken together, this broadly suggests that transiently expressed genes at the 6 hr time-point showing bimodality of expression may later become differentially expressed between populations to drive cell fate decisions in culture.

5.5 An adaptation period in culture frames a cell fate decision

Taken together, we propose that the 6 hr time-point encompasses an adaptation period where LT-HSCs are responding to the overwhelming change in external mitogenic stimuli forced by *ex vivo* culture. From our data we cannot speculate whether this adaptation period is detrimental or beneficial for later stem cell functional capacity *ex vivo*. Heightened TNF signalling and ER stress programs may preserve survival of LT-HSCs. Supporting this concept, addition of small molecules which increase activity of the cell stress regulator HSF1 over *ex vivo* culture restores HSC proteostasis and preserves long-term repopulation capacity equivalent to that of uncultured HSCs (Kruta et al., 2021). Indeed, we observe that LT-HSCs cultured for 6 hr show a trend for higher viability in response to the proteostatic stress agent tunicamycin when compared to 24 hr of culture.

However, it is well established that increased survival in culture is not necessarily beneficial for stem cell function and often an amplification of phenotypic HSC numbers is associated with a real term attrition in LT-HSC function (Xie et al., 2019). This adaptation period to culture may therefore not be beneficial for LT-HSC functional capacity *ex vivo*.

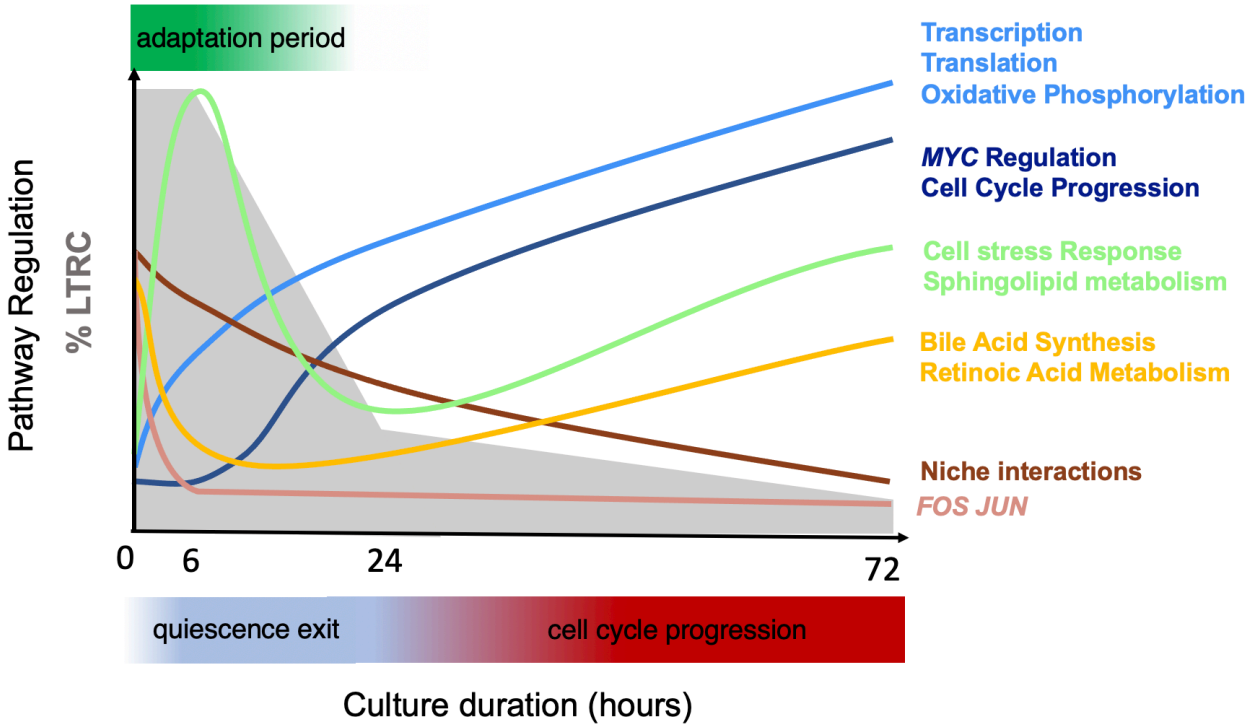

Figure 5.1: Model of molecular and functional regulation of CB LT-HSC activation during 72 hr in culture. The first 72 hr of culture is associated with dynamic transcriptional patterns of gene expression from the uncultured quiescent state. The first 6 hr of culture encompass an adaptation period where LT-HSCs are responding to the overwhelming change induced by the extrinsic environment. The time-point marking the end of this adaptation period is unknown (indicated by colour gradient). The frequency of the long-term repopulating cell (% LTRC; grey colour block scaled by HSC frequency in transplanted population) is maintained for the first 6 hr of culture before sharply dropping (≈ 9 fold reduction) by 24 hr corresponding to the completion of quiescence exit (G_0-early G_1). The % LTRC further reduces from 24 to 72 hr in culture (≈ 2 fold reduction) as LT-HSCs are progressing through the cell cycle until mitosis (late G_1 – M). The exact time of quiescence exit varies within the LT-HSC population (indicated by colour gradient).

Within the 6 to 24 hr transition in culture, we observe a dramatic reduction in the LT-HSC frequency of transplanted CB and this is functionally associated with increased cell size and increased mitochondrial activity. In mPB, this transition is associated with increased cell death and a trend for a reduced ability to respond to proteostatic stress. Taken together, we propose that this adaptation period is followed by a cell fate decision; LT-HSCs must decide whether to undergo apoptosis, lose functional self-renewal capacity or rarely, remain as a LT-HSC. (Fig.5.2).

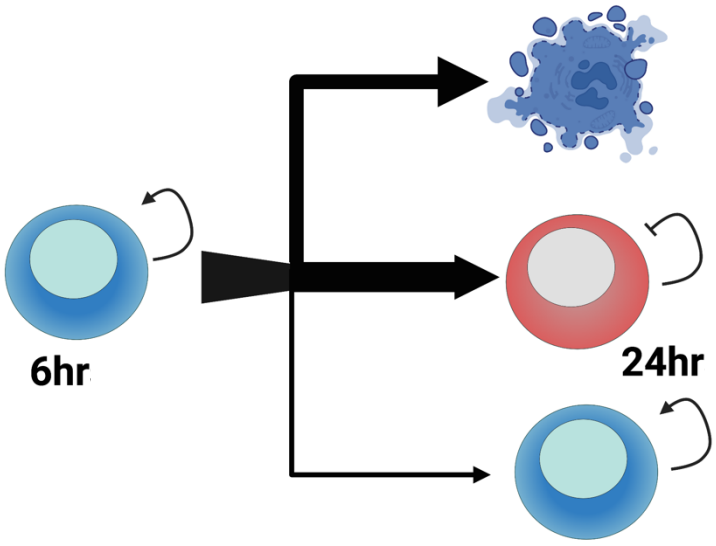

Figure 5.2: Schematic illustration of LT-HSC fate *ex vivo*. Cultured LT-HSCs at the 6 to 24 hr transition may undergo apoptosis (29.1%; top branch), loose self-renewal capacity (60.1%; middle branch) or maintain self-renewal capacity (10.8%; lower branch). Arrow thickness illustrates the calculated proportion of cells which exhibit the indicated fate decision in culture based on long-term *in vivo* transplantation and *in vitro* apoptosis assays. (Proportion of LT-HSCs undergoing apoptosis is calculated from mPB LT-HSCs in gene therapy media whereas the maintenance/loss of self-renewal capacity *ex vivo* is calculated from CB LT-HSCs cultured in differentiation facilitating conditions). Created with Biorender.

We further speculate that the observed increase in cell death may represent a bottleneck in population heterogeneity whereby the instructive signals from *ex vivo* culture effectively select for cells most suited to the *ex vivo* environment. In line with this concept, the 0 hr time-point displays the highest transcriptional heterogeneity as calculated by an entropy based statistic, and this is rapidly lost by the 6 hr time-point. Collectively, loss of the *in vivo* HSC diversity over culture may bias our study of HSCs *ex vivo* and limit the full repertoire of cells for therapy following culture.

5.6 Potential molecular drivers of self-renewal capacity loss *ex vivo*

Our time-course scRNA-Seq dataset matched with functional analysis of repopulation capacity is a key strength of our approach and allows us to pinpoint the molecular changes occurring within the timeframe of the most extreme loss of self-renewal capacity in culture. Furthermore, we reveal dynamic patterns of expression which would be oversimplified by comparison of only quiescent and activated populations; both the amplitude and duration of fluctuations in gene expression may influence cell fate decisions in culture.

At the molecular level loss of long-term repopulation capacity is associated with downregulation of transiently upregulated pathways such as TNF-α signalling, apoptosis signalling (including *CFLAR*) and sphingolipid biosynthesis (including *DEGS1*) coupled with continuous downregulation of HSC niche maintenance factors and continuous upregulation of transcription, translation and oxidative phosphorylation pathways. It is interesting that continuously changed pathways over culture such as upregulated oxidative phosphorylation and downregulated *in vivo* niche regulatory signalling do not exhibit a drastic change in amplitude upon self-renewal loss in culture. This may indicate that this pathway activation/repression is prescribed and does not drive loss of functional HSC capacity only after 6 hr of culture. However, this contrasts recent literature demonstrating that mitochondrial activation itself (Hinge et al., 2020; Vannini et al., 2016) and associated shifts in metabolism (Ho et al., 2017; Loeffler et al., 2019) can directly influence cell fate decisions.

Perhaps threshold levels of key genes exist which may "tip the balance" and drive loss of self-renewal capacity *ex vivo*. Of interest, we also observe sharp activation of *Myc* and subsequent *Myc* target pathways only after 6 hr of culture. Upregulation of *Myc* is associated with loss dormancy and HSC activation *in vivo* (Cabezas-Wallscheid et al., 2017) although to our knowledge, the 0 – 6 hr transition is the first characterisation of "pre-Myc" HSC activation. Mouse HSCs deficient in *Myc* are quiescent and serially transplantable although do not give rise to functional differentiated progeny (Laurenti et al., 2008) suggesting that Myc functionally regulates decisions involving stem cell fate. Recently, MYC has been shown to have a direct role mediating stem cell fate by inhibiting lysosomal recycling of environment sensing cell surface receptors (García-Prat et al., 2021). In turn, this drives HSC activation and subsequent proliferation with a loss of self-renewal capacity (García-Prat et al., 2021). The rapid upregulation of *MYC* between 6 hr to 24 hr is observed with a simultaneous drop in stem cell repopulation capacity and may therefore be indicative of a driving molecular mechanism of self-renewal loss *ex vivo*.

5.7 Towards identification of novel cultured LT-HSC markers

In spite of long-term repopulation capacity loss, a subset of cultured LT-HSCs are thought to persist following *ex vivo* modification in gene therapy protocols, as evidenced by long-term haematopoietic reconstitution up to 13.4 years following therapy (Cicalese et al., 2016). Gene therapy protocols are limited by the fact that this subset cannot be easily identified because cell surface markers used to isolate phenotypic populations change over culture. Identification of a reliable cell surface marker of LT-HSCs after the gene therapy *ex vivo* protocol may be a powerful prognostic indicator of therapy

efficacy and this additional information would allow for cell dose or conditioning regimen modulation, further aided by the increased characterization time afforded through the use of cryopreserved cells in recent trials (Fumagalli et al., 2020). Recently, Integrin-3α in combination with EPCR has been identified to mark LT-HSCs in culture, which represents significant progress (Tomellini et al., 2019) although this has only been tested formally *in vivo* within the context of UM171 mediated expansion in CB.

To prospectively identify markers of LT-HSCs which persist following culture, genes which were differentially expressed in LT-HSCs compared to ST-HSCs and $CD34^+$ cells were identified at 0 hr and 62 hr. mPB LT-HSCs showed higher expression of MHC Class II regulators at both time-points and higher expression of MHC Class I regulators at 62 hr, including significantly increased expression of the invariant chain *CD74* and the alpha chain of HLA-DR, *HLA-DRA* at both time-points. *CD74* and MHC Class II protein expression have recently been used to subset populations of intestinal stem cells (ISCs) with distinct cell cycle kinetics and heightened functional properties in antigen presentation which influences the self-renewal capacity of the ISC pool (Biton et al., 2018). Moreover, a subset of mouse HSCs with elevated MHC II cell surface expression have recently been identified which show reduced cell cycle kinetics and resistance to 5-FU treatment (Li & Williams et al. Blood, *in revision*). I contributed to this study by re-analyzing a dataset from our lab (Belluschi et al., 2018) to identify higher MHC Class II expression within a subset of LT-HSCs with slower cell cycle kinetics and multi-lineage repopulation capacity ($CD34^{lo}CLEC9A^{hi}$).

In the present study we find that a broad antibody for HLA-DR/DP/DQ showed consistently higher expression in cultured LT-HSCs with respect to ST-HSCs and $CD34^+$ cells. Recently, HLA-DR expression has been positively correlated with proliferative capacity in long-term *in vitro* culture of CB LT-HSCs (Knapp et al., 2018) further suggesting merit in investigating MHC Class II molecules as cell surface markers of cultured LT-HSCs. In contrast, cell surface expression of CD74 was not reliably higher in cultured LT-HSCs with respect to ST-HSCs or $CD34^+$ cells (data not shown), perhaps attributed multiple functional roles of this protein both intracellularly and as a cell surface receptor (reviewed in Schroder, 2016). Upon integrating the mPB and CB scRNA-Seq datasets, we also identified higher *CD74* expression in a subset of 62 hr mPB LT-HSCs with the highest transcriptional similarity to earlier time-points in the CB time-course dataset. This suggests that mPB LT-HSCs with high CD74 expression may be associated at the molecular level with increased long-term repopulation capacity. Taken together, *CD74* may represent a candidate transcriptional marker for cultured LT-HSCs. Importantly, it is very possible that the particular markers identified may not reliably mark cultured LT-HSCs, although more broadly our data point to a largely unexplored area in *ex vivo* HSC regulation.

5.8 Cell cycle progression does not drive loss of self-renewal capacity in cultured LT-HSCs

Studies in mouse have conclusively shown that an increased frequency of division *in vivo* reduces repopulation capacity (Bernitz et al., 2016; Foudi et al., 2009; Wilson et al., 2008) with it even suggested that each consecutive division drives a successive loss in self-renewal capacity (Bernitz et al., 2016). Indeed, in the present study we observe a significant drop in the long-term repopulation capacity of mPB and CB HSCs within a timeframe equating to completion of quiescence exit and cell cycle progression, which is further supported by late up (upregulation after 6 hr) expression dynamics of cell cycle progression pathways. However, it has been presumed from current *in vivo* literature (Bernitz et al., 2017; Laurenti et al., 2008; Passegué et al., 2005) that cell cycle progression itself drives the loss of self-renewal capacity associated with HSC culture. However, no study to date has formally tackled this question in a purified HSC subset.

To address this, I used a LT-HSC specific *in vitro* system to reversibly arrest cell cycle progression past early G_1 through treatment with the CDK6 inhibitor PD (Palbociclib). PD treatment inhibits the catalytic activity of CDK4/CDK6-Cyclin D complexes with additional roles indirectly inhibiting CDK2/Cyclin E (Pack et al., 2021) to synergistically arrest cells in G_1. The majority of transcriptional changes acquired over culture and functional hallmarks of *ex vivo* HSC activation, including an increased cell size and increased mitochondrial activity (Ito et al., 2006; Simsek et al., 2010; Takubo et al., 2013) are determined as independent of cell cycle progression. Furthermore, through scRNA-Seq in mPB and CB culture systems, we show that the transcriptional differences between PD treated and UNTR LT-HSCs at matched culture durations are predominantly related to G_2M specific processes. This observation therefore suggests that PD treatment has relatively limited off target effects. It should be noted that upon PD treatment, CDK6 is still able to perform kinase independent functions and can directly regulate the transcription of selective genes such as VEGF and $p16^{INK4A}$ (Kollmann et al., 2013) in addition to EGR1 (Scheicher et al., 2015). Kinase independent functions of CDK6 may therefore have a resulting influence on stem cell function and should be considered. Although importantly, kinase independent functions of CDK6 are independent to the mechanism mediating hyperphosphorylation of Rb and subsequent cell division, therefore conclusions drawn within this study relate directly to cell cycle progression.

The findings generated from *in vivo* studies conclusively demonstrate that pharmacological G_1 arrest during culture does not significantly change the HSC frequency of the transplanted population in comparison to untreated, cycling cells cultured for matched durations. We conclusively demonstrate this

finding for CB cultured in differentiation facilitating conditions for 24 or 72 hr and for mPB cultured in a 62 hr *ex vivo* gene therapy protocol. This demonstrates the validity of our results irrespective of culture conditions or HSC source. Early *ex vivo* studies which have attempted to disentangle the impact of cell cycle progression *ex vivo* on HSC functional capacity have shown a higher proliferative output in G_0 phase $CD34^+$ cells upon long-term culture when compared to G_1 phase $CD34^+$ cells (Gothot et al., 1998). Additionally, CB $CD34^+$ cells exclusively in the G_0/G_1 phases of the cell cycle have been reported to generate grafts upon transplantation, in contrast to cycling counterparts (Glimm et al., 2000). These studies do not unpick the differences between cells in G_0-early G_1 compared to late G_1 and study the $CD34^+$ fraction which is mostly comprised of progenitors, reading out short-term engraftment (Glimm et al., 2000) or *in vitro* proliferative capacity (Gothot et al., 1998) and this does not reflect LT-HSC function. In addition, the lack of *in vivo* short-term engraftment in cycling $CD34^+$ cells may be explained by a reduced homing capacity in cycling cells (Kallinikou et al., 2012; Larochelle et al., 2012). Importantly, this literature broadly demonstrates that the G_0/G_1 fraction of the transplanted population contains the highest frequency of repopulating cells. Our results do not discount this finding, and instead complement this work by showing that arresting cells in G_1 *ex vivo* is not sufficient to maintain long-term repopulation capacity, and therefore, mechanisms other than cell cycle status itself dictate the repopulation status of LT-HSCs *ex vivo*. Overall, our results suggest that we can effectively uncouple cell cycle progression from the 11 fold and 27 fold reduction in long-term repopulation capacity observed within the mPB and CB culture systems respectively.

5.9 Investigating the mechanistic relationship between cell cycle progression and differentiation *ex vivo*

The link between the cell cycle and stem cell capacity has been extensively studied with interest (reviewed in Boward et al., 2016; reviewed in Orford and Scadden, 2008). The G_1 phase is viewed as an important time window for exposure to differentiation signals. ESCs show a truncated G_1 allowing cells to rapidly cycle between DNA synthesis and mitosis with a lengthening of G_1 observed upon differentiation (reviewed in Dalton and Coverdell, 2015). This is also true in neural stem cells where the length of G_1 increases by 4 fold over neurogenesis (Takahashi et al., 1995). In contrast, much less is known within the blood system about the mechanistic interplay between the cell cycle and stem cell fate.

In the present study, we show that LT-HSCs pharmacologically arrested in early G_1 upregulate lineage specification programs over culture, however, to a significantly reduced degree when compared to cycling cells. Our data suggests that cell cycle progression and lineage specification are not inextricably

dependent on each other. Supporting this concept, cytokine stimulation in cultured undivided HSCs leads to upregulation of the myeloid specification factor PU.1 (Mossadegh-Keller et al., 2013). Furthermore, both megakaryocytic and erythroid lineage specification genes are expressed in G_1 phase mouse HSCs *in vivo* (Yang et al., 2017) and *in vivo* transplanted HSCs can convincingly differentiate into restricted progenitors and lose repopulation capacity even before entering S phase (Grinenko et al., 2018).

Our data do however show that key lineage specification genes including *ELANE, PRTN3* and *CTSG* which are required for myeloid differentiation (Yáñez et al., 2017) and triphosphate isomerase *TPI1*, which is necessary for correct erythroid function (Conway et al., 2018) are significantly increased upon cell cycle progression. Cell cycle progression has recently been shown to be required for the generation of HSCs during endothelial to haematopoietic transition (EHT) using an *in vitro* culture system to recapitulate development (Canu et al., 2020). Endothelial cells reversibly arrested in cell cycle progression can later undergo proliferation, although haemogenic identity is lost and replaced with endothelial or mesenchymal fate (Canu et al., 2020). Therefore, cell cycle progression may indeed be required in the context of developmental haematopoiesis for cell fate. Furthermore, if *ex vivo* differentiation itself is dependent on cell cycle progression whereas loss of self-renewal capacity *ex vivo* is not, these results likely suggest that distinct molecular mechanisms regulate differentiation and self-renewal. This further frames the question of whether these are, in fact, independent processes. Within this context, differentiation could be initiated independently of the loss of self-renewal capacity, as proposed in an earlier study in the mouse system (Kent et al., 2008). This is supported by work demonstrating that an undifferentiated status can be maintained *ex vivo* with a net decline in self-renewal capacity (Magnusson et al., 2013). Overall, our results within the present study hint that there may be a degree of cell cycle independent lineage priming but suggest that differentiation is facilitated through progression past early G_1. Importantly, our data warrant further study over a longer timeframe of culture to better disentangle this question.

5.10 Clinical Relevance of Study

The results presented in this thesis have direct relevance for improving current gene therapy protocols. First, the *ex vivo* gene therapy field is experiencing a shortcoming in targeting a sufficient LT-HSC population as evidenced across many clinical trials. Lentiviral transduction within the 62 hr protocol is placed at 24 and 48 hr of culture. To easily improve on the current protocol performed in this thesis, our results suggest that transduction should be moved to earlier time-points with the aim of

successfully targeting a larger LT-HSC fraction *ex vivo*, and this approach is now being employed in *ex vivo* gene therapy clinical trials for Hurler Syndrome (Gentner et al., 2021).

Next, despite holding great clinical promise, gene editing in HSCs remains a challenge partly due to an absence of functioning HDR machinery in non-cycling HSCs (HSCs which are not in S-G_2 cell cycle phases) (reviewed in Ferrari et al., 2021). We conclusively demonstrate that cell cycle progression itself does not drive loss of long-term repopulation capacity *ex vivo*. Therefore, if the gene therapy field adopts a culture system which can better preserve repopulation capacity beyond the 24 hr time-point, we suggest that HDR mediated gene editing is performed in S-G_2 phase cells with active HDR machinery to improve editing efficiency.

Finally, my data more broadly suggests that it is critical to better understand the mechanistic underpinnings of the adaptation period and the consequent cell fate determination point where a dramatic reduction in long-term repopulation capacity is observed between 6 – 24 hr of culture. At this decision point, human HSCs either lose self-renewal capacity, undergo apoptosis, or in rare cases maintain self-renewal capacity. We have further speculated that in addition to a loss of absolute HSC numbers, this bottleneck in HSC function may be effectively selecting for HSCs with characteristics most suited to the extrinsic environment of *ex vivo* culture. Therefore, HSCs which are "unfit" for culture may undergo apoptosis, reducing the heterogeneity of the HSC population. We believe that future research from this work should be carried out with the aim of rescuing the observed bottleneck in HSC functional capacity after culture and this will be explored in the next section (Discussion 5.10).

5.11 Future Directions

To deliver on the proposed objective of rescuing long-term repopulation capacity loss in culture, we suggest modulation of candidate genes/pathways during the first cell cycle *ex vivo*. First, a list of candidates should be generated from the 10010 differentially expressed genes in CB LT-HSCs during 72 hr of culture in differentiation facilitating conditions. Optimal targets may either represent individual genes or entire cell signalling pathways identified by GSVA over 72 hr. As our dataset presents a wealth of potential targets, we suggest investigating candidates which are specifically changed during the 6 to 24 hr window where ≈ 9 fold loss in self-renewal capacity is observed.

Transient upregulation of apoptosis and cell stress response signalling is observed at 6 hr during an initial adaptation period and this is subsequently followed by a drop in cell survival and a reduced ability to resolve proteostatic stress at 24 hr. A comprehensive body of literature demonstrates that

HSCs have higher quality control thresholds (Galen et al., 2014, 2018) with a distinct proteostatic control network (Xie et al., 2019; Kruta et al., 2021; Liu et al., 2020b; Xu et al., 2020), permitting preferential apoptosis in response to cell stress (Galen et al., 2014, 2018; Milyavsky et al., 2010; Yahata et al., 2011). Taken together with our present findings, we suggest that the candidate list should include key regulators of cell stress response pathways (such as the UPR and downstream ISR) and cell survival/apoptosis pathways. It is tempting to speculate that modulation of cell stress response and cell survival networks would shift the balance towards HSC maintenance during the 6 to 24 hr culture window.

Once a comprehensive list of candidate genes/pathways is established, individual targets will be systematically screened by gain of function (GOF) and loss of function (LOF) studies *ex vivo* to investigate potential changes to HSC survival, differentiation capacity and self-renewal capacity. Crucially, we suggest that modulation of candidates should be performed before quiescence exit (before 24 hr). Commonly used genetic modification strategies mediated by lentiviral vectors cannot be used during this early time window of *ex vivo* culture and CRISPR-mediated editing would also be ineffective. Therefore, small molecule inhibitors and transient mRNA overexpression through electroporation should instead be used for the manipulation of candidates. mRNA electroporation in human HSCs has been shown to be effective following a total culture duration of 72 hr in gene therapy conditions, with significantly increased protein expression detected 24 hr after delivery (Omer-Javed et al., 2022), therefore this strategy should allow for candidate interrogation at early time-points of culture.

As long-term repopulation capacity can only be analysed in transplantation experiments spanning ≥16 weeks of study, it is suggested that initial screening of HSC self-renewal capacity is evaluated by *in vitro* serial replating assays and cell survival assays. *In vitro* serial repopulation can be used as a surrogate for long-term repopulation capacity and this technique was employed to discover maintained repopulation capacity in human HSCs cultured with 4HPR (Xie et al., 2019). Candidates showing a response in initial testing should then be subject to further mechanistic examination and long-term *in vivo* transplantation approaches should be used to characterise the potential changes to HSC functional output. It can be speculated that modulation of a limited number of targets during the first cell cycle *ex vivo* may show a benefit to long-term repopulation capacity and be clinically implementable in gene therapy trials. Further study should then determine whether candidates show a benefit to the quality of the long-term graft in pre-clinical models of HSC *ex vivo* gene therapy. Assessment of graft quality would include measuring graft size, graft multipotency (contribution of erythroid, myeloid and lymphoid cells to graft) and persistence of the graft (engraftment in secondary animals). Finally, given

that gene editing strategies hold great promise for the future of HSC *ex vivo* gene therapy (reviewed in Ferrari et al., 2021) we suggest investigating whether the candidate strategies are compatible with CRISPR/Cas9 HDR-mediated gene editing in human HSCs using the *ex vivo* barcoding strategy, Bar-Seq (Ferrari et al., 2020). When paired with *in vivo* transplantation approaches, Bar-Seq allows for simultaneous assessment of gene editing efficiency and graft polyclonality (number of HSCs contributing to engraftment).

Importantly, the approaches suggested to modulate candidates have several benefits. Firstly, both small molecule inhibitors and mRNA overexpression technology permit rapid manipulation of targets which is determined as crucial from the present study. Secondly, small molecule inhibitors can be used to target entire cell signalling pathways and mRNA overexpression technology permits modulation of several mRNAs at once, therefore providing broad targeting of a pathway. Finally, there is little clinical risk of these strategies because no insertion of DNA will be performed and approaches will not impact gene/pathway expression after the *ex vivo* culture protocol. Importantly, regardless of the results obtained from these future studies, this avenue of work will provide a fundamental assessment of how HSC fitness and cell fate decisions are regulated *ex vivo*.

6. FINAL CONCLUSIONS

Altogether, data presented in this thesis comprehensively characterises single HSC activation *ex vivo* over an entire cell cycle in a time resolved manner. I identify significant loss in HSC long-term repopulation capacity over an *ex vivo* gene therapy protocol aimed to preserve self-renewal capacity. Furthermore, I discover that such loss in HSC function is in fact, not linear with time and preceded by an adaptation period where LT-HSCs are responding to the overwhelming change in extrinsic environment mediated by instructive cell culture signals. LT-HSC functional loss *ex vivo* is correlated with the timings of completed quiescence exit and framed by dynamic transcriptional patterns of gene expression. Finally, I formally tackle the dependency of self-renewal capacity loss in culture to cell cycle progression. Contrary to presumed from current literature, my data demonstrate that cell cycle progression is not responsible for the loss of long-term repopulation capacity associated with human HSC culture. Overall, data presented in this thesis have relevance in the improvement of current clinical gene therapy protocols and establish a clear rationale for further study into the very early changes associated with HSC activation in culture.

7. BIBLIOGRAPHY

Acar, M., Kocherlakota, K.S., Murphy, M.M., Peyer, J.G., Oguro, H., Inra, C.N., Jaiyeola, C., Zhao, Z., Luby-Phelps, K., and Morrison, S.J. (2015). Deep imaging of bone marrow shows non-dividing stem cells are mainly perisinusoidal. Nature *526*, 126–130. https://doi.org/10.1038/nature15250.

Adli, M. (2018). The CRISPR tool kit for genome editing and beyond. Nat. Commun. *9*, 1911. https://doi.org/10.1038/s41467-018-04252-2.

Adolfsson, J., Månsson, R., Buza-Vidas, N., Hultquist, A., Liuba, K., Jensen, C.T., Bryder, D., Yang, L., Borge, O.-J., Thoren, L.A.M., et al. (2005). Identification of Flt3+ Lympho-Myeloid Stem Cells Lacking Erythro-Megakaryocytic Potential. Cell *121*, 295–306. https://doi.org/10.1016/j.cell.2005.02.013.

Aiuti, A., Slavin, S., Aker, M., Ficara, F., Deola, S., Mortellaro, A., Morecki, S., Andolfi, G., Tabucchi, A., Carlucci, F., et al. (2002). Correction of ADA-SCID by stem cell gene therapy combined with nonmyeloablative conditioning. Science *296*, 2410–2413. https://doi.org/10.1126/science.1070104.

Aiuti, A., Cattaneo, F., Galimberti, S., Benninghoff, U., Cassani, B., Callegaro, L., Scaramuzza, S., Andolfi, G., Mirolo, M., Brigida, I., et al. (2009). Gene therapy for immunodeficiency due to adenosine deaminase deficiency. N. Engl. J. Med. *360*, 447–458. https://doi.org/10.1056/NEJMoa0805817.

Aiuti, A., Biasco, L., Scaramuzza, S., Ferrua, F., Cicalese, M.P., Baricordi, C., Dionisio, F., Calabria, A., Giannelli, S., Castiello, M.C., et al. (2013). Lentiviral Hematopoietic Stem Cell Gene Therapy in Patients with Wiskott-Aldrich Syndrome. Science *341*, 1233151–1233151. https://doi.org/10.1126/science.1233151.

Aiuti, A., Roncarolo, M.G., and Naldini, L. (2017). Gene therapy for ADA-SCID, the first marketing approval of an ex vivo gene therapy in Europe: paving the road for the next generation of advanced therapy medicinal products. EMBO Mol. Med. *9*, 737–740. https://doi.org/10.15252/emmm.201707573.

Akashi, K., Traver, D., Miyamoto, T., and Weissman, I.L. (2000). A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. Nature *404*, 193–197. https://doi.org/10.1038/35004599.

de Almeida, M.J., Luchsinger, L.L., Corrigan, D.J., Williams, L.J., and Snoeck, H.-W. (2017). Dye-Independent Methods Reveal Elevated Mitochondrial Mass in Hematopoietic Stem Cells. Cell Stem Cell *21*, 725-729.e4. https://doi.org/10.1016/j.stem.2017.11.002.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166–169. https://doi.org/10.1093/bioinformatics/btu638.

Andrews, S. (2010). FastQC. Available at: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Andrews, R.G., Singer, J.W., and Bernstein, I.D. (1990). Human hematopoietic precursors in long-term culture: single CD34+ cells that lack detectable T cell, B cell, and myeloid cell antigens produce multiple colony-forming cells when cultured with marrow stromal cells. J. Exp. Med. *172*, 355–358. https://doi.org/10.1084/jem.172.1.355.

Asada, N., Katayama, Y., Sato, M., Minagawa, K., Wakahashi, K., Kawano, H., Kawano, Y., Sada, A., Ikeda, K., Matsui, T., et al. (2013). Matrix-embedded osteocytes regulate mobilization of hematopoietic stem/progenitor cells. Cell Stem Cell *12*, 737–747. https://doi.org/10.1016/j.stem.2013.05.001.

Bai, T., Li, J., Sinclair, A., Imren, S., Merriam, F., Sun, F., O'Kelly, M.B., Nourigat, C., Jain, P., Delrow, J.J., et al. (2019). Expansion of primitive human hematopoietic stem cells by culture in a zwitterionic hydrogel. Nat. Med. *25*, 1566–1575. https://doi.org/10.1038/s41591-019-0601-5.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. Science *315*, 1709–1712. https://doi.org/10.1126/science.1138140.

Baum, C.M., Weissman, I.L., Tsukamoto, A.S., Buckle, A.M., and Peault, B. (1992). Isolation of a candidate human hematopoietic stem-cell population. Proc. Natl. Acad. Sci. U. S. A. *89*, 2804–2808. .

Becker, A.J., McCULLOCH, E.A., and Till, J.E. (1963). Cytological Demonstration of the Clonal Nature of Spleen Colonies Derived from Transplanted Mouse Marrow Cells. Nature *197*, 452–454. https://doi.org/10.1038/197452a0.

Belluschi, S., Calderbank, E.F., Ciaurro, V., Pijuan-Sala, B., Santoro, A., Mende, N., Diamanti, E., Sham, K.Y.C., Wang, X., Lau, W.W.Y., et al. (2018). Myelo-lymphoid lineage restriction occurs in the human haematopoietic stem cell compartment before lymphoid-primed multipotent progenitors. Nat. Commun. *9*, 4100. https://doi.org/10.1038/s41467-018-06442-4.

Bendall, L.J., and Bradstock, K.F. (2014). G-CSF: From granulopoietic stimulant to bone marrow stem cell mobilizing agent. Cytokine Growth Factor Rev. *25*, 355–367. https://doi.org/10.1016/j.cytogfr.2014.07.011.

Bernitz, J.M., Kim, H.S., MacArthur, B., Sieburg, H., and Moore, K. (2016). Hematopoietic Stem Cells Count and Remember Self-Renewal Divisions. Cell *167*, 1296-1309.e10. https://doi.org/10.1016/j.cell.2016.10.022.

Bernitz, J.M., Daniel, M.G., Fstkchyan, Y.S., and Moore, K. (2017). Granulocyte colony-stimulating factor mobilizes dormant hematopoietic stem cells without proliferation in mice. Blood *129*, 1901–1912. https://doi.org/10.1182/blood-2016-11-752923.

Bhatia, M., Bonnet, D., Murdoch, B., Gan, O.I., and Dick, J.E. (1998). A newly discovered class of human hematopoietic cells with SCID-repopulating activity. Nat. Med. *4*, 1038–1045. https://doi.org/10.1038/2023.

Biasco, L., Pellin, D., Scala, S., Dionisio, F., Basso-Ricci, L., Leonardelli, L., Scaramuzza, S., Baricordi, C., Ferrua, F., Cicalese, M.P., et al. (2016). In Vivo Tracking of Human Hematopoiesis Reveals Patterns of Clonal Dynamics during Early and Steady-State Reconstitution Phases. Cell Stem Cell *19*, 107–119. https://doi.org/10.1016/j.stem.2016.04.016.

Biffi, A., Bartolomae, C.C., Cesana, D., Cartier, N., Aubourg, P., Ranzani, M., Cesani, M., Benedicenti, F., Plati, T., Rubagotti, E., et al. (2011). Lentiviral vector common integration sites in preclinical models and a clinical trial reflect a benign integration bias and not oncogenic selection. Blood *117*, 5332–5339. https://doi.org/10.1182/blood-2010-09-306761.

Biffi, A., Montini, E., Lorioli, L., Cesani, M., Fumagalli, F., Plati, T., Baldoli, C., Martino, S., Calabria, A., Canale, S., et al. (2013). Lentiviral Hematopoietic Stem Cell Gene Therapy Benefits Metachromatic Leukodystrophy. Science *341*, 1233158–1233158. https://doi.org/10.1126/science.1233158.

Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., Fridman, W.-H., Pagès, F., Trajanoski, Z., and Galon, J. (2009). ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics *25*, 1091–1093. https://doi.org/10.1093/bioinformatics/btp101.

Biton, M., Haber, A.L., Rogel, N., Burgin, G., Beyaz, S., Schnell, A., Ashenberg, O., Su, C.-W., Smillie, C., Shekhar, K., et al. (2018). T Helper Cell Cytokines Modulate Intestinal Stem Cell Renewal and Differentiation. Cell *175*, 1307-1320.e22. https://doi.org/10.1016/j.cell.2018.10.008.

bluebird bio (2021a). bluebird bio Reports Second Quarter Financial Results and Provides Operational Update. Available at: https://investor.bluebirdbio.com/news-releases/news-release-details/bluebird-bio-reports-second-quarter-financial-results-and.

bluebird bio (2021b). Press release on Acute Myeloid Leukemia in SCD gene therapy treated patient. Available at: https://investor.bluebirdbio.com/news-releases/news-release-details/bluebird-bio-provides-updated-findings-reported-case-acute.

Boelens, J.J., Aldenhoven, M., Purtill, D., Ruggeri, A., Defor, T., Wynn, R., Wraith, E., Cavazzana-Calvo, M., Rovelli, A., Fischer, A., et al. (2013). Outcomes of transplantation using various hematopoietic cell sources in children with Hurler syndrome after myeloablative conditioning. Blood *121*, 3981–3987. https://doi.org/10.1182/blood-2012-09-455238.

Boitano, A.E., Wang, J., Romeo, R., Bouchez, L.C., Parker, A.E., Sutton, S.E., Walker, J.R., Flaveny, C.A., Perdew, G.H., Denison, M.S., et al. (2010). Aryl hydrocarbon receptor antagonists promote the expansion of human hematopoietic stem cells. Science *329*, 1345–1348. https://doi.org/10.1126/science.1191536.

Bolotin, A., Quinquis, B., Sorokin, A., and Ehrlich, S.D. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiol. Read. Engl. *151*, 2551–2561. https://doi.org/10.1099/mic.0.28048-0.

Bordignon, C., Notarangelo, L.D., Nobili, N., Ferrari, G., Casorati, G., Panina, P., Mazzolari, E., Maggioni, D., Rossi, C., Servida, P., et al. (1995). Gene therapy in peripheral blood lymphocytes and bone marrow for ADA- immunodeficient patients. Science *270*, 470–475. https://doi.org/10.1126/science.270.5235.470.

Bosma, G.C., Custer, R.P., and Bosma, M.J. (1983). A severe combined immunodeficiency mutation in the mouse. Nature *301*, 527–530. https://doi.org/10.1038/301527a0.

Boward, B., Wu, T., and Dalton, S. (2016). Control of cell fate through cell cycle and pluripotency networks. Stem Cells Dayt. Ohio *34*, 1427–1436. https://doi.org/10.1002/stem.2345.

Boztug, K., Schmidt, M., Schwarzer, A., Banerjee, P.P., Díez, I.A., Dewey, R.A., Böhm, M., Nowrouzi, A., Ball, C.R., Glimm, H., et al. (2010). Stem-Cell Gene Therapy for the Wiskott–Aldrich Syndrome. N. Engl. J. Med. *363*, 1918–1927. https://doi.org/10.1056/NEJMoa1003548.

Bruns, I., Lucas, D., Pinho, S., Ahmed, J., Lambert, M.P., Kunisaki, Y., Scheiermann, C., Schiff, L., Poncz, M., Bergman, A., et al. (2014). Megakaryocytes regulate hematopoietic stem cell quiescence through CXCL4 secretion. Nat. Med. *20*, 1315–1320. https://doi.org/10.1038/nm.3707.

Burke, J.R., Deshong, A.J., Pelton, J.G., and Rubin, S.M. (2010). Phosphorylation-induced Conformational Changes in the Retinoblastoma Protein Inhibit E2F Transactivation Domain Binding. J. Biol. Chem. *285*, 16286–16293. https://doi.org/10.1074/jbc.M110.108167.

Burns, J.C., Friedmann, T., Driever, W., Burrascano, M., and Yee, J.K. (1993). Vesicular stomatitis virus G glycoprotein pseudotyped retroviral vectors: concentration to very high titer and efficient gene transfer into mammalian and nonmammalian cells. Proc. Natl. Acad. Sci. U. S. A. *90*, 8033–8037. https://doi.org/10.1073/pnas.90.17.8033.

Cabezas-Wallscheid, N., Buettner, F., Sommerkamp, P., Klimmeck, D., Ladel, L., Thalheimer, F.B., Pastor-Flores, D., Roma, L.P., Renders, S., Zeisberger, P., et al. (2017). Vitamin A-Retinoic Acid Signaling Regulates Hematopoietic Stem Cell Dormancy. Cell *169*, 807-823.e19. https://doi.org/10.1016/j.cell.2017.04.018.

Canu, G., Athanasiadis, E., Grandy, R.A., Garcia-Bernardo, J., Strzelecka, P.M., Vallier, L., Ortmann, D., and Cvejic, A. (2020). Analysis of endothelial-to-haematopoietic transition at the single cell level identifies cell cycle regulation as a driver of differentiation. Genome Biol. *21*, 157. https://doi.org/10.1186/s13059-020-02058-4.

Carrelha, J., Meng, Y., Kettyle, L.M., Luis, T.C., Norfo, R., Alcolea, V., Boukarabila, H., Grasso, F., Gambardella, A., Grover, A., et al. (2018). Hierarchically related lineage-restricted fates of multipotent haematopoietic stem cells. Nature *554*, 106–111. https://doi.org/10.1038/nature25455.

Cavazzana-Calvo, M., Hacein-Bey, S., de Saint Basile, G., Gross, F., Yvon, E., Nusbaum, P., Selz, F., Hue, C., Certain, S., Casanova, J.L., et al. (2000). Gene therapy of human severe combined immunodeficiency (SCID)-X1 disease. Science *288*, 669–672. .

Chagraoui, J., Girard, S., Spinella, J.-F., Simon, L., Bonneil, E., Mayotte, N., MacRae, T., Coulombe-Huntington, J., Bertomeu, T., Moison, C., et al. (2021). UM171 Preserves Epigenetic Marks that Are Reduced in Ex Vivo Culture of Human HSCs via Potentiation of the CLR3-KBTBD4 Complex. Cell Stem Cell *28*, 48-62.e6. https://doi.org/10.1016/j.stem.2020.12.002.

Chalancon, G., Ravarani, C.N.J., Balaji, S., Martinez-Arias, A., Aravind, L., Jothi, R., and Babu, M.M. (2012). Interplay between gene expression noise and regulatory network architecture. Trends Genet. *28*, 221–232. https://doi.org/10.1016/j.tig.2012.01.006.

Challen, G.A., Boles, N.C., Chambers, S.M., and Goodell, M.A. (2010). Distinct hematopoietic stem cell subtypes are differentially regulated by TGF-beta1. Cell Stem Cell *6*, 265–278. https://doi.org/10.1016/j.stem.2010.02.002.

Chen, Q., Khoury, M., and Chen, J. (2009). Expression of human cytokines dramatically improves reconstitution of specific human-blood lineage cells in humanized mice. Proc. Natl. Acad. Sci. U. S. A. *106*, 21783–21788. https://doi.org/10.1073/pnas.0912274106.

Cheng, T., Rodrigues, N., Shen, H., Yang, Y., Dombkowski, D., Sykes, M., and Scadden, D.T. (2000). Hematopoietic stem cell quiescence maintained by p21cip1/waf1. Science *287*, 1804–1808. https://doi.org/10.1126/science.287.5459.1804.

Christopher, M.J., Rao, M., Liu, F., Woloszynek, J.R., and Link, D.C. (2011). Expression of the G-CSF receptor in monocytic cells is sufficient to mediate hematopoietic progenitor mobilization by G-CSF in mice. J. Exp. Med. *208*, 251–260. https://doi.org/10.1084/jem.20101700.

Chung, M., Liu, C., Yang, H.W., Köberlin, M.S., Cappell, S.D., and Meyer, T. (2019). Transient Hysteresis in CDK4/6 Activity Underlies Passage of the Restriction Point in G1. Mol. Cell *76*, 562-573.e4. https://doi.org/10.1016/j.molcel.2019.08.020.

Cicalese, M.P., Ferrua, F., Castagnaro, L., Pajno, R., Barzaghi, F., Giannelli, S., Dionisio, F., Brigida, I., Bonopane, M., Casiraghi, M., et al. (2016). Update on the safety and efficacy of retroviral gene therapy for immunodeficiency due to adenosine deaminase deficiency. Blood *128*, 45–54. https://doi.org/10.1182/blood-2016-01-688226.

Cohen, S., Roy, J., Lachance, S., Delisle, J.-S., Marinier, A., Busque, L., Roy, D.-C., Barabé, F., Ahmad, I., Bambace, N., et al. (2020). Hematopoietic stem cell transplantation using single UM171-expanded cord blood: a single-arm, phase 1-2 safety and feasibility study. Lancet Haematol. *7*, e134–e145. https://doi.org/10.1016/S2352-3026(19)30202-9.

Conway, A.J., Brown, F.C., Hortle, E.J., Burgio, G., Foote, S.J., Morton, C.J., Jane, S.M., and Curtis, D.J. (2018). Bone marrow transplantation corrects haemolytic anaemia in a novel ENU mutagenesis mouse model of TPI deficiency. Dis. Model. Mech. *11*, dmm034678. https://doi.org/10.1242/dmm.034678.

Craig, W., Kay, R., Cutler, R.L., and Lansdorp, P.M. (1993). Expression of Thy-1 on human hematopoietic progenitor cells. J. Exp. Med. *177*, 1331–1342. https://doi.org/10.1084/jem.177.5.1331.

Csaszar, E., Kirouac, D.C., Yu, M., Wang, W., Qiao, W., Cooke, M.P., Boitano, A.E., Ito, C., and Zandstra, P.W. (2012). Rapid Expansion of Human Hematopoietic Stem Cells by Automated Control of Inhibitory Feedback Signaling. Cell Stem Cell *10*, 218–229. https://doi.org/10.1016/j.stem.2012.01.003.

Dahiya, A., Wong, S., Gonzalo, S., Gavin, M., and Dean, D.C. (2001). Linking the Rb and polycomb pathways. Mol. Cell *8*, 557–569. https://doi.org/10.1016/s1097-2765(01)00346-x.

Dalton, S., and Coverdell, P.D. (2015). Linking the cell cycle to cell fate decisions. Trends Cell Biol. *25*, 592–600. https://doi.org/10.1016/j.tcb.2015.07.007.

Danet, G.H., Pan, Y., Luongo, J.L., Bonnet, D.A., and Simon, M.C. (2003). Expansion of human SCID-repopulating cells under hypoxic conditions. J. Clin. Invest. *112*, 126–135. https://doi.org/10.1172/JCI200317669.

De Ravin, S.S., Li, L., Wu, X., Choi, U., Allen, C., Koontz, S., Lee, J., Theobald-Whiting, N., Chu, J., Garofalo, M., et al. (2017). CRISPR-Cas9 gene repair of hematopoietic stem cells from patients with X-linked chronic granulomatous disease. Sci. Transl. Med. *9*, eaah3480. https://doi.org/10.1126/scitranslmed.aah3480.

Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z.A., Eckert, M.R., Vogel, J., and Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature *471*, 602–607. https://doi.org/10.1038/nature09886.

Dever, D.P., Bak, R.O., Reinisch, A., Camarena, J., Washington, G., Nicolas, C.E., Pavel-Dinu, M., Saxena, N., Wilkens, A.B., Mantri, S., et al. (2016). CRISPR/Cas9 β-globin gene targeting in human haematopoietic stem cells. Nature *539*, 384–389. https://doi.org/10.1038/nature20134.

Dexter, T.M., Allen, T.D., and Lajtha, L.G. (1977). Conditions controlling the proliferation of haemopoietic stem cells in vitro. J. Cell. Physiol. *91*, 335–344. https://doi.org/10.1002/jcp.1040910303.

Dhaliwal, N.K., Miri, K., Davidson, S., Tamim El Jarkass, H., and Mitchell, J.A. (2018). KLF4 Nuclear Export Requires ERK Activation and Initiates Exit from Naive Pluripotency. Stem Cell Rep. *10*, 1308–1323. https://doi.org/10.1016/j.stemcr.2018.02.007.

Dhar, S.S., Ongwijitwat, S., and Wong-Riley, M.T.T. (2008). Nuclear Respiratory Factor 1 Regulates All Ten Nuclear-encoded Subunits of Cytochrome c Oxidase in Neurons. J. Biol. Chem. *283*, 3120–3129. https://doi.org/10.1074/jbc.M707587200.

Ding, L., Saunders, T.L., Enikolopov, G., and Morrison, S.J. (2012). Endothelial and perivascular cells maintain haematopoietic stem cells. Nature *481*, 457–462. https://doi.org/10.1038/nature10783.

Doan, P.L., Himburg, H.A., Helms, K., Russell, J.L., Fixsen, E., Quarmyne, M., Harris, J.R., Deoliviera, D., Sullivan, J.M., Chao, N.J., et al. (2013). Epidermal Growth Factor Regulates Hematopoietic Regeneration Following Radiation Injury. Nat. Med. *19*, 295–304. https://doi.org/10.1038/nm.3070.

Doudna, J.A. (2020). The promise and challenge of therapeutic genome editing. Nature *578*, 229–236. https://doi.org/10.1038/s41586-020-1978-5.

Doulatov, S., Notta, F., Eppert, K., Nguyen, L.T., Ohashi, P.S., and Dick, J.E. (2010). Revised map of the human progenitor hierarchy shows the origin of macrophages and dendritic cells in early lymphoid development. Nat. Immunol. *11*, 585–593. https://doi.org/10.1038/ni.1889.

Dull, T., Zufferey, R., Kelly, M., Mandel, R.J., Nguyen, M., Trono, D., and Naldini, L. (1998). A third-generation lentivirus vector with a conditional packaging system. J. Virol. *72*, 8463–8471. https://doi.org/10.1128/JVI.72.11.8463-8471.1998.

Durinck, S., Moreau, Y., Kasprzyk, A., Davis, S., De Moor, B., Brazma, A., and Huber, W. (2005). BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. Bioinforma. Oxf. Engl. *21*, 3439–3440. https://doi.org/10.1093/bioinformatics/bti525.

DuRose, J.B., Tam, A.B., and Niwa, M. (2006). Intrinsic capacities of molecular sensors of the unfolded protein response to sense alternate forms of endoplasmic reticulum stress. Mol. Biol. Cell *17*, 3095–3107. https://doi.org/10.1091/mbc.e06-01-0055.

Dykstra, B., Kent, D., Bowie, M., McCaffrey, L., Hamilton, M., Lyons, K., Lee, S.-J., Brinkman, R., and Eaves, C. (2007). Long-Term Propagation of Distinct Hematopoietic Differentiation Programs In Vivo. Cell Stem Cell *1*, 218–229. https://doi.org/10.1016/j.stem.2007.05.015.

Eberwine, J., Yeh, H., Miyashiro, K., Cao, Y., Nair, S., Finnell, R., Zettel, M., and Coleman, P. (1992). Analysis of gene expression in single live neurons. Proc. Natl. Acad. Sci. *89*, 3010–3014. https://doi.org/10.1073/pnas.89.7.3010.

Eichler, F.S., Duncan, C., Amartino, H., Aziz-Bose, R., Baruchel, A., Dansereau, C., Dalle, J.-H., Gaspar, H.B., Gissen, P., Kohn, D.B., et al. (2016). 250. A Phase 2/3 Study of the Efficacy and Safety of Ex Vivo Gene Therapy with Lenti-DTM Lentiviral Vector for the Treatment of Cerebral Adrenoleukodystrophy. Mol. Ther. *24*, S98–S99. https://doi.org/10.1016/S1525-0016(16)33059-3.

El Omar, R., Julien, E., Biasch, K., Guffroy, B., Lioure, B., Vallat, L., Gross, I., Domon-Dell, C., Lanza, F., Gachet, C., et al. (2021). CDX2 regulates ACE expression in blood development and leukemia cells. Blood Adv. *5*, 2012–2016. https://doi.org/10.1182/bloodadvances.2020003563.

Eling, N., Richard, A.C., Richardson, S., Marioni, J.C., and Vallejos, C.A. (2019). Correcting the Mean-Variance Dependency for Differential Variability Testing Using Single-Cell RNA Sequencing Data. Cell Syst. *9*, 401–413. https://doi.org/10.1016/j.cels.2019.08.003.

Ema, H., Takano, H., Sudo, K., and Nakauchi, H. (2000). In Vitro Self-Renewal Division of Hematopoietic Stem Cells. J. Exp. Med. *192*, 1281–1288. .

Eppert, K., Takenaka, K., Lechman, E.R., Waldron, L., Nilsson, B., van Galen, P., Metzeler, K.H., Poeppl, A., Ling, V., Beyene, J., et al. (2011). Stem cell gene expression programs influence clinical outcome in human leukemia. Nat. Med. *17*, 1086–1093. https://doi.org/10.1038/nm.2415.

Fares, I., Chagraoui, J., Gareau, Y., Gingras, S., Ruel, R., Mayotte, N., Csaszar, E., Knapp, D.J.H.F., Miller, P., Ngom, M., et al. (2014). Cord blood expansion. Pyrimidoindole derivatives are agonists of human hematopoietic stem cell self-renewal. Science *345*, 1509–1512. https://doi.org/10.1126/science.1256337.

Ferrari, G., Rossini, S., Giavazzi, R., Maggioni, D., Nobili, N., Soldati, M., Ungers, G., Mavilio, F., Gilboa, E., and Bordignon, C. (1991). An in Vivo Model of Somatic Cell Gene Therapy for Human Severe Combined Immunodeficiency. Science *251*, 1363–1366. https://doi.org/10.1126/science.1848369.

Ferrari, S., Jacob, A., Beretta, S., Unali, G., Albano, L., Vavassori, V., Cittaro, D., Lazarevic, D., Brombin, C., Cugnata, F., et al. (2020). Efficient gene editing of human long-term hematopoietic stem cells validated by clonal tracking. Nat. Biotechnol. *38*, 1298–1308. https://doi.org/10.1038/s41587-020-0551-y.

Ferrari, S., Vavassori, V., Canarutto, D., Jacob, A., Castiello, M.C., Javed, A.O., and Genovese, P. (2021). Gene Editing of Hematopoietic Stem Cells: Hopes and Hurdles Toward Clinical Translation. Front. Genome Ed. *3*, 9. https://doi.org/10.3389/fgeed.2021.618378.

Ford, C.E., Hamerton, J.L., Barnes, D.W., and Loutit, J.F. (1956). Cytological identification of radiation-chimaeras. Nature *177*, 452–454. https://doi.org/10.1038/177452a0.

Foudi, A., Hochedlinger, K., Van Buren, D., Schindler, J.W., Jaenisch, R., Carey, V., and Hock, H. (2009). Analysis of histone 2B-GFP retention reveals slowly cycling hematopoietic stem cells. Nat. Biotechnol. *27*, 84–90. https://doi.org/10.1038/nbt.1517.

Frangoul, H., Altshuler, D., Cappellini, M.D., Chen, Y.-S., Domm, J., Eustace, B.K., Foell, J., de la Fuente, J., Grupp, S., Handgretinger, R., et al. (2021). CRISPR-Cas9 Gene Editing for Sickle Cell Disease and β-Thalassemia. N. Engl. J. Med. *384*, 252–260. https://doi.org/10.1056/NEJMoa2031054.

Friedmann, T., and Roblin, R. (1972). Gene therapy for human genetic disease? Science *175*, 949–955. https://doi.org/10.1126/science.175.4025.949.

Fumagalli, F., Calbi, V., Sessa, M., Zambon, A., Baldoli, C., Rancoita, P.M.V., Acquati, S., de Mattia, F., Tucci, F., Gallo, V., et al. (2020). Lentiviral hematopoietic stem and progenitor cell gene therapy (HSPC-GT) for metachromatic leukodystrophy (MLD): Clinical outcomes from 33 patients. Mol. Genet. Metab. *129*, S59. https://doi.org/10.1016/j.ymgme.2019.11.135.

Galen, P. van, Kreso, A., Mbong, N., Kent, D.G., Fitzmaurice, T., Chambers, J.E., Xie, S., Laurenti, E., Hermans, K., Eppert, K., et al. (2014). The unfolded protein response governs integrity of the haematopoietic stem-cell pool during stress. Nature *510*, 268–272. https://doi.org/10.1038/nature13228.

Galen, P. van, Mbong, N., Kreso, A., Schoof, E.M., Wagenblast, E., Ng, S.W.K., Krivdova, G., Jin, L., Nakauchi, H., and Dick, J.E. (2018). Integrated Stress Response Activity Marks Stem Cells in Normal Hematopoiesis and Leukemia. Cell Rep. *25*, 1109-1117.e5. https://doi.org/10.1016/j.celrep.2018.10.021.

Gan, O.I., Murdoch, B., Larochelle, A., and Dick, J.E. (1997). Differential maintenance of primitive human SCID-repopulating cells, clonogenic progenitors, and long-term culture-initiating cells after incubation on human bone marrow stromal cells. Blood *90*, 641–650. .

Gándara, C., Affleck, V., and Stoll, E.A. (2018). Manufacture of Third-Generation Lentivirus for Preclinical Use, with Process Development Considerations for Translation to Good Manufacturing Practice. Hum. Gene Ther. Methods *29*, 1–15. https://doi.org/10.1089/hgtb.2017.098.

García-Prat, L., Sousa-Victor, P., and Muñoz-Cánoves, P. (2017). Proteostatic and Metabolic Control of Stemness. Cell Stem Cell *20*, 593–608. https://doi.org/10.1016/j.stem.2017.04.011.

García-Prat, L., Kaufmann, K.B., Schneiter, F., Voisin, V., Murison, A., Chen, J., Chan-Seng-Yue, M., Gan, O.I., McLeod, J.L., Smith, S.A., et al. (2021). TFEB-mediated endolysosomal activity controls human hematopoietic stem cell fate. Cell Stem Cell *28*, 1838-1850.e10. https://doi.org/10.1016/j.stem.2021.07.003.

Gaspar, H.B., Aiuti, A., Porta, F., Candotti, F., Hershfield, M.S., and Notarangelo, L.D. (2009). How I treat ADA deficiency. Blood *114*, 3524–3532. https://doi.org/10.1182/blood-2009-06-189209.

Genovese, P., Schiroli, G., Escobar, G., Di Tomaso, T., Firrito, C., Calabria, A., Moi, D., Mazzieri, R., Bonini, C., Holmes, M.C., et al. (2014). Targeted genome editing in human repopulating haematopoietic stem cells. Nature *510*, 235–240. https://doi.org/10.1038/nature13420.

Gentner, B., Tucci, F., Galimberti, S., Fumagalli, F., De Pellegrin, M., Silvani, P., Camesasca, C., Pontesilli, S., Darin, S., Ciotti, F., et al. (2021). Hematopoietic Stem- and Progenitor-Cell Gene Therapy for Hurler Syndrome. N. Engl. J. Med. *385*, 1929–1940. https://doi.org/10.1056/NEJMoa2106596.

Giralt, S., and Bishop, M.R. (2009). Principles and Overview of Allogeneic Hematopoietic Stem Cell Transplantation. Cancer Treat. Res. *144*, 1–21. https://doi.org/10.1007/978-0-387-78580-6_1.

Gläser, R., Meyer-Hoffert, U., Harder, J., Cordes, J., Wittersheim, M., Kobliakova, J., Fölster-Holst, R., Proksch, E., Schröder, J.-M., and Schwarz, T. (2009). The Antimicrobial Protein Psoriasin (S100A7) Is Upregulated in Atopic Dermatitis and after Experimental Skin Barrier Disruption. J. Invest. Dermatol. *129*, 641–649. https://doi.org/10.1038/jid.2008.268.

Glimm, H., Oh, I.H., and Eaves, C.J. (2000). Human hematopoietic stem cells stimulated to proliferate in vitro lose engraftment potential during their S/G(2)/M transit and do not reenter G(0). Blood *96*, 4185–4193. .

Gothot, A., van der Loo, J.C., Clapp, D.W., and Srour, E.F. (1998). Cell cycle-related changes in repopulating capacity of human mobilized peripheral blood CD34(+) cells in non-obese diabetic/severe combined immune-deficient mice. Blood *92*, 2641–2649. .

Greenbaum, A., Hsu, Y.-M.S., Day, R.B., Schuettpelz, L.G., Christopher, M.J., Borgerding, J.N., Nagasawa, T., and Link, D.C. (2013). CXCL12 in early mesenchymal progenitors is required for haematopoietic stem-cell maintenance. Nature *495*, 227–230. https://doi.org/10.1038/nature11926.

Grinenko, T., Eugster, A., Thielecke, L., Ramasz, B., Krüger, A., Dietz, S., Glauche, I., Gerbaulet, A., von Bonin, M., Basak, O., et al. (2018). Hematopoietic stem cells can differentiate into restricted myeloid progenitors before cell division in mice. Nat. Commun. *9*, 1898. https://doi.org/10.1038/s41467-018-04188-7.

Grün, D. (2020). Revealing dynamics of gene expression variability in cell state space. Nat. Methods *17*, 45–49. https://doi.org/10.1038/s41592-019-0632-3.

Guo, G., Yang, J., Nichols, J., Hall, J.S., Eyres, I., Mansfield, W., and Smith, A. (2009). Klf4 reverts developmentally programmed restriction of ground state pluripotency. Dev. Camb. Engl. *136*, 1063–1069. https://doi.org/10.1242/dev.030957.

Haas, S., Trumpp, A., and Milsom, M.D. (2018). Causes and Consequences of Hematopoietic Stem Cell Heterogeneity. Cell Stem Cell *22*, 627–638. https://doi.org/10.1016/j.stem.2018.04.003.

Hacein-Bey-Abina, S., Le Deist, F., Carlier, F., Bouneaud, C., Hue, C., De Villartay, J.-P., Thrasher, A.J., Wulffraat, N., Sorensen, R., Dupuis-Girod, S., et al. (2002). Sustained correction of X-linked severe combined immunodeficiency by ex vivo gene therapy. N. Engl. J. Med. *346*, 1185–1193. https://doi.org/10.1056/NEJMoa012616.

Hacein-Bey-Abina, S., Garrigue, A., Wang, G.P., Soulier, J., Lim, A., Morillon, E., Clappier, E., Caccavelli, L., Delabesse, E., Beldjord, K., et al. (2008). Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. J. Clin. Invest. *118*, 3132–3142. https://doi.org/10.1172/JCI35700.

Hall (2014). Extreme Limiting Dilution Analysis software. Available at: https://bioinf.wehi.edu.au/software/elda/.

Hannun, Y.A., and Obeid, L.M. (2018). Sphingolipids and their metabolism in physiology and disease. Nat. Rev. Mol. Cell Biol. *19*, 175–191. https://doi.org/10.1038/nrm.2017.107.

Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: gene set variation analysis for microarray and RNA-Seq data. BMC Bioinformatics *14*, 7. https://doi.org/10.1186/1471-2105-14-7.

Hao, Q.L., Shah, A.J., Thiemann, F.T., Smogorzewska, E.M., and Crooks, G.M. (1995). A functional comparison of CD34 + CD38- cells in cord blood and bone marrow. Blood *86*, 3745–3753. .

Hao, S., Chen, C., and Cheng, T. (2016). Cell cycle regulation of hematopoietic stem or progenitor cells. Int. J. Hematol. *103*, 487–497. https://doi.org/10.1007/s12185-016-1984-4.

Hasegawa, Y., Taylor, D., Ovchinnikov, D.A., Wolvetang, E.J., Torrenté, L. de, and Mar, J.C. (2015). Variability of Gene Expression Identifies Transcriptional Regulators of Early Human Embryonic Development. PLOS Genet. *11*, e1005428. https://doi.org/10.1371/journal.pgen.1005428.

Hauber, I., Beschorner, N., Schrödel, S., Chemnitz, J., Kröger, N., Hauber, J., and Thirion, C. (2018). Improving Lentiviral Transduction of CD34+ Hematopoietic Stem and Progenitor Cells. Hum. Gene Ther. Methods *29*, 104–113. https://doi.org/10.1089/hgtb.2017.085.

Hesselton, R.M., Greiner, D.L., Mordes, J.P., Rajan, T.V., Sullivan, J.L., and Shultz, L.D. (1995). High levels of human peripheral blood mononuclear cell engraftment and enhanced susceptibility to human immunodeficiency virus type 1 infection in NOD/LtSz-scid/scid mice. J. Infect. Dis. *172*, 974–982. https://doi.org/10.1093/infdis/172.4.974.

Hinge, A., He, J., Bartram, J., Javier, J., Xu, J., Fjellman, E., Sesaki, H., Li, T., Yu, J., Wunderlich, M., et al. (2020). Asymmetrically Segregated Mitochondria Provide Cellular Memory of Hematopoietic Stem Cell Replicative History and Drive HSC Attrition. Cell Stem Cell *26*, 420-430.e6. https://doi.org/10.1016/j.stem.2020.01.016.

Hirata, Y., Furuhashi, K., Ishii, H., Li, H.W., Pinho, S., Ding, L., Robson, S.C., Frenette, P.S., and Fujisaki, J. (2018). CD150high Bone Marrow Tregs Maintain Hematopoietic Stem Cell Quiescence and Immune Privilege via Adenosine. Cell Stem Cell *22*, 445-453.e5. https://doi.org/10.1016/j.stem.2018.01.017.

Ho, T.T., Warr, M.R., Adelman, E.R., Lansinger, O.M., Flach, J., Verovskaya, E.V., Figueroa, M.E., and Passegué, E. (2017). Autophagy maintains the metabolism and function of young and old stem cells. Nature *543*, 205–210. https://doi.org/10.1038/nature21388.

Hock, H., Hamblen, M.J., Rooke, H.M., Schindler, J.W., Saleque, S., Fujiwara, Y., and Orkin, S.H. (2004). Gfi-1 restricts proliferation and preserves functional integrity of haematopoietic stem cells. Nature *431*, 1002–1007. https://doi.org/10.1038/nature02994.

Hodgson, G.S., and Bradley, T.R. (1979). Properties of haematopoietic stem cells surviving 5-fluorouracil treatment: evidence for a pre-CFU-S cell? Nature *281*, 381–382. https://doi.org/10.1038/281381a0.

Hogge, D.E., Lansdorp, P.M., Reid, D., Gerhard, B., and Eaves, C.J. (1996). Enhanced detection, maintenance, and differentiation of primitive human hematopoietic cells in cultures containing murine fibroblasts engineered to produce human steel factor, interleukin-3, and granulocyte colony-stimulating factor. Blood *88*, 3765–3773. .

Hosokawa, K., Arai, F., Yoshihara, H., Iwasaki, H., Nakamura, Y., Gomei, Y., and Suda, T. (2010). Knockdown of N-cadherin suppresses the long-term engraftment of hematopoietic stem cells. Blood *116*, 554–563. https://doi.org/10.1182/blood-2009-05-224857.

Howe, S.J., Mansour, M.R., Schwarzwaelder, K., Bartholomae, C., Hubank, M., Kempski, H., Brugman, M.H., Pike-Overzet, K., Chatters, S.J., de Ridder, D., et al. (2008). Insertional mutagenesis combined with acquired somatic mutations causes leukemogenesis following gene therapy of SCID-X1 patients. J. Clin. Invest. *118*, 3143–3150. https://doi.org/10.1172/JCI35798.

Hu, W., Dooley, J., Chung, S.S., Chandramohan, D., Cimmino, L., Mukherjee, S., Mason, C.E., de Strooper, B., Liston, A., and Park, C.Y. (2015). miR-29a maintains mouse hematopoietic stem cell self-renewal by regulating Dnmt3a. Blood *125*, 2206–2216. https://doi.org/10.1182/blood-2014-06-585273.

Hur, J., Choi, J.-I., Lee, H., Nham, P., Kim, T.-W., Chae, C.-W., Yun, J.-Y., Kang, J.-A., Kang, J., Lee, S.E., et al. (2016). CD82/KAI1 Maintains the Dormancy of Long-Term Hematopoietic Stem Cells through Interaction with DARC-Expressing Macrophages. Cell Stem Cell *18*, 508–521. https://doi.org/10.1016/j.stem.2016.01.013.

Ieyasu, A., Ishida, R., Kimura, T., Morita, M., Wilkinson, A.C., Sudo, K., Nishimura, T., Ohehara, J., Tajima, Y., Lai, C.-Y., et al. (2017). An All-Recombinant Protein-Based Culture System Specifically Identifies Hematopoietic Stem Cell Maintenance Factors. Stem Cell Rep. *8*, 500–508. https://doi.org/10.1016/j.stemcr.2017.01.015.

Ikonomi, N., Kühlwein, S.D., Schwab, J.D., and Kestler, H.A. (2020). Awakening the HSC: Dynamic Modeling of HSC Maintenance Unravels Regulation of the TP53 Pathway and Quiescence. Front. Physiol. *11*, 848. https://doi.org/10.3389/fphys.2020.00848.

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., and Nakata, A. (1987). Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. J. Bacteriol. *169*, 5429–5433. https://doi.org/10.1128/jb.169.12.5429-5433.1987.

Itkin, T., Gur-Cohen, S., Spencer, J.A., Schajnovitz, A., Ramasamy, S.K., Kusumbe, A.P., Ledergor, G., Jung, Y., Milo, I., Poulos, M.G., et al. (2016). Distinct bone marrow blood vessels differentially regulate haematopoiesis. Nature *532*, 323–328. https://doi.org/10.1038/nature17624.

Ito, K., Hirao, A., Arai, F., Takubo, K., Matsuoka, S., Miyamoto, K., Ohmura, M., Naka, K., Hosokawa, K., Ikeda, Y., et al. (2006). Reactive oxygen species act through p38 MAPK to limit the lifespan of hematopoietic stem cells. Nat. Med. *12*, 446–451. https://doi.org/10.1038/nm1388.

Ito, K., Turcotte, R., Cui, J., Zimmerman, S.E., Pinho, S., Mizoguchi, T., Arai, F., Runnels, J.M., Alt, C., Teruya-Feldstein, J., et al. (2016). Self-renewal of a purified Tie2+ hematopoietic stem cell population relies on mitochondrial clearance. Science *354*, 1156–1160. https://doi.org/10.1126/science.aaf5530.

Ito, M., Hiramatsu, H., Kobayashi, K., Suzue, K., Kawahata, M., Hioki, K., Ueyama, Y., Koyanagi, Y., Sugamura, K., Tsuji, K., et al. (2002). NOD/SCID/gamma(c)(null) mouse: an excellent recipient mouse model for engraftment of human cells. Blood *100*, 3175–3182. https://doi.org/10.1182/blood-2001-12-0207.

Jacobson, L.O., Simmons, E.L., Marks, E.K., and Eldredge, J.H. (1951). Recovery from Radiation Injury. Science *113*, 510–511. https://doi.org/10.1126/science.113.2940.510.

Janky, R., Verfaillie, A., Imrichová, H., Sande, B.V. de, Standaert, L., Christiaens, V., Hulselmans, G., Herten, K., Sanchez, M.N., Potier, D., et al. (2014). iRegulon: From a Gene List to a Gene Regulatory Network Using Large Motif and Track Collections. PLOS Comput. Biol. *10*, e1003731. https://doi.org/10.1371/journal.pcbi.1003731.

Jansen, S., Podschun, R., Leib, S.L., Grötzinger, J., Oestern, S., Michalek, M., Pufe, T., and Brandenburg, L.-O. (2013). Expression and Function of Psoriasin (S100A7) and Koebnerisin (S100A15) in the Brain. Infect. Immun. *81*, 1788–1797. https://doi.org/10.1128/IAI.01265-12.

Jawaid, W. (2017). bglab lackage. Available at: https://rdrr.io/github/wjawaid/bglab/.

Johnson, C., Belluschi, S., and Laurenti, E. (2020). Beyond “to divide or not to divide”: Kinetics matters in hematopoietic stem cells. Exp. Hematol. *92*, 1-10.e2. https://doi.org/10.1016/j.exphem.2020.11.003.

Kallinikou, K., Anjos-Afonso, F., Blundell, M.P., Ings, S.J., Watts, M.J., Thrasher, A.J., Linch, D.C., Bonnet, D., and Yong, K.L. (2012). Engraftment defect of cytokine-cultured adult human mobilized CD34(+) cells is related to reduced adhesion to bone marrow niche elements. Br. J. Haematol. *158*, 778–787. https://doi.org/10.1111/j.1365-2141.2012.09219.x.

Katayama, Y., Battista, M., Kao, W.-M., Hidalgo, A., Peired, A.J., Thomas, S.A., and Frenette, P.S. (2006). Signals from the Sympathetic Nervous System Regulate Hematopoietic Stem Cell Egress from Bone Marrow. Cell *124*, 407–421. https://doi.org/10.1016/j.cell.2005.10.041.

Kaufmann, K.B., Zeng, A.G.X., Coyaud, E., Garcia-Prat, L., Papalexi, E., Murison, A., Laurent, E.M.N., Chan-Seng-Yue, M., Gan, O.I., Pan, K., et al. (2021). A latent subset of human hematopoietic stem cells resists regenerative stress to preserve stemness. Nat. Immunol. *22*, 723–734. https://doi.org/10.1038/s41590-021-00925-1.

Kennedy, D.R., McLellan, K., Moore, P.F., Henthorn, P.S., and Felsburg, P.J. (2009). Effect of ex vivo culture of CD34+ bone marrow cells on immune reconstitution of XSCID dogs following allogeneic bone marrow transplantation. Biol. Blood Marrow Transplant. J. Am. Soc. Blood Marrow Transplant. *15*, 662–670. https://doi.org/10.1016/j.bbmt.2009.03.014.

Kent, D.G., Dykstra, B.J., Cheyne, J., Ma, E., and Eaves, C.J. (2008). Steel factor coordinately regulates the molecular signature and biologic function of hematopoietic stem cells. Blood *112*, 560–567. https://doi.org/10.1182/blood-2007-10-117820.

Kent, D.G., Copley, M.R., Benz, C., Wöhrer, S., Dykstra, B.J., Ma, E., Cheyne, J., Zhao, Y., Bowie, M.B., Zhao, Y., et al. (2009). Prospective isolation and molecular characterization of hematopoietic stem cells with durable self-renewal potential. Blood *113*, 6342–6350. https://doi.org/10.1182/blood-2008-12-192054.

Khurana, S., Schouteden, S., Manesia, J.K., Santamaria-Martínez, A., Huelsken, J., Lacy-Hulbert, A., and Verfaillie, C.M. (2016). Outside-in integrin signalling regulates haematopoietic stem cell function via Periostin-Itgav axis. Nat. Commun. *7*, 13500. https://doi.org/10.1038/ncomms13500.

Kim, K., Bang, S.-Y., Lee, H.-S., and Bae, S.-C. (2017). Update on the genetic architecture of rheumatoid arthritis. Nat. Rev. Rheumatol. *13*, 13–24. https://doi.org/10.1038/nrrheum.2016.176.

Kirouac, D.C., Ito, C., Csaszar, E., Roch, A., Yu, M., Sykes, E.A., Bader, G.D., and Zandstra, P.W. (2010). Dynamic interaction networks in a hierarchically organized tissue. Mol. Syst. Biol. *6*, 417. https://doi.org/10.1038/msb.2010.71.

Knapp, D.J.H.F., Hammond, C.A., Hui, T., van Loenhout, M.T.J., Wang, F., Aghaeepour, N., Miller, P.H., Moksa, M., Rabu, G.M., Beer, P.A., et al. (2018). Single-cell analysis identifies a CD33+ subset of human cord blood cells with high regenerative potential. Nat. Cell Biol. *20*, 710–720. https://doi.org/10.1038/s41556-018-0104-5.

Kohn, D.B., Weinberg, K.I., Nolta, J.A., Heiss, L.N., Lenarsky, C., Crooks, G.M., Hanley, M.E., Annett, G., Brooks, J.S., and el-Khoureiy, A. (1995). Engraftment of gene-modified umbilical cord blood cells in neonates with adenosine deaminase deficiency. Nat. Med. *1*, 1017–1023. .

Kohn, D.B., Hershfield, M.S., Puck, J.M., Aiuti, A., Blincoe, A., Gaspar, H.B., Notarangelo, L.D., and Grunebaum, E. (2019). Consensus approach for the management of severe combined immune deficiency caused by adenosine deaminase deficiency. J. Allergy Clin. Immunol. *143*, 852–863. https://doi.org/10.1016/j.jaci.2018.08.024.

Kohn, D.B., Booth, C., Kang, E.M., Pai, S.-Y., Shaw, K.L., Santilli, G., Armant, M., Buckland, K.F., Choi, U., De Ravin, S.S., et al. (2020). Lentiviral gene therapy for X-linked chronic granulomatous disease. Nat. Med. *26*, 200–206. https://doi.org/10.1038/s41591-019-0735-5.

Kollmann, K., Heller, G., Schneckenleithner, C., Warsch, W., Scheicher, R., Ott, R.G., Schäfer, M., Fajmann, S., Schlederer, M., Schiefer, A.-I., et al. (2013). A Kinase-Independent Function of CDK6 Links the Cell Cycle to Tumor Angiogenesis. Cancer Cell *24*, 167–181. https://doi.org/10.1016/j.ccr.2013.07.012.

Kondo, M., Weissman, I.L., and Akashi, K. (1997). Identification of Clonogenic Common Lymphoid Progenitors in Mouse Bone Marrow. Cell *91*, 661–672. https://doi.org/10.1016/S0092-8674(00)80453-5.

Kruta, M., Sunshine, M.J., Chua, B.A., Fu, Y., Chawla, A., Dillingham, C.H., Jose, L.H.S., Jong, B.D., Zhou, F.J., and Signer, R.A.J. (2021). Hsf1 promotes hematopoietic stem cell fitness and proteostasis in response to ex vivo culture stress and aging. Cell Stem Cell *28*, 1950-1965.e6. https://doi.org/10.1016/j.stem.2021.07.009.

Kume, A., Koremoto, M., Mizukami, H., Okada, T., Hanazono, Y., Sugamura, K., and Ozawa, K. (2002). Selective growth advantage of wild-type lymphocytes in X-linked SCID recipients. Bone Marrow Transplant. *30*, 113–118. https://doi.org/10.1038/sj.bmt.1703613.

Lambolez, B., Audinat, E., Bochet, P., Crépel, F., and Rossier, J. (1992). AMPA receptor subunits expressed by single Purkinje cells. Neuron *9*, 247–258. https://doi.org/10.1016/0896-6273(92)90164-9.

Lansdorp, P., Sutherland, H., and Eaves, C.J. (1990). Selective expression of CD45 isoforms on functional subpopulations of CD34+ hemopoietic cells from human bone marrow. J. Exp. Med. *172*, 363–366. .

Larochelle, A., Gillette, J.M., Desmond, R., Ichwan, B., Cantilena, A., Cerf, A., Barrett, A.J., Wayne, A.S., Lippincott-Schwartz, J., and Dunbar, C.E. (2012). Bone marrow homing and engraftment of human hematopoietic stem and progenitor cells is mediated by a polarized membrane domain. Blood *119*, 1848–1855. https://doi.org/10.1182/blood-2011-08-371583.

Laurenti, E., and Gottgens, B. (2018). From haematopoietic stem cells to complex differentiation landscapes. 9. .

Laurenti, E., Varnum-Finney, B., Wilson, A., Ferrero, I., Blanco-Bose, W.E., Ehninger, A., Knoepfler, P.S., Cheng, P.-F., MacDonald, H.R., Eisenman, R.N., et al. (2008). Hematopoietic Stem Cell Function and Survival Depend on c-Myc and N-Myc Activity. Cell Stem Cell *3*, 611–624. https://doi.org/10.1016/j.stem.2008.09.005.

Laurenti, E., Doulatov, S., Zandi, S., Plumb, I., Chen, J., April, C., Fan, J.-B., and Dick, J.E. (2013). The transcriptional architecture of early human hematopoiesis identifies multilevel control of lymphoid commitment. Nat. Immunol. *14*, 756–763. https://doi.org/10.1038/ni.2615.

Laurenti, E., Frelin, C., Xie, S., Ferrari, R., Dunant, C.F., Zandi, S., Neumann, A., Plumb, I., Doulatov, S., Chen, J., et al. (2015). CDK6 Levels Regulate Quiescence Exit in Human Hematopoietic Stem Cells. Cell Stem Cell *16*, 302–313. https://doi.org/10.1016/j.stem.2015.01.017.

Lechman, E.R., Gentner, B., van Galen, P., Giustacchini, A., Saini, M., Boccalatte, F.E., Hiramatsu, H., Restuccia, U., Bachi, A., Voisin, V., et al. (2012). Attenuation of miR-126 Activity Expands HSC In Vivo without Exhaustion. Cell Stem Cell *11*, 799–811. https://doi.org/10.1016/j.stem.2012.09.001.

Lemieux, M.E., Rebel, V.I., Lansdorp, P.M., and Eaves, C.J. (1995). Characterization and purification of a primitive hematopoietic cell type in adult mouse marrow capable of lymphomyeloid differentiation in long-term marrow "switch" cultures. Blood *86*, 1339–1347. .

Li, Q.V., Dixon, G., Verma, N., Rosen, B.P., Gordillo, M., Luo, R., Xu, C., Wang, Q., Soh, C.-L., Yang, D., et al. (2019). Genome-scale screens identify JNK–JUN signaling as a barrier for pluripotency exit and endoderm differentiation. Nat. Genet. *51*, 999–1010. https://doi.org/10.1038/s41588-019-0408-9.

Lim, J., Bae, S.-C., and Kim, K. (2019). Understanding HLA associations from SNP summary association statistics. Sci. Rep. *9*, 1337. https://doi.org/10.1038/s41598-018-37840-9.

Liu, B., Li, C., Li, Z., Wang, D., Ren, X., and Zhang, Z. (2020a). An entropy-based metric for assessing the purity of single cell populations. Nat. Commun. *11*, 3155. https://doi.org/10.1038/s41467-020-16904-3.

Liu, L., Inoki, A., Fan, K., Mao, F., Shi, G., Jin, X., Zhao, M., Ney, G., Jones, M., Sun, S., et al. (2020b). ER-associated degradation preserves hematopoietic stem cell quiescence and self-renewal by restricting mTOR activity. Blood *136*, 2975–2986. https://doi.org/10.1182/blood.2020007975.

Lo Celso, C., Fleming, H.E., Wu, J.W., Zhao, C.X., Miake-Lye, S., Fujisaki, J., Côté, D., Rowe, D.W., Lin, C.P., and Scadden, D.T. (2009). Live-animal tracking of individual haematopoietic stem/progenitor cells in their niche. Nature *457*, 92–96. https://doi.org/10.1038/nature07434.

Loeffler, D., Wehling, A., Schneiter, F., Zhang, Y., Müller-Bötticher, N., Hoppe, P.S., Hilsenbeck, O., Kokkaliaris, K.D., Endele, M., and Schroeder, T. (2019). Asymmetric lysosome inheritance predicts activation of haematopoietic stem cells. Nature *573*, 426–429. https://doi.org/10.1038/s41586-019-1531-6.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*. https://doi.org/10.1186/s13059-014-0550-8.

Lucas, D., Battista, M., Shi, P.A., Isola, L., and Frenette, P.S. (2008). Mobilized hematopoietic stem cell yield depends on species-specific circadian timing. Cell Stem Cell *3*, 364–366. https://doi.org/10.1016/j.stem.2008.09.004.

Ludin, A., Itkin, T., Gur-Cohen, S., Mildner, A., Shezen, E., Golan, K., Kollet, O., Kalinkovich, A., Porat, Z., D'Uva, G., et al. (2012). Monocytes-macrophages that express α-smooth muscle actin preserve primitive hematopoietic cells in the bone marrow. Nat. Immunol. *13*, 1072–1082. https://doi.org/10.1038/ni.2408.

Macaulay, I.C., Svensson, V., Labalette, C., Ferreira, L., Hamey, F., Voet, T., Teichmann, S.A., and Cvejic, A. (2016). Single-Cell RNA-Sequencing Reveals a Continuous Spectrum of Differentiation in Hematopoietic Cells. Cell Rep. *14*, 966–977. https://doi.org/10.1016/j.celrep.2015.12.082.

Magnusson, M., Sierra, M.I., Sasidharan, R., Prashad, S.L., Romero, M., Saarikoski, P., Handel, B.V., Huang, A., Li, X., and Mikkola, H.K.A. (2013). Expansion on Stromal Cells Preserves the Undifferentiated State of Human Hematopoietic Stem Cells Despite Compromised Reconstitution Ability. PLOS ONE *8*, e53912. https://doi.org/10.1371/journal.pone.0053912.

Mamcarz, E., Zhou, S., Lockey, T., Abdelsamed, H., Cross, S.J., Kang, G., Ma, Z., Condori, J., Dowdy, J., Triplett, B., et al. (2019). Lentiviral Gene Therapy Combined with Low-Dose Busulfan in Infants with SCID-X1. N. Engl. J. Med. *380*, 1525–1534. https://doi.org/10.1056/NEJMoa1815408.

Manesia, J.K., Xu, Z., Broekaert, D., Boon, R., van Vliet, A., Eelen, G., Vanwelden, T., Stegen, S., Van Gastel, N., Pascual-Montano, A., et al. (2015). Highly proliferative primitive fetal liver hematopoietic stem cells are fueled by oxidative metabolic pathways. Stem Cell Res. *15*, 715–721. https://doi.org/10.1016/j.scr.2015.11.001.

Marktel, S., Scaramuzza, S., Cicalese, M.P., Giglio, F., Galimberti, S., Lidonnici, M.R., Calbi, V., Assanelli, A., Bernardo, M.E., Rossi, C., et al. (2019). Intrabone hematopoietic stem cell gene therapy for adult and pediatric patients affected by transfusion-dependent ß-thalassemia. Nat. Med. *25*, 234–241. https://doi.org/10.1038/s41591-018-0301-6.

Mason, E.A., Mar, J.C., Laslett, A.L., Pera, M.F., Quackenbush, J., Wolvetang, E., and Wells, C.A. (2014). Gene Expression Variability as a Unifying Element of the Pluripotency Network. Stem Cell Rep. *3*, 365–377. https://doi.org/10.1016/j.stemcr.2014.06.008.

Matreyek, K.A., and Engelman, A. (2013). Viral and Cellular Requirements for the Nuclear Entry of Retroviral Preintegration Nucleoprotein Complexes. Viruses *5*, 2483–2511. https://doi.org/10.3390/v5102483.

Matsumoto, A., Takeishi, S., Kanie, T., Susaki, E., Onoyama, I., Tateishi, Y., Nakayama, K., and Nakayama, K.I. (2011). p57 Is Required for Quiescence and Maintenance of Adult Hematopoietic Stem Cells. Cell Stem Cell *9*, 262–271. https://doi.org/10.1016/j.stem.2011.06.014.

May, C., Rivella, S., Callegari, J., Heller, G., Gaensler, K.M.L., Luzzatto, L., and Sadelain, M. (2000). Therapeutic haemoglobin synthesis in β-thalassaemic mice expressing lentivirus-encoded human β-globin. Nature *406*, 82–86. https://doi.org/10.1038/35017565.

Mayani, H., Dragowska, W., and Lansdorp, P.M. (1993). Characterization of functionally distinct subpopulations of CD34+ cord blood cells in serum-free long-term cultures supplemented with hematopoietic cytokines. Blood *82*, 2664–2672. .

Mende, N., Bastos, H.P., Santoro, A., Mahbubani, K.T., Ciaurro, V., Calderbank, E.F., Quiroga Londoño, M., Sham, K., Mantica, G., Morishima, T., et al. (2022). Unique molecular and functional features of extramedullary hematopoietic stem and progenitor cell reservoirs in humans. Blood *139*, 3387–3401. https://doi.org/10.1182/blood.2021013450.

Méndez-Ferrer, S., Lucas, D., Battista, M., and Frenette, P.S. (2008). Haematopoietic stem cell release is regulated by circadian oscillations. Nature *452*, 442–447. https://doi.org/10.1038/nature06685.

Méndez-Ferrer, S., Michurina, T.V., Ferraro, F., Mazloom, A.R., MacArthur, B.D., Lira, S.A., Scadden, D.T., Ma'ayan, A., Enikolopov, G.N., and Frenette, P.S. (2010). Mesenchymal and haematopoietic stem cells form a unique bone marrow niche. Nature *466*, 829–834. https://doi.org/10.1038/nature09262.

Menendez-Gonzalez, J.B., Vukovic, M., Abdelfattah, A., Saleh, L., Almotiri, A., Thomas, L., Agirre-Lizaso, A., Azevedo, A., Menezes, A.C., Tornillo, G., et al. (2019). Gata2 as a Crucial Regulator of Stem Cells in Adult Hematopoiesis and Acute Myeloid Leukemia. Stem Cell Rep. *13*, 291–306. https://doi.org/10.1016/j.stemcr.2019.07.005.

Mian, S.A., Anjos-Afonso, F., and Bonnet, D. (2021). Advances in Human Immune System Mouse Models for Studying Human Hematopoiesis and Cancer Immunotherapy. Front. Immunol. *11*, 3753. https://doi.org/10.3389/fimmu.2020.619236.

Miharada, K., Sigurdsson, V., and Karlsson, S. (2014). Dppa5 Improves Hematopoietic Stem Cell Activity by Reducing Endoplasmic Reticulum Stress. Cell Rep. *7*, 1381–1392. https://doi.org/10.1016/j.celrep.2014.04.056.

Milyavsky, M., Gan, O.I., Trottier, M., Komosa, M., Tabach, O., Notta, F., Lechman, E., Hermans, K.G., Eppert, K., Konovalova, Z., et al. (2010). A Distinctive DNA Damage Response in Human Hematopoietic Stem Cells Reveals an Apoptosis-Independent Role for p53 in Self-Renewal. Cell Stem Cell *7*, 186–197. https://doi.org/10.1016/j.stem.2010.05.016.

Min, I.M., Pietramaggiori, G., Kim, F.S., Passegué, E., Stevenson, K.E., and Wagers, A.J. (2008). The Transcription Factor EGR1 Controls Both the Proliferation and Localization of Hematopoietic Stem Cells. Cell Stem Cell *2*, 380–391. https://doi.org/10.1016/j.stem.2008.01.015.

Mitchell, E., Spencer Chapman, M., Williams, N., Dawson, K.J., Mende, N., Calderbank, E.F., Jung, H., Mitchell, T., Coorens, T.H.H., Spencer, D.H., et al. (2022). Clonal dynamics of haematopoiesis across the human lifespan. Nature *606*, 343–350. https://doi.org/10.1038/s41586-022-04786-y.

Miyamoto, K., Araki, K.Y., Naka, K., Arai, F., Takubo, K., Yamazaki, S., Matsuoka, S., Miyamoto, T., Ito, K., Ohmura, M., et al. (2007). Foxo3a Is Essential for Maintenance of the Hematopoietic Stem Cell Pool. Cell Stem Cell *1*, 101–112. https://doi.org/10.1016/j.stem.2007.02.001.

Mohrin, M., Bourke, E., Alexander, D., Warr, M.R., Barry-Holson, K., Le Beau, M.M., Morrison, C.G., and Passegué, E. (2010). Hematopoietic Stem Cell Quiescence Promotes Error-Prone DNA Repair and Mutagenesis. Cell Stem Cell *7*, 174–185. https://doi.org/10.1016/j.stem.2010.06.014.

Mohrin, M., Shin, J., Liu, Y., Brown, K., Luo, H., Xi, Y., Haynes, C.M., and Chen, D. (2015). Stem cell aging. A mitochondrial UPR-mediated metabolic checkpoint regulates hematopoietic stem cell aging. Science *347*, 1374–1377. https://doi.org/10.1126/science.aaa2361.

Moignard, V., Macaulay, I.C., Swiers, G., Buettner, F., Schütte, J., Calero-Nieto, F.J., Kinston, S., Joshi, A., Hannah, R., Theis, F.J., et al. (2013). Characterisation of transcriptional networks in blood stem and progenitor cells using high-throughput single cell gene expression analysis. Nat. Cell Biol. *15*, 363–372. https://doi.org/10.1038/ncb2709.

Mojica, F.J., Juez, G., and Rodríguez-Valera, F. (1993). Transcription at different salinities of Haloferax mediterranei sequences adjacent to partially modified PstI sites. Mol. Microbiol. *9*, 613–621. https://doi.org/10.1111/j.1365-2958.1993.tb01721.x.

Mojica, F.J.M., Díez-Villaseñor, C., García-Martínez, J., and Soria, E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. J. Mol. Evol. *60*, 174–182. https://doi.org/10.1007/s00239-004-0046-3.

Montini, E., Cesana, D., Schmidt, M., Sanvito, F., Bartholomae, C.C., Ranzani, M., Benedicenti, F., Sergi, L.S., Ambrosi, A., Ponzoni, M., et al. (2009). The genotoxic potential of retroviral vectors is strongly modulated by vector design and integration site selection in a mouse model of HSC gene therapy. J. Clin. Invest. *119*, 964–975. https://doi.org/10.1172/JCI37630.

Morgan, R.A., Gray, D., Lomova, A., and Kohn, D.B. (2017). Hematopoietic Stem Cell Gene Therapy: Progress and Lessons Learned. Cell Stem Cell *21*, 574–590. https://doi.org/10.1016/j.stem.2017.10.010.

Morita, Y., Ema, H., and Nakauchi, H. (2010). Heterogeneity and hierarchy within the most primitive hematopoietic stem cell compartment. J. Exp. Med. *207*, 1173–1182. https://doi.org/10.1084/jem.20091318.

Morrison, S.J., Wright, D.E., and Weissman, I.L. (1997). Cyclophosphamide/granulocyte colony-stimulating factor induces hematopoietic stem cells to proliferate prior to mobilization. Proc. Natl. Acad. Sci. U. S. A. *94*, 1908–1913. https://doi.org/10.1073/pnas.94.5.1908.

Mossadegh-Keller, N., Sarrazin, S., Kandalla, P.K., Espinosa, L., Stanley, E.R., Nutt, S.L., Moore, J., and Sieweke, M.H. (2013). M-CSF instructs myeloid lineage fate in single haematopoietic stem cells. Nature *497*, 239–243. https://doi.org/10.1038/nature12026.

Müller-Sieburg, C.E., Cho, R.H., Thoman, M., Adkins, B., and Sieburg, H.B. (2002). Deterministic regulation of hematopoietic stem cell self-renewal and differentiation. Blood *100*, 1302–1309. https://doi.org/10.1182/blood.V100.4.1302.h81602001302_1302_1309.

Muller-Sieburg, C.E., Cho, R.H., Karlsson, L., Huang, J.-F., and Sieburg, H.B. (2004). Myeloid-biased hematopoietic stem cells have extensive self-renewal capacity but generate diminished lymphoid progeny with impaired IL-7 responsiveness. Blood *103*, 4111–4118. https://doi.org/10.1182/blood-2003-10-3448.

Nakahara, F., Borger, D.K., Wei, Q., Pinho, S., Maryanovich, M., Zahalka, A.H., Suzuki, M., Cruz, C.D., Wang, Z., Xu, C., et al. (2019). Engineering a haematopoietic stem cell niche by revitalizing mesenchymal stromal cells. Nat. Cell Biol. *21*, 560–567. https://doi.org/10.1038/s41556-019-0308-3.

Nakajima, H., Ito, M., Smookler, D.S., Shibata, F., Fukuchi, Y., Morikawa, Y., Ikeda, Y., Arai, F., Suda, T., Khokha, R., et al. (2010). TIMP-3 recruits quiescent hematopoietic stem cells into active cell cycle and expands multipotent progenitor pool. Blood *116*, 4474–4482. https://doi.org/10.1182/blood-2010-01-266528.

Nakamura-Ishizu, A., Takizawa, H., and Suda, T. (2014). The analysis, roles and regulation of quiescence in hematopoietic stem cells. Development *141*, 4656–4666. https://doi.org/10.1242/dev.106575.

Naldini, L., Blömer, U., Gallay, P., Ory, D., Mulligan, R., Gage, F.H., Verma, I.M., and Trono, D. (1996). In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. Science *272*, 263–267. https://doi.org/10.1126/science.272.5259.263.

Nestorowa, S., Hamey, F.K., Pijuan Sala, B., Diamanti, E., Shepherd, M., Laurenti, E., Wilson, N.K., Kent, D.G., and Göttgens, B. (2016). A single-cell resolution map of mouse hematopoietic stem and progenitor cell differentiation. Blood *128*, e20–e31. https://doi.org/10.1182/blood-2016-05-716480.

Nicolini, F.E., Cashman, J.D., Hogge, D.E., Humphries, R.K., and Eaves, C.J. (2004). NOD/SCID mice engineered to express human IL-3, GM-CSF and Steel factor constitutively mobilize engrafted human progenitors and compromise human stem cell regeneration. Leukemia *18*, 341–347. https://doi.org/10.1038/sj.leu.2403222.

Notta, F., Doulatov, S., Laurenti, E., Poeppl, A., Jurisica, I., and Dick, J.E. (2011). Isolation of Single Human Hematopoietic Stem Cells Capable of Long-Term Multilineage Engraftment. Science *333*, 218–221. https://doi.org/10.1126/science.1201219.

Notta, F., Zandi, S., Takayama, N., Dobson, S., Gan, O.I., Wilson, G., Kaufmann, K.B., McLeod, J., Laurenti, E., Dunant, C.F., et al. (2016). Distinct routes of lineage development reshape the human blood hierarchy across ontogeny. Science *351*, aab2116. https://doi.org/10.1126/science.aab2116.

Oguro, H., Ding, L., and Morrison, S.J. (2013). SLAM family markers resolve functionally distinct subpopulations of hematopoietic stem cells and multipotent progenitors. Cell Stem Cell *13*, 102–116. https://doi.org/10.1016/j.stem.2013.05.014.

Ohbo, K., Suda, T., Hashiyama, M., Mantani, A., Ikebe, M., Miyakawa, K., Moriyama, M., Nakamura, M., Katsuki, M., Takahashi, K., et al. (1996). Modulation of Hematopoiesis in Mice With a Truncated Mutant of the Interleukin-2 Receptor γ Chain. Blood *87*, 956–967. https://doi.org/10.1182/blood.V87.3.956.bloodjournal873956.

Okada, S., Fukuda, T., Inada, K., and Tokuhisa, T. (1999). Prolonged Expression of c-fos Suppresses Cell Cycle Entry of Dormant Hematopoietic Stem Cells. 11. .

Omer-Javed, A., Pedrazzani, G., Albano, L., Ghaus, S., Latroche, C., Manzi, M., Ferrari, S., Fiumara, M., Jacob, A., Vavassori, V., et al. (2022). Mobilization-based chemotherapy-free engraftment of gene-edited human hematopoietic stem cells. Cell *185*, 2248-2264.e21. https://doi.org/10.1016/j.cell.2022.04.039.

Orchard therapeutics (2020). Press release on T cell leukemia in ADA-SCID gene therapy treated patient. Available at: https://ir.orchard-tx.com/news-releases/news-release-details/orchard-statement-strimvelisr-gamma-retroviral-vector-based-gene.

Orford, K.W., and Scadden, D.T. (2008a). Deconstructing stem cell self-renewal: genetic insights into cell-cycle regulation. Nat. Rev. Genet. *9*, 115–128. https://doi.org/10.1038/nrg2269.

Orford, K.W., and Scadden, D.T. (2008b). Deconstructing stem cell self-renewal: genetic insights into cell-cycle regulation. Nat. Rev. Genet. *9*, 115–128. https://doi.org/10.1038/nrg2269.

Orlic, D., Fischer, R., Nishikawa, S., Nienhuis, A.W., and Bodine, D.M. (1993). Purification and characterization of heterogeneous pluripotent hematopoietic stem cell populations expressing high levels of c-kit receptor. Blood *82*, 762–770. .

Osawa, M., Hanada, K., Hamada, H., and Nakauchi, H. (1996). Long-term lymphohematopoietic reconstitution by a single CD34-low/negative hematopoietic stem cell. Science *273*, 242–245. .

Ott, M.G., Schmidt, M., Schwarzwaelder, K., Stein, S., Siler, U., Koehl, U., Glimm, H., Kühlcke, K., Schilz, A., Kunkel, H., et al. (2006). Correction of X-linked chronic granulomatous disease by gene therapy, augmented by insertional activation of MDS1-EVI1, PRDM16 or SETBP1. Nat. Med. *12*, 401–409. https://doi.org/10.1038/nm1393.

Pack, L.R., Daigh, L.H., Chung, M., and Meyer, T. (2021). Clinical CDK4/6 inhibitors induce selective and immediate dissociation of p21 from cyclin D-CDK4 to inhibit CDK2. Nat. Commun. *12*, 3356. https://doi.org/10.1038/s41467-021-23612-z.

Pardee, A.B. (1974). A Restriction Point for Control of Normal Animal Cell Proliferation. Proc. Natl. Acad. Sci. *71*, 1286–1290. https://doi.org/10.1073/pnas.71.4.1286.

Passegué, E., Wagner, E.F., and Weissman, I.L. (2004). JunB Deficiency Leads to a Myeloproliferative Disorder Arising from Hematopoietic Stem Cells. Cell *119*, 431–443. https://doi.org/10.1016/j.cell.2004.10.010.

Passegué, E., Wagers, A.J., Giuriato, S., Anderson, W.C., and Weissman, I.L. (2005). Global analysis of proliferation and cell cycle gene expression in the regulation of hematopoietic stem and progenitor cell fates. J. Exp. Med. *202*, 1599–1611. https://doi.org/10.1084/jem.20050967.

Patano, L. (2017). degPatterns. Available at: https://rdrr.io/github/lpantano/DEGreport/man/degPatterns.html.

Paul, F., Arkin, Y., Giladi, A., Jaitin, D.A., Kenigsberg, E., Keren-Shaul, H., Winter, D., Lara-Astiaso, D., Gury, M., Weiner, A., et al. (2015). Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors. Cell *163*, 1663–1677. https://doi.org/10.1016/j.cell.2015.11.013.

Pei, W., Shang, F., Wang, X., Fanti, A.-K., Greco, A., Busch, K., Klapproth, K., Zhang, Q., Quedenau, C., Sauer, S., et al. (2020). Resolving Fates and Single-Cell Transcriptomes of Hematopoietic Stem Cell Clones by PolyloxExpress Barcoding. Cell Stem Cell *27*, 383-395.e8. https://doi.org/10.1016/j.stem.2020.07.018.

Pennycook, B.R., and Barr, A.R. (2020). Restriction point regulation at the crossroads between quiescence and cell proliferation. FEBS Lett. *594*, 2046–2060. https://doi.org/10.1002/1873-3468.13867.

Petzer, A.L., Hogge, D.E., Landsdorp, P.M., Reid, D.S., and Eaves, C.J. (1996). Self-renewal of primitive human hematopoietic cells (long-term-culture-initiating cells) in vitro and their expansion in defined medium. Proc. Natl. Acad. Sci. *93*, 1470–1474. https://doi.org/10.1073/pnas.93.4.1470.

Picelli, S., Faridani, O.R., Björklund, A.K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. Nat. Protoc. *9*, 171–181. https://doi.org/10.1038/nprot.2014.006.

Pinho, S., and Frenette, P.S. (2019). Haematopoietic stem cell activity and interactions with the niche. Nat. Rev. Mol. Cell Biol. *20*, 303–320. https://doi.org/10.1038/s41580-019-0103-9.

Popescu, D.-M., Botting, R.A., Stephenson, E., Green, K., Webb, S., Jardine, L., Calderbank, E.F., Polanski, K., Goh, I., Efremova, M., et al. (2019). Decoding human fetal liver haematopoiesis. Nature *574*, 365–371. https://doi.org/10.1038/s41586-019-1652-y.

Prosper, F., Vanoverbeke, K., Stroncek, D., and Verfaillie, C.M. (1997). Primitive Long-Term Culture Initiating Cells (LTC-ICs) in Granulocyte Colony-Stimulating Factor Mobilized Peripheral Blood Progenitor Cells Have Similar Potential for Ex Vivo Expansion as Primitive LTC-ICs in Steady State Bone Marrow. Blood *89*, 3991–3997. https://doi.org/10.1182/blood.V89.11.3991.

Psaila, B., Barkas, N., Iskander, D., Roy, A., Anderson, S., Ashley, N., Caputo, V.S., Lichtenberg, J., Loaiza, S., Bodine, D.M., et al. (2016). Single-cell profiling of human megakaryocyte-erythroid progenitors identifies distinct megakaryocyte and erythroid differentiation pathways. Genome Biol. *17*, 83. https://doi.org/10.1186/s13059-016-0939-7.

Qian, H., Buza-Vidas, N., Hyland, C.D., Jensen, C.T., Antonchuk, J., Månsson, R., Thoren, L.A., Ekblom, M., Alexander, W.S., and Jacobsen, S.E.W. (2007). Critical Role of Thrombopoietin in Maintaining Adult Quiescent Hematopoietic Stem Cells. Cell Stem Cell *1*, 671–684. https://doi.org/10.1016/j.stem.2007.10.008.

Qiu, J., Papatsenko, D., Niu, X., Schaniel, C., and Moore, K. (2014). Divisional History and Hematopoietic Stem Cell Function during Homeostasis. Stem Cell Rep. *2*, 473–490. https://doi.org/10.1016/j.stemcr.2014.01.016.

Robertson, K.D., Ait-Si-Ali, S., Yokochi, T., Wade, P.A., Jones, P.L., and Wolffe, A.P. (2000). DNMT1 forms a complex with Rb, E2F1 and HDAC1 and represses transcription from E2F-responsive promoters. Nat. Genet. *25*, 338–342. https://doi.org/10.1038/77124.

Rodgers, J.T., King, K.Y., Brett, J.O., Cromie, M.J., Charville, G.W., Maguire, K.K., Brunson, C., Mastey, N., Liu, L., Tsai, C.-R., et al. (2014). mTORC1 controls the adaptive transition of quiescent stem cells from G0 to GAlert. Nature *510*, 393–396. https://doi.org/10.1038/nature13255.

Rodriguez-Fraticelli, A.E., Wolock, S.L., Weinreb, C.S., Panero, R., Patel, S.H., Jankovic, M., Sun, J., Calogero, R.A., Klein, A.M., and Camargo, F.D. (2018). Clonal analysis of lineage fate in native haematopoiesis. Nature *553*, 212–216. https://doi.org/10.1038/nature25168.

Rodriguez-Fraticelli, A.E., Weinreb, C., Wang, S.-W., Migueles, R.P., Jankovic, M., Usart, M., Klein, A.M., Lowell, S., and Camargo, F.D. (2020). Single-cell lineage tracing unveils a role for TCF15 in haematopoiesis. Nature *583*, 585–589. https://doi.org/10.1038/s41586-020-2503-6.

Romero, Z., Lomova, A., Said, S., Miggelbrink, A., Kuo, C.Y., Campo-Fernandez, B., Hoban, M.D., Masiuk, K.E., Clark, D.N., Long, J., et al. (2019). Editing the Sickle Cell Disease Mutation in Human Hematopoietic Stem Cells: Comparison of Endonucleases and Homologous Donor Templates. Mol. Ther. J. Am. Soc. Gene Ther. *27*, 1389–1406. https://doi.org/10.1016/j.ymthe.2019.05.014.

Rongvaux, A., Willinger, T., Martinek, J., Strowig, T., Gearty, S.V., Teichmann, L.L., Saito, Y., Marches, F., Halene, S., Palucka, A.K., et al. (2014). Development and function of human innate immune cells in a humanized mouse model. Nat. Biotechnol. *32*, 364–372. https://doi.org/10.1038/nbt.2858.

Rossi, L., Lin, K.K., Boles, N.C., Yang, L., King, K.Y., Jeong, M., Mayle, A., and Goodell, M.A. (2012). Less is More: unveiling the functional core of hematopoietic stem cells through knockout mice. Cell Stem Cell *11*, 302–317. https://doi.org/10.1016/j.stem.2012.08.006.

Sagoo, P., and Gaspar, H.B. (2021). The transformative potential of HSC gene therapy as a genetic medicine. Gene Ther. 1–19. https://doi.org/10.1038/s41434-021-00261-x.

Sanjuan-Pla, A., Macaulay, I.C., Jensen, C.T., Woll, P.S., Luis, T.C., Mead, A., Moore, S., Carella, C., Matsuoka, S., Jones, T.B., et al. (2013). Platelet-biased stem cells reside at the apex of the haematopoietic stem-cell hierarchy. Nature *502*, 232–236. https://doi.org/10.1038/nature12495.

Santaguida, M., Schepers, K., King, B., Sabnis, A.J., Forsberg, E.C., Attema, J.L., Braun, B.S., and Passegué, E. (2009). JunB protects against myeloid malignancies by limiting hematopoietic stem cell proliferation and differentiation without affecting self-renewal. Cancer Cell *15*, 341–352. https://doi.org/10.1016/j.ccr.2009.02.016.

Satija lab (2017). Cell-cycle scoring and regression. Available at: https://satijalab.org/seurat/archive/v3.1/cell_cycle_vignette.html.

Scala, S., Basso-Ricci, L., Dionisio, F., Pellin, D., Giannelli, S., Salerio, F.A., Leonardelli, L., Cicalese, M.P., Ferrua, F., Aiuti, A., et al. (2018). Dynamics of genetically engineered hematopoietic stem and progenitor cells after autologous transplantation in humans. Nat. Med. *24*, 1683–1690. https://doi.org/10.1038/s41591-018-0195-3.

Scheicher, R., Hoelbl-Kovacic, A., Bellutti, F., Tigan, A.-S., Prchal-Murphy, M., Heller, G., Schneckenleithner, C., Salazar-Roa, M., Zöchbauer-Müller, S., Zuber, J., et al. (2015). CDK6 as a key regulator of hematopoietic and leukemic stem cell activation. Blood *125*, 90–101. https://doi.org/10.1182/blood-2014-06-584417.

Schiroli, G., Conti, A., Ferrari, S., Volpe, L. della, Jacob, A., Albano, L., Beretta, S., Calabria, A., Vavassori, V., Gasparini, P., et al. (2019). Precise Gene Editing Preserves Hematopoietic Stem Cell Function following Transient p53-Mediated DNA Damage Response. Cell Stem Cell *24*, 551-565.e8. https://doi.org/10.1016/j.stem.2019.02.019.

Schönberger, K., Obier, N., Romero-Mulero, M.C., Cauchy, P., Mess, J., Pavlovich, P.V., Zhang, Y.W., Mitterer, M., Rettkowski, J., Lalioti, M.-E., et al. (2021). Multilayer omics analysis reveals a non-classical retinoic acid signaling axis that regulates hematopoietic stem cell identity. Cell Stem Cell *0*. https://doi.org/10.1016/j.stem.2021.10.002.

Schreiber, M., Kolbus, A., Piu, F., Szabowski, A., Möhle-Steinlein, U., Tian, J., Karin, M., Angel, P., and Wagner, E.F. (1999). Control of cell cycle progression by c-Jun is p53 dependent. Genes Dev. *13*, 607–619. .

Schroder, B. (2016). The multifaceted roles of the invariant chain CD74--More than just a chaperone. Biochim. Biophys. Acta *1863*, 1269–1281. https://doi.org/10.1016/j.bbamcr.2016.03.026.

Seita, J., Ema, H., Ooehara, J., Yamazaki, S., Tadokoro, Y., Yamasaki, A., Eto, K., Takaki, S., Takatsu, K., and Nakauchi, H. (2007). Lnk negatively regulates self-renewal of hematopoietic stem cells by modifying thrombopoietin-mediated signal transduction. Proc. Natl. Acad. Sci. *104*, 2349–2354. https://doi.org/10.1073/pnas.0606238104.

Shaulian, E., and Karin, M. (2001). AP-1 in cell proliferation and survival. Oncogene *20*, 2390–2400. https://doi.org/10.1038/sj.onc.1204383.

Shen, H., Cheng, T., Olszak, I., Garcia-Zepeda, E., Lu, Z., Herrmann, S., Fallon, R., Luster, A.D., and Scadden, D.T. (2001). CXCR-4 Desensitization Is Associated with Tissue Localization of Hemopoietic Progenitor Cells. J. Immunol. *166*, 5027–5033. https://doi.org/10.4049/jimmunol.166.8.5027.

Shin, J.Y., Hu, W., Naramura, M., and Park, C.Y. (2014). High c-Kit expression identifies hematopoietic stem cells with impaired self-renewal and megakaryocytic bias. J. Exp. Med. *211*, 217–231. https://doi.org/10.1084/jem.20131128.

Shultz, L.D., Lyons, B.L., Burzenski, L.M., Gott, B., Chen, X., Chaleff, S., Kotb, M., Gillies, S.D., King, M., Mangada, J., et al. (2005). Human lymphoid and myeloid cell development in NOD/LtSz-scid IL2R gamma null mice engrafted with mobilized human hemopoietic stem cells. J. Immunol. Baltim. Md 1950 *174*, 6477–6489. https://doi.org/10.4049/jimmunol.174.10.6477.

Signer, R.A.J., Magee, J.A., Salic, A., and Morrison, S.J. (2014). Haematopoietic stem cells require a highly regulated protein synthesis rate. Nature *509*, 49–54. https://doi.org/10.1038/nature13035.

Signer, R.A.J., Qi, L., Zhao, Z., Thompson, D., Sigova, A.A., Fan, Z.P., DeMartino, G.N., Young, R.A., Sonenberg, N., and Morrison, S.J. (2016). The rate of protein synthesis in hematopoietic stem cells is limited partly by 4E-BPs. Genes Dev. https://doi.org/10.1101/gad.282756.116.

Sigurdsson, V., Takei, H., Soboleva, S., Radulovic, V., Galeev, R., Siva, K., Leeb-Lundberg, L.M.F., Iida, T., Nittono, H., and Miharada, K. (2016). Bile Acids Protect Expanding Hematopoietic Stem Cells from Unfolded Protein Stress in Fetal Liver. Cell Stem Cell *18*, 522–532. https://doi.org/10.1016/j.stem.2016.01.002.

Simsek, T., Kocabas, F., Zheng, J., DeBerardinis, R.J., Mahmoud, A.I., Olson, E.N., Schneider, J.W., Zhang, C.C., and Sadek, H.A. (2010). The Distinct Metabolic Profile of Hematopoietic Stem Cells Reflects Their Location in a Hypoxic Niche. Cell Stem Cell *7*, 380–390. https://doi.org/10.1016/j.stem.2010.07.011.

Spangrude, G.J., Heimfeld, S., and Weissman, I.L. (1988). Purification and characterization of mouse hematopoietic stem cells. Science *241*, 58–62. https://doi.org/10.1126/science.2898810.

Staber, P.B., Zhang, P., Ye, M., Welner, R.S., Nombela-Arrieta, C., Bach, C., Kerenyi, M., Bartholdy, B.A., Zhang, H., Alberich-Jordà, M., et al. (2013). Sustained PU.1 Levels Balance Cell-Cycle Regulators to Prevent Exhaustion of Adult Hematopoietic Stem Cells. Mol. Cell *49*, 934–946. https://doi.org/10.1016/j.molcel.2013.01.007.

Stepanova, L., and Sorrentino, B.P. (2005). A limited role for p16Ink4a and p19Arf in the loss of hematopoietic stem cells during proliferative stress. Blood *106*, 827–832. https://doi.org/10.1182/blood-2004-06-2242.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. Cell *177*, 1888-1902.e21. https://doi.org/10.1016/j.cell.2019.05.031.

Suda, T., Takubo, K., and Semenza, G.L. (2011). Metabolic Regulation of Hematopoietic Stem Cells in the Hypoxic Niche. Cell Stem Cell *9*, 298–310. https://doi.org/10.1016/j.stem.2011.09.010.

Sun, J., Ramos, A., Chapman, B., Johnnidis, J.B., Le, L., Ho, Y.-J., Klein, A., Hofmann, O., and Camargo, F.D. (2014). Clonal dynamics of native haematopoiesis. Nature *514*, 322–327. https://doi.org/10.1038/nature13824.

Sutherland, H.J., Eaves, C.J., Eaves, A.C., Dragowska, W., and Lansdorp, P.M. (1989). Characterization and Partial Purification of Human Marrow Cells Capable of Initiating Long-Term Hematopoiesis In Vitro. Blood *74*, 1563–1570. https://doi.org/10.1182/blood.V74.5.1563.1563.

Svensson, V., Vento-Tormo, R., and Teichmann, S.A. (2018). Exponential scaling of single-cell RNA-seq in the past decade. Nat. Protoc. *13*, 599–604. https://doi.org/10.1038/nprot.2017.149.

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663–676. https://doi.org/10.1016/j.cell.2006.07.024.

Takahashi, T., Nowakowski, R.S., and Caviness, V.S. (1995). The cell cycle of the pseudostratified ventricular epithelium of the embryonic murine cerebral wall. J. Neurosci. *15*, 6046–6057. https://doi.org/10.1523/JNEUROSCI.15-09-06046.1995.

Takayama, N., Murison, A., Takayanagi, S., Arlidge, C., Zhou, S., Garcia-Prat, L., Chan-Seng-Yue, M., Zandi, S., Gan, O.I., Boutzen, H., et al. (2021). The Transition from Quiescent to Activated States in Human Hematopoietic Stem Cells Is Governed by Dynamic 3D Genome Reorganization. Cell Stem Cell *28*, 488-501.e10. https://doi.org/10.1016/j.stem.2020.11.001.

Takizawa, H., Regoes, R.R., Boddupalli, C.S., Bonhoeffer, S., and Manz, M.G. (2011). Dynamic variation in cycling of hematopoietic stem cells in steady state and inflammation. J. Exp. Med. *208*, 273–284. https://doi.org/10.1084/jem.20101643.

Takubo, K., Goda, N., Yamada, W., Iriuchishima, H., Ikeda, E., Kubota, Y., Shima, H., Johnson, R.S., Hirao, A., Suematsu, M., et al. (2010). Regulation of the HIF-1alpha level is essential for hematopoietic stem cells. Cell Stem Cell *7*, 391–402. https://doi.org/10.1016/j.stem.2010.06.020.

Takubo, K., Nagamatsu, G., Kobayashi, C.I., Nakamura-Ishizu, A., Kobayashi, H., Ikeda, E., Goda, N., Rahimi, Y., Johnson, R.S., Soga, T., et al. (2013). Regulation of Glycolysis by Pdk Functions as a Metabolic Checkpoint for Cell Cycle Quiescence in Hematopoietic Stem Cells. Cell Stem Cell *12*, 49–61. https://doi.org/10.1016/j.stem.2012.10.011.

Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B.B., Siddiqui, A., et al. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. Nat. Methods *6*, 377–382. https://doi.org/10.1038/nmeth.1315.

Theodore, L.N., Hagedorn, E.J., Cortes, M., Natsuhara, K., Liu, S.Y., Perlin, J.R., Yang, S., Daily, M.L., Zon, L.I., and North, T.E. (2017). Distinct Roles for Matrix Metalloproteinases 2 and 9 in Embryonic Hematopoietic Stem Cell Emergence, Migration, and Niche Colonization. Stem Cell Rep. *8*, 1226–1241. https://doi.org/10.1016/j.stemcr.2017.03.016.

Thio, C.L., Thomas, D.L., Goedert, J.J., Vlahov, D., Nelson, K.E., Hilgartner, M.W., O'Brien, S.J., Karacki, P., Marti, D., Astemborski, J., et al. (2001). Racial Differences in HLA Class II Associations with Hepatitis C Virus Outcomes. J. Infect. Dis. *184*, 16–21. https://doi.org/10.1086/321005.

Thompson, A.A., Walters, M.C., Kwiatkowski, J., Rasko, J.E.J., Ribeil, J.-A., Hongeng, S., Magrin, E., Schiller, G.J., Payen, E., Semeraro, M., et al. (2018). Gene Therapy in Patients with Transfusion-Dependent β-Thalassemia. N. Engl. J. Med. *378*, 1479–1493. https://doi.org/10.1056/NEJMoa1705342.

Till, J.E., and McCULLOCH, E.A. (1961). A direct measurement of the radiation sensitivity of normal mouse bone marrow cells. Radiat. Res. *14*, 213–222. .

Tirosh, I., Izar, B., Prakadan, S.M., Wadsworth, M.H., Treacy, D., Trombetta, J.J., Rotem, A., Rodman, C., Lian, C., Murphy, G., et al. (2016). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science *352*, 189–196. https://doi.org/10.1126/science.aad0501.

Tomellini, E., Fares, I., Lehnertz, B., Chagraoui, J., Mayotte, N., MacRae, T., Bordeleau, M.-È., Corneau, S., Bisaillon, R., and Sauvageau, G. (2019). Integrin-α3 Is a Functional Marker of Ex Vivo Expanded Human Long-Term Hematopoietic Stem Cells. Cell Rep. *28*, 1063-1073.e5. https://doi.org/10.1016/j.celrep.2019.06.084.

Tong, P., Chen, Y., Su, X., and Coombes, K.R. (2013). SIBER: systematic identification of bimodally expressed genes using RNAseq data. Bioinforma. Oxf. Engl. *29*, 605–613. https://doi.org/10.1093/bioinformatics/bts713.

Tsuchiya, Y., Nakabayashi, O., and Nakano, H. (2015). FLIP the Switch: Regulation of Apoptosis and Necroptosis by cFLIP. Int. J. Mol. Sci. *16*, 30321–30341. https://doi.org/10.3390/ijms161226232.

Tucci, F., Scaramuzza, S., Aiuti, A., and Mortellaro, A. (2021). Update on Clinical Ex Vivo Hematopoietic Stem Cell Gene Therapy for Inherited Monogenic Diseases. Mol. Ther. *29*, 489–504. https://doi.org/10.1016/j.ymthe.2020.11.020.

Vallejos, C.A., Marioni, J.C., and Richardson, S. (2015). BASiCS: Bayesian Analysis of Single-Cell Sequencing Data. PLOS Comput. Biol. *11*, e1004333. https://doi.org/10.1371/journal.pcbi.1004333.

Vannini, N., Girotra, M., Naveiras, O., Nikitin, G., Campos, V., Giger, S., Roch, A., Auwerx, J., and Lutolf, M.P. (2016). Specification of haematopoietic stem cell fate via modulation of mitochondrial activity. Nat. Commun. *7*, 13125. https://doi.org/10.1038/ncomms13125.

Velten, L., Haas, S.F., Raffel, S., Blaszkiewicz, S., Islam, S., Hennig, B.P., Hirche, C., Lutz, C., Buss, E.C., Nowak, D., et al. (2017). Human haematopoietic stem cell lineage commitment is a continuous process. Nat. Cell Biol. *19*, 271–281. https://doi.org/10.1038/ncb3493.

Velthoven, C.T.J. van, and Rando, T.A. (2019). Stem Cell Quiescence: Dynamism, Restraint, and Cellular Idling. Cell Stem Cell *24*, 213–225. https://doi.org/10.1016/j.stem.2019.01.001.

Viatour, P., Somervaille, T.C., Venkatasubrahmanyam, S., Kogan, S., McLaughlin, M.E., Weissman, I.L., Butte, A.J., Passegué, E., and Sage, J. (2008). Hematopoietic stem cell quiescence is maintained by compound contributions of the retinoblastoma gene family. Cell Stem Cell *3*, 416–428. https://doi.org/10.1016/j.stem.2008.07.009.

Wagers, A.J., Allsopp, R.C., and Weissman, I.L. (2002). Changes in integrin expression are associated with altered homing properties of Lin(-/lo)Thy1.1(lo)Sca-1(+)c-kit(+) hematopoietic stem cells following mobilization by cyclophosphamide/granulocyte colony-stimulating factor. Exp. Hematol. *30*, 176–185. https://doi.org/10.1016/s0301-472x(01)00777-9.

Wagner, J.E., Brunstein, C.G., Boitano, A.E., DeFor, T.E., McKenna, D., Sumstad, D., Blazar, B.R., Tolar, J., Le, C., Jones, J., et al. (2016). Phase I/II Trial of StemRegenin-1 Expanded Umbilical Cord Blood Hematopoietic Stem Cells Supports Testing as a Stand-Alone Graft. Cell Stem Cell *18*, 144–155. https://doi.org/10.1016/j.stem.2015.10.004.

Walters, M.C. (2015). Update of hematopoietic cell transplantation for sickle cell disease. Curr. Opin. Hematol. *22*, 227–233. https://doi.org/10.1097/MOH.0000000000000136.

Wang, S. (2020). smqpp. Available at: https://github.com/SharonWang/smqpp.

Wang, J.C.Y., and Dick, J.E. (2005). Cancer stem cells: lessons from leukemia. Trends Cell Biol. *15*, 494–501. https://doi.org/10.1016/j.tcb.2005.07.004.

Wang, G.P., Berry, C.C., Malani, N., Leboulch, P., Fischer, A., Hacein-Bey-Abina, S., Cavazzana-Calvo, M., and Bushman, F.D. (2010). Dynamics of gene-modified progenitor cells analyzed by tracking retroviral integration sites in a human SCID-X1 gene therapy trial. Blood *115*, 4356–4366. https://doi.org/10.1182/blood-2009-12-257352.

Wang, J.H., Zheng, X., Ke, X., Dorak, M.T., Shen, J., Boodram, B., O'Gorman, M., Beaman, K., Cotler, S.J., Hershow, R., et al. (2009). Ethnic and geographical differences in HLA associations with the outcome of hepatitis C virus infection. Virol. J. *6*, 46. https://doi.org/10.1186/1743-422X-6-46.

Warr, M.R., Binnewies, M., Flach, J., Reynaud, D., Garg, T., Malhotra, R., Debnath, J., and Passegué, E. (2013). FOXO3A directs a protective autophagy program in haematopoietic stem cells. Nature *494*, 323–327. https://doi.org/10.1038/nature11895.

Warren, L., Bryder, D., Weissman, I.L., and Quake, S.R. (2006). Transcription factor profiling in individual hematopoietic progenitors by digital RT-PCR. Proc. Natl. Acad. Sci. U. S. A. *103*, 17807–17812. https://doi.org/10.1073/pnas.0608512103.

Watcham, S., Kucinski, I., and Gottgens, B. (2019). New insights into hematopoietic differentiation landscapes from single-cell RNA sequencing. Blood *133*, 1415–1426. https://doi.org/10.1182/blood-2018-08-835355.

Wein, F., Pietsch, L., Saffrich, R., Wuchter, P., Walenda, T., Bork, S., Horn, P., Diehlmann, A., Eckstein, V., Ho, A.D., et al. (2010). N-Cadherin is expressed on human hematopoietic progenitor cells and mediates interaction with human mesenchymal stromal cells. Stem Cell Res. *4*, 129–139. https://doi.org/10.1016/j.scr.2009.12.004.

Wilkinson, A.C., Ishida, R., Kikuchi, M., Sudo, K., Morita, M., Crisostomo, R.V., Yamamoto, R., Loh, K.M., Nakamura, Y., Watanabe, M., et al. (2019). Long-term ex vivo haematopoietic-stem-cell expansion allows nonconditioned transplantation. Nature *571*, 117–121. https://doi.org/10.1038/s41586-019-1244-x.

Wilkinson, A.C., Igarashi, K.J., and Nakauchi, H. (2020). Haematopoietic stem cell self-renewal in vivo and ex vivo. Nat. Rev. Genet. *21*, 541–554. https://doi.org/10.1038/s41576-020-0241-0.

Will, B., Vogler, T.O., Bartholdy, B., Garrett-Bakelman, F., Mayer, J., Barreyro, L., Pandolfi, A., Todorova, T.I., Okoye-Okafor, U.C., Stanley, R.F., et al. (2013). Satb1 regulates the self-renewal of hematopoietic stem cells by promoting quiescence and repressing differentiation commitment. Nat. Immunol. *14*, 437–445. https://doi.org/10.1038/ni.2572.

Wilson, A., Laurenti, E., Oser, G., van der Wath, R.C., Blanco-Bose, W., Jaworski, M., Offner, S., Dunant, C.F., Eshkind, L., Bockamp, E., et al. (2008). Hematopoietic Stem Cells Reversibly Switch from Dormancy to Self-Renewal during Homeostasis and Repair. Cell *135*, 1118–1129. https://doi.org/10.1016/j.cell.2008.10.048.

Winkler, I.G., Barbier, V., Nowlan, B., Jacobsen, R.N., Forristal, C.E., Patton, J.T., Magnani, J.L., and Lévesque, J.-P. (2012). Vascular niche E-selectin regulates hematopoietic stem cell dormancy, self renewal and chemoresistance. Nat. Med. *18*, 1651–1657. https://doi.org/10.1038/nm.2969.

Wolf, R., Mirmohammadsadegh, A., Walz, M., Lysa, B., Tartler, U., Remus, R., Hengge, U., Michel, G., and Ruzicka, T. (2003). Molecular cloning and characterization of alternatively spliced mRNA isoforms from psoriatic skin encoding a novel member of the S100 family. FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol. *17*, 1969–1971. https://doi.org/10.1096/fj.03-0148fje.

Wolf, R., Howard, O.M.Z., Dong, H.-F., Voscopoulos, C., Boeshans, K., Winston, J., Divi, R., Gunsior, M., Goldsmith, P., Ahvazi, B., et al. (2008). Chemotactic activity of S100A7 (psoriasin) is mediated by RAGE and potentiates inflammation with highly homologous but functionally distinct S100A15. J. Immunol. Baltim. Md 1950 *181*, 1499–1506. .

Worton, R.G., McCulloch, E.A., and Till, J.E. (1969). Physical separation of hemopoietic stem cells differing in their capacity for self-renewal. J. Exp. Med. *130*, 91–103. https://doi.org/10.1084/jem.130.1.91.

Wright, D.E., Cheshier, S.H., Wagers, A.J., Randall, T.D., Christensen, J.L., and Weissman, I.L. (2001). Cyclophosphamide/granulocyte colony-stimulating factor causes selective mobilization of bone marrow hematopoietic stem cells into the blood after M phase of the cell cycle. Blood *97*, 2278–2285. https://doi.org/10.1182/blood.V97.8.2278.

Wu, L., Timmers, C., Maiti, B., Saavedra, H.I., Sang, L., Chong, G.T., Nuckolls, F., Giangrande, P., Wright, F.A., Field, S.J., et al. (2001). The E2F1-3 transcription factors are essential for cellular proliferation. Nature *414*, 457–462. https://doi.org/10.1038/35106593.

Wu, T.D., Reeder, J., Lawrence, M., Becker, G., and Brauer, M.J. (2016). GMAP and GSNAP for Genomic Sequence Alignment: Enhancements to Speed, Accuracy, and Functionality. Methods Mol. Biol. Clifton NJ *1418*, 283–334. https://doi.org/10.1007/978-1-4939-3578-9_15.

Xie, S.Z., Garcia-Prat, L., Voisin, V., Ferrari, R., Gan, O.I., Wagenblast, E., Kaufmann, K.B., Zeng, A.G.X., Takayanagi, S., Patel, I., et al. (2019). Sphingolipid Modulation Activates Proteostasis Programs to Govern Human Hematopoietic Stem Cell Self-Renewal. Cell Stem Cell *25*, 639-653.e7. https://doi.org/10.1016/j.stem.2019.09.008.

Xu, L., Liu, X., Peng, F., Zhang, W., Zheng, L., Ding, Y., Gu, T., Lv, K., Wang, J., Ortinau, L., et al. (2020). Protein quality control through endoplasmic reticulum-associated degradation maintains haematopoietic stem cell identity and niche interactions. Nat. Cell Biol. *22*, 1162–1169. https://doi.org/10.1038/s41556-020-00581-x.

Yahata, T., Takanashi, T., Muguruma, Y., Ibrahim, A.A., Matsuzawa, H., Uno, T., Sheng, Y., Onizuka, M., Ito, M., Kato, S., et al. (2011). Accumulation of oxidative DNA damage restricts the self-renewal capacity of human hematopoietic stem cells. Blood *118*, 2941–2950. https://doi.org/10.1182/blood-2011-01-330050.

Yamashita, M., and Passegué, E. (2019). TNF-α Coordinates Hematopoietic Stem Cell Survival and Myeloid Regeneration. Cell Stem Cell *25*, 357-372.e7. https://doi.org/10.1016/j.stem.2019.05.019.

Yáñez, A., Coetzee, S.G., Olsson, A., Muench, D.E., Berman, B.P., Hazelett, D.J., Salomonis, N., Grimes, H.L., and Goodridge, H.S. (2017). Granulocyte-monocyte progenitors and monocyte-dendritic cell progenitors independently produce functionally distinct monocytes. Immunity *47*, 890-902.e4. https://doi.org/10.1016/j.immuni.2017.10.021.

Yang, J., Tanaka, Y., Seay, M., Li, Z., Jin, J., Garmire, L.X., Zhu, X., Taylor, A., Li, W., Euskirchen, G., et al. (2017). Single cell transcriptomics reveals unanticipated features of early hematopoietic precursors. Nucleic Acids Res. *45*, 1281–1296. https://doi.org/10.1093/nar/gkw1214.

Yao, G., Lee, T.J., Mori, S., Nevins, J.R., and You, L. (2008). A bistable Rb–E2F switch underlies the restriction point. Nat. Cell Biol. *10*, 476–482. https://doi.org/10.1038/ncb1711.

Yoshikawa, Y., Ikebuchi, K., Ohkawara, J., Hirayama, F., Yamaguchi, M., Sato, N., Mori, K.J., Kasai, M., and Sekiguchi, S. (1999). A clonal culture assay for human cord blood lymphohematopoietic progenitors. Hum. Immunol. *60*, 75–82. https://doi.org/10.1016/s0198-8859(98)00094-9.

Yu, S.F., von Rüden, T., Kantoff, P.W., Garber, C., Seiberg, M., Rüther, U., Anderson, W.F., Wagner, E.F., and Gilboa, E. (1986). Self-inactivating retroviral vectors designed for transfer of whole genes into mammalian cells. Proc. Natl. Acad. Sci. U. S. A. *83*, 3194–3198. .

Yuan, Y., Shen, H., Franklin, D.S., Scadden, D.T., and Cheng, T. (2004). In vivo self-renewing divisions of haematopoietic stem cells are increased in the absence of the early G1-phase inhibitor, p18INK4C. Nat. Cell Biol. *6*, 436–442. https://doi.org/10.1038/ncb1126.

Zeng, H., Yücel, R., Kosan, C., Klein-Hitpass, L., and Möröy, T. (2004). Transcription factor Gfi1 regulates self-renewal and engraftment of hematopoietic stem cells. EMBO J. *23*, 4116–4125. https://doi.org/10.1038/sj.emboj.7600419.

Zhang, H.S., Gavin, M., Dahiya, A., Postigo, A.A., Ma, D., Luo, R.X., Harbour, J.W., and Dean, D.C. (2000). Exit from G1 and S phase of the cell cycle is regulated by repressor complexes containing HDAC-Rb-hSWI/SNF and Rb-hSWI/SNF. Cell *101*, 79–89. https://doi.org/10.1016/S0092-8674(00)80625-X.

Zhang, Y., Parmigiani, G., and Johnson, W.E. (2020). ComBat-seq: batch effect adjustment for RNA-seq count data. NAR Genomics Bioinforma. *2*, lqaa078. https://doi.org/10.1093/nargab/lqaa078.

Zhao, Y., Lin, Y., Zhan, Y., Yang, G., Louie, J., Harrison, D.E., and Anderson, W.F. (2000). Murine hematopoietic stem cell characterization and its regulation in BM transplantation. Blood *96*, 3016–3022. .

Zonari, E., Desantis, G., Petrillo, C., Boccalatte, F.E., Lidonnici, M.R., Kajaste-Rudnitski, A., Aiuti, A., Ferrari, G., Naldini, L., and Gentner, B. (2017). Efficient Ex Vivo Engineering and Expansion of Highly Purified Human Hematopoietic Stem and Progenitor Cell Populations for Gene Therapy. Stem Cell Rep. *8*, 977–990. https://doi.org/10.1016/j.stemcr.2017.02.010.