Aus dem Deutschen Krebsforschungszentrum Heidelberg

Geschäftsführende Direktoren: Prof. Dr. med. Dr. h. c. Michael Baumann und Ursula Weyrich

Abteilung B-Zell-Immunologie

Abteilungsleiterin: Prof. Dr. Hedda Wardemann

Stem cell heterogeneity and clonal dynamics explored by *PolyloxExpress* RNA barcoding

Inauguraldissertation
zur Erlangung des Doctor scientiarum humanarum
(Dr. sc. hum.)
an der
Medizinischen Fakultät Heidelberg
der
Ruprecht-Karls-Universität

Vorgelegt von
Fuwei Shang
aus
Kunming, China
2023

Dekan: Herr Prof. Dr. med. Dr. h. c. Hans-Georg Kräusslich

Doktormutter: Frau Prof. Dr. Hedda Wardemann

Table of Contents

Abbreviations

4-OHT	4-hydroxy-tamoxifen
AGM	Aorta-gonad-mesonephros
AUC	Area under curve
bp	Base pair
CD	Cluster of differentiation
CCS	Circular consensus sequencing
cDNA	Complementary DNA
Cas9	CRISPR associated protein 9
CITEseq	Cellular indexing of transcriptomes and epitopes by sequencing
CLP	Common lymphoid progenitor
CMP	Common myeloid progenitor
CRISPR	Clustered regularly interspaced palindromic repeats
DC	Diffusion component
DEG	Differentially expressed gene
DMEM	Dulbecco's modified eagle medium
DMSO	Dimethyl sulfoxide
DN	Double negative pre-T cell
dNTP	Deoxyribonucleotide triphosphate
DP	Double positive pre-T cell
DPBS	Dulbecco's phosphate-buffered saline
E9.5	Embryonic day 9.5
eMPP	Embryonic multipotent progenitor
ECCITEseq	Expanded CRISPR compatible cellular indexing of transcriptomes and epitopes by sequencing
EDTA	Ethylenediaminetetraacetic acid
ESC	Embryonic stem cell
EtOH	Ethanol
EryP	Erythroid progenitor
FACS	Fluorescence-activated cell sorting

FBS	Fetal bovine serum
FL	Fetal liver
FDR	False discovery rate
gDNA	Genomic DNA
GPRC5C	G protein-coupled receptor class C group 5 membrane C
GMP	Granulocyte-monocyte progenitor
Gr	Granulocyte
HE	Hemogenic endothelium
HSC	Hematopoietic stem cell
HSPC	Hematopoietic stem and progenitor cell
IgG	Immunoglobulin G
Lin	Lineage
LIF	Leukemia inhibitory factor
LMPP	Lymphoid-primed multipotent progenitor
LoxP	Locus of X-over P1
LSK	Lin^- $Sca\text{-}1^+$ Kit^+
LT-HSC	Long-term hematopoietic stem cell
MEF	Mouse embryonic fibroblast
MEP	Megakaryocyte-erythrocyte progenitor
Mer	Murine estrogen receptor
MiCM	Mer-iCre-Mer
Mk	Megakaryocyte
MkP	Megakaryocyte progenitor
Mono	Monocyte
MPP	Multipotent progenitor
mRNA	Messenger RNA
NHEJ	Non-homologous mediated end joining
PB	Periphery blood
PCA	Principal component analysis
PCR	Polymerase chain reaction
Pgen	Probability of generation

RBC	Red blood cell
RT-PCR	Reverse transcription polymerase chain reaction
SEM	Standard error of mean
sgRNA	Single guide RNA
Spl	Spleen
SLAM	Signaling lymphocyte activation molecule
ST-HSC	Short-term hematopoietic stem cell
SMRT	Single-molecule real-time sequencing
scRNAseq	Single-cell RNA sequencing
scNMT	Single-cell Nucleosome, methylation and transcription sequencing
t-SNE	t-Distributed stochastic neighbor embedding
UTR	Untranslated region
vWF	von Willebrand Factor

1 Introduction

1.1 Hematopoiesis and characteristics of hematopoietic stem cells

Hematopoiesis is a continuous process of erythroid, myeloid and lymphoid cell formation that begins at the embryonic stage of development and continues throughout life. The activity of hematopoietic stem cells (HSCs) was first documented in the 1960s when McCulloch and colleagues observed spleen colony forming units after transplanting normal bone marrow cells into lethally irradiated mice (Becker et al. 1963; Till and Mc 1961). With the identification of lineage-restricted progenitors, transplantation-based studies further demonstrated that hematopoiesis is organized in a hierarchical structure and the generation of blood and immune cells follows a step-wise process (Akashi et al. 2000; Kondo et al. 1997) (Figure 1). In this hierarchy, HSCs reside at the top and are characterized by their self-renewing capacity and multipotency (Laurenti and Göttgens 2018) (Figure 1).

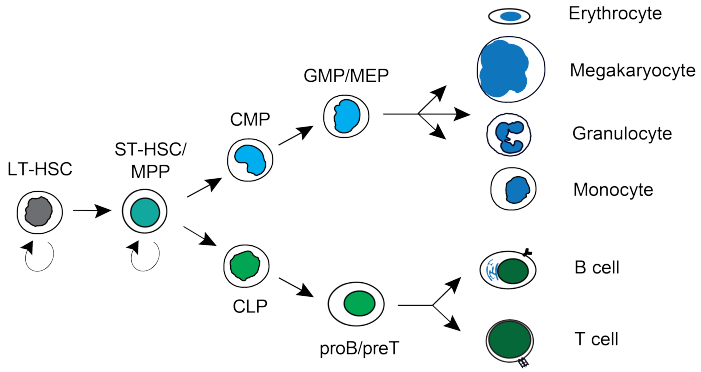


Figure 1. The classical bifurcating tree model of hematopoiesis. In this hierarchical model, the blood system is built and maintained from multipotent, self-renewing long-term hematopoietic stem cells (LT-HSC) and short-term hematopoietic stem cells (ST-HSC) that constantly differentiate via intermediate progenitor stages into mature blood cell types. Along this hierarchy, the potential to produce different cell lineages is gradually lost. Multipotent progenitors (MPP) can still either differentiate to common myeloid progenitors (CMPs) or common lymphoid progenitors (CLPs). CMPs produce myeloid cells through a series of further differentiation steps via granulocyte-monocyte progenitors (GMPs) or megakaryocyte-erythrocyte progenitors (MEPs). CLPs either commit into proB or preT cells that finally generate mature lymphoid B or T cells.

1.1.1 Purification of hematopoietic stem cells

Since the identification of HSCs, the quest for purifying HSCs to their functional homogeneity has been ongoing. Surface marker identification played an important role in early HSC studies. The combination of Thy-1^{lo} Sca-1^{+} Lineage (Lin)$^{-}$ Kit^{+}, originally reported over three decades ago, was the first to enrich HSCs in mice (Spangrude et al. 1988). Subsequently, dye excluding cells in Hoechst staining or Sca-1^{+} Lin^{-} Kit^{+} CD34$^{lo/-}$ bone marrow cells were also shown to enrich for HSCs (Goodell et al. 1996; Osawa et al. 1996). More recently, the combination of Lin^{-} Sca-1^{+} Kit^{+} (LSK) with signaling lymphocyte activation molecule (SLAM) markers CD150^{+} CD48^{-} was established and has now been commonly accepted by most laboratories for the purification of HSCs (Kiel et al. 2005). However, only around half of the cells purified by SLAM markers possess long-term multilineage reconstitution capacity, suggesting that functional heterogeneity remains in this surface-marker-defined compartment (Kiel et al. 2005).

With the development of high-throughput sequencing in the recent decade, further HSC specific genes were characterized and genetic reporter mouse lines were generated (Busch et al. 2015; Chen et al. 2016; Christodoulou et al. 2020; Gazit et al. 2014; Sawai et al. 2016). Typically, in these studies, a subset of HSCs was enriched from the SLAM marker-defined compartment. When combined with fluorescent reporter or Cre recombinase, these mouse models allow purification of functional HSCs from total bone marrow cells (e.g., Fgd5 reported by Gazit et al. 2014), quantification of HSC output in native hematopoiesis (e.g., Tie2 reported by Busch et al. 2015, Pdzk1ip1 reported by Sawai et al. 2016) or determining the spatial localization of HSCs (e.g., Hoxb5 by Chen et al. 2016, Mds1 reported by Christodoulou et al. 2020).

1.1.2 Hematopoietic stem cell heterogeneity

Despite advances in the phenotypic characterization of HSCs, studies investigating lineage outputs based on single-HSC transplantation revealed heterogeneous HSC behavior over the past two decades (Dykstra et al. 2007; Muller-Sieburg et al. 2004; Müller-Sieburg et al. 2002; Sanjuan-Pla et al. 2013; Yamamoto et al. 2013). It has been shown that, apart from types with multilineage contribution, distinct subsets of HSCs exist that could maintain a lineage-biased fate even upon serial transplantation (Dykstra et al. 2007). These different HSC fates were clustered into four categories based on their lineage contribution: α, β, γ and δ cells (Dykstra et al. 2007). More recently, a bypass pathway linking HSCs directly to the megakaryocyte (Mk) lineage was reported (Carrelha et al. 2018; Sanjuan-Pla et al. 2013). Together, transplantation studies demonstrated that HSC fate is not homogeneous and, importantly, fate commitment may already happen at the HSC stage (Figure 2).

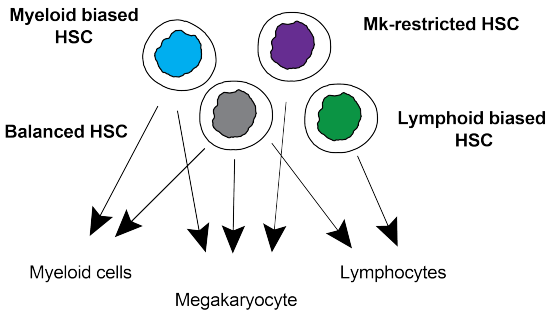


Figure 2. Schematic diagram of HSC fate heterogeneity. Hematopoietic stem cells (HSCs) with different fate potentials are highlighted in color and arrows indicate cell types they generate. Mk, megakaryocyte.

HSCs may not only differ in fate, but also in surface marker expression. It has been suggested that multiple surface markers (e.g., CD201 and CD41) are selectively expressed in a fraction of $CD150^{+}$ $CD48^{-}$ LSK HSCs (Balazs et al. 2006; Kent et al. 2009; Kiel et al. 2005). However, few studies were able to correlate marker expression with distinct HSC functions. To this end, Nakauchi and colleagues performed surface marker screening on HSCs (defined by Sca-1^{+} Lin^{-} Kit^{+} $CD34^{lo/-}$), followed by transplantation of marker-defined HSC subsets into lethally irradiated mice to realize their fates (Morita et

al. 2010). However, none of the markers examined were able to enrich for HSCs with specific lineage output, except that the intensity of CD150 expression correlated with the erythroid/megakaryocyte differentiation potential (Morita et al. 2010). More recently, the group of Sean Morrison showed that CD229 and CD244 SLAM family markers could, to certain extent, separate HSCs and MPPs with distinct functions (Oguro et al. 2013). Of note, in most studies, lineage output was the main assay to characterize the function of marker-defined HSC subsets. Whether such markers can distinguish other HSC functional aspects (e.g., self-renewal capacity) would require further investigation.

Besides their ability to generate blood and immune cells, one of the key features of HSCs is their capacity to self-renew and maintain the stem cell compartment. Using a label retention assay, the group of Andreas Trumpp reported a subset of rarely proliferative HSCs within the $CD150^{+}$ $CD48^{-}$ $CD34^{-}$ LSK compartment, which they termed “dormant” HSCs (Wilson et al. 2008). Subsequent studies showed that this subset can be enriched using the G protein-coupled receptor class C group 5 membrane C (GPRC5C) and that these cells require retinoic acid signaling for their maintenance of homeostasis (Cabezas-Wallscheid et al. 2017; Zhang et al. 2022). It is not yet clear, however, how HSC self-renewal potential correlates with HSC differentiation.

Collectively, heterogeneity of HSCs is not only reflected in lineage output and surface marker expression, but also in their ability to self-renew. Notably, the kinetics of HSC contribution to the downstream compartment has been shown to differ between native hematopoiesis and transplantation (Busch et al. 2015; Busch and Rodewald 2016; Höfer et al. 2016; Sun et al. 2014). However, it remains elusive, to what extent the heterogeneity of HSCs observed in transplantation experiments resembles the properties of steady-state hematopoiesis. Therefore, new tools had to be established to minimize artificial perturbation and to study the functions of HSCs in their native environment.

1.2 Lineage tracing of hematopoietic stem cells

Understanding the complex development of a multicellular organism is an interesting but challenging topic in biology. During the past decades, different approaches have been applied to label cells with inheritable markers (Baron and van Oudenaarden 2019; Kebschull and Zador 2018; Woodworth et al. 2017). By recording the propagation of these markers, scientists were able to dissect the lineage relationships and differentiation pathways of cells. Typically, these methods are known as “fate mapping” or “lineage tracing”.

Since the 1980s, viral barcoding approaches have been introduced to study HSCs (Dick et al. 1985; Lemischka et al. 1986). This typically involves the transduction of mixtures of short DNA sequences (so-called “barcodes”) by retroviral or lentiviral gene transfer approaches, with the aim of integrating a unique barcode into the genome of each HSC. Barcoded HSCs are subsequently transplanted into recipients where the distribution of the barcodes in the progeny cells is recorded (Lu et al. 2011) (Figure 3A). Results obtained from these studies revealed that HSCs do not equally contribute to hematopoiesis, but that high variability exists within the HSC compartment (Lu et al. 2011; Verovskaya et al. 2013). Moreover, compared to single-HSC transplantation, viral barcoding experiments profiled HSC fate at a larger scale (more than hundreds of cells) and confirmed the existence of HSC fate heterogeneity upon transplantation (Lu et al. 2011).

To understand the nature of steady state hematopoiesis, several groups devised genetic reporter mouse lines to label HSCs specifically with fluorescent proteins and characterize their functions in situ (Busch et al. 2015; Christodoulou et al. 2020; Morcos et al. 2022; Sawai et al. 2016; Upadhaya et al. 2018; Yu et al. 2016) (Figure 3B). Of interest, using an inducible Cre reporter system driven by Tie2 to label HSCs (Busch et al. 2015), the Rodewald lab followed steady state HSC output kinetically (Busch et al. 2015). In contrast to transplantation, where HSCs were stimulated to differentiate at a substantial rate, in native hematopoiesis the contribution of HSCs to downstream progeny was much slower

than expected (Busch et al. 2015). Although continuously ongoing, the equilibration of fluorescent reporter labeling in downstream progeny to the initial rate of HSCs would exceed the life-span of the mouse, highlighting a key difference between native hematopoiesis and transplantation (Busch et al. 2015). In a separate study, Yu and colleagues generated a more complex multicolor mouse model and revealed that HSCs with the same color code shared similar differentiation pattern in both native and stress conditions (Yu et al. 2016). Moreover, by mapping the chromatin accessibility of different HSC clones, the study showed that stereotypic HSC behaviors could be explained by epigenetic imprinting (Yu et al. 2016).

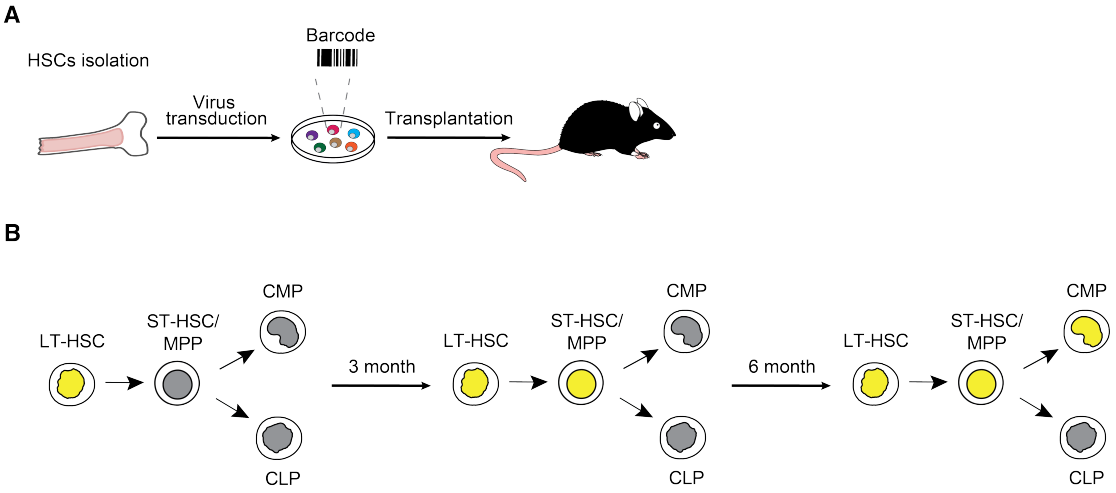


Figure 3. Lineage tracing by viral barcoding or fluorescent proteins.

(A) Hematopoietic stem cells (HSCs) are purified from mouse bone marrow cells and cultured *in vitro*. Unique DNA sequences serving as barcodes are inserted into the genome of each HSC by viral transduction. Barcoded HSCs are then transplanted back into lethally irradiated recipients to study their fate.

(B) Using HSC specific reporter mice, long term HSCs (LT-HSCs) are labelled by fluorescent protein expression *in vivo*. By recording the propagation of the reporter into downstream compartments at different time points, the output kinetic of steady-state hematopoiesis can be revealed. Short term HSC (ST-HSC); MPP, multipotent progenitor; CMP, common myeloid progenitor; CLP, common lymphoid progenitor.

In sum, although viral barcoding has the advantage of studying hundreds to thousands of unique barcoded HSCs in the same recipient, it is obvious that random DNA insertion and transplantation may inevitably introduce artificial perturbation to HSCs, and hence alter HSC functions. Taking the advantage of labeling HSCs at steady state in situ, fluorescent reporter-based fate mapping circumvents the need for transplantation and revealed fundamental differences between native hematopoiesis and transplantation. However, given that heterogeneity of HSC functions is observed even at single-cell level, these methods are still limited in their capacity to resolve steady state HSC functions at clonal resolution.

To circumvent the need for transplantation and to determine native hematopoiesis with high resolution at the clonal level, DNA barcoding systems have been developed (Kalhor et al. 2018; McKenna et al. 2016; Pei et al. 2017; Sun et al. 2014). Taking the advantage of an inducible transposase system, Camargo's lab devised a mouse model capable of tagging individual cells through random DNA insertion (Sun et al. 2014). By comparing tags shared between HSCs and progeny 40 weeks after induction, this method led to the finding that native hematopoiesis is maintained mostly by hematopoietic progenitor cells rather than HSCs (Sun et al. 2014). In a subsequent study that included megakaryocyte progenitors in the sampling, they found that most HSCs had a fate restricted to megakaryocytes (Rodriguez-Fraticelli et al. 2018). Nevertheless, given the risk of under sampling and the lack of specificity for transposase induction, conclusions drawn from this system require careful interpretation. In contrast, another high-resolution DNA barcoding system developed by the Rodewald lab circumvents the shortcomings of transposon barcoding. The system, named "*Polylox*", deploys an artificial DNA cassette consisting of ten Locus of X-over P1 (LoxP) sites and nine interspersing unique DNA sequences (non-inverted segments are named 1-9 and inverted segments are named A-I) (Pei et al. 2017) (Figure 4). Upon Cre mediated recombination, random rearrangement of the *Polylox* cassette segments could generate up to 1.8 million different sequences called "barcodes" (Figure 4). Of note, in the *Polylox* system, not all barcodes are generated at equal frequencies. By calculating the probability of generation (*Pgen*), rare barcodes that are most likely generated only once in the population of interest are

selected to define clonal cell fate (Pei et al. 2017; Pei et al. 2019). Application of this system to study the hematopoietic stem cell compartment led to the identification of multilineage, myeloid-erythroid-restricted and differentiation-inactive HSC clones (Pei et al. 2017). Therefore, fate heterogeneity is unlikely an artifact of transplantation, but rather an intrinsic property of HSCs.

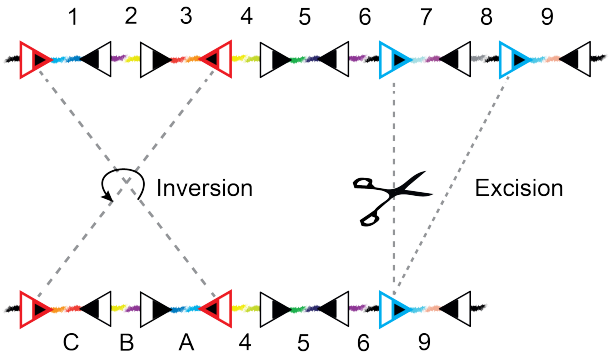


Figure 4. Schematic diagram of the *Polylox* cassette. The first line represents the unrecombined *Polylox* substrate, the second line shows a possible barcode that originated by one inversion and one excision. Colored segments depict unique DNA sequences that are interspersed by loxP sites. Random barcodes are generated by Cre-mediated excision (red line) or inversion (blue line). Segments 1-9 are named A-I when inverted. (Figure adopted and modified from Pei et al. 2020)

1.3 Mapping hematopoiesis with single-cell RNA sequencing

Advances in high-throughput transcriptome profiling over the past decade have opened the possibility to investigate heterogeneity within cell types and to determine gene expression signatures within cell subsets (Gawad et al. 2016; Huang et al. 2015; Tanay and Regev 2017). Mostly, this is achieved by profiling RNA expression at single-cell resolution using plate-based or microfluidic platforms (Klein et al. 2015; Picelli et al. 2013; Picelli et al. 2014; Zheng et al. 2017; Ziegenhain et al. 2017; Zilionis et al. 2017). In particular, recent hematopoietic studies using single-cell RNA sequencing (scRNAseq) suggest that phenotypically defined hematopoietic stem and progenitor cell (HSPC) populations are transcriptionally heterogeneous (Giladi et al. 2018; Jacobsen and Nerlov 2019; Laurenti and Göttgens 2018; Nestorowa et al. 2016; Olsson et al. 2016; Paul et al. 2015; Velten et al. 2017; Watcham et al. 2019; Wilson et al. 2015). In contrast to the classical bifurcating tree model, in which cells are assumed to make binary fate choices at different stages, scRNAseq experiments have demonstrated that multiple lineage priming states coexist within the same phenotypically defined HSPC compartment (Olsson et al. 2016; Paul et al. 2015). Hence, the complex hematopoietic process may be oversimplified by a few surface markers used for cell purification and functional testing (Liggett and Sankaran 2020; Triana et al. 2021). In support of this, the functional heterogeneity of the MPP compartment could be further resolved if more markers were included to identify subsets (Pietras et al. 2015; Sommerkamp et al. 2021; Terszowski et al. 2005).

In recent years, several bioinformatic algorithms have been established to predict differentiation pathways based on single-cell transcriptomes (Haghverdi et al. 2016; Kester and van Oudenaarden 2018; Trapnell et al. 2014; Wang et al. 2022). Considering cells with similar gene expression profiles being more closely related during their development than others, the so called “pseudo-time trajectories” are calculated (Haghverdi et al. 2016). By organizing HSPCs with different molecular states along the pseudo-time trajectory, these studies proposed that HSC differentiation is a continuous

process in which lineage-restricted cells gradually arise from less committed progenitors (Buenrostro et al. 2018; Velten et al. 2017).

Although scRNAseq is a comprehensive and powerful approach for identifying cell states and calculating the point of fate divergence, it is not suitable for capturing the dynamic differentiation process of the same cell. Instead, only a snapshot from fluctuating cell states is taken to infer the ongoing developmental process. Efforts have been made to capture transcriptome information from live cells, yet the method is still far from analyzing primary cells in situ (Chen et al. 2022). Moreover, a differentiation trajectory of HSCs is calculated at the population scale, which does not necessarily reflect the developmental pathway of individual HSCs or HSC clones (Wagner and Klein 2020). For instance, differentiation-inactive HSCs, defined by their lack of hematopoietic cell productivity, are unlikely to follow the same developmental “trajectory” as multilineage HSCs, although both fate-defined HSCs coexist within the less committed progenitor state (Pei et al. 2017; Velten et al. 2017). Therefore, it is mandatory that developmental paths are functionally demonstrated, which may or may not agree with paths predicted from pseudo-time trajectories

Taken together, single-cell snapshot data lack precursor-product information and therefore cannot inform on fate. On the other hand, DNA barcoding fails to provide transcriptional information and hence fate patterns of HSC clones cannot be linked to gene expression, precluding insights into molecular mechanisms. Understanding cell fate determination requires paired lineage tracing and transcriptional profiling for single cells, which remains a major challenge.

1.4 Analyzing cell fate and fate determinants by RNA barcoding

To link gene expression profiles directly to cell fate, several RNA barcoding techniques have been developed. The principle of these techniques is to label cells and their progeny with barcodes that are expressed as RNA and can be capture together with the whole transcriptome using scRNAseq (Kester and van Oudenaarden 2018; Liggett and Sankaran 2020; Shang and Rodewald 2022; Wagner and Klein 2020). One of the most common strategies to achieve RNA barcoding is to introduce an artificial DNA cassette into the cell genome using transfection. These cassettes typically contain a promoter sequence at the 5' end, followed by a reporter gene (e.g., fluorescent protein) and a random nucleotide sequence, which serves as unique tag for cellular barcoding (Figure 5). Applying this strategy to *in vitro* cultured cells led to the identification of important regulators associated with cell state maintenance and transition (Biddy et al. 2018; Fennell et al. 2022; Frieda et al. 2017; Rodriguez-Fraticelli et al. 2020; Weinreb et al. 2020). However, the artificial perturbation inherent in such a strategy limits the extrapolation of results to more physiological relevant conditions.

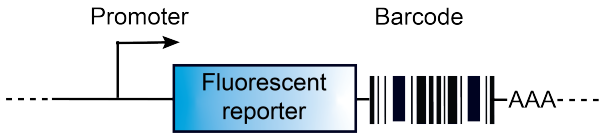


Figure 5. Schematic diagram for the design of RNA barcoding. Most RNA barcoding is achieved by inserting a barcode sequence into the 3' untranslated region (UTR) of a fluorescent reporter. Driven by a promoter at the 5' end, the reporter and the barcode are expressed as messenger RNA (mRNA). AAA represents polyA tail. (Figure adopted and modified from Shang and Rodewald 2022)

The initial in situ RNA barcoding experiments were performed with clustered regularly interspaced palindromic repeats (CRISPR) gene editing in zebrafish (Alemany et al. 2018; Raj et al. 2018; Spanjaard et al. 2018). By injecting CRISPR associated protein 9 (Cas 9) and single guide RNA (sgRNA) into the embryo at the one-cell stage, templates within the genome (fluorescent reporter sequence or artificial DNA cassette) were targeted to form double strand breaks. These breaks are subsequently repaired through the

mechanism of non-homologous mediated end joining (NHEJ), resulting in insertions and deletions that are inherited to progeny cells as genomic “scars” (Alemany et al. 2018; Raj et al. 2018; Spanjaard et al. 2018). Such scars, when expressed as RNA, could be used as unique barcodes for tracking cell fate and identifying fate-related transcriptional signatures in early embryonic development (Alemany et al. 2018). A rather more complex system modified the current CRISPR strategy by manipulating mutation rates through sgRNA mismatches (Chan et al. 2019). Owing to the differences in the degree of sgRNA mismatch, template mutation rates differ (Chan et al. 2019). In this way, barcodes will be continuously generated over a relatively long period of time as cells proliferate and differentiate. Therefore, this approach provides a unique opportunity to record embryonic development on a time scale.

Nevertheless, these reported methods always require early-stage embryonic injection and are unable to induce barcode generation specifically in the cell of interest. Ideally, a powerful RNA barcoding system applicable for more general biological fields should allow barcoding at a given time (time specific) within a given cell (tissue specific) *in vivo*.

1.5 Generation of the *PolyloxExpress* RNA barcoding system

To devise a model that allows joint read out of gene expression and cell fate *in vivo*, Thorsten Feyerabend and Weike Pei in the Rodewald lab designed the next generation of the *Polylox* lineage tracing system "*PolyloxExpress*". In this system, barcodes are generated at the DNA level as described for *Polylox* in section 1.2 (Figure 4), but due to the design of the *PolyloxExpress* locus, barcodes are also transcribed into RNA, driven by the ubiquitous *Rosa26* promoter (Figure 6). These RNA molecules contain the coding sequence of the red fluorescent protein "tdTomato" and in the 3' UTR the *PolyloxExpress* barcodes (Figure 6). In combination with scRNAseq, lineage information and gene expression profile can be retrieved at the single-cell level. Moreover, tissue specificity can be achieved by selecting a suitable conditional Cre-recombinase.

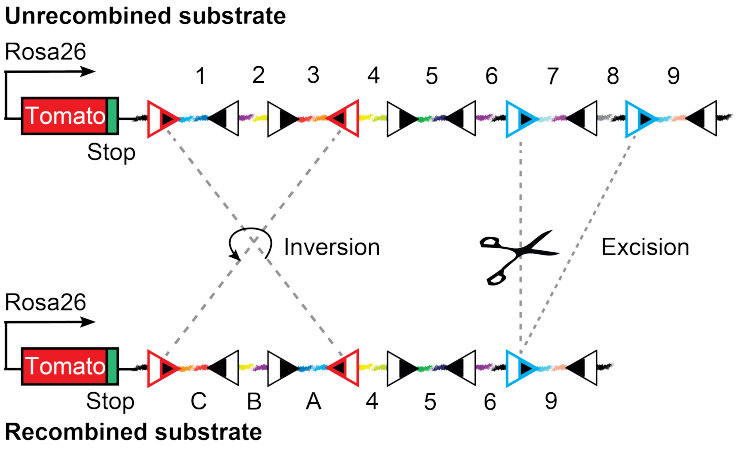


Figure 6. Design of the *PolyloxExpress* cassette. The working principle for the *PolyloxExpress* barcode sequence is exactly the same as for *Polylox* (described in Figure 4), except that the unrecombined barcode substrate is inserted into the 3' UTR of a tdTomato reporter. A stop codon (stop) is placed immediately after the tdTomato sequence. The *PolyloxExpress* cassette was inserted into the *Rosa26* locus of mouse embryonic stem cells and the expression of the barcode is therefore driven by the intrinsic *Rosa26* promoter. (Figure adopted and modified from Pei et al. 2020)

To generate a mouse model for *in vivo* analysis, the *PolyloxExpress* sequence was first inserted into the *Rosa26* locus of murine embryonic stem cells (ESCs) by gene targeting. *Rosa26*$^{PolyloxExpress/+}$ knock in mice were further generated through blastocyst injection of

targeted ESCs and germ line transmission. Upon crossing with an inducible, tissue-specific Cre mouse line (e.g., *Tie2*$^{MiCM/+}$), *PolyloxExpress* barcodes can be induced by the application of Tamoxifen.

1.6 Aim of the project

(1) Establishing the *PolyloxExpress* System:

- Identification of recombined *PolyloxExpress* barcodes from mRNA

- Proof of consistency between DNA and RNA barcodes in single cell

- Develop an experimental workflow for combined scRNAseq and *PolyloxExpress* barcode sequencing

- Recover full-length *PolyloxExpress* barcodes and transcriptome from the same cell

(2) Application of the *PolyloxExpress* System to answer the following questions:

- At which developmental stages does HSC fate specification occur?

- How stable are HSC fates over time and is the intrinsic fate stability maintained in transplantation?

- Do fate-defined HSCs possess specific gene expression signatures?

2 Materials and Methodology

2.1 Materials

2.1.1 Plastics and consumables

Table 1. List of plastics and consumables

Name	Source	Catalog Number
10 µl Filtered Tips	Starlab	S1120-3810
20 µl Filtered Tips	Starlab	S1120-1810
200 µl Filtered Tips	Starlab	S1120-8810
1000 µl Filtered Tips	Starlab	S1126-7810
DNA LoBind Tubes (1.5 ml)	Eppendorf	0030108051
Eppendorf Safe-Lock Tubes (1.5 ml)	Eppendorf	0030120086
Eppendorf Safe-Lock Tubes (2 ml)	Eppendorf	0030121880
Serological pipette tubes (5 ml)	Greiner bio-one	606180
Serological pipette tubes (10 ml)	Greiner bio-one	607180
Serological pipette tubes (25 ml)	Greiner bio-one	760180
Falcon 15ml centrifuge tubes	Corning	352095
Falcon 50ml centrifuge tubes	Corning	352070
5 ml Round Bottom Polystyrene Test Tubes (with Cell Strainer Snap Cap)	Corning	352235
5 ml Round Bottom Polystyrene Tubes	Corning	352003
1.4 ml Round bottom tubes	Micronic	MP32022
PCR tubes and caps (0.2 ml)	Thermo Fisher Scientific	AB0266
PCR strips with individual caps (0.2 ml)	Starlab	A1402-3700
Sigma cryogenic vial	Sigma-Aldrich	SIAL0659

Micro sample container K3 EDTA (1.3 ml)	SARSTEDT	41.1504.005
Qubit Assay Tubes	Thermo Fisher Scientific	Q32856
1 ml Syringe	B.Braun	9166017V
BD Discardit II Syringe 10 ml	BD Biosciences	300928
Syringe (20 ml)	B.Braun	4606205V
Sterican insulin cannula 26G	B.Braun	4665457
Syringe filter (0.2 µm)	SARSTEDT	83.1826.001
Bottle Top Vacuum Filter (0.22 µm, 150 ml)	Corning	430624
Cell culture dish (100 mm)	Corning	353003
Cell culture multiwell plate (6 well)	Greiner bio-one	657160
Cell culture multiwell plate (6 well)	Thermo Fisher Scientific	150239
Cell culture multiwell plate (24 well)	Thermo Fisher Scientific	142475
Cell culture multiwell plate (96 well)	Thermo Fisher Scientific	161093
Pasteur pipettes	Brand	747715
Costar reagent reservoirs	Corning	4870
40 µm Cell Strainer	Corning	352340
100 µm Cell Strainer	Corning	352360

2.1.2 Chemicals and cell culture medium

Table 2. List of chemicals and cell culture medium

Name	Source	Catalog Number
4-Hydroxytamoxifen	Sigma-Aldrich	H7904-25MG
Tamoxifen free base	Sigma-Aldrich	T5648-1G
Peanut oil	Sigma-Aldrich	P2144-250ML
Progesterone	Sigma-Aldrich	P0130
Ethanol absolute	Fisher Scientific	E/0650DF/08

Agarose	Bio-Rad	161-3104EDU
Ethidium Bromide Solution 0.07% Dropper Bottle	AppliChem	A22730015
Nuclease-Free Water	Thermo Fisher Scientific	AM9937
Qiagen Buffer EB	Qiagen	19086
Fetal bovine serum (FBS)	Sigma-Aldrich	F7524-500ML
Gelatin solution (2%)	Sigma-Aldrich	G1393-100ML
Dulbecco's Phosphate Buffered Saline (DPBS)	Sigma-Aldrich	D8537-500ML
Dimethyl sulfoxide (DMSO)	Sigma-Aldrich	D2650
Recombinant Mouse LIF Protein	Sigma-Aldrich	ESG1107
DMEM (no glutamine)	Gibco	11960044
Sodium Pyruvate (100 mM)	Gibco	11360039
MEM Non-Essential Amino Acids Solution (100X)	Gibco	11140035
2-Mercaptoethanol (50 mM)	Gibco	31350010
Penicillin-Streptomycin (10,000 U/ml)	Gibco	15140122
GlutaMAX Supplement	Gibco	35050038
Trypsin-EDTA (0.25%)	Gibco	25200056
KnockOut DMEM	Gibco	10829018
Embryonic stem-cell FBS	Gibco	16141079

2.1.3 Oligonucleotides

Table 3. List of oligonucleotides

Name	Sequence (5' to 3')
972	GAGGCAGCATCTGTCTACAAGAGATGG
984	CATCGCATACCATACATAGGTGGAGG

HL14	AATCAAGGGTCCCCAAACTCAC
1116	AAGGGAGCTGCAGTGGAGTA
1117	TAAGCCTGCCCAGAAGACTCC
1118	AAGACCGCGAAGAGTTTGTCC
2426	CGACGACACTGCCAAAGATTTC
2427	CATACCTTAGAGAAAGCCTGTCGAG
2652	GCATGGACGAGCTGTACAAG
2653	CGAACTTCTCTCTGTTAAAGCAAG
2674	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTC
2676	AATGATACGGCGACCACCGA
2702	GAAAGCCTGTCGAGATATGTTTCTC
2999	CGTTCTCATTAACGCCGACGACAC
ISPCR-2999	AAGCAGTGGTATCAACGCAGAGTCGTTCTCATTAACGCCGACGACAC
2426-A	TCTCCATCCGACGACACTGCCAAAGATTTC
2426-B	TCTGGTTGCGACGACACTGCCAAAGATTTC
2426-C	TCCGACAACGACGACACTGCCAAAGATTTC
2426-D	TCACGAACCGACGACACTGCCAAAGATTTC
2426-E	TACGGTACCGACGACACTGCCAAAGATTTC
2426-F	TAGCCAAGCGACGACACTGCCAAAGATTTC
2426-G	CTCTTCGACGACGACACTGCCAAAGATTTC
2426-H	CCTTAGACCGACGACACTGCCAAAGATTTC
2426-I	AAGGTCCTCGACGACACTGCCAAAGATTTC
2426-J	TTCGAGGTCGACGACACTGCCAAAGATTTC
2426-K	TTCACCGACGACGACACTGCCAAAGATTTC
2426-L	TTGTGAGGCGACGACACTGCCAAAGATTTC
2426-M	AGGTTGCACGACGACACTGCCAAAGATTTC
2426-N	TTGCTGTCCGACGACACTGCCAAAGATTTC
2426-O	CATCATGGCGACGACACTGCCAAAGATTTC

2426-P	GCATCCTTCGACGACACTGCCAAAGATTTC
2426-Q	GGTCTTGTCGACGACACTGCCAAAGATTTC
2426-R	AGCTTCACCGACGACACTGCCAAAGATTTC
2426-S	AACGCACTCGACGACACTGCCAAAGATTTC
2426-T	AAGTAGCCCGACGACACTGCCAAAGATTTC
2426-U	TTGCTCCACGACGACACTGCCAAAGATTTC
2426-V	GGAAGTGACGACGACACTGCCAAAGATTTC
2426-W	GACAAGTGCGACGACACTGCCAAAGATTTC
2426-X	AATAGCGGCGACGACACTGCCAAAGATTTC
2426-OV	CGTGAGTACGACGACACTGCCAAAGATTTC
2426-ER	GCTAGATGCGACGACACTGCCAAAGATTTC
2426-IN	TACCAGCACGACGACACTGCCAAAGATTTC
2426-TE	GGATCTCACGACGACACTGCCAAAGATTTC
2426-LL	CACTGTCCCGACGACACTGCCAAAGATTTC
2426-EC	CTGAAATCCGACGACACTGCCAAAGATTTC
2426-TU	ATTGCCTCCGACGACACTGCCAAAGATTTC
2426-AL	GTTCACTGCGACGACACTGCCAAAGATTTC
2426-IZ	TCAGAGTCCGACGACACTGCCAAAGATTTC
2426-AT	ACACAATCCGACGACACTGCCAAAGATTTC
2426-IO	ATCCGTATCGACGACACTGCCAAAGATTTC
2426-NS	TGGCAGCTCGACGACACTGCCAAAGATTTC

2.1.4 Antibodies

Table 4. List of antibodies

Name	Clone	Dilution	Source	Catalog Number
BP1-FITC	6C3	400	Life Technologies	11-5891-82
CD3ε-BV605	145-2C11	100	BD Pharmingen	563004

CD4-APC	RM4-5	400	Life Technologies	17-0042-83
CD4-BV421	GK1.5	800	Biolegend	100443
CD8a-BV421	53-6.7	800	Biolegend	100753
CD8a-FITC	53-6.7	200	BD Pharmingen	553030
CD11b-PE-Cy7	M1/70	400	Life Technologies	25-0112
CD11b-BV421	M1/70	800	Biolegend	101251
CD16/32-BV605	2.4G2	100	BD Pharmingen	563006
CD19-FITC	1D3	400	BD Pharmingen	557398
CD19-BV421	6D5	200	Biolegend	115549
CD24-APC-eFluor780	M1/69	50	Life Technologies	47-0242
CD25-PE-Cy7	PC61	1600	BD Pharmingen	552880
CD34-FITC	RAM34	25	BD Pharmingen	553733
CD41-BV711	MWReg30	200	BD Pharmingen	740712
CD42d-APC	1C2	200	Biolegend	148505
CD43-PE-Cy7	S7	200	BD Pharmingen	562866
CD44-Alexa-Fluor700	IM7	100	BD Pharmingen	560567
CD44-APC-Cy7	IM7	200	Biolegend	103028
CD45R-APC	RA3-6B2	200	BD Pharmingen	553092
CD48-Alexa-Fluor700	HM48-1	200	Biolegend	103426
CD48-FITC	HM48-1	200	Life Technologies	11-0481
CD71-BV786	C2	200	BD Pharmingen	740856
CD115-BV605	AFS98	400	Biolegend	135517
CD117-APC-eFluor780	2B8	200	Life Technologies	47-1171-80
CD117-BV711	2B8	800	Biolegend	105835
CD127-PE-Cy7	A7R34	100	Life Technologies	25-1271
CD135-APC	A2F10	100	Life Technologies	17-1351-80
CD150-BV605	TC15-12F12.2	100	Biolegend	115927

CD150-PE-Cy7	TC15-12F12.2	200	Biolegend	115914
Gr1-BV421	RB6-8C5	800	Biolegend	108445
Ly6C-APC-Fire750	HK1.4	200	Biolegend	128046
Ly6C-PerCP-Cy5.5	HK1.4	100	Biolegend	128011
Ly6G-PerCP-Cy5.5	1A8	200	BD Pharmingen	560602
Ly6G-BV421	1A8	200	BD Pharmingen	562737
NK1.1-BV421	PK136	100	Biolegend	108731
Sca1-PerCP-Cy5.5	D7	200	Life Technologies	45-5981
Ter119-BV421	TER-119	200	Biolegend	116234
Ter119-BV711	TER-119	200	BD Pharmingen	740686
ChromPure Mouse IgG, whole molecule	/	20	Jackson ImmunoResearch	015-000-003

2.1.5 Reagents and kits

Table 5. List of reagents and kits

Name	Source	Catalog Number
Agilent High Sensitivity DNA Kit	Agilent	5067-4626
High Sensitivity D1000 ScreenTape/Reagents	Agilent	5067-5592/ 5067-5593
High Sensitivity D5000 ScreenTape/Reagents	Agilent	5067-5584/ 5067-5585
RBC lysis buffer	Biolegend	420301
AMPure XP Reagents (60 ml)	Beckman Coulter	A63881
SPRIselect (60 ml)	Beckman Coulter	B23318
NextSeq 2000 P3 Reagents (100 Cycles)	illumina	20040559

NextSeq 1000/2000 P2 Reagents (100 Cycles) v3	illumina	20046811
NextSeq PhiX Control Kit	illumina	FC-110-3002
DirectPCR Lysis Reagent Tail	Peqlab	31-101-T
SMRT Cell 1M v3 Tray (4/Tray)	Pacific Bioscience	101-531-000
Sequel sequencing kit 3.0 (4 rxns)/Reagent plate	Pacific Bioscience	101-597-900
SMRTbell Express Template Prep Kit 2.0 (18 rxns)	Pacific Bioscience	100-938-900
Sequel Binding and Internal Ctrl Kit 3.0	Pacific Bioscience	101-626-600
Barcoded Overhang Adapter Kit- 8A	Pacific Bioscience	101-628-400
Barcoded Overhang Adapter Kit- 8B	Pacific Bioscience	101-628-500
Ampure PB, 5 ml	Pacific Bioscience	100-265-900
RNeasy Plus Mini Kit	Qiagen	74134
Expand Long Template PCR system	Roche	11759060001
SuperScript First-Strand Synthesis System for RT-PCR	Thermo Fisher Scientific	11904018
dNTP Mix (10 mM each)	Thermo Fisher Scientific	R0192
Recombinant Proteinase K Solution (20 mg/mL)	Thermo Fisher Scientific	AM2546
TriTrack DNA Loading Dye (6X)	Thermo Fisher Scientific	R1161
GeneRuler DNA Ladder Mix	Thermo Fisher Scientific	SM0331
Dynabeads Untouched Mouse CD4 Cells Kit	Thermo Fisher Scientific	11415D
SYTOX Blue Dead Cell Stain	Thermo Fisher Scientific	S34857
OneComp eBeads compensation beads	Thermo Fisher Scientific	01-1111-42
Qubit dsDNA HS Assay Kit	Thermo Fisher Scientific	Q32854
Cell count plastic slide for automated cell counter	Nexcelom	CHT4-PD100-002

Chromium Single Cell 3' Library & Gel Bead Kit v2, 16 rxns	10x Genomics	PN-120237
Chromium Single Cell A Chip Kit, 48 rxns	10x Genomics	PN-120236
Chromium Single Cell 3' GEM, Library & Gel Bead Kit v3, 16 rxns	10x Genomics	PN-1000075
Chromium Chip B Single Cell Kit, 48 rxns	10x Genomics	PN-1000073
Chromium i7 Multiplex Kit, 96 rxns	10x Genomics	PN-120262

2.1.6 Cell lines

Table 6. List of cell lines

Name	Source
Cell lines	
Rosa26$^{PolyloxExpress/+}$ ESC (*Mer-iCre-Mer* transfected)	Weike Pei/Thorsten Feyerabend, DKFZ
Mouse embryonic fibroblast (MEF)	Weike Pei/Thorsten Feyerabend, DKFZ

2.1.7. Mice

Table 7. List of mice

Name	Source
Tie2$^{MeriCreMer/+}$	Katrin Busch, DKFZ
Rosa26$^{PolyloxExpress/+}$	Weike Pei/Thorsten Feyerabend, DKFZ
Rag2$^{-/-}$ *γc*$^{-/-}$ *Kit*$^{W/Wv}$	Katrin Busch, DKFZ
CD1	DKFZ
C57BL/6	DKFZ

2.1.8 Tools and equipment

Table 8. List of tools and equipment

Name	Source	Catalog Number
PIPETMAN Classic P10	Gilson	F144802
PIPETMAN Classic P20	Gilson	F123600
PIPETMAN Classic P200	Gilson	F123601
PIPETMAN Classic P1000	Gilson	F123602
PIPETBOY acu 2	Integra Biosciences	155000
Multichannel microliter pipette (0.5 - 10 µl)	Brand	705920
Multichannel microliter pipette (20 - 200 µl)	Brand	705930
Mortars and Pestles	VWR	N.A
Solofix blood lancets	B.Braun	6185002
Heracell 240i CO_2 Incubator	Thermo Fisher Scientific	51032875
Safe 2020 Biological Safety Cabinets	Thermo Fisher Scientific	N.A
Water Bath	GFL	1003
VACUSAFE safe aspiration system	Integra Biosciences	158310
Megafuge 40R TX-750 Cell Culture Centrifuge	Thermo Fisher Scientific	50126352
Fresco 17 Microcentrifuge	Thermo Fisher Scientific	75002402
Gammacell 40 Exactor	Best Theratronics	N.A
Cellometer Auto 2000 Cell Viability Counter	Nexcelom	N.A
FACS Fortessa	BD Pharmingen	N.A
FACS Aria III	BD Pharmingen	N.A
Vortex-Genie 2	Scientific Industries	SI-0236

ThermoMixer Comfort	Eppendorf	5355 000.011
Azure C200 Gel Imaging Workstations	Azure Biosystems	N.A
Horizontal Electrophoresis Systems	Bio-Rad	N.A
Dynabeads MPC-6 Magnetic Particle Concentrator	Thermo Fisher Scientific	12002D
10x Magnetic Separator	10x Genomics	PN-120250
10x Vortex Adapter	10x Genomics	PN-120251
10x Chip Holder	10x Genomics	PN-120252
10x Chromium Controller	10x Genomics	N.A
C1000 Touch Thermal Cycler with 96–Deep Well Reaction	Bio-Rad	1851197
T3 thermocycler	Biometra	N.A
Qubit 2.0 Fluorometer	Thermo Fisher Scientific	N.A
Qubit 4.0 Fluorometer	Thermo Fisher Scientific	Q33226
2100 Bioanalyzer Instrument	Agilent	G2939BA
4200 TapeStation System	Agilent	G2991AA
PacBio Sequel system	Pacific Bioscience	N.A
ALPS 50 V-Manual Heat Sealer	Thermo Fisher Scientific	AB-1443A
NextSeq 1000 & NextSeq 2000 Systems	illumina	N.A
HiSeq4000 platform	illumina	N.A

2.1.9 Key software

Table 9. List of key software

Name	Source
EndNote 20	Clarivate
Adobe Illustrator 2022	Adobe

GraphPad Prism 8	GraphPad
FlowJo 10.8.1	BD Biosciences
FACSDIVA version 6.1.3	BD Biosciences
Gene Construction Kit (v4.0)	Textco

2.2 Methods

2.2.1 Cell culture

2.2.1.1 MEF cell culture, passaging and cryopreservation

Embryonic stem cells were routinely cultured on monolayers of murine embryonic feeder cells (MEFs). Therefore, six cell culture plates (10 cm) were gelatinized with 8 ml 0.1% gelatin solution for 30 min at room temperature (Table 10). In the meantime, a vial of MEF was taken from N_2 storage, quickly thawed at 37 °C in a water bath and cells were resuspended in 60 ml MEF culture medium (Table 11). After aspirating the gelatin solution from the coated cell culture plates, resuspended MEFs were equally split into these plates (10 ml/plate) and kept at 37 °C, 10% CO_2 until confluent.

MEFs can be split and passaged around 3 times at a ratio of 1:3 - 1:4. To split and expand MEFs, 18 - 24 cell culture plates (10 cm) were gelatinized with 8 ml 0.1% gelatin solution for 30 min at room temperature. In parallel, trypsin solution was pre-warmed in a 37 °C water bath (Table 12). Next, MEF culture medium was aspirated and the confluent 10 cm plates were gently washed once with 10 ml DPBS to remove residual medium. The cells were then treated with 2 ml trypsin solution at 37 °C for 5 min, followed by adding 8 ml MEF culture medium per plate to stop the trypsinization process. MEFs were resuspended and collected in a final volume of 180 - 240 ml MEF culture medium. After removing excess gelatin solution from the coated plates, MEF suspension was equally split onto the plates (10 ml/plate), which were then left at 37 °C and 10% CO_2 in the cell incubator until confluent. Prior to their use in ESC culture, plates with confluent MEF layers were irradiated for a total of 3200 rad in a Gammacell 40 Caesium-137 source to inactivate further growth.

To cryopreserve MEFs, four confluent MEF culture plates (10 cm) were trypsinized, resuspended and collected in a 50 ml Falcon tube. Next, cells were pelleted by centrifugation (500 g, 5 min at 4 °C), resuspended in 1 ml freezing medium and stored in

a cryogenic tube (Table 13). The tube was transferred to - 80°C or liquid nitrogen for long-term storage.

Table 10. Gelatin solution (0.1%)

Component	Volume
Gelatin solution (2%)	25 ml
DPBS	500 ml

Table 11. MEF culture medium

Component	Volume
Dulbecco's modified eagle medium (DMEM, no glutamine)	500 ml
FBS	60 ml
Sodium Pyruvate (100 mM)	6 ml
MEM Non-Essential Amino Acids Solution (100X)	6 ml
2-Mercaptoethanol (50 mM)	1.2 ml
Penicillin-Streptomycin (10,000 U/ml)	6 ml
GlutaMAX Supplement	6 ml

Table 12. Trypsin solution

Component	Volume
Trypsin-EDTA (0.25%)	/

Table 13. Freezing medium

Component	Volume
FBS	12.5 ml
Cell culture medium (MEF/ESC)	10 ml

DMSO	2.5 ml

2.2.1.2 ESC culture, passaging and cryopreservation

A vial of ESCs was gently thawed in a 37 °C water bath. Cells were resuspended in 10 ml ESC culture medium and pelleted by centrifugation (500 g, 5 min at 4 °C). Next, the cell pellet was resuspended in 20 - 30 ml ESC culture medium (Table 14) and gently distributed onto confluent and irradiated (3200 rad) 10 cm MEF plates (10 ml/plate). The plates were kept at 37 °C, 10% CO_2. To maintain ESC stemness, the culture medium was changed daily and cells were passaged every 2 - 3 days.

Table 14. ESC culture medium

Component	Volume
KnockOut DMEM	500 ml
Embryonic stem-cell FBS	60 ml
MEM Non-Essential Amino Acids Solution (100X)	6 ml
2-Mercaptoethanol (50 mM)	1.2 ml
Penicillin-Streptomycin (10,000 U/ml)	6 ml
GlutaMAX Supplement	6 ml
Recombinant Mouse LIF Protein	60 µl

To split and expand ESCs at the ratio of 1:4 - 1:10, 2 - 3 days after seeding, ESC culture plates were washed once with 10 ml DPBS and trypsinized with 2 ml trypsin solution at 37 °C for 5 min. The dissociation was stopped by addition of 8 ml ESC culture medium and the cells were thoroughly resuspended and transferred into 15 ml Falcon tubes. The tubes were left for 10 - 15 min at room temperature to let remaining cell clumps settle down. Finally, 8 ml of the supernatant were mixed with 40 – 100 ml ESC culture medium, redistributed onto 4 - 10 MEF plates (10 ml/plate) and grown at 37 °C and 10% CO_2. For

PolyloxExpress barcode induction *in vitro*, a detailed protocol is described in section 2.2.3.1.

To cryopreserve ESCs, 1 - 2 culture plates were washed once with 10 ml DPBS and trypsinized with 2 ml trypsin solution at 37 °C for 5 min. After adding 8 ml ESC culture medium to stop the trypsinization process, ESCs were pelleted by centrifugation (500 g, 5 min at 4 °C), resuspended in 1 ml freezing medium and transferred to a cryogenic tube. The tube was placed at - 80°C for at least 2 - 3 days before the tube was transferred to liquid nitrogen for long-term storage.

2.2.2 Mice

Mice were housed in individually ventilated cages with free access to food and water at the German Cancer Research Center. Hygiene status was controlled through regular hygiene monitoring with contact sentinels or by direct sampling from mice feces or blood. Both male and female mice were used without randomization or blinding. All mice experiments were conducted by following the institutional and governmental regulations, with approval from local authority (Regierungspräsidium, Karlsruhe, Germany).

2.2.2.1 Timed pregnancy

Experimental mice were bred by mating of *Tie2*$^{MeriCreMer/+}$ and *Rosa26*$^{PolyloxExpress/PolyloxExpress}$ mice. To determine the timing of pregnancy, one male heterozygous *Tie2*$^{MeriCreMer/+}$ mouse and two females homozygous *Rosa26*$^{PolyloxExpress/PolyloxExpress}$ mice were put together in the evening and left in the same cage overnight. On the next morning, between 7 - 8 am, male and female mice were separated and mating success was determined by observing vaginal plugs at the individual female mice. The date of the plug was documented as embryonic day 0.5 (E0.5). Foster mothers, for raising experimental mice that were born by cesarian section, were generated by timed mating of CD1 mice, started 1 - 2 days before the experimental mating. Details of *in vivo* barcode induction are described in section 2.2.3.2.

2.2.2.2 Genotyping of mice

Tissues from 4-week-old mice ear punch biopsies were collected in 1.5 ml eppendorf tubes, resuspended in 100 µl ‘‘DirectPCR Lysis Reagent’’ supplemented with 0.5 mg/ml proteinase K. The lysates were then incubated overnight on a shaker at 56 °C (800 rpm), followed by a proteinase K inactivation step by incubating the lysates for 10 - 20 min at 95 °C (800 rpm). For genotyping PCR, 2 µl aliquots of the lysates were used as DNA template (Table 15-18).

Table 15. *Tie2*$^{MeriCreMer/+}$ allele genotyping PCR master mix

Master Mix	Volume (1 reaction)
Buffer2* (Roche expand long template PCR system)	2 µl
dNTPs (2.5 mM)	1.6 µl
972 (10 µM)	1 µl
984 (10 µM)	1 µl
HL14 (10 µM)	1 µl
DNA Polymerase (Roche expand long template PCR system)	0.5 µl
Ear punch lysate	2 µl
Nuclease Free H_2O	10.9 µl
Total	**20 µl**

* Thaw and equilibrate at 54 °C until fully dissolved.

Table 16. *Tie2*$^{MeriCreMer/+}$ allele genotyping PCR program

Program	Cycle
95 °C, 2 min	1 cycle
95 °C, 20 s	35 cycles total
62 °C, 30 s	

72 °C, 1 min	
72 °C, 2 min	1 cycle
4 °C	hold

Table 17. *Rosa26*$^{PolyloxExpress/+}$ allele genotyping PCR master mix

Master Mix	Volume (1 reaction)
Buffer1* (Roche expand long template PCR system)	2.5 µl
dNTPs (2.5 mM)	2.5 µl
1116 (10 µM)	1 µl
1117 (10 µM)	1 µl
1118 (10 µM)	1 µl
DNA Polymerase (Roche expand long template PCR system)	0.5 µl
Ear punch lysate	2 µl
Nuclease Free H_2O	14.5 µl
Total	**25 µl**

* Thaw and equilibrate at 54 °C until fully dissolved.

Table 18. *Rosa26*$^{PolyloxExpress/+}$ allele genotyping PCR program

Program	Cycle
95 °C, 2 min	1 cycle
95 °C, 20 s	35 cycles total
60 °C, 30 s	
72 °C, 1 min	
72 °C, 2 min	1 cycle
4 °C	hold

PCR results were observed by gel electrophoresis (2% agarose, 110 V, 30 min). Expected fragment sizes are 746 bp for the *Tie2*^{MeriCreMer} allele and 527 bp for wild type. In the *Rosa26*^{PolyloxExpress} PCR, band sizes are 210 bp for wild type and 311 bp for the knock in allele.

2.2.2.3 Serial bleeding

Barcoded *Tie2*$^{MeriCreMer/+}$ *Rosa26*$^{PolyloxExpress/+}$ mice were bled monthly. By punching the submandibular vein on the cheek with a lancet, around 4 - 5 drops of blood were collected into EDTA-coated tubes. Samples were kept at 4 °C until further red blood cell lysis and fluorescence-activated cell sorting (FACS) staining were performed as described in section 2.2.4.

2.2.2.4 HSC transplantation

For the HSC transplantation experiments, FACS-purified LT-HSCs (see section 2.2.4 for details) were equally split into two portions. One specimen was used for DNA barcode PCR, as described in section 2.2.6.1, cells from the second were collected by centrifugation (400 g, 5 min, 4 °C) and resuspended in 200 µl sterile DPBS. This second cell sample was again equally split into two halves and each half was transferred to a 1-ml syringe and transplanted into one *Rag2*$^{-/-}$ *γc*$^{-/-}$ *Kit*$^{W/Wv}$ mouse through tail vein injection.

2.2.3 *PolyloxExpress* barcode induction *in vitro* and *in vivo*

2.2.3.1 *PolyloxExpress* barcode induction *in vitro*

Mer-iCre-Mer-transfected *Rosa26*$^{PolyloxExpress/+}$ ESCs were grown in ESC medium. For the induction of *PolyloxExpress* barcodes, 4-hydroxytamoxifen (4-OHT) was added to the ESC culture medium at a final concentration of 800 nM. For uninduced control, only ethanol, the solvent was added. Cells were kept in the media for the indicated time and afterwards washed three times with 10 ml DPBS before replacing with normal ESC culture

medium. Induced ESCs were chased for the indicated time and aliquots (0.5 - 1 million cells/aliquot) were harvested for subsequent bulk genomic DNA (gDNA) extraction or bulk mRNA extraction and *PolyloxExprss* barcode PCRs. Detailed *PolyloxExpress* PCR protocols are listed in section 2.2.6.

2.2.3.2 *PolyloxExpress* barcode induction *in vivo*

Vaginal plug positive mice were treated with 2.5 mg tamoxifen and 1.25 mg progesterone at embryonic day 9.5 (E9.5) by oral gavage. 10 days after the treatment (E19.5), pups were delivered by caesarean section and raised to 4-week-old by CD1 foster mothers.

The protocol for preparing 20 mg/ml tamoxifen stock is listed below:
- Sterilize pipettes and surface of the laminar flow cabinet with 70% ethanol.
- Prepare sterile 50 ml tubes, 2 ml Eppendorf tubes and 1000 µl filter tips.
- Add 1 g of tamoxifen powder into a sterile 50 ml tube in the laminar flow cabinet.
- Add 45 ml peanut oil and 5 ml absolute ethanol to the 50 ml tube, vortex and mix well.
- Incubate the stock solution at 55 °C and vortex in between until no precipitates are visible.
- Distribute aliquots of 1.5 ml tamoxifen solution into 2 ml Eppendorf tubes and store at -20 °C.

The protocol for preparing 10 mg/ml progesterone stock is listed below:
- Sterilize pipettes and surface of the laminar flow cabinet with 70% ethanol.
- Prepare sterile 50 ml tubes, 2 ml Eppendorf tubes and 1000 µl filter tips.
- Add 0.5 g of progesterone powder into a sterile 50 ml tube in the laminar flow cabinet.
- Add 45 ml peanut oil and 5 ml absolute ethanol to the 50 ml tube, vortex and mix well.
- Incubate the stock solution at 55 °C and vortex in between until no visible precipitates.
- Distribute aliquots of 1.5 ml progesterone solution to 2 ml Eppendorf tubes and store at -20 °C.

To prepare tamoxifen and progesterone mixture for oral gavage, 1 ml tamoxifen stock and 1 ml progesterone stock were mixed in a 2 ml Eppendorf tube under sterile environment. 250 µl of the mixed solution, containing 2.5 mg tamoxifen and 1.25 mg progesterone, were used for oral gavage per mouse.

2.2.4 Fluorescence-activated cell sorting

2.2.4.1 Organ preparation and cell isolation

For adult analysis, mice were sacrificed by cervical dislocation or CO_2 inhalation. For E18.5 analysis, pregnant female mice were sacrificed by cervical dislocation and embryos were sacrificed by decapitation. Spleen, thymus, bones (adult mice) and fetal liver (embryos) were harvested and immediately kept in 5% FACS buffer on ice (Table 19). Bone marrow cells were released by crushing the bones with mortar and pestle in 3 - 5 ml 5% FACS buffer. Cells were collected in a 50ml tube after passing through the 40 µm cell strainer (100 µm cell strainer was used for bone marrow cells if megakaryocytes were sampled). To maximize cell recovery, the crushed bones were washed at least 3 - 5 times with 5% FACS buffer until no red cell clumps were visible. Spleen and thymus were cut into pieces using sterile surgical scissors, smashed with the plunger of a 5 ml syringe by grinding on a 40 µm filter and flushed through the filter with FACS staining buffer to receive a single cell suspension.

Red blood cell (RBC) lysis buffer (1x concentrated) was used to deplete red blood cells from spleen, blood or fetal liver samples. Single cell suspensions from spleen and liver were prepared as described above, cells were collected by centrifugation (500 g, 5 min, 4 °C), and resuspended in 5 ml RBC lysis buffer at room temperature. After 180 sec the lysis was stopped by addition of 30 ml 5% FACS buffer and the suspension was filtered again through a 40 µm cell strainer. For blood collected in EDTA-coated tube, the sample was resuspended in 3 ml RBC lysis buffer, transferred to a 15 ml tube and kept at room temperature for around 10 min. The tube was then filled up to 15 ml with 5% FACS buffer to stop the lysis.

Table 19. 5% FACS buffer

Component	Volume
DPBS	500 ml
FBS	25 ml

2.2.4.2 Cell staining and optional lineage depletion

Cell numbers of the different samples were determined by automated cell counting. Desired numbers of cells were incubated at a concentration of 1 million cells/ 10 µl in FACS buffer containing 300 µg/ml whole mouse immunoglobulin G (IgG) for 15 min on ice to block Fc receptors. Cells were collected by centrifugation (500 g, 5 min, 4 °C) and resuspended in cocktails of the antibody mixtures described below (Table 20-23). Staining mixtures were kept on ice in the dark for 45 min. Afterwards, cells were washed twice with 1 ml FACS buffer, collected by centrifugation (500 g, 5 min, 4 °C) and either used for lineage depletion (HSC staining) or resuspended for flow cytometric analysis in filtered FACS buffer containing the live-dead discriminating dye Sytox^{TM} blue (1:10,000 dilution).

Table 20. *T cell/B cell/Granulocyte/Monocyte staining* (For spleen cells and blood)

Antibodies	Dilution	Clone
CD4 APC	400	RAM4-5
CD8a BV421	800	53-6.7
CD19 FITC	400	1D3
CD11b PE-Cy7	400	M1/70
Ly6G PerCP-Cy5.5	200	1A8
Ly6C APC-Fire750	200	HK1.4
CD115 BV605	400	AFS98
Ter119 BV711	200	TER-119

Table 21. ***Granulocyte/Monocyte/Erythroid Progenitor/proB cell staining*** **(For bone marrow cells)**

Antibodies	Dilution	Clone
Ly6G BV421	200	1A8
Ly6C PerCP-Cy5.5	100	HK1.4
CD115 BV605	400	AFS98
CD45R APC	200	RA3-6B2
CD43 PE-Cy7	200	S7
BP1 FITC	400	6C3
CD24 APC-eFluor780	50	M1/69
CD71 BV786	200	C2
Ter119 BV711	200	TER-119
CD44 Alexa-Fluor700	100	IM7

Table 22. ***preT cell staining*** **(For thymus)**

Antibodies	Dilution	Clone
CD11b BV421	800	M1/70
CD19 BV421	200	6D5
Gr1 BV421	800	RB6-8C5
Ter119 BV421	200	TER-119
NK1.1 BV421	100	PK136
CD4 APC	400	RM4-5
CD8a FITC	200	53-6.7
CD3ε BV605	100	145-2C11
CD44 APC-Cy7	200	IM7
CD25 PE-Cy7	1600	PC61

Table 23. *HSPC staining* (For bone marrow cells)

Table 23.1. 1st round staining

Antibodies	Dilution	Clone
CD4 BV421	800	GK1.5
CD8a BV421	800	53-6.7
CD11b BV421	800	M1/70
CD19 BV421	400	6D5
Gr1 BV421	800	RB6-8C5
Ter119 BV421	200	TER-119

- Lineage depletion: for the enrichment of rare progenitor cells, bone marrow samples were first blocked and stained with the 1st round staining solution as described above. Afterwards, cells were washed, resuspended in 10 ml FACS buffer containing dynabeads (2 beads/cell, or 5 µl beads/1 million cells) and incubated on a roller for 40 min at 4 °C. Lineage positive cells bound to the beads were removed using a magnetic stand, while lineage negative cells (not bound to beads) were collected with the supernatant and used for a 2nd round staining.

Table 23.2. 2nd round staining after lineage depletion (for HSCs, MPPs, Myeloid progenitors and Megakaryocytes)

Antibodies	Dilution	Clone
CD4 BV421	800	GK1.5
CD8a BV421	800	53-6.7
CD11b BV421	800	M1/70
CD19 BV421	400	6D5
Gr1 BV421	800	RB6-8C5
Ter119 BV421	200	TER-119
Nk1.1 BV421	100	PK136

CD117 APC-eFluor780	200	2B8
Sca-1 PerCP-Cy5.5	200	D7
CD150 PE-Cy7	200	TC15-12F12.2
CD48 Alexa-Fluor700	200	HM48-1
CD41 BV711	200	MWReg30
CD42d APC	200	1C2
CD34 FITC	25	RAM34
CD16/32 BV605	100	2.4G2

Table 23.3. 2nd round staining after lineage depletion (for HSCs, MPPs and Lymphoid progenitors)

Antibodies	Dilution	Clone
CD4 BV421	800	GK1.5
CD8a BV421	800	53-6.7
CD11b BV421	800	M1/70
CD19 BV421	400	6D5
Gr1 BV421	800	RB6-8C5
Ter119 BV421	200	TER-119
Nk1.1 BV421	100	PK136
CD117 BV711	800	2B8
Sca-1 PerCP-Cy5.5	200	D7
CD48 FITC	200	HM48-1
CD150 BV605	100	TC15-12F12.2
CD135 APC	100	A2F10
CD127 PE-Cy7	100	A7R34

2.2.4.3 Cell sorting

Purification of fluorescent-labeled cells was performed on a BD FACSAriaIII instrument (100 µm nozzle). Compensation values for the spectral overlap between the different dyes

were determined using single-color stained compensation beads and a fraction of stained cells was recorded to set the gating strategies for the cell of interests (Table 24). Cells were sorted into 0.2 μm-filtered 20% FACS buffer (Table 25) and kept on ice for subsequent scRNAseq or *PolyloxExpress* PCR. For mature cells, myeloid progenitor cells and lymphoid progenitor cells, 2 - 3 sample repeats (30,000 cell/sample) were sorted. For MPPs, 50 - 75% of the entire cells were collected in 1 - 2 specimen. For HSCs and megakaryocytes, all available cells were collected. From each sample, around 1 - 2% of the sorted cells were used for reanalysis to determine sort purity.

Table 24. Hematopoietic cell gating strategies

Cell type	FACS phenotypes
LT-HSC	Lin^- Sca-1^+ c-Kit^+ $CD150^+$ $CD48^-$
ST-HSC	Lin^- Sca-1^+ c-Kit^+ $CD150^-$ $CD48^-$
MPP	Lin^- Sca-1^+ c-Kit^+ $CD150^-$ $CD48^+$
CMP	Lin^- Sca-1^- c-Kit^+ $CD16/32^{med}$ $CD34^{med}$
GMP	Lin^- Sca-1^- c-Kit^+ $CD16/32^+$ $CD34^+$
MEP	Lin^- Sca-1^- c-Kit^+ $CD16/32^-$ $CD34^-$
Megakaryocyte Progenitors	Lin^- Sca-1^- c-Kit^+ $CD150^+$ $CD41^+$
CLP	Lin^- Sca-1^{lo} c-Kit^{lo} $CD127^+$ $CD135^+$
proB cells (Fraction B and C)	$Ter119^-$ $Ly6G^-$ $Ly6C^-$ $CD43^{med}$ $CD45R^{med}$ $CD24^+$ BP1^- and BP1^+
preT cells (DP)	$CD11b^-$ $Gr1^-$ $Ter119^-$ $CD19^-$ $Nk1.1^-$ $CD4^+$ $CD8^+$
preT cells (DN2 and DN3)	$CD11b^-$ $Gr1^-$ $Ter119^-$ $CD19^-$ $Nk1.1^-$ $CD4^-$ $CD8^-$ $CD3\varepsilon^-$ $CD25^+$ $CD44^+$ (DN2) and $CD25^+$ $CD44^-$ (DN3)
Erythroid progenitors (stage II)	$Ly6G^-$ $Ly6C^-$ $CD45R^-$ $Ter119^+$ $CD71^+$ $CD44^+$ FSC^+

Erythroid progenitors (stage III)	$Ly6G^-$ $Ly6C^-$ $CD45R^-$ $Ter119^+$ $CD71^+$ $CD44^{med}$ FSC^{med}
Megakaryocytes	Lin^- $Sca\text{-}1^-$ $CD41^+$ $CD42d^+$ FSC^{hi} SSC^{hi} $Sytox^+$
CD4+ T cells	$Ter119^-$ $CD11b^-$ $CD19^-$ $CD8^-$ $CD4^+$
CD8+ T cells	$Ter119^-$ $CD11b^-$ $CD19^-$ $CD4^-$ $CD8^+$
B cells	$Ter119^-$ $CD11b^-$ $CD4^-$ $CD8^-$ $CD19^+$
Granulocytes	$Ter119^-$ $CD19^-$ $CD4^-$ $CD8^-$ $CD11b^+$ $Ly6C^-$ $Ly6G^+$
Monocytes	$Ter119^-$ $CD19^-$ $CD4^-$ $CD8^-$ $CD11b^+$ $Ly6G^-$ $Ly6C^+$ $CD115^+$

Table 25. 20% FACS buffer

Component	Volume
5% FACS buffer	12.6ml
FBS	2.4ml

2.2.5 single-cell RNA sequencing

2.2.5.1 scRNAseq with standard 10x "Chromium Single Cell 3' Gene Expression" protocol

Sorted cells (maximum 20,000) were collected by centrifugation (500 g, 5 min twice at 4 °C) and resuspended in 44.6 µl filtered and pre-chilled FACS buffer. All the subsequent steps were conducted according to the manufacturer's instructions (CG00052 RevE & CG000183 RevA). In brief, the cell suspension was loaded on a chromium chip and encapsulated with indexed beads plus reverse transcription (RT) reagents using the 10x chromium controller. Next, the encapsulated cells were incubated in a thermal cycler for cell lysis and mRNA reverse transcription. complementary DNA (cDNA) library was then purified and amplified for 13 - 14 cycles using 10x common adapters (PCR extension time was increased to 3 min). After successful PCR amplification was confirmed using Bioanalyzer or Tapestation, 25% of 10x cDNA was used for Illumina transcriptome library

preparation according the manufacturer's instructions. The other fraction of the cDNA (5 - 10 ng) was used for *PolyloxExpress* RNA barcode PCR (see section 2.2.6.2 for details)

2.2.5.2 Modified scRNAseq protocol for target enrichment

To increase *PolyloxExpress* barcode capture efficiency, a modification was made to the standard 10x 3' scRNAseq protocol. Specifically, all the steps before global cDNA amplification were conducted exactly the same as described in section 2.2.5.1. After the clean-up step, cDNA was resuspended in 34 µl elution solution. Next, 1 µl *PolyloxExpress* targeting oligo (ISPCR-2999, 6 µM stock), 15 µl 10x cDNA primers and 50 µl 10x amplification mix were added to the cDNA elute for a total of 100 µl PCR reaction volume. PCR and all the following steps were conducted exactly the same as described in section 2.2.5.1.

2.2.6 *PolyloxExpress* barcode PCR

2.2.6.1 DNA barcode PCR (optional with sample index primer)

Bulk sorted mature cell and progenitor cell samples were collected by centrifugation (1500 g, twice for 5 min at 4 °C) and resuspended in 25 µl small scale lysis buffer (Table 26). The samples were lysed for 1 hour at 55 °C and proteinase K was inactivated for 10 min at 95 °C.

Table 26. Small scale lysis buffer

Component	Volume (1 reaction)
Buffer1* (Roche expand long template PCR system)	2.5 µl
Proteinase K (20 mg/ml)	0.63 µl
Nuclease Free H_2O	21.87 µl
Total	**25 µl**

* Thaw and equilibrate at 54 °C until fully dissolved.

For the amplification of DNA barcodes, the lysates were transferred to PCR strips and 20 µl freshly prepared master mix was added to each sample (Table 27). For the scRNAseq experiments (Experiment #1-4, section 3.3), DNA barcode PCRs were done using primer combination 2653 and 2427 (conditions see Table 27,28). To facilitate PacBio library preparation for large number of samples, DNA barcodes from the E18.5 experiments (Section 3.2.1), clonal dynamic experiments (Section 3.2.2) and transplantation experiments (Section 3.2.3) were amplified using one of the 2426 index primers (Table 3) and the common primer 2702 (conditions see Table 27,29).

Table 27. DNA barcode PCR master mix

Master Mix	Volume (1 reaction)
Buffer1* (Roche expand long template PCR system)	2.5 µl
dNTPs (2.5 mM)	5 µl
2653 (10 µM) or 2426 (10 µM, index)	3 µl
2427 (10 µM) or 2702 (10 µM)	3 µl
DNA Polymerase (Roche expand long template PCR system)	0.75 µl
Nuclease Free H_2O	5.75 µl
Total	**20 µl**

* Thaw and equilibrate at 54 °C until fully dissolved.

Table 28. PCR program for 2653+2427 primer combination

Program	Cycle
95 °C, 5 min	1 cycle
95 °C, 30 s	35 cycles total
54 °C, 30 s	

72 °C, 3 min	
72 °C, 10 min	1 cycle
4 °C	hold

Table 29. PCR program for 2426(index)+2702 primer combination

Program	Cycle
95 °C, 5 min	1 cycle
95 °C, 30 s	35 cycles total
60 °C, 30 s	
72 °C, 3 min	
72 °C, 10 min	1 cycle
4 °C	hold

2.2.6.2 RNA barcode PCR

To amplify full-length *PolyloxExpress* RNA barcodes from 10x cDNA libraries, a nested PCR protocol was applied and two aliquots of cDNA (5 - 10 ng each) were used as input for PCR replicates. Initially when setting up the nested PCR, the standard 10x protocol was applied with oligo 2652 as one of the primers for the 1st round nested PCR (Section 3.1.2). However, in the 10x target enrichment protocol, a primer at the inner end of primer 2652 (3' direction), 2999, was used in the target enrichment 10x protocol. Therefore, 2999 instead of 2652 was used for amplifying RNA barcodes from target enrichment protocol. A detailed PCR protocol, compatible with both standard or target enrichment 10x protocol, using primer 2999 is listed below (Table 30-33).

Table 30. 1st round RNA barcode PCR

Master Mix	Volume (1 reaction)
Buffer3* (Roche expand long template PCR system)	5 µl

dNTPs (2.5 mM)	10 µl
2674 (10 µM)	5 µl
2999 (10 µM)**	5 µl
DNA Polymerase (Roche expand long template PCR system)	0.75 µl
cDNA	3 µl
Nuclease Free H_2O	21.25 µl
Total	**50 µl**

* Thaw and equilibrate at 54 °C until fully dissolved.

** Can be replaced with 2652 for cDNA generated with standard 10x protocol.

Table 31. 1st round PCR program

Program	Cycle
95 °C, 5 min	1 cycle
95 °C, 30 s	12 cycles total
60 °C, 30 s	
72 °C, 3 min	
72 °C, 10 min	1 cycle
4 °C	hold

The 1st round PCR was purified by 0.7x AMPure XP beads, eluted in 25 µl nuclease free H_2O and used for 2nd round PCR (Table 32-33).

Table 32. 2nd round RNA barcode PCR

Master Mix	Volume (1 reaction)
Buffer3* (Roche expand long template PCR system)	5 µl
dNTPs (2.5 mM)	10 µl
2676 (10 µM)	5 µl

2426 (10 µM)	5 µl
DNA Polymerase (Roche expand long template PCR system)	0.75 µl
1st round PCR elute	24.25 µl
Total	**50 µl**

* Thaw and equilibrate at 54 °C until fully dissolved.

Table 33. 2nd round PCR program

Program	Cycle
95 °C, 5 min	1 cycle
95 °C, 30 s	16 cycles total*
60 °C, 30 s	
72 °C, 3 min	
72 °C, 10 min	1 cycle
4 °C	hold

* 16 cycles are sufficient and create less PCR bias.

PCR quality was determined by gel electrophoresis (1.5% agarose, 110 V, 55 min) using 12.5 µl of the PCR amplicons. The remaining PCR products were purified by 0.7x AMPure XP beads and eluted in 50 - 70 µl buffer EB for a final concentration of 30 - 40 ng/µl.

2.2.7 Next generation sequencing

2.2.7.1 Long read *PolyloxExpress* barcode sequencing (PacBio platform)

Purified full-length *PolyloxExpress* barcodes were sequenced using PacBio single-molecule real-time sequencing (SMRT). Samples amplified with 2426 index primer were processed using Pacbio library preparation protocol for barcoded primers (part number 101-791-800 version 02), samples amplified without using 2426 index primer (include RNA barcode PCR) were processed using Pacbio library preparation protocol for

barcoded overhang adapters (part number 101-791-700 version 05). All the steps were conducted according to the manufacturer's instructions. For each Pacbio library, up to eight different indexed samples amplified from 30,000 cells were pooled (100 - 120 ng/sample) to ensure a minimum of 40,000 - 50,000 polymerase reads per sample. After library preparation and polymerase binding, the library was quantified using Qubit and loaded on SMRT cell for sequencing (45 pM loading concentration). The sequencing parameters are as follows: movie time 10 hours, pre-extension time 0.6 hour, immobilization time 2 hours. After the sequencing was done, circular consensus sequencing (CCS) reads were retrieved using SMRT Link software and subjected to bioinformatic analysis.

2.2.7.2 Short read transcriptome sequencing (Illumina platform)

Transcriptome libraries for experiments #1-4 (section 3.3) were sequenced at the DKFZ genomics and proteomics core facility using the HiSeq4000 platform with the following sequencing parameters: paired-end sequencing, read1 28bp, i7 index 8bp, read2 74bp. Libraries for target enrichment set up experiments (section 3.1.3) were sequenced at the DKFZ sequencing open lab using the Nextseq1000/2000 platform with the following sequencing parameters: paired-end sequencing, read1 28bp, i7 index 8bp, read2 91bp.

3 Results

3.1 Establishing the *PolyloxExpress* system

3.1.1 *PolyloxExpress* DNA and RNA barcodes are consistent

The *PolyloxExpress* system generates barcode on both DNA and RNA level in the same cell. They are supposed to be identical because the RNA barcodes are transcribed from the Cre-recombined DNA locus. However, whether intrinsic cellular factors (e.g., RNA modification post transcription) could interfere with the generation of the correct RNA barcode sequences remains unknown. Therefore, it is important to demonstrate that *PolyloxExpress* DNA and RNA barcodes are consistent.

To address this experimentally, I performed a pulse-chase experiment together with Weike Pei using *Rosa26*$^{PolyloxExpress/+}$ targeted embryonic stem cells (ESCs) that were transfected with a *Mer-iCre-Mer* (a tamoxifen inducible Cre, designated MiCM) expression vector. The cells were treated for 16 hours with 4-hydroxy-tamoxifen (4-OHT), the active metabolite of tamoxifen, to induce barcode generation. Next, 16 days post-induction, bulk gDNA and mRNA were harvested (Figure 7A). I reasoned that after this long chase, DNA recombination should have been completed, and that RNA barcodes generated during intermediate recombination steps would have been degraded. Hence, only RNA barcode identical to DNA barcode should be expressed in each individual ESC. As indicated in Figure 7B, barcode polymerase chain reaction (PCR) amplicons obtained from bulk gDNA or mRNA revealed five recombination bands with comparable distribution patterns, suggesting that barcodes were successfully induced at both DNA and RNA level. Due to the inversion nature of the Cre-loxP system, each of the five bands contains highly diverse barcode sequences. To molecularly define the association between DNA and RNA barcodes, PCR repeats of both origins were profiled on the Pacbio sequencing platform. The PCR repeats of barcodes derived from DNA or RNA showed strong correlation and overlap, indicating that the majority of barcodes were faithfully amplified and recovered at bulk DNA or RNA level (Figure 7C). Importantly, a strong correlation

was also observed when DNA and RNA barcodes were compared against each other, suggesting that barcodes from both origins are highly similar at the bulk level (Figure 7C).

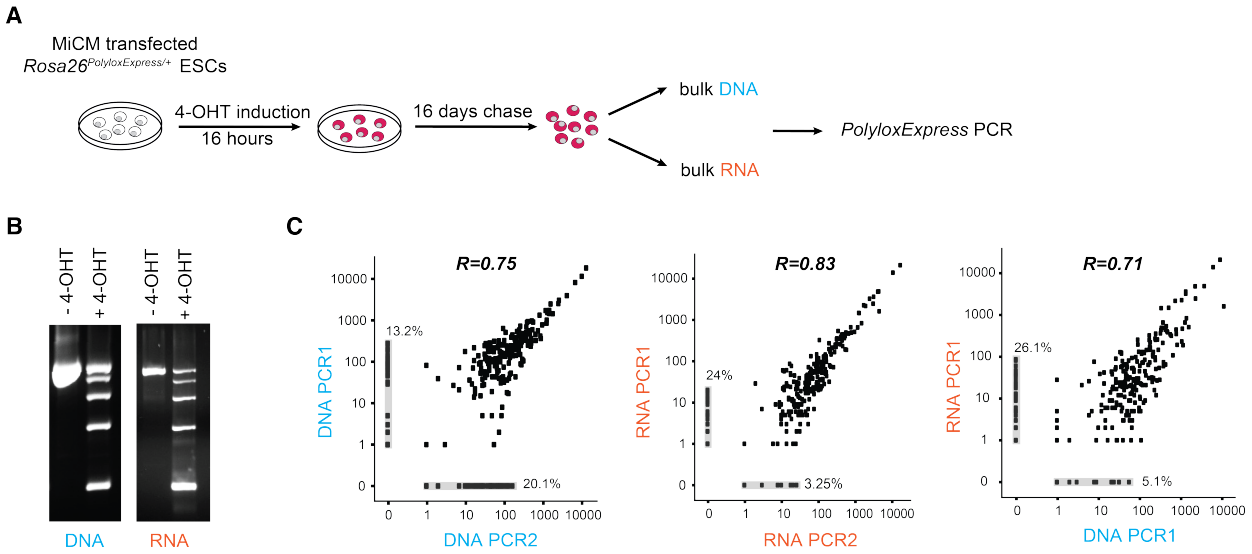


Figure 7. Correlation analysis of DNA and RNA barcodes.

(A) Experimental design of a pulse-chase experiment testing for DNA and RNA barcode correlation. After 16 hours of 4-OHT induction, MiCM-transfected *Rosa26*$^{PolyloxExpress/+}$ embryonic stem cells (ESCs) were chased for 16 days and bulk gDNA and mRNA were extracted from aliquots of the cells to compare barcode correlation.

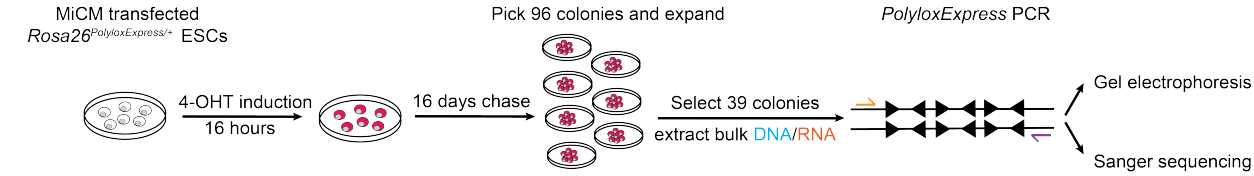
(B) *PolyloxExpress* PCR on gDNA and RT-PCR (reverse transcription-PCR) on mRNA of ESCs treated with (+ 4-OHT) or without 4-OHT (- 4-OHT).

(C) Barcode correlation between PCR repeats from DNA (left) or RNA (middle), as well as between PCR samples from DNA and RNA based on barcode read counts obtained by Pacbio sequencing. Each dot corresponds to a *PolyloxExpress* barcode. Unique barcodes, i.e., those found in one sample but not in the other, are depicted in grey boxes at the axes. Spearman's rank correlation coefficients (R) are indicated above the graphs. (Bioinformatic calculation by Xi Wang)

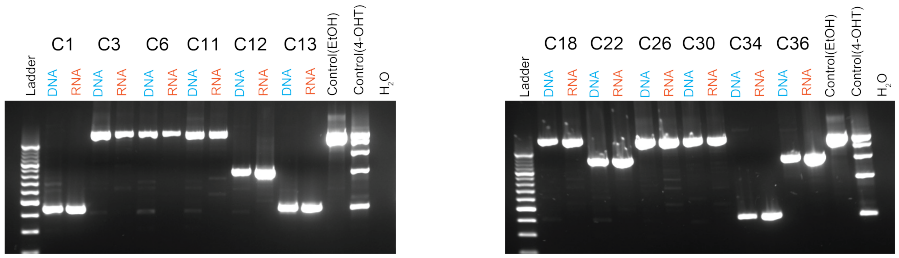
Correlation analysis on bulk samples does not necessarily mean that DNA and RNA barcodes are consistent in each individual cell. To specifically address this question, ESCs from the previous pulse-chase experiment (Figure 7A) were grown after the chase at low density to form individual colonies (Figure 8A). I picked 96 of these colonies and expanded the clones in separate wells. In the end, 39 clones were randomly selected to extract gDNA and mRNA for amplifying *PolyloxExpress* barcodes (Figure 8A). Consistent

with the idea that recombination has completed after a long chase, only a single DNA or RNA barcode was observed from all colonies examined (Figure 8B). Of note, the size of the DNA and RNA barcodes differed between several clones, but was identical within the same clone (Figure 8B). Sanger sequencing of all 78 PCR products confirmed that the DNA and RNA barcode of individual clones were identical not only by size but also by sequence (Figure 8C). Therefore, *PolyloxExpress* DNA and RNA barcodes are both accessible and consistent at the single cell level.

A



B



C

Colony ID	DNA Barcode	RNA Barcode	Colony ID	DNA Barcode	RNA Barcode
C1	1	1	C18	123456789	123456789
C3	GFEDCBA89	GFEDCBA89	C22	12345	12345
C6	123456789	123456789	C26	123456789	123456789
C11	123456789	123456789	C30	123456789	123456789
C12	129	129	C34	I	I
C13	G	G	C36	14789	14789

Figure 8. Validating DNA and RNA barcode consistency at single-cell resolution.

(A) Workflow of the clonal DNA and RNA barcode analysis experiment. From 39 barcoded *Rosa26*$^{PolyloxExpress/+}$ embryonic stem cell (ESC) clones, gDNA and mRNA was extracted and *PolyloxExpress* barcodes were amplified by PCR.

(B) PCR products were observed for fragment lengths by gel electrophoresis. A representative example of 12 *Rosa26*$^{PolyloxExpress/+}$ ESC clones is shown. C, colony; EtOH, ethanol.
(C) Shown are barcode sequences identified by Sanger sequencing from the 12 exemplary ESC clones in **(B)**.

3.1.2 Establishing a customized *PolyloxExpress* enrichment workflow compatible with whole transcriptome single-cell RNA sequencing

The advancements of microfluidic-based scRNAseq techniques (e.g., Chromium Single Cell Gene Expression platform from 10x Genomics) have allowed high throughput transcriptome profiling at single-cell resolution. In the standard protocol, each cell is encapsulated with one indexed bead in a single-droplet (Figure 9A). All beads are conjugated with oligos that contain a common adapter sequence (blue bar, Figure 9A) and a bead-specific index sequence (purple bar, Figure 9A). In each droplet, all mRNA molecules from a single cell are captured by the conjugated oligos on the bead and therefore receive a cell-specific index (cell identifier) during the subsequent reverse transcription (Figure 9A). Afterwards, cDNA molecules from different cells are pooled and globally amplified using common adapters (Figure 9A). Finally, the cDNA library will be fragmented, ligated with sequencing adapters and sequenced (Figure 9A). However, the fragmentation step in the standard protocol cuts full-length cDNA molecules into short fragments, which would preclude the possibility of recovering full-length *PolyloxExpress* barcode sequences experimentally. In addition, the standard scRNAseq protocol does not provide a solution for enriching genes of interest from pooled cDNA libraries, which may result in low capture efficiency of important targets (e.g., *PolyloxExpress* barcode). Therefore, customization of the current scRNAseq protocol was required to obtain the full-length *PolyloxExpress* RNA barcode while maintaining robust transcriptome information.

To this end, I hypothesized that one could split the full-length cDNA library generated after global amplification into representative aliquots for different subsequent protocols and purposes. One cDNA aliquot would be used for profiling the global transcriptome with the standard scRNAseq protocol and short read Illumina sequencing, while the other

aliquot could be used for full-length *PolyloxExpress* barcode enrichment (Figure 9B). By reconciliation of the cell indices obtained in the two different methods, it should be possible to recover and match the *PolyloxExpress* barcode and transcriptome of each individual cell (Figure 9B).

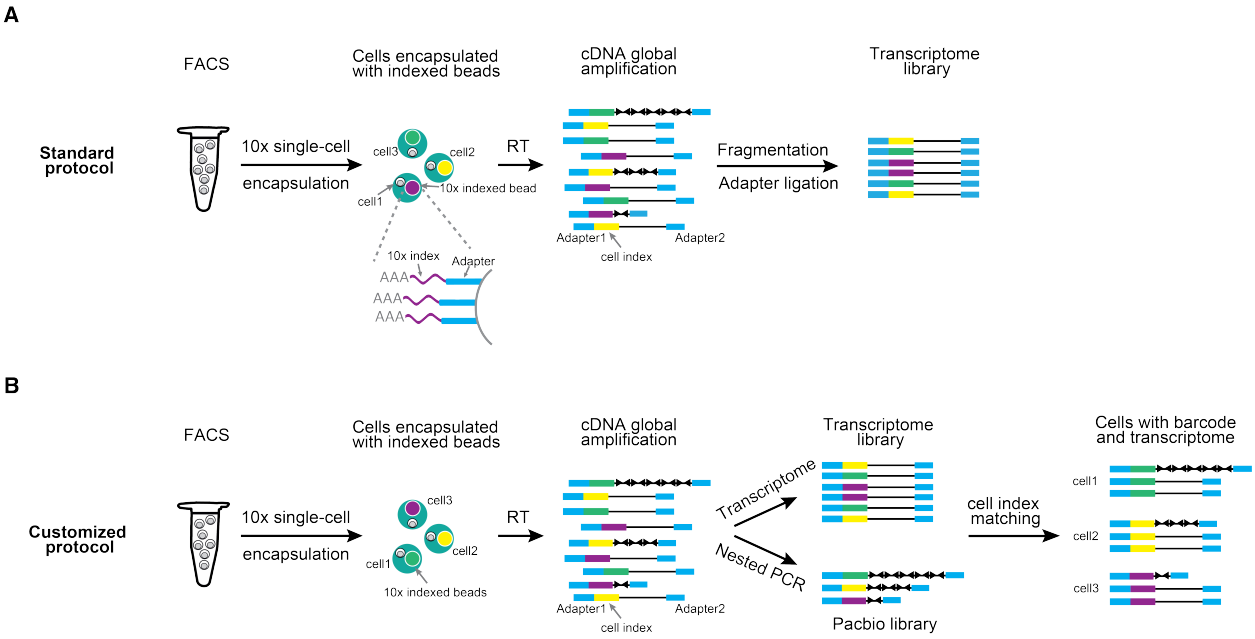


Figure 9. Customizing scRNAseq protocol to recover full-length *PolyloxExpress* RNA barcodes.
(A) Workflow of the standard 10x (10x Genomics) scRNAseq protocol. FACS, fluorescence-activated cell sorting; RT, reverse transcription.
(B) Workflow of customized scRNAseq protocol in compatible with full-length RNA barcode recovery. cDNA is split into two aliquots, one for transcriptome library preparation and one for *PolyloxExpress* barcode amplification using nested PCR. Transcriptome and barcode information are matched for individual cells via the bead specific 10x index (cell identifier).

In order to test this hypothesis, I generated two single-cell cDNA libraries from 15,000 uninduced and 15,000 induced *Rosa26*$^{PolyloxExpress/+}$ ESCs using the 10x "Chromium Single Cell 3' Gene Expression" protocol (Figure 10A). A portion of ESCs from the same dish in both conditions was also taken to extract gDNA and amplify *PolyloxExpress* barcodes for determining the ''reference'' recombination pattern. From each single-cell cDNA library, aliquots were taken for establishing an RNA barcode enrichment PCR and preparing the transcriptome library (Figure 10A).

In a preliminary experiment, I found that a normal PCR protocol was not sufficient for amplifying *PolyloxExpress* barcodes with good yield and high specificity because on the 3' end a common primer has to be used (data not shown). To improve this, I designed a nested PCR protocol. The difficulty was that the cell index locates at the 3' end of the common adapter (Figure 10B) and therefore the two 3' nested PCR primers 2674 and 2676 have to be chosen to preserve this cell index information (Figure 10B). Primer 2674 was designed in a way that it recognizes the common adapter binding sequence at its 3' end and contains an overhang sequence at its 5' end (Figure 10B). The 1st round nested PCR amplified molecules will contain the overhang sequence, which provides a binding site for primer 2676 in the 2nd round of nested PCR (Figure 10B). To increase PCR specificity to the *PolyloxExpress* cDNA, two primers 2652 and 2426 that specifically bind to the 5' end of the *PolyloxExpress* cDNA were designed (Figure 10B). Hence, the nested PCR is conducted using the combination of ''2674+2652'' for the 1st round and ''2676+2426'' for the 2nd round (Figure 10B).

To obtain sufficient PCR amplicons and at the same time minimize non-specific amplification, different nested PCR cycle combinations were tested. Compared to the reference barcode distribution obtained by bulk gDNA PCR (Figure 10C), the PCR conditions ''5+30'' or ''10+25'' (cycle numbers in the 1st + 2nd round of PCR) with a total of 35 cycles failed to amplify the barcodes with high specificity (Figure 10D). Strong bias towards the shortest barcode was observed, regardless of the different cycle combinations applied (Figure 10D). Therefore, I decreased the total PCR cycle number to 30 with the cycle combination set to ''10+20'' or ''12+18''. As a result, the typical pattern of five *PolyloxExpress* barcode bands was successfully recovered from the induced ESC library (Figure 10E). In addition, only one band with expected size was amplified from the uninduced ESC library, suggesting that PCR specificity was significantly improved (Figure 10E). Of note, given that the primers 2674/2676 are not *PolyloxExpress* cDNA specific, background amplification of other cDNA molecules, characterized by a faint smear around the size of the longest barcode, is inevitable (Figure 10F). In the subsequent sequencing steps, reads assigned to these background amplicons will be filtered and discarded.

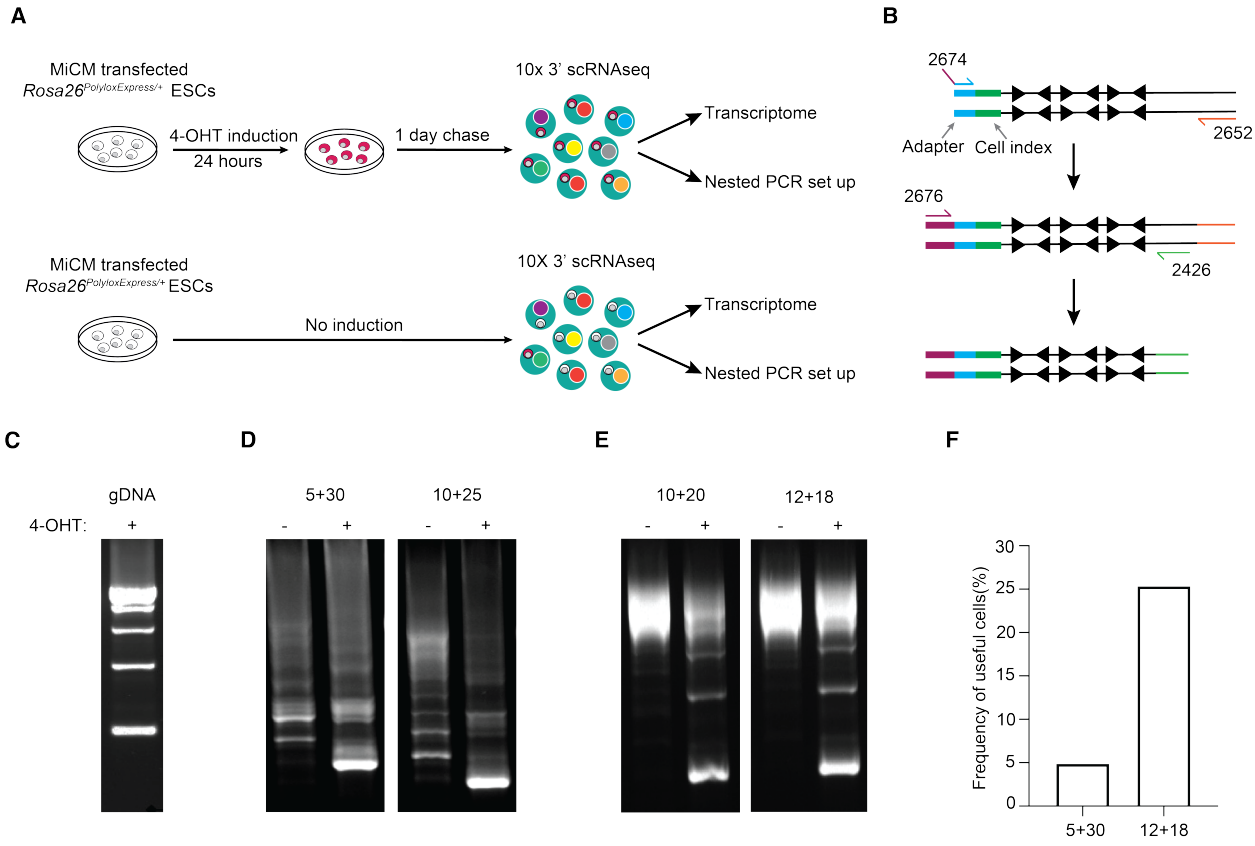


Figure 10. Establishing a nested PCR for specific barcode amplification from 10x cDNA library.

(A) Experimental design for nested PCR set up based on scRNAseq cDNA library. 15,000 inducible cre (MiCM) transfected embryonic stem cells (ESCs) from each condition were used as input for 10x 3' scRNAseq.

(B) Schematic drawing of the nested PCR strategy and primers used for amplifying *PolyloxExpress* barcodes. 2674 and 2676 are common primers for preserving cell index information, 2652 and 2426 are primers specific to the *PolyloxExpress* sequence.

(C) Gel electrophoresis of *PolyloxExpress* barcodes amplified form bulk gDNA.

(D) Gel electrophoresis of nested PCR test results on ESC scRNAseq cDNA libraries (with or without barcodes induction) with different 1^{st}+2^{nd} round cycle combinations and a total of 35 cycles.

(E) Gel electrophoresis of nested PCR test results with different 1^{st}+2^{nd} round cycle combinations and a total of 30 cycles.

(F) The proportion of useful cells among 10x captured cells was significantly improved by reducing the total nested PCR cycles. (Bioinformatic calculation by Xi Wang)

To determine the quality of the PCR results, I first sequenced the transcriptome libraries generated from the single-cell cDNA aliquot on Illumina. The sequencing data revealed that around 1/3 of the FACS-purified cells could be captured and transcribed into cDNA library with the 10x protocol, which is in agreement with the expected statistical distribution of single cells in limiting dilution. Next, I sequenced induced ESC amplicons from the ‘‘5+30’’ (Figure 10D) and ‘‘12+18’’ (Figure 10E) nested PCR using Pacbio. By matching the cell indices detected in both transcriptome and *PolyloxExpress* barcode amplicons, the number of cells that fulfill the following criteria (designated ‘‘useful’’ cells) were determined: 1) Containing an intact *PolyloxExpress* barcode 2) Only one *PolyloxExpress* barcode per cell index, and 3) a match of the cell index for both *PolyloxExpress* barcode and transcriptome. Consistent with the gel electrophoresis picture, the number of useful cells obtained from the ‘‘12+18’’ PCR was about 5-fold higher than from the ‘‘5+30’’ PCR (Figure 10F). Compared to the 4.9% useful cells in the initial test, around 25.4% of 10x captured cells successfully passed the stringency filtering with the ‘‘12+18’’ enrichment protocol (Figure 10F).

In summary, the specific design of nested primers compatible with the 10x adapter sequences and the titration of PCR amplification cycles enabled the establishment of a customized library preparation protocol. It is compatible with scRNAseq and allows specific enrichment and amplification of full-length *PolyloxExpress* bardcodes (Details of the nested PCR protocol can be found in the ‘‘Materials and Methodology’’ section). Of note, *PolyloxExpress* barcodes cannot be obtained from every 10x captured cell, probably due to limited mRNA capture efficiency of 10x beads or a relative low expression level of *PolyloxExpress* barcodes from the *Rosa26* locus.

3.1.3 Improving *in vivo* single-cell RNA barcode capture efficiency by target enrichment

With the advancement in scRNAseq techniques, several recent studies reported that target-specific oligo spike-in during the global cDNA amplification step could increase target capture efficiency without interfering with the transcriptome information (Giustacchini et al. 2017; Mimitou et al. 2019; Nam et al. 2019). Therefore, to test whether the previously established customized *PolyloxExpress* protocol could be further improved using target enrichment approach, I decided to modify the standard 10x protocol and spike in a *PolyloxExpress*-specific oligo at the global cDNA amplification step (Figure 11A). *Tie2*$^{MeriCreMer/+}$ and *Rosa26*$^{PolyloxExpress/PolyloxExpress}$ mice were crossed to obtain *Tie2*$^{MeriCreMer/+}$*Rosa26*$^{PolyloxExpress/+}$ reporter mice. In the *Tie2*$^{MeriCreMer/+}$ allele, a codon-improved Cre fused to two modified estrogen receptor binding domains is expressed under the control of the Tie2 locus (Busch et al. 2015). Upon tamoxifen treatment, Cre is activated and induces *PolyloxExpress* barcode generation in Tie2 positive cells. I treated *Tie2*$^{MeriCreMer/+}$*Rosa26*$^{PolyloxExpress/+}$ mice with tamoxifen in utero at E9.5 to label hematopoietic progenitors (Figure 11B). While the blood system is built up from labelled cells, the recombined *PolyloxExpress* barcodes are propagated into HSCs and their downstream progeny. At 17 weeks of age, one mouse was sacrificed to harvest barcoded hematopoietic progenitors (LSKs) for scRNAseq (Figure 11B). Of note, FACS purified LSKs were separated into three aliquots containing the same number of cells to test standard and modified protocols. Specifically, Aliquot 1 went through the previous customized *PolyloxExpress* protocol without any modifications (Figure 11C). Aliquot 2 and 3 went through the standard RT step as aliquot 1, but received a *PolyloxExpress*-specific oligo (ISPCR-2999) spike in to the standard 10x cDNA primer mix (Aliquot 2: 0.03 µM final; Aliquot 3: 0.06 µM final) before cDNA amplification (Figure 11C).

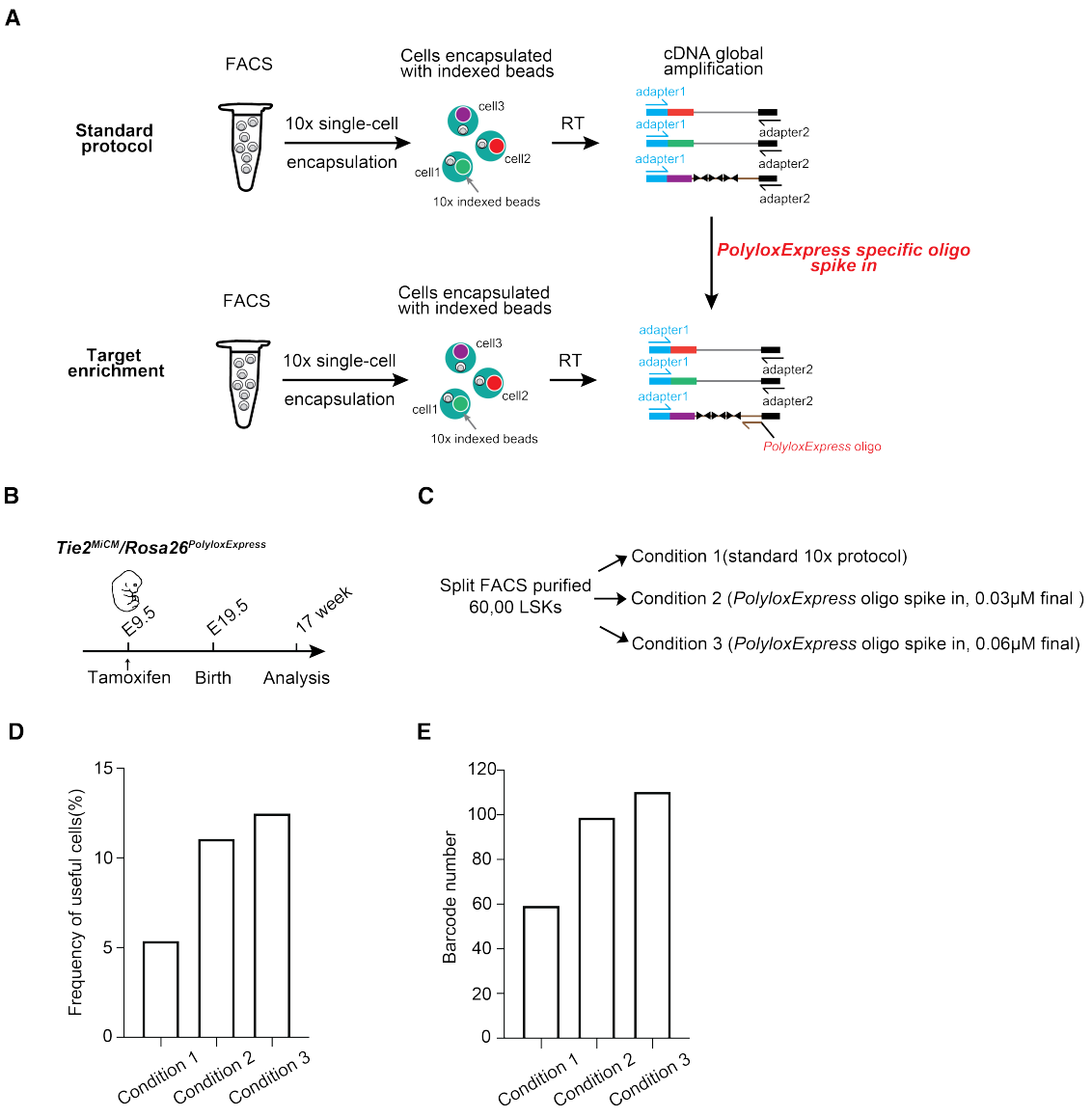


Figure 11. Improving barcode capture efficiency by target enrichment.

(A) Schematic diagram of standard and target enrichment scRNAseq protocols. In the target enrichment protocol, a small fraction of *PolyloxExpress*-specific oligos was spiked into the PCR mix before global cDNA amplification. FACS, fluorescence-activated cell sorting; RT, reverse transcription.

(B) *Tie2*$^{MeriCreMer/+}$*Rosa26*$^{PolyloxExpress/+}$ mice were induced at E9.5 to label HSC progenitors and analyzed at the age of 17-weeks.

(C) 60,000 Lin^{-} Sca-1^{+} Kit^{+} (LSK) cells from mouse bone marrow were sorted into one tube and then split into three equal number aliquots for scRNAseq. Different conditions were applied to the aliquots to compare barcode capture efficiency.

(D) The frequency of useful cells among 10x captured cells was improved by *PolyloxExpress*-specific oligo spike in.

(E) The number of *PolyloxExpress* barcodes recovered from nested PCR was improved by *PolyloxExpress*-specific oligo spike in. (Bioinformatic calculation by Xi Wang)

In all three conditions *PolyloxExpress* barcodes could be amplified, indicating that the established nested PCR protocol is compatible with the target enriched cDNA libraries (data not shown). Interestingly, the spike in of a *PolyloxExpress*-specific oligo increased the proportion of cells for which transcriptome and barcode could be obtained by more than two-fold (Figure 11D). The absolute number of detected *PolyloxExpress* barcode number was also higher in the modified protocols (Figure 11E). Of note, increasing the amount of the spike in oligo only slightly enhanced the proportion of recovered ‘‘useful’’ cells or absolute number of recovered barcodes (Figure 11D, 11E). Hence, this target enrichment approach significantly improved *PolyloxExpress* barcode capture efficiency (Details of the modified protocol can be found in the ‘‘Materials and Methodology’’ section).

Altogether, I have demonstrated that the *PolyloxExpress* DNA barcodes are expressed and transcribed into mRNA, and I confirmed that DNA and RNA barcode are identical within a single cell. In combination with high-throughput scRNAseq, I successfully set up a customized protocol for full-length *PolyloxExpress* barcode retrieval in thousands of single cells. Therefore, I concluded that the mouse model and experimental protocols of the *PolyloxExpress* system were successfully set up.

3.2 Clonal dynamic analysis of HSC clones

3.2.1 HSC fate during embryonic hematopoiesis

Using single-HSC transplantation or high-resolution barcoding, HSC fate heterogeneity has been well documented in adult mice (Dykstra et al. 2007; Muller-Sieburg et al. 2004; Pei et al. 2017). However, given that HSCs expand in the fetal liver during embryogenesis and colonize the bone marrow around birth (Mikkola and Orkin 2006), it remains elusive whether HSC fate is determined embryonically or postnatally. To answer this question, I decided to label HSC progenitors in utero, observe the barcode distribution of different blood and immune cell lineages shortly before HSC bone marrow colonization (Figure 12), and compare it to the fate patterns known from adult mice (Pei et al. 2017).

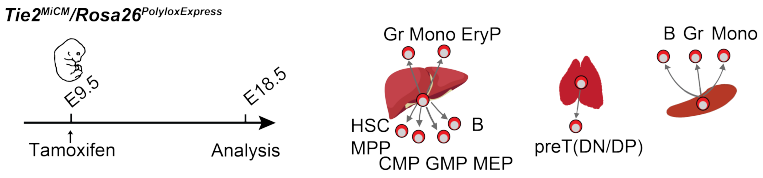


Figure 12. Experimental design for the embryonic fate mapping experiments. HSPCs and the indicated mature cells were harvested from fetal liver, thymus and spleen for DNA barcode retrieval. Gr, granulocytes; EryP, erythroid-progenitors; Mono, monocytes; B, B cell; HSC, hematopoietic stem cell; MPP, multipotent progenitor; CMP, common myeloid progenitor; GMP, granulocyte-monocyte progenitor; MEP, megakaryocyte-erythrocyte progenitor; DN, double negative (CD4-CD8-) preT cells; DP, double positive (CD4+CD8+) preT cells.

Tie2$^{MeriCreMer/+}$*Rosa26*$^{PolyloxExpress/+}$ mice were treated with tamoxifen in utero at E9.5. Nine days later, after phenotypic HSCs (Lin$^-$ Sca-1$^+$ Kit$^+$ CD150$^+$ CD48$^-$) emerged and migrated to the fetal liver, I retrieved DNA barcodes from fetal liver HSCs, hematopoietic progenitors and mature cells (Figure 12). A total of four embryos from two pregnant mice were analyzed. Of note, the accuracy of HSC fate annotation highly depends on the sampling depth of their products in downstream compartments. Previous work has shown that two sample repeats (30,000 cells/sample) of each mature cell type is sufficient to

represent the majority of barcodes in the corresponding lineage (Pei et al. 2017). Therefore, two sample repeats of each mature cell type representing the myeloid-erythroid lineage (Granulocyte, Gr; Monocyte, Mono; Erythroid progenitor, EryP) and lymphoid lineage (B cell, B) were taken to determine the sampling depth in the current experiment (Figure 12). Indeed, Spearman rank correlation analysis revealed good overlap between sample repeats (R>0.6 is considered as good overlap), indicating that the sampled barcodes faithfully represent the total barcode distribution of the corresponding lineages (Figure 13A). As for the progenitors, sample repeats were also taken for some myeloid (CMP, GMP, MEP) and lymphoid progenitors (CD4+CD8+/double positive pre-T, DP), but not for the ST-HSC, MPP and CD4-CD8-/double negative pre-T cells (DN) due to their low abundance.

To reveal clonal HSC fate, rare barcodes were filtered based on their probability of generation. *Pgen* cut off was set to $5x10^{-4}$ to ensure the majority of barcodes were induced in a single cell (Pei et al. 2019). Of note, non-rare HSC barcodes that showed inactive or myeloid-erythroid-restricted fate are more likely to be clonal origin and were also included for analysis. Interestingly, barcode comparison between HSCs and their progeny revealed that differentiation-inactive, myeloid-erythroid-restricted and multilineage fates are reproducibly present in HSCs shortly before birth (Figure 13B). These fate patterns are consistent with HSC fates identified postnatally and in adult mice (Pei et al. 2017), indicating that HSC fate is determined embryonically, and maintained in the fetus prior to bone marrow colonization.

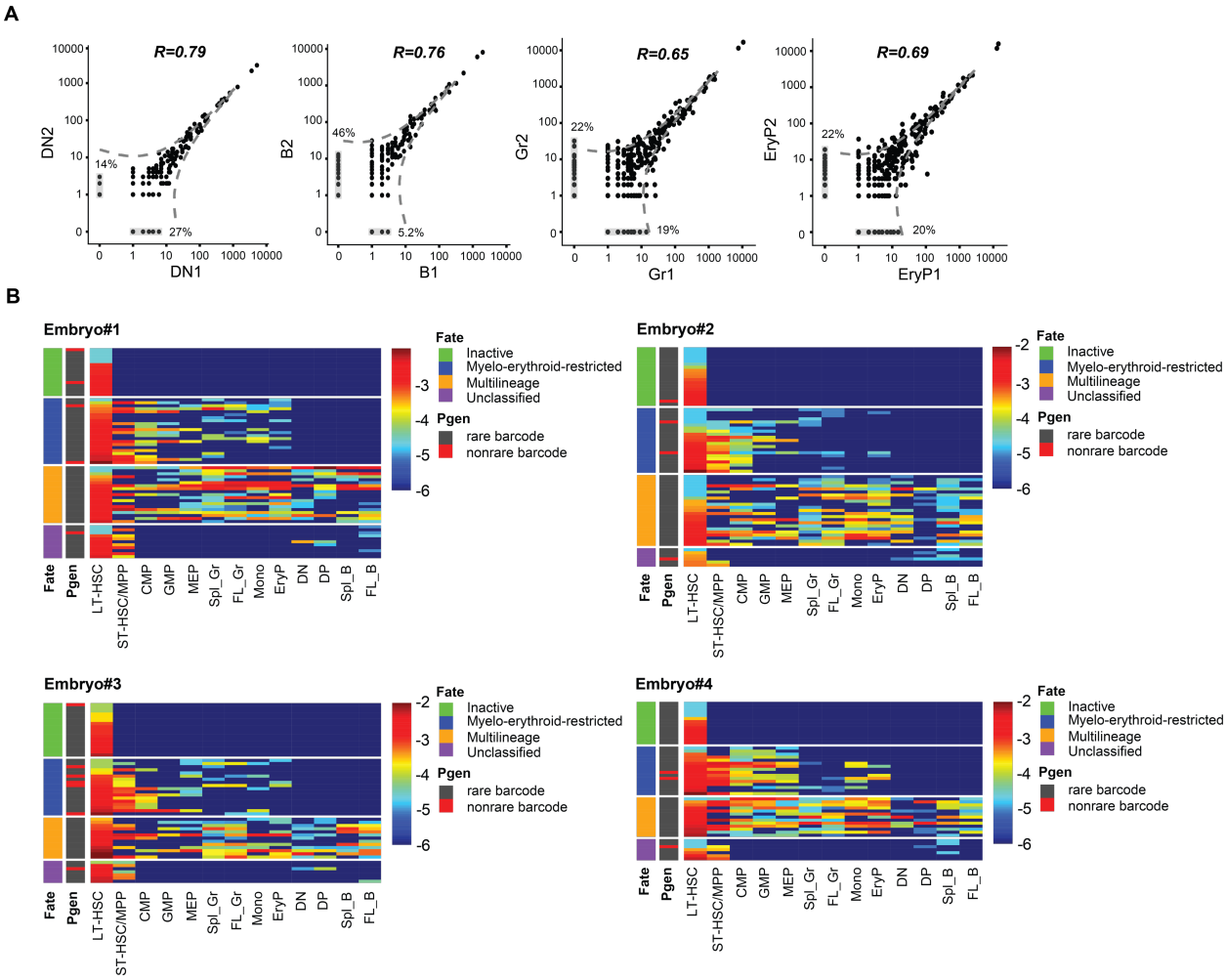


Figure 13. *PolyloxExpress* barcoding reveals HSC fates at embryonic stage.

(A) Scatter plot showing barcode correlation between two sample repeats of the indicated cell type based on barcode read counts obtained by Pacbio sequencing. Each dot corresponds to a *PolyloxExpress* barcode. Unique barcodes found in one but not the other sample position along the axes and were highlighted with grey background. The dashed line represents the boundary for 95% confidence. Spearman's rank correlation coefficients (R) were calculated and shown on top of each panel.

(B) HSC barcode propagation heatmap of E18.5 analyzed embryos. In the heatmaps, each row represents a unique barcode and each column represents a cell type. On the left part of each heatmap, HSCs were grouped into differentiation-inactive, myeloid-erythroid-restricted and multilineage clones (highlighted by different colors in the fate column) based on the presence of their barcodes in downstream progenies and indicated by colored bars. A small fraction of HSC clones with unclassified fate was also observed. Barcodes of sample repeats were merged based on the cell type and then displayed as individual columns on the heatmap. Barcode frequencies were represented by color scale. In addition to rare barcodes (*Pgen*<$5x10^{-4}$, black color in *Pgen* column), non-rare barcodes (*Pgen*>$5x10^{-4}$, red color in *Pgen* column)

with an inactive or restricted fate were also shown on the heatmaps. Gr, granulocyte; EryP, erythroid-progenitor; Mono, monocyte; B, B cell; HSC, hematopoietic stem cell; MPP, multipotent progenitor; CMP, common myeloid progenitor; GMP, granulocyte-monocyte progenitor; MEP, megakaryocyte-erythrocyte progenitor; DN, double negative (CD4-CD8-) preT cell; DP, double positive (CD4+CD8+) preT cell; FL, fetal liver; Spl, spleen. (Bioinformatic calculation by Xi Wang)

3.2.2 Time-resolved HSC fate at steady state

Since similar HSC fate patterns are observed in both embryonic and postnatal stages, it is conceivable that HSC fate remains consistent over time. However, the current lineage tracing experiments are mostly end-point analysis, reflecting only a snapshot of the dynamic developmental process (Pei et al. 2017; Sun et al. 2014). HSC fate patterns found across development might be maintained only at the population scale and, therefore, current data could not preclude the possibility of fate conversion between individual HSC clones.

To study whether embryonically established HSC fates are propagated throughout adult life and remain stable over time, I labeled HSC progenitors in utero and repeatedly took peripheral blood (PB) before comprehensive end-point analysis. Comparison of these samples reveals whether HSC fates are stable over time. *Tie2*$^{MeriCreMer/+}$*Rosa26*$^{PolyloxExpress/+}$ mice were treated with tamoxifen at E9.5 to label HSC progenitors (Figure 14A, 14B). During the postnatal stage, I repeatedly took blood samples over a period of 4-7 months from barcoded mice, either with a young or an old age (Figure 14A, 14B). At each bleeding timepoint, barcodes were retrieved from FACS-purified mature lymphocytes (T, B cells) as well as myeloid cells (Granulocytes, Monocytes) (Figure 14A, 14B). At the end point of analysis, mice were sacrificed to harvest barcodes of HSCs and their progeny from bone marrow, spleen, blood and thymus (Figure 14A, 14B). Sampling of progenitors and mature cells was performed as described in section 3.2.1.

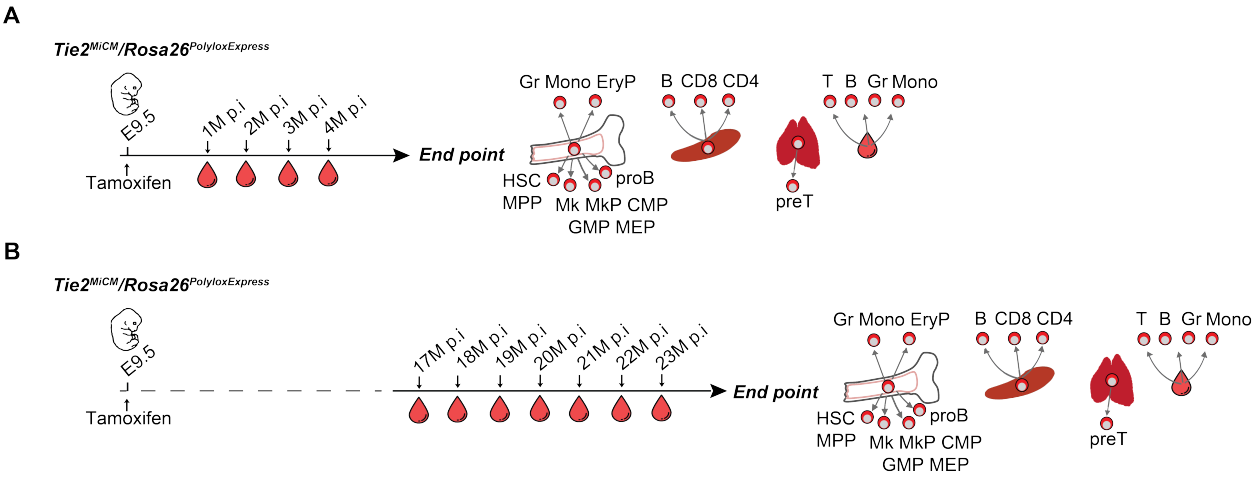


Figure 14. Experimental design for the clonal dynamic experiments.

(A, B) *Tie2*$^{MeriCreMer/+}$*Rosa26*$^{PolyloxExpress/+}$ mice were induced at E9.5 and analyzed at adult. Peripheral blood was taken monthly at the indicated time points from young **(A)** or old mice **(B)**. At the endpoint, HSPCs and the indicated mature cells were harvested from bone marrow, thymus and spleen for DNA barcode retrieval. 1M, 1 month; p.i, post induction; Gr, granulocyte; EryP, erythroid-progenitor; Mono, monocyte; B, B cell; T, T cell; HSC, hematopoietic stem cell; MPP, multipotent progenitor; CMP, common myeloid progenitor; GMP, granulocyte-monocyte progenitor; MEP, megakaryocyte-erythrocyte progenitor; Mk, megakaryocyte; MkP, megakaryocyte progenitor; preT, preT cell; proB, proB cell.

I first focused on the mature cell barcodes (*Pgen*<5x10^{-4}) recovered from blood at different time points. Using t-Distributed Stochastic Neighbor Embedding (t-SNE) for dimensionality reduction and visualization, Xi and I found that mature cell barcodes of the same lineage clustered closely with each other, regardless of the timing of barcode retrieval (Figure 15A). This indicates that the majority of barcodes detected in blood was reliably retrieved at each time point, and hence good sampling depth was achieved during the kinetic experiment. This is supported by an additional analysis showing that at least 50% of HSC barcodes were recovered in all lineages at each bleeding time point (Figure 15B). Notably, in the same t-SNE analysis, mature cell barcodes from myeloid and lymphoid lineages were found in separate clusters, suggesting distinct origins of mature myeloid and lymphoid cells (Figure 15A).

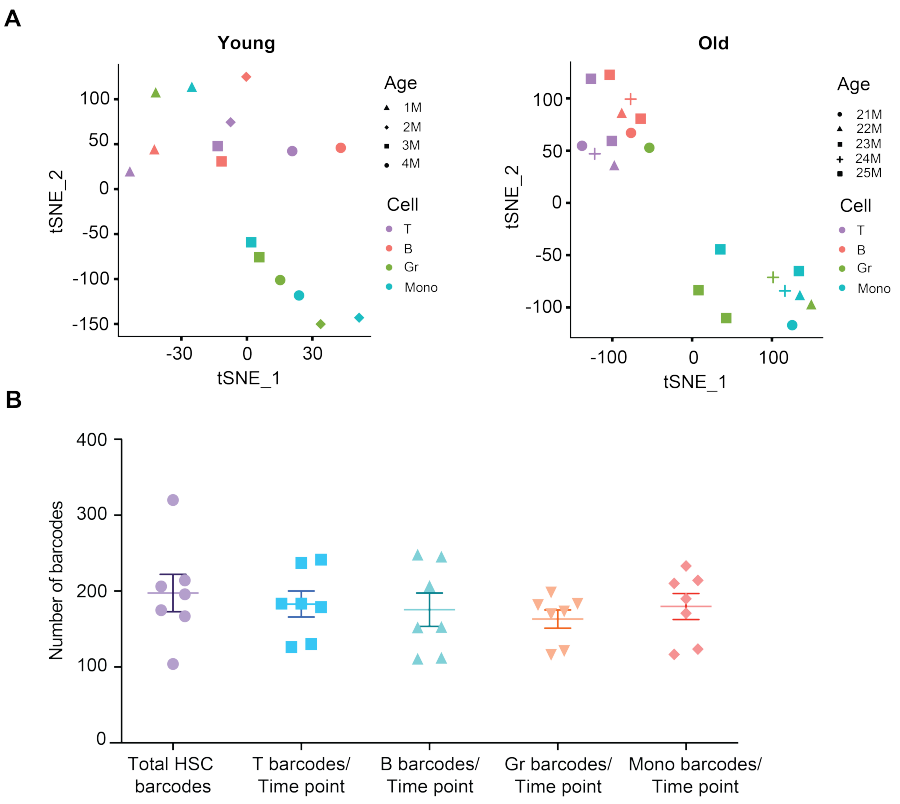


Figure 15. Clonal dynamic of peripheral blood lineages.

(A) t-distributed stochastic neighbor embedding (t-SNE) analysis of blood samples, data shown for one young mouse and one old mouse. 1M, 1 month.

(B) Average number of barcodes detected for each blood timepoint compared to total hematopoietic stem cell (HSC) barcodes detected at endpoint analysis (mean ± SEM). Data shown were derived from three young mice and four old mice. T, T cell; B, B cell; Gr, Granulocyte; Mono, Monocyte. (Bioinformatic calculation by Xi Wang)

Next, I asked whether all HSC clones contributed equally to the downstream progeny. Therefore, the number of blood barcodes (*Pgen* filtered) that could be found at more than one bleeding time point for each mature cell type was quantified. Interestingly, the analysis showed that around 40% of clonal barcodes were detected at least two times and around 20% of clonal barcodes were detected at least three times in young mice (Figure 16A). The frequency was higher in old mice, with around 50% of clonal barcodes been detected at least two times and around 40% of clonal barcodes been detected at least three times (Figure 16B). This data indicates that embryonically derived HSC clones may not contribute to hematopoiesis equally. Some clones contributed more to the

system and their products could be detected repetitively in blood, while other clones contributed less and hence less products were sampled.

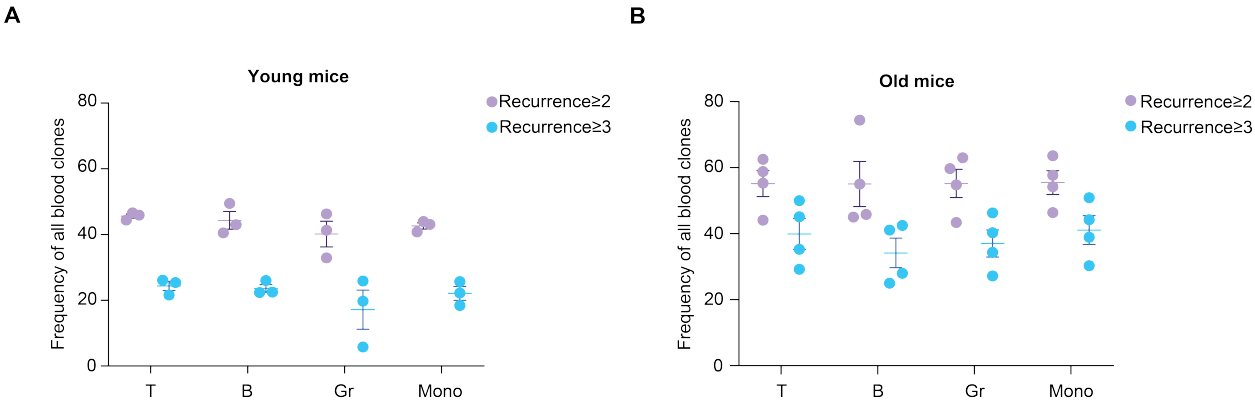


Figure 16. Frequency of recurrent peripheral blood clones.

(A) The frequency of recurrent blood clones in all periphery blood (PB) clones detected in young mice (n=3, mean ± SEM). T, T cell; B, B cell; Gr, Granulocyte; Mono, Monocyte.

(B) The frequency of recurrent blood clones in all PB clones detected in old mice (n=4, mean ± SEM).

To test whether differences in HSC clone size could explain the varied hematopoietic contribution, barcode sequencing read frequency (*Pgen* filtered) was used to infer clone size distribution. The analysis showed that recurrent clones of mature blood cells (found at least three times in blood) had larger clones sizes compared to non-recurrent clones in both young (Figure 17A) and old mice (Figure 17B). Notably, the clone size difference was not restricted to mature blood cells, but could also be found in the corresponding progenitors that gave rise to them, including HSCs (Figure 17C, 17D). Therefore, this data supports the previous hypothesis that the extent of HSC contribution to hematopoiesis positively correlates with the size of individual HSC clones.

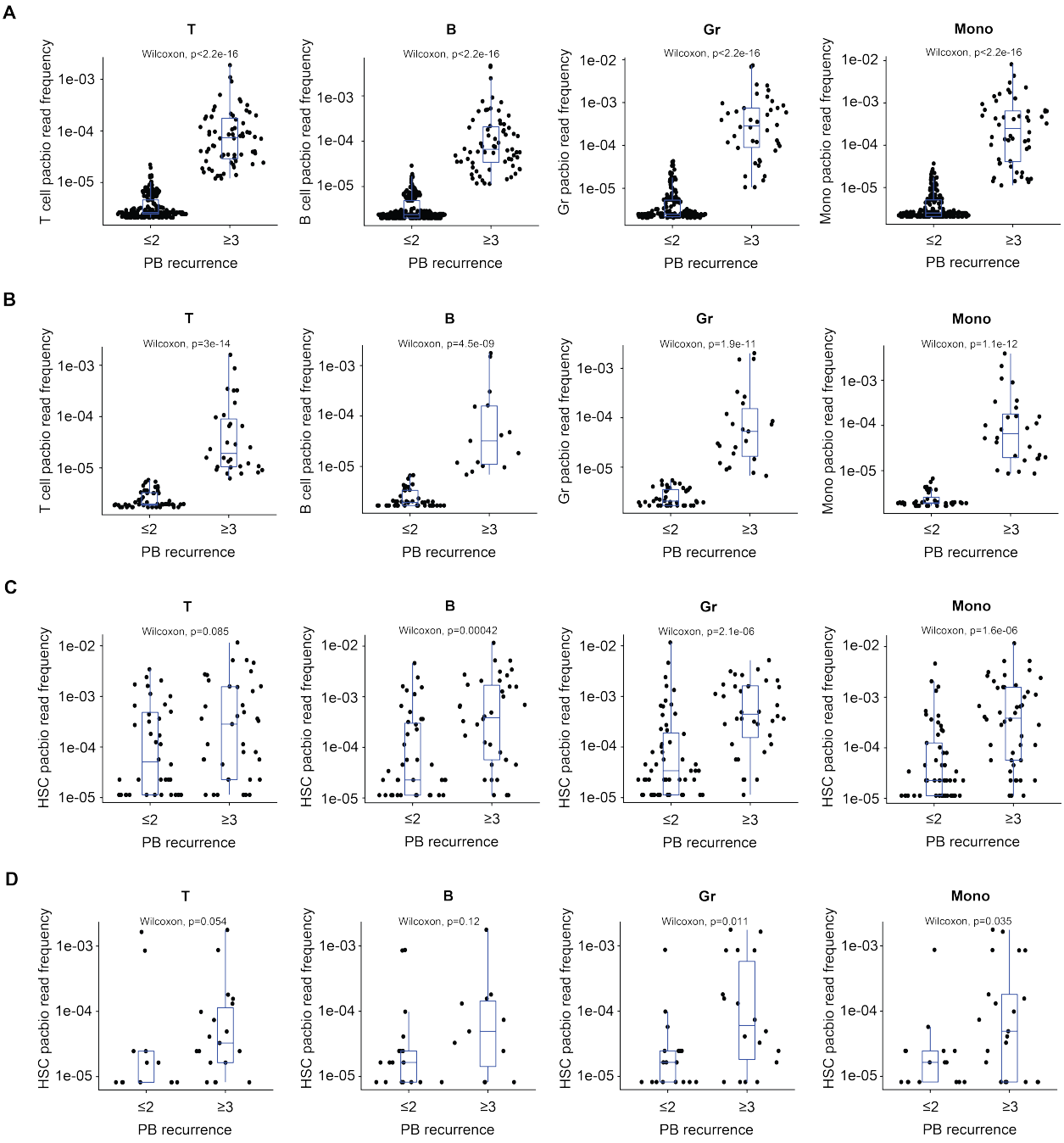


Figure 17. Read count analysis of recurrent peripheral blood clones.

(A) Pacbio sequencing read frequency of recurrent vs. non-recurrent periphery blood (PB) clones of each mature cell lineage. Data shown for one representative young mouse. T, T cell; B, B cell; Gr, Granulocyte; Mono, Monocyte. Each dot corresponds to a *Pgen* filtered *PolyloxExpress* barcode.

(B) Pacbio sequencing read frequency of recurrent vs. non-recurrent clones of each mature cell lineage. Data shown for one representative old mouse.

(C) Pacbio sequencing read frequency of HSC clones that share barcodes with recurrent or non-recurrent PB clones. Data shown for one representative young mouse.
(D) Pacbio sequencing read frequency of HSC clones that share barcodes with recurrent or non-recurrent PB clones. Data shown for one representative old mouse.
p values were defined by 2-sided Wilcoxon-Mann-Whitney rank-sum test. Middle line in each box plot represents median value, upper and lower lines in each box plot represent the first and third quartile. Vertical line in the middle represents standard deviation. (Bioinformatic calculation by Xi Wang)

Finally, I asked whether fate-defined HSC clones identified at end point analysis had a coherent fate usage across time. To this end, HSC barcodes (*Pgen* filtered) were compared with their downstream progeny sampled at the end point (left half of the heatmap in Figure 18A, 18B), as well as blood barcodes retrieved from multiple time points in the past (right half of the heatmap in Figure 18A, 18B). Interestingly, I found that the fate of distinct HSC clones remained largely stable over time in both young (Figure 18A) and old mice (Figure 18B). In order to determine fate coherency quantitatively, I calculated the number of inactive and myeloid-erythroid-restricted HSC clones with and without fate violation during the course of the kinetics for each individual mouse (Figure 18C, 18D). Quantifying the proportion of inactive clones with no violation within the total inactive HSC clones revealed that, on average 81.1% of the inactive HSC clones from young mice and 80.5% of the inactive HSC clones from old mice did not produce progenitor or mature cells in the course of the kinetics (Figure 18C). Similarly, on average 79.8% of the HSC clones identified as myeloid-erythroid-restricted from young mice and 72.5% of such clones from old mice never produced any lymphocytes (Figure 18D). Given the long life-span of lymphoid cells, it is conceivable that these fate-defined HSC clones maintain their fate propagation for an even longer period of time.

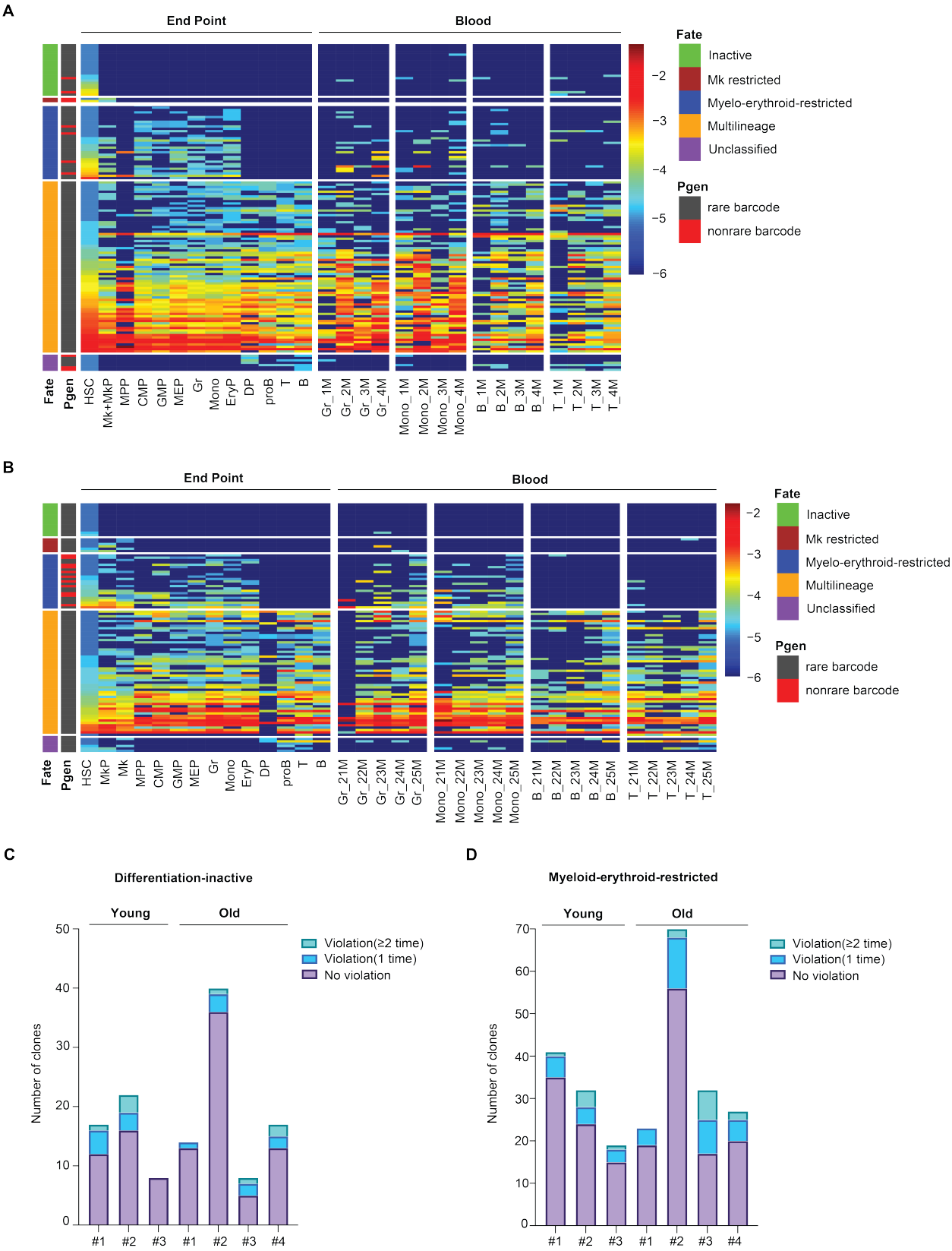

A
End Point
Blood
Fate
Inactive
Mk restricted
Myelo-erythroid-restricted
Multilineage
Unclassified
Pgen
rare barcode
nonrare barcode
−2
−3
−4
−5
−6
Fate
Pgen
HSC
Mk+MkP
MPP
CMP
GMP
MEP
Gr
Mono
EryP
DP
proB
T
B
Gr_1M
Gr_2M
Gr_3M
Gr_4M
Mono_1M
Mono_2M
Mono_3M
Mono_4M
B_1M
B_2M
B_3M
B_4M
T_1M
T_2M
T_3M
T_4M
B
End Point
Blood
Fate
Inactive
Mk restricted
Myelo-erythroid-restricted
Multilineage
Unclassified
Pgen
rare barcode
nonrare barcode
−2
−3
−4
−5
−6
Fate
Pgen
HSC
MkP
Mk
MPP
CMP
GMP
MEP
Gr
Mono
EryP
DP
proB
T
B
Gr_21M
Gr_22M
Gr_23M
Gr_24M
Gr_25M
Mono_21M
Mono_22M
Mono_23M
Mono_24M
Mono_25M
B_21M
B_22M
B_23M
B_24M
B_25M
T_21M
T_22M
T_23M
T_24M
T_25M
C
Differentiation-inactive
Young
Old
Violation(≥2 time)
Violation(1 time)
No violation
Number of clones
0
10
20
30
40
50
#1
#2
#3
#1
#2
#3
#4
D
Myeloid-erythroid-restricted
Young
Old
Violation(≥2 time)
Violation(1 time)
No violation
Number of clones
0
10
20
30
40
50
60
70
#1
#2
#3
#1
#2
#3
#4

Figure 18. Clonal dynamic analysis of HSC clones at steady state.

(A, B) Hematopoietic stem cell (HSC) barcode propagation heatmap of one representative young mouse **(A)** and one representative old mouse **(B)**. Each row represents a unique barcode. On the left part of the heatmap, HSCs were grouped by their fates and indicated by colored bars. A small fraction of HSC clones with unclassified fate was also observed. On the right part of the heatmap, the appearance of HSC barcodes in mature cells from peripheral blood at different time points was shown. Barcodes of sample repeats were merged based on the cell type and then displayed as individual columns on the heatmap. Barcode frequencies were represented by color scale. In addition to rare barcodes (*Pgen*<$5x10^{-4}$, black color in *Pgen* column), non-rare barcodes (*Pgen*>$5x10^{-4}$, red color in *Pgen* column) with an inactive or restricted fate were also shown on the heatmap. 1M, 1 month. Mk, megakaryocyte; MkP, megakaryocyte progenitor; MPP, multipotent progenitor; CMP, common myeloid progenitor; GMP, granulocyte-monocyte progenitor; MEP, megakaryocyte-erythrocyte progenitor; Gr, granulocyte; Mono, monocyte; EryP, erythroid-progenitor; DP, double positive (CD4+CD8+) preT cell; proB, proB cell; T, T cell; B, B cell.

(C) Total number of HSC clones defined as inactive based on the endpoint analysis and distinction of the number of clones not, once or twice violating this classification by appearance of their barcodes in any mature blood cells. (n=3 mice for the young group and n=4 mice for the old group)

(D) As in **(C)** but for myeloid-erythroid-restricted HSC clones counting their violations by appearance of their barcodes in any blood lymphocytes (T or B cells). (n=3 mice for the young group and n=4 mice for the old group) (Bioinformatic calculation by Xi Wang)

3.2.3 Time-resolved HSC fate upon transplantation

Having shown that HSC fates are largely stable during native steady-state hematopoiesis, I next explored whether coherent fates are also maintained after transplantation. Therefore, I first induced two *Tie2*$^{MeriCreMer/+}$*Rosa26*$^{PolyloxExpress/+}$ donor mice at E9.5 and determined clonal barcode distribution across the different lineages by four monthly blood analyses (Figure 19A). At the four-month time point, HSCs were harvested and the sample was split. One half was used for barcode recovery to determine native HSC fates, and the other half was transplanted into two genetically conditioned recipient mice (*Rag2*$^{-/-}$*gc*$^{-/-}$*Kit*$^{W/Wv}$, Waskow et al. 2009) to study HSC fates after transplantation (Figure 19A). Over another four-month period, recipients were bled four times and then sacrificed for final analysis (Figure 19A). Sampling of progenitors and mature cells from the donors and recipients was performed as described in section 3.2.1.

To compare native and transplanted HSC fate, HSC barcodes (*Pgen* filtered) and their progeny were first defined for the donor mice (left part of Figure 19B, 19C). Next, donor HSC barcodes that were also found in any samples of their recipients were profiled (middle and right part of Figure 19B, 19C). In the first analysis, I found that not all donor HSC clones were detected in the recipients (Figure 19B, 19C), suggesting that either a fraction of HSC clones was lost during the transplantation process, or that only certain donor HSC clones engrafted the recipients. Interestingly, when focusing on donor HSC clones that were also identified in recipients, I found that most donor HSCs maintained similar fate usage upon transplantation (marked by green asterisk in Figure 19B, 19C). However, inconsistent fate usage was also observed for some clones (marked by black asterisk in Figure 19B, 19C). To precisely quantify fate coherency across transplantation, I calculated the number of myeloid-erythroid-restricted and multilineage HSC clones (defined by native fate) with and without fate violation (green and black asterisk) during the course of transplantation (Figure 19D, 19E). Quantifying the proportion of non-violating myeloid-erythroid-restricted clones within the total myeloid-erythroid-restricted clones revealed that, on average 77.7% of myeloid-erythroid-restricted HSCs (two experiments, n=4 recipients) maintained a coherent fate usage (Figure 19D). Likewise, 76.5% of multilineage HSC clones (two experiments, n=4 recipients) detected in recipients remained multilineage after transplantation (Figure 19E). Only few barcodes could be retrieved in the recipients from inactive HSCs, suggesting that either these cells were also not productive after transplantation or that they failed to engraft and expand in the recipients (Figure 19B, 19C). In conclusion, most fate-defined HSC clones retained their fate potential throughout transplantation, which implies that intrinsic, but not extrinsic, properties are driving HSC fate determination. However, fate conversion into lymphoid-biased fate was observed for 3 myeloid-erythroid-restricted and multilineage clones (Figure 19B, 19C) pointing at potential differences of HSC fates observed in situ (through lineage tracing) or after transplantation.

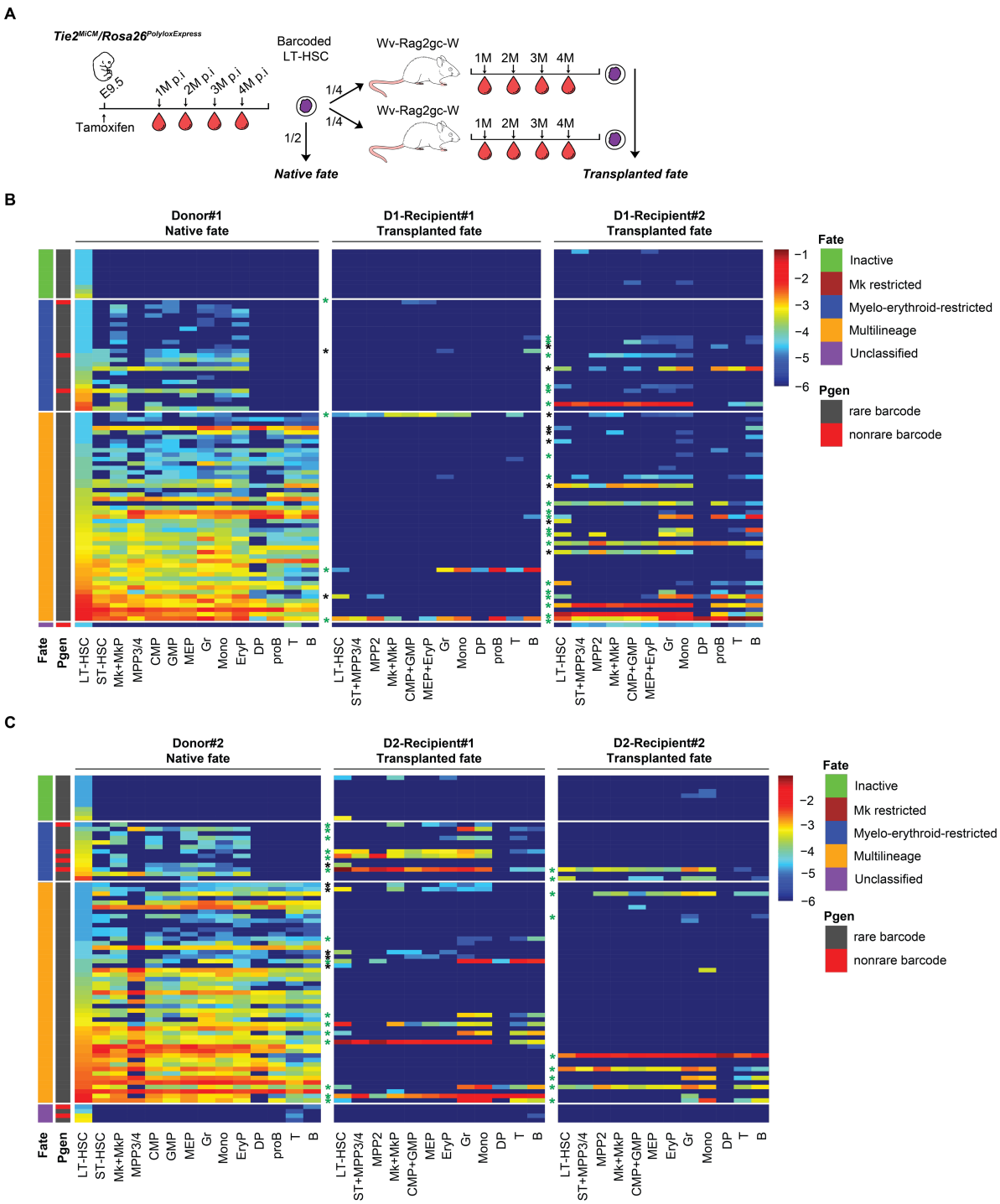

A
Tie2MiCM/Rosa26PolyloxExpress
E9.5
1M p.i
2M p.i
3M p.i
4M p.i
Tamoxifen
Barcoded LT-HSC
1/2
1/4
1/4
Native fate
Wv-Rag2gc-W
Wv-Rag2gc-W
1M 2M 3M 4M
1M 2M 3M 4M
Transplanted fate
B
Donor#1 Native fate
D1-Recipient#1 Transplanted fate
D1-Recipient#2 Transplanted fate
Fate
Inactive
Mk restricted
Myelo-erythroid-restricted
Multilineage
Unclassified
Pgen
rare barcode
nonrare barcode
Fate Pgen LT-HSC ST-HSC Mk+MkP MPP3/4 CMP GMP MEP Gr Mono EryP DP proB T B
LT-HSC ST+MPP3/4 MPP2 Mk+MkP CMP+GMP MEP+EryP Gr Mono DP proB T B
LT-HSC ST+MPP3/4 MPP2 Mk+MkP CMP+GMP MEP+EryP Gr Mono DP proB T B
C
Donor#2 Native fate
D2-Recipient#1 Transplanted fate
D2-Recipient#2 Transplanted fate
Fate
Inactive
Mk restricted
Myelo-erythroid-restricted
Multilineage
Unclassified
Pgen
rare barcode
nonrare barcode
Fate Pgen LT-HSC ST-HSC Mk+MkP MPP3/4 CMP GMP MEP Gr Mono EryP DP proB T B
LT-HSC ST+MPP3/4 MPP2 Mk+MkP CMP+GMP MEP EryP Gr Mono DP T B
LT-HSC ST+MPP3/4 MPP2 Mk+MkP CMP+GMP MEP EryP Gr Mono DP T B
D
Donor#1
Donor#2
Myelo-erythroid-restricted/biased
Other fates
Number of clones
R1 R2 R1 R2
E
Donor#1
Donor#2
Multilineage
Other fates
Number of clones
R1 R2 R1 R2

Figure 19. Clonal dynamic analysis of HSC clones upon transplantation.

(A) Experimental design for the hematopoietic stem cell (HSC) transplantation experiments. 1M, 1 month; p.i, post induction; LT-HSC, long-term HSC.

(B, C) HSC barcodes propagation heatmap of donor (left) and their two recipients (middle and right). Each row represents a unique barcode. HSC clones derived from donors were grouped by their fates and indicated by colored bar. Donor HSC clones that maintained similar fate usage in recipients were marked by green asterisk (*) and clones that converted their fates were marked by black asterisk (*). Only donor HSC clones contributed to at least two cell types in recipients were marked by asterisk and quantified. Barcodes of sample repeats were merged based on the cell type and then displayed as individual columns on the heatmap. Barcode frequencies were represented by color scale. In addition to rare barcodes (*Pgen*<$5x10^{-4}$, black color in *Pgen* column), non-rare barcodes (*Pgen*>$5x10^{-4}$, red color in *Pgen* column) with an inactive or restricted fate were also shown on the heatmap. D1/D2, donor1/donor2. ST-HSC, short-term HSC; Mk, megakaryocyte; MkP, megakaryocyte progenitor; MPP, multipotent progenitor; CMP, common myeloid progenitor; GMP, granulocyte-monocyte progenitor; MEP, megakaryocyte-erythrocyte progenitor; Gr, granulocyte; Mono, monocyte; EryP, erythroid-progenitor; DP, double positive (CD4+CD8+) preT cell; proB, proB cell; T, T cell; B, B cell.

(D, E) Quantification of donor HSC clones that maintained (green bar) or converted (black bar) their fate after transplantation for myeloid-erythroid-restricted **(D)** and multilineage HSCs **(E)**. R1/R2, recipient#1 and recipient#2. (Bioinformatic calculation by Xi Wang)

In summary, using high-resolution barcoding mice and repetitive sampling, I demonstrated that HSC fates are set embryonically and remain remarkably stable from birth into adulthood. Most HSC clones even maintained a stable use of their fate potential after transplantation, but about 20-25% of the HSC clones displayed an altered fate usage in the recipients, underscoring the differences of fate analysis in situ or via transplantation.

3.3 Combining transcriptome and fate analysis of HSC clones

3.3.1 Dissecting HSC fates *in vivo* by *PolyloxExpress* RNA barcodes

The fact that HSCs have shown fate stability in the clonal dynamic experiments argues against extrinsic fate regulation or stochastic fate usage, but implies intrinsic HSC fate determination. To better understand the mechanisms behind HSC fate regulation, I decided to investigate the gene expression programs of HSCs related to their fate using *Tie2*$^{MeriCreMer/+}$*Rosa26*$^{PolyloxExpress/+}$ mice.

Four mice were treated with tamoxifen at E9.5 to label hematopoietic progenitor cells and sacrificed at different time points of adulthood (Figure 20A). Bulk sampling of progenitors and mature cells for determination of DNA barcodes and lineage specification was performed as described in section 3.2.1. Additionally, all HSCs and a fraction of myeloid progenitors (CMP, GMP, MEP, 20,000 cells/population) and lymphoid progenitors (CLP, proB, preT, 20,000 cells/population) were analyzed by scRNAseq (10x genomics) to simultaneously recover RNA barcodes and transcriptome (Figure 20A). DNA and RNA barcode PCR products were sequenced on the Pacbio platform, while transcriptome libraries were sequenced on the Illumina platform.

The first analysis was determined to show whether RNA barcoding recapitulates HSC fate heterogeneity like the previous DNA barcoding experiments (Figure 18). Therefore, HSC-derived RNA barcodes (*Pgen* filtered) were compared with DNA barcodes from downstream progeny to determine HSC fates. As shown in Figure 20B, the *PolyloxExpress* RNA barcoding system indeed robustly revealed distinct HSC fates in all four mice analyzed, including differentiation-inactive, myeloid-erythroid-restricted and multilineage HSCs. In addition, to refine the spectrum of fate-defined HSCs in the subsequent transcriptome analysis, clones with strong myeloid bias or vague fate output were also defined (Figure 20B).

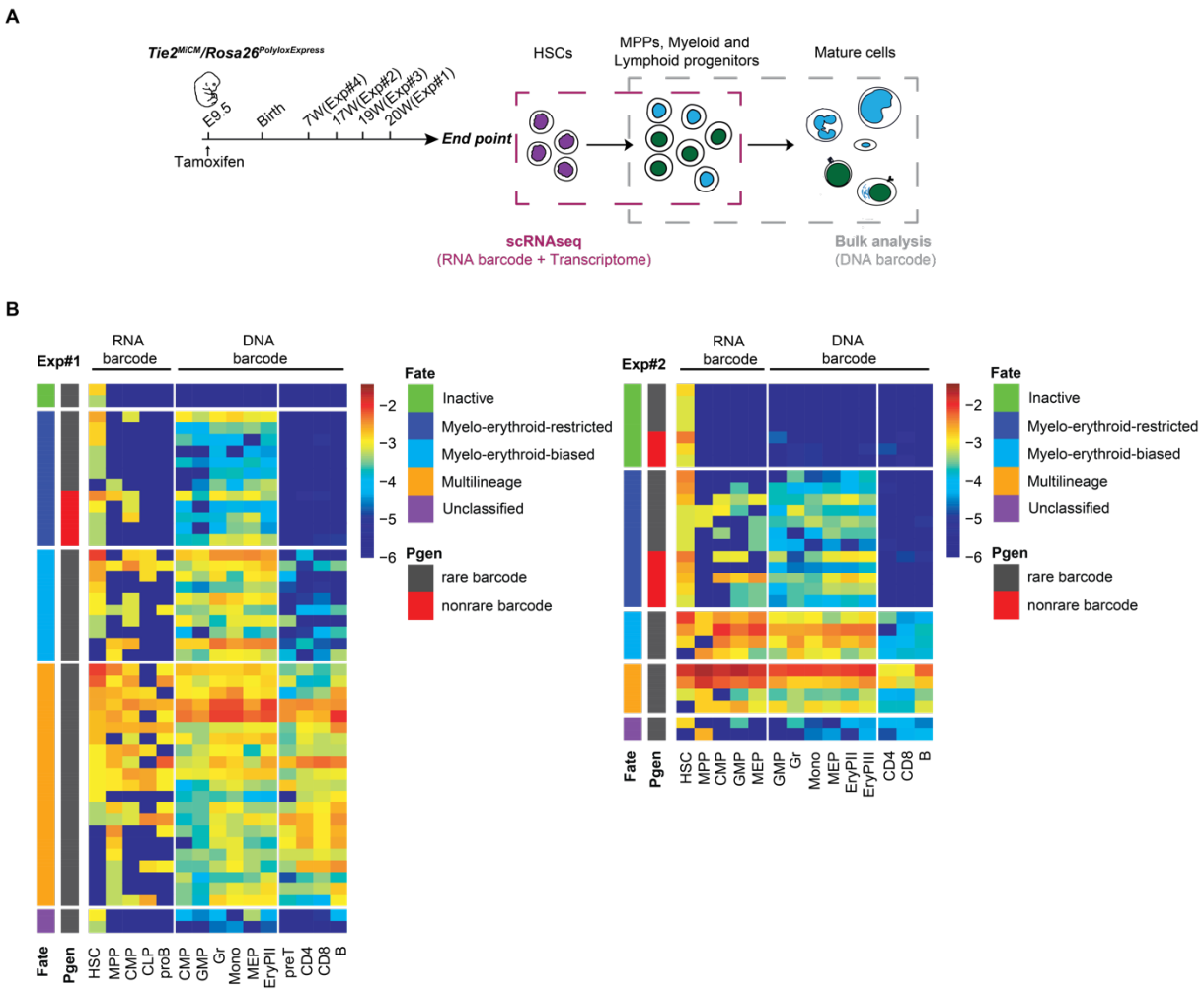


Figure 20. Dissecting HSC fates *in vivo* by *PolyloxExpress* RNA barcoding.

(A) Experimental design of RNA barcoding experiments. Red box indicates samples that were used for scRNAseq to recover RNA barcodes and transcriptome. Grey box indicates samples that were used for bulk analysis to recover DNA barcodes. Exp, experiment; scRNAseq, single-cell RNA sequencing; HSC, hematopoietic stem cell; MPP, multipotent progenitor.

(B) HSC and MPP barcode propagation heatmap. RNA barcodes retrieved from HSC and progenitors were listed on the left part of each heatmap, DNA barcodes retrieved from bulk samples were listed on the right part of each heatmap. Each row represents a unique barcode. HSC clones were grouped by their fates and indicated by colored bar. Barcodes of sample repeats were merged based on the cell type and then displayed as individual column on the heatmap. Barcode frequencies were represented by color scale. In addition to rare barcodes (*Pgen*<$5x10^{-4}$, black color in *Pgen* column), non-rare barcodes (*Pgen*>$5x10^{-4}$, red color in *Pgen* column) with an inactive or restricted fate were also shown on the heatmap. Data shown for Experiment #1-#2. *Pgen*, probability of generation; CMP, common myeloid progenitor; GMP, granulocyte-monocyte progenitor; MEP, megakaryocyte-erythrocyte progenitor; CLP, common lymphoid progenitor; Gr,

granulocyte; Mono, monocyte; EryP, erythroid progenitor; CD4, CD4+ T cell; CD8, CD8+ T cell; B, B cell. (Bioinformatic calculation by Xi Wang; Figure adopted and modified from Pei et al. 2020)

3.3.2 Principal component analysis of fate-defined HSC clones reveals differentially expressed genes related to fate

To dissect underlying mechanisms governing fate heterogeneity, all HSCs with gene expression information derived from experiment #1-3 (experiment #4 was omitted from this analysis because of the low number of informative barcodes) were analyzed using principal component analysis (PCA, n= 22,565 cells). Transcriptome clustering divided HSCs into three clusters along the PC1 but not PC2 axis (Figure 21A). With the RNA barcode information obtained in addition to the transcriptome, I asked whether HSCs also cluster on the PCA plot according to their fates. Therefore, HSCs with both RNA barcode and transcriptome information were highlighted on the PCA plot, with different colors representing different fates, to reveal their positions (Figure 21B). Unexpectedly, there was no cluster separation along the PC1 axis and HSCs with different fates had similar PC1 scores (Figure 21C). However, on the PC2 axis, HSCs were scored differently depending on their fate, with differentiation-inactive HSCs on average receiving the highest and multilineage HSCs receiving the lowest PC2 scores (Figure 21D), although the groups of HSCs representing different fates were still overlapping. Therefore, single-cell transcriptome-based PCA analysis alone is not sufficient to predict cell fate.

Next, I asked what were the differentially expressed genes (DEGs) that contributed to PC1 and PC2 separation. Analysis revealed that PC1 mainly enriched for translation-related genes that were unrelated to fate (Figure 21E). PC2, however, enriched genes associated with HSC quiescence and self-renewal (e.g., Mllt3) (Calvanese et al. 2019) on the positive axis and genes correlated with multilineage output (e.g., CD34) (Krause et al. 1994; Osawa et al. 1996; Yang et al. 2005) on the negative axis (Figure 21F). Interestingly, vWF, a megakaryocyte associated gene, was found to enrich in differentiation-inactive HSCs (Figure 21F). Previous studies reported that vwf could mark platelet-biased HSCs upon transplantation (Sanjuan-Pla et al. 2013). Therefore, differentiation-inactive HSC clones may have generated progeny cells not sampled in the current experimental setting.

Of note, a number of novel targets not yet characterized in HSCs (e.g., Bex4, H2afy) were also identified along the PC2 axis and hence provided a unique resource for subsequent validation (Figure 21F).

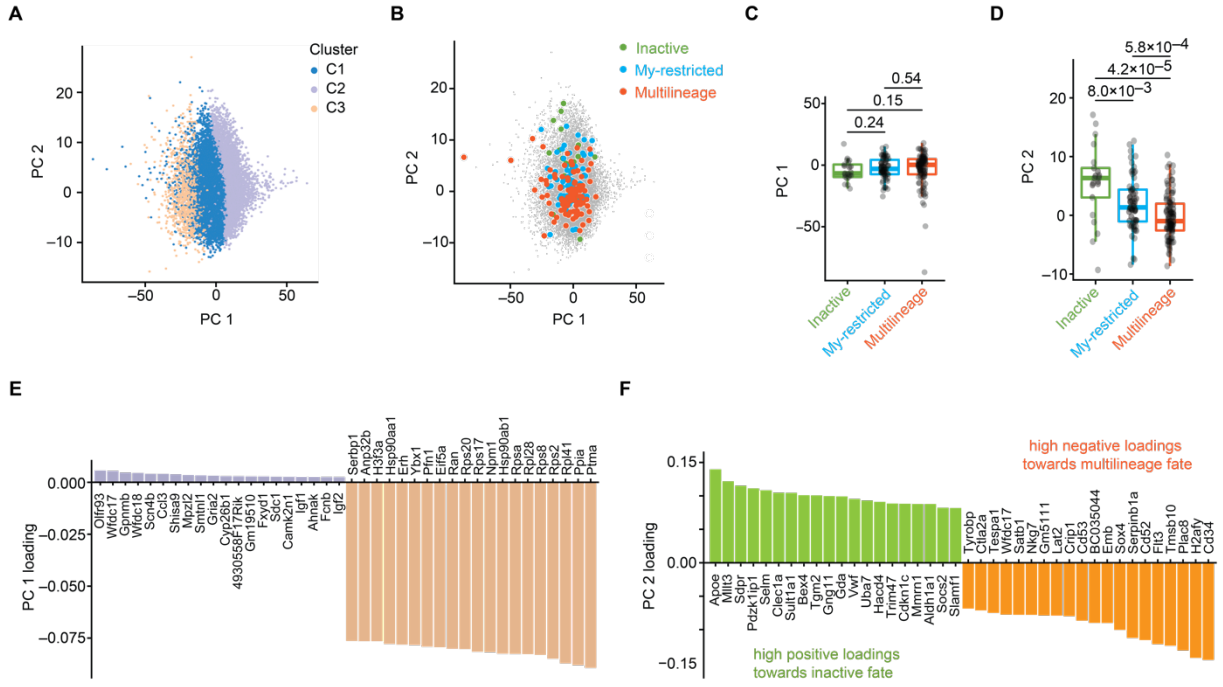


Figure 21. Principal component analysis of HSC clones.

(A) HSC transcriptomes (colored dots, n = 22,565 from Exp #1-3) were analyzed by principal component analysis and clustered along PC1. PC, principal component; C1/2/3, cluster 1/2/3.

(B) HSCs with both RNA barcodes and transcriptome were projected onto the PCA plot and highlighted in colors depending on their fate. HSCs for which only transcriptome but not RNA barcodes were available were indicated by grey dots (n = 22,565 from Experiment #1-3). Inactive, differentiation-inactive; My-restricted, myeloid-erythroid-restricted.

(C) HSCs with different fates (n = 85) had similar PC1 scores. p values were defined by 2-sided Wilcoxon-Mann-Whitney rank-sum test and indicated on the top. In each box plot middle line represents median value, upper and lower lines represent the first and third quartile. Vertical line in the middle represents ± 2.7 standard deviation.

(D) HSCs with different fates (n = 85) had different PC2 scores. p value calculation and box plot characters are the same as described in **(C)**.

(E) Differentially expressed genes related to transcriptome clustering along PC1 axis are shown.

(F) Differentially expressed genes related to fate separation along PC2 axis are shown. Genes related to inactive fate had high positive PC2 loading and genes related to multilineage fate had high negative PC2 loading. (Bioinformatic calculation by Xi Wang; Figure adopted and modified from Pei et al. 2020)

3.3.3 Diffusion pseudotime analysis of HSC and progenitor transcriptomes

To visualize the transcriptional transition from HSCs to their immediate progeny, ‘‘diffusion pseudotime’’, an unsupervised computational approach (Haghverdi et al. 2016) processing single-cell transcriptome data was applied. Based on gene expression similarity between individual cells, the diffusion pseudotime algorithm computed a diffusion map to infer developmental trajectories from HSCs to lineage-primed progenitors (Figure 22A). On this developmental landscape, HSCs locate to the tip of the trajectory, MPPs reside in the central area, while CMPs and CLPs are segregated into two separate branches (Figure 22A). To gain insights into the state of lineage priming of HSPCs, I extracted lineage-specific marker genes from the literature (Table 34) and Xi Wang analyzed their representation on the transcriptional landscape (Figure 22B). The analysis revealed that stem-cell-associated genes were highly enriched in the common branch, which was mainly composed of HSCs and MPPs (Figure 22B). Lineage specification genes, on the other hand, were only found downstream of the common branch and positioned at the end of the trajectory (Figure 22B). Therefore, our analysis confirmed that pseudotime trajectory analysis could robustly recapitulate known hematopoietic developmental pathways from HSCs to lineage-primed progenitors (Figure 22B). Finally, to determine the location of each cell on the landscape, the pseudotime rank from HSCs to more lineage primed progenitors was calculated (Figure 22C). Less differentiated cells that positioned closer to the tip of the trajectory acquired lower rank scores, while more differentiated cells closer to the end of the branches acquired higher rank scores (Figure 22C). In this way, the differentiation states of each cell as well as transcriptome similarity between different cells could be quantified.

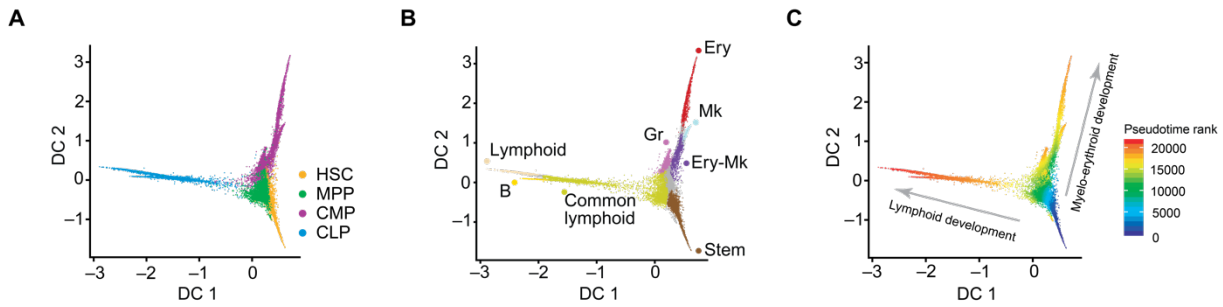


Figure 22. Diffusion pseudotime analysis of HSC and progenitor transcriptomes.

(A) Diffusion map, constructed from single-cell transcriptomes of HSPCs, showing the position and developmental trajectory of HSPC samples (HSPC samples are highlighted in color). DC, diffusion component; HSC, hematopoietic stem cell; MPP, multipotent progenitor; CMP, common myeloid progenitor; CLP, common lymphoid progenitor.

(B) Diffusion map branches were annotated using marker genes listed in Table 34 and highlighted in color. Gr, Granulocyte; Ery, Erythroid; Mk, Megakaryocyte; B, B cell.

(C) Pseudotime rank scores were color coded and shown for each single-cell on the diffusion map.

All data shown in this figure was derived from one of the representative mice (Experiment #1). (Bioinformatic calculation by Xi Wang; Figure adopted and modified from Pei et al. 2020)

Table 34. Marker genes for lineage annotation on diffusion map

Branches	**Marker genes**
Stem cell	Hlf, Ifitm1, Ly6a
Megakaryocyte	Pf4, Vwf, Itga2b
Erythroid	Mki67, Gata1, Car2, Hbb-bs, Hba-a2, Car1, Apoe, Gata2
Monocyte	F13a1, Ly86, Csf1r, C1qb
Granulocyte	Ifitm1, Mpo, Elane, Cebpe, Prss34, Prg2, Gstm1, Fcnb, Ltf, Gfi1, Mmp8, Itgam, Il1b, Ccl6
Lymphoid	Dntt, Satb1, Flt3, Ly6d, Rag1, Ebf1, Cd79a, Gata3, Pax5, Vpreb1, Vpreb2, Vpreb3, Igll1, Fcrla, Ccl5, Ncr1, Cd3d
B cell	Pax5, Vpreb1, Vpreb2, Vpreb3, Igll1, Fcrla

3.3.4 HSC clones exhibit different diffusion pseudotime ranks depending on their fate

Next, transcriptional differences between differentiation-inactive and multilineage HSCs were analyzed based on diffusion pseudotime. HSCs identified as inactive or multilineage by RNA barcodes were projected onto the transcriptional landscape (Figure 23A).

Interestingly, the analysis revealed that inactive HSCs were positioned closer to the tip of the pseudotime trajectory, while multilineage HSCs were positioned further downstream (Figure 23A). Pseudotime rank comparison confirmed the observation that HSCs with different fate occupied distinct transcriptional territories (Figure 23B), supporting the previous PCA results in which fate-defined HSCs could be transcriptionally separated along the PC2 axis (Figure 21B, 21D). Together, the data indicates that distinct HSC differentiation outcomes (inactive or multilineage) are driven by unique fate-associated transcriptional programs.

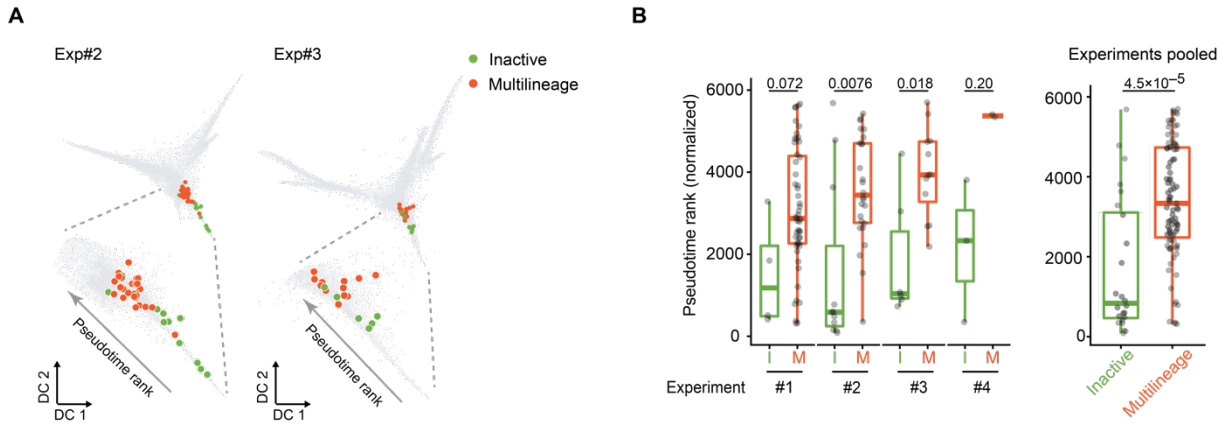


Figure 23. Correlation of HSC fates and diffusion pseudotime ranks.

(A) HSC clones with inactive (colored in green) and multilineage (colored in orange) fate were projected onto the diffusion map, with a zoomed in view of the common trunk shown on the bottom left. Data from two representative experiments is shown (Experiment #2-3). Exp, experiment; DC, diffusion component.

(B) Pseudotime rank comparison of inactive (I, colored in green) and multilineage (M, colored in red) HSC clones. The left panel shows the analysis for individual experiment (n=4, 11, 6, 3 for inactive HSCs from experiment#1-4 and n=58, 29, 13, 3 for multilineage HSCs from experiment#1-4) and the right panel shows the pooled statistics for the four experiments (n=24 for inactive HSCs and n=103 for multilineage HSCs). p values were defined by 2-sided Wilcoxon-Mann-Whitney rank-sum test and indicated on the top. Middle line in each box plot represents median value, upper and lower lines in each box plot represent the first and third quartile. Vertical line in the middle represents ± 2.7 standard deviation. (Bioinformatic calculation by Xi Wang; Figure adopted and modified from Pei et al. 2020)

3.3.5 LT-HSC gene expression signature but not cell cycle genes distinguish differentiation-inactive from multilineage HSCs

LT-HSCs are superior to ST-HSCs in their ability to reconstitute lethally irradiated recipients in serial transplantation (Kiel et al. 2005; Yang et al. 2005). Taking into account that the HSC analyses so far did not distinguish between LT- and ST-HSCs, I asked whether these two phenotypically defined subsets might contribute differently to the fate-associated transcriptional program. To this end, the relative expression intensity of highly variable genes from scRNAseq data of LT- and ST-HSCs (unpublished data from Rodewald lab) was determined for inactive and multilineage HSCs. The analysis revealed that the LT-HSC signature was more closely related to the expression profile of differentiation-inactive HSCs than that of multilineage HSCs (Figure 24A). In contrast, the expression of ST-HSC signature genes was lower in the inactive and higher in the multilineage HSCs (Figure 24A). These findings were further confirmed using a published data set (Kowalczyk et al. 2015) (Figure 24B). In conclusion, although LT- and ST-HSC signatures could be found in both inactive and multilineage HSCs, their distribution frequency varied significantly (Figure 24C).

To assess whether HSCs with distinct fates have different self-renewal potential, the expression level of cell-cycle related genes (S/G2/M phase, retrieved from Tirosh et al. 2016) was determined. The analysis showed that a fraction of HSCs from both fates was cycling and that the average cell cycle scores were not different between inactive and multilineage HSCs (Figure 24D). Using label retention assay, several studies proposed that a subset of ‘‘dormant’’ HSCs, exists that rarely self-renews (Cabezas-Wallscheid et al. 2017; Walter et al. 2015; Wilson et al. 2008). Therefore, I asked whether such a ‘‘dormant’’ gene expression signature (DEGs extracted by Xi Wang from Cabezas-Wallscheid et al. 2017) could be found in fate-associated HSC transcriptomes. Interestingly, the analysis showed that the ‘‘dormant’’ HSC signature was higher in multilineage HSCs compared to inactive HSCs, which is counterintuitive but would suggest that dormant HSCs may also contribute to hematopoiesis (Figure 24E).

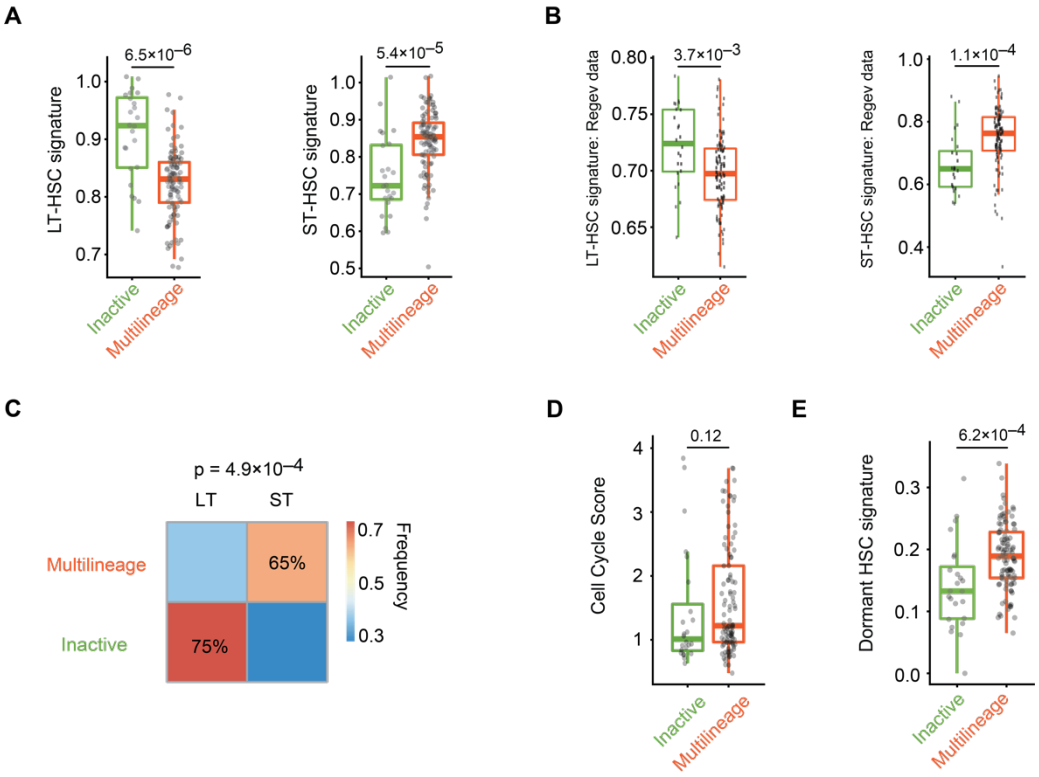


Figure 24. Characterizing differentiation-inactive and multilineage HSC clones by gene expression signatures.

(A) Comparison of long-term (LT) and short-term (ST) hematopoietic stem cell (HSC) gene expression signatures between inactive and multilineage HSCs using unpublished data from the Rodewald lab.

(B) Comparison of LT- and ST- HSC gene expression signatures between inactive and multilineage HSCs using data from Kowalczyk et al. 2015.

(C) Determining the frequency of fate-defined HSCs found in phenotypically defined LT- and ST- HSCs by their expression of LT- and ST- HSC signatures.

(D) Comparing cell cycle score enrichment between inactive and multilineage HSCs.

(E) Comparing dormant HSC signature between inactive and multilineage HSCs.

Data shown were derived from four experiments pooled (Experiment #1-4, n=24 cells for inactive HSCs and n=103 cells for multilineage HSCs). p values, indicated on the top of the figure, were defined by 2-sided Wilcoxon-Mann-Whitney rank-sum test for **(A) (B) (D) (E)** and Fisher's exact test for **(C)**. Middle line in each box plot represents median value, upper and lower lines in each box plot represent the first and third quartile. Vertical line in the middle represents ± 2.7 standard deviation. (Bioinformatic calculation by Xi Wang; Figure adopted and modified from Pei et al. 2020)

3.3.6 Extraction of gene expression signatures that discriminate differentiation-inactive and multilineage HSCs

Finally, DEGs between inactive and multilineage HSCs were determined (Figure 25A). In line with previous results (Figure 21F), several highly variable genes identified in PCA were also found in the current analysis, including Apoe, Pdzk1ip1, Bex4 and Cdkn1c for inactive HSCs and Cd34, H2afy, Plac8, Flt3, Serpinb1a, Cd53 for multilineage HSCs. Mycn, a known regulator of HSC quiescence, was enriched in inactive HSCs (Laurenti et al. 2008). Notably, interesting targets with unknown function in HSCs were also identified. For instance, Hoxb2, enriched in inactive HSCs, was reported to be highly expressed in leukemic cells but remains uncharacterized in steady-state hematopoiesis (Cabezas-Wallscheid et al. 2014; Mullighan et al. 2007). Given Hoxb5, another member of the HOX family gene, was previously reported to mark LT-HSCs (Chen et al. 2016), it is conceivable that Hoxb2 may play a role in HSC fate determination. Other genes like Bex1 and Bex4 may also serve as interesting future targets, given that these genes are critical in stem cell function of non-hematopoietic tissues (Ito et al. 2014). To validate the fate-associated genes in silico, Xi Wang performed supervised random forest classification on the transcriptome of HSCs (Figure 25B). Indeed, DEGs robustly divided HSCs into inactive or multilineage fates, suggesting that DEGs may serve as signatures to enrich HSCs with distinct fates.

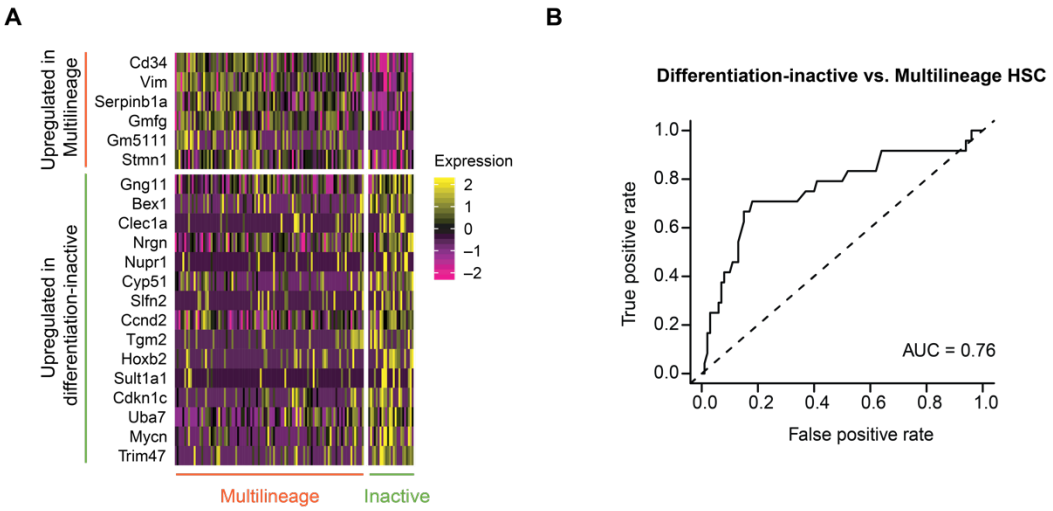


Figure 25. Gene expression analysis of differentiation-inactive and multilineage HSC clones.
(A) Heatmap showing DEGs of inactive and multilineage HSCs. Each row represents a DEG and each column represents a fate-defined HSC. Gene expression level is represented by color scale.
(B) Distinguishing inactive and multilineage HSCs by random forest algorithm using identified fate-associated transcriptome. The result was shown by receiver operating characteristic curve. Dashed line represents random classification with area under curve (AUC) = 0.5.
Data shown are derived from four experiments pooled (Experiment #1-4, n=24 cells for inactive HSCs and n=103 cells for multilineage HSCs). (Bioinformatic calculation by Xi Wang; Figure adopted and modified from Pei et al. 2020)

3.3.7 Transcriptome analysis of myeloid-erythroid-restricted and multilineage HSCs

To determine the transcriptome differences between myeloid-erythroid-restricted and multilineage HSCs, pseudotime rank analysis was repeated with inactive HSCs included as reference. The results revealed that restricted HSCs were positioned between inactive and multilineage HSCs (Figure 26A). Furthermore, no difference in the cell cycle scores between restricted and multilineage HSCs was detected (Figure 26B). Of note, a key difference between multilineage and myeloid-erythroid-restricted HSCs is their lymphoid potential. Therefore, in order to assess whether lymphoid priming could already be detected at the HSC level, highly variable genes from CMPs and CLPs were derived to generate myeloid-erythroid and lymphoid signatures, respectively. Next, the expression level of these lineage signatures was analyzed in HSCs. Interestingly, the results showed that relative expression of the lymphoid but not the myeloid-erythroid signature differed between myeloid-erythroid-restricted and multilineage HSCs (Figure 26C), suggesting that lymphoid priming already takes place at the HSC stage. A similar result was observed at the MPP stage with an even more significant difference (Figure 26D), which was supported by random forest analysis (Figure 26E). Finally, DEGs of restricted and multilineage HSCs or MPPs were determined (Figure 26F, 26G). Flt3, a marker found to be express on lymphoid-primed MPPs (LMPPs), was expressed significantly higher in MPPs generated from multilineage HSCs, which implies a potential distinct developmental pathway from multilineage HSCs to LMPPs (Figure 26G). In summary, the transcriptional characterization of restricted and multilineage clones suggested that the

separation between the two different HSC fates is due to differential lymphoid priming. Furthermore, the conclusion that fate specification occurs at the HSC stage is supported by the fact that lineage priming was already observed at the HSC level and became more prominent at the MPP stage (Figure 26C, 26D, 26E).

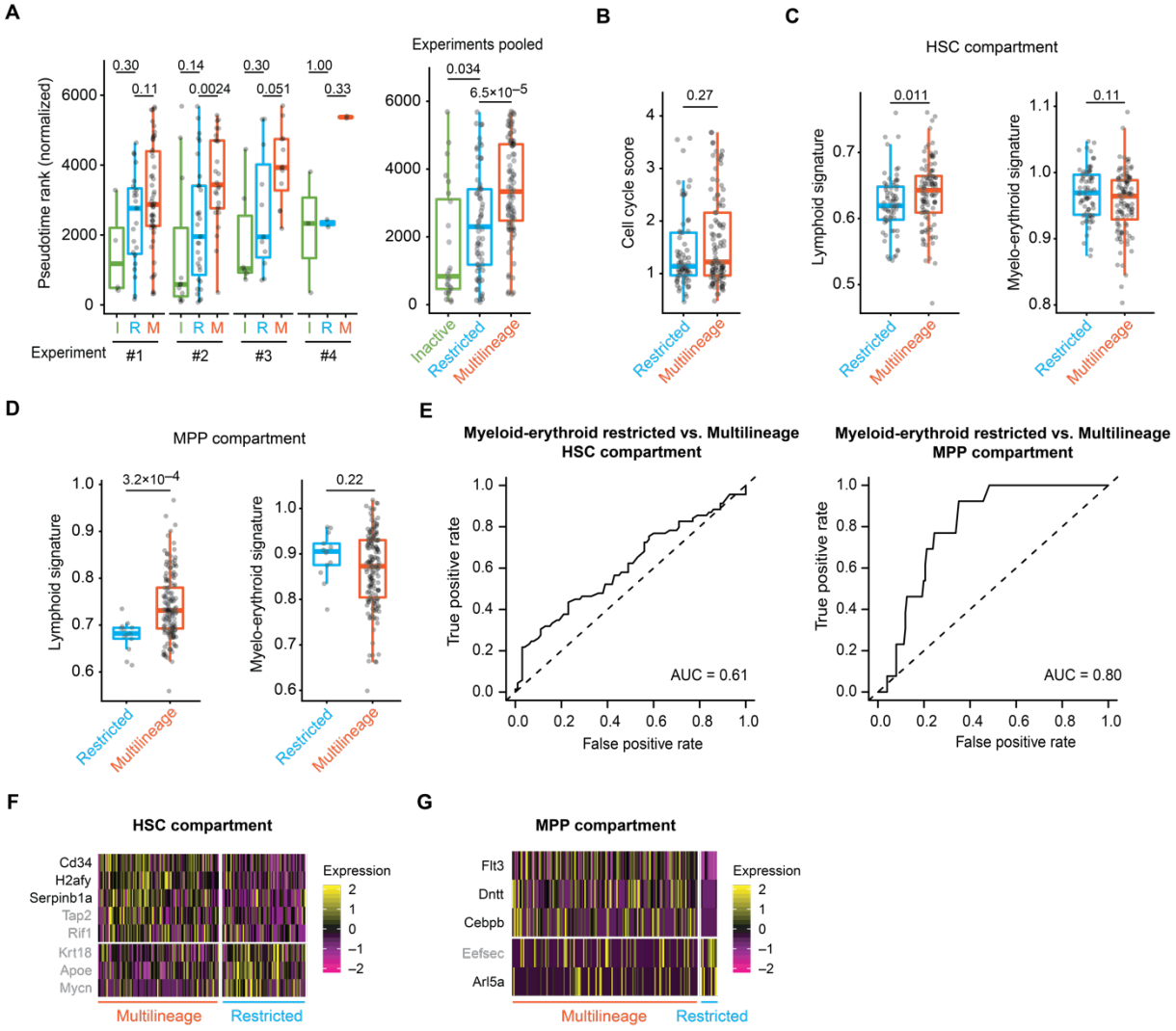


Figure 26. Transcriptome analysis of myeloid-erythroid restricted and multilineage HSC clones.

(A) Pseudotime rank comparison between inactive (I, colored in green), myeloid-erythroid-restricted (M, colored in blue) and multilineage (M, colored in red) HSC clones. The left panel shows the analysis for individual experiments (n=4, 11, 6, 3 for inactive, n=25, 31, 11, 2 for restricted and n=58, 29, 13, 3 for multilineage HSCs from Experiment #1-4) and the right panel shows the pooled analysis of the experiments (n=24 for inactive, n=69 for restricted and n=103 for multilineage HSCs).

(B) Cell cycle score comparison between myeloid-erythroid-restricted and multilineage HSCs.

(C) Comparing the expression of lymphoid signature as well as myeloid-erythroid signature between restricted and multilineage HSCs. HSC, hematopoietic stem cell.

(D) Comparing the expression of lymphoid signature as well as myeloid-erythroid signature between restricted and multilineage MPPs. MPP, multipotent progenitor.

(E) Distinguishing restricted and multilineage HSCs and MPPs by random forest algorithm using identified fate-associated transcriptomes. The result is shown by a receiver operating characteristic curve. Dashed line represents random classification with area under curve (AUC) =0.5.

(F) Heatmap showing DEGs of restricted and multilineage HSCs. Each row represents a DEG and each column represents a fate-defined HSC. Gene expression level is represented by color scale. DEGs with false discovery rate (FDR) below 5% are labeled in black and DEGs with FDR below 20% are labelled in grey.

(G) Heatmap showing DEGs of restricted and multilineage MPPs. Each row represents a DEG and each column represents a fate-defined MPP.

Data shown for **(B)(C)(D)(E)(F)(G)** are derived from four experiments pooled (Experiment # 1-4, n=69 for restricted and n=103 for multilineage HSCs). p values were defined by 2-sided Wilcoxon-Mann-Whitney rank-sum test for **(A) (B) (C) (D)** and indicated on the top of each figure. Middle line in each box plot represents median value, upper and lower lines in each box plot represent the first and third quartile. Vertical line in the middle represents ± 2.7 standard deviation. (Bioinformatic calculation by Xi Wang; Figure adopted and modified from Pei et al. 2020)

4 Discussion

4.1 *PolyloxExpress*: from DNA to RNA, from fates to fate determinants

More than a century ago, the French artist Paul Gauguin created the famous painting named ‘‘Where Do We Come From? What Are We? Where Are We Going?’’. The name itself pointed at fundamental questions of human beings which could also be extended to the field of biology: What is the origin of a cell? What defines the identity of a cell? What is the fate of a cell (Figure 27)? To answer these questions, different methods have been established in the past, including fluorescent protein-based lineage tracing, high-resolution DNA barcoding and transcriptome-based fate prediction (introduction section 1.2 and 1.3). However, due to various limitations, none of these methods are capable of linking cell fate with fate-associated regulators. Therefore, during my PhD work, I focused on the establishment and application of the *PolyloxExpress* barcoding system, which allows direct profiling of cell fate (RNA barcode) and fate determinants (fate-associated gene expression signatures) at single-cell resolution.

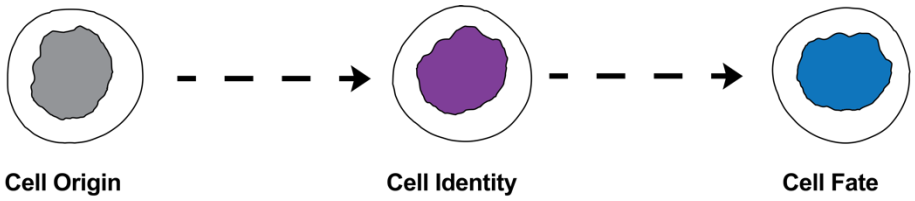


Figure 27. Schematic diagram of cell developmental history.

In the meantime, other RNA barcoding approaches based on virus-mediated transfer of barcodes have also been reported (Fennell et al. 2022; Rodriguez-Fraticelli et al. 2020; Weinreb et al. 2020). Another recent study reported *in vivo* RNA barcoding using CRISPR based editing (Bowling et al. 2020). However, barcode induction could not be restricted to the cells of interest in this approach (Bowling et al. 2020). Key advantages of *PolyloxExpress* compared to these methods are 1) *in vivo* barcoding (the barcoding platform and the modifying Cre enzyme are present in all cells of the organism), 2) tissue

specificity and time of recombination can be determined by selection of a Cre allele with an appropriate promoter and the timing of tamoxifen injection, respectively. This minimizes artificial perturbations and allows non-invasive tracking of the cell differentiation history. This is of particular importance for cell types that need to be maintained strictly at an undifferentiated state, which is usually difficult to achieve by *in vitro* culture (e.g., stem cells).

The *PolyloxExpress* system also has limitations. The first limitation is that due to the inversion nature of the loxP sites, *PolyloxExpress* only allows for one-time barcode induction. A second induction to the same organism may generate barcode sequences that are indistinguishable from the first induction, which prevents application in longitudinal studies that require consecutive barcoding (Chan et al. 2019). The second limitation is RNA barcode drop out in the scRNAseq workflow due to detection sensitivity. By modifying the standard protocol with a target enrichment approach, I improved barcode capture efficiency in primary cells. Nevertheless, more than half of the cells with transcriptome still lack barcode information due to drop out. On the one hand, this relates to the technical constrains of microfluidic based scRNAseq methods when capturing low abundance genes (Zhang et al. 2019; Zheng et al. 2017). On the other hand, the intrinsic *Rosa26* promoter governing *PolyloxExpress* barcode expression may not be the strongest driver. Replacing *Rosa26* with a stronger promoter (e.g., CAG promoter) could further increase RNA barcode capture efficiency. Finally, using tdTomato fluorescence intensity level as read out, Thorsten Feyerabend and I realized that barcode length (hence length of the tdTomato reporter 3' UTR) negatively correlates with the barcode expression level (data not shown), which might be explained by faster RNA degradation of long barcodes compared to short barcodes. Therefore, higher barcode capture efficiency can be achieved by analyzing mice with stronger barcode recombination.

4.2 HSC fate heterogeneity

Using single-HSC transplantation technique, several studies suggested that HSC fate is not homogeneous (Dykstra et al. 2007; Muller-Sieburg et al. 2004; Yamamoto et al. 2013). By labelling HSC progenitors at E9.5 using *Tie2*$^{MeriCreMer/+}$*Rosa26*$^{PolyloxExpress/+}$ mice, I confirmed that fate heterogeneity is not an artifact of transplantation, but rather an intrinsic character of the HSC compartment. However, the HSC fate patterns observed in my experiments were not exactly consistent with the reported transplantation fates (Dykstra et al. 2007; Muller-Sieburg et al. 2004; Yamamoto et al. 2013). For instance, in my steady state barcoding experiments, inactive HSCs were robustly observed across all mice analyzed. However, such HSC subset was rarely found in the transplantation setting of published literature, which may suggest that transplantation imposes stress on HSCs and alters their functions. Indeed, when comparing fate of the same HSC clones before and after transplantation, I observed fate conversion in a fraction of HSC clones. Therefore, analyzing HSCs in their native environment is critical for understanding their ‘‘true’’ functions.

In transplantation studies, megakaryocyte-restricted HSCs have also been described (Carrelha et al. 2018; Notta et al. 2016; Sanjuan-Pla et al. 2013). In fact, a recent DNA barcoding study from Carmago’s laboratory concluded that most HSCs are megakaryocyte-restricted at steady-state (Rodriguez-Fraticelli et al. 2018). While I confirmed the presence of this restricted HSC subset with *PolyloxExpress*, I detected them at a much lower frequency. Importantly, this discrepancy is not a result of different labelling times (E9.5 labelling in my experiment and adult labelling in Rodriguez-Fraticelli et al. 2018), because when *Tie2*$^{MeriCreMer/+}$ *Rosa26*$^{PolyloxExpress/+}$ mice were labelled at the age of eight weeks and chased for one year, only about 10% of the HSCs revealed megakaryocyte-restricted fate (data not shown). Therefore, the high percentage of reported megakaryocyte-restricted HSCs might be a result of under sampling other hematopoietic lineage cells (Rodriguez-Fraticelli et al. 2018). Consistently, however, current studies and my own data support the view that HSC fate is heterogeneous.

Moreover, the presence of HSCs with distinct fates indicates that fate commitment is already established at the HSC stage (Figure 28).

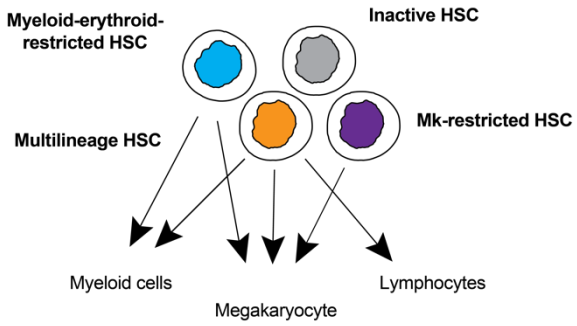


Figure 28. Schematic diagram of HSC fate heterogeneity.
Fate-defined hematopoietic stem cells (HSCs) are highlighted in color and arrows indicate cell types they generate. Mk, megakaryocyte. (Figure adopted and modified from Shang and Rodewald 2022)

4.3 HSC fate determination in embryonic hematopoiesis

When HSCs emerge, they will expand, migrate to the fetal liver and eventually colonize the bone marrow around birth (Mikkola and Orkin 2006). Although HSC fate heterogeneity was observed at steady state and through transplantation, the timing of HSC fate specification (embryonically or postnatally) remains unknown. Transplanting single HSCs purified from both fetal liver and bone marrow cells revealed that myeloid-biased and multilineage HSCs exist at both prenatal and postnatal stages (Benz et al. 2012). Using E9.5 labelled *Tie2*$^{MeriCreMer/+}$ *Rosa26*$^{PolyloxExpress/+}$ mice, I studied embryonic HSC fates (E18.5) *in vivo* and found similar fates as with adult HSCs, suggesting that HSC fates are determined at early developmental stage. Furthermore, consistent fate patterns found in both fetal liver and adult bone marrow questioned the role of the niche in HSC fate determination. Of note, since barcodes were induced at E9.5, and hence during HSC emergence, fate decision making for HSC subsets likely took place in HSC progenitors (Figure 29A). If fate determination had not occurred prior to the emergence of HSCs, it would be unlikely that HSCs with inactive or myeloid-erythroid-restricted fates could be distinguished by barcoding because they would then share barcodes with other fate-determined HSCs due to their common clonal origin (Figure 29B). Therefore, the current

data suggests that subsets of hemogenic endothelium (HE) that share the same Tie2 expression history but differ in their fates co-exist in the embryos. It has been shown that apart from the aorta-gonad-mesonephros (AGM) region, the placenta also serves as a hematopoietic organ in the early embryonic stage (Gekas et al. 2005; Ottersbach and Dzierzak 2005). Whether spatial separation of embryonic hematopoiesis can lead to differential determination of HE fates remains largely unknown.

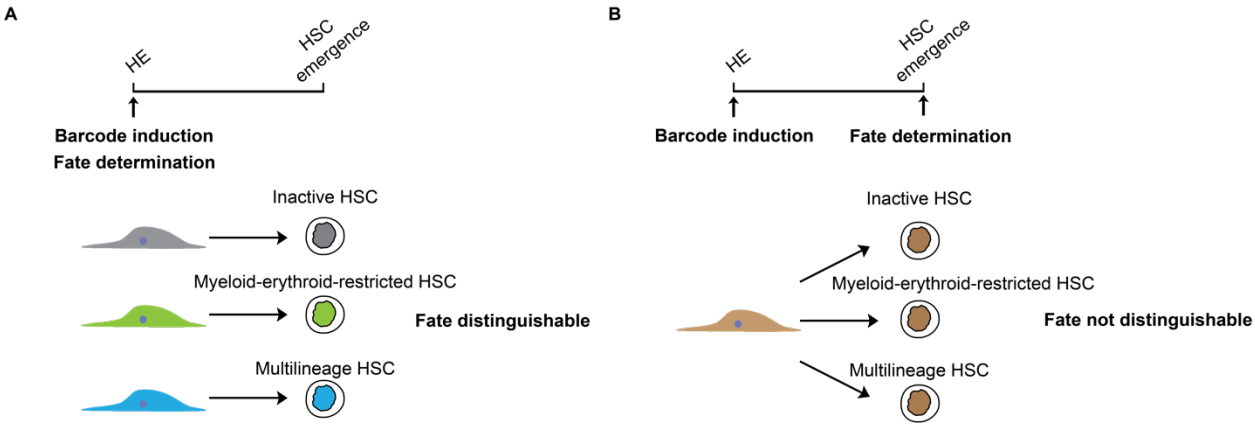


Figure 29. Schematic diagram of fate determination during embryonic hematopoiesis.

(**A**) Barcodes are induced in hemogenic endothelium (HE) cells. Fate determination happens at the hemogenic endothelium stage. Distinct hematopoietic stem cell (HSC) fates are distinguishable. Each color represents a unique fate.

(**B**) Barcodes are induced in hemogenic endothelium cells. Fate determination happens after HSC emergence. Distinct hematopoietic stem cell (HSC) fates are not distinguishable. Each color represents a unique fate.

While it has been thought that HSC progenitors, emerging in mice around E9.5 of development, are the ancestors of definitive HSCs, novel lineage tracing studies suggest that bone marrow HSCs may not be the only source of native hematopoiesis (Dignum et al. 2021; Patel et al. 2022; Yokomizo et al. 2022). In contrast to the classic model (Figure 30A), these studies proposed that hematopoietic progenitor cells with a MPP phenotype, called embryonic multipotent progenitor (eMPP), emerge directly from the hemogenic endothelium of the AGM region and contribute to embryonic and adult hematopoiesis (Patel et al. 2022; Yokomizo et al. 2022) (Figure 30B). While it remains to be

demonstrated whether Tie2 is also expressed in hemogenic endothelium cells, it could be that these cells were also barcoded in my experiments and hence gave rise to both HSCs and eMPPs. In this case, some HSC clones would share barcodes with MPPs that were generated independent from HSCs, and the observed clonal fate could be a mixture of ‘‘true’’ HSC fate and eMPP fate (Figure 30B). Therefore, further experiments labeling HSCs and MPPs separately at a later stage of development (e.g., induction at E12.5 after HSPCs emergence) would be critical to solve the current controversy (Figure 30).

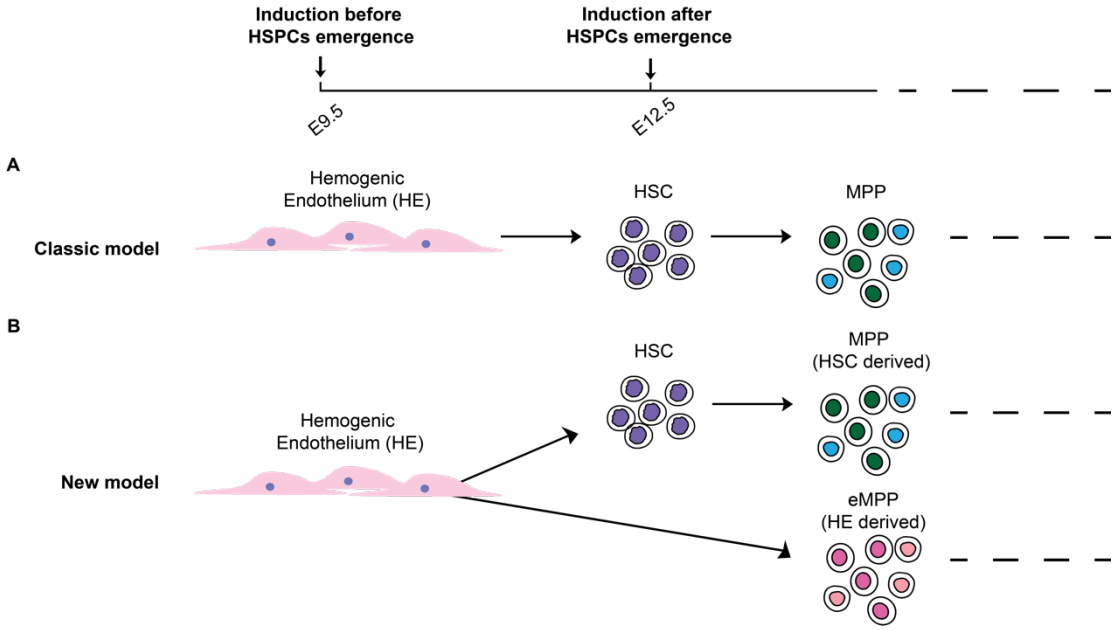


Figure 30. Schematic diagram of embryonic hematopoiesis models.

(**A**) In the classic model, hemogenic endothelium cells generate LT-HSCs, which then give rise to ST-HSCs and MPPs.

(**B**) In the new model, hemogenic endothelium cells could generate both LT-HSCs as well as ST-HSCs and MPPs. Barcode induction before HSPCs emergence may lead to mixed LT-HSC and ST-HSC/MPP fates in the new model but not the classic model.

HSPC, hematopoietic stem and progenitor cell; HE, hemogenic endothelium; HSC, hematopoietic stem cell; MPP, multipotent progenitor; eMPP, embryonic multipotent progenitor.

4.4 HSC fate stability and fate-associated transcriptome signatures

It has long been debated whether the fate of HSCs is cell-intrinsically or -extrinsically determined. Several studies implied that niche cells (Lymperi et al. 2010) and cytokines (Mossadegh-Keller et al. 2013) may instruct HSC fate. However, in serial transplantation studies using single HSCs, lineage output was not random across different recipients (Dykstra et al. 2007; Muller-Sieburg et al. 2004). A recent study combined multicolor fluorescent labeling with serial transplantation and showed stereotypic behaviors of HSCs both at steady state and after inflammatory stress (Yu et al. 2016). Nevertheless, most of the published HSC fate mapping experiments are end-point analyses, only reflecting a snapshot of the ongoing differentiation process. To study the dynamic process of HSC differentiation, I repeatedly collected blood from several barcoded mice. Intriguingly, the data suggests that HSC fates are mostly coherent over time in steady state hematopoiesis and even after transplantation, which is direct evidence for HSC fate stability at clonal resolution. Therefore, intrinsic mechanisms, yet to be determined, are the driving force of stereotypic HSC behaviors.

Over the past decade, scRNAseq has been widely used to uncover cellular subsets with distinct transcriptomes in different organs (Gawad et al. 2016; Huang et al. 2015; Tanay and Regev 2017). However, how the transcriptome correlates with cell function (e.g., cell fate) remains to be investigated. Combining fate-mapping with scRNAseq using *PolyloxExpress*, I found that deducing from the HSC transcriptome alone is not sufficient to predict experimentally validated fate heterogeneity. Only when the barcode information is integrated, fate-associated transcriptome signatures could be revealed. Indeed, the transcriptome acquired by scRNAseq only reflects the current cellular state, but not the dynamic differentiation processes of the past or the future. Although different algorithms have been computed to infer cell differentiation pathways based on transcriptome data alone (Haghverdi et al. 2016; Kester and van Oudenaarden 2018; Trapnell et al. 2014), such methods can only reflect hypothetical cellular behavior at the population scale, but cannot resolve the developmental pathway of individual clones (Wagner and Klein 2020). For instance, in the diffusion pseudotime analysis, HSCs with distinct fates all reside in

the common branch of the diffusion map and gradually transfer towards myeloid and lymphoid branches (Figure 31). However, in reality, not all HSC clones will follow the inferred developmental trajectories and give rise to myeloid and lymphoid cells. Instead, inactive HSC clones will reside at the tip and rarely go through the trajectory. Myeloid-erythroid-restricted HSC clones will only take the myeloid-erythroid pathway but not the lymphoid pathway (Figure 31). Therefore, transcriptome alone, even at single-cell resolution, is not enough to predict clonal behavior. Functional validation with barcoding experiments is essential to determine the *bona fide* developmental pathways (Shang and Rodewald 2022).

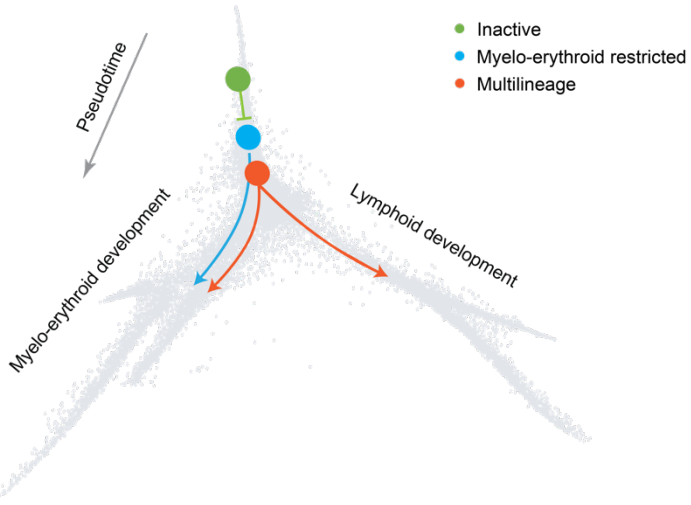


Figure 31. Limitations of transcriptome-based developmental pathway inference algorithms.
Schematic diagram showing HSC clonal behaviors on the diffusion map calculated based on single-cell transcriptome. Each small grey dot represents a cell, each big colored dot represents a member of an HSC clone with the color indicating the clone's fate. Inactive HSC clones, highlighted in green, reside at the tip of the trajectory and do not propagate their barcodes to the downstream, as shown by the green line. Myeloid-erythroid-restricted HSC clones, highlighted in blue, only propagate their barcodes to the myeloid-erythroid cells but not lymphoid cells, as shown by the blue arrowed line. Multilineage HSC clones, highlighted in orange, propagate their barcodes to both myeloid-erythroid and lymphoid cells, as shown by the two orange arrowed lines. (Figure adopted and modified from Pei et al. 2020)

The identification of HSC fate-associated transcriptome signatures not only testified for the presence of heterogeneous HSC fates, but also provided molecular insights into fate

determination. Detailed analysis of the fate-related transcriptome signatures of HSCs revealed that both LT- and ST-HSCs could fall into the same fate category, suggesting that SLAM markers, widely used to characterize HSCs, are not sufficient to uncover distinct HSC functions. It has been claimed that a fraction of rarely proliferating dormant HSCs exists (Cabezas-Wallscheid et al. 2017; Wilson et al. 2008), but how cell cycle correlates with differentiation of HSCs remains largely unknown. Comparing fate-associated signatures with the signature of dormant HSCs may shed light on this question. Interestingly, multilineage HSCs, but not inactive HSCs, were enriched for a published dormant signature (Cabezas-Wallscheid et al. 2017). This is counterintuitive because HSCs must proliferate to generate blood cells while retaining their identity. A possible explanation of what might be wrong with the dormancy concept comes from a study by Morcos and colleagues, who found that HSC label retention may have been overestimated due to leakiness and nonspecific GFP accumulation in the H2B-GFP model commonly used to determine dormant HSCs (Morcos et al. 2022). Therefore, the correlation between HSC differentiation and cell cycle remains to be formally tested.

4.5 Outlook: Combining *PolyloxExpress* RNA barcoding with single-cell multiomics

In my PhD work, I demonstrated the power of combining *PolyloxExpress* with scRNAseq, but there are further potential applications for the system. Given the remarkable stability of HSC fate at steady state and upon transplantation, it is conceivable that certain epigenetic mechanisms (e.g., chromatin accessibility and DNA methylation) not captured by scRNAseq may play an important role in fate maintenance. Therefore, advances in single-cell multiomics experimental approaches now provide unique opportunities to link fate and fate determinants in multiple dimensions (Argelaguet et al. 2019; Mimitou et al. 2019; Mimitou et al. 2021; Stoeckius et al. 2017) (Figure 32).

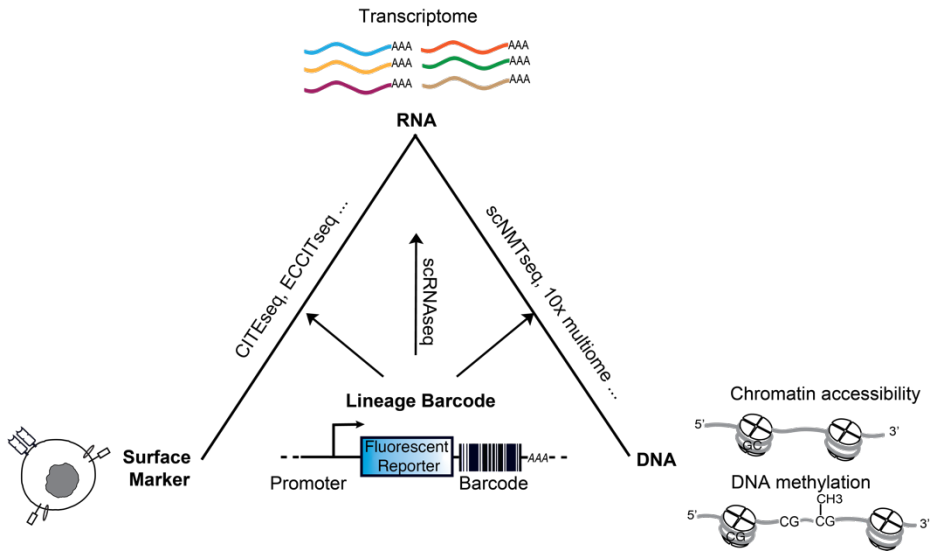


Figure 32. Schematic diagram of combining RNA barcoding with single-cell multiomics approaches. The common design for the RNA barcoding approach is shown at the bottom of the triangle. Single-cell multiomics techniques available for linking RNA and DNA expression or linking RNA and surface marker expression are shown on both sides of the triangle. CITEseq, Cellular Indexing of Transcriptomes and Epitopes by sequencing; ECCITEseq, Expanded CRISPR compatible Cellular Indexing of Transcriptomes and Epitopes by sequencing; scNMT, single-cell nucleosome, methylation and transcription sequencing. (Figure adopted and modified from Shang and Rodewald 2022)

For instance, over the past two decades, the combination of Lin^- $Sca\text{-}1^+$ Kit^+ surface markers with the SLAM family markers $CD150^+$ $CD48^-$ has been established and commonly accepted for the purification of HSCs by most laboratories (Busch et al. 2015; Carrelha et al. 2018; Kiel et al. 2005; Oguro et al. 2013; Sun et al. 2014; Yu et al. 2016). Although functional heterogeneity exists within this phenotypic compartment, there are so far no good markers for the purification of fate-defined HSC subsets. As a result, identification of HSC subsets still largely relies on transplantation or barcode sequencing, which are tedious, time consuming and expensive. In recent years, a new technique named CITEseq (Cellular Indexing of Transcriptomes and Epitopes by sequencing) has been introduced (Stoeckius et al. 2017). This technique enables paralleled quantitative profiling of surface proteins and RNA expression at single-cell resolution by using oligonucleotide-conjugated surface-antigen-specific antibodies to label cells before

scRNAseq (Stoeckius et al. 2017). Adapting this protocol to the *PolyloxExpress* workflow may provide a unique opportunity to uncover novel markers for HSC subsets purification.

In conclusion, during my PhD, I have established the *PolyloxExpress* RNA barcoding system workflow, which allows simultaneous read-out of cell fate and fate-associated transcriptome at single-cell resolution. Applying this system to study hematopoiesis, I have uncovered multiple key HSC characteristics, including HSC fate determination, HSC fate heterogeneity and HSC fate stability. Transcriptome analysis of HSCs with distinct fates provides molecular insights into HSC fate regulation. The versatile compatibility and feasibility of *PolyloxExpress* will not only shed light on HSC biology, but also beyond. Eventually, the paths towards the understanding of cell origin, cell identity, and cell fate shall be uncovered.

5 Summary

All blood and immune cells share the same developmental origin but have diverse functions. Hematopoietic stem cells play a central role in hematopoiesis, as reflected by their ability of multilineage reconstitution of lethally irradiated mice. Recent advancements in DNA barcoding showed that hematopoietic stem cells are not functionally uniform but have heterogenous fates (multilineage, myeloid-erythroid restricted, or differentiation-inactive). The time of hematopoietic stem cells fate specification (embryonically or postnatally) and the stability of distinct fates (at steady state or during transplantation) remain largely unknown. Of note, while DNA barcoding can reveal physiological fates in vivo, it cannot provide molecular insights into fate regulation. To approach these key questions, a next generation barcoding system, *PolyloxExpress*, was developed. By transcribing the DNA barcodes into RNA, *PolyloxExpress* allows high-resolution barcoding at both DNA and RNA levels. Applying the *PolyloxExpress* system in mice and focusing on DNA barcodes, I analyzed hematopoietic stem cell fates at embryonic stages, and found the same fate patterns as in adult stem cells, indicating very early fate determination. To temporally resolve hematopoietic stem cell output in adult life, peripheral blood was repeatedly collected from individual barcoded mice. Distinct stem cell fates remained largely stable over time in both native hematopoiesis and after transplantation, suggesting that hematopoietic stem cell fates are determined cell-intrinsically. To link fates to transcriptomes, I combined *PolyloxExpress* with single-cell RNA sequencing and established a customized protocol for simultaneous recovery of full-length RNA barcodes and transcriptome. These experiments uncovered that fate-defined hematopoietic stem cell occupied different regions on the transcriptional landscape and have hence, at least partially, distinct transcriptome signatures. These fate-associated signatures were not apparent when analyzing transcriptome alone, which underscores the importance of obtaining fate and transcriptome information together. These data also contain transcripts with uncharacterized functions in stem cells, which may provide a unique resource for future studies. Collectively, *PolyloxExpress* barcoding revealed coherent hematopoietic stem cell fates which are cell-intrinsically determined and arise around the time of hematopoietic stem cell emergence in the embryo. Fate-associated transcriptome signatures could be identified. Combining *PolyloxExpress* with single-cell multiomics techniques will provide further insights into molecular aspects of cell fate determination.

6 Zusammenfassung

Alle Blut- und Immunzellen haben denselben Entwicklungsursprung, zeigen aber unterschiedliche Funktionen. Hämatopoetische Stammzellen spielen eine zentrale Rolle bei der Hämatopoese, wie ihre Fähigkeit zur Rekonstitution der unterschiedlichen Zelllinien in letal bestrahlten Mäusen zeigt. Jüngste Fortschritte bei der Technologie der experimentellen Zellmarkierung durch DNA-Barcodes haben gezeigt, dass hämatopoetische Stammzellen keine einheitliche Population bilden, sondern selbst unter natürlichen Bedingungen einem heterogenen Differenzierungsschicksal unterliegen. Der Zeitpunkt der Spezifikation des hämatopoetischen Stammzellschicksals (embryonal oder postnatal) und die Stabilität der verschiedenen hämatopoetischen Differenzierungsschicksale (im natürlichen Gleichgewicht und bei Transplantation oder Infektion) sind jedoch noch weitgehend unbekannt. Außerdem können DNA-Barcodes zwar physiologische Schicksale *in vivo* aufzeigen, aber keine molekularen Einblicke in die Schicksalsregulierung liefern. Um diese wichtigen Fragen zu klären, wurde *PolyloxExpress*, als nächste Generation des zellulären Barcode-Reportersystems *Polylox*, entwickelt. Durch die Transkription der DNA-Barcodes in RNA ermöglicht *PolyloxExpress* hochauflösende Barcode-Markierung sowohl auf DNA- als auch auf RNA-Ebene. Unter Anwendung des *PolyloxExpress* Systems in Mäusen und der Fokussierung auf DNA-Barcodes habe ich die Schicksale hämatopoetischer Stammzellen im Embryonalstadium analysiert und dieselben Muster wie bei adulten hämatopoetischen Stammzellen gefunden, was auf eine sehr frühe Festlegung des Schicksals hinweist. Um die Produktion hämatopoetischer Stammzellen im Erwachsenenalter zeitlich aufzulösen, wurde peripheres Blut wiederholt von einzelnen mit Barcodes markierten Mäusen entnommen. Interessanterweise blieben die verschiedenen hämatopoetischen Stammzellschicksale sowohl in der nativen Hämatopoese als auch nach Transplantation weitgehend stabil, was darauf hindeutet, dass die hämatopoetischen Stammzellschicksale zell-intrinsisch festgelegt werden. Um die Schicksale und ihre molekularen Bestimmungsfaktoren in Verbindung zu bringen, habe ich das *PolyloxExpress* Reportersystem mit der Einzelzell-RNA-Sequenzierung kombiniert und ein maßgeschneidertes Protokoll, zur gleichzeitigen Gewinnung von RNA-Barcodes in voller Länge und des Transkriptoms aus hämatopoetischen Stammzellen, entwickelt. Die Analyse von Barcode-markierten Mäusen mit diesem Protokoll ergab, dass hämatopoetische Stammzellen abhängig von ihrem

Differenzierungsschicksal verschiedene Regionen einer Computer-generierten Karte des Transkriptoms besetzen und daher zumindest teilweise unterschiedliche Genexpressions-signaturen aufweisen. Diese Schicksals-assoziierten Signaturen waren bei der alleinigen Analyse des Transkriptoms nicht erkennbar, was die Bedeutung der Gewinnung von Schicksalsinformationen in Verbindung mit der Transkriptomanalyse unterstreicht. Diese Daten könnten für zukünftige Studien eine wertvolle Ressource darstellen, da sie auch Transkripte enthalten, deren Funktion in Stammzellen bislang unbekannt ist. Zusammengefasst enthüllte das *PolyloxExpress* Barcode System kohärente hämatopoetische Stammzellschicksale, die zell-intrinsisch determiniert sind und um den Zeitpunkt der Entstehung der hämatopoetischen Stammzellen im Embryo entstehen. Schicksals-assoziierte Transkriptom-Signaturen konnten identifiziert werden. Zukünftige Arbeiten, bei denen das *PolyloxExpress* Reportersystem mit neuesten genomischen Einzelzelltechniken kombiniert wird, werden weitere tiefe Einblicke in molekulare Aspekte der Bestimmung des Zellschicksals ermöglichen.

7 References List

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 (6774)*, 193-197, doi: 10.1038/35004599.

Alemany, A., Florescu, M., Baron, C. S., Peterson-Maduro, J. and van Oudenaarden, A. (2018). **Whole-organism clone tracing using single-cell sequencing**. Nature *556 (7699)*, 108-112, doi: 10.1038/nature25969.

Argelaguet, R., Clark, S. J., Mohammed, H., Stapel, L. C., Krueger, C., Kapourani, C. A., Imaz-Rosshandler, I., Lohoff, T., Xiang, Y., Hanna, C. W., Smallwood, S., Ibarra-Soria, X., Buettner, F., Sanguinetti, G., Xie, W., Krueger, F., Gottgens, B., Rugg-Gunn, P. J., Kelsey, G., Dean, W., Nichols, J., Stegle, O., Marioni, J. C. and Reik, W. (2019). **Multi-omics profiling of mouse gastrulation at single-cell resolution**. Nature *576 (7787)*, 487-491, doi: 10.1038/s41586-019-1825-8.

Balazs, A. B., Fabian, A. J., Esmon, C. T. and Mulligan, R. C. (2006). **Endothelial protein C receptor (CD201) explicitly identifies hematopoietic stem cells in murine bone marrow**. Blood *107 (6)*, 2317-2321, doi: 10.1182/blood-2005-06-2249.

Baron, C. S. and van Oudenaarden, A. (2019). **Unravelling cellular relationships during development and regeneration using genetic lineage tracing**. Nat Rev Mol Cell Biol *20 (12)*, 753-765, doi: 10.1038/s41580-019-0186-3.

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 (4866)*, 452-454, doi: 10.1038/197452a0.

Benz, C., Copley, M. R., Kent, D. G., Wohrer, S., Cortes, A., Aghaeepour, N., Ma, E., Mader, H., Rowe, K., Day, C., Treloar, D., Brinkman, R. R. and Eaves, C. J. (2012). **Hematopoietic stem cell subtypes expand differentially during development and display distinct lymphopoietic programs**. Cell Stem Cell *10 (3)*, 273-283, doi: 10.1016/j.stem.2012.02.007.

Biddy, B. A., Kong, W., Kamimoto, K., Guo, C., Waye, S. E., Sun, T. and Morris, S. A. (2018). **Single-cell mapping of lineage and identity in direct reprogramming**. Nature *564 (7735)*, 219-224, doi: 10.1038/s41586-018-0744-4.

Bowling, S., Sritharan, D., Osorio, F. G., Nguyen, M., Cheung, P., Rodriguez-Fraticelli, A., Patel, S., Yuan, W. C., Fujiwara, Y., Li, B. E., Orkin, S. H., Hormoz, S. and Camargo, F. D. (2020). **An Engineered CRISPR-Cas9 Mouse Line for Simultaneous Readout of Lineage Histories and Gene Expression Profiles in Single Cells**. Cell *181 (6)*, 1410-1422.e1427, doi: 10.1016/j.cell.2020.04.048.

Buenrostro, J. D., Corces, M. R., Lareau, C. A., Wu, B., Schep, A. N., Aryee, M. J., Majeti, R., Chang, H. Y. and Greenleaf, W. J. (2018). **Integrated Single-Cell Analysis Maps the Continuous Regulatory Landscape of Human Hematopoietic Differentiation**. Cell *173 (6)*, 1535-1548 e1516, doi: 10.1016/j.cell.2018.03.074.

Busch, K., Klapproth, K., Barile, M., Flossdorf, M., Holland-Letz, T., Schlenner, S. M., Reth, M., Höfer, T. and Rodewald, H. R. (2015). **Fundamental properties of unperturbed haematopoiesis from stem cells in vivo**. Nature *518 (7540)*, 542-546, doi: 10.1038/nature14242.

Busch, K. and Rodewald, H. R. (2016). **Unperturbed vs. post-transplantation hematopoiesis: both in vivo but different**. Curr Opin Hematol *23 (4)*, 295-303, doi: 10.1097/MOH.0000000000000250.

Cabezas-Wallscheid, N., Buettner, F., Sommerkamp, P., Klimmeck, D., Ladel, L., Thalheimer, F. B., Pastor-Flores, D., Roma, L. P., Renders, S., Zeisberger, P., Przybylla, A., Schonberger, K., Scognamiglio, R., Altamura, S., Florian, C. M., Fawaz, M., Vonficht, D., Tesio, M., Collier, P., Pavlinic, D., Geiger, H., Schroeder, T., Benes, V., Dick, T. P., Rieger, M. A., Stegle, O. and Trumpp, A. (2017). **Vitamin A-Retinoic Acid Signaling Regulates Hematopoietic Stem Cell Dormancy**. Cell *169 (5)*, 807-823 e819, doi: 10.1016/j.cell.2017.04.018.

Cabezas-Wallscheid, N., Klimmeck, D., Hansson, J., Lipka, D. B., Reyes, A., Wang, Q., Weichenhan, D., Lier, A., von Paleske, L., Renders, S., Wunsche, P., Zeisberger, P., Brocks, D., Gu, L., Herrmann, C., Haas, S., Essers, M. A. G., Brors, B., Eils, R., Huber, W., Milsom, M. D., Plass, C., Krijgsveld, J. and Trumpp, A. (2014). **Identification of regulatory networks in HSCs and their immediate progeny via integrated proteome, transcriptome, and DNA methylome analysis**. Cell Stem Cell *15 (4)*, 507-522, doi: 10.1016/j.stem.2014.07.005.

Calvanese, V., Nguyen, A. T., Bolan, T. J., Vavilina, A., Su, T., Lee, L. K., Wang, Y., Lay, F. D., Magnusson, M., Crooks, G. M., Kurdistani, S. K. and Mikkola, H. K. A. (2019). **MLLT3 governs human haematopoietic stem-cell self-renewal and engraftment**. Nature *576 (7786)*, 281-286, doi: 10.1038/s41586-019-1790-2.

Carrelha, J., Meng, Y., Kettyle, L. M., Luis, T. C., Norfo, R., Alcolea, V., Boukarabila, H., Grasso, F., Gambardella, A., Grover, A., Högstrand, K., Lord, A. M., Sanjuan-Pla, A., Woll, P. S., Nerlov, C. and Jacobsen, S. E. W. (2018). **Hierarchically related lineage-restricted fates of multipotent haematopoietic stem cells**. Nature *554 (7690)*, 106-111, doi: 10.1038/nature25455.

Chan, M. M., Smith, Z. D., Grosswendt, S., Kretzmer, H., Norman, T. M., Adamson, B., Jost, M., Quinn, J. J., Yang, D., Jones, M. G., Khodaverdian, A., Yosef, N., Meissner, A. and Weissman, J. S. (2019). **Molecular recording of mammalian embryogenesis**. Nature *570 (7759)*, 77-82, doi: 10.1038/s41586-019-1184-5.

Chen, J. Y., Miyanishi, M., Wang, S. K., Yamazaki, S., Sinha, R., Kao, K. S., Seita, J., Sahoo, D., Nakauchi, H. and Weissman, I. L. (2016). **Hoxb5 marks long-term haematopoietic stem cells and reveals a homogenous perivascular niche**. Nature *530 (7589)*, 223-227, doi: 10.1038/nature16943.

Chen, W., Guillaume-Gentil, O., Rainer, P. Y., Gäbelein, C. G., Saelens, W., Gardeux, V., Klaeger, A., Dainese, R., Zachara, M., Zambelli, T., Vorholt, J. A. and Deplancke, B. (2022). **Live-seq enables temporal transcriptomic recording of single cells**. Nature *608 (7924)*, 733-740, doi: 10.1038/s41586-022-05046-9.

Christodoulou, C., Spencer, J. A., Yeh, S. A., Turcotte, R., Kokkaliaris, K. D., Panero, R., Ramos, A., Guo, G., Seyedhassantehrani, N., Esipova, T. V., Vinogradov, S. A., Rudzinskas, S., Zhang, Y., Perkins, A. S., Orkin, S. H., Calogero, R. A., Schroeder, T., Lin, C. P. and Camargo, F. D. (2020). **Live-animal imaging of native haematopoietic stem and progenitor cells**. Nature *578 (7794)*, 278-283, doi: 10.1038/s41586-020-1971-z.

Dick, J. E., Magli, M. C., Huszar, D., Phillips, R. A. and Bernstein, A. (1985). **Introduction of a selectable gene into primitive stem cells capable of long-term reconstitution of the hemopoietic system of W/Wv mice**. Cell *42 (1)*, 71-79, doi: https://doi.org/10.1016/S0092-8674(85)80102-1.

Dignum, T., Varnum-Finney, B., Srivatsan, S. R., Dozono, S., Waltner, O., Heck, A. M., Ishida, T., Nourigat-McKay, C., Jackson, D. L., Rafii, S., Trapnell, C., Bernstein, I. D. and Hadland, B. (2021). **Multipotent progenitors and hematopoietic stem cells arise independently from hemogenic endothelium in the mouse embryo**. Cell Rep *36 (11)*, 109675, doi: 10.1016/j.celrep.2021.109675.

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 (2)*, 218-229, doi: 10.1016/j.stem.2007.05.015.

Fennell, K. A., Vassiliadis, D., Lam, E. Y. N., Martelotto, L. G., Balic, J. J., Hollizeck, S., Weber, T. S., Semple, T., Wang, Q., Miles, D. C., MacPherson, L., Chan, Y. C., Guirguis, A. A., Kats, L. M., Wong, E. S., Dawson, S. J., Naik, S. H. and Dawson, M. A. (2022). **Non-genetic determinants of malignant clonal fitness at single-cell resolution**. Nature *601 (7891)*, 125-131, doi: 10.1038/s41586-021-04206-7.

Frieda, K. L., Linton, J. M., Hormoz, S., Choi, J., Chow, K. K., Singer, Z. S., Budde, M. W., Elowitz, M. B. and Cai, L. (2017). **Synthetic recording and in situ readout of lineage information in single cells**. Nature *541 (7635)*, 107-111, doi: 10.1038/nature20777.

Gawad, C., Koh, W. and Quake, S. R. (2016). **Single-cell genome sequencing: current state of the science**. Nat Rev Genet *17 (3)*, 175-188, doi: 10.1038/nrg.2015.16.

Gazit, R., Mandal, P. K., Ebina, W., Ben-Zvi, A., Nombela-Arrieta, C., Silberstein, L. E. and Rossi, D. J. (2014). **Fgd5 identifies hematopoietic stem cells in the murine bone marrow**. J Exp Med *211 (7)*, 1315-1331, doi: 10.1084/jem.20130428.

Gekas, C., Dieterlen-Lievre, F., Orkin, S. H. and Mikkola, H. K. (2005). **The placenta is a niche for hematopoietic stem cells**. Dev Cell *8 (3)*, 365-375, doi: 10.1016/j.devcel.2004.12.016.

Giladi, A., Paul, F., Herzog, Y., Lubling, Y., Weiner, A., Yofe, I., Jaitin, D., Cabezas-Wallscheid, N., Dress, R., Ginhoux, F., Trumpp, A., Tanay, A. and Amit, I. (2018). **Single-cell characterization of haematopoietic progenitors and their trajectories in homeostasis and perturbed haematopoiesis**. Nat Cell Biol *20 (7)*, 836-846, doi: 10.1038/s41556-018-0121-4.

Giustacchini, A., Thongjuea, S., Barkas, N., Woll, P. S., Povinelli, B. J., Booth, C. A. G., Sopp, P., Norfo, R., Rodriguez-Meira, A., Ashley, N., Jamieson, L., Vyas, P., Anderson, K., Segerstolpe, A., Qian, H., Olsson-Stromberg, U., Mustjoki, S., Sandberg, R., Jacobsen, S. E. W. and Mead, A. J. (2017). **Single-cell transcriptomics uncovers distinct molecular signatures of stem cells in chronic myeloid leukemia**. Nat Med *23 (6)*, 692-702, doi: 10.1038/nm.4336.

Goodell, M. A., Brose, K., Paradis, G., Conner, A. S. and Mulligan, R. C. (1996). **Isolation and functional properties of murine hematopoietic stem cells that are replicating in vivo**. J Exp Med *183 (4)*, 1797-1806, doi: 10.1084/jem.183.4.1797.

Haghverdi, L., Buttner, M., Wolf, F. A., Buettner, F. and Theis, F. J. (2016). **Diffusion pseudotime robustly reconstructs lineage branching**. Nat Methods *13 (10)*, 845-848, doi: 10.1038/nmeth.3971.

Höfer, T., Busch, K., Klapproth, K. and Rodewald, H. R. (2016). **Fate Mapping and Quantitation of Hematopoiesis In Vivo**. Annu Rev Immunol *34*, 449-478, doi: 10.1146/annurev-immunol-032414-112019.

Huang, L., Ma, F., Chapman, A., Lu, S. and Xie, X. S. (2015). **Single-Cell Whole-Genome Amplification and Sequencing: Methodology and Applications**. Annu Rev Genomics Hum Genet *16*, 79-102, doi: 10.1146/annurev-genom-090413-025352.

Ito, K., Yamazaki, S., Yamamoto, R., Tajima, Y., Yanagida, A., Kobayashi, T., Kato-Itoh, M., Kakuta, S., Iwakura, Y., Nakauchi, H. and Kamiya, A. (2014). **Gene Targeting Study Reveals Unexpected Expression of Brain-expressed X-linked 2 in Endocrine and Tissue Stem/Progenitor Cells in Mice***. Journal of Biological Chemistry *289 (43)*, 29892-29911, doi: https://doi.org/10.1074/jbc.M114.580084.

Jacobsen, S. E. W. and Nerlov, C. (2019). **Haematopoiesis in the era of advanced single-cell technologies**. Nature Cell Biology *21 (1)*, 2-8, doi: 10.1038/s41556-018-0227-8.

Kalhor, R., Kalhor, K., Mejia, L., Leeper, K., Graveline, A., Mali, P. and Church, G. M. (2018). **Developmental barcoding of whole mouse via homing CRISPR**. Science *361 (6405)*, eaat9804, doi: 10.1126/science.aat9804.

Kebschull, J. M. and Zador, A. M. (2018). **Cellular barcoding: lineage tracing, screening and beyond**. Nat Methods *15 (11)*, 871-879, doi: 10.1038/s41592-018-0185-x.

Kent, D. G., Copley, M. R., Benz, C., Wohrer, S., Dykstra, B. J., Ma, E., Cheyne, J., Zhao, Y., Bowie, M. B., Zhao, Y., Gasparetto, M., Delaney, A., Smith, C., Marra, M. and Eaves, C. J. (2009). **Prospective isolation and molecular characterization of hematopoietic stem cells with durable self-renewal potential**. Blood *113 (25)*, 6342-6350, doi: 10.1182/blood-2008-12-192054.

Kester, L. and van Oudenaarden, A. (2018). **Single-Cell Transcriptomics Meets Lineage Tracing**. Cell Stem Cell *23 (2)*, 166-179, doi: 10.1016/j.stem.2018.04.014.

Kiel, M. J., Yilmaz, Ö. H., Iwashita, T., Yilmaz, O. H., Terhorst, C. and Morrison, S. J. (2005). **SLAM Family Receptors Distinguish Hematopoietic Stem and Progenitor Cells and Reveal Endothelial Niches for Stem Cells**. Cell *121 (7)*, 1109-1121, doi: 10.1016/j.cell.2005.05.026.

Klein, A. M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D. A. and Kirschner, M. W. (2015). **Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells**. Cell *161 (5)*, 1187-1201, doi: 10.1016/j.cell.2015.04.044.

Kondo, M., Weissman, I. L. and Akashi, K. (1997). **Identification of clonogenic common lymphoid progenitors in mouse bone marrow**. Cell *91 (5)*, 661-672, doi: 10.1016/s0092-8674(00)80453-5.

Kowalczyk, M. S., Tirosh, I., Heckl, D., Rao, T. N., Dixit, A., Haas, B. J., Schneider, R. K., Wagers, A. J., Ebert, B. L. and Regev, A. (2015). **Single-cell RNA-seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells**. Genome Research *25 (12)*, 1860-1872, doi: 10.1101/gr.192237.115.

Krause, D. S., Ito, T., Fackler, M. J., Smith, O. M., Collector, M. I., Sharkis, S. J. and May, W. S. (1994). **Characterization of Murine CD34, a Marker for Hematopoietic Progenitor and Stem Cells**. Blood *84 (3)*, 691-701, doi: https://doi.org/10.1182/blood.V84.3.691.691.

Laurenti, E. and Göttgens, B. (2018). **From haematopoietic stem cells to complex differentiation landscapes**. Nature *553 (7689)*, 418-426, doi: 10.1038/nature25022.

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., Bernstein, I. D. and Trumpp, A. (2008). **Hematopoietic Stem Cell Function and Survival Depend on c-Myc and N-Myc Activity**. Cell Stem Cell *3 (6)*, 611-624, doi: https://doi.org/10.1016/j.stem.2008.09.005.

Lemischka, I. R., Raulet, D. H. and Mulligan, R. C. (1986). **Developmental potential and dynamic behavior of hematopoietic stem cells**. Cell *45 (6)*, 917-927, doi: 10.1016/0092-8674(86)90566-0.

Liggett, L. A. and Sankaran, V. G. (2020). **Unraveling Hematopoiesis through the Lens of Genomics**. Cell *182 (6)*, 1384-1400, doi: 10.1016/j.cell.2020.08.030.

Lu, R., Neff, N. F., Quake, S. R. and Weissman, I. L. (2011). **Tracking single hematopoietic stem cells in vivo using high-throughput sequencing in conjunction with viral genetic barcoding**. Nat Biotechnol *29 (10)*, 928-933, doi: 10.1038/nbt.1977.

Lymperi, S., Ferraro, F. and Scadden, D. T. (2010). **The HSC niche concept has turned 31. Has our knowledge matured?** Ann N Y Acad Sci *1192*, 12-18, doi: 10.1111/j.1749-6632.2009.05223.x.

McKenna, A., Findlay, G. M., Gagnon, J. A., Horwitz, M. S., Schier, A. F. and Shendure, J. (2016). **Whole-organism lineage tracing by combinatorial and cumulative genome editing**. Science *353 (6298)*, aaf7907, doi: 10.1126/science.aaf7907.

Mikkola, H. K. A. and Orkin, S. H. (2006). **The journey of developing hematopoietic stem cells**. Development *133 (19)*, 3733-3744, doi: 10.1242/dev.02568.

Mimitou, E. P., Cheng, A., Montalbano, A., Hao, S., Stoeckius, M., Legut, M., Roush, T., Herrera, A., Papalexi, E., Ouyang, Z., Satija, R., Sanjana, N. E., Koralov, S. B. and Smibert, P. (2019). **Multiplexed detection of proteins, transcriptomes, clonotypes and CRISPR perturbations in single cells**. Nat Methods *16 (5)*, 409-412, doi: 10.1038/s41592-019-0392-0.

Mimitou, E. P., Lareau, C. A., Chen, K. Y., Zorzetto-Fernandes, A. L., Hao, Y., Takeshima, Y., Luo, W., Huang, T. S., Yeung, B. Z., Papalexi, E., Thakore, P. I., Kibayashi, T., Wing, J. B., Hata, M., Satija, R., Nazor, K. L., Sakaguchi, S., Ludwig, L. S., Sankaran, V. G., Regev, A. and Smibert, P. (2021). **Scalable, multimodal profiling of chromatin accessibility, gene expression and protein levels in single cells**. Nat Biotechnol *39 (10)*, 1246-1258, doi: 10.1038/s41587-021-00927-2.

Morcos, M. N. F., Li, C., Munz, C. M., Greco, A., Dressel, N., Reinhardt, S., Sameith, K., Dahl, A., Becker, N. B., Roers, A., Höfer, T. and Gerbaulet, A. (2022). **Fate mapping of**

hematopoietic stem cells reveals two pathways of native thrombopoiesis. Nat Commun *13 (1)*, 4504, doi: 10.1038/s41467-022-31914-z.

Morita, Y., Ema, H. and Nakauchi, H. (2010). **Heterogeneity and hierarchy within the most primitive hematopoietic stem cell compartment**. J Exp Med *207 (6)*, 1173-1182, doi: 10.1084/jem.20091318.

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 (7448)*, 239-243, doi: 10.1038/nature12026.

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 (11)*, 4111-4118, doi: 10.1182/blood-2003-10-3448.

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 (4)*, 1302-1309.

Mullighan, C. G., Kennedy, A., Zhou, X., Radtke, I., Phillips, L. A., Shurtleff, S. A. and Downing, J. R. (2007). **Pediatric acute myeloid leukemia with NPM1 mutations is characterized by a gene expression profile with dysregulated HOX gene expression distinct from MLL-rearranged leukemias**. Leukemia *21 (9)*, 2000-2009, doi: 10.1038/sj.leu.2404808.

Nam, A. S., Kim, K. T., Chaligne, R., Izzo, F., Ang, C., Taylor, J., Myers, R. M., Abu-Zeinah, G., Brand, R., Omans, N. D., Alonso, A., Sheridan, C., Mariani, M., Dai, X., Harrington, E., Pastore, A., Cubillos-Ruiz, J. R., Tam, W., Hoffman, R., Rabadan, R., Scandura, J. M., Abdel-Wahab, O., Smibert, P. and Landau, D. A. (2019). **Somatic mutations and cell identity linked by Genotyping of Transcriptomes**. Nature *571 (7765)*, 355-360, doi: 10.1038/s41586-019-1367-0.

Nestorowa, S., Hamey, F. K., Pijuan Sala, B., Diamanti, E., Shepherd, M., Laurenti, E., Wilson, N. K., Kent, D. G. and Gottgens, B. (2016). **A single-cell resolution map of mouse hematopoietic stem and progenitor cell differentiation**. Blood *128 (8)*, e20-31, doi: 10.1182/blood-2016-05-716480.

Notta, F., Zandi, S., Takayama, N., Dobson, S., Gan, O. I., Wilson, G., Kaufmann, K. B., McLeod, J., Laurenti, E., Dunant, C. F., McPherson, J. D., Stein, L. D., Dror, Y. and Dick, J. E. (2016). **Distinct routes of lineage development reshape the human blood hierarchy across ontogeny**. Science *351 (6269)*, aab2116, doi: 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 (1)*, 102-116, doi: 10.1016/j.stem.2013.05.014.

Olsson, A., Venkatasubramanian, M., Chaudhri, V. K., Aronow, B. J., Salomonis, N., Singh, H. and Grimes, H. L. (2016). **Single-cell analysis of mixed-lineage states leading to a binary cell fate choice**. Nature *537 (7622)*, 698-702, doi: 10.1038/nature19348.

Osawa, M., Hanada, K. I., Hamada, H. and Nakauchi, H. (1996). **Long-Term Lymphohematopoietic Reconstitution by a Single CD34-Low/Negative Hematopoietic Stem Cell**. Science *273 (5272)*, 242-245, doi: 10.1126/science.273.5272.242.

Ottersbach, K. and Dzierzak, E. (2005). **The murine placenta contains hematopoietic stem cells within the vascular labyrinth region**. Dev Cell *8 (3)*, 377-387, doi: 10.1016/j.devcel.2005.02.001.

Patel, S. H., Christodoulou, C., Weinreb, C., Yu, Q., da Rocha, E. L., Pepe-Mooney, B. J., Bowling, S., Li, L., Osorio, F. G., Daley, G. Q. and Camargo, F. D. (2022). **Lifelong multilineage contribution by embryonic-born blood progenitors**. Nature *606 (7915)*, 747-753, doi: 10.1038/s41586-022-04804-z.

Paul, F., Arkin, Y., Giladi, A., Jaitin, D. A., Kenigsberg, E., Keren-Shaul, H., Winter, D., Lara-Astiaso, D., Gury, M., Weiner, A., David, E., Cohen, N., Lauridsen, F. K., Haas, S., Schlitzer, A., Mildner, A., Ginhoux, F., Jung, S., Trumpp, A., Porse, B. T., Tanay, A. and Amit, I. (2015). **Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors**. Cell *163 (7)*, 1663-1677, doi: 10.1016/j.cell.2015.11.013.

Pei, W., Feyerabend, T. B., Rössler, J., Wang, X., Postrach, D., Busch, K., Rode, I., Klapproth, K., Dietlein, N., Quedenau, C., Chen, W., Sauer, S., Wolf, S., Höfer, T. and Rodewald, H.-R. (2017). **Polylox barcoding reveals haematopoietic stem cell fates realized in vivo**. Nature *548 (7668)*, 456-460, doi: 10.1038/nature23653.

Pei, W., Shang, F., Wang, X., Fanti, A. K., Greco, A., Busch, K., Klapproth, K., Zhang, Q., Quedenau, C., Sauer, S., Feyerabend, T. B., Höfer, T. and Rodewald, H. R. (2020). **Resolving Fates and Single-Cell Transcriptomes of Hematopoietic Stem Cell Clones by PolyloxExpress Barcoding**. Cell Stem Cell *27 (3)*, 383-395.e388, doi: 10.1016/j.stem.2020.07.018.

Pei, W., Wang, X., Rössler, J., Feyerabend, T. B., Höfer, T. and Rodewald, H.-R. (2019). **Using Cre-recombinase-driven Polylox barcoding for in vivo fate mapping in mice**. Nature Protocols *14 (6)*, 1820-1840, doi: 10.1038/s41596-019-0163-5.

Picelli, S., Bjorklund, A. K., Faridani, O. R., Sagasser, S., Winberg, G. and Sandberg, R. (2013). **Smart-seq2 for sensitive full-length transcriptome profiling in single cells**. Nat Methods *10 (11)*, 1096-1098, doi: 10.1038/nmeth.2639.

Picelli, S., Faridani, O. R., Bjorklund, A. K., Winberg, G., Sagasser, S. and Sandberg, R. (2014). **Full-length RNA-seq from single cells using Smart-seq2**. Nat Protoc *9 (1)*, 171-181, doi: 10.1038/nprot.2014.006.

Pietras, E. M., Reynaud, D., Kang, Y. A., Carlin, D., Calero-Nieto, F. J., Leavitt, A. D., Stuart, J. M., Gottgens, B. and Passegue, E. (2015). **Functionally Distinct Subsets of Lineage-Biased Multipotent Progenitors Control Blood Production in Normal and Regenerative Conditions**. Cell Stem Cell *17 (1)*, 35-46, doi: 10.1016/j.stem.2015.05.003.

Raj, B., Wagner, D. E., McKenna, A., Pandey, S., Klein, A. M., Shendure, J., Gagnon, J. A. and Schier, A. F. (2018). **Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain**. Nat Biotechnol *36 (5)*, 442-450, doi: 10.1038/nbt.4103.

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 (7817)*, 585-589, doi: 10.1038/s41586-020-2503-6.

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 (7687)*, 212-216, doi: 10.1038/nature25168.

Sanjuan-Pla, A., Macaulay, I. C., Jensen, C. T., Woll, P. S., Luis, T. C., Mead, A., Moore, S., Carella, C., Matsuoka, S., Bouriez Jones, T., Chowdhury, O., Stenson, L., Lutteropp, M., Green, J. C., Facchini, R., Boukarabila, H., Grover, A., Gambardella, A., Thongjuea, S., Carrelha, J., Tarrant, P., Atkinson, D., Clark, S. A., Nerlov, C. and Jacobsen, S. E. (2013). **Platelet-biased stem cells reside at the apex of the haematopoietic stem-cell hierarchy**. Nature *502 (7470)*, 232-236, doi: 10.1038/nature12495.

Sawai, C. M., Babovic, S., Upadhaya, S., Knapp, D., Lavin, Y., Lau, C. M., Goloborodko, A., Feng, J., Fujisaki, J., Ding, L., Mirny, L. A., Merad, M., Eaves, C. J. and Reizis, B. (2016). **Hematopoietic Stem Cells Are the Major Source of Multilineage Hematopoiesis in Adult Animals**. Immunity *45 (3)*, 597-609, doi: 10.1016/j.immuni.2016.08.007.

Shang, F. and Rodewald, H. R. (2022). **Toward the dissection of hematopoietic stem cell fates and their determinants**. Curr Opin Genet Dev *75*, 101945, doi: 10.1016/j.gde.2022.101945.

Sommerkamp, P., Romero-Mulero, M. C., Narr, A., Ladel, L., Hustin, L., Schönberger, K., Renders, S., Altamura, S., Zeisberger, P., Jäcklein, K., Klimmeck, D., Rodriguez-Fraticelli,

A., Camargo, F. D., Perié, L., Trumpp, A. and Cabezas-Wallscheid, N. (2021). **Mouse multipotent progenitor 5 cells are located at the interphase between hematopoietic stem and progenitor cells**. Blood *137 (23)*, 3218-3224, doi: 10.1182/blood.2020007876.

Spangrude, G., Heimfeld, S. and Weissman, I. (1988). **Purification and characterization of mouse hematopoietic stem cells**. Science *241 (4861)*, 58-62, doi: 10.1126/science.2898810.

Spanjaard, B., Hu, B., Mitic, N., Olivares-Chauvet, P., Janjuha, S., Ninov, N. and Junker, J. P. (2018). **Simultaneous lineage tracing and cell-type identification using CRISPR-Cas9-induced genetic scars**. Nat Biotechnol *36 (5)*, 469-473, doi: 10.1038/nbt.4124.

Stoeckius, M., Hafemeister, C., Stephenson, W., Houck-Loomis, B., Chattopadhyay, P. K., Swerdlow, H., Satija, R. and Smibert, P. (2017). **Simultaneous epitope and transcriptome measurement in single cells**. Nature Methods *14 (9)*, 865-868, doi: 10.1038/nmeth.4380.

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 (7522)*, 322-327, doi: 10.1038/nature13824.

Tanay, A. and Regev, A. (2017). **Scaling single-cell genomics from phenomenology to mechanism**. Nature *541 (7637)*, 331-338, doi: 10.1038/nature21350.

Terszowski, G., Waskow, C., Conradt, P., Lenze, D., Koenigsmann, J., Carstanjen, D., Horak, I. and Rodewald, H.-R. (2005). **Prospective isolation and global gene expression analysis of the erythrocyte colony-forming unit (CFU-E)**. Blood *105 (5)*, 1937-1945, doi: https://doi.org/10.1182/blood-2004-09-3459.

Till, J. E. and Mc, C. E. (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., Fallahi-Sichani, M., Dutton-Regester, K., Lin, J.-R., Cohen, O., Shah, P., Lu, D., Genshaft, A. S., Hughes, T. K., Ziegler, C. G. K., Kazer, S. W., Gaillard, A., Kolb, K. E., Villani, A.-C., Johannessen, C. M., Andreev, A. Y., Van Allen, E. M., Bertagnolli, M., Sorger, P. K., Sullivan, R. J., Flaherty, K. T., Frederick, D. T., Jané-Valbuena, J., Yoon, C. H., Rozenblatt-Rosen, O., Shalek, A. K., Regev, A. and Garraway, L. A. (2016). **Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq**. Science *352 (6282)*, 189-196, doi: 10.1126/science.aad0501.

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N. J., Livak, K. J., Mikkelsen, T. S. and Rinn, J. L. (2014). **The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells**. Nat Biotechnol *32 (4)*, 381-386, doi: 10.1038/nbt.2859.

Triana, S., Vonficht, D., Jopp-Saile, L., Raffel, S., Lutz, R., Leonce, D., Antes, M., Hernandez-Malmierca, P., Ordonez-Rueda, D., Ramasz, B., Boch, T., Jann, J. C., Nowak, D., Hofmann, W. K., Muller-Tidow, C., Hubschmann, D., Alexandrov, T., Benes, V., Trumpp, A., Paulsen, M., Velten, L. and Haas, S. (2021). **Single-cell proteo-genomic reference maps of the hematopoietic system enable the purification and massive profiling of precisely defined cell states**. Nat Immunol *22 (12)*, 1577-1589, doi: 10.1038/s41590-021-01059-0.

Upadhaya, S., Sawai, C. M., Papalexi, E., Rashidfarrokhi, A., Jang, G., Chattopadhyay, P., Satija, R. and Reizis, B. (2018). **Kinetics of adult hematopoietic stem cell differentiation in vivo**. J Exp Med *215 (11)*, 2815-2832, doi: 10.1084/jem.20180136.

Velten, L., Haas, S. F., Raffel, S., Blaszkiewicz, S., Islam, S., Hennig, B. P., Hirche, C., Lutz, C., Buss, E. C., Nowak, D., Boch, T., Hofmann, W. K., Ho, A. D., Huber, W., Trumpp, A., Essers, M. A. and Steinmetz, L. M. (2017). **Human haematopoietic stem cell lineage commitment is a continuous process**. Nat Cell Biol *19 (4)*, 271-281, doi: 10.1038/ncb3493.

Verovskaya, E., Broekhuis, M. J., Zwart, E., Ritsema, M., van Os, R., de Haan, G. and Bystrykh, L. V. (2013). **Heterogeneity of young and aged murine hematopoietic stem cells revealed by quantitative clonal analysis using cellular barcoding**. Blood *122 (4)*, 523-532, doi: 10.1182/blood-2013-01-481135.

Wagner, D. E. and Klein, A. M. (2020). **Lineage tracing meets single-cell omics: opportunities and challenges**. Nature Reviews Genetics *21 (7)*, 410-427, doi: 10.1038/s41576-020-0223-2.

Walter, D., Lier, A., Geiselhart, A., Thalheimer, F. B., Huntscha, S., Sobotta, M. C., Moehrle, B., Brocks, D., Bayindir, I., Kaschutnig, P., Muedder, K., Klein, C., Jauch, A., Schroeder, T., Geiger, H., Dick, T. P., Holland-Letz, T., Schmezer, P., Lane, S. W., Rieger, M. A., Essers, M. A., Williams, D. A., Trumpp, A. and Milsom, M. D. (2015). **Exit from dormancy provokes DNA-damage-induced attrition in haematopoietic stem cells**. Nature *520 (7548)*, 549-552, doi: 10.1038/nature14131.

Wang, S.-W., Herriges, M. J., Hurley, K., Kotton, D. N. and Klein, A. M. (2022). **CoSpar identifies early cell fate biases from single-cell transcriptomic and lineage information**. Nature Biotechnology *40 (7)*, 1066-1074, doi: 10.1038/s41587-022-01209-1.

Waskow, C., Madan, V., Bartels, S., Costa, C., Blasig, R. and Rodewald, H.-R. (2009). **Hematopoietic stem cell transplantation without irradiation**. Nature Methods *6 (4)*, 267-269, doi: 10.1038/nmeth.1309.

Watcham, S., Kucinski, I. and Gottgens, B. (2019). **New insights into hematopoietic differentiation landscapes from single-cell RNA sequencing**. Blood *133 (13)*, 1415-1426, doi: 10.1182/blood-2018-08-835355.

Weinreb, C., Rodriguez-Fraticelli, A., Camargo, F. D. and Klein, A. M. (2020). **Lineage tracing on transcriptional landscapes links state to fate during differentiation**. Science *367 (6479)*, eaaw3381, doi: 10.1126/science.aaw3381.

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., Lio, P., Macdonald, H. R. and Trumpp, A. (2008). **Hematopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair**. Cell *135 (6)*, 1118-1129, doi: 10.1016/j.cell.2008.10.048.

Wilson, N. K., Kent, D. G., Buettner, F., Shehata, M., Macaulay, I. C., Calero-Nieto, F. J., Sanchez Castillo, M., Oedekoven, C. A., Diamanti, E., Schulte, R., Ponting, C. P., Voet, T., Caldas, C., Stingl, J., Green, A. R., Theis, F. J. and Gottgens, B. (2015). **Combined Single-Cell Functional and Gene Expression Analysis Resolves Heterogeneity within Stem Cell Populations**. Cell Stem Cell *16 (6)*, 712-724, doi: 10.1016/j.stem.2015.04.004.

Woodworth, M. B., Girskis, K. M. and Walsh, C. A. (2017). **Building a lineage from single cells: genetic techniques for cell lineage tracking**. Nat Rev Genet *18 (4)*, 230-244, doi: 10.1038/nrg.2016.159.

Yamamoto, R., Morita, Y., Ooehara, J., Hamanaka, S., Onodera, M., Rudolph, K. L., Ema, H. and Nakauchi, H. (2013). **Clonal analysis unveils self-renewing lineage-restricted progenitors generated directly from hematopoietic stem cells**. Cell *154 (5)*, 1112-1126, doi: 10.1016/j.cell.2013.08.007.

Yang, L., Bryder, D., Adolfsson, J., Nygren, J., Mansson, R., Sigvardsson, M. and Jacobsen, S. E. (2005). **Identification of Lin(-)Sca1(+)kit(+)CD34(+)Flt3- short-term hematopoietic stem cells capable of rapidly reconstituting and rescuing myeloablated transplant recipients**. Blood *105 (7)*, 2717-2723, doi: 10.1182/blood-2004-06-2159.

Yokomizo, T., Ideue, T., Morino-Koga, S., Tham, C. Y., Sato, T., Takeda, N., Kubota, Y., Kurokawa, M., Komatsu, N., Ogawa, M., Araki, K., Osato, M. and Suda, T. (2022). **Independent origins of fetal liver haematopoietic stem and progenitor cells**. Nature *609 (7928)*, 779-784, doi: 10.1038/s41586-022-05203-0.

Yu, V. W. C., Yusuf, R. Z., Oki, T., Wu, J., Saez, B., Wang, X., Cook, C., Baryawno, N., Ziller, M. J., Lee, E., Gu, H., Meissner, A., Lin, C. P., Kharchenko, P. V. and Scadden, D. T. (2016). **Epigenetic Memory Underlies Cell-Autonomous Heterogeneous Behavior of Hematopoietic Stem Cells**. Cell *167 (5)*, 1310-1322.e1317, doi: 10.1016/j.cell.2016.10.045.

Zhang, X., Li, T., Liu, F., Chen, Y., Yao, J., Li, Z., Huang, Y. and Wang, J. (2019). **Comparative Analysis of Droplet-Based Ultra-High-Throughput Single-Cell RNA-Seq Systems**. Molecular Cell *73 (1)*, 130-142.e135, doi: 10.1016/j.molcel.2018.10.020.

Zhang, Y. W., Mess, J., Aizarani, N., Mishra, P., Johnson, C., Romero-Mulero, M. C., Rettkowski, J., Schönberger, K., Obier, N., Jäcklein, K., Woessner, N. M., Lalioti, M. E., Velasco-Hernandez, T., Sikora, K., Wäsch, R., Lehnertz, B., Sauvageau, G., Manke, T., Menendez, P., Walter, S. G., Minguet, S., Laurenti, E., Günther, S., Grün, D. and Cabezas-Wallscheid, N. (2022). **Hyaluronic acid-GPRC5C signalling promotes dormancy in haematopoietic stem cells**. Nat Cell Biol *24 (7)*, 1038-1048, doi: 10.1038/s41556-022-00931-x.

Zheng, G. X., Terry, J. M., Belgrader, P., Ryvkin, P., Bent, Z. W., Wilson, R., Ziraldo, S. B., Wheeler, T. D., McDermott, G. P., Zhu, J., Gregory, M. T., Shuga, J., Montesclaros, L., Underwood, J. G., Masquelier, D. A., Nishimura, S. Y., Schnall-Levin, M., Wyatt, P. W., Hindson, C. M., Bharadwaj, R., Wong, A., Ness, K. D., Beppu, L. W., Deeg, H. J., McFarland, C., Loeb, K. R., Valente, W. J., Ericson, N. G., Stevens, E. A., Radich, J. P., Mikkelsen, T. S., Hindson, B. J. and Bielas, J. H. (2017). **Massively parallel digital transcriptional profiling of single cells**. Nat Commun *8*, 14049, doi: 10.1038/ncomms14049.

Ziegenhain, C., Vieth, B., Parekh, S., Reinius, B., Guillaumet-Adkins, A., Smets, M., Leonhardt, H., Heyn, H., Hellmann, I. and Enard, W. (2017). **Comparative Analysis of Single-Cell RNA Sequencing Methods**. Mol Cell *65 (4)*, 631-643 e634, doi: 10.1016/j.molcel.2017.01.023.

Zilionis, R., Nainys, J., Veres, A., Savova, V., Zemmour, D., Klein, A. M. and Mazutis, L. (2017). **Single-cell barcoding and sequencing using droplet microfluidics**. Nature Protocols *12 (1)*, 44-73, doi: 10.1038/nprot.2016.154.

8 Personal Publications

For section 3.1.1-3.1.2, I designed and performed the experiments with feedback and intellectual input from Weike Pei, Thorsten Feyerabend, Xi Wang and Qin Zhang. Pacbio sequencing was performed by Claudia Quedenau and Sascha Sauer. Xi Wang performed sequencing related bioinformatic calculation. For section 3.1.3, I designed and performed all the experiments, conducted Pacbio and Illumina sequencing, analyzed the data and interpreted the results, with the help from Xi Wang for sequencing related bioinformatic calculation. For section 3.2, I designed and performed all the experiments, conducted Pacbio sequencing, analyzed the data and interpreted the results, with help from Sven Schäfer for blood collection, Thorsten Feyerabend for cell sorting, Larissa Johanna Frank for HSC transplantation and Xi Wang for sequencing related bioinformatic calculation. For section 3.3, I induced mice with tamoxifen, harvested organs, stained cells for sorting, performed scRNAseq and nested PCR for all experiments. Weike Pei helped with cell staining and performed DNA barcode PCR on bulk sorted cells. Thorsten Feyerabend performed cell sorting. Katrin Busch and Ann-Kathrin Fanti helped with mice timed mating organization and planning. Claudia Quedenau and Sascha Sauer performed Pacbio sequencing. DKFZ genomics and proteomics core facility performed Illumina sequencing. Xi Wang performed bioinformatic calculation related to sequencing, with help from Alessandro Greco. All the work in this dissertation was supervised by Hans-Reimer Rodewald, with collaborative work with Thomas Höfer.

Partial of this dissertation have already been published in the following articles:

1. Pei, W.*, **Shang, F.***, Wang, X.*, Fanti, A. K., Greco, A., Busch, K., Klapproth, K., Zhang, Q., Quedenau, C., Sauer, S., Feyerabend, T. B., Höfer, T.# and Rodewald, H. R.# (2020). **Resolving Fates and Single-Cell Transcriptomes of Hematopoietic Stem Cell Clones by PolyloxExpress Barcoding**. Cell Stem Cell *27 (3)*, 383-395.e388, doi: 10.1016/j.stem.2020.07.018.
** Equal contributions, # Corresponding authors*

2. **Shang, F.** and Rodewald, H. R. (2022). **Toward the dissection of hematopoietic stem cell fates and their determinants**. Curr Opin Genet Dev *75*, 101945, doi: 10.1016/j.gde.2022.101945.

Publication 1 is based on the results in section 3.1.1-3.1.2 and 3.3. The discussions about *PolyloxExpress*, HSC fate heterogeneity and stability (Section 4.1, 4.2 and 4.4) are also depicted in this publication. I contributed to the experimental design, data collection and interpretation, manuscript writing and peer-review process of this publication.

Publication 2 is a review article and its contents are briefly incorporated into the discussion of this dissertation (Section 4). My personal contribution to this publication consists of the research of the background literature, draft manuscript writing and peer-review process of this publication.

1. Dietlein, N., Wang, X., Metz, J., Disson, O., **Shang, F.**, Beyersdörffer, C., Rodríguez Correa, E., Lipka, D. B., Begus-Nahrmann, Y., Kosinsky, R. L., Johnsen, S. A., Lecuit, M., Höfer, T. and Rodewald, H. R. (2022). **Usp22 is an intracellular regulator of systemic emergency hematopoiesis**. Science Immunology 7 (78), eabq2061, doi: 10.1126/sciimmunol.abq2061.

2. Fanti, A. K., Busch, K., Greco, A., Wang, X., Cirovic, B., **Shang, F.**, Nizharadze, T., Frank, L., Barile, M., Feyerabend, T. B., Höfer, T. and Rodewald, H. R. (2023). **Flt3- and Tie2- Cre tracing identifies regeneration in sepsis from multipotent progenitors but not hematopoietic stem cells**. Cell Stem Cell, accepted for publication

3. Plum, T., Binzberger, R., Stakenborg, N., Fung, C., Tappe-Theodor, A., Wang, Z., Fortea, M., Postrach, D., **Shang, F.**, Thiele, R., Poth, T., MacLaren, D., Boeckxstaens, G., Vanden Berghe, P., Kuner, R., Tanaka, S., Voehringer, D., Strid, J., Monyer, H., Feyerabend, T. B. and Rodewald, H. R. **Mast cells link immune sensing to antigen avoidance behavior**. Manuscript submitted for publication.

4. Ma, S.*, Sandhoff, R.*, Luo, X.*, **Shang, F.***, Li, Z., Wu, J., Schwarz, F., Ming, Y., Madi, A., Weisshaar, N., Mieg, A., Hering, M., Mohr, K., Ten Bosch, N., Li, Z., Poschet, G., Rodewald, H. R., Wang, X.#, Gao, P.# and Cui, G.#. **Serine enrichment in tumor promotes regulatory T cell accumulation through sphinganine**. Manuscript submitted for publication.
** Equal contributions, # Corresponding authors*

Curriculum Vitae

Personal Information

Name:	Fuwei Shang
Date of Birth:	08.11.1994
Place of Birth:	Kunming City, Yunnan Province, People's Republic of China
Nationality:	Chinese
Marital Status:	Single

Primary and Secondary Schools

2000 – 2006	Primary School Affiliated to Yunnan Agricultural University
2006 – 2009	Yunnan University Secondary School
2009 – 2012	High School Affiliated to Yunnan Normal University
07.06.2012-08.06.2012	National College Entrance Examination Exam Result: 641/750 (rank 170/210,000, top 0.8‰)

Universities

2012 – 2017 Shanghai Jiao Tong University School of Medicine

- Bachelor in Clinical Medicine
- *Graduation with honor: Outstanding College Graduate, Shanghai Ministry of Education (Highest honor for undergraduate)*

2017 – 2018 Heidelberg University

- Master of science in Translational Medical Research

2018 – 2023 Heidelberg University

- PhD (Dr. sc. hum), Division of Cellular Immunology, German Cancer Research Center (DKFZ)
- *PhD fellowship awarded by the Helmholtz International Graduate School of Cancer Research (2018)*

Professional Experience

06.2013 – 08.2013	University of Sydney - Exchange student for medical study
06.2016 – 08.2016	Seoul National University College of Medicine/ Seoul National University Hospital (SNUH) - Medical internship: General Surgery, Orthopedics, Cardiothoracic Surgery
03.2016 – 05.2017	Shanghai Renji Hospital / Shanghai Children's Medical Center - Medical internship: Internal medicine, Surgery, Obstetrics and Gynecology, Pediatrics
03.2018 – 09.2018	Heidelberg University - Master thesis, Division of Vascular Oncology and Metastasis, German Cancer Research Center (DKFZ)
09.2019	20th International Summer School on Immunology - Poster presentation
09.2022	International Society of Experimental Hematology (ISEH) - Poster presentation at 51st Annual Scientific Meeting
11.2022	1st International School on Advanced Immunology - Poster and oral presentation - *Scholarship awarded by Deutsche Gesellschaft für Immunologie (DGfI)*

Acknowledgments

Giving up the opportunity for medical residency training and pursuing for scientific training instead was a difficult decision for me to make. Luckily, my PhD journey was inspired and supported by so many great people throughout the way, and I'm more than grateful to thank all of them.

My deepest gratitude goes to Prof. Dr. Hans-Reimer Rodewald. It is my great honor to get the chance of joining his lab and work on the interesting barcoding project. Hans-Reimer created an encouraging environment in the lab so that I could have the freedom to explore science while, in the same time, having his timely supervision. I was also constantly encouraged by his enthusiasm in science and inspired by his critical thinking and clear judgement. These qualities guided me throughout the PhD and shaped my thoughts and behaviors as a young scientist.

I would also like to show my faithful gratitude to Dr. Thorsten Feyerabend for his patient guidance and continuous support along the way. The *PolyloxExpress* project could have been much more difficult without his design and input for experiments. It is said that details determine success or failure. There were many times when I overlooked important details in experiments while Thorsten was the one who reminded me and helped me. This thesis also benefited and improved a lot from his criticisms and corrections.

I was so lucky to have a period of my PhD working with Dr. Weike Pei and Dr. Xi Wang together. They were not only collaborators but also friends to me. I was constantly encouraged by the discussions we had about science. There were many times when the experiments had to run from early morning to midnight, Weike shared the efforts with me and made the long day much easier. My PhD also benefited a lot from the insightful data analysis discussions with Xi. His patient and straightforward explanations made complicated bioinformatic analysis much easier to comprehend.

I'm very grateful for Prof. Dr. Thomas Höfer. The insightful discussions and computational analysis from him undoubtedly brought the *PolyloxExpress* project to a much higher level. His explanations of mathematical and computational methods made seemingly complicated analysis reasonable and understandable. I would also like to show my sincere gratitude to Dr. Guoliang Cui, not only for the great opportunity to collaborate and learn about T cell biology, but also for his kind suggestions and feedback about my PhD and future career. I also want to thank my great TAC members, Prof. Dr. Hedda Wardemann, Prof. Dr. Gergana Dobreva and Dr. Jan-Philipp Mallm for giving me feedback and guidance throughout my PhD. Specifically, I want to thank Hedda and Lindsay Murrells for their tremendous support on my PhD registration in the faculty. Things could have been way more difficult and complicated without their help.

I can't imagine a PhD life without the great colleagues in the lab. The friendly and helpful environment created by everybody in the lab is something that set the tone for the great PhD experience. In particular I want to thank Larissa Johanna Frank. It was a great pleasure for me to do experiments with Larissa together because she's always very responsible and reliable. I also want to thank Dr. Katrin Busch and Dr. Thomas Plum for their discussions and feedback on my PhD, as well as Günter Küblbeck and Sven Schäfer for their support and help to make difficult experiments easier to handle.

Last but not the least, I would like to express my deepest gratitude for my parents and grandparents. I left China on October 2018 to start my PhD in Germany, since then, more than four years have passed and I was not able to go home and visit them. The COVID-19 outbreak changed everything in a way I could have never imagined. It was a hard time for everybody, and perhaps the best out of the worst, I finished my PhD thesis on time and my parents and grandparents were healthy. Words could not convey my gratitude for the love and support I got from them, and I could only hope to see them soon in 2023.

Eidesstattliche Versicherung

1. Bei der eingereichten Dissertation zu dem Thema

Stem cell heterogeneity and clonal dynamics explored by *PolyloxExpress* RNA barcoding

handelt es sich um meine eigenständig erbrachte Leistung.

2. Ich habe nur die angegebenen Quellen und Hilfsmittel benutzt und mich keiner unzulässigen Hilfe Dritter bedient. Insbesondere habe ich wörtlich oder sinngemäß aus anderen Werken übernommene Inhalte als solche kenntlich gemacht.

3. Die Arbeit oder Teile davon habe ich bislang nicht an einer Hochschule des In- oder Auslands als Bestandteil einer Prüfungs- oder Qualifikationsleistung vorgelegt.

4. Die Richtigkeit der vorstehenden Erklärungen bestätige ich.

5. Die Bedeutung der eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unrichtigen oder unvollständigen eidesstattlichen Versicherung sind mir bekannt. Ich versichere an Eides statt, dass ich nach bestem Wissen die reine Wahrheit erklärt und nichts verschwiegen habe.

_____ _____

Ort, Datum Unterschrift