Combining single-cell lineage tracing with transcriptomics

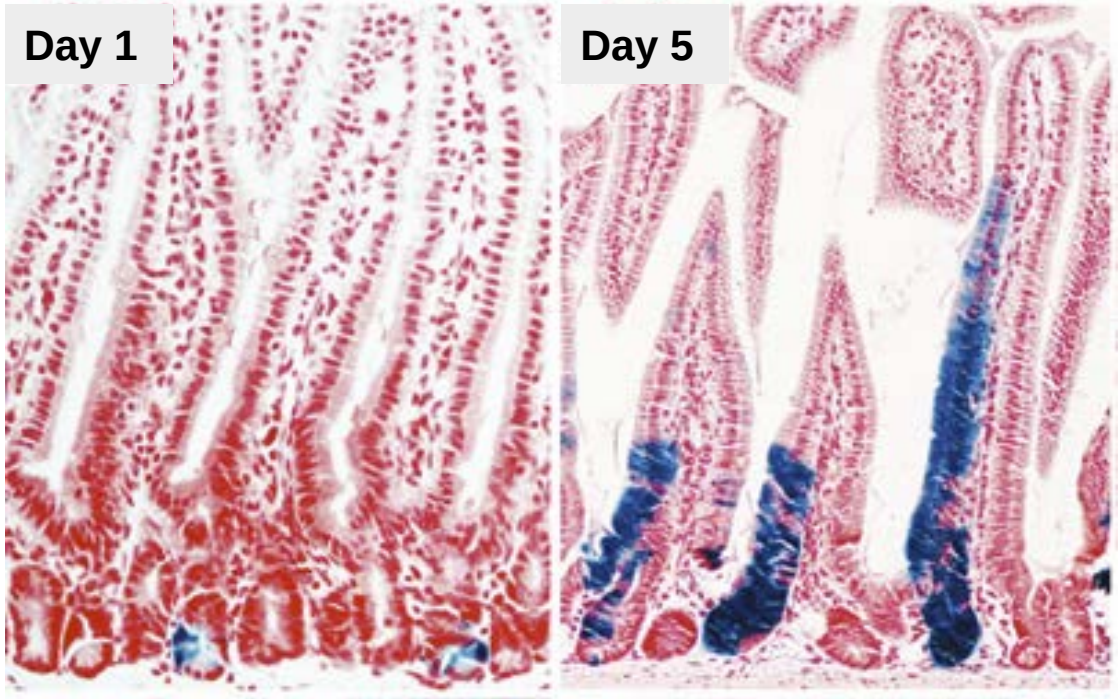
Technical Journal Club

Lukas Frick

15 September 2020

In traditional lineage tracing, the fate of a **whole cohort** of cells is observed over time.

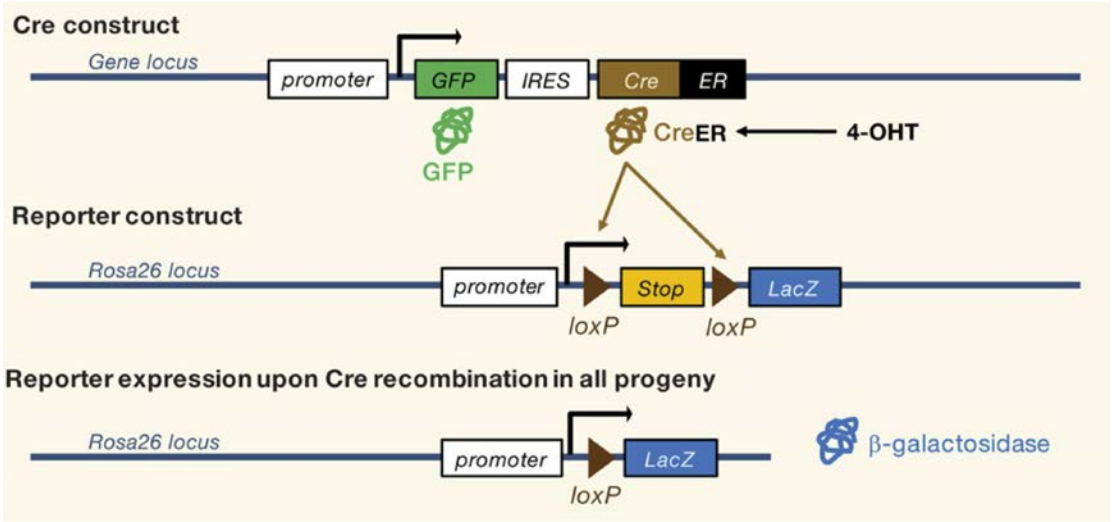
For example, cells expressing a specific progenitor marker are labelled.



Identification of stem cells in small intestine and colon by marker gene Lgr5

Clevers et al, Nature, 2007

Kretzschmar and Watt, 2011



A tamoxifen-inducible Cre recombinase is inserted into the Lgr5 locus (heterozygous knockin mice).

Upon tamoxifen treatment, $CreER^{T2}$ excises a stop cassette enabling *LacZ* expression in all progeny cells.

Lineage tracing can address some of the most basic questions in tissue development and regeneration

- **What are the progenitor cells**, and which cell types do they give rise to?
- **How often does a cell divide** during tissue homeostasis and regeneration?
- What influences the **cell fate decisions** of a cell to divide or differentiate?
  - Environmental niche, growth factors
  - Cell-autonomous, transcription factors

However, traditional lineage tracing has its limitations…

1. Constructing reporter animals is **laborious**
2. There may be **no truly specific markers** for some progenitor-cell types
3. Only a small number of clones can be analysed at once
4. Individual clones cannot be tracked reliably
   → difficult to study stem cell heterogeneity

Genetic barcodes can be used for lineage tracing of **single cells**

- Thousands of individual cells are marked using unique barcodes.
- After cell division, all daughter cells will carry the barcode.

Dissecting T cell lineage relationships by cellular barcoding

JEM, 2008

- Retroviral transduction of a library of barcodes
- Requires *ex vivo* manipulation of haematopoietic progenitor cells and subsequent transplantation

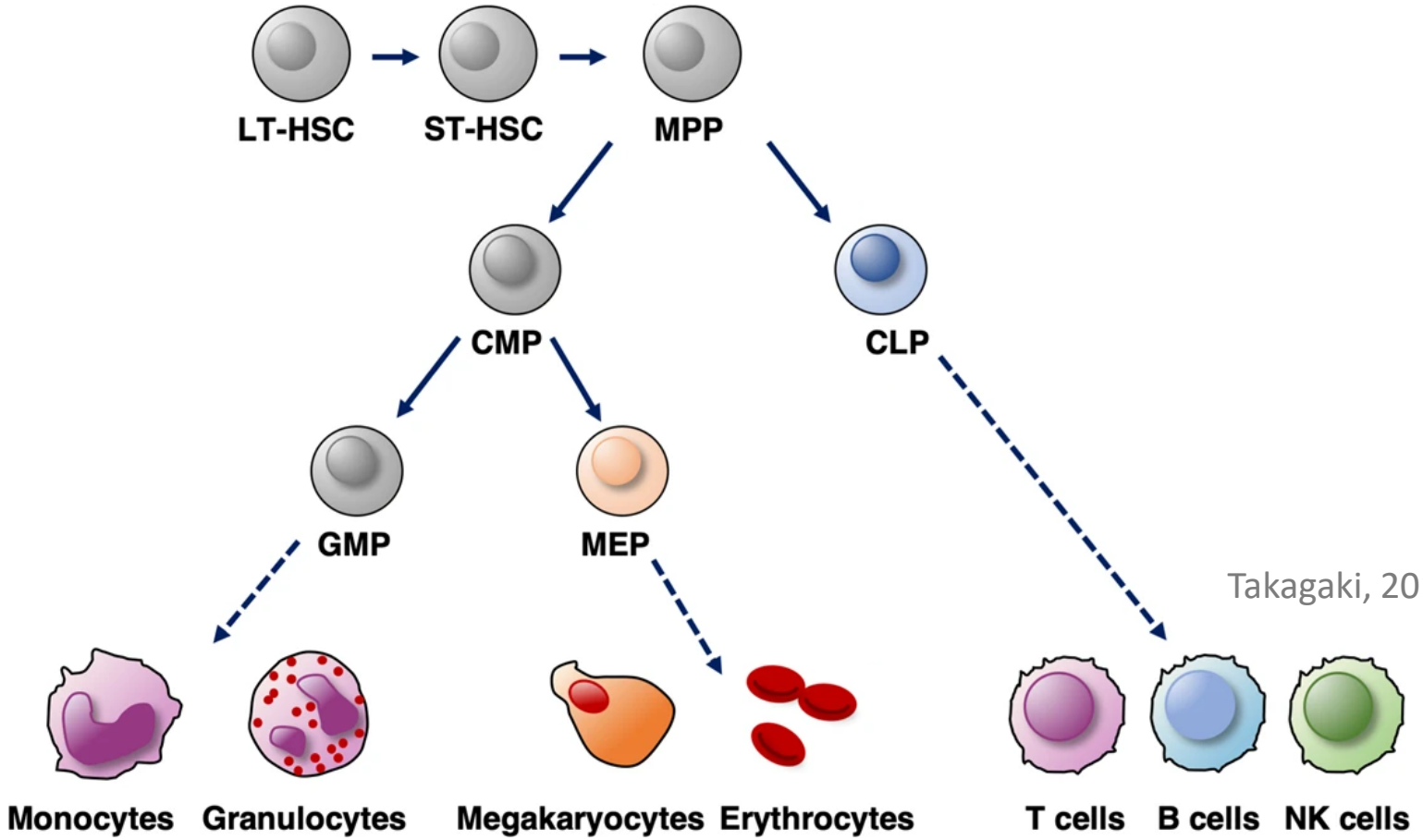
Clonal dynamics of native haematopoiesis

Nature, 2014

- A hyperactive Sleeping Beauty transposase leads to random insertions
- Sequencing DNA surrounding transposon insertion site → **stable genetic tag**
- Doxycycline-inducible *in vivo*

The hematopoietic system can be considered a well-accessible model of stem cell renewal and differentiation

- The bone marrow can be ablated and replaced with transplanted cells
- Serial sampling is easy (just draw blood)
- Easy to culture *ex vivo*

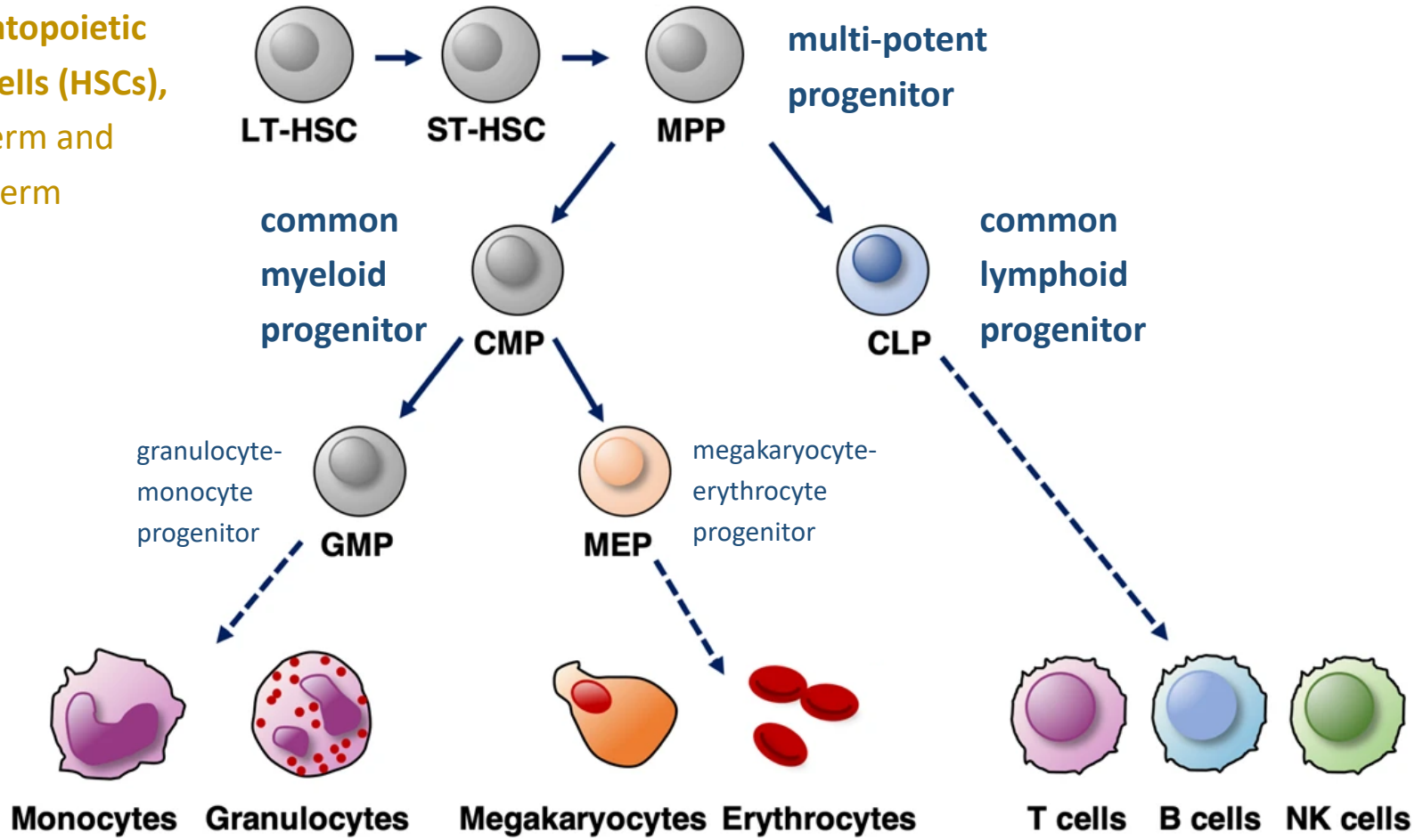


Takagaki, 2019

In the classical hierarchical view of haematopoiesis, stem and progenitor cells differentiate into mature cells in **discrete** steps

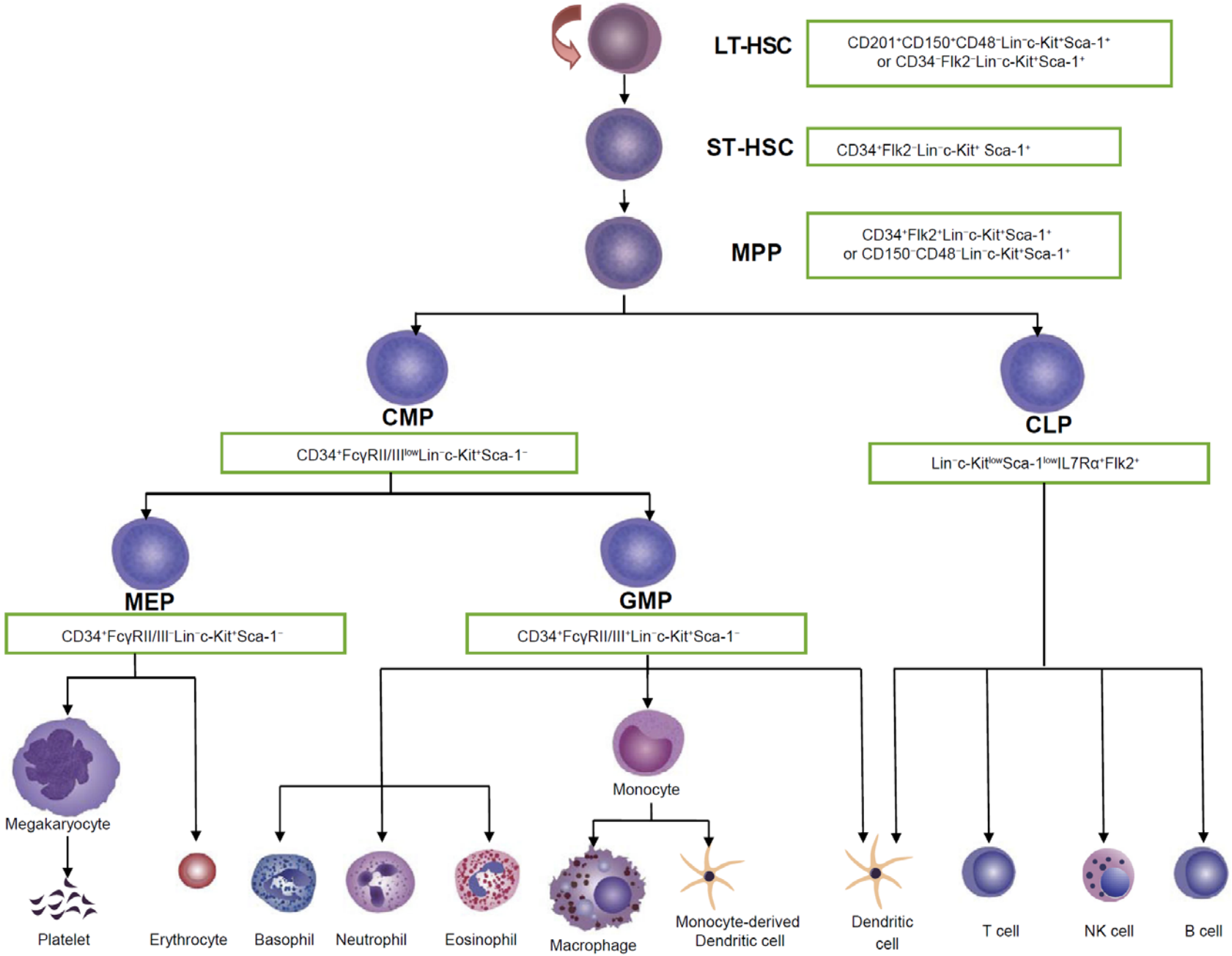
Haematopoietic stem cells (HSCs), long-term and short-term

- **multipotent**
- **indefinite self-renewal**



These cell types have been **defined using** combinations of cell surface markers with **flow cytometry**

Lin⁻ = lacking lineage-specific markers of mature cells, e.g. CD3e⁻, CD19⁻, Gr1⁻, Mac1⁻ and Ter119⁻ in mice



Cheng, 2020

- Previous approaches to genetic barcoding required analysis of genomic DNA …
  - … either sequencing of
    - bulk DNA (mix of clones)
    - or single colonies grown in culture
    - or DNA from sorted single cells amplified by single-cell PCR

- **If the barcodes were expressed** (transcribed as mRNA), they could be analysed by standard RNA-seq!

To study the mouse hematopoietic system, three papers published in 2020 used **expressed barcodes** to combine single-cell lineage tracing with scRNA-seq

RESEARCH ARTICLE

Lineage tracing on transcriptional landscapes links state to fate during differentiation

Science

Article

Single-cell lineage tracing unveils a role for TCF15 in haematopoiesis

Nature

Cell

Resource

An Engineered CRISPR-Cas9 Mouse Line for Simultaneous Readout of Lineage Histories and Gene Expression Profiles in Single Cells

LARRY – lentivirus-based

→ *ex vivo* manipulation + transplantation

CARLIN – CRISPR-based

→ doxycycline-inducible *in situ*

RESEARCH ARTICLE

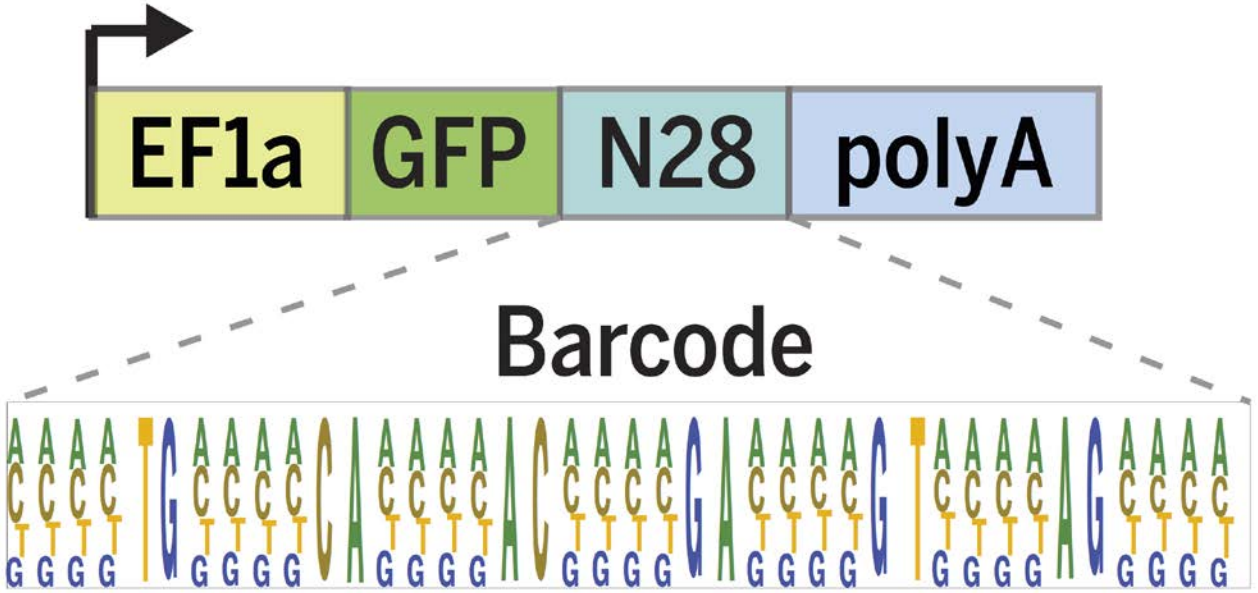
DEVELOPMENTAL BIOLOGY

Lineage tracing on transcriptional landscapes links state to fate during differentiation

Caleb Weinreb[1]*, Alejo Rodriguez-Fraticelli[2,3]*, Fernando D. Camargo[2,3]†, Allon M. Klein[1]†‡

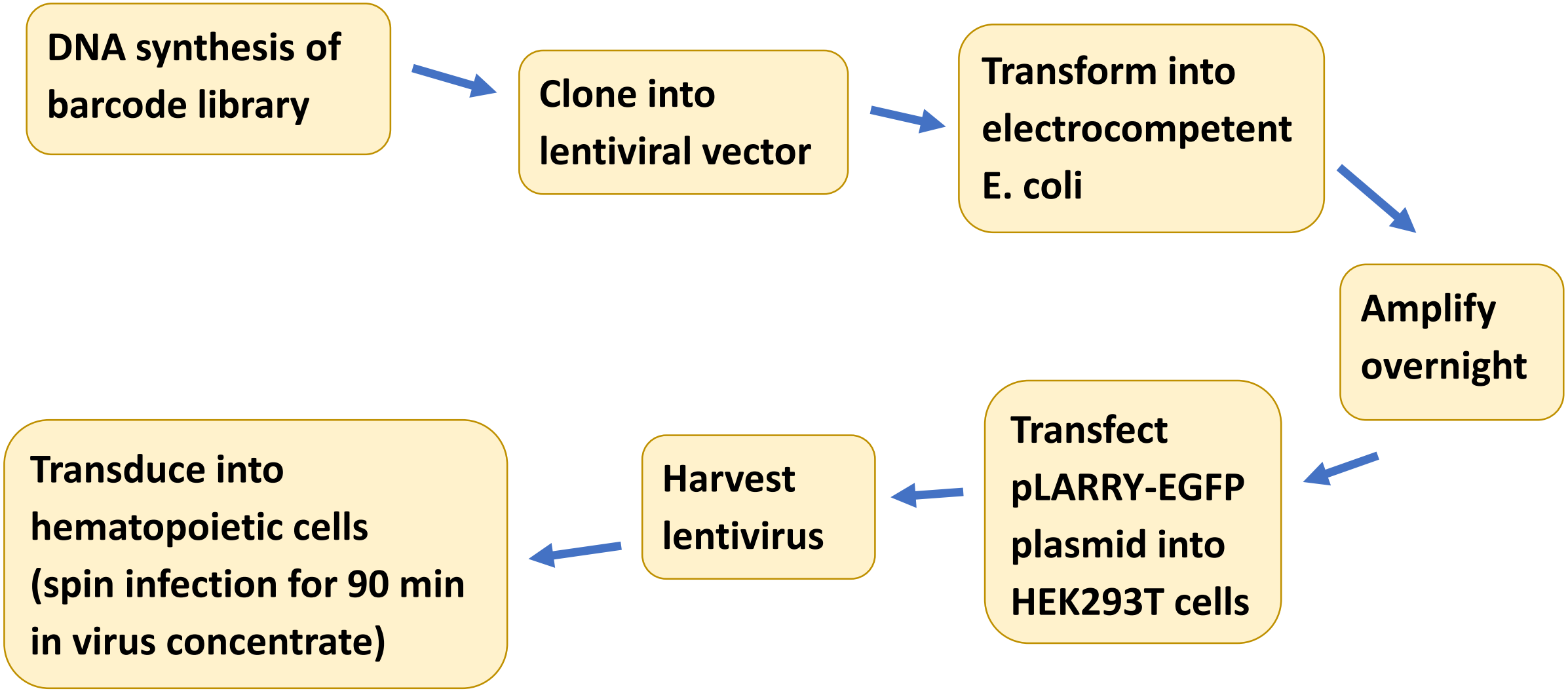
Science, February 2020

They authors introduce LARRY (Lineage and RNA recovery) – a tool involving lentiviral delivery of inherited DNA barcodes



A random 28-mer is placed in the 3’ UTR of an eGFP transgene under an ubiquitous EF1a promoter

LARRY library synthesis, amplification, packaging and delivery

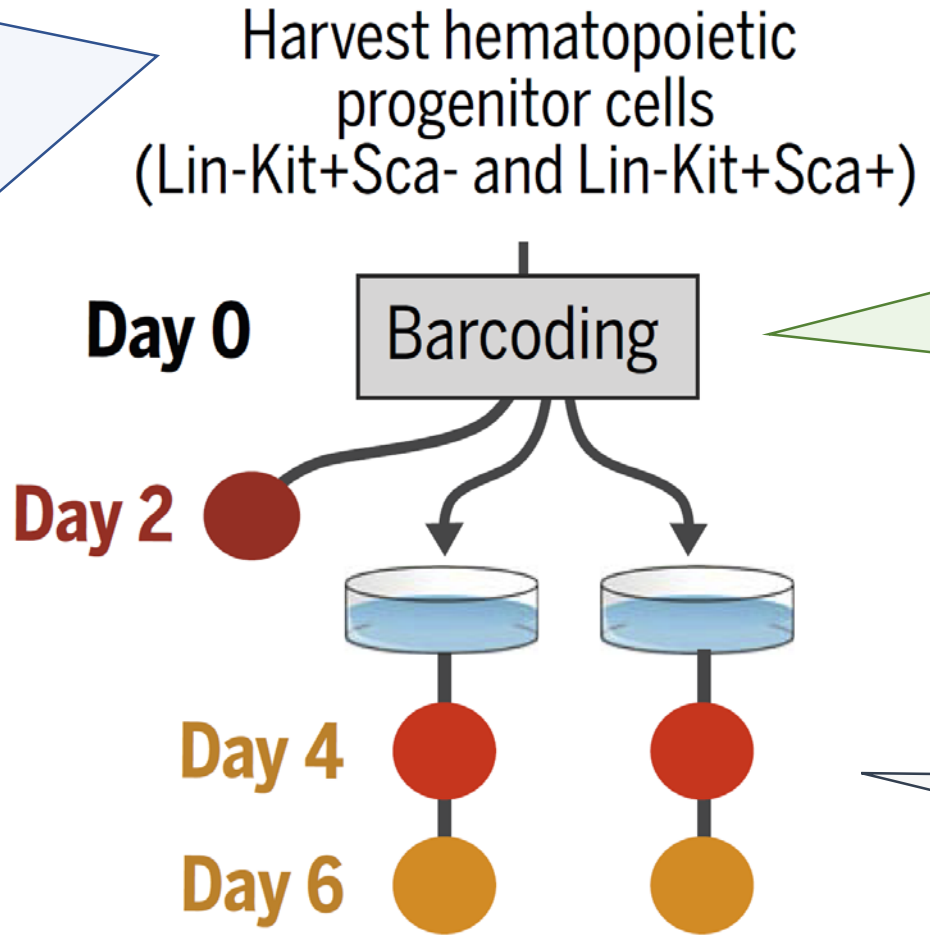


How many unique clones can we differentiate using LARRY?

- In theory, a 28-bp random sequence can produce 4^{28} different barcodes.
- The actual diversity is much lower, because each barcode must be transformed, amplified and harvested from a bacterial colony.
- By looking for barcodes shared across experiments (that cannot come from the same clone), the barcode diversity was estimated at ~500’000.

  → **<1% barcode overlap** between clones

1. Donor mice are euthanized
2. Bones crushed, marrow filtered through a strainer
3. Mature Lin^+ cells depleted with antibody-coupled magnetic beads
4. FACS-sorting to isolate stem and progenitor cells



Culture for 2 days to allow viral integration (cells divide ~3×)

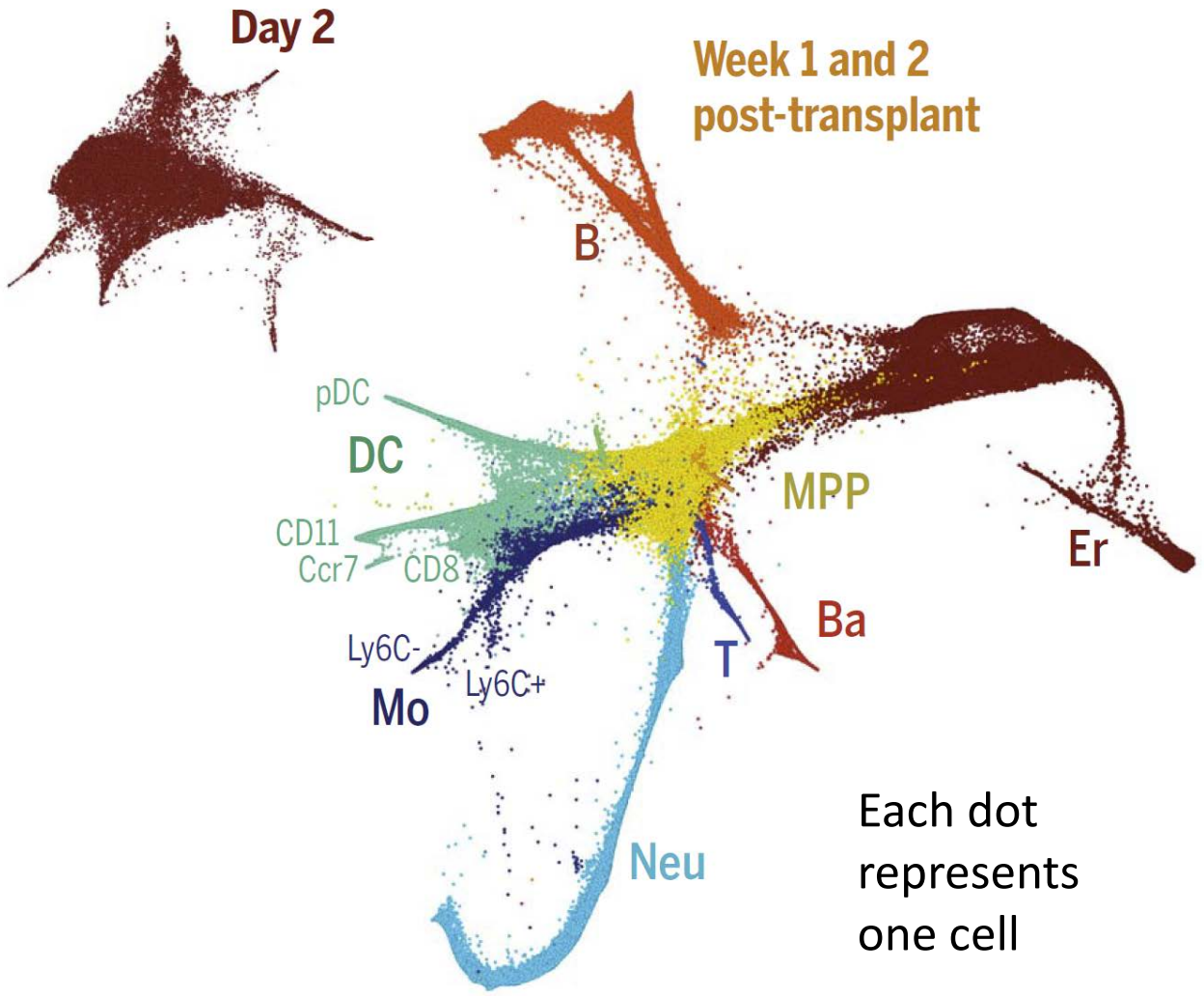
Analyze with **scRNA-seq**

Standard droplet-based single-cell RNA-sequencing was done using the inDrop platform

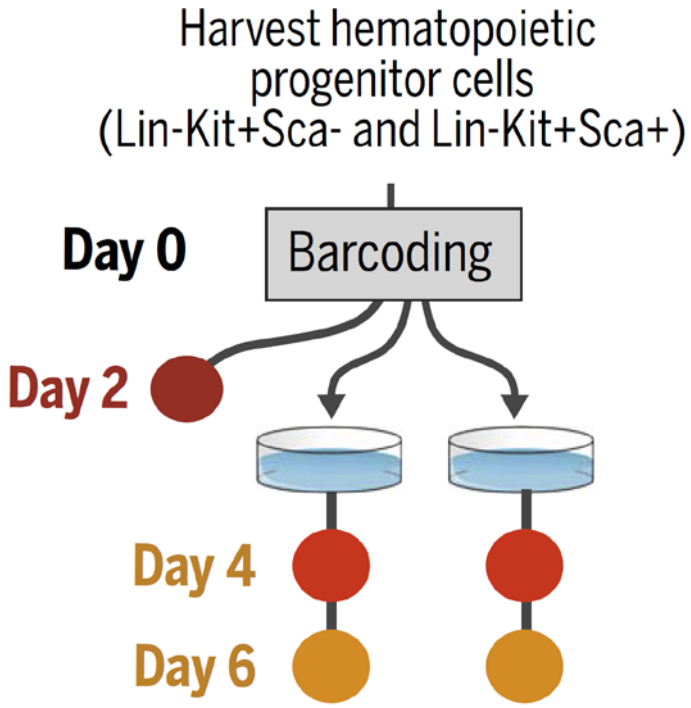
- There was only one modification:
  - After reverse transcription, half of the sample was used for a targeted PCR to selectively amplify LARRY barcode transcripts.
  - The product was then pooled with the remaining library for sequencing.

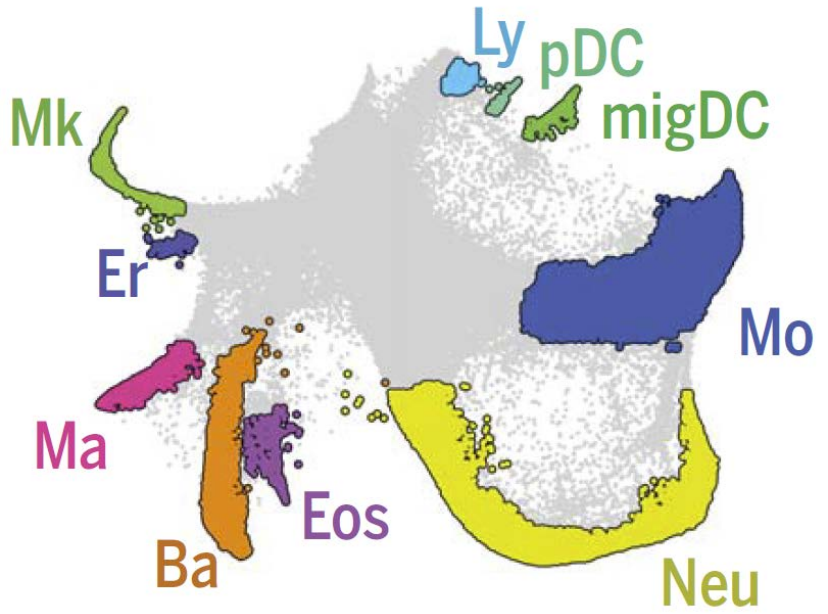
SPRING plots were used to visualize single-cell RNA-seq data

- SPRING plots are force-directed layouts of k-nearest neighbour graphs (Weinberg et al, 2018).
- Based on a distance matrix of the top 50 principal components
  - Using only highly variable genes (cell cycle genes are excluded)
- Compared to t-SNE: Less likely to form disconnected clusters (better for continuous expression trajectories)



In a first experiment, hematopoietic stem and progenitor cells remained in culture, in medium that supported pan-myeloid differentiation

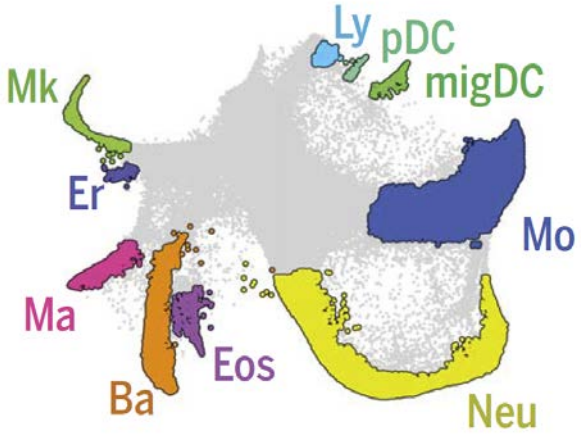




Regions were annotated manually based on marker genes

scRNA-seq data of cultured cells confirms that in this setting, we see differentiation into all major lineages

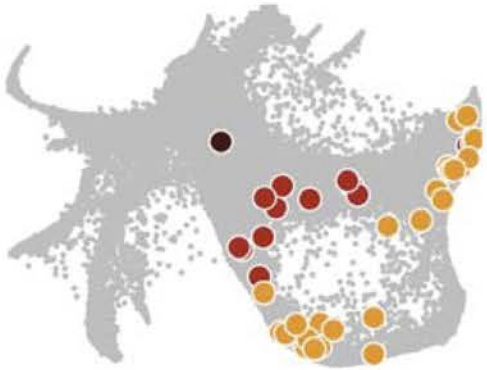
Differentiation of the progeny of individual progenitor cells could be tracked over time



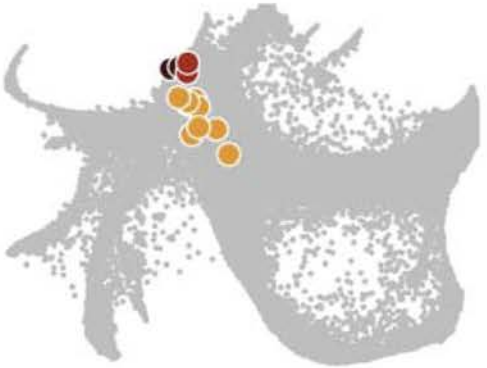
All cells

Clone A

Clone B

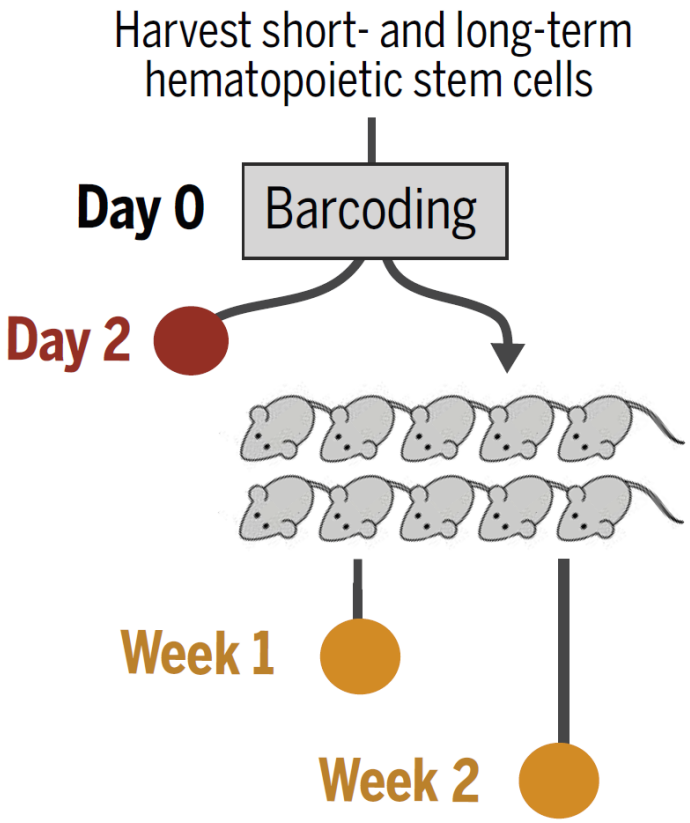


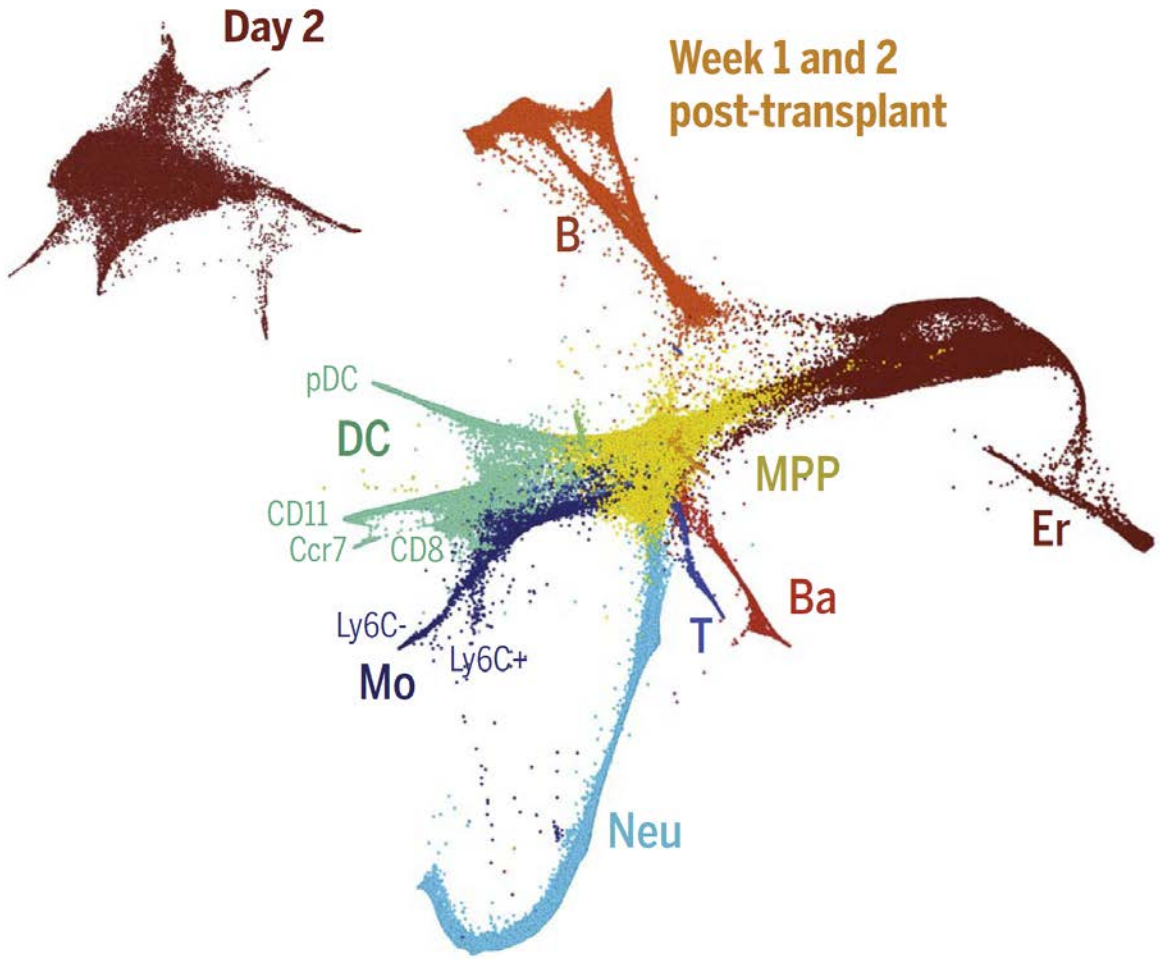
Clone C



Day 2 ● Day 4 ● Day 6 ●

scRNA-seq of hematopoietic stem cells transplanted into mice also showed differentiation into various mature cell types within two weeks





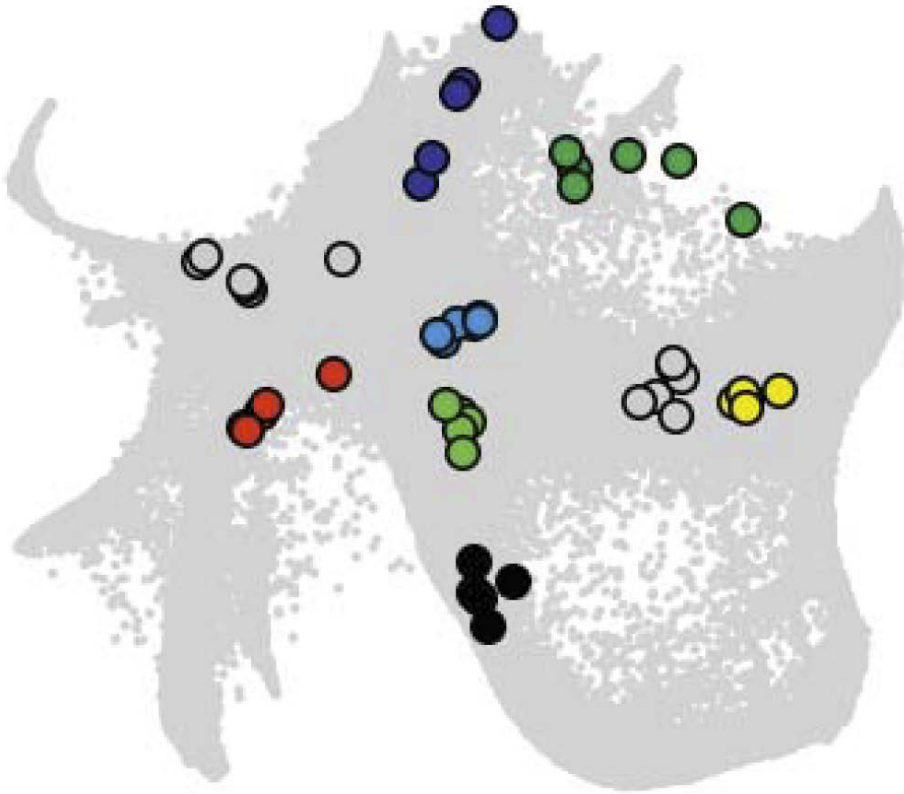
In vivo data

Total cells	182,173
Cells in multi-cell clone	115,570
Total multi-cell clones	7,751
State-fate clones	817

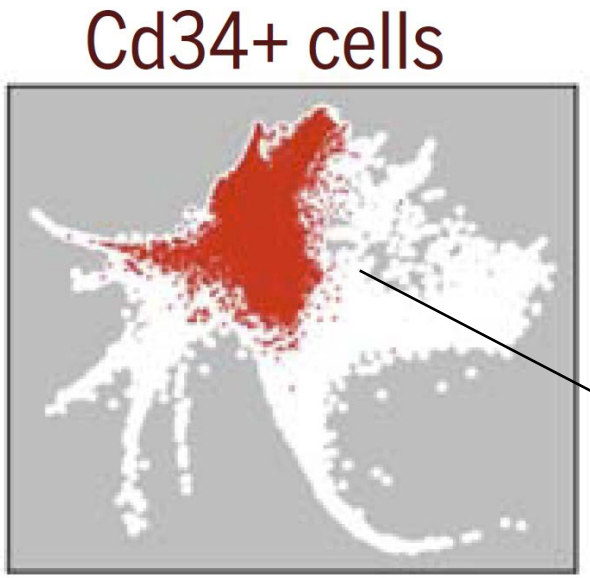
817 clones were sampled at both early and late timepoints!

With LARRY, we can estimate how a single cell changes over time by sampling a single clone over multiple timepoints

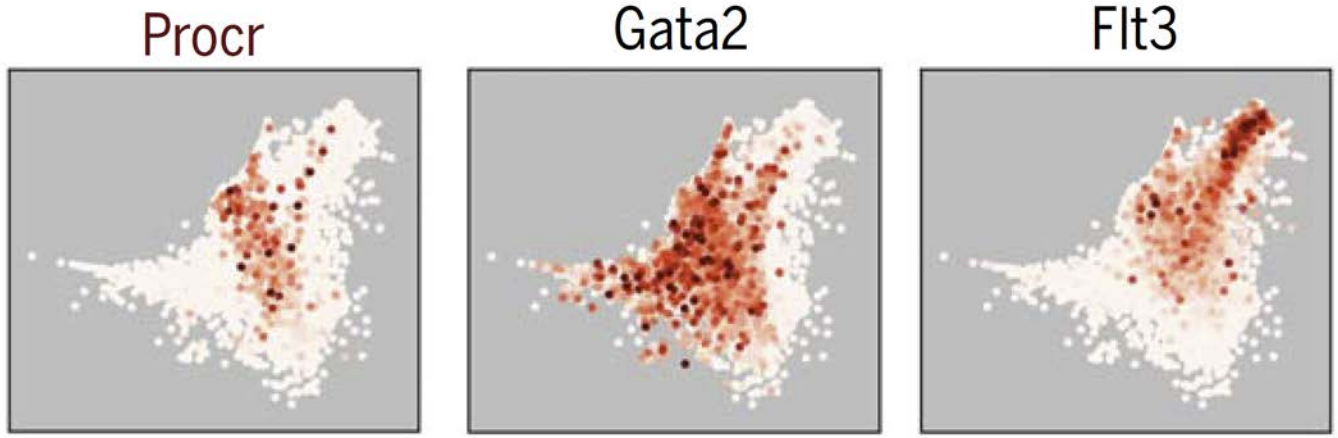
- For example, we can examine a cell's sister cells at day 2 to approximate its baseline state.
- However, this assumes that sister cells at early timepoints are very similar.
- Luckily, at day 2, sister cells had correlated gene expression (R ≈ 0.85) and clustered close together.



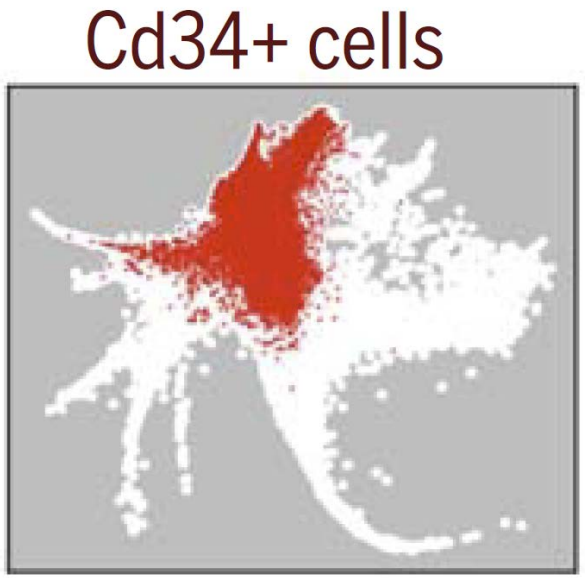
We can superimpose gene expression data with clonal outcomes – “link state to fate”



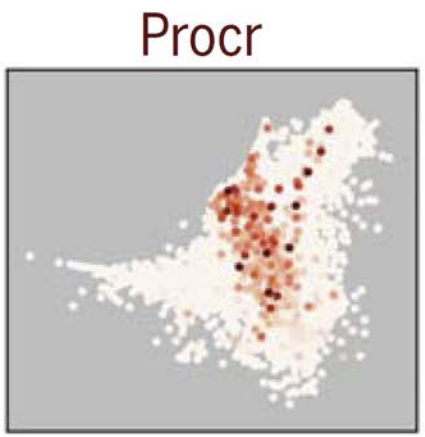
As an example, we will examine Day 2 cells within the **CD34^{+} multipotent progenitor region** in the *in vitro* dataset.



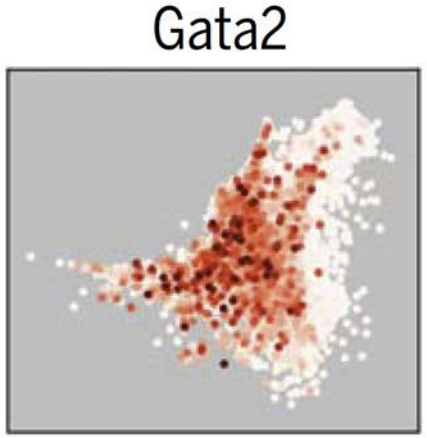
We can superimpose gene expression data with clonal outcomes – “link state to fate”



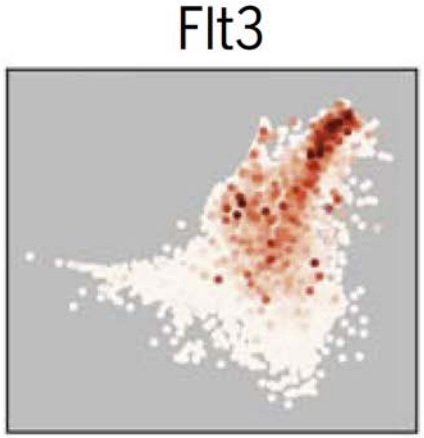
The region can be broadly subdivided using marker genes.



stem cell marker

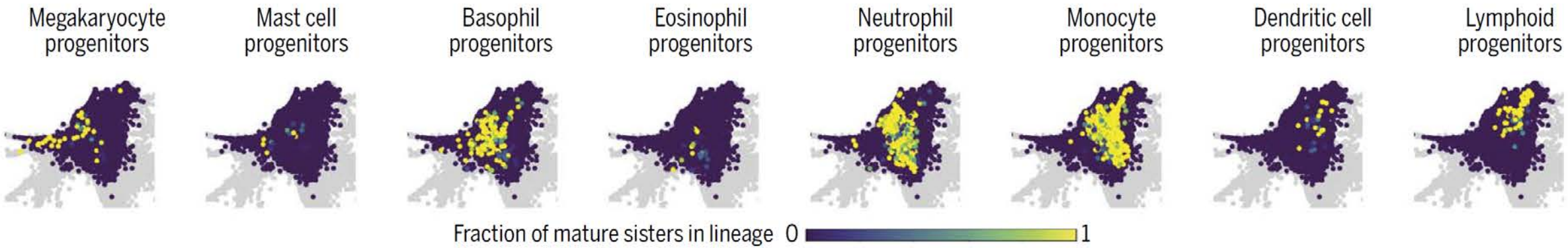


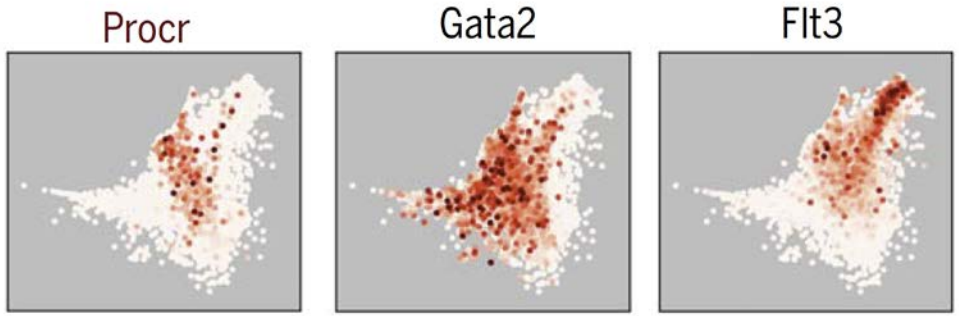
erythroid and stem cell marker



lymphoid priming

We can appreciate finer subdivisions by superimposing clonal fates



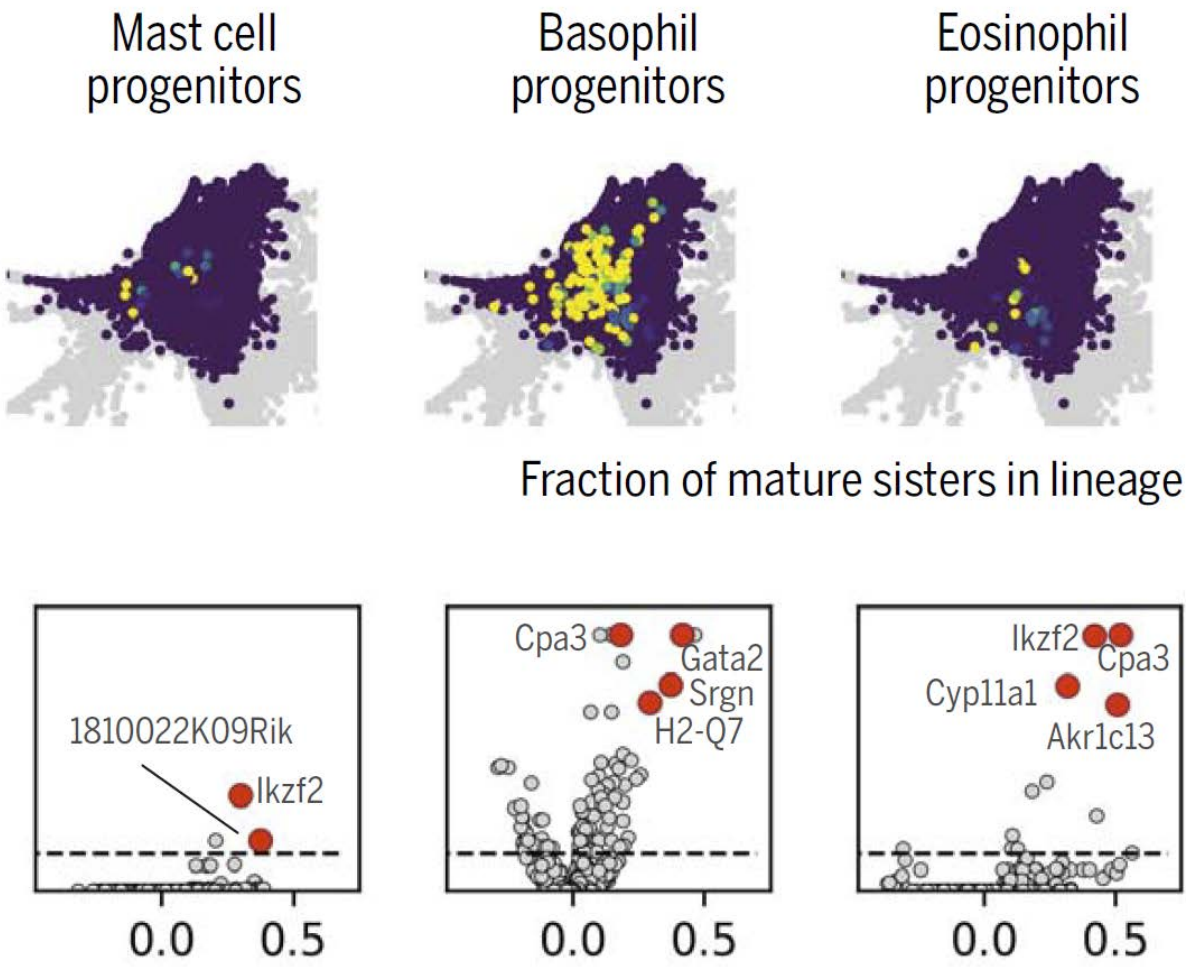


stem cell marker

erythroid and stem cell marker

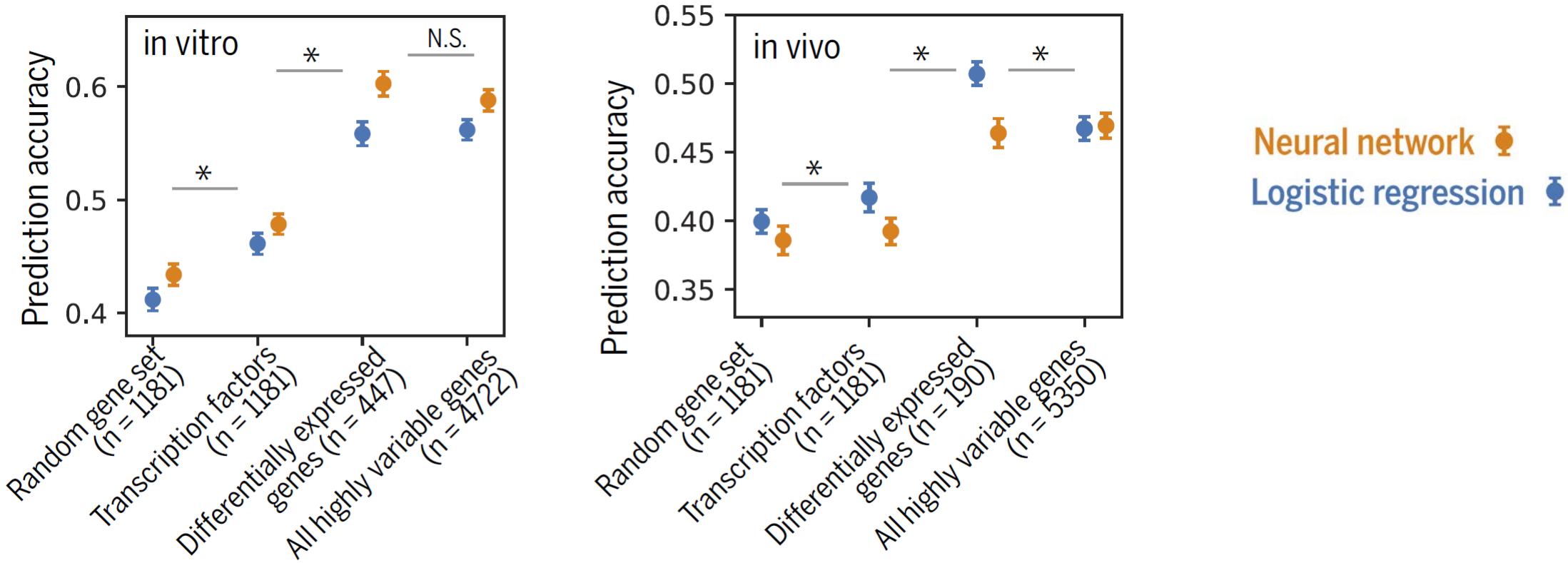
lymphoid priming

The dataset can be used to find new markers for progenitor cells



For example, *Ikzf2* is overexpressed in progenitors of mast cells and eosinophils.

Can we predict cell fate from scRNA-seq data, using machine learning algorithms?



⟶ Yes, but not very well…

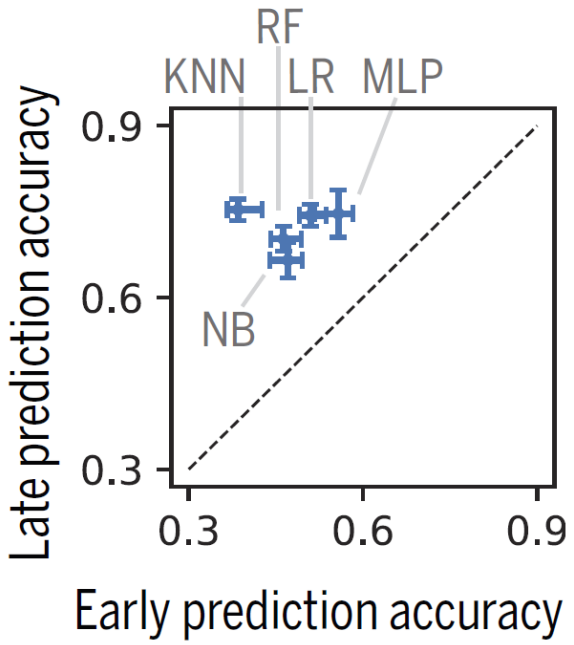
Are there “hidden variables” influencing cell fate that we cannot observe in RNA-seq data at early timepoints?

- If no “hidden variables” exist, scRNA-seq profiles of sister cells will only become more dissimilar as time passes.
- However, if stable hidden properties exist, then sister cells will become more alike over time, as these properties manifest in cell fate.

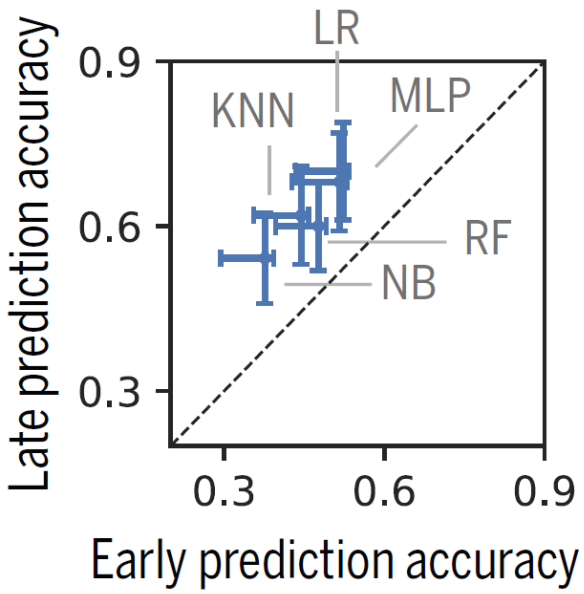
The authors examined **separated sister cells**

(plated into separate wells / transplanted into different mice)

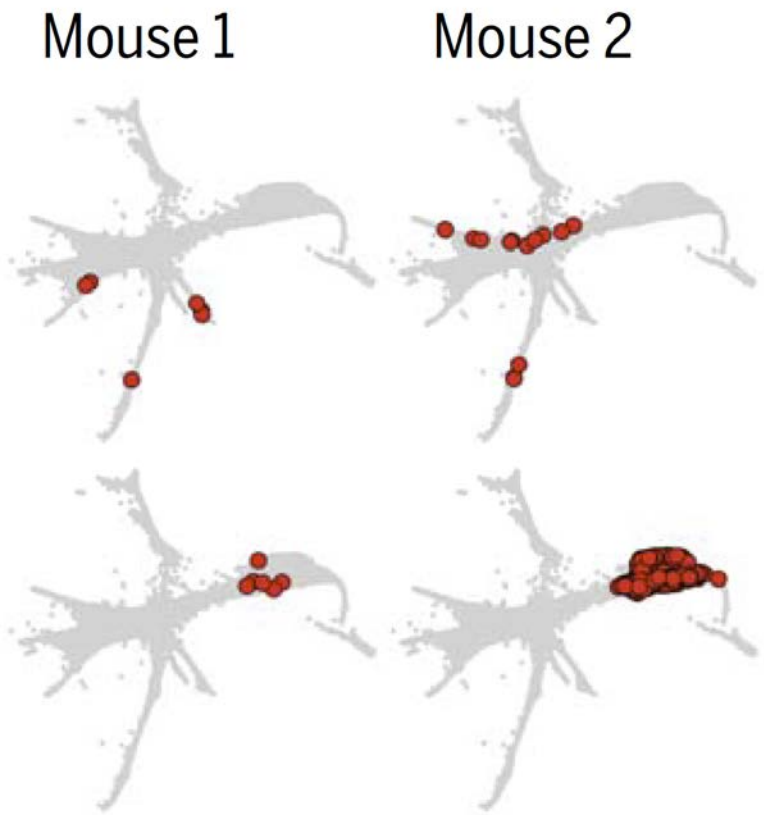
In vitro



In vivo



Indeed, RNA-seq data from **separated sister cells at late timepoints were better** at predicting fate than data from day 2 sisters!



Sister cells transplanted into separate mice shared the same dominant fate 71% of the time (23% expected by chance)

Thus, cells commit to a cell fate earlier than we might expect, and we may not capture this using scRNA-seq!

- Explanations?
  - scRNA-seq data is noisy or incomplete
  - The information on cell fate decisions cannot be found in mRNA
    - Chromatin state, protein levels, cell organization…?

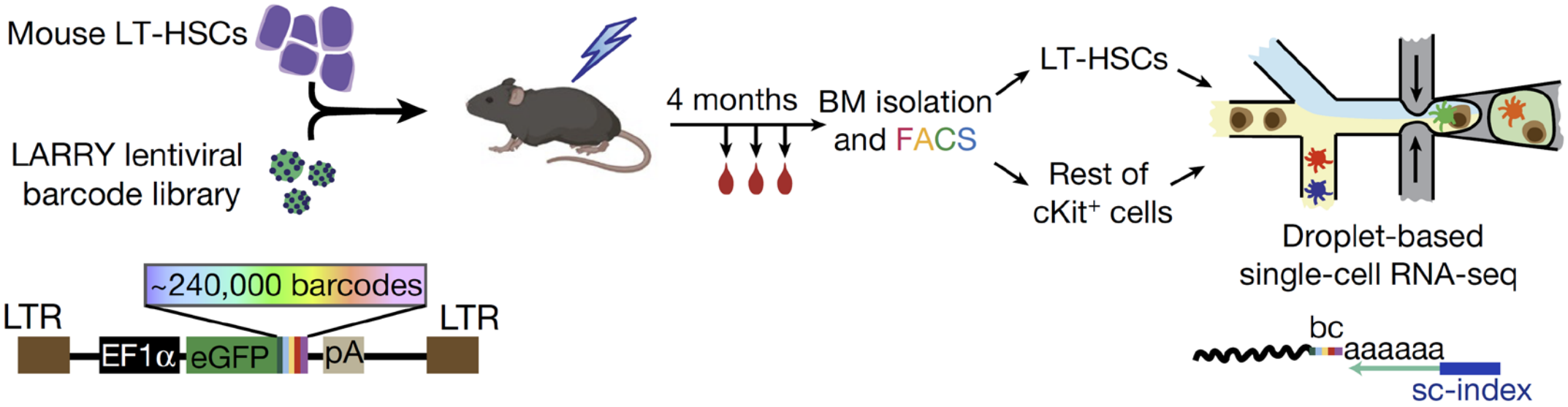
Article

Single-cell lineage tracing unveils a role for TCF15 in haematopoiesis

Alejo E. Rodriguez-Fraticelli[1,2,3,4], Caleb Weinreb[5], Shou-Wen Wang[5], Rosa P. Migueles[6], Maja Jankovic[1,2], Marc Usart[1,2], Allon M. Klein[5], Sally Lowell[6] & Fernando D. Camargo[1,2,3,4]✉

Nature, July 2020

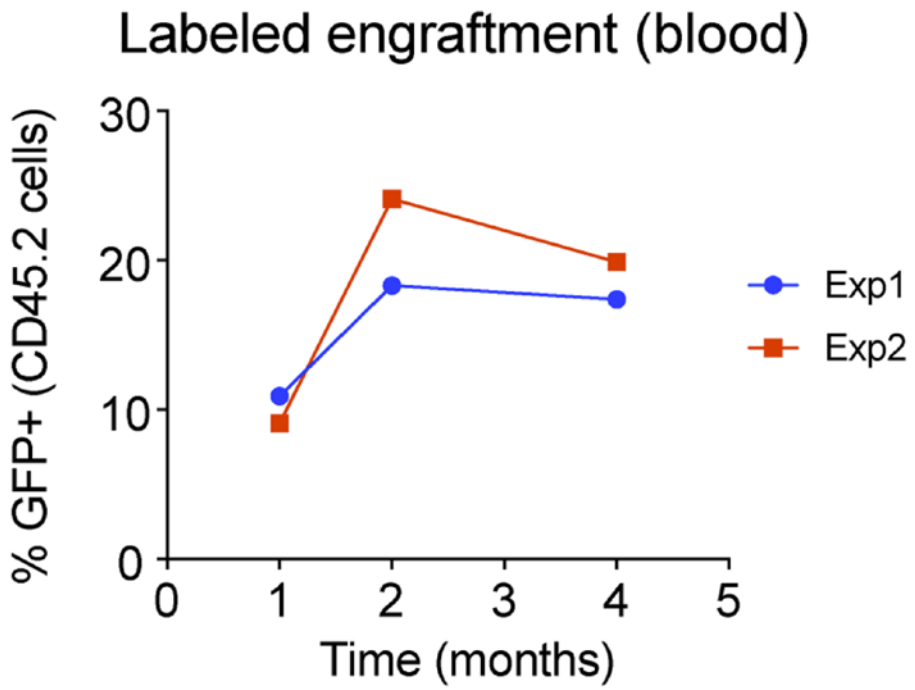
The authors used LARRY to track the fate of hematopoietic stem cells (HSCs) over the **long term**



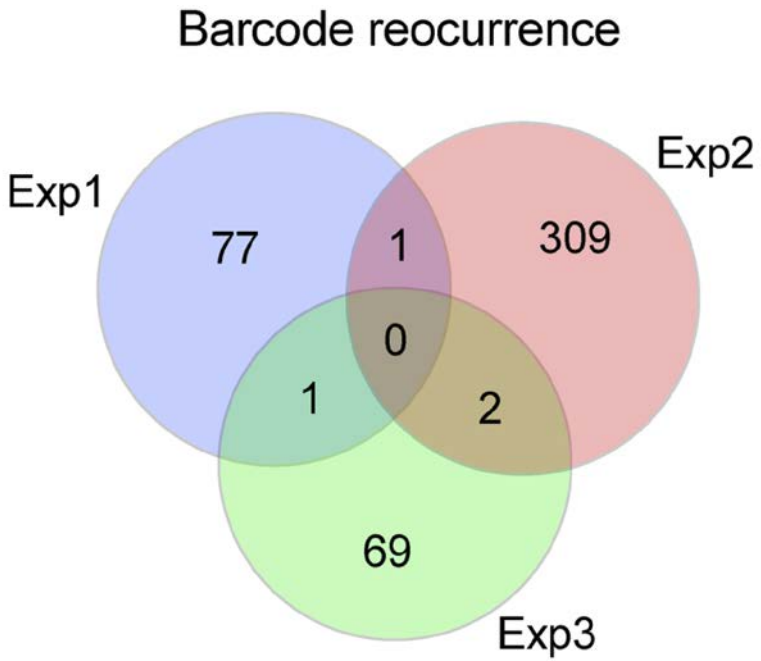
LARRY-barcoded LT-HSCs were transplanted into lethally irradiated 8-week old mice.

FACS: $Lin^{-}cKit^{+}Sca1^{+}CD150^{+}CD48^{-}$

Technical checks



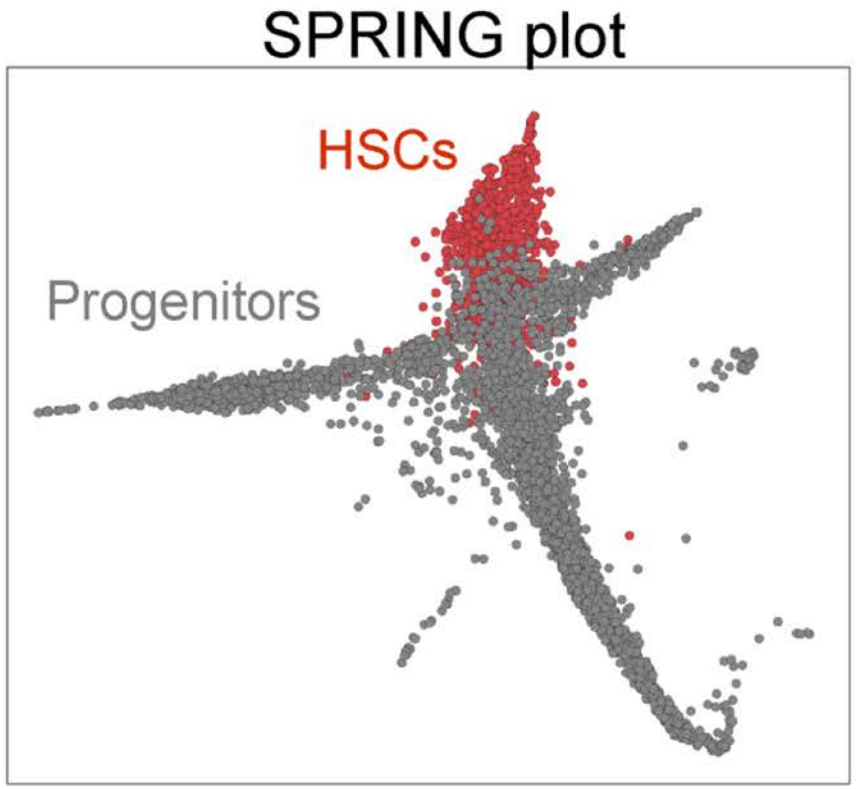
About 20% of cells had integrated a LARRY construct.

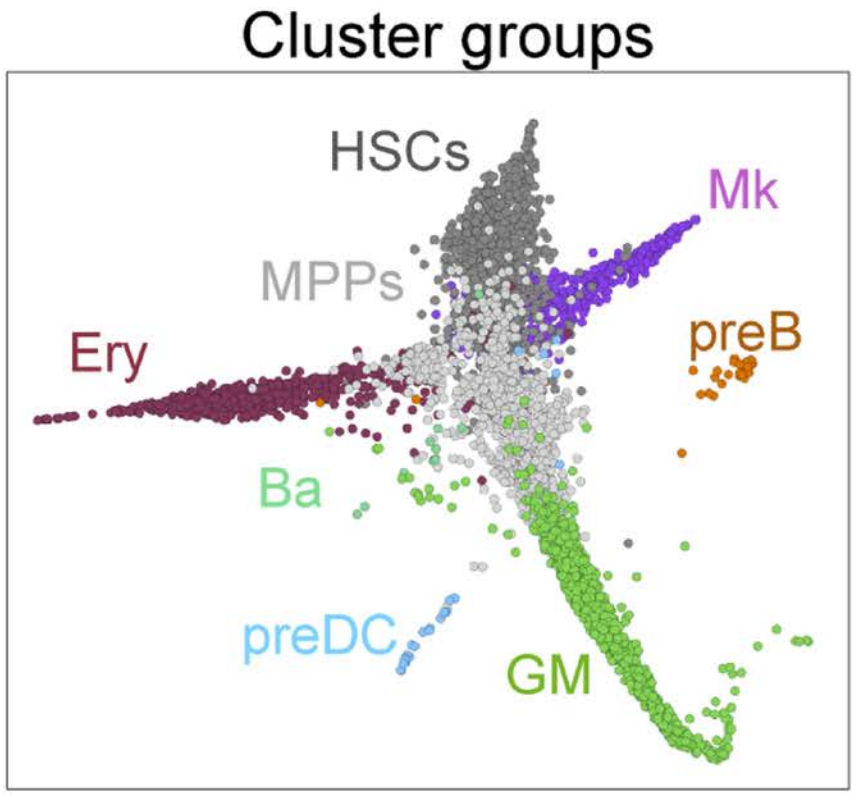


Barcode diversity was adequate

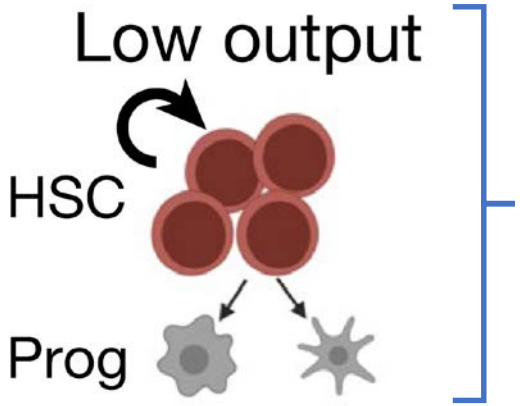
For most (~80%) clones, at least one barcoded HSC was detected.

The functional behaviour of 227 HSCs and their associated gene expression programmes could be analyzed.

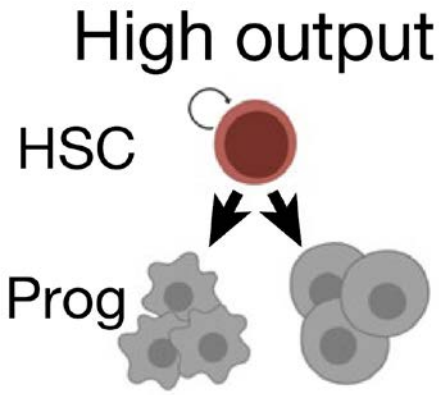


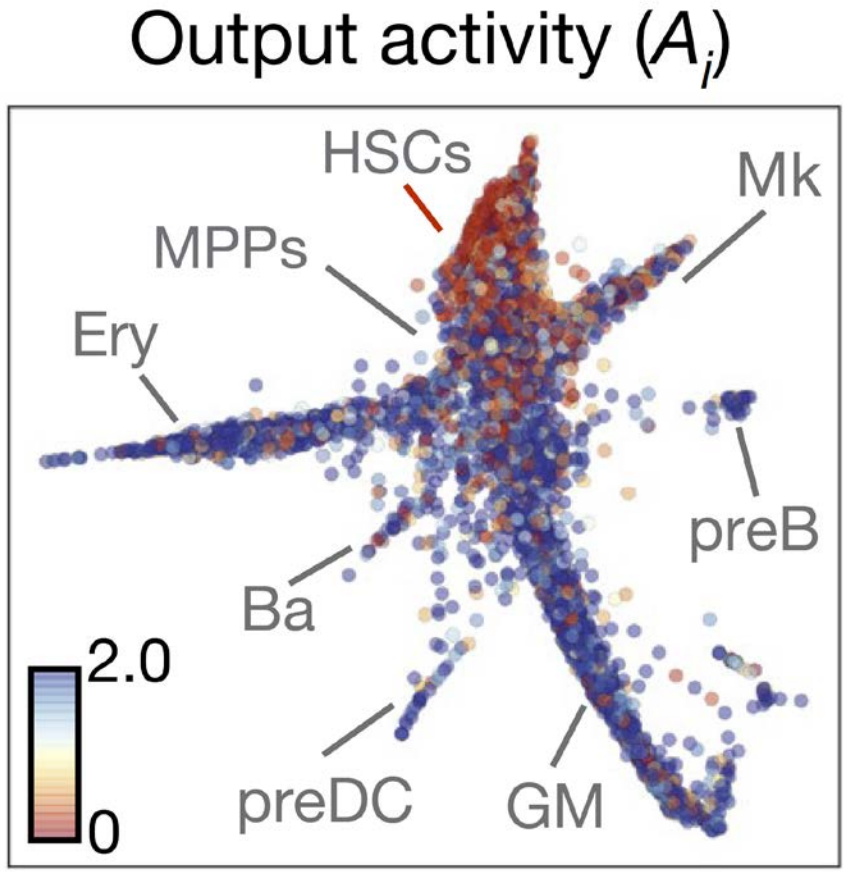


There was a large degree of clonal heterogeneity – ~60% of HSCs were “low-output”

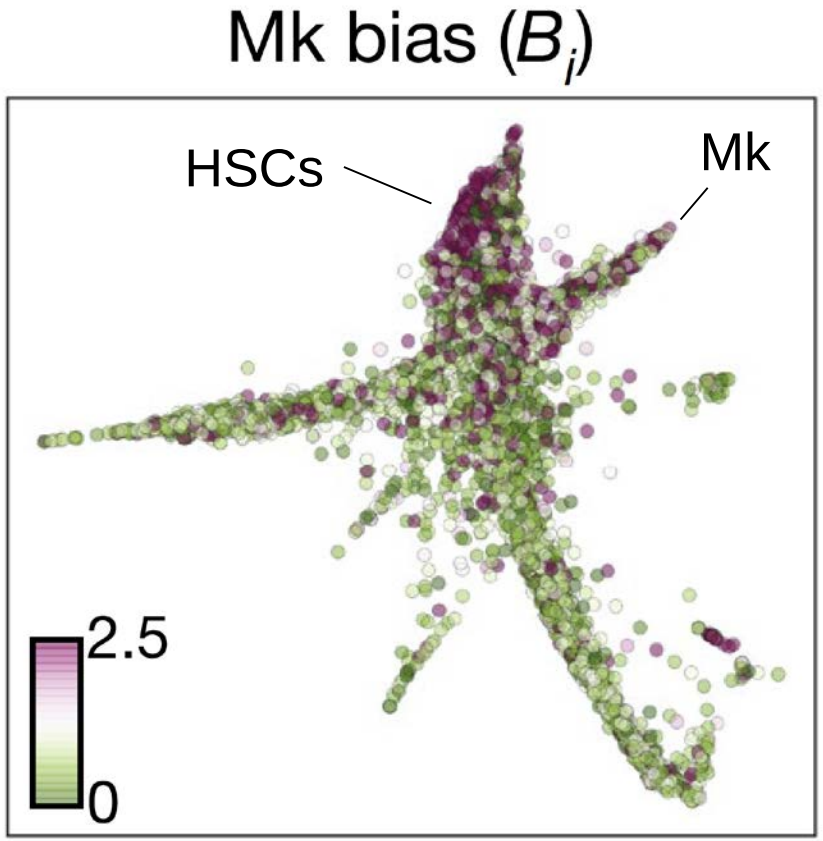


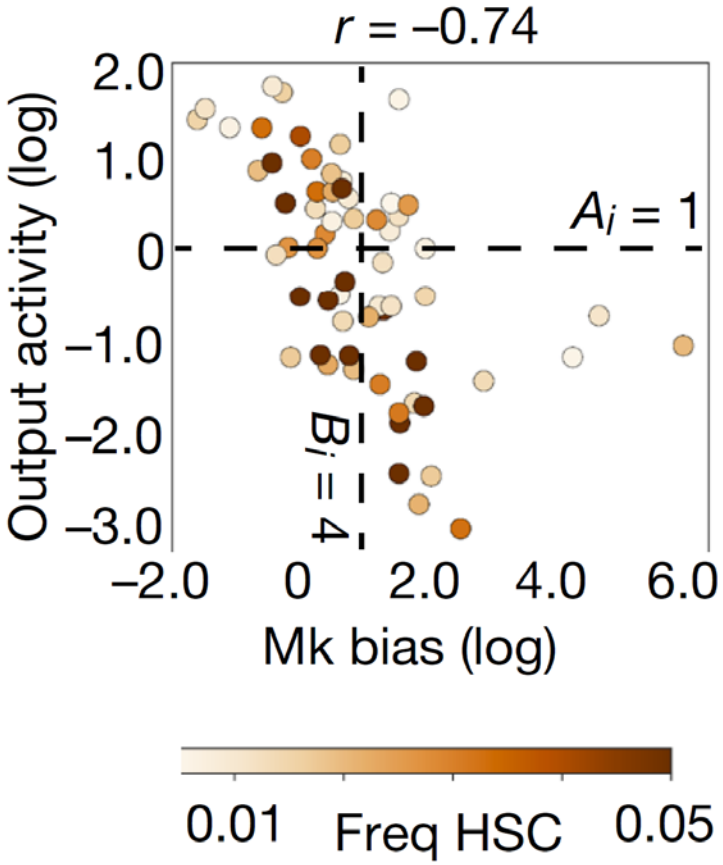
Self-renewal > contribution to differentiated progeny



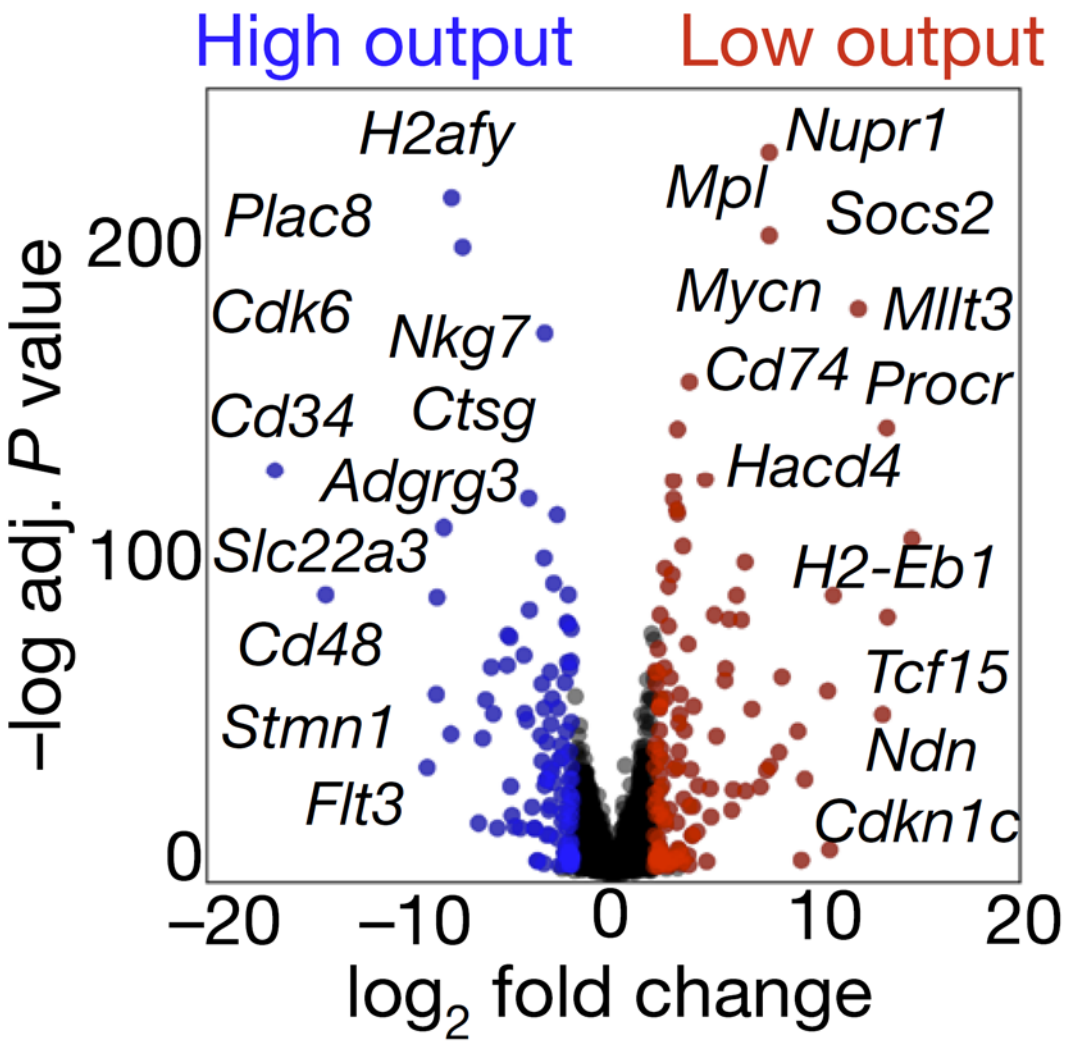


~30% of HSC clones were megakaryocyte-biased, and produced 50-60% of megakaryocyte progeny



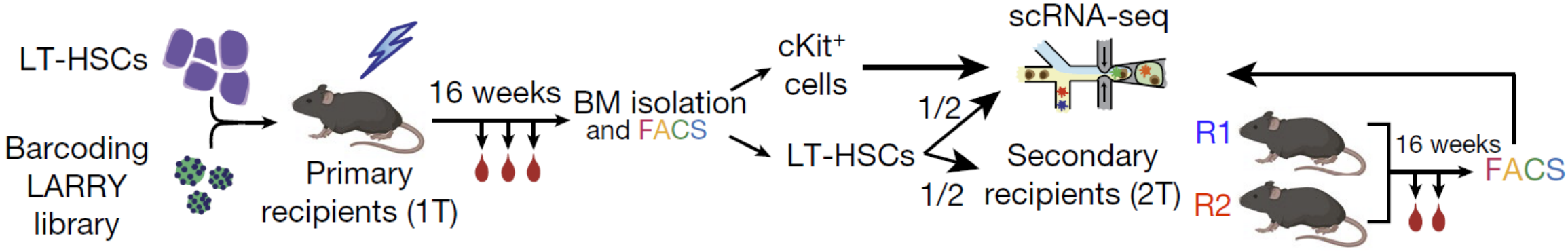


Mk bias correlated with low output.

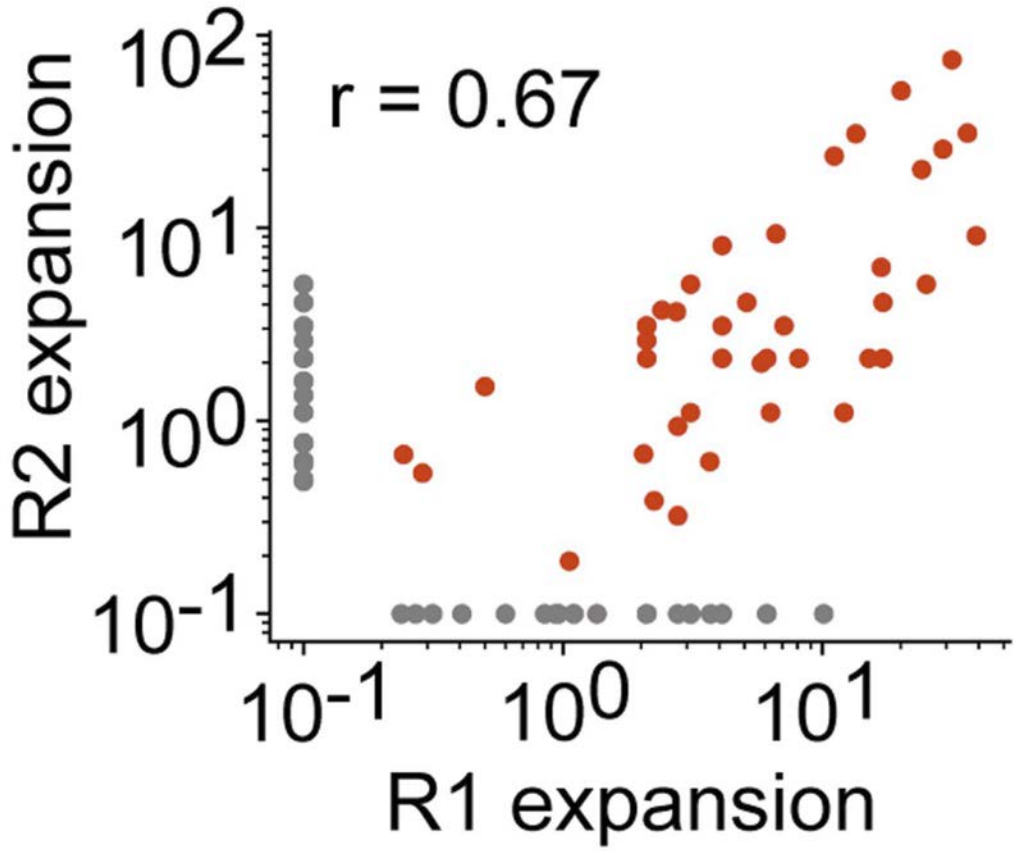


Genes differentiating low-output from high-output HSCs could be identified.

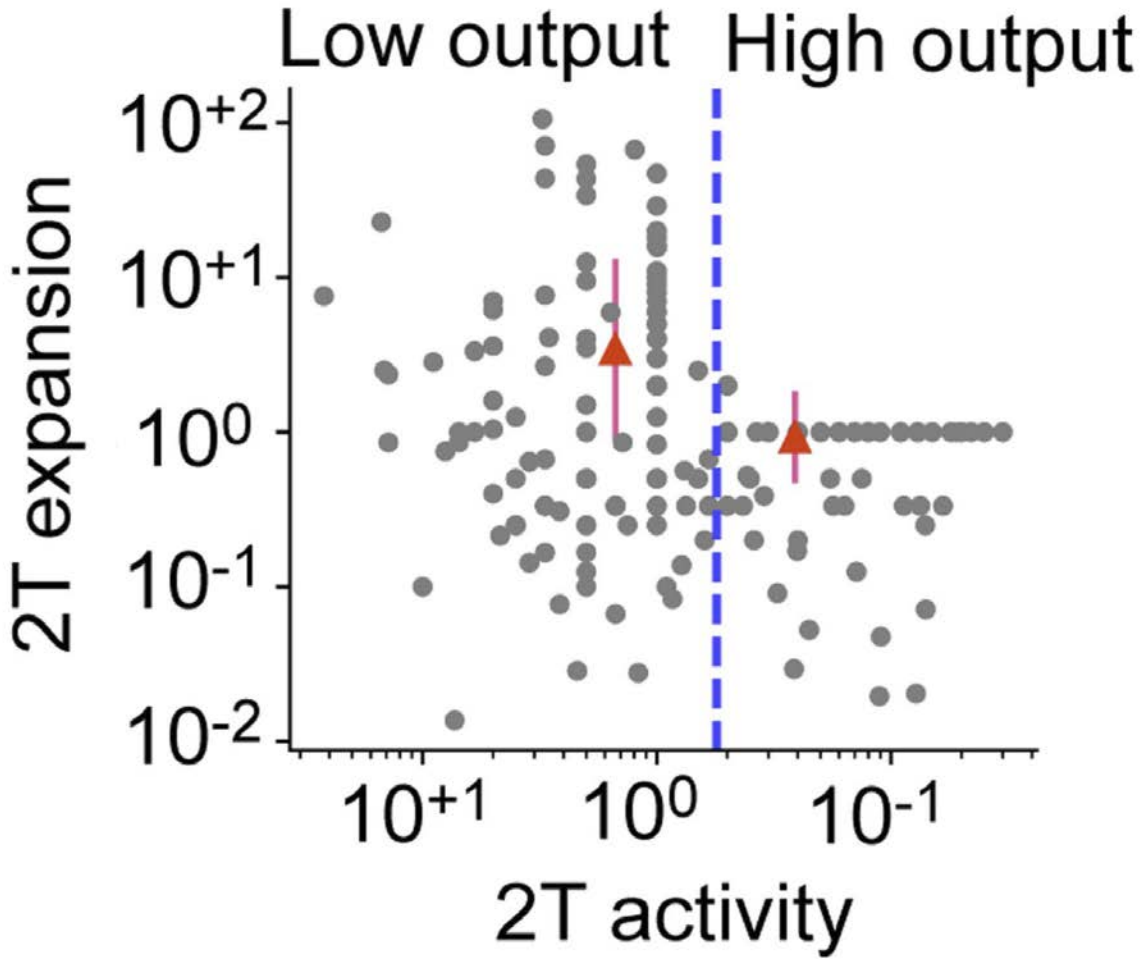
To assess the long-term repopulation capacity of HSCs, secondary transplantation experiments were done



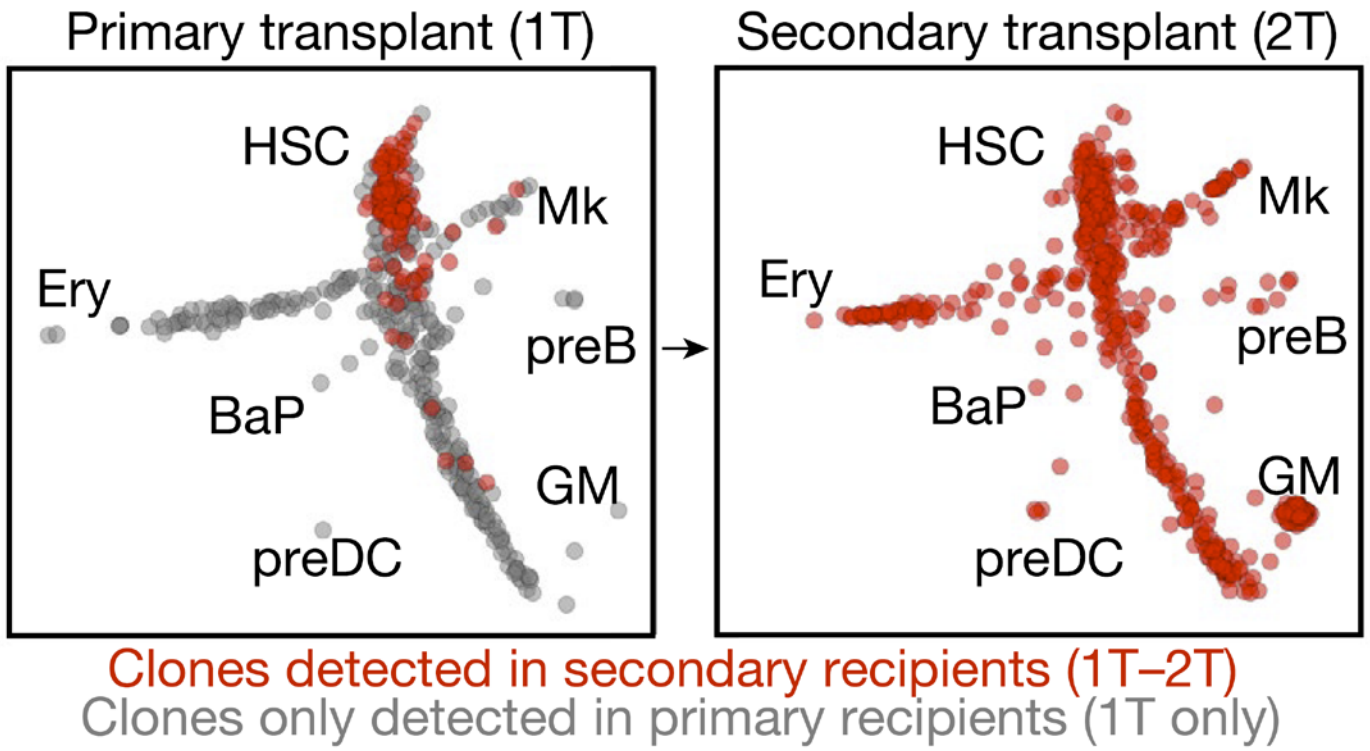
→ How well do clones engraft and expand?



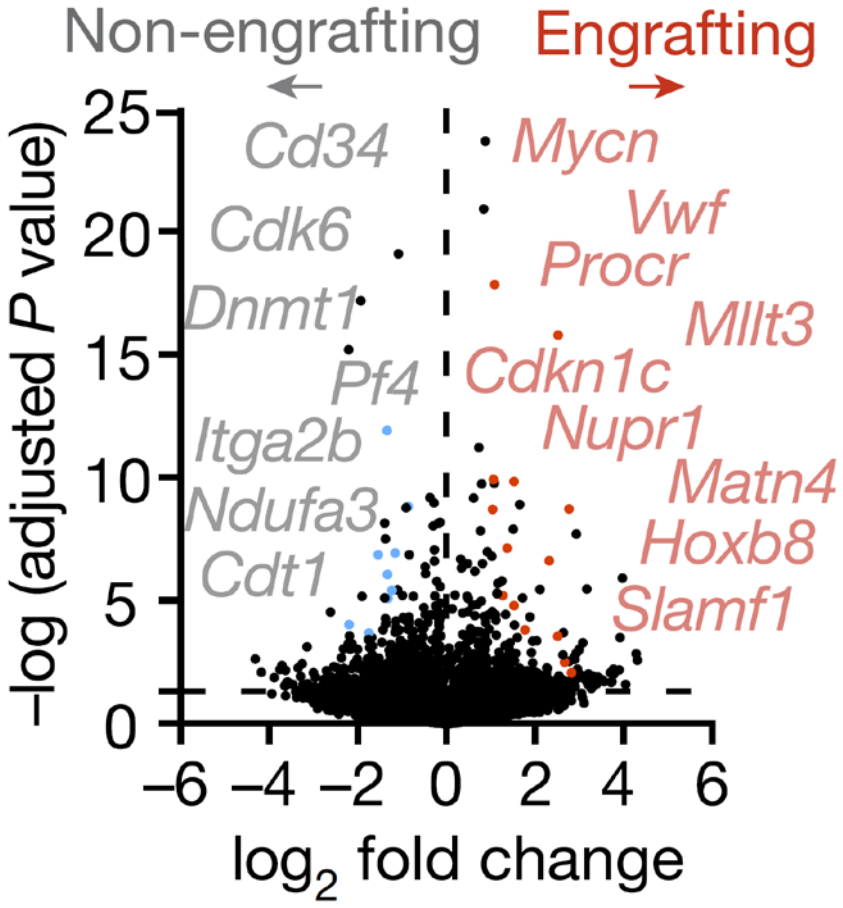
There was a good correlation of clonal expansion between the two recipient mice (split experiment).



Low-output clones showed better expansion after transplantation!

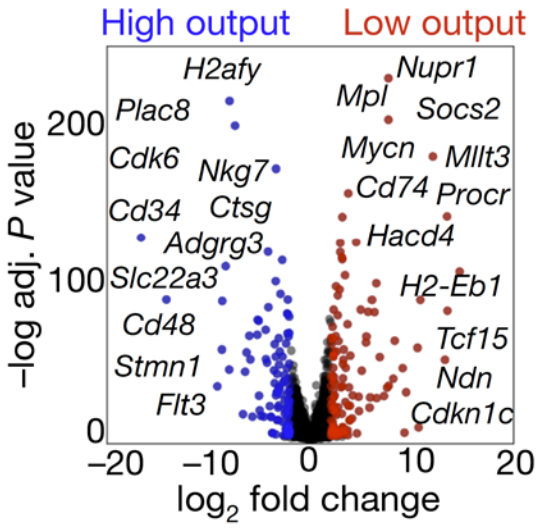


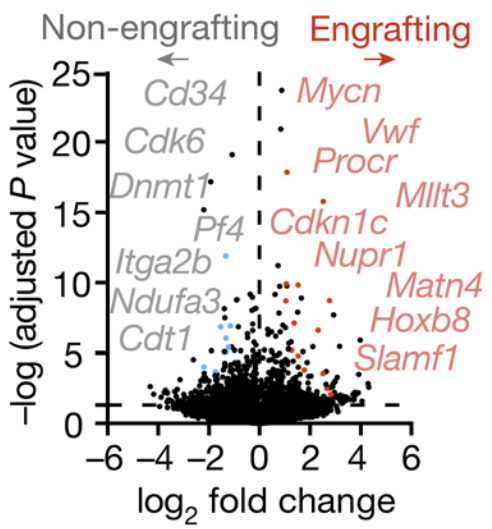
High-output clones (which had been responsible for most mature progeny) were lost…



Genes associated with clonal engraftment were identified.

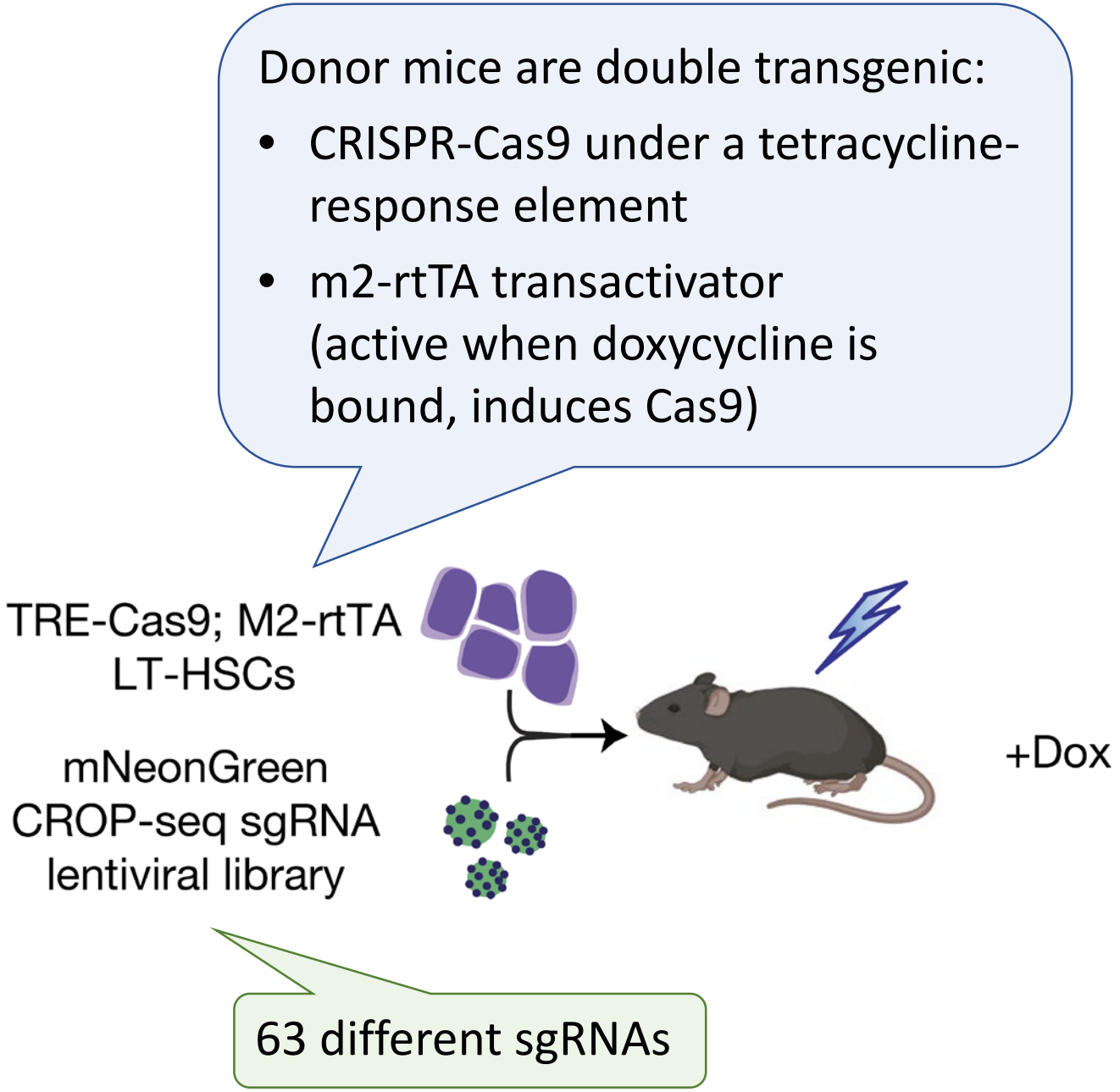
- The authors merged the two gene lists (output and engraftment) to select **63 genes** for further analysis.
  - Genes not specific to LT-HSCs were excluded.
  - Most genes already implicated in HSC maintenance were excluded → novelty



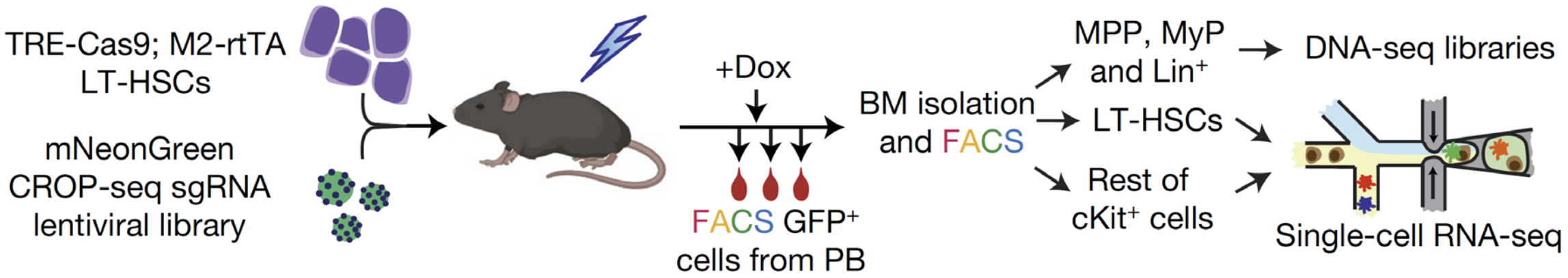


- The genes were used for an in **vivo CRISPR screen** (based on CROP-seq).

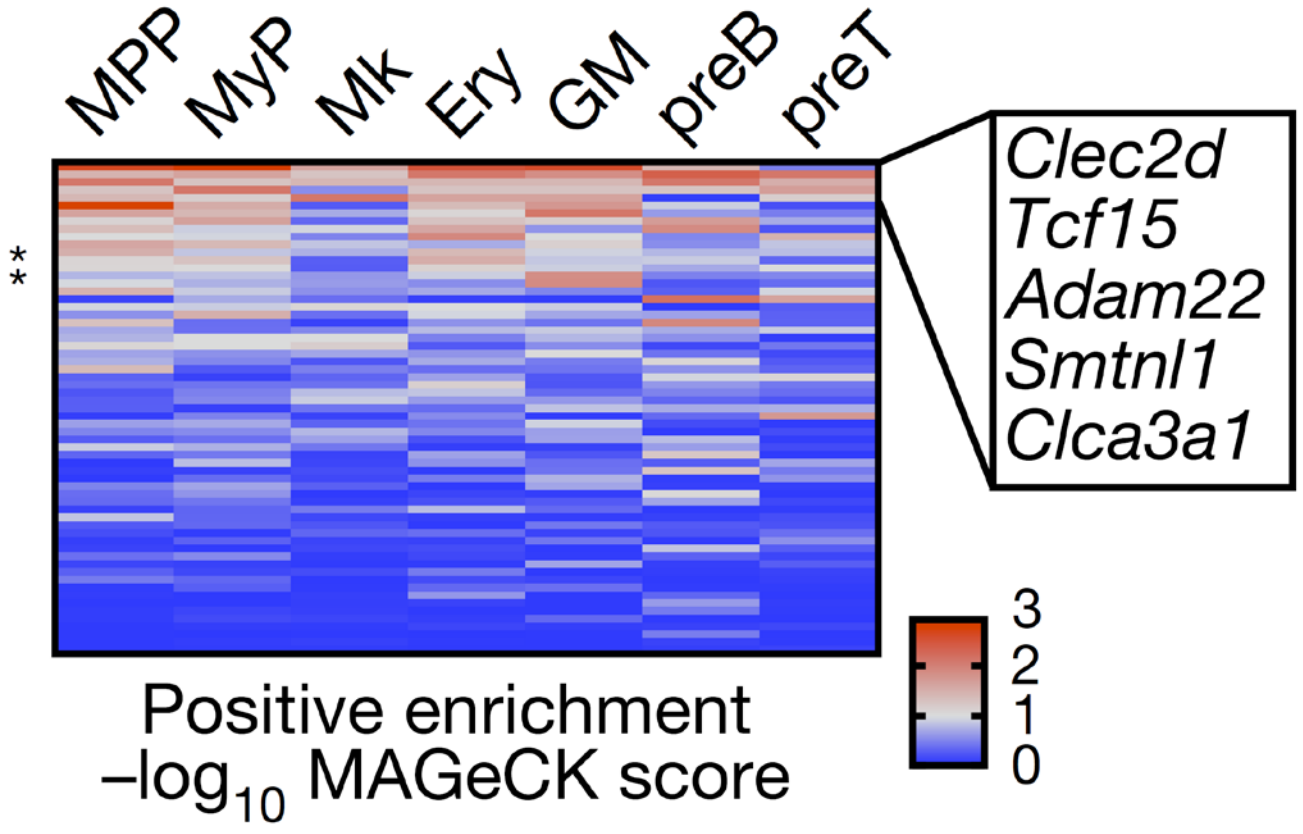
- CROP-seq allows the readout of pooled CRISPR screens using standard single-cell RNA-seq.
- With the CROP-seq vector, the gRNA becomes part of a larger mRNA transcribed by RNA Polymerase II (poly-A tail → detectable).



The screen was to designed to discover genes whose knockout causes HSCs to **increase mature/progenitor output** *in vivo*



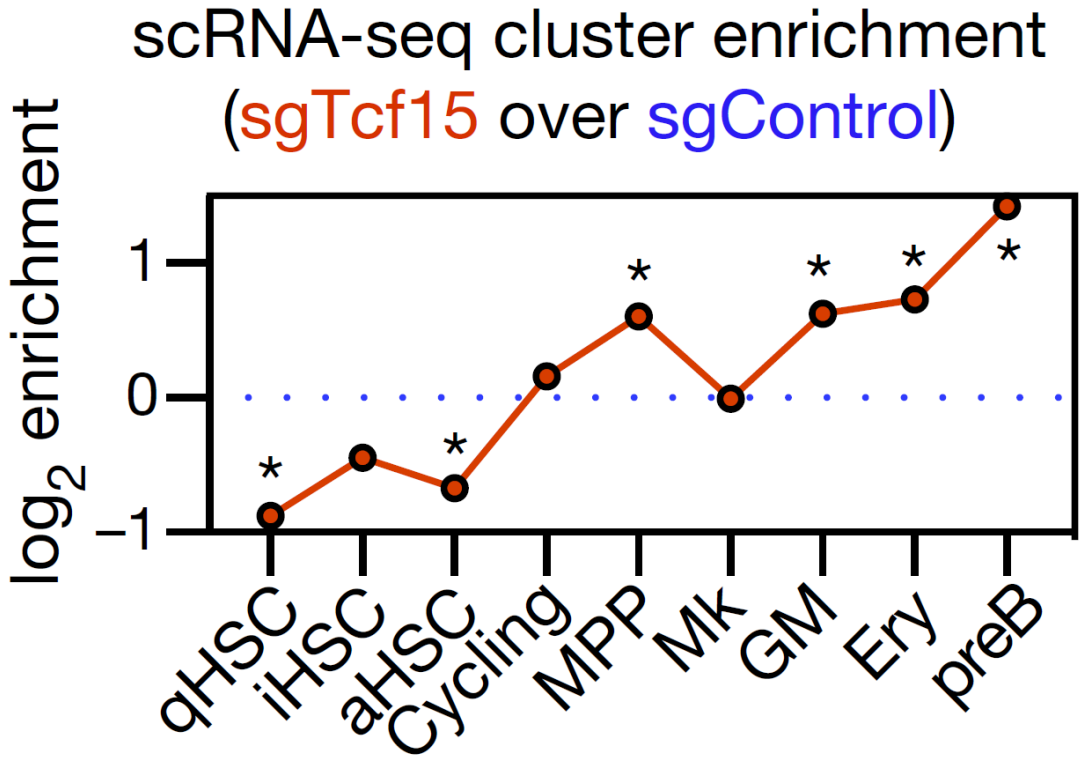
Doxycycline treatment (to induce Cas9) began 16 weeks after transplantation and continued for ~ 2 months.

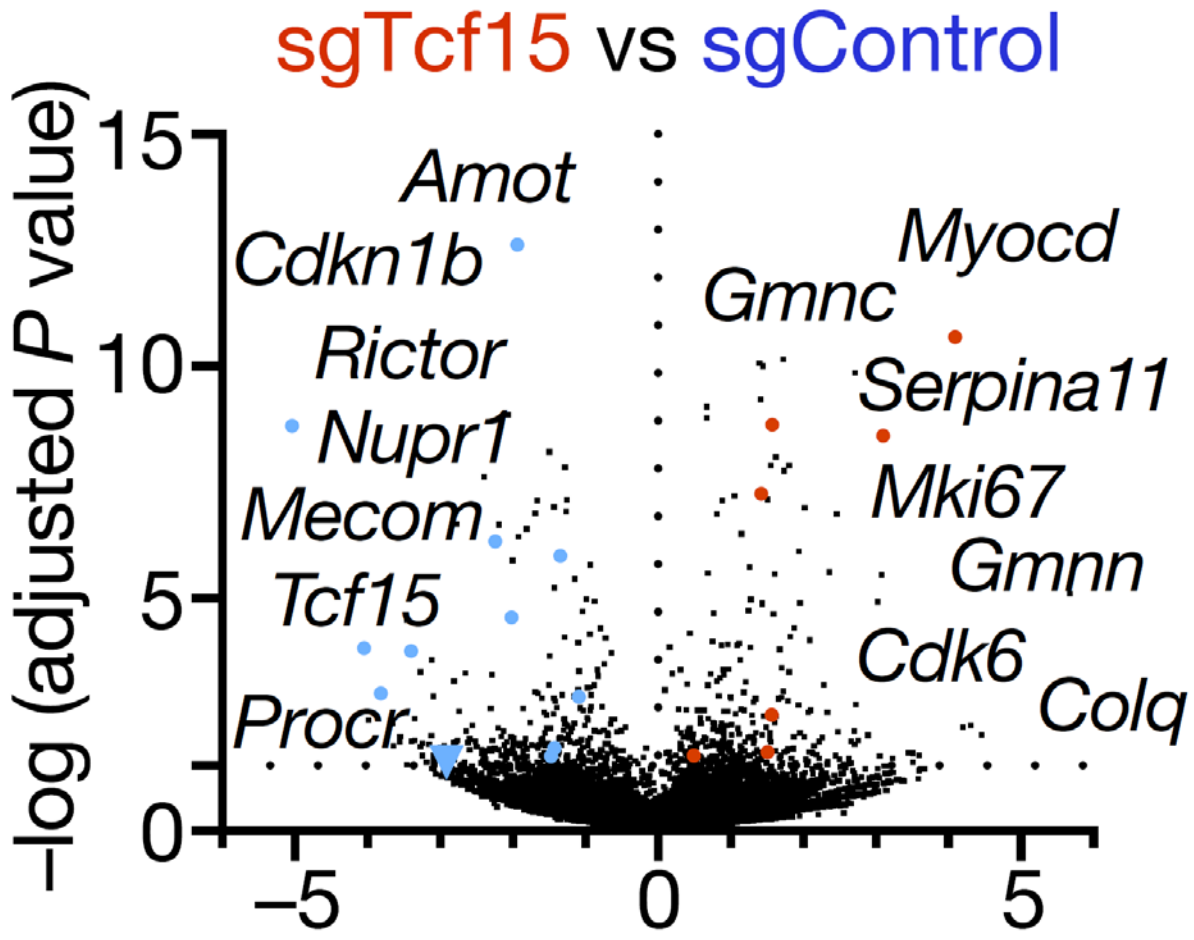


5 top genes were identified whose knockout increased HSC output (sgRNAs found to be enriched in progenitor or mature cells using deep DNA-seq)

The transcription factor **Tcf15** was the most consistent hit across replicates, and has not been described in haematopoiesis so far.

Analysis of scRNA-seq data showed enrichment of Tcf15 gRNA in mature cells, and depletion in HSCs.

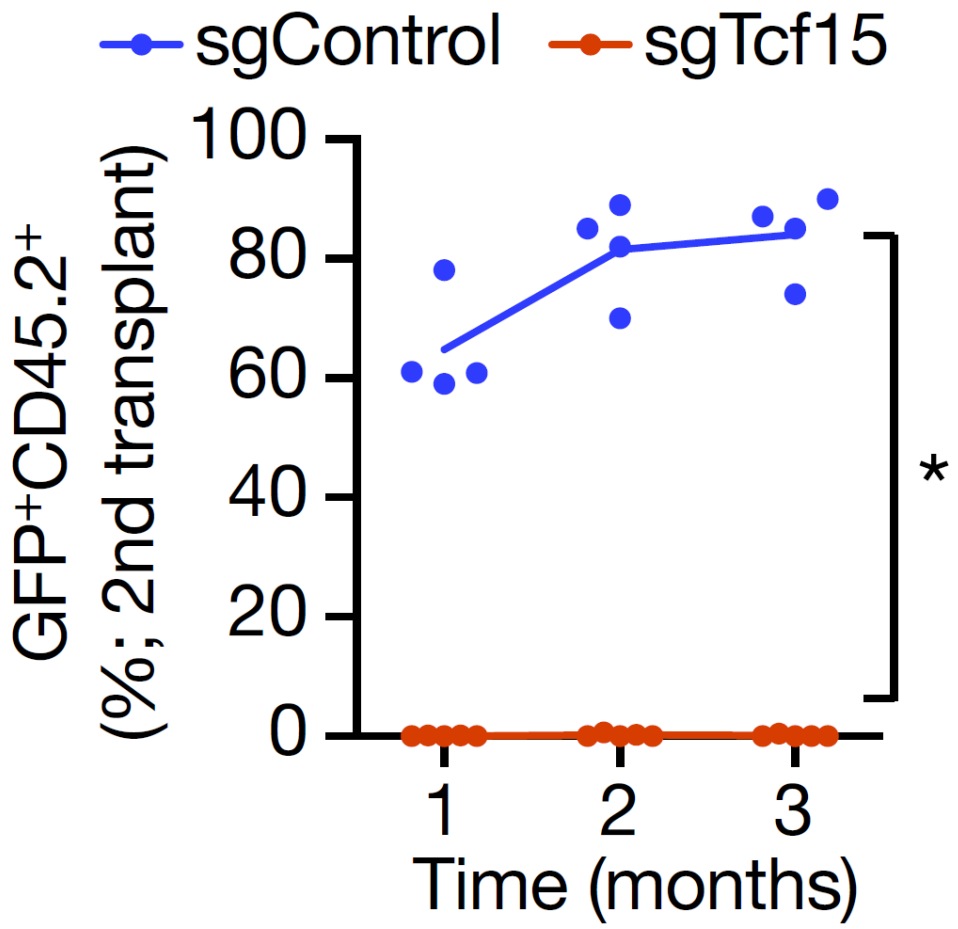




In Tcf15 knockout cells, Tcf15 is downregulated, along with quiescent HSC markers.

Cell cycle genes and markers of active HSCs are upregulated.

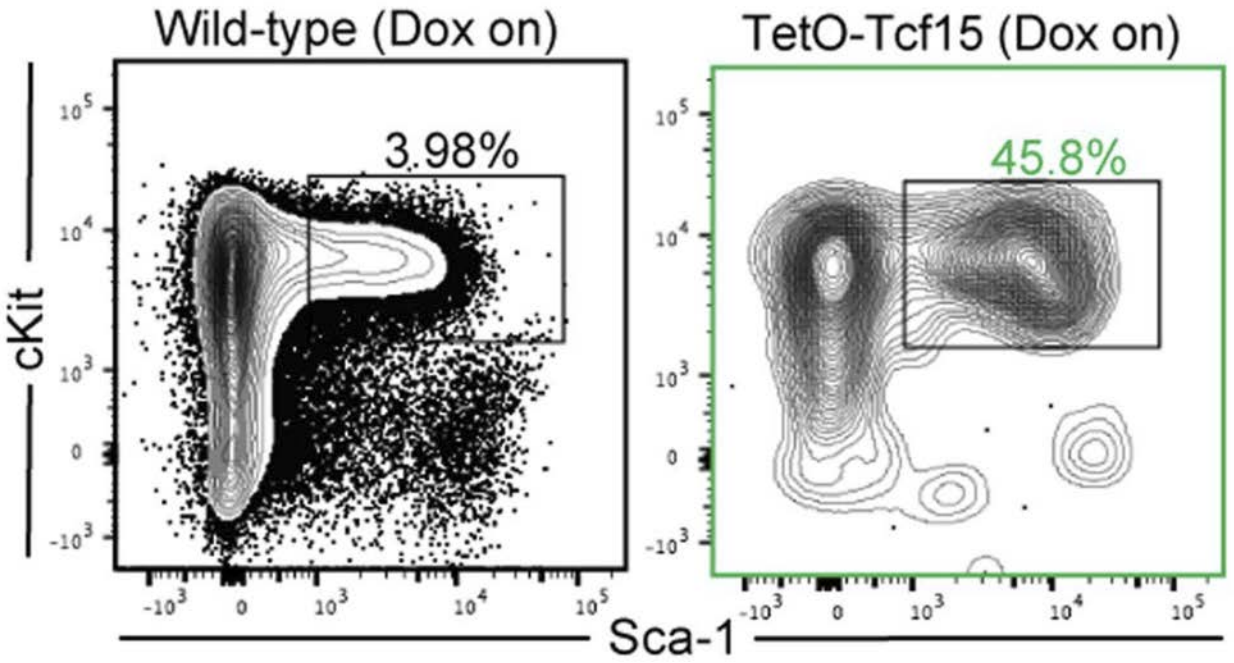
Tcf15 is necessary for HSC quiescence and long-term maintenance



Lentiviral-mediated CRISPR knockout of Tcf15 in HSCs abolished engraftment after a secondary transplantation.

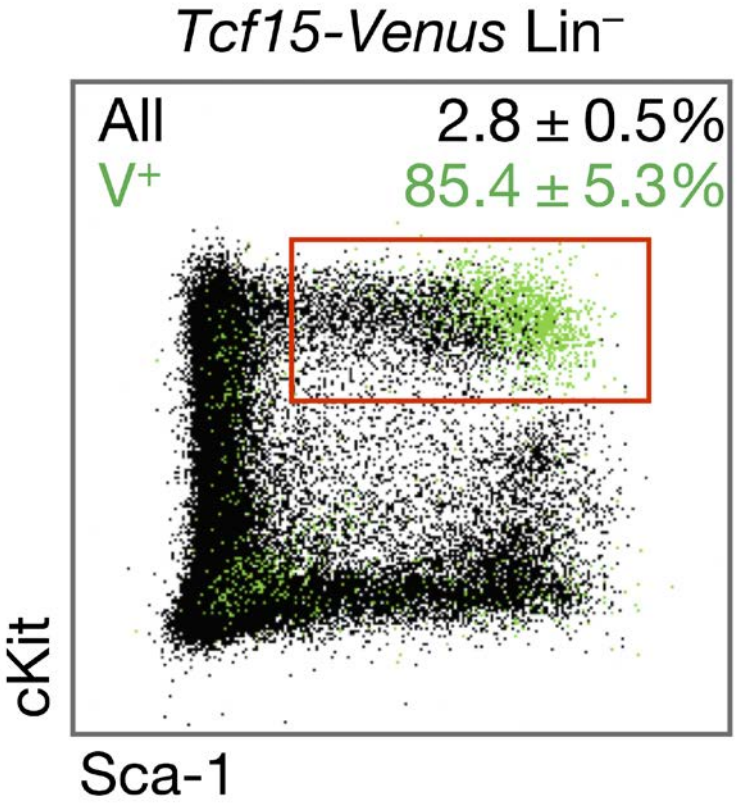
Tcf15 is sufficient for HSC quiescence

- Tcf15 was overexpressed using a doxycline-inducible lentiviral transgene.
  - The proportion of HSCs increased.
  - More mature populations were depleted.



Finally, a TCF15 reporter mouse was created → isolation of Tcf15^{+} cells

Heterozygous knockin mouse (TCF+ cells express YFP)

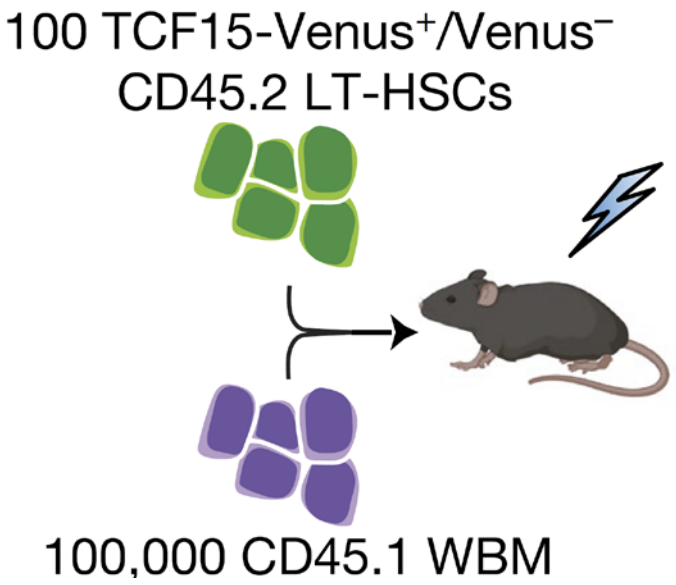


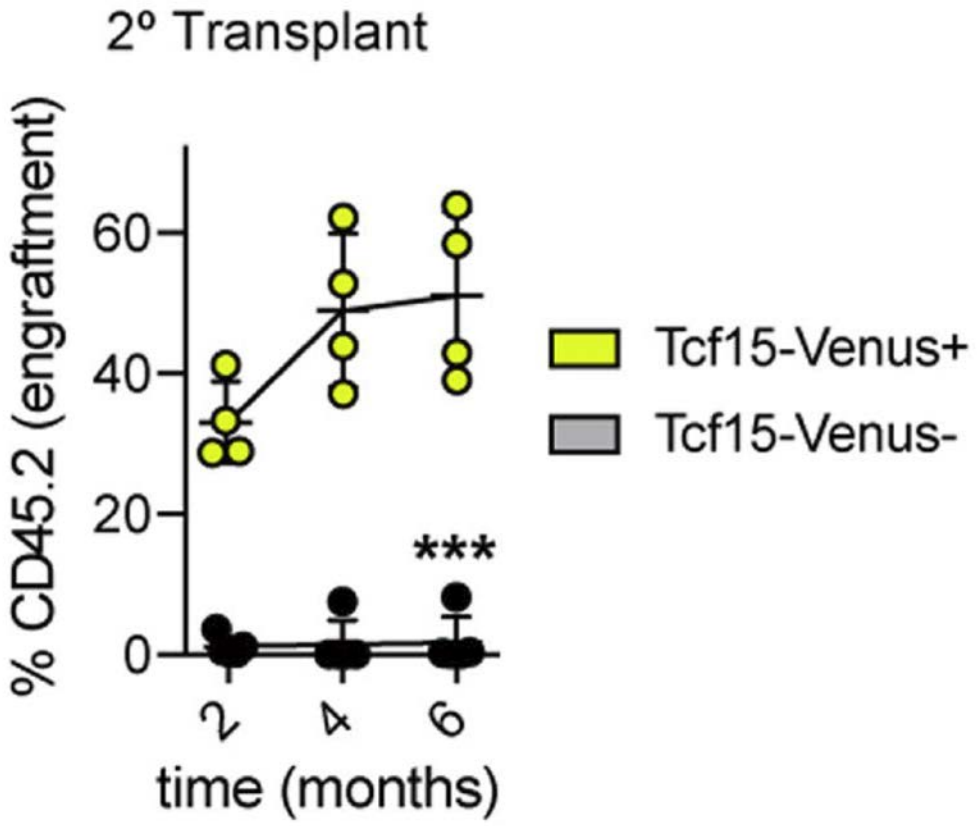
As expected,

TCF-V+ cells are

cKIT^{+} Sca-1^{+} HSCs

Competitive transplantation experiments show that only Tcf^{+} HSCs maintain their ability to reconstitute the hematopoietic system over the long term





Conclusions

- LARRY is an efficient system for lineage barcoding ex vivo.
  → useful for short- and long-term experiments
- Combining single-cell lineage and scRNA-seq data is an unbiased approach to discover new progenitor cell subtypes and differentiation trajectories.
  → more granular than scRNA data alone
- The functional relevance of progenitor-cell markers can be probed using CRISPR screens.
  → discover genes that influence a cell's decision
  (stay quiescent, proliferate, differentiate…)

Cell

Resource

An Engineered CRISPR-Cas9 Mouse Line for Simultaneous Readout of Lineage Histories and Gene Expression Profiles in Single Cells

Sarah Bowling, Duluxan Sritharan,
Fernando G. Osorio, ..., Stuart H. Orkin,
Sahand Hormoz, Fernando D. Camargo

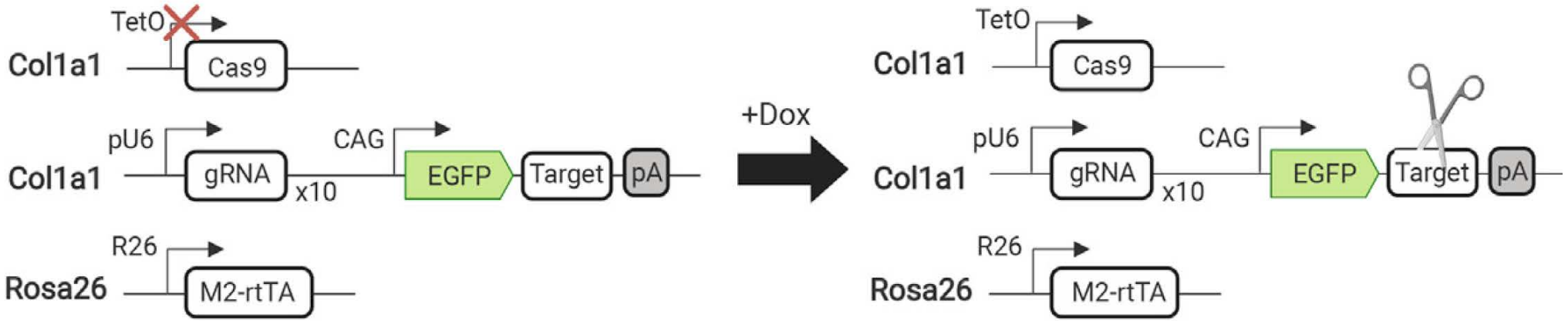
June 2020

The authors introduce CARLIN, a doxycycline-inducible system for *in vivo* barcoding in mice

The “barcode” is generated using CRISPR-Cas9.

- All cells expresses the same **10 guide RNAs**.
- They also contain a target region (**CARLIN array**) containing perfect-match target sites for each of the 10 gRNAs.
  - The CARLIN array is transcribed → readout by scRNA-seq
- **Efficiency of Cas9 cutting is < 100%**. Therefore, a different combination of target sites are modified in each cell.
- The repair outcomes of CRISPR-induced double-strand breaks are unpredictable.
  - The same target site may contain various mutations, insertions or deletions in different cells.

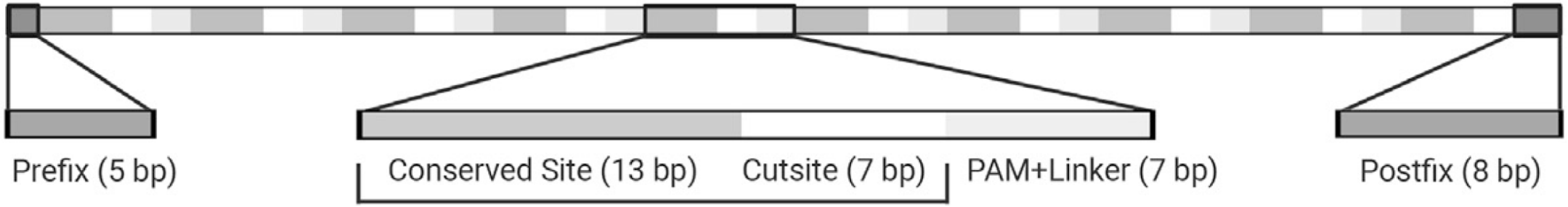
Treatment with doxycycline induces Cas9 expression.



By varying the dose and duration of doxycline treatment, one can modulate the percentage of cells and target sites that are modified.

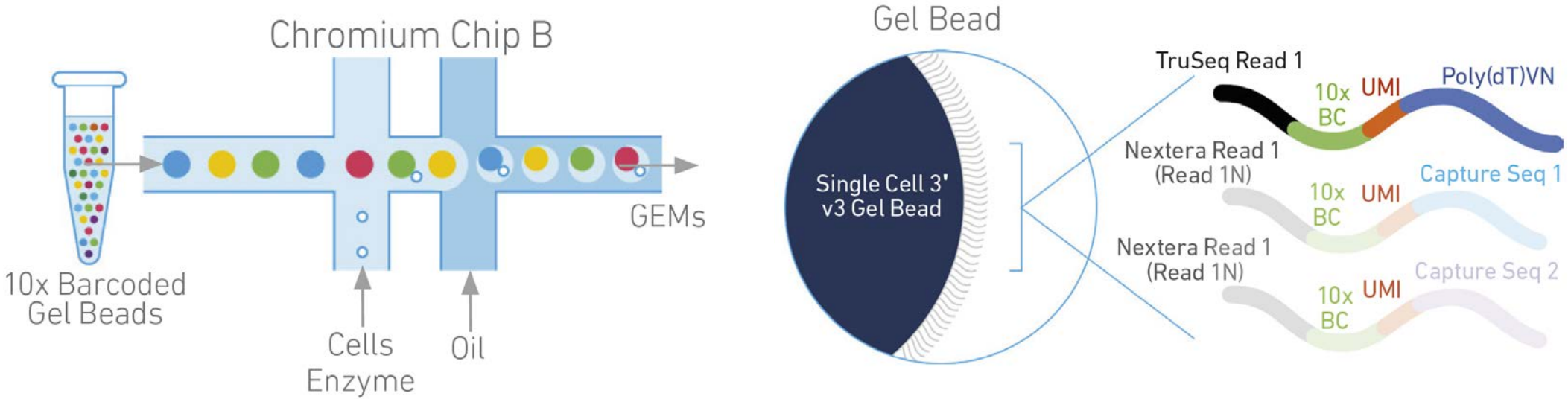
Sequential pulses of doxycycline treatment allows us to examine **clonal hierarchies**.

The sgRNA target sites are closely spaced within the CARLIN array



- Because some alterations are more common than others, and deletions spanning multiple target sites are frequent, many barcodes are ambiguous.
- The authors constructed a large database of barcode frequencies using bulk DNA-seq. Using this as a reference, the **significance of a barcode** can be established. Only significant clones are used in downstream analysis.

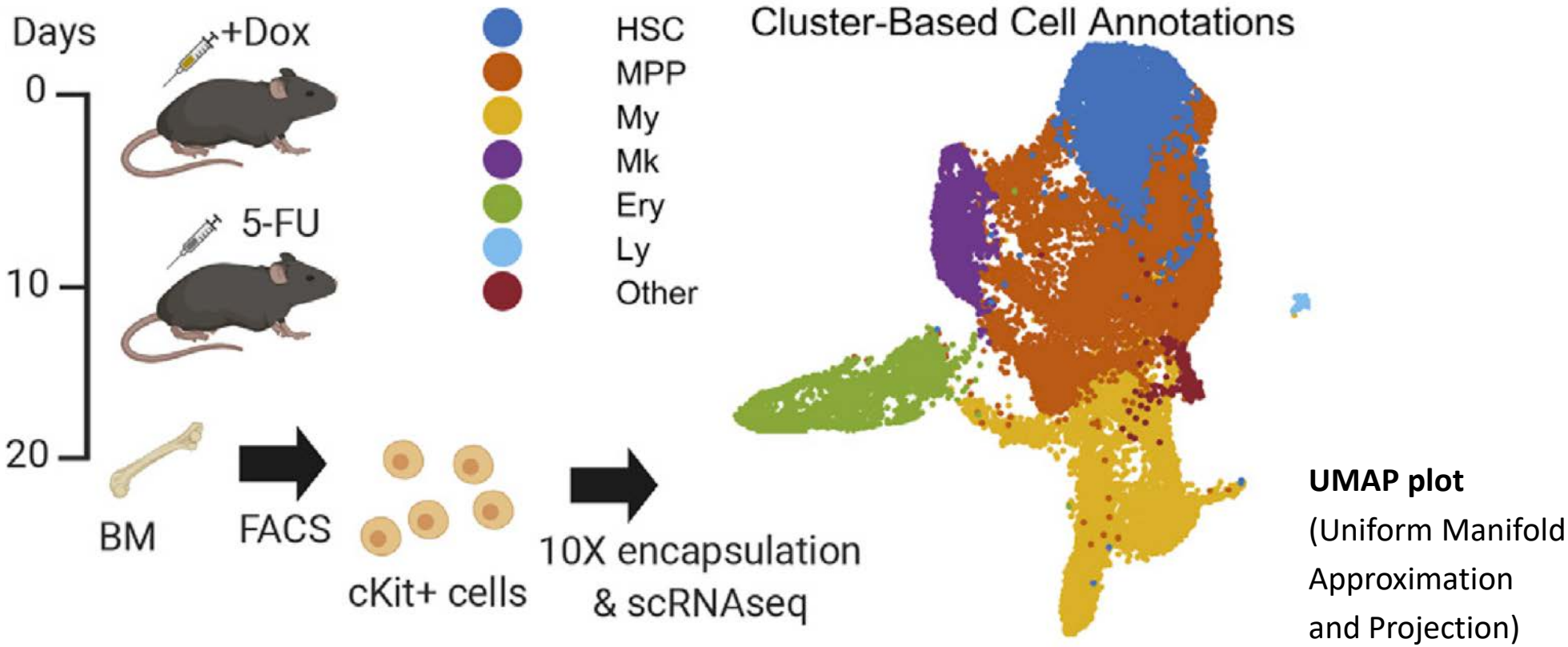
The authors wrote custom algorithms for denoising and filtering of cell barcodes, transcript barcodes (unique molecular identifiers) and CARLIN lineage barcodes in scRNA-seq data



Chromium microfluidics-based droplet scRNA-seq was used.

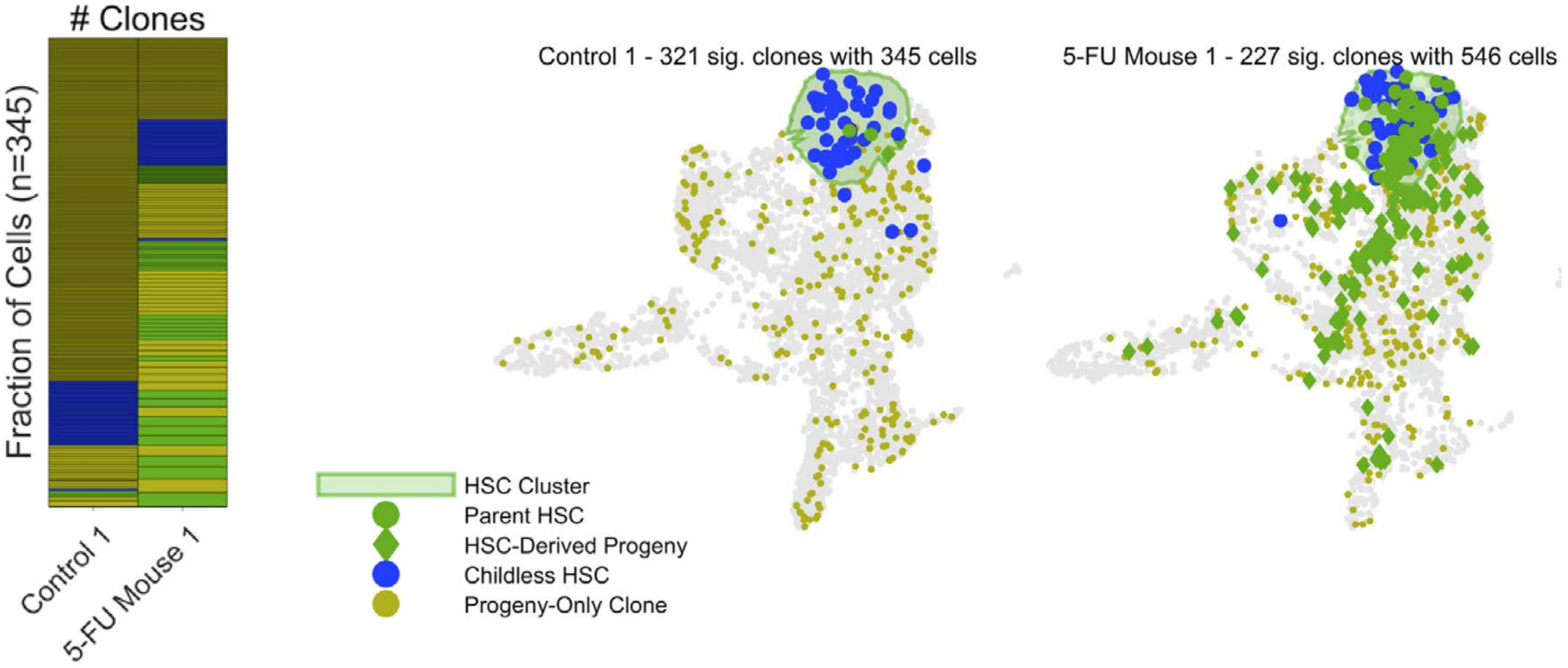
After cDNA transcription, CARLIN barcodes were selectively amplified using targeted PCR.

The 5-fluoro-uracil (5-FU) model of bone marrow ablation was used to examine clonal heterogeneity in marrow regeneration

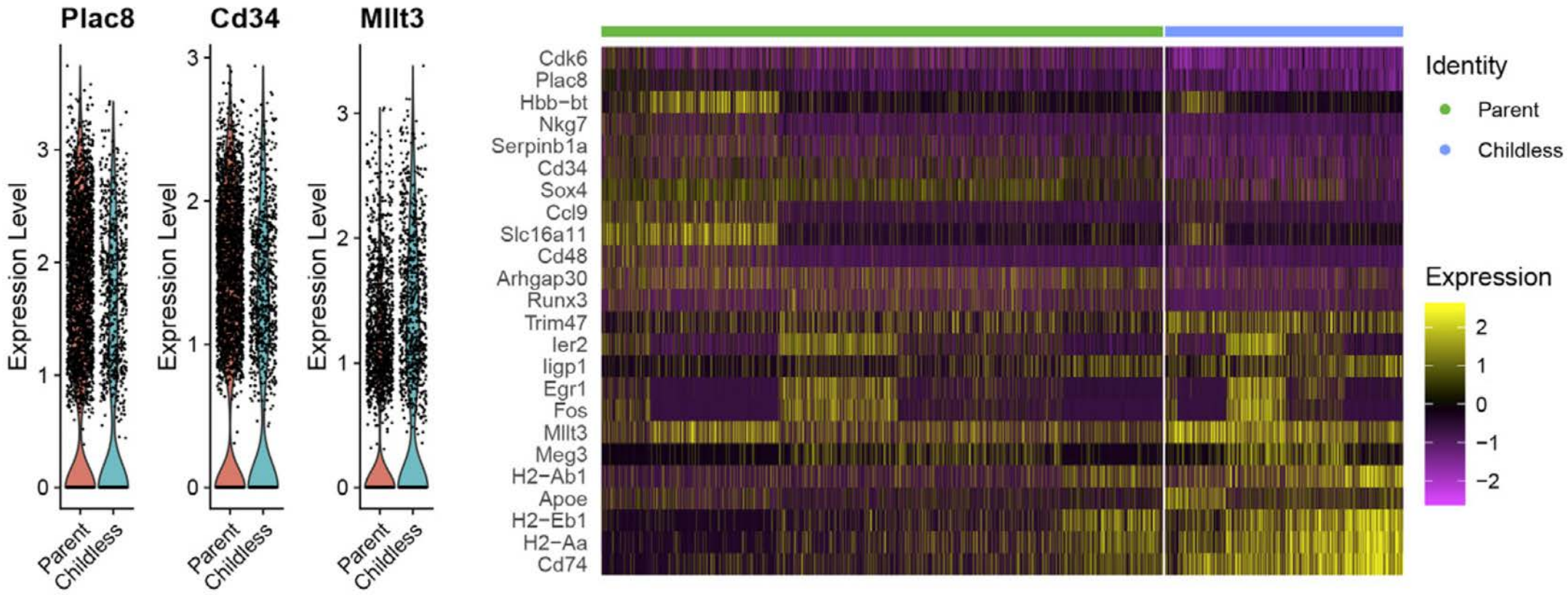


UMAP plot
(Uniform Manifold Approximation and Projection)

After 5-FU, the number of clones *decreased*, the size of individual clones *increased*, and HSC clones generated more mature progeny



By comparing “parent” and “childless” HSC clusters, authors identified signature genes of the “active” HSC state



The combined dataset was analysed (untreated and 5-FU mice).

Conclusions

- The bioinformatics analysis of CARLIN barcodes is far more complex.
- CARLIN has the advantage of being inducible *in vivo*, thus allowing single-cell lineage tracing without *ex vivo* manipulation and transplantation.

Thank you for your attention!