



# Identification and characterization of *in vitro* expanded hematopoietic stem cells

James L C Che<sup>1,2,3,†</sup> , Daniel Bode<sup>1,2,3,†</sup>, Iwo Kucinski<sup>1,2</sup> , Alyssa H Cull<sup>3</sup>, Fiona Bain<sup>3</sup> , Hans J Becker<sup>4,5</sup>, Maria Jassinskaja<sup>3</sup>, Melania Barile<sup>1,2</sup>, Grace Boyd<sup>3</sup> , Miriam Belmonte<sup>1,2</sup>, Andy G X Zeng<sup>6,7</sup>, Kyomi J Igarashi<sup>8</sup> , Juan Rubio-Lara<sup>3</sup>, Mairi S Shepherd<sup>1,2</sup>, Anna Clay<sup>1</sup>, John E Dick<sup>6,7</sup> , Adam C Wilkinson<sup>9</sup>, Hiromitsu Nakauchi<sup>4,5,8</sup>, Satoshi Yamazaki<sup>10,11</sup> , Berthold Göttgens<sup>1,2</sup>  & David G Kent<sup>1,2,3,\*</sup> 

## Abstract

Hematopoietic stem cells (HSCs) cultured outside the body are the fundamental component of a wide range of cellular and gene therapies. Recent efforts have achieved > 200-fold expansion of functional HSCs, but their molecular characterization has not been possible since the majority of cells are non-HSCs and single cell-initiated cultures have substantial clone-to-clone variability. Using the *Fgd5* reporter mouse in combination with the EPCR surface marker, we report exclusive identification of HSCs from non-HSCs in expansion cultures. By directly linking single-clone functional transplantation data with single-clone gene expression profiling, we show that the molecular profile of expanded HSCs is similar to proliferating fetal HSCs and reveals a gene expression signature, including *Esam*, *Prdm16*, *Fstl1*, and *Palld*, that can identify functional HSCs from multiple cellular states. This “repopulation signature” (RepopSig) also enriches for HSCs in human datasets. Together, these findings demonstrate the power of integrating functional and molecular datasets to better derive meaningful gene signatures and opens the opportunity for a wide range of functional screening and molecular experiments previously not possible due to limited HSC numbers.

**Keywords** hematopoiesis; hematopoietic stem cells; HSC expansion; self-renewal gene signature; single cell biology

**Subject Categories** Haematology; Methods & Resources; Stem Cells & Regenerative Medicine

DOI 10.15252/embr.202255502 | Received 27 May 2022 | Revised 21 July 2022 | Accepted 27 July 2022 | Published online 16 August 2022

EMBO Reports (2022) 23: e55502

## Introduction

Achieving efficient and controlled *in vitro* HSC expansion and defined mature cell production would have substantial therapeutic implications. HSC transplantation has been the bedrock of therapy in hematological malignancies for over 60 years and its success strongly correlates with the number of HSCs transplanted (Eaves, 2015). Increasing the purity of transplanted HSCs relative to mature cells would help reduce the likelihood of graft-versus-host disease (Lang & Handgretinger, 2008). Similarly, the expansion of functional HSCs outside the body would benefit gene therapy efforts for congenital hematopoietic diseases by preserving HSC function during genetic manipulation (Naldini, 2015) while also seeding *in vitro* production of virtually limitless numbers of mature blood cells, alleviating the need for blood cell donations (Batta *et al*, 2016).

Decades of research have identified a wide range of intrinsic genetic regulators that substantially increase HSC expansion *in vitro*, including *Hoxb4*, *Fbxw7*, *Dppa5*, *Prdm16*, among others (Antonchuk *et al*, 2002; Deneault *et al*, 2009; Iriuchishima *et al*, 2011; Miharada *et al*, 2014). Despite significant increases in functional HSCs, these strategies uniformly required genetic integration, resulting in a risk of leukemic initiation via over-activation of

1 Wellcome MRC Cambridge Stem Cell Institute, University of Cambridge, Cambridge, UK

2 Department of Haematology, University of Cambridge, Cambridge, UK

3 Department of Biology, York Biomedical Research Institute, University of York, York, UK

4 Division of Stem Cell Biology, Distinguished Professor Unit, The Institute of Medical Science, The University of Tokyo, Tokyo, Japan

5 Institute for Stem Cell Biology and Regenerative Medicine, Stanford University School of Medicine, Stanford, CA, USA

6 Princess Margaret Cancer Centre, University Health Network, Toronto, ON, Canada

7 Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada

8 Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA

9 MRC Weatherall Institute of Molecular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford, UK

10 Division of Stem Cell Biology, Center for Stem Cell Biology and Regenerative Medicine, The Institute of Medical Science, The University of Tokyo, Tokyo, Japan

11 Laboratory of Stem Cell Therapy, Faculty of Medicine, University of Tsukuba, Ibaraki, Japan

\*Corresponding author. Tel: (+44) 1904 328847; E-mail: david.kent@york.ac.uk

†These authors contributed equally to this work

self-renewal programs or blockage of differentiation. To overcome this, numerous groups have explored transient use of extrinsic regulators such as hematopoietic cytokines, growth factors, and small molecules to increase HSC self-renewal (Zhang & Lodish, 2008; Fares et al, 2014; Wohrer et al, 2014; Wen et al, 2020). These efforts have culminated in a fully defined culture system that expands mouse HSCs > 200-fold over a 28-day period (Wilkinson et al, 2019).

Despite this significant breakthrough, several outstanding issues remain. First, single HSCs cultured under these conditions display considerable heterogeneity in terms of their expansion and functional transplantation outcomes, and there is currently no way to prospectively identify clones containing functional HSCs. Second, since the large majority of cells are not HSCs (even in a successfully expanded HSC culture), it is challenging to undertake any molecular experiments on purified populations of expanded HSCs. As a result, despite the urgent need to understand and manipulate the molecular program of an expanded HSC, there are no methods currently available to undertake these studies amidst a lack of robust markers to isolate functional HSCs *in vitro* (Zhang & Lodish, 2005).

Here, we describe novel methods for prospectively isolating and characterizing *in vitro* expanded HSCs. Using the *Fgd5* reporter mouse (Gazit et al, 2014) in combination with the HSC cell surface marker EPCR (Kent et al, 2009), we present a specific purification strategy for expanded HSCs and validate its functional utility in transplantation experiments. Combining the functional outcomes of these experiments with transcriptional profiling of the same clones split into expanded HSCs and their non-HSC counterparts, we report the first molecular profile of expanded HSCs. Finally, by integrating single-cell gene expression profiles of cycling and *in vitro* hibernating HSCs with a freshly isolated hematopoietic cell landscape, we identify a robust transcriptional signature that can identify functional HSCs irrespective of cellular source or activation status, including highly efficient enrichment of functional human HSCs within human RNA-seq datasets.

## Results

### *Fgd5* and EPCR mark stem cells *in vitro*

EPCR has previously been identified as a highly selective marker for HSCs *in vivo* (Kent et al, 2009) and has also been demonstrated to track expanded human HSCs in culture (Fares et al, 2017), but on its own, it is insufficient to obtain highly purified HSCs. The advent of numerous mouse HSC reporter strains (Gazit et al, 2014; Busch et al, 2015; Chen et al, 2016; Cabezas-Wallscheid et al, 2017; Tajima et al, 2017; Pinho et al, 2018) represents a potentially novel set of tools for improving the identification of expanded HSCs *in vitro*. Because *Fgd5* was previously described as a highly specific reporter for HSCs (Gazit et al, 2014) and to enrich for a subset of primitive HSCs in EPCR<sup>+</sup> cell fractions as well as immune-activated cells (Bujanover et al, 2018; Rabe et al, 2019), we investigated whether *Fgd5* and EPCR could, in combination, mark functional mouse HSCs *in vitro* after periods of culture.

Since expanded HSCs are actively cycling—unlike adult HSCs—we first assessed whether *Fgd5* was expressed on actively cycling HSCs *in vivo* by studying fetal liver (FL) HSCs. All phenotypic HSCs

(as defined by the E-SLAM gating strategy; Benz et al, 2012; Kent et al, 2009) in both bone marrow (BM) and FL were *Fgd5*-ZsGreen<sup>+</sup> (Fig 1A and B), suggesting that *Fgd5* and EPCR expression could mark actively cycling HSCs. Next, we utilized 10-day single-cell HSC expansion cultures to compare *Fgd5* and EPCR expression with traditional *in vitro* markers of expanded HSCs. We observed a strong correlation between the percentage of *Fgd5*-ZsGreen<sup>+</sup>EPCR<sup>+</sup> (FE<sup>+</sup>) cells and the Lin<sup>-</sup>Sca1<sup>+</sup>Kit<sup>+</sup> (LSK) phenotype in 10-day clones (Fig 1C). However, some clones with high LSK percentages had lower FE<sup>+</sup> percentages, suggesting that FE<sup>+</sup> could be more selective for phenotypic HSCs than LSK alone (Fig 1C). Similarly, 10-day single-cell cultures (Shepherd et al, 2018) with a wide-range of clone sizes showed that the FE<sup>+</sup> fraction of cells contained significantly higher proportions of LSK cells and that all of the larger, more differentiated clones lost both *Fgd5* and EPCR expression (Fig EV1A).

To test the functional HSC content of FE<sup>+</sup> cells within the culture, we cultured E-SLAM HSCs for 3 days in conditions that maintain HSC function (Kent et al, 2008) and re-sorted *Fgd5*-ZsGreen<sup>high</sup>EPCR<sup>high</sup> (F<sup>hi</sup>E<sup>hi</sup>) and *Fgd5*-ZsGreen<sup>low</sup>EPCR<sup>low</sup> (F<sup>lo</sup>E<sup>lo</sup>) cells for transplantation into irradiated recipients (Fig EV1B). *Fgd5* and EPCR expression were correlated ( $r^2 = 0.5$ ; Fig EV1C) and even though fewer F<sup>hi</sup>E<sup>hi</sup> cells were transplanted (583 vs. 1,509 cells), only mice receiving F<sup>hi</sup>E<sup>hi</sup> cells displayed long-term multilineage reconstitution, indicating that *Fgd5* and EPCR expression are retained on functional HSCs *in vitro* (Figs 1D and EV1D). In addition, FE<sup>+</sup> cells were also more numerous in single HSC cultures with higher levels of expansion (e.g., cultures containing PVA compared to those containing HSA; Wilkinson et al, 2019; Figs 1E, and EV1E and F), thereby supporting the use of FE<sup>+</sup> as a simple two-color screening tool for functional HSC content *in vitro*.

Recently, a fully defined cell culture protocol was reported to expand HSCs between 236- and 899-fold over a 28-day period (Wilkinson et al, 2019). However, since individual clones showed substantial heterogeneity in clone size and functional output, we set out to determine whether the FE<sup>+</sup> strategy could help prospectively identify clones containing functional HSCs. After 28 days of culture, single HSC clones were harvested for transplantation (Fig EV1G–I) and 10% of each clone was collected for flow cytometric analysis. At day 28, the percentage of FE<sup>+</sup> cells strongly correlated ( $r^2 = 0.6658$ ) with functional HSC content as measured by transplantation (Fig 1F). The addition of LSK markers further enhanced the correlation ( $r^2 = 0.7425$ ; Fig 1G and H). Taken together, these results indicate that *Fgd5* and EPCR can reliably identify clones containing functional HSCs in both short- and long-term cultures.

### Linked functional and gene expression assays enable analyses of heterogeneous HSC clones

To understand the molecular drivers of heterogeneity in functional HSC expansion, we performed simultaneous functional and molecular assessment of 20 single HSC-derived 28-day clones. Individual clones were expanded for 28 days and cells were fractionated into phenotypic HSCs and non-HSCs, with 50% of the sorted cells used for transplantation and 50% for RNA-sequencing (RNA-seq; Fig 2A). For these experiments, although *Fgd5*<sup>ZsGreen<sup>+</sup>ZsGreen<sup>+</sup></sup> mice were used for immunophenotypic profiling of the clones, *Fgd5* was not used in the gating strategy of the re-sort for two reasons: (i)

EPCR and *Fgd5* levels correlated strongly within the LSK fraction and (ii) reliance on a reporter mouse would decrease the broader applicability of the results and method. Therefore, for the re-isolation post culture, phenotypic HSCs were defined as EPCR<sup>+</sup>LSK (ELSK) cells and non-HSCs represented the remaining cell fraction (nonELSK; Fig EV2A).

Recipients of phenotypic HSCs from 8 of 20 (40%) clones displayed high levels of multilineage engraftment, which accords with the previously reported single clone engraftment frequency of 28.5% (Wilkinson *et al*, 2019; Fig 2B and C). Donor cell contribution in recipient mice did not correlate with absolute live cell numbers or absolute numbers of phenotypic HSCs within each clone

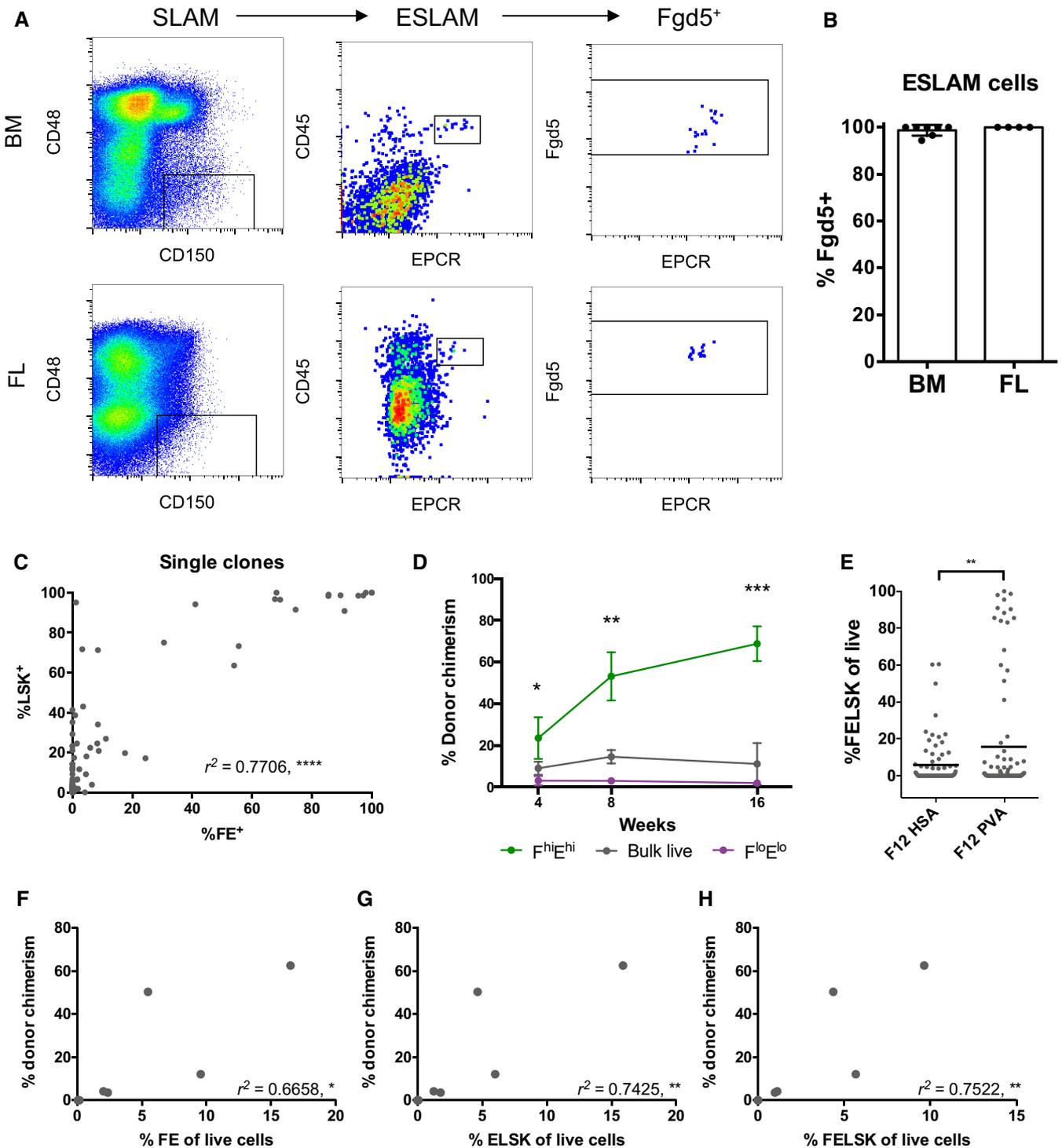


Figure 1.

**Figure 1. *Fgd5* and EPCR mark HSCs in short and long-term cultures.**

- A Representative flow analysis for adult BM and FL ESLAM HSCs.
- B Percentage of ESLAM HSCs that are *Fgd5*<sup>+</sup> in adult BM and FL ( $n = 7$  and 4 biological replicates, respectively). Error bars represent SD.
- C The correlation between %LSK and %*Fgd5*<sup>+</sup>EPCR<sup>+</sup> (FE<sup>+</sup>) cells in single-cell clones cultured for 10 days ( $n = 2$ ). Pearson correlation, \*\*\*\* $P < 0.001$ .
- D The percentage of donor chimerism in primary recipients over 16 weeks is displayed for F<sup>hi</sup>E<sup>hi</sup> cells ( $n = 3$ ), F<sup>lo</sup>E<sup>lo</sup> cells ( $n = 3$ ), and bulk live cells ( $n = 2$ ). t-Test, \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ .
- E The percentage of cells that are FLSK within each clone is displayed in the graph for 10-day F12 cultures containing HSA ( $n = 81$ ) or PVA ( $n = 89$ ). t-Test, \*\* $P < 0.01$ .
- F–H The correlation between donor chimerism and different phenotypic gating strategies for clones transplanted after 28-days of culture in F12 PVA medium supplemented with 10 ng/ml SCF, 100 ng/ml TPO, and 20 ng/ml IL-11. Pearson correlation,  $r^2$  values displayed, \* $P < 0.05$ , \*\* $P < 0.01$ .

( $r^2 = 0.0096$  and  $r^2 = 0.1894$ ), suggesting that HSC self-renewal is not linked with overall clonal proliferation (Fig 2D and E). In contrast, donor cell contribution was highly correlated with the percentage of ELSK cells present in the clone ( $r^2 = 0.8279$ ) and there was an even stronger correlation of %ELSK with donor cell contribution to the granulocyte-monocyte (GM) lineage, which is a reliable indicator of serial repopulating ability (Dykstra *et al*, 2007; Fig 2F and G). The addition of *Fgd5* to the gating strategy resulted in a slight improvement to the correlation ( $r^2 = 0.8642$ ; Fig EV2B), but for the reasons outlined above it was not used for cell isolation. Using these data, we identified a conservative > 20% cutoff for the %ELSK proportion that could reliably identify clones with functional HSCs. Transplantation of non-HSCs (nonELSK cells) from multiple clones was also performed and despite transplanting an average of 26-fold more cells per mouse, nonELSK cells largely lacked multilineage reconstitution capacity (Fig EV2C and D) with a calculated estimate of HSC frequency within the nonELSK fraction being less than 1 in 750,494 cells (Fig EV2E), strongly indicating that the vast majority of functional HSCs are contained within the ELSK fraction. The frequency of functional HSCs in the ELSK fraction, on the other hand, was much higher with single cell transplantation data from 28-day bulk cultures showing a very high rate of success. At 12 weeks post-transplantation, 33 of 50 (66%) single cell transplantations gave donor chimerism values of > 1% (Fig EV2F). Seventeen of 50 single-cell transplantations (34%) contributed > 0.5% to each of the GM, B, and T cell lineages (Fig EV2G–I). These data indicate that the ELSK fraction is highly enriched for HSCs (> 100,000 fold) compared to non-ELSK cells although a small minority of functional HSCs likely still exist outside of the ELSK gate.

The above experiments were performed by transplanting 50% of the HSCs from each clone. In order to further test the HSC expansion capacity of single cell-derived cultures, we selected seven clones for transplantation using just 5% of sorted ELSK cells (ranging from 3 to 48% of total cells in the clone). Out of the seven transplanted clones, only the two with the highest ELSK content showed successful multilineage reconstitution at 5% doses (Fig EV2J–L). Overall, our data strongly indicate that the ELSK phenotype can be used to reliably track functional HSC content in heterogeneous expansion cultures, and therefore acts as a robust tool for functional validation of the molecular characterization of expanded HSCs described below.

#### Expanded HSC clones share molecular features with freshly isolated HSCs

To characterize the molecular state of *in vitro* expanded HSCs and identify potential drivers of repopulating versus nonrepopulating

clones, RNA-seq was performed on 12 clones which were split into phenotypic HSCs (ELSK) and non-HSCs (nonELSK; Fig EV3A) and coupled with transplantation assays for functional validation. Clones were selected to represent a range of phenotypic HSC content and donor chimerism in transplantation assays, directly linking the functional HSC content to the transcriptional profile. To simplify the analysis, the samples were categorized into four groups: (i) ELSK cells from clones that repopulated mice (PosELSK); (ii) ELSK cells from clones that did not repopulate mice (NegELSK); (iii) nonELSK cells from clones that repopulated mice (PosNonELSK); and (iv) nonELSK cells from clones that did not repopulate mice (NegNonELSK; Fig 3A).

Of the 24 cell fractions, two samples failed quality control due to low read counts (Table 1). After removal of lowly expressed genes, 16,648 genes were detected across 22 unique samples. We first performed multiple dimensional scaling (MDS), which showed clear separation between samples originating from clones which repopulated mice and clones that did not (Fig 3B). Samples could be further resolved by whether they were phenotypic HSCs (ELSK cells) or not (nonELSK cells). Notably, the nonELSK fraction of repopulating clones overlapped with the profiles of ELSK cells from nonrepopulating clones (negELSK), suggesting that molecular profiles are more closely linked to cellular function than cell surface immunophenotype (Fig 3B). In line with this observation, posELSK samples with an increasing proportion of donor chimerism and GM contribution clustered separately from negELSK cells (Fig EV3B).

In order to assess the similarity of repopulating ELSK cells to freshly isolated HSCs, we first computed the geometric mean for a previously described gene signature for freshly isolated HSCs (termed molecular overlap, or “MoIo”; Wilson *et al*, 2015). Here, increasing repopulation potency was closely correlated with the MoIo geometric mean score, and nonELSK cell fractions had significantly lower MoIo scores than ELSK cells (Figs 3B and EV3C). NonELSK cells from repopulating clones (PosNonELSK) showed higher MoIo scores than ELSK cells obtained from clones without functional HSCs (NegELSK), suggesting that the MoIo signature correlated with functional HSC content (Fig EV3C). Of note, several MoIo signature genes were below the minimum threshold of expressed genes across all samples (*Cldn10*, *Ramp2*, *Smtnl1*, *Sox18*, and *Sqrdl*), indicating that although these genes are expressed in freshly isolated HSCs (and may play a biological role in those cells), they are not highly expressed in *ex vivo* cultured HSCs. Overall, these data highlight the prospect of identifying functional HSCs with durable self-renewal and repopulation potency based on their transcriptional profiles.

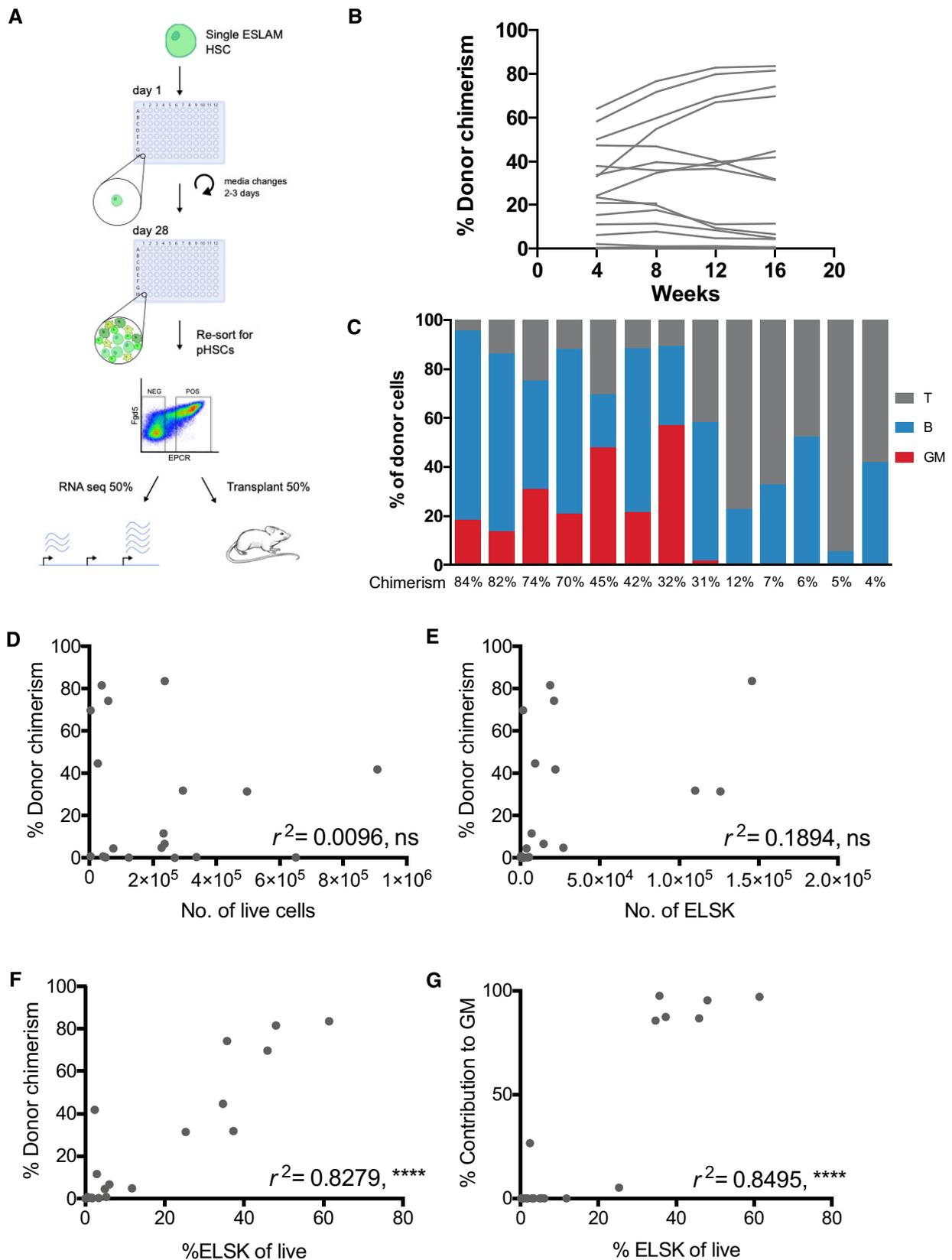


Figure 2.

**Figure 2. Reporter strategy deciphers clonal heterogeneity in expansion cultures.**

- A Schematic of experimental design in which single ESLAM HSCs were cultured for 28 days in F12 media containing PVA, 10 ng/ml SCF, and 100 ng/ml TPO. At day 28, 20 clones were harvested and re-sorted for phenotypic HSCs, defined as EPCR<sup>+</sup>Lin<sup>-</sup>Sca-1<sup>+</sup>C-kit<sup>-</sup> (ELSK) cells and remaining nonELSK cells. The two fractions were each split in half, 50% for transplantation and 50% for bulk RNA-sequencing. On average for each clone, 22,382 ELSK cells were sorted compared to 90,700 nonELSK cells.
- B The donor chimerism in animals receiving ELSK cells re-sorted from the 20 clones (45–50% dose). One mouse was culled for health reasons before experimental endpoint.
- C The proportion of donor GM, B, and T cells in each clone above 1% donor chimerism at week 16 with overall donor chimerism listed underneath each bar.
- D–G The correlation between donor chimerism and absolute numbers and proportions of live cells or FLSK cells. Pearson correlation, \*\*\*\**P* < 0.001.

### Nonrepopulating clones express mature cell gene signatures and lose HSC gene expression signatures

To first identify the dominant cell types in isolated cell fractions, we computed the correlation of each fraction with previously defined gene expression profiles for a broad spectrum of hematopoietic stem and progenitor cell types (curated within the ImmGen database; Shay & Kang, 2013). While all ELSK cell fractions were correlated with short-term HSCs (ST34F), only repopulating ELSKs were specifically correlated with long-term HSCs (LT-HSC; Fig 3C). Despite the stark differences in HSC functional content and associated gene signatures, a direct comparison of posELSK to negELSK cells revealed a limited set of 44 differentially expressed genes (Figs EV3D and E), with negELSK enriched for differentiation-associated gene ontology (GO) terms and posELSKs enriched for cell surface GO terms (Fig EV3F).

The non-HSC fraction (nonELSK) on the other hand, were subject to a wider range of transcriptional differences between repopulating and nonrepopulating clones (Appendix Fig S1A), suggesting that the cellular composition of each clone might affect HSC expansion. In order to gain a better understanding of the cellular compositions of the *in vitro* expanded clones, we performed 10× RNA-seq on a culture initiated with 50 LT-HSCs. In accordance with the earlier ImmGen analysis, the majority of cells shared a transcriptomic signature with primitive hematopoietic cells, with a small minority of single cells expressing mature myeloid genes and an even smaller proportion expressing lymphoid genes (Fig 3D). Cell state transitions of this *in vitro* progenitor cell production were also predicted to share a high degree of similarity with previously mapped *in vivo* differentiation trajectories, including the HSC fraction bearing the molecular signature of the G<sub>1</sub>/G<sub>0</sub> phases of the cell cycle (Appendix Fig S1B and C). A customized online resource allows full exploration of these data ([http://128.232.224.252/gp\\_apps/CheBode2021/](http://128.232.224.252/gp_apps/CheBode2021/)). In order to infer cell identities from the bulk transcriptomes, we computed direction of state transition (DoT) scores (Kucinski *et al*, 2020) for PosNonELSK and NegNonELSK fractions and projected these onto a previously defined single-cell hematopoietic landscape, using differentially expressed genes (DEGs) of matched ELSK and nonELSK cells within each clone (Fig 3E and Appendix Fig S1D). While both ELSK fractions were enriched for genes expressed in HSCs, both nonELSK fractions showed enrichment for genes associated with myeloid cell types such as monocytes, neutrophils, and basophils (Fig 3E, Appendix Fig S1E). Interestingly, only the non-HSC fractions from nonrepopulating clones (NegNonELSK) showed enrichment of megakaryocyte and erythrocyte genes (Fig 3E). Overall, these results indicate that nonrepopulating clones undergo increased myeloid and megakaryocyte-erythroid differentiation, and that these cell types may negatively regulate HSC self-renewal.

### A molecular signature for expanded HSCs

In order to identify a gene signature most strongly associated with *ex vivo* expanded functional HSCs, we derived the PCA-based dimensionality reduction for all bulk transcriptome samples and computed Pearson correlations of transplantation metadata with each principal component (PC) and the associated loading plots (Fig 4A). In line with the previous MDS plots, repopulating and nonrepopulating samples showed distinct clustering (Fig 4B). Intriguingly, a single PC (PC1, 35.62% of variation) was significantly correlated with donor chimerism, GM contribution, and a binary repopulation outcome score (Figs 4C and EV4A). We identified the top 50 genes driving PC1 (Fig 4D) and further curated the gene signature using fitted logistic and linear regression models for each transplantation parameter (Figs 4E, and EV4B and C) to identify the most significant drivers of repopulation potential. The resulting “repopulation signature” gene list (RepopSig) contains 23 genes (Fig 4F), including previously described HSC markers and self-renewal regulators, such as *Esam*, *Slamf1* (CD150), and *Prdm16* (Deneault *et al*, 2009; Yokota *et al*, 2009; Oguro *et al*, 2013; Gudmundsson *et al*, 2020), as well as novel genes that were not previously associated with HSC self-renewal, such as *Klhl4*, *Mpdz* and *Insyn1*. Next, we computed the geometric mean for the RepopSig across all samples, confirming a robust identification of repopulating clones (Figs 4G and EV4D). Intriguingly, the RepopSig gene signature score improved the distinction of repopulating samples when compared to the MoIO signature (Figs 4H, and EV4D and E). Of note, a subset of MoIO signature genes was not correlated with the RepopSig, possibly indicating their limited role for the function of *ex vivo* expanded HSCs (Fig EV4F).

The top signaling pathway enriched among genes closely correlated with the RepopSig (*r* > 0.75) was sphingolipid signaling, which was recently implicated in human HSC self-renewal (Xie *et al*, 2019; Fig 4I). A number of other enriched pathways, namely Hippo, FoxO, Ras, and VEGF, involve RhoGTPase signaling and several previous studies in mouse HSCs have specifically implicated CDC42 (Florian *et al*, 2018; Liu *et al*, 2019) and ARHGAP5 (Hinge *et al*, 2017) in mouse HSC function (Fig 4I). To test the role of RhoGTPase signaling in regulating HSC expansion, we undertook expansion cultures with or without various RhoGTPase inhibitors (CASIN, NSC23766, Rhosin) or an activator (ML099; Gao *et al*, 2004; Surviladze *et al*, 2010; Shang *et al*, 2013; Liu *et al*, 2019). Inhibitors uniformly decreased the percentage of phenotypic HSCs (ELSK cells) in a dose-dependent manner and in some cases resulted in substantially reduced survival (Figs EV4G and H). Activating RhoGTPase signaling with ML099, on the other hand, did not alter HSC expansion, suggesting that increased RhoGTPase

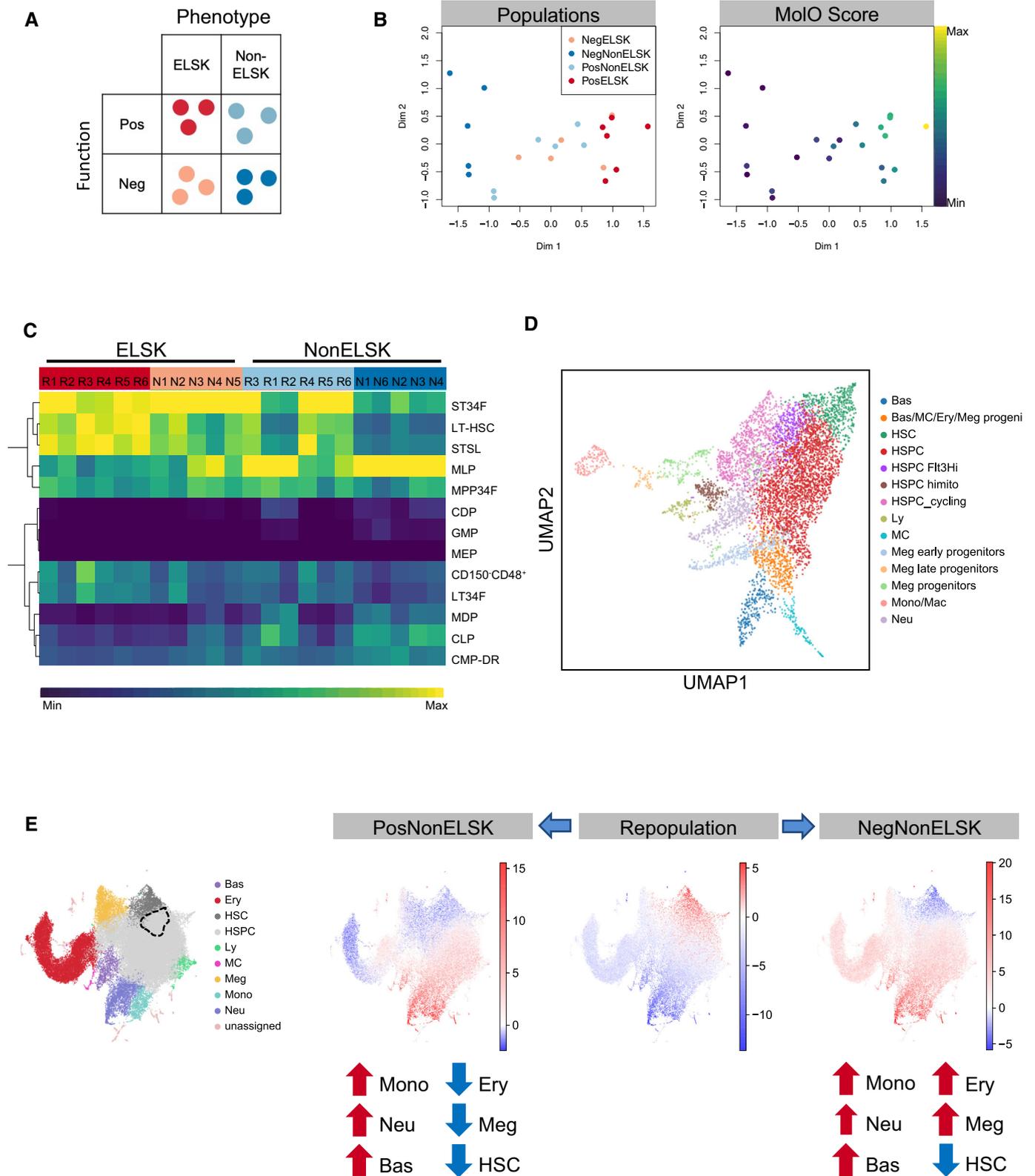


Figure 3.

**Figure 3. Expanded HSC clones are transcriptionally similar to freshly isolated HSCs.**

- A Schematic of color-coded cell population categories. Repopulation was defined as having > 1% donor chimerism and > 1% contribution to GM at 16 weeks.
- B MDS plot of bulk RNA sequencing samples colored by their population categories and their corresponding MoIO score.
- C Correlation of each sample to gene expression profiles of various hematopoietic stem and progenitor cell populations, as defined in the Immgen database.
- D UMAP depiction of single cell profiling of a 28-day culture initiated by 50 HSCs. Cell type annotations were derived using marker gene signatures.
- E UMAP representation of mouse LK/LSK (Dahlin *et al*, 2018) single cell transcriptomes colored by DoT scores computed using DEGs of PosELSK against PosNonELSK (left); NegELSK (middle); NegNonELSK samples (right). Dominant differentiation trajectories are indicated by positive DoT scores (red), while negative DoT scores (blue) outline underrepresented lineages in expanded clones. Enriched populations are indicated by red arrows and underrepresented populations indicated by blue arrows. The point of origin is marked by a dotted line.

**Table 1. Clones with matched ELSK and nonELSK samples were chosen for RNA-seq.**

Clone	Repop-ulation	Fraction	Group	16 weeks Donor Chimerism	16 weeks contribution to GM
R1	Y	<sup>a</sup> ELSK	<sup>a</sup> PosELSK	83.50%	97.00%
		NonELSK	PosNonELSK		
R2	Y	ELSK	PosELSK	81.50%	95.40%
		NonELSK	PosNonELSK		
R3	Y	<sup>a</sup> ELSK	<sup>a</sup> PosELSK	74.20%	97.50%
		NonELSK	PosNonELSK		
R4	Y	ELSK	PosELSK	44.60%	85.60%
		NonELSK	PosNonELSK		
R5	Y	ELSK	PosELSK	41.80%	26.60%
		NonELSK	PosNonELSK		
R6	Y	ELSK	PosELSK	31.80%	87.40%
		NonELSK	PosNonELSK		
N1	N	ELSK	NegELSK	0.67%	0.02%
		NonELSK	NegNonELSK		
N2	N	ELSK	NegELSK	0.40%	0.02%
		NonELSK	NegNonELSK		
N3	N	ELSK	NegELSK	0.22%	0%
		<sup>a</sup> NonELSK	<sup>a</sup> NegNonELSK		
N4	N	ELSK	NegELSK	0.17%	0.02%
		<sup>b</sup> NonELSK	<sup>b</sup> NegNonELSK		
N5	N	ELSK	NegELSK	0.11%	0%
		NonELSK	NegNonELSK		
N6	N	<sup>b</sup> ELSK	<sup>b</sup> NegELSK	0%	0%
		NonELSK	NegNonELSK		

<sup>a</sup>Indicates sample was run twice as a technical repeat.

<sup>b</sup>Indicates sample failed QC.

signaling is not sufficient to actively drive HSC self-renewal beyond the current limitations of the expansion system (Fig EV4G and H). These experiments further underscore the power of the ELSK reporter system to replace lengthy and expensive functional transplantation assays for validating such molecules for their effect on HSC expansion.

#### Repopulation signature identifies HSCs from multiple cellular states

Compared to the MoIO signature, the RepopSig score was better able to separate repopulating clones from nonrepopulating cells in our

initial experiments (Figs 4H and EV4D). However, since the RepopSig was initially derived from this training dataset, we next generated a validation dataset by an additional series of 28-day single-cell cultures with concomitant qPCR, flow cytometry, and transplantation assays. We selected 9 clones with > 20% ELSK and 10 with < 1% ELSK for parallel qPCR and transplantation assays. Functional HSC activity was exclusively restricted to clones with > 20% ELSK, where all nine had robust multilineage contribution in recipient animals (Fig 5A and B). Signature gene expression strongly correlated with clones that had a high percentage of phenotypic HSCs (> 20% ELSK) and genes that associated with negative transplantation outcomes were more highly expressed in clones

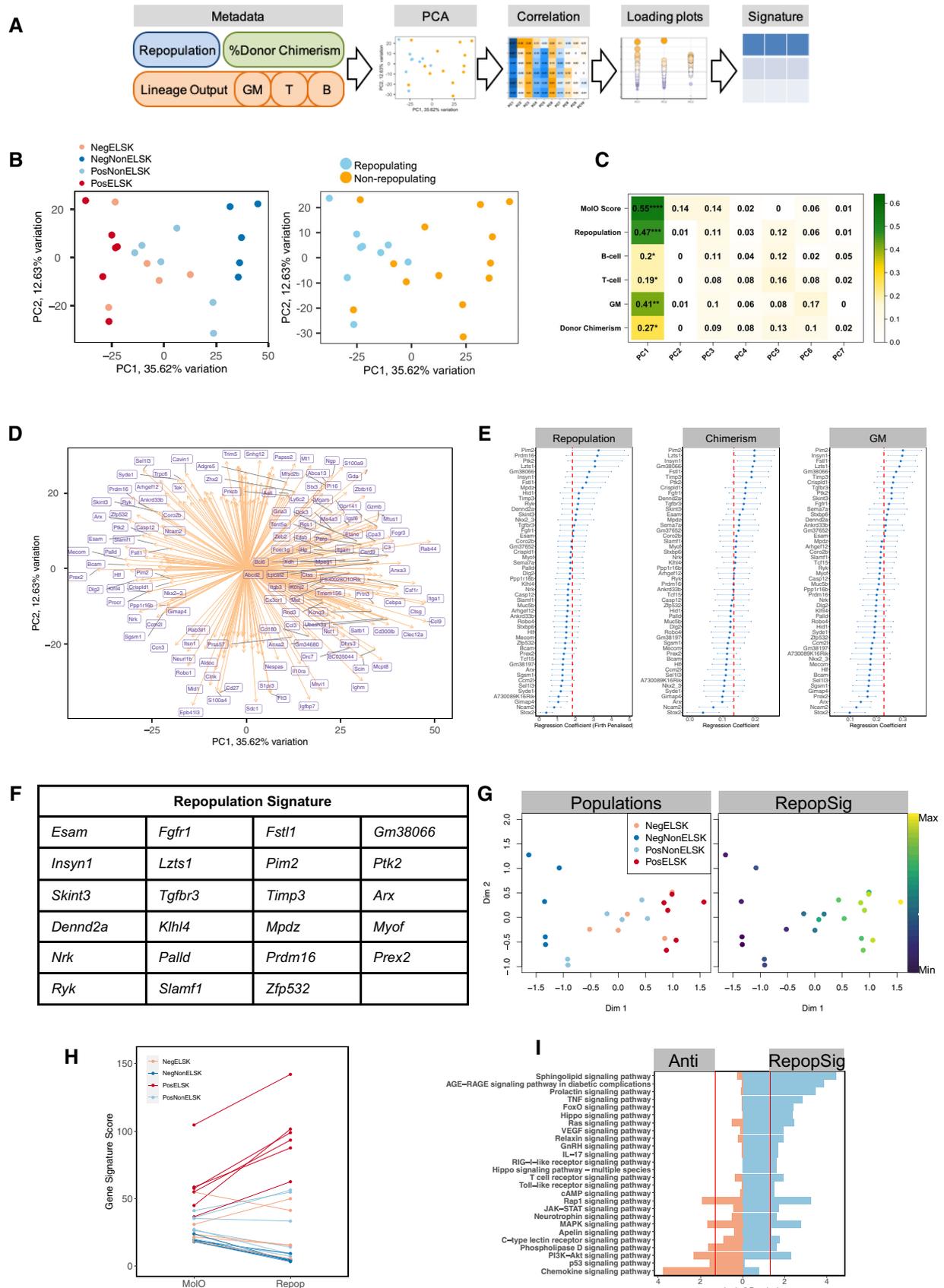


Figure 4.

**Figure 4. A molecular signature (RepopSig) of expanded HSCs.**

- A Schematic of how the repopulation signature (RepopSig) was derived using PCA analysis and transplantation-associated metadata.
- B PCA plot of samples colored by their categories and by functional outcome.
- C Correlation between the first 7 principal components and the metadata, showing  $r^2$  values and significance. Pearson correlation, \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ , \*\*\*\* $P < 0.0001$ .
- D PCA loading plot for PC1 and PC2.
- E Regression coefficients of top 50 repopulation-associated genes. Logistic regression coefficients (Firth penalized) depicted for repopulation outcome and linear regression of chimerism and GM contribution. Cut-off for signature inclusion is indicated by the red dotted line.  $N = 22$  individual clones, error bars represent SD.
- F List of RepopSig genes.
- G MDS plots depicting sample categories and RepopSig signature scores.
- H MoLO and RepopSig scores of each sample colored by their categories.
- I KEGG pathway analysis of genes correlated with the RepopSig ( $r > 0.7$ , Signature) and genes least correlated with the RepopSig ( $r < -0.7$ ).

with fewer phenotypic HSCs (< 1% ELSK; Fig 5C and D). To further demonstrate the applicability of RepopSig genes in the identification of repopulating clones, we probed 28-day single cell-derived clones for cell surface expression of ESAM (one of the RepopSig genes) by flow cytometric profiling. Interestingly, we observed a strong correlation ( $r^2 = 0.951$ ) between the ESAM-ELSK (EELSK) and *Fgd5*-ELSK phenotypes, indicating that *Fgd5* can be replaced by ESAM, thus removing the need for using the reporter mouse in screening experiments (Fig 5E). The utility of RepopSig is further indicated by the 10 $\times$  single cell RNA-sequencing data where the HSC compartment is more specifically identified by the RepopSig compared to the MoLO gene signature, with a multitude of RepopSig genes exclusively expressed in the HSC cluster (Fig 5F, Appendix Fig S2A and B). Overall, these data underscore the robustness of both the ELSK phenotype and the RepopSig score for identifying cultures with high numbers of functional HSCs.

Following its validation in transplantation assays, we next assessed the general applicability of the RepopSig for identifying HSCs from a variety of cellular states using published single-cell RNA-sequencing (scRNA-seq) datasets (Nestorowa *et al*, 2016; Oedekoven *et al*, 2021). We first assessed data from ~1,600 freshly isolated stem and progenitor cells from Nestorowa *et al*, where the RepopSig was able to distinguish LT-HSCs from HSPCs and progenitors (Figs 6A and B, and EV5A and B). Although the MoLO score outperforms the RepopSig score in this dataset (largely due to cell cycle genes in the MoLO signature), we hypothesized that the RepopSig might perform better for uniformly marking HSCs in culture and in cell cycle. To test this, we generated new scRNA-seq libraries for cycling FL HSCs and 7-day *in vitro* hibernating HSCs (Oedekoven *et al*, 2021). Whereas the MoLO score consistently ranked hibernating HSCs higher than FL HSCs (Fig 6C and D), the RepopSig scored both FL HSCs and hibernating HSCs similarly despite their distinct cell cycle status (Fig 6C and D). In addition, higher RepopSig scores were also observed in freshly isolated and hibernating HSCs compared to cytokine-stimulated cells with reduced HSC frequency (Figs 6E and F, and EV5C). In addition, we observed RepopSig enrichment in freshly isolated HSCs against a broad range of hematopoietic cell types (Appendix Fig S3A and B). Overall, this suggests that the RepopSig can mark HSCs from multiple distinct cellular states ranging from active versus quiescent, freshly isolated versus cultured, and adult versus fetal origin.

Finally, to test the applicability of the HSC RepopSig across species and to explore its potential translatability to humans, human hematopoietic cell datasets were obtained and assessed for the

expression of RepopSig homologs. Out of 23 repopulation signature genes, only *Skint3* and *Gm38066* lacked human homologs and the expression of the remaining 21 genes was highly enriched within human HSCs compared to all other examined hematopoietic cell subsets (Fig 6G). This further affirms the strength of RepopSig for identifying HSCs from a wide range of cellular states and sets the stage for translating these findings to the human system, including an exploration of individual genes for their impact on HSC function.

## Discussion

*Ex vivo* HSC expansion has been a long-standing goal in the field, with substantial clinical implications for improving stem cell transplantation, production of limitless populations of mature blood cells, and the base cellular product for gene therapy. While the recent report of 200–899 fold mouse HSC expansion *ex vivo* represents a major breakthrough (Wilkinson *et al*, 2019), the substantial heterogeneity in single cell-derived clones has thus far precluded the molecular characterization of expanded HSCs or high throughput screening for cultures containing large numbers of functional HSCs. Here, we report an *in vitro* reporter strategy that overcomes these issues by using *Fgd5* and EPCR as markers of functional HSCs in culture and by prospectively separating HSCs from non-HSCs, thus allowing molecular profiling. By integrating single-clone functional transplantation data with gene expression profiling from the same clones, we report that (i) EPCR and *Fgd5* are reliable *in vitro* markers for enriching functional HSCs; (ii) ESAM can substitute for the *Fgd5* reporter as a reliable *in vitro* marker; (iii) expanded HSCs share a core molecular program with freshly isolated HSCs; (iv) megakaryocytic and erythroid genes are over-represented in non-HSCs in nonrepopulating clones, which may provide a source of negative feedback signals; (v) the molecular profile of expanded HSCs can be defined with a new repopulation signature, which can also identify HSCs in multiple cellular states. This reporter system represents a highly efficient way of identifying functional HSCs *in vitro* and avoids the costly and time-consuming *in vivo* transplantation step, thereby setting the stage for large-scale screening that has been previously impossible to undertake.

One of the main barriers that has hindered the study of HSC expansion has been the lack of robust markers to isolate stem cells *in vitro* (Zhang & Lodish, 2005). Fares *et al*, 2017 first indicated that EPCR expression tracked with human HSC content *in vitro*, and

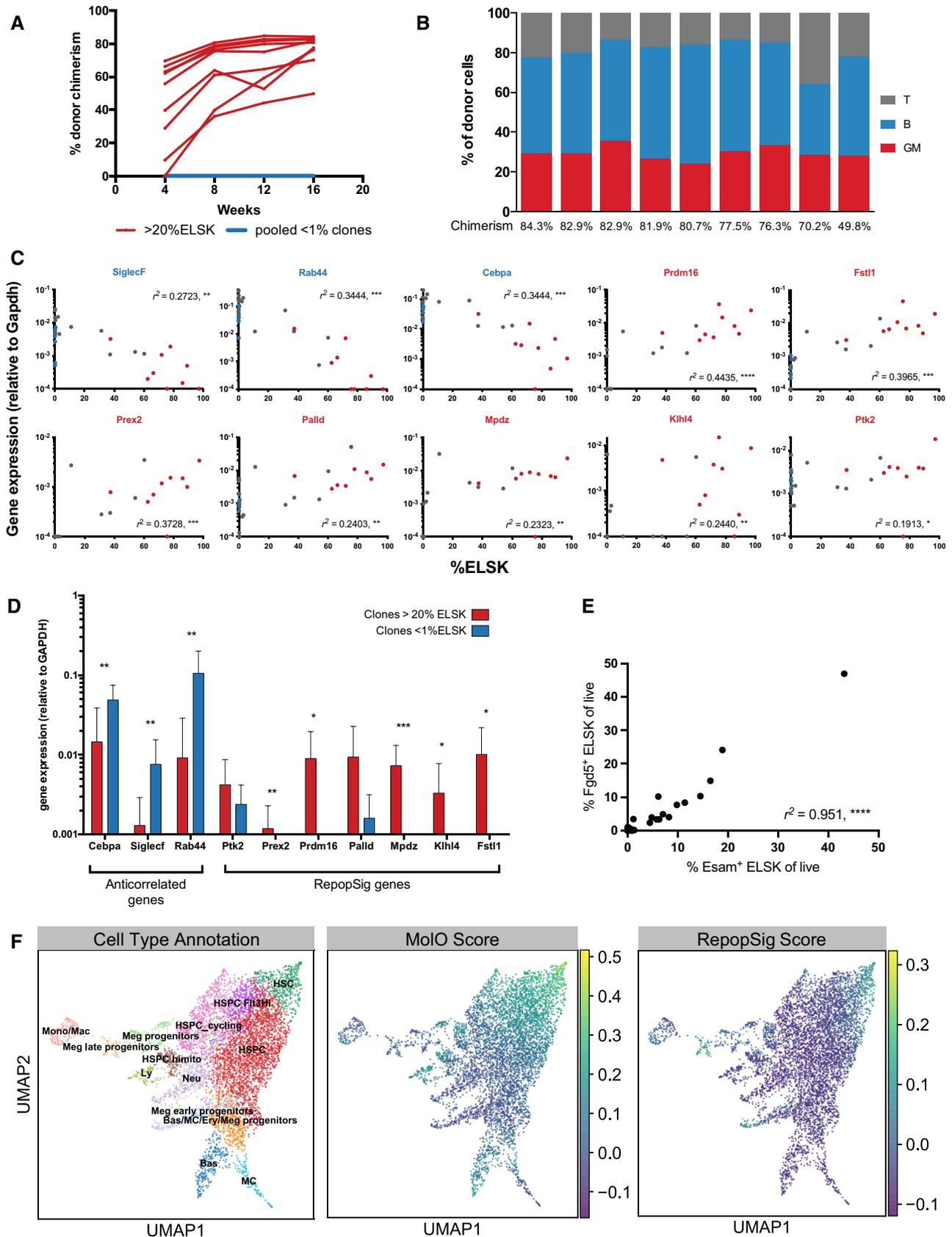


Figure 5.

**Figure 5. Reporter strategy and RepopSig gene signature reliably mark clones with functional HSCs.**

- A A selection of clones was transplanted into irradiated recipients using 50% of the cells harvested at day 28 and donor chimerism of mice from clones with > 20% ELSK (red,  $n = 9$ ) and < 1% ELSK (blue, pooled from 10 clones) are displayed.
- B Corresponding lineage output of clones as a percentage of donor cells at week 12 post-transplantation with donor chimerism displayed under each bar.
- C Correlation of the relative gene expression against the ELSK percentage of the clones. Red and blue dots indicate clones that were transplanted in Fig 5C. Pearson correlation, \*\*\*\* $P < 0.0001$ , \*\*\* $P < 0.001$ , \*\* $P < 0.01$ , \* $P < 0.05$ .
- D Single ESLAM HSCs were cultured for 28 days and 10% of the cultures were analyzed by flow cytometry on day 27. Clones with above 20% ( $n = 13$ ) and below 1% ELSK cells ( $n = 15$ ) were analyzed for their relative gene expression of RepopSig genes (both positive and negative markers) \*\*\* $P < 0.001$ , \*\* $P < 0.01$ , \* $P < 0.05$ . Error bars represent SD.
- E The correlation between the proportion of Esam + ELSK (%EELSK) and FELSK cells in single-cell clones cultured for 28 days, Pearson correlation, \*\*\*\* $P < 0.0001$ .
- F Projections of geometric mean scores for MoIO (middle) and RepopSig (right) gene signatures onto the previously derived (Fig 3D) single cell landscape of 28 day-expanded bulk HSC cultures (left).

our study combines EPCR expression with *Fgd5* and LSK markers to deliver a robust tool for marking a highly HSC-enriched fraction in long-term expansion cultures. However, our strategy still does not isolate functionally pure HSCs and recently developed reporter mouse strains may be usefully exploited to further purify HSCs from multiple distinct cellular states (Gazit *et al*, 2014; Busch *et al*, 2015; Chen *et al*, 2016; Cabezas-Wallscheid *et al*, 2017; Tajima *et al*, 2017; Pinho *et al*, 2018). The ability to expand, and subsequently highly enrich, HSCs also enables a wide range of previously impossible experiments requiring large numbers of input cells, including global proteomics, metabolomics, and ChIP-sequencing. Moreover, the faithful tracking of the FELSK and EELSK phenotypes with functional HSC content now permits large-scale functional screens (small molecules, CRISPR, etc.) and directed differentiation experiments on a previously unimaginable scale. Importantly, while a > 20% ELSK cutoff provided a reliable metric for identifying single cell-derived 28-day expanded clones with robust expansion of functional HSCs, it would require validation and potential adjustment if culture length and/or input cells were altered.

Whereas the molecular profile of freshly isolated HSCs from different isolation strategies and developmental timepoints have been firmly established (Kent *et al*, 2009; Wilson *et al*, 2015; Nestorowa *et al*, 2016; Dong *et al*, 2020), our data represent the first description of the molecular profile of *in vitro* expanded HSCs. A complete understanding of the molecular machinery governing HSC self-renewal *ex vivo* will be instrumental in the improvement of gene therapy and directed differentiation protocols. Our data implicate a wide range of signaling pathways and cellular feedback mechanisms that could be key to unlocking this clinical potential.

At present, extensive clonal heterogeneity remains in 28-day cultures and our dataset accords with previous data from Wilkinson *et al*, demonstrating that single HSCs can expand in F12 PVA-based conditions. Surprisingly, considering the length of culture, the molecular profile of expanded HSCs resembles freshly isolated HSCs to a high degree. Interestingly, the total cell number of a clone did not correlate significantly with its repopulation potential, further affirming the established negative relationship between HSC self-renewal and proliferation (Wilson *et al*, 2008). Our data suggest that irrespective of clone size, the percentage of phenotypic HSCs in a 28-day clone is the best predictor of its repopulation potential. The identification of previously established self-renewal regulators such as *Prdm16* (Deneault *et al*, 2009; Chuikov *et al*, 2010; Gudmundsson *et al*, 2020), *Esam* (Ooi *et al*, 2009; Yokota *et al*, 2009) and *Fstl1* (Holmfeldt *et al*, 2016) in the RepopSig is perhaps not

surprising; however, the RepopSig additionally includes a number of genes that have not been implicated in HSC biology previously, such as *Klhl4* and *Mpdz*, offering exciting new targets for potentially regulating HSC expansion. Although filtered out from the final RepopSig by linear regression analysis, *Tcf15*, which was recently shown to be enriched in mouse HSCs, was also identified within the loading plot of PC1 to be associated with repopulation (Rodriguez-Fraticelli *et al*, 2020). Pathway analysis of the RepopSig also accords with recent reports on the role of sphingolipid signaling in the self-renewal of human HSCs *ex-vivo* (Xie *et al*, 2019).

Our data also suggest that clones which no longer contain functional HSCs at 28 days predominantly generate cells with molecular characteristics of the megakaryocyte and erythrocyte lineages. Future expansion strategies might take advantage of this by targeted removal of such cells, and some efforts (such as fed-batch cultures; Csaszar *et al*, 2012) have already demonstrated that dilution of exogenous factors can increase expansion efficiency. Additionally, the RepopSig can be combined with FELSK markers as a quality control tool for rapid monitoring of long-term HSC content via qPCR and flow cytometry, respectively. While previous HSC gene signatures such as MoIO are biased toward freshly isolated, quiescent HSCs, the RepopSig appears to represent a more general functional HSC signature, capable of identifying cycling as well as cultured HSCs. Our strategy also extends on the growing number of studies that demonstrate the power of linking functional and molecular data to better impart biological meaning behind transcriptomic information (Wilson *et al*, 2015; Psaila *et al*, 2016; Shepherd *et al*, 2018; Shepherd & Kent, 2019).

Ultimately, such approaches will be applied to human HSC expansion by adding combinations of extrinsic self-renewal regulators, utilizing fed-batch negative feedback regulation (Csaszar *et al*, 2012), or even engineering artificial 3D niches with ECM proteins and functionalized hydrogels (Bai *et al*, 2019). To exemplify the direct applicability and translatability of mouse HSC expansion profiling to the human system, we also demonstrate that homologs to genes with high RepopSig scores are highly enriched in human HSCs, including a number of genes having previously described associations with human HSC biology, such as *Hlf* and EPCR, which have recently been reported to also mark expanded human HSCs in culture (Fares *et al*, 2017; Garg *et al*, 2019; preprint: Lehnertz *et al*, 2020). This, in combination with the early indication that F12 PVA-based cultures can modestly expand human HSCs, bodes well for moving these findings rapidly into the human system. Similarly, fed-batch systems are already being applied with novel small

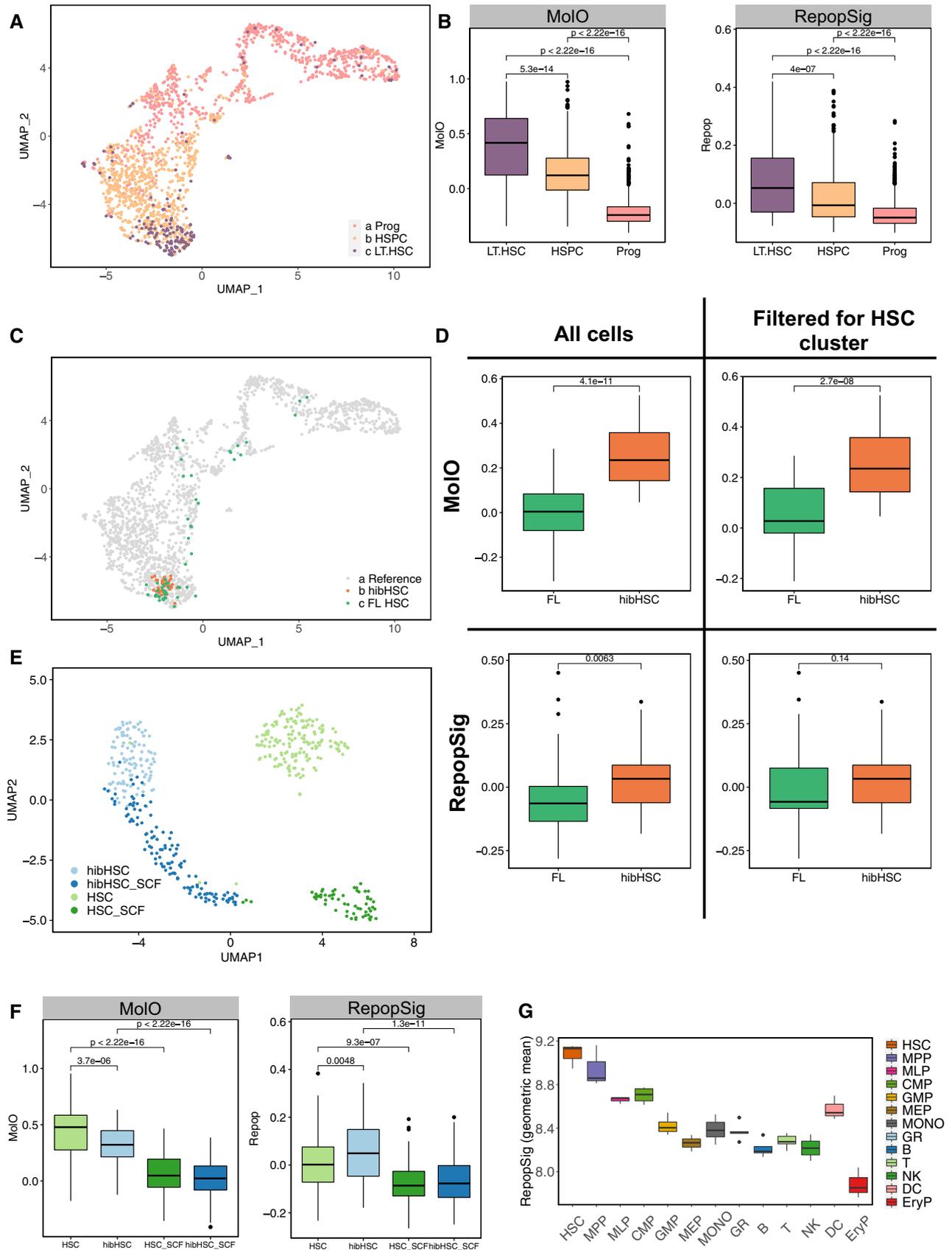


Figure 6.

**Figure 6. RepopSig identifies HSCs from multiple cellular sources and cell cycle states.**

- A UMAP representation of mouse HSC transcriptomes (Nestorowa *et al*, 2016), colored by their cell type.
- B Boxplot of MoIo and RepopSig scores for each cell group with both signatures being able to identify LT-HSCs.
- C Projections of hibernating (hibHSC) and fetal liver (FL) HSC scRNA-seq profiles onto the single-cell landscape showing the majority of cells in both cases localizing to the LT-HSC region.
- D MoIo and RepopSig scores for FL HSCs and hibHSCs where MoIo preferentially associates with quiescent hibHSCs and RepopSig associates with both equally. This is true both when (I) all single cells and (II) excluding cells falling outside the LT-HSC compartment are assessed.
- E UMAP landscape for unstimulated and SCF-stimulated freshly isolated HSCs and hibHSCs (Oedekoven *et al*, 2021).
- F MoIo and RepopSig scores for freshly isolated HSCs and hibHSCs, as outlined in (E) where again RepopSig identifies populations with high proportions of functional HSCs.
- G Geometric mean signature scoring for human homologs of RepopSig genes across human stem, progenitor, and mature cell populations (HSC, hematopoietic stem cells; MPP, multipotent progenitors; MLP, multilymphoid progenitors; CMP, common myeloid progenitor; GMP, granulocyte-monocyte progenitors; MEP, megakaryocyte-erythroid progenitors; MONO, monocytes; GR, granulocyte; B, B cells; T, T cells; NK, natural killer cells; DC, dendritic cells; EryP, erythroid progenitors. For all boxplots, central band represents median, boxes represent first and second quartile, whiskers represent third and fourth quartile, and dots represent outliers. T-test. Individual geometric means were computed using all single cells within the relevant defined cell clusters.

molecules such as UM171 and SR1 (Fares *et al*, 2014) to achieve modest levels of expansion. Combining such promising avenues will undoubtedly lead to success in future clinical scale human HSC expansion.

## Materials and Methods

### Mice

*Fgd5*<sup>ZsGreen-ZsGreen/+</sup> knock in/knock out mice were purchased from Jackson Laboratories and wild-type (WT) mice were either *Fgd5*<sup>+/+</sup> litter mates or CD45.2 C57BL/6. All transplantation recipients were C57BL/6<sup>W41/W41</sup>-Ly5.1 (W41). All mice were kept in microisolator cages in Central Biomedical Service animal facility of University of Cambridge and University of York, and provided continuously with sterile food, water, and bedding. All mice were kept in specified pathogen-free conditions, and all procedures performed according to the United Kingdom Home Office regulations, in accordance with the Animal Scientific Procedure Act.

### Isolation and analysis ESLAM Sca1<sup>+</sup> HSCs

Mice were sacrificed by dislocation of the neck. BM cells were isolated from the tibia, femur and sternum of both hind legs, by crushing bones in PBS (Sigma) supplemented with 2% Fetal Calf Serum (FCS (Sigma) or STEMCELL Technologies (SCT)). Samples were filtered through 20- $\mu$ m sterile filters before further processing. Red cell lysis was performed using ammonium chloride (NH<sub>4</sub>Cl, SCT) and HSPC were enriched by magnet separation using EasySep Mouse Hematopoietic Progenitor Cell Enrichment Kit (SCT). ESLAM Sca-1<sup>+</sup> cells were isolated by fluorescence-activated cell sorting (FACS) using CD45 BV421 (Clone 30-F11, Biolegend), CD150 PE/Cy7 (Clone TC15-12F12.2, Biolegend), CD48 APC (Clone HM48-1, Biolegend), Sca-1 BV605 (Clone D7, Biolegend), EPCR PE (Clone RMEPCR1560, SCT), and 7-Aminoactinomycin D (7AAD; Life Technologies). The cells were sorted on an Influx (BD) using the following filter sets 530/40 (for *Fgd5*), 585/29 (for PE), 670/30 (for APC), 460/50 (for BV421), 670/30 (for 7AAD), and 610/20 (for BV605). When single HSCs were required, the single-cell deposition unit of the sorter was used to place 1 cell per well into 96-well plates, each well having been preloaded with 50  $\mu$ l or 100  $\mu$ l medium. E14.5 FLs were prepared and stained as above and analyzed as above.

### Stemspan (SS) based HSC cultures

Bulk HSCs were cultured in 96 well U-bottom plates (Corning) containing 100  $\mu$ l of StemSpan Serum-Free Expansion Medium (SS, SCT) supplemented with 1% Penicillin/Streptomycin (Sigma), 1% L-Glutamine (Sigma), 0.2% Beta-Mercaptoethanol (Life Technologies), 300 ng/ml of mouse SCF (SCT or Bio-Techne), and 20 ng/ml human IL-11 (SCT or Bio-Techne) at 37°C with 5% CO<sub>2</sub>. All SS-based cultures are performed serum-free.

### F12-based 28-day HSC cultures

F12-based cultures performed as described previously (Wilkinson *et al*, 2020). Briefly, single or bulk HSCs were cultured on BioCoat fibronectin 96 well plates (Corning) in 200  $\mu$ l of Ham's F12 nutrient mix (Thermo) supplemented with 1% Insulin-Transferrin-Selenium-Ethanolamine (ITSX, Gibco), 10 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES, Gibco), 1% Penicillin/Streptomycin/L-Glutamate (P/S/G, Gibco), 100 ng/ml mouse TPO (Preprotech), 10 ng/ml mouse SCF (Preprotech), and 0.1% PVA (Sigma) or HSA (Albumin Bioscience) at 37°C with 5% CO<sub>2</sub>. Where indicated, 20 ng/ml of human IL-11 (SCT or Bio-Techne) were also used. Complete medium changes were made every 2–3 days after the first 5–6 days. Where indicated, 10% of the cultures were taken out for flow cytometric analysis detailed below. For RhoGTPase inhibitor and activator cultures, indicated concentrations of CASIN (Tocris), NSC23766 (Tocris), Rhosin (Tocris), and ML-099 (Merck) were used for the entirety of the culture, with medium changes performed as normal.

### F12-based short-term (< 10 days) cultures

For short-term cultures up to 10-days, cells were cultured as above, except 96 well U-bottom plates (Corning) were used and no media changes were performed.

### Flow cytometric analysis of *in vitro* cultures

At the indicated experimental endpoint, cultured cells (cultured from bulk or single clones) were stained with EPCR PE (Clone RMEPCR1560, SCT), Sca-1 BV605 (Clone D7, Biolegend), CD11b APC (Clone M1/70, Biolegend), Gr-1 PE/Cy7 (Clone RB6-8C5, Biolegend), c-Kit APC/Cy7 (Clone 2B8, Biolegend), CD45 BV421 (Biolegend), and 7AAD (Life Technologies). To enumerate cells, a defined

number of fluorescent beads (Trucount Control Beads, BD) were added to each well and each sample was back calculated to the proportion of the total that were run through the cytometer. Flow cytometry was performed on an LSRFortessa (BD) with a High Throughput Sampler (BD; for single clone analysis).

### Bone marrow transplantation assay

Recipient mice were W41 mice as described previously. Recipient mice were sub-lethally irradiated with a single dose (400 cGy) of Cesium irradiation. All transplantations were performed by intravenous tail vein injection of cell fractions suspended in 200–300  $\mu$ l PBS using a 29.5G insulin syringe. Repopulation was defined as having > 1% donor chimerism and > 1% contribution to GM at 16 weeks.

### Peripheral blood analysis

Peripheral blood samples were collected from the tail vein at indicated timepoints using EDTA coated microvette tubes (Sarstedt AGF & Co, Nuembrecht, Germany). Red cell lysis was performed by using  $\text{NH}_4\text{Cl}$  (SCT) and samples were subsequently analyzed for repopulation levels as previously described (Wilson *et al*, 2015; Kent *et al*, 2016). Cells were stained for lineage markers using Ly6g BV421 (Clone 1A8), B220 APC (Clone RA3-6B2), CD3e PE (Clone 17A2), CD11b PE-Cy7 or BV605 (Clone M1/70), CD45.1 AF700 (Clone A20), and CD45.2 APC-Cy7 (Clone 104). All antibodies were obtained from Biolegend. Samples were acquired on LSR Fortessa (BD) and flow cytometry data analyzing by using FlowJo (Treestar, Ashland, OR, USA).

### Bulk RNA sequencing

RNA was extracted using the Picopure RNA Isolation Kit (Thermo) according to the manufacturer's protocol. Libraries were prepared using the SMARTer Stranded Total RNA-seq Kit v2—Pico Input mammalian (Takara Bio, CA, USA) according to manufacturer's protocol. Quality control (QC) steps were performed using Qubit RNA HS Assay Kit and bioanalyzer. Sequencing was run at the Cancer Research UK Cambridge Institute Genomics core on a Novaseq 6000 (Illumina), using 50 bp paired-end reads. Reads were trimmed using trim\_galore (parameters: --paired --quality 30 --clip\_R2 3) and aligned to the *Mus musculus* genome build (mm10) using STAR (default parameters). Gene counts were acquired using HTSeq (parameters: --format = bam --stranded = reverse --type = exon --mode = intersection-nonempty --additional-attr = gene\_name). Raw data and processed gene count tables are available via GEO accession number: GSE175400. Raw counts were processed using edgeR (version 3.28.1; Robinson *et al*, 2010; McCarthy *et al*, 2012). First, lowly expressed genes were excluded from downstream analysis. Here, genes with fewer than two libraries expressing a minimum of 1 CPM (counts per million) were considered lowly expressed. Subsequently, read counts were normalized using the trimmed mean of M values (TMM) method (Robinson & Oshlack, 2010). Where there are multiple sequencing runs across an experiment, technical replicates were used to inform batch correction, performed with Limma (version 3.42.2; Ritchie *et al*, 2015). With little variation between Batch1 and Batch2, batch correction was performed on Batch1 and Batch3, where a significant variation of technical

replicates was identified. Log-transformed and batch corrected values were subsequently used for downstream analysis.

### Single-Cell RNA sequencing – Smart-Seq2 Data

scRNA-seq analysis was performed according to the previously described Smart-seq2 protocol (Picelli *et al*, 2013). Freshly isolated fetal liver (FL) HSCs and 7-day cultured hibernating HSCs (hibHSC) were deposited into 96-well plates, containing lysis buffer [0.2% Triton X-100 (Sigma), RNase inhibitor (SUPERase, Thermo Fisher), nuclease-free water (ThermoFisher)]. The Illumina Nextera XT DNA preparation kit was used for library construction. The pooled library (single end, 50 bp reads) was sequenced on the Illumina HiSeq 4000 at the Cancer Research UK Cambridge Institute Genomics core. Raw data and processed gene count tables are available via GEO accession number: GSE175400. Raw reads were aligned to the *Mus musculus* genome build (mm10) using STAR and read counts were computed using featureCounts. Cells not passing quality control thresholds below were excluded. First, a threshold of mapped reads was set to > 1e5 and < 3e7, with mapped reads comprising nuclear genes, mitochondrial genes and ERCCs. A minimum threshold of 20% for reads mapping to known genes was set, in order to exclude empty wells and dead cells. In addition, the threshold for reads mapping to mitochondrial genes was > 0.075, to ensure a minimum of 7.5% of reads to map to nonmitochondrial genes. Finally, an ERCC cutoff of 5% and a high gene cutoff of 1,800 were selected.

Besides newly-generated scRNA-seq data for fetal liver and hibernating HSCs, the following previously published datasets were used: (i) Hematopoietic stem and progenitor compartment and (ii) freshly isolated, hibernating, and stimulated HSCs. All datasets were processed using the Seurat R package (version 4.0.0). Data were normalized using a scaling factor of 10,000 and 7,500 variable features were computed. Data were scaled using default parameters. Gene signature scoring and visualizations were performed using Seurat (version 4.0.0), ggplot2 (version 3.3.3) and native R functions). FL HSC and hibHSC single cells were projected onto the single-cell hematopoietic stem and progenitor landscape using default settings for finding anchors between the reference landscape and query datasets (version 4.0.0). RNA-seq profiles of human stem, progenitor, and mature cell types were retrieved from Xie *et al* (2019, 2021). Gene expression profiles were normalized using variance-stabilizing transformation (performed using DESeq2; Love *et al*, 2014), prior to Repopulation Signature geometric mean scoring. Human homologs for 19 out of 23 Repopulation Signature genes were identified and utilized for gene scoring. The four missing genes included *Gm38066*, *INSYNI*, *SKINT3*, and *ZFP532*.

### Single-cell RNA sequencing—10 $\times$ genomics data

A 30-day expansion culture (as previously described Wilkinson *et al*, 2020) was collected for single-cell RNA-seq analysis from 12-week-old male C57BL/6 mice. HSCs were isolated by FACS as previously described using the following strategy: CD150<sup>+</sup> CD34<sup>-/lo</sup> c-Kit<sup>+</sup> Sca1<sup>+</sup> Lineage<sup>-</sup> (Wilkinson *et al*, 2020). Droplet scRNA-Seq experiment was performed using the 10 $\times$  Genomics Single Cell 3' v3 kit according to manufacturer's protocol. The UMI counts per cell were calculated using the cellranger package (v3.1.0) and downstream analysis was performed using the scanpy package (Wolf *et al*,

2018). Low quality cells were excluded with following criteria per cell: < 1,500 genes and < 5,000 UMI counts. Putative doublet cells were removed using the scrublet tool (Wolock *et al*, 2019) (threshold of 0.3). The scRNA-Seq landscape was generated by: cell count log-normalization, selection of 5,000 top variable genes, computation of 40 principal components, nearest neighbor search ( $k = 10$ ), leiden clustering (resolution = 1), and finding UMAP embedding. To minimize cell cycle effects a set of genes correlated with cell cycle signature were removed from the highly variable gene set (as used previously (Dahlin *et al*, 2018)) and residual cell-cycle effects were regressed out by using the cell cycle phase as covariate (regress\_out function). Cell cycle phases were computed using the score\_genes\_cell\_cycle function. Xist and Y-chromosome genes were also excluded. Clusters were manually annotated on markers identified in the literature. Cluster connectivity and putative trajectories were computed using the PAGA method (Wolf *et al*, 2019).

### Statistical analysis

Differential expression was performed using a likelihood ratio test approach. For this purpose, a negative binomial generalized linear model (GLM) was fitted. Multidimensional scaling (MDS) plots were computed using Limma (version 3.42.2). Genes were considered differentially expressed when a  $\text{LogFC} \geq 2$  and  $\text{FDR} < 0.05$ .

To compute gene ontology (GO) enrichment and KEGG pathway enrichment, gene symbols were converted to Entrez gene identifiers, using the mouse genome annotation database (org.Mm.eg.db, version 3.10.0). GO terms were extracted from the GO annotation database (GO.db, version 3.10.0) and GO term enrichment was computed using the Limma package (version 3.42.2). Biological process GO terms with a  $P$ -value < 0.05 were considered enriched. KEGG pathways were extracted from the KEGG annotation database (version 3.2.3) and were also computed using the Limma package (version 3.42.2).

Principal component analysis (PCA) was performed using the PCATools R package (version 1.2.0). To ensure a Gaussian distribution of gene expression values for PCA computations, lowly expressed genes were removed based on a cumulative cut-off > 40CPM across all samples per gene. During PCA computation, 10% of the most nonvariable genes were excluded from analysis. To identify key genes driving separation of principal components, loadings plots were computed using the top 15% variable genes. Subsequently, a 0.05 cut-off irrespective of directionality was applied to select genes. Pearson correlation coefficients and the respective  $r^2$  values were computed to determine the correlation of transplantation metadata with principal components.

A molecular overlap (MoIo) gene signature associated with freshly isolated LT-HSCs was previously described. MoIo signature genes which passed the threshold for expressed genes (minimum 1 CPM in at least 2 libraries) were extracted from the dataset. The geometric mean was computed on log-transformed expression values for all MoIo genes to derive the MoIo score for each sample. A geometric mean was also computed for a novel repopulation gene signature, derived from the loading plots of the PCA.

To identify dominant cell types of each sample library, the scRNA-seq-based cell type recognition tool SingleR (version 1.0.6) was repurposed and applied to the bulk RNA-seq dataset at hand (Aran *et al*, 2019). Default parameters were used to compute the correlation of each sample against the curated ImmGen reference

dataset (Jojic *et al*, 2013; Shay & Kang, 2013; Aguilar *et al*, 2020). In particular, subtypes within the broad hematopoietic stem cell compartment were used as reference.

Pathway analysis was performed based on the curated Reactome pathway database, using the ReactomePA tool (version 1.30.0; Yu & He, 2016). Entrez gene identifiers for genes of interest were used as input. A  $P$ -value cut-off < 0.05 was applied. Gene Set Enrichment Analysis was performed using the GSEA software (US San Diego and Broad Institute; Subramanian *et al*, 2005; Mootha *et al*, 2003). Gene sets for hematopoietic cell types were retrieved from Chambers *et al* (2007).

To determine the cell type composition of single HSC-derived clones and deconvolute bulk transcriptomes, the direction of state transition (DoT) score was computed (Kucinski *et al*, 2020). Differentially expressed genes between (i) PosELSK vs NegELSK, (ii) PosNonELSK vs PosELSK, and (iii) NegNonELSK vs PosELSK were used for computing DoT scores. The previously described scRNA-seq data of mouse LK and LSK cells (Dahlin *et al*, 2018) was used as a reference landscape. The DoT score was computed as described previously. The point of origin was set to a naive stem and progenitor compartment (Figs 3G and Appendix Fig S3I).

Logistic and linear regression models were fitted to curate the repopulation gene signature for a binary repopulation outcome, donor chimerism, and GM contribution. Logistic regression models were fitted using logistf (version 1.24) using Firth's penalized maximum likelihood and  $\alpha = 0.05$ . Linear models were fitted using native R functions. Coefficients and standard errors for each model were extracted. A signature inclusion cutoff was set to the lower bound of the standard error of the gene with the highest coefficient for each transplantation parameter.

### qPCR validation

RNA was extracted as above, cDNA was synthesized using the SuperScript™ III First-Strand Synthesis System (Invitrogen). TaqMan™ Fast Advanced Master Mix (Applied Biosystems) was used with the following Taqman probes (Thermo): *Prdm16* (Mm00712556\_m1), *Fstl1* (Mm00433371\_m1), *Prex2* (Mm02747802\_s1), *Mpdz* (Mm00447870\_m1), *Cebpa* (Mm00514283\_s1), *Rab44* (Mm01306199\_m1), *Siglecf* (Mm00523987\_m1), *Klhl4* (Mm00555463\_m1), *Gapdh* (Mm99999915\_g1), *Palld* (Mm01341202\_m1), *Ptk2* (Mm00433209\_m1). Reactions were run on the ViiA 7 Real-Time PCR System (Applied Biosystems).

## Data availability

The datasets produced in this study are available in the following databases: RNA-Seq data: Gene Expression Omnibus GSE175400 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE175400>). Flow cytometry data and code used for statistical analyses are available upon request.

**Expanded View** for this article is available online.

### Acknowledgements

The authors thank Reiner Schulte, Chiara Cossetti, and Gabriela Grondys-Kotarba of the CIMR Flow Cytometry core; Peter Ashton, Sally James, John Davey, Katherine Newling, Karen Hogg, Graeme Park, and Peter O'Toole of the

York Biosciences Technology Facility; Maïke Paramor and Vicki Murray of the CSCI Genomics core; Tina Hamilton, Dean Pask, James Baye, Carys Johnson, Nicola Wilson, Fernando Calero-Nieto, James Fox, and Sarah Kinston, for technical assistance; Sally Thomas and the Central Biomedical Services unit staff; Chan Zuckerberg Biohub for next generation sequencing (10×); Haley Daniels and the York BSF for animal care and maintenance; Elisa Laurenti, Ian Hitchcock, Katherine Bridge, Steve Johnson, Thomas Krauss, Michael Milsom, Marc de la Roche, Cédric Ghevaert, Jose Silva and all members of the Kent lab for helpful discussion. DB was supported by a Wellcome PhD Studentship, MSS was the recipient of a Biotechnology and Biological Sciences Research Council (BBSRC) Industrial Collaborative Award in Science and Engineering PhD Studentship, FB was supported by an NC3Rs PhD Studentship (NC/V001922/1) and JLCC was supported by an MRC PhD Studentship under the University of Cambridge Doctoral Training Programme. MJ was supported by an International Postdoc grant from the Swedish Research Council (2021-00185). Research in the BG lab is supported by Wellcome, Blood Cancer UK, and an MRC-AMED joint award (MR/V005502/1). The DGK laboratory is supported by a Blood Cancer UK Bennett Fellowship (15008), an ERC Starting Grant (ERC-2016-STG-715371), an MRC-AMED joint award (MR/V005502/1), and a Cancer Research UK Programme Foundation Award (DCRPGF\100008). This research was funded in whole, or in part, by the Wellcome Trust [203151/Z/16/Z] and the UKRI Medical Research Council [MC\_PC\_17230]. For the purpose of open access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

## Author contributions

**James LC Che:** Conceptualization; resources; data curation; formal analysis; validation; investigation; visualization; methodology; writing – original draft; writing – review and editing. **Daniel Bode:** Conceptualization; resources; data curation; software; formal analysis; writing – original draft; writing – review and editing. **Iwo Kucinski:** Data curation; formal analysis; visualization. **Alyssa H Cull:** Conceptualization; investigation; writing – review and editing. **Fiona Bain:** Formal analysis; investigation; methodology. **Hans J Becker:** Formal analysis; investigation. **Maria Jassinskaja:** Formal analysis; writing – review and editing. **Melania Barile:** Formal analysis. **Grace Boyd:** Investigation; methodology. **Miriam Belmonte:** Investigation; methodology. **Andy GX Zeng:** Formal analysis; visualization. **Kyomi J Igarashi:** Formal analysis. **Juan Rubio-Lara:** Formal analysis; investigation; methodology. **Mairi S Shepherd:** Investigation; methodology. **Anna Clay:** Investigation; methodology. **John E Dick:** Formal analysis; supervision; funding acquisition; writing – review and editing. **Adam C Wilkinson:** Formal analysis; supervision; methodology; writing – review and editing. **Hiromitsu Nakauchi:** Conceptualization; supervision; funding acquisition; writing – review and editing. **Satoshi Yamazaki:** Formal analysis; supervision; funding acquisition; investigation; methodology; writing – review and editing. **Berthold Göttgens:** Resources; software; supervision; funding acquisition; methodology; writing – review and editing. **David G Kent:** Conceptualization; resources; formal analysis; supervision; funding acquisition; investigation; methodology; project administration; writing – review and editing.

## Disclosure and competing interests statement

The authors declare that they have no conflict of interest.

## References

- Aguilar SV, Aguilar O, Allan R, Amir EAD, Angeli V, Artyomov MN, Asinovsky N, Astarita J, Austen KF, Bajpai G et al (2020) ImmGen at 15. *Nat Immunol* 21: 700–703
- Antonchuk J, Sauvageau G, Humphries RK (2002) HOXB4-induced expansion of adult hematopoietic stem cells ex vivo. *Cell* 109: 39–45
- Aran D, Looney AP, Liu L, Wu E, Fong V, Hsu A, Chak S, Naikawadi RP, Wolters PJ, Abate AR et al (2019) Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat Immunol* 20: 163–172
- Bai T, Li J, Sinclair A, Imren S, Merriam F, Sun F, O’Kelly MB, Nourigat C, Jain P, Delrow JJ et al (2019) Expansion of primitive human hematopoietic stem cells by culture in a zwitterionic hydrogel. *Nat Med* 25: 1566–1575
- Batta K, Menegatti S, Garcia-Alegria E, Florkowska M, Lacaud G, Kouskoff V (2016) Concise review: recent advances in the in vitro derivation of blood cell populations. *Stem Cells Transl Med* 5: 1330–1337
- Benz C, Copley MR, Kent DG, Wohrer S, Cortes A, Aghaepour N, Ma E, Mader H, Rowe K, Day C et al (2012) Cell stem cell hematopoietic stem cell subtypes expand differentially during development and display distinct Lymphopoietic programs. *Cell Stem Cell* 10: 273–283
- Bujanover N, Goldstein O, Greenspan Y, Turgeman H, Klainberger A, Scharff Y, Gazit R (2018) Identification of immune-activated hematopoietic stem cells. *Leukemia* 32: 2016–2020
- Busch K, Klapproth K, Barile M, Flossdorf M, Holland-Letz T, Schlenner SM, Reth M, Höfer T, Rodewald HR (2015) Fundamental properties of unperturbed haematopoiesis from stem cells in vivo. *Nature* 518: 542–546
- Cabezas-Wallscheid N, Buettner F, Sommerkamp P, Klimmeck D, Ladel L, Thalheimer FB, Pastor-Flores D, Roma LP, Renders S, Zeisberger P et al (2017) Vitamin A-retinoic acid signaling regulates hematopoietic stem cell dormancy. *Cell* 169: 807–823
- Chambers SM, Boles NC, Lin K-YK, Tierney MP, Bowman TV, Bradfute SB, Chen AJ, Merchant AA, Sirin O, Weksberg DC et al (2007) Hematopoietic fingerprints: an expression database of stem cells and their progeny. *Cell Stem Cell* 1: 578–591
- Chen JY, Miyaniishi M, Wang SK, Yamazaki S, Sinha R, Kao KS, Seitza J, Sahoo D, Nakauchi H, Weissman IL (2016) Hoxb5 marks long-term haematopoietic stem cells and reveals a homogenous perivascular niche. *Nature* 530: 223–227
- Chukov S, Levi BP, Smith ML, Morrison SJ (2010) Prdm16 promotes stem cell maintenance in multiple tissues, partly by regulating oxidative stress. *Nat Cell Biol* 12: 999–1006
- Csaszar E, Kirouac DC, Yu M, Wang W, Qiao W, Cooke MP, Boitano AE, Ito C, Zandstra PW (2012) Rapid expansion of human hematopoietic stem cells by automated control of inhibitory feedback signaling. *Cell Stem Cell* 10: 218–229
- Dahlin JS, Hamey FK, Pijuan-Sala B, Shepherd M, Lau WWY, Nestorowa S, Weinreb C, Wolock S, Hannah R, Diamanti E et al (2018) A single-cell hematopoietic landscape resolves 8 lineage trajectories and defects in kit mutant mice. *Blood* 131: e1–e11
- Deneault E, Cellot S, Faubert A, Laverdure J-P, Fréchette M, Chagraoui J, Mayotte N, Sauvageau M, Ting SB, Sauvageau G (2009) A functional screen to identify novel effectors of hematopoietic stem cell activity. *Cell* 137: 369–379
- Dong F, Hao S, Zhang S, Zhu C, Cheng H, Yang Z, Hamey FK, Wang X, Gao A, Wang F et al (2020) Differentiation of transplanted haematopoietic stem cells tracked by single-cell transcriptomic analysis. *Nat Cell Biol* 22: 630–639
- Dykstra B, Kent D, Bowie M, McCaffrey L, Hamilton M, Lyons K, Lee S-J, Brinkman R, Eaves C (2007) Long-term propagation of distinct hematopoietic differentiation programs in vivo. *Cell Stem Cell* 1: 218–229
- Eaves CJ (2015) Hematopoietic stem cells: Concepts, definitions, and the new reality. *Blood* 125: 2605–2613

- Fares I, Chagraoui J, Gareau Y, Gingras S, Ruel R, Mayotte N, Cszaszar E, Knapp DJHFFH, Miller P, Ngom M *et al* (2014) Pyrimidoindole derivatives are agonists of human hematopoietic stem cell self-renewal. *Science* 345: 1509–1512
- Fares I, Chagraoui J, Lehnertz B, MacRae T, Mayotte N, Tomellini E, Aubert L, Roux PP, Sauvageau G (2017) EPCR expression marks UM171-expanded CD34<sup>+</sup> cord blood stem cells. *Blood* 129: 3344–3351
- Florian MC, Klose M, Sacma M, Jablanovic J, Knudson L, Nattamai KJ, Marka G, Vollmer A, Soller K, Sakk V *et al* (2018) Aging alters the epigenetic asymmetry of HSC division. *PLoS Biol* 16: e2003389
- Gao Y, Dickerson JB, Guo F, Zheng J, Zheng Y (2004) Rational design and characterization of a Rac GTPase-specific small molecule inhibitor. *Proc Natl Acad Sci USA* 101: 7618–7623
- Garg S, Reyes-Palomares A, He L, Bergeron A, Lavallée V-P, Lemieux S, Gendron P, Rohde C, Xia J, Jagdhane P *et al* (2019) Hepatic leukemia factor is a novel leukemic stem cell regulator in DNMT3A, NPM1, and FLT3-ITD triple-mutated AML. *Blood* 134: 263–276
- Gazit R, Mandal PK, Ebina W, Ben-Zvi A, Nombela-Arrieta C, Silberstein LE, Rossi DJ (2014) Fgd5 identifies hematopoietic stem cells in the murine bone marrow. *J Exp Med* 211: 1315–1331
- Gudmundsson KO, Nguyen N, Oakley K, Han Y, Gudmundsdottir B, Liu P, Tessarollo L, Jenkins NA, Copeland NG, Du Y (2020) Prdm16 is a critical regulator of adult long-term hematopoietic stem cell quiescence. *Proc Natl Acad Sci* 117: 31945–31953
- Hinge A, Xu J, Javier J, Mose E, Kumar S, Kapur R, Srour EF, Malik P, Aronow BJ, Filippi M-D (2017) p190-B RhoGAP and intracellular cytokine signals balance hematopoietic stem and progenitor cell self-renewal and differentiation. *Nat Commun* 8: 14382
- Holmfeldt P, Ganuza M, Marathe H, He B, Hall T, Kang G, Moen J, Pardieck J, Saulsberry AC, Cico A *et al* (2016) Functional screen identifies regulators of murine hematopoietic stem cell repopulation. *J Exp Med* 213: 433–449
- Iriuchishima H, Takubo K, Matsuoka S, Onoyama I, Nakayama KI, Nojima Y, Suda T (2011) Ex vivo maintenance of hematopoietic stem cells by quiescence induction through Fbxw7 $\alpha$  overexpression. *Blood* 117: 2373–2377
- Jojic V, Shay T, Sylvia K, Zuk O, Sun X, Kang J, Regev A, Koller D, Best AJ, Knell J *et al* (2013) Identification of transcriptional regulators in the mouse immune system. *Nat Immunol* 14: 633–643
- Kent DG, Copley MR, Benz C, Wöhler S, Dykstra BJ, Ma E, Cheyne J, Zhao Y, Bowie MB, Zhao Y *et al* (2009) Prospective isolation and molecular characterization of hematopoietic stem cells with durable self-renewal potential. *Blood* 113: 6342–6350
- Kent DG, Dykstra BJ, Cheyne J, Ma E, Eaves CJ (2008) Steel factor coordinately regulates the molecular signature and biologic function of hematopoietic stem cells. *Blood* 112: 560–567
- Kent DG, Dykstra BJ, Eaves CJ (2016) Isolation and assessment of single long-term reconstituting hematopoietic stem cells from adult mouse bone marrow. In *Current protocols in stem cell biology*, M Bhatia, AG Elefanti, SJ Fisher, R Patient, T Schlaeger, EY Snyder (eds), pp 2A.4.1–2A.4.24. Hoboken, NJ: John Wiley & Sons, Inc
- Kucinski I, Wilson NK, Hannah R, Kinston SJ, Cauchy P, Lenaerts A, Grosschedl R, Göttgens B (2020) Interactions between lineage-associated transcription factors govern haematopoietic progenitor states. *EMBO J* 39: e104983
- Lang P, Handgretinger R (2008) Haploidentical SCT in children: an update and future perspectives. *Bone Marrow Transplant* 42: S54–S59
- Lehnertz B, MacRae T, Chagraoui J, Tomellini E, Corneau S, Mayotte N, Boivin I, Sauvageau G (2020) HLF expression defines the human Haematopoietic stem cell state. *bioRxiv* <https://doi.org/10.1101/2020.06.29.177709> [PREPRINT]
- Liu W, Du W, Shang X, Wang L, Evelyn C, Florian MC, Ryan MA, Rayes A, Zhao X, Setchell K *et al* (2019) Rational identification of a Cdc42 inhibitor presents a new regimen for long-term hematopoietic stem cell mobilization. *Leukemia* 33: 749–761
- Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15: 550
- McCarthy DJ, Chen Y, Smyth GK (2012) Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res* 40: 4288–4297
- Miharada K, Sigurdsson V, Karlsson S (2014) Dppa5 improves hematopoietic stem cell activity by reducing endoplasmic reticulum stress. *Cell Rep* 7: 1381–1392
- Mootha VK, Lindgren CM, Eriksson K-F, Subramanian A, Sihag S, Lehar J, Puigserver P, Carlsson E, Ridderstråle M, Laurila E *et al* (2003) PGC-1 $\alpha$ -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* 34: 267–273
- Naldini L (2015) Gene therapy returns to Centre stage. *Nature* 526: 351–360
- Nestorowa S, Hamey FK, Pijuan Sala B, Diamanti E, Shepherd M, Laurenti E, Wilson NK, Kent DG, Göttgens B (2016) A single-cell resolution map of mouse hematopoietic stem and progenitor cell differentiation. *Blood* 128: e20–e31
- Oedekoven CA, Belmonte M, Bode D, Hamey FK, Shepherd MS, Che JLC, Boyd G, McDonald C, Belluschi S, Diamanti E *et al* (2021) Hematopoietic stem cells retain functional potential and molecular identity in hibernation cultures. *Stem Cell Reports* 16: 1614–1628
- Oguro H, Ding L, Morrison SJ (2013) SLAM family markers resolve functionally distinct subpopulations of hematopoietic stem cells and multipotent progenitors. *Cell Stem Cell* 13: 102–116
- Ooi AGL, Karsunky H, Majeti R, Butz S, Vestweber D, Ishida T, Quattermost T, Weissman IL, Forsberg EC (2009) The adhesion molecule esam1 is a novel hematopoietic stem cell marker. *Stem Cells* 27: 653–661
- Picelli S, Björklund ÅK, Faridani OR, Sagasser S, Winberg G, Sandberg R (2013) Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat Methods* 10: 1096–1100
- Pinho S, Marchand T, Yang E, Wei Q, Nerlov C, Frenette PS (2018) Lineage-biased hematopoietic stem cells are regulated by distinct niches. *Dev Cell* 44: 634–641
- Psaila B, Barkas N, Iskander D, Roy A, Anderson S, Ashley N, Caputo VS, Lichtenberg J, Loaiza S, Bodine DM *et al* (2016) Single-cell profiling of human megakaryocyte-erythroid progenitors identifies distinct megakaryocyte and erythroid differentiation pathways. *Genome Biol* 17: 83
- Rabe JL, Hernandez G, Chavez JS, Mills TS, Nerlov C, Pietras EM (2019) CD34 and EPCR coordinately enrich functional murine hematopoietic stem cells under normal and inflammatory conditions. *Exp Hematol* 81: 1–15
- Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 43: e47
- Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26: 139–140
- Robinson MD, Oshlack A (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* 11: R25
- Rodríguez-Fraticelli AE, Weinreb C, Wang S-W, Migueles RP, Jankovic M, Usart M, Klein AM, Lowell S, Camargo FD (2020) Single-cell lineage tracing unveils a role for TCF15 in haematopoiesis. *Nature* 583: 585–589

- Shang X, Marchioni F, Evelyn CR, Sipes N, Zhou X, Seibel W, Wortman M, Zheng Y (2013) Small-molecule inhibitors targeting G-protein-coupled rho guanine nucleotide exchange factors. *Proc Natl Acad Sci USA* 110: 3155–3160
- Shay T, Kang J (2013) Immunological genome project and systems immunology. *Trends Immunol* 34: 602–609
- Shepherd MS, Kent DG (2019) Emerging single-cell tools are primed to reveal functional and molecular heterogeneity in malignant hematopoietic stem cells. *Curr Opin Hematol* 26: 214–221
- Shepherd MS, Li J, Wilson NK, Oedekoven CA, Li J, Belmonte M, Fink J, Prick JCM, Pask DC, Hamilton TL et al (2018) Single-cell approaches identify the molecular network driving malignant hematopoietic stem cell self-renewal. *Blood* 132: 791–803
- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES et al (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 102: 15545–15550
- Surviladze Z, Ursu O, Miscioscia F, Curpan R, Halip L, Bologna C, Oprea T, Waller A, Strouse J, Salas V et al (2010) *Three small molecule pan activator families of Ras-related GTPases national*. Bethesda, MD: Center for Biotechnology Information (US)
- Tajima Y, Ito K, Umino A, Wilkinson AC, Nakauchi H, Yamazaki S (2017) Continuous cell supply from Krt7-expressing hematopoietic stem cells during native hematopoiesis revealed by targeted in vivo gene transfer method. *Sci Rep* 7: 1–10
- Wen R, Dong C, Xu C, Zhao L, Yang Y, Zhang Z, Chen Y, Duan L, Chen H, Yang Z et al (2020) UM171 promotes expansion of autologous peripheral blood hematopoietic stem cells from poorly mobilizing lymphoma patients. *Int Immunopharmacol* 81: 106266
- Wilkinson AC, Ishida R, Kikuchi M, Sudo K, Morita M, Crisostomo RV, Yamamoto R, Loh KM, Nakamura Y, Watanabe M et al (2019) Long-term ex vivo haematopoietic-stem-cell expansion allows nonconditioned transplantation. *Nature* 571: 117–121
- Wilkinson AC, Ishida R, Nakauchi H, Yamazaki S (2020) Long-term ex vivo expansion of mouse hematopoietic stem cells. *Nat Protoc* 15: 628–648
- Wilson A, Laurenti E, Oser G, van der Wath RC, Blanco-Bose W, Jaworski M, Offner S, Dunant CF, Eshkind L, Bockamp E et al (2008) Hematopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair. *Cell* 135: 1118–1129
- Wilson NK, Kent DG, Theis FJ, Göttgens B (2015) Combined single-cell functional and gene expression analysis resolves heterogeneity within stem cell populations. *Cell Stem Cell* 16: 712–724
- Wohrer S, Knapp DJHF, Copley MR, Benz C, Kent DG, Rowe K, Babovic S, Mader H, Oostendorp RAJ, Eaves CJ et al (2014) Distinct stromal cell factor combinations can separately control hematopoietic stem cell survival, proliferation, and self-renewal. *Cell Rep* 7: 1956–1967
- Wolf FA, Angerer P, Theis FJ (2018) SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol* 19: 15
- Wolf FA, Hamey FK, Plass M, Solana J, Dahlin JS, Göttgens B, Rajewsky N, Simon L, Theis FJ (2019) PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol* 20: 59
- Wolock SL, Lopez R, Klein AM (2019) Scrublet: computational identification of cell doublets in single-cell transcriptomic data. *Cell Syst* 8: 281–291.e9
- Xie SZ, Garcia-Prat L, Voisin V, Bader GD, Laurenti E, Dick JE (2019) Sphingolipid modulation activates proteostasis programs to govern human hematopoietic stem cell self-renewal. *Cell Stem Cell* 25: 639–653
- Xie SZ, Kaufmann KB, Wang W, Chan-Seng-Yue M, Gan OI, Laurenti E, Garcia-Prat L, Takayanagi S, Ng SWK, Xu C et al (2021) Sphingosine-1-phosphate receptor 3 potentiates inflammatory programs in normal and leukemia stem cells to promote differentiation. *Blood Cancer Discov* 2: 32–53
- Yokota T, Oritani K, Butz S, Kokame K, Kincade PW, Miyata T, Vestweber D, Kanakura Y (2009) The endothelial antigen ESAM marks primitive hematopoietic progenitors throughout life in mice. *Blood* 113: 2914–2923
- Yu G, He Q-Y (2016) ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. *Mol Biosyst* 12: 477–479
- Zhang CC, Lodish HF (2005) Murine hematopoietic stem cells change their surface phenotype during ex vivo expansion. *Blood* 105: 4314–4320
- Zhang CC, Lodish HF (2008) Cytokines regulating hematopoietic stem cell function. *Curr Opin Hematol* 15: 307–311



**License:** This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.