Iterative Single-Cell Analyses Define the Transcriptome of the First Functional Hematopoietic Stem Cells

Graphical Abstract

Highlights

- Single-cell iterations capture the transcriptome of the first functional HSCs in mouse
- Gata2 connects the “inner” HSC regulatory network with single-cell function
- Specific Gata2, cKit, and CD27 levels define all HSCs in embryonic aortic clusters
- HSCs emerge from endothelium as single cells within aortic clusters of 1–2 cells

Authors

Chris Sebastiaan Vink, Fernando Jose Calero-Nieto, Xiaonan Wang, ..., Wajid Jawaid, Berthold Göttgens, Elaine Dzierzak

Correspondence

elaine.dzierzak@ed.ac.uk

In Brief

Vink and colleagues capture the transcriptome of the first functional HSCs in mouse by single-cell RNA-seq, index-sorting, and in vivo and in vitro hematopoietic analyses. The HSC transcriptome is unique compared to HPCs, and heterogeneous expression of pivotal genes suggests that establishment of functional hematopoietic fate during cell emergence from embryonic aortic endothelium is stochastic.
Iterative Single-Cell Analyses Define the Transcriptome of the First Functional Hematopoietic Stem Cells

Chris Sebastiaan Vink, Fernando Jose Calero-Nieto, Xiaonan Wang, Antonio Maglittio, Samanta Antonella Mariani, Wajid Jawaid, Berthold Göttgens, and Elaine Dzierzak

1Centre for Inflammation Research, Queens Medical Research Institute, University of Edinburgh, Edinburgh, Midlothian, Scotland EH16 4TJ, UK
2Department of Haematology, Wellcome & MRC Cambridge Stem Cell Institute, University of Cambridge, Cambridge, Cambridgeshire, England CB2 0AW, UK
3Lead Contact
*Correspondence: elaine.dzierzak@ed.ac.uk
https://doi.org/10.1016/j.celrep.2020.107627

SUMMARY

Whereas hundreds of cells in the mouse embryonic aorta transdifferentiate to hematopoietic cells, only very few establish hematopoietic stem cell (HSC) identity at a single time point. The Gata2 transcription factor is essential for HSC generation and function. In contrast to surface-marker-based cell isolation, Gata2-based enrichment provides a direct link to the internal HSC regulatory network. Here, we use iterations of index-sorting of Gata2-expressing intra-aortic hematopoietic cluster (IAHC) cells, single-cell transcriptomics, and functional analyses to connect HSC identity to specific gene expression. Gata2-expressing IAHC cells separate into 5 major transcriptomic clusters. Iterative analyses reveal refined CD31, cKit, and CD27 phenotypic parameters that associate specific molecular profiles in one cluster with distinct HSC and multipotent progenitor function. Thus, by iterations of single-cell approaches, we identify the transcriptome of the first functional HSCs as they emerge in the mouse embryo and localize them to aortic clusters containing 1–2 cells.

INTRODUCTION

Hematopoietic stem cells (HSCs) arise through a developmental fate process known as endothelial-to-hematopoietic transition (EHT) (Dzierzak and Bigas, 2018; Dzierzak and Speck, 2008; Jaffredo et al., 2005). EHT is best characterized by the emergence of intra-aortic hematopoietic cluster (IAHC) cells within the aorta-gonad-mesonephros (AGM) region that transdifferentiate from specialized hemogenic endothelial cells (de Bruijn et al., 2002; North et al., 2002; Yokomizo and Dzierzak, 2010; Zovein et al., 2008). Between mouse embryonic day (E)10.5 and E12.5, IAHCs contain HSCs, hematopoietic progenitor cells (HPCs), and other hematopoietic cells (HCs). Time-lapse imaging gave the first visual in vivo proof of EHT in the mouse and zebrafish embryonic aorta (Bertrand et al., 2010; Boisset et al., 2010; Kissa and Herbomel, 2010). This transition involves changes from an endothelial transcriptional program to a program promoting HSC identity, morphology, and function (Swiers et al., 2013). Although the precise program directing the generation of HSCs during EHT is as yet unknown, important molecular and physiological aspects of HSC generation are conserved between zebrafish, chick, Xenopus, mouse, and human (Ciau-Uitz and Patient, 2019; Dzierzak and Bigas, 2018; Ivanovs et al., 2017).

Quantitation of IAHC cells in whole mouse embryos (Yokomizo and Dzierzak, 2010) shows a peak of ~700 phenotypic (cKit-positive) HCs at E10.5. Multicolor lineage tracing shows that clusters are polyclonal, with about 545 Vav+ nascent hematopoietic progenitor/stem cells (HP/SCs) emerging between E11.5 and 14.5 to establish lifelong hematopoiesis (Ganuza et al., 2017b). However, in the E10.5 aorta at the time of embryo isolation, very few of these cells function as bona fide HSCs (Kumaravelu et al., 2002; Medvinsky and Dzierzak, 1996; Müller et al., 1994), as determined by direct in vivo transplantation and long-term hematopoietic reconstitution of adult recipients (a standard test for HSC function and clinical relevance). Recent in vivo marking methods question the precise role of phenotypic HSCs defined by in vivo repopulating activity and, instead, suggest that multipotent progenitors are responsible for adult steady-state hematopoiesis (Busch et al., 2015; Pei et al., 2017; Rodriguez-Fraticelli et al., 2018; Schoedel et al., 2016; Sun et al., 2014). Out of the hundreds of IAHC cells in E10.5 embryos, colony-forming unit-culture (CFU-C) studies show that about half (350) of the cluster cells are HPCs (de Pater et al., 2013). Thus, the very low frequency of HSCs in the IAHCs highlights the complexity in programming HSC identity as defined by repopulating function and raises the questions: what processes influence acquisition of this rare HSC identity rather than the more abundant HPC or...
HC identities, and can we capture the transcriptome of the first HSCs?

Many studies, including our own, have set out to describe the genetic program of HC transdifferentiation from that of embryonic aortic endothelium. Published transcriptome databases are available for phenotypically (surface marker) enriched endothelial cells, hemogenic endothelial cells, transitioning cells, IAHCs, and HP/SCs, at both the population and single-cell levels (Baron et al., 2018; Li et al., 2014, 2017; Lichtinger et al., 2012; McKinney-Freeman et al., 2012; Moignard et al., 2015; Solaimani Kartalaei et al., 2015; Swiers et al., 2013; Zhou et al., 2016). However, no unique transcriptional profile has yet been ascribed to AGM HSCs, since the HSC transcriptome is represented in the single-cell datasets only at a very low frequency compared with the high frequency of HPCs and HCs. Interestingly, all these datasets show the upregulated expression of several hematopoietic (“heptad”) transcription factors (TFs) (Wilson et al., 2010) during EHT, including Gata2, which is pivotal to HSC and HPC generation.

Gata2 is required for generation of IAHCs and functional HSCs (de Pater et al., 2013; Ling et al., 2004; Tsai et al., 1994; Tsai and Orkin, 1997). Haploinsufficiency perturbs EHT and the timing and quantitative generation of HSCs (de Pater et al., 2013; Ling et al., 2004), and its overexpression blocks HSC function (Guu et al., 2013; Tipping et al., 2009). Our recent demonstration of pulsatile Gata2 expression level changes in aortic endothelial cells undergoing EHT (Eich et al., 2018) reveals a previously unexplored dynamic regulatory aspect in hematopoietic fate acquisition. Studies in other systems have shown an unstable program of gene expression in cells as they take on specific lineage fate. These include regulators nuclear factor κB (NF-κB), Hes, and Nanog, which show pulsatile expression behavior in single cells (Abranches et al., 2014; Chang et al., 2008; Cohen-Saidon et al., 2009; Huang, 2009; Kueh et al., 2016; Lahav, 2004; Miyawaki and Torres-Padilla, 2012; Purvis and Lahav, 2013; Ryu et al., 2016).

To date, transcriptomic analyses of HSCs have relied on cell-surface-marker-mediated enrichments. As TFs are responsible for promoting the expression of these markers, hematopoietic identities may be better understood through direct enrichment based on TF expression. Our previously described Gata2Venus (G2V) mouse model reports real-time transcription and protein expression of Gata2 in vivo without affecting hematopoietic development or function (Kaimakis et al., 2016). To understand whether the transcriptome of single IAHC cells (Baron et al., 2018; Zhou et al., 2016) correlates with hematopoietic function, and whether HSC identity is responsible for promoting the expression of these markers, IAHCs, and HP/SCs were used to examine IAHC cells by iterative single-cell transcriptomics and functional analyses. Here, we show that functional heterogeneity in highly enriched single index-sorted IAHC cells expressing medium levels of Gata2 (G2med) corresponds with heterogeneity in TF expression levels. Following this, we reveal that specific expression levels of Gata2, cKit, and CD27 define all in vivo repopulating HSCs and that HSCs localize to IAHCs containing 1–2 cells. This iterative methodology captures the transcriptome of the first functional HSCs and closely related HPCs and suggests a stochastic process of quantitative molecular states leading to the establishment of HSC identity.

RESULTS

Specific Gata2 and cKit Levels Define All HSCs and the Majority of Multipotent HPCs

Previously, we showed that Venus fluorescence faithfully reports Gata2 expression in Gata2iresVenus reporter (G2V) mice (Eich et al., 2018; Kaimakis et al., 2016). In G2V embryos, the spatio-temporal expression of Gata2 and the development of EHT and HSC are normal. At E10.5, all HSCs and 70% of HPCs express G2med. Some HPCs are Gata2 negative (G2−) or high (G2+) (Eich et al., 2018; Kaimakis et al., 2016), and mature mast cells express the highest levels (Kauts et al., 2018; Tsai and Orkin, 1997). To focus on the most potent Gata2-expressing progenitors and HSCs, we followed a multi-pronged approach combining the refinement of sorting gates and single-cell index-sorting with functional in vivo and in vitro assays. G2medCD31+cKit+ IAHC cells (all IAHC cells co-express CD31 and cKit) were separated into low (G2lo) and intermediate (G2int) Gata2 fractions and further partitioned into cKitlo and cKit+ fractions (Figure 1A). In vitro progenitor assays (Figure 1B) of the four fractions revealed that most HPCs were distributed equally between the CD31+cKit+G2lo and -G2int fractions at both E10 and E11. Whereas, at E10, few or no CFU-Cs were found in the cKitlo fractions, these fractions at E11 contained some CFU-Cs. Multipotent CFU-GEMMs (colony forming unit-granulocyte-erythrocyte-macrophage-megakaryocytes) were highly enriched or exclusively found in the CD31+cKit+G2lo and -G2int fractions (the latter containing slightly more GEMMs) at both embryonic stages.

Index-sorted CD31+cKit+G2lo (G2lo) cells (Figures 1B and 1C) and CD31+cKit+G2hi (G2hi) cells (as control; Figure 1D) were examined in single-cell CFU assays to test whether HPC lineage potential was related to Gata2 level. After 9–11 days, both G2lo and G2hi cells produced CFU-GEMMs, -GMs (granulocyte-macrophages), -Gs (granulocytes), and -Ms (macrophages) (Figures 1Ci and 1Di; plating efficiency, 1:5–6 and 1:2–4, respectively). However, the proportion of GEMMs was higher in the G2hi fraction (22% versus 9%). The CFU-C type arising from each single-index-sorted cell was plotted against log-normalized mean fluorescent intensities (MFIs) for G2 (Figures 1Ci and 1Di), cKit (Figures 1Cii and 1Dii), and CD31 (Figures 1Ciii and 1Diii). Cells that produced CFU-Gs were widely varying in levels of G2 expression (MFI = 842–8,315) compared to those that produced CFU-GEMMs and -Ms (MFI = 850–4,845). The average MFI for cKit expression on G2hi cells yielding CFU-GEMMs was significantly higher than for cells yielding CFU-Gs (Figure 1Ci), and the average MFI for cKit expression on G2int cells yielding CFU-GEMMs was significantly higher than for the other three CFU types (Figure 1Dii). Both G2lo and G2hi fractions contained cells yielding CFU-GEMMs with a significantly higher average MFI for cKit expression on G2hi cells yielding CFU-GEMMs than for cells yielding CFU-Gs (Figures 1Ciii and 1Diii). Thus, the most immature HPCs are cKithi and CD31hi.
more (multipotent) HPCs, both CD31+cKithiG2lo and -G2int cells provided HSC engraftment at 4 months post-transplantation (Figure 1E). No HSCs were detected in the CD31+cKitloG2med fraction (Figure 1E) or, as we previously reported, in the G2hi fraction (Eich et al., 2018).

Transcriptome Heterogeneity and Clustering of Single CD31+cKithiG2med Cells

Considering that, at E11, all GEMMs were found in the G2lo, Gint, and G2hi fractions and that HSCs are exclusively in the G2lo and G2int fractions, we index-sorted CD31+cKithiG2med (G2lo+G2int = G2med) IAHC cells and analyzed them by single-cell RNA sequencing (RNA-seq) using SmartSeq2 (Picelli et al., 2014). Out of 1,015 CD31+cKithiG2med cells, 968 passed quality control (95%). Of note, the mapping of reads was performed using a custom reference that included the mouse genome, the sequence of Venus and External RNA Controls Consortium (ERCC) sequences, allowing confirmation of the co-regulation of Gata2 and Venus expression levels (Figure S1A; correlation coefficient, 0.7). We obtained the SPRING visualization (Weinreb et al., 2018) of the 968 cells and defined 11 clusters (L1.1–L1.11) using the Louvain method (Figure 2A; Figure S1Bi). Datasets obtained from several biological samples, or analyzed on different days, integrated well and showed no batch effect (Figure S1Bii). Hence, cell acquisition parameters and results were consistent between experiments.

To investigate cluster composition, cells were colored according to the expression of well-known hematopoietic (Figure 2B) and endothelial (Figure 2C) genes as well as several pivotal hematopoietic TFs (heptad TFs; Figure 2D). In this way, we could merge clusters to define 5 main groups of cells within our highly enriched HSC population (Figure 2A). The endothelial-like cluster (EC; L1.1 and L1.2) contained 335 cells expressing endothelial genes such as Flk1 (VEGFR-2), Cdh5 (VE-cadherin), and AplnR. Runx1, which would be expected to be expressed in all cells (See et al., 2009), was downregulated in the EC when compared...
to the other clusters. The hematopoietic cluster (HC) contained 529 cells (L1.7–L1.11) expressing well-known hematopoietic marker genes such as **CD41**, **CD45**, **PU.1**, and **Gpr56**. HCs could be subdivided into HC1 (405 cells; L1.7 and L1.8) and HC2 (124 cells; L1.9–L1.11) cells, according to the expression of genes such as **VE-cadherin** and **Ly6a**. Finally, two smaller unknown clusters were defined: UC1 and UC2. UC1 contained 85 cells (L1.5–L1.6) expressing some hematopoietic markers such as **CD45** and **PU.1** and lacking expression of **CD41**, **Gpr56**, and **endothelial markers Flk1, Cdh5, and AplnR**. Cells within this cluster also expressed high levels of **Runx1**. UC2 contained 19 cells (L1.3–L1.4) that expressed **Gata2** consistently, although there were too few cells for confident analysis. In a more detailed analysis of the expression of the heptad TFs, Gata2 showed some heterogeneity in expression levels in all clusters except UC1, in which it was downregulated (see Discussion). Tal1 also showed heterogeneous expression in all clusters. **Erg** and **Lmo2** were downregulated in HC2. These data demonstrate heterogeneity in heptad TF expression among the single cells of highly enriched HPCs and HSCs. While some cell clusters show consistent levels of one TF, other TFs vary in expression within a cluster, suggesting an unstable regulation and/or dynamic expression between highly similar single cells.

We also calculated the similarity of each cell to the gene expression profile of validated adult bone marrow HSCs using hscScore (Hamney and Göttgens, 2019) (Figure S1C). The cells within HC1 obtained the highest score, suggesting that this cluster most likely contains the HSC fraction. Our datasets were further examined in the context of a published single-cell dataset (Zhou et al., 2016) of phenotypically enriched endothelial cells, pre-HSCs (T1+T2), and E12/E14 HSCs (Zhou et al., 2016). Louvain identity (L1.1 to L1.11) and merged clusters (EC, HC1, HC2, UC1, and UC2) are also indicated (top).
(CD41, Mpl, and Ly6a) (mainly cells within L1.7) and cells that expressed predominantly only HSC-specific genes (mainly cells within L1.8). HC2 represented a more mature hematopoietic profile where cells expressed all HSC-specific and no pre-HSC-specific genes. UC1 had some endothelial and hematopoietic features and expressed mature hematopoietic genes, including Cd53, Cd84/SLAM, Mndal/myeloid, and Sp100/PML; almost no pre-HSC-specific genes; and a subset of endothelial-specific genes (Pmp22, Anxa2, Anxa3, Ophn1, Marcks, and Tcf4).

Phenotypic Characteristics Derived from Index-Sorting Guide Further Enrichment of Multipotent HCs

To examine whether the 5 transcriptomic clusters of CD31+ cKithiG2med cells correspond to specific hematopoietic activities, index-sorting data were analyzed for phenotypic characteristics that would provide for cluster-distinct cell isolation. EC cells were CD31hi and side-scatter high (SSChi), UC1+UC2+HC2 cells were CD31lo, and HC1 cells were CD31hi and SSClo (Figure 3A). No clear further separation was observed within the cKithi fraction (Figure 3B), although HC1 cells expressed higher...
levels of cKit than EC and UC1+UC2+HC2 cells, which showed variable levels. Based on these new sorting parameters (CD31hiSSClo, CD31hi, and CD31loSSClo), three fractions were sorted and tested in vitro for CFU-C formation and in vivo for long-term HSC repopulation.

No CFU-Cs were generated from EC cells (CD31hiSSClo; Figure 3C) as expected. CD31lo cKit+G2med cells (UC1+UC2+HC2) contained 50 CFU-Cs per embryo equivalent (ee) of various CFU-C types. CD31loSSClo cKit+G2med cells (HC1) contained 30 CFU-Cs per ee. Interestingly, a trend of more multipotent HPCs (7 GEMMs per ee), fewer bi-potent HPCs (17 GMs per ee), and few or no unipotent HPCs (Gs and burst-forming unit-erythroid [BFU-Es]) was detected in HC1 cells as compared to the UC1+UC2+HC2 cells (5 GEMMs per ee and 24.5 GMs per ee). The results of long-term in vivo transplantsations showed that the HC1 fraction contains all HSCs (Figure 3D) and is 10.5-fold enriched compared to the CD31+ cKit+G2med fraction.

To provide scope for further enrichment of HSCs and multipotent HPCs, HC1 cells (Figure 3Ei) were index-sorted and analyzed for single-cell–CFU-C lineage output and G2, cKit, and CD31 MFIs. A total of 86 CFU-Cs was obtained (plating efficiency, 1:9–11). Ninety-two percent (21% GEMMs + 71% GMs) of the CFU-Cs were multipotent/bipotent (Figure 3Eii). Cells yielding GEMMs showed significantly higher average MFIs for G2 expression as compared to GM and G cells (Figure 3Eiii). MFIs for cKit (Figure 3Eiii) and CD31 (Figure 3Eiv) were not significantly different when GEMM and GM cells were compared. Together, these data indicate that all HSCs and the most multipotent HPCs are enriched within the CD31hiSSClo cKit+G2med (HC1) fraction and that cells that produce GEMMs had higher G2 MFIs as compared to the more widely varying G2 MFIs for GM-producing cells.

**Iterative Analyses Link Expression of Specific Genes with HSC and Multipotent HPC Function**

Gene lists were examined for new candidate markers of HC1 cells as compared to the other clusters (Table S1). Violin plots showed significantly higher Tnfrsf7 (CD27) expression in HC1 (Figure 4A; Figure S2A), as well as Selp and Mpl (Figure S2B). Index-sorting parameters revealed that 70% of E11 CD31hiSSClo cKit+G2med cells were isolated and transplanted into irradiated recipients. High-level, long-term multilineage in vivo repopulation by the CD27+HC1 donor cells was detected in 2 of 2 recipients (Figure 4C, open squares). Secondary transplantsations confirmed that these were self-renewing HSCs (Figure S3A). For further enrichment, a higher and narrower (+N) CD27 gate (MFI = 150–7,000), CD27hi cells were isolated and transplanted into irradiated recipients. High-level, long-term multilineage in vivo repopulation by the CD27+HC1+ donor cells was detected in 2 of 2 recipients (Figure 4C, open squares). Secondary transplantsations confirmed that these were self-renewing HSCs (Figure S3A). For further enrichment, a higher and narrower (+N) CD27 gate (MFI = 500–7,000) was set (Figure 4B), and CD27+ (19%) and CD27−/lo (81%) HC1 cells were transplanted. The CD27+ HC1 cells yielded high-level, long-term multilineage repopulation in both primary (Figure 4C, black squares) and secondary (Figure S3A) transplantsations. No HSCs were found in the CD27−/lo HC1 fraction, even with injection of high numbers of cells. As compared to the CD31hiSSClo cKit+Gata2+G2med sorted fraction, the CD27+ HC1 gate enriches HSCs by a further 5.2-fold.

Single-cell RNA-seq of 119 CD31hiSSClo cKit+Gata2+G2med CD27+HC1 cells (from 8 AGMs) was performed, and data were combined with the 968 cells to produce a new SPRING plot of 1,087 cells. Louvain clustering showed a cluster distribution (L2.1–L2.12) of the 1,087 cells (Figures S2B and S3C) almost identical to that obtained from the 968 cells (Figure S1B). The majority (71%) of the 119 CD27+HC1 cells (Figure 4D, red) localized to HC1, and 58% localized to L2.8 (Figure 4D, circled).

Interestingly, when index-sorted CD27+HC1 cells were analyzed in single-cell CFU-C assays (plating efficiency, 1:3–4 cells), GEMM-forming cells showed a significantly lower average CD27 MFI (767.1 ± 41.8; Figure 4Ei) and a significantly higher G2 MFI (1,022.0 ± 60.1; Figure 4Eii) than those of GM-forming cells (945.2 ± 44.6 and 858.3 ± 31.8, respectively). Average cKit and CD31 MFIs were not significantly different (Figures 4Eii and 4Eiv). Notably, 95% of immature/HSC-like progenitors forming GEMMs fall within the smaller CD27 MFI range of 513–1,084 (Figure 4F, red dotted area), which is now referred to as CD27med. The CD27med fraction is 15.5% ± 1.3% of HC1 and 86% of CD27+HC1 cells. There are 2 ± 2 CD27med HC1 cells per AGM, of which 1–2 were HSCs at the time of analysis. In addition to our HC1 enrichment (10.5-fold), a further 6.5-fold enrichment of HSCs was achieved in the CD27med fraction to yield a combined enrichment of 68.3-fold (summarized in Figure 4G).

**Defining the AGM HSC Transcriptome**

The increased enrichment of multipotent cells within the CD27med MFI range provided a means for dataset refinement in which the molecular profiles of different functional subsets could be examined. Out of the 119 CD31hiSSClo cKit+Gata2+G2med cells previously isolated for single-cell RNA-seq (Figure 4D), 85 belonged to HC1, of which 14 showed CD27+ and 71 CD27med expression. We analyzed the 85 cells using iterative Clustering and Guide-gene Selection (ICGS) (Olsson et al., 2016), which can define groups of genes that behave similarly and identify groups of cells with similar gene-expression patterns. Three groups of genes and 3 groups of cells that we named ICGS1, ICGS2, and ICGS3 contained 54, 29, and 2 cells, respectively (Figures S4 and S5). When the distribution of the ICGS groups within the previously defined L2.7 and L2.8 clusters was examined, 81% of the 85 CD27med cells fell within L2.8, and 19% fell within L2.7. ICGS1 cells were mainly in L2.8, whereas ICGS2 cells were equally distributed in L2.7 and L2.8, suggesting that these cells are acquiring HSC identity. The hscScore for each cell was calculated and plotted against its CD27 MFI value (Figure 5A). ICGS2 cells were almost exclusively in the CD27med and high hscScore range, whereas ICGS1 cells were within the full CD27 MFI range and showed a low hscScore. Together, these data confirm that ICGS2 cells are the most multipotent or HSC-like.

To more precisely identify the transcriptome of the few HSCs within the 85 cell profiles, other known HSC characteristics were examined. Ly6a is known to be expressed by all in vivo adult repopulating AGM HSCs (de Brujin et al., 2002). Ly6a can be detected in 27 of the 85 cells within HC1 (Figure 5Bi), 23 of 71 CD27+ cells (Figure 5Bii), and 4 of 14 CD27hi cells (Figure 5Biii), consistent with previous findings that Ly6a expression is not exclusive to HSCs. We then focused our attention on the 23 Ly6a+ cells within the “multipotent/HSC” CD27med MFI range (Figure 5C). These cells separated into two ICGSs: 15 belonged

---

6 Cell Reports 31, 107627, May 12, 2020
to ICGS1, and 8 belonged to ICGS2. The 8 cells contained in ICGS2 had the highest hscScores: 4 belonged to L2.7, and 4 belonged to L2.8. Thus, the stratified results of single-cell index-sorting, transcriptome profiling, and functional analyses strongly suggest a profile for HSCs that is unique but closely related to that of multipotent HPCs.

Considering the expression of several genes relevant to HSC generation, the profile for the 8 Ly6a+CD27med ICGS2 single cells stands out (Figure 5C). Notch pathway genes (Bigas and Porcheri, 2018) are highly represented in these cells and include Hes1 and Hey1 (negative regulators), Dll4 (Notch ligand), and Notch4. Sox17, which is known to be expressed in fetal HSCs and maintain stem cell potency (He et al., 2011), is highly expressed, as is Il6st, which is part of the receptor complex for hematopoietic and stem cell factors interleukin (IL)-6, leukemia inhibitory factor (LIF), and Oncostatin M (Mihara et al., 2012). Also, Nos3 (endothelial nitric oxide synthetase), which functions in AGM HSC and HP/SC generation (North et al., 2009), is highly expressed. VEGF-R, Flt1, and Kdr are expressed along with other HSC genes.

Figure 4. All Functional AGM HSCs Are within the HC1 CD27med Fraction
(A) Violin plots of Tnfrsf7 (CD27) expression in the 5 major transcriptomic clusters. EC, endothelial-like cluster; HC1, hematopoietic cluster 1; HC2, hematopoietic cluster 2; UC1, unknown cluster 1; UC2, unknown cluster 2.
(B) Enrichment sorting parameters for CD27+ HC1 cell isolation. Side scatter (SSC-A), CD31, cKit, Gata2 (G2), and CD27 sorting gates and percentage (±SEM) of cells. CD27+ cells (%) are indicated within a wide (+W) and a narrow (+N) mean fluorescent intensity (MFI) gate.
(C) Analysis of HSC function within CD27 fractions (CD27+W, CD27+N, and CD27−/lo) of HC1 cells by in vivo transplantation. Open squares denote CD27+W HC1 cell recipients (n = 2) of 442 cells (2.5×10^3) and 479 cells (3.3×10^3) from 15 and 9 (42–50 somite pairs, sp) embryos, respectively. Closed squares (n = 4; 42–49 sp; embryos/experiment: 25, 11, 22, 28) denote recipients injected with CD27+N HC1 cells (238–957; 3.7–9.8 ee), EC cells (899–2,333; 3.9–8.6 ee), CD27−/lo HC1 cells (1,042–4,325; 4.1–9.8 ee), and UC1+UC2+HC2 cells (1,361–3,741; 4.2–9.9 ee). Percentage of donor cell chimerism in peripheral blood of recipients at 4 months post-injection. Black line indicates mean.
(D) Single-cell RNA-seq of 119 CD31hiSSC-hi cKithiGata2medCD27+N sorted cells (all passed quality control) from 6 embryos (41–47 sp). These data combined with the previous dataset (gray) resulted in an extended dataset SPRIING plot of 1,087 cells in total. Overlay of CD27−/hi cells (red) showing localization to one HC1 subcluster (circled).
(E) Single CD27−/hi HC1 cells from E11 G2V IAHCs (50 embryos; 42–49 sp) deposited into methylcellulose and colony-forming unit-culture (CFU-C) counts at 8–10 days (138 total CFU-Cs). Colony type (individual dot) plotted against log10-normalized mean fluorescence intensities (MFIs) for (Ei) CD27, (Eii) G2, (Eiii) cKit, and (Eiv) CD31 expression. Red lines indicate mean ± SEM.
(F) Representative FACS plot showing newly defined CD27med MFI range (red dotted line) within the CD27−/hi fraction. Percentage ± SEM.
(G) Summary of enrichment strategy. The HC1 fraction containing all HSCs is 10.5-fold enriched compared to previous sorting strategies (Eich et al., 2018). Further enrichments are based on gating of specific CD27−/hi and CD27med fractions to give an additional 6.5-fold enrichment. After iterative analyses and refinement of sorting gates, a total combined HSC enrichment of 68.3-fold is achieved.
such as Cdkn1c, Pdzk1ip1, Procr, Ramp2, Trim47, and Vwf.

Thus, the near-equivalence of the transcriptomic profiles of the 8 Ly6a+CD27med ICGS2 cells provides a likely genetic identity for the first cells achieving HSC function in the mouse embryo.

Functional HSCs Are Localized in 1- to 2-Cell IAHCs

Having defined the transcriptome of functional HSCs, we examined the localization of CD27-, Ly6a-, and Gata2-expressing cells within IAHCs. E11 G2V cryosections immunostained for CD27 and CD34 showed rare G2+CD34+CD27+ cells almost exclusively within the smallest ventrally located IAHCs. Figure 6A shows two CD34+G2+CD27+ cells (arrowhead) attached to one CD34+G2+CD27+ cell within the endothelium, and a single rounded/bulging CD34+G2+CD27+ cell (arrow). Another section shows a ventral cluster (Figure 6Bi) with two CD34+G2+CD27+ cells (one appearing to be bulging out of the endothelium) surrounded by 3-4 CD34+G2+CD27- cells. Also found ventrally (Figure 6Bii) was a 2-cell cluster with a single CD34+G2+CD27+ cell closely associated with a CD34+G2+CD27+ cell (Video S1). Figure 6Biii shows a single rounded bulging CD34+G2+CD27+ cell on the dorsal aspect of the aorta (likely an HPC; Taoudi and Medvinsky, 2007). Examination of CD27+ cell-containing IAHCs revealed that 80% (56 out of 70; 42 in 1-cell clusters and 14 in 2-cell clusters; n = 4 embryos) were 1- to 2-cell clusters (Figure S6). Since all HSCs and most multipotent progenitors are CD27+, we conclude that these functional cells are localized closely to the ventral aortic endothelium within small or single-cell IAHCs.

E11 Ly6aGFP transgenic embryo sections were also immunostained to examine whether this HSC marker could subset and/or provide further validation into the localization of CD27+ HSCs. A single GFP+CD27+ cell was observed at the apex of a bulging GFP+ cell within the ventral endothelium (Figure 6Ci, arrowhead); another GFP+ cell (arrow) was observed in the ventral endothelium, but it was CD27-. Figure 6Cii indicates two round adjacent CD27+ cells, of which...
one is GFPlo (arrowhead) and attached to interspersed GFP+ endothelial cells. Three CD27+ cells, one of which is GFP+ (Figure 6Ciii, arrowheads), were found in very close association with the ventral endothelium. The juxtaposition of single CD27+Ly6aGFP+ cells to the ventral aortic endothelium indicates that HSCs emerge directly from embryonic endothelium as single cells within intra-aortic clusters of 1–2 cells.

**DISCUSSION**

Here, we identified the transcriptome of the first functional adult repopulating HSCs. Starting from Gata2-based single-cell index-sorting and iterative single-cell analyses, the “inner” HSC transcriptional regulatory network is connected directly with multipotent hematopoietic function in vivo and in vitro. Specific levels of Gata2, cKit, and CD27 expression define all functional in vivo repopulating HSCs, and, together with Ly6aGFP expression, CD27 localizes the first HSC to the smallest intra-aortic clusters of 1–2 cells. The surprising variability in gene expression levels within highly enriched single cells suggests that establishment of hematopoietic function is stochastic, occurring through dynamic regulation and/or unstable expression of groups of pivotal genes.

**Emerging HCIs Are Heterogeneous in Phenotype, Transcriptome, and Function**

EHT occurs during a short 2- to 3-day period during mouse midgestation (Yokomizo and Dzierzak, 2010; Zovein et al., 2008) and extends over the length of the aorta to produce IAHCs that are predominantly ventral (Peeters et al., 2009; Soulholt et al., 2016; Taoudi and Medvinsky, 2007) and more abundant near the juncture of the vitelline artery (Mascarenhas et al., 2009). Individual intra-aortic clusters range from 1 to 19 cells at E10.5 and from 1 to 11 cells at E11.5 (Yokomizo and Dzierzak, 2010). Single-cell quantitation of functional HPCs and HSCs within the clusters shows that the number of lineage-restricted HPCs is the highest, followed by that of multipotent HPCs and rare HSCs. This low frequency of HSCs implies that stem cell identity/function is a difficult hematopoietic fate to achieve.

Whereas all functional HSCs are in ventral clusters and near the vitelline juncture, it cannot be predicted which emerging cell in the aorta is an HSC. The variability in phenotypic marker expression, and the degree of heterogeneity found in the transcriptomes of individual cells during the embryonic development of HPCs, pre-HSCs, and HSCs (Zhou et al., 2016), has precluded the identification of the transcriptome of rare HSCs. Rather than dependence on surface molecules not required for HSC
function, Gata2-based index-sorting more directly connected with the “inner” hematopoietic regulatory network. G2medCD31hi cKit+ cells highly enriched in multipotent HPCs and containing all functional repopulating HSCs show unexpected heterogeneity between single-cell transcriptomic profiles. Louvain analysis of >1,000 sequenced cells revealed 5 major cell clusters. Unexpectedly, within this highly enriched set of IAHC cells, one major cluster still expressed endothelial markers and was deemed endothelial-like (an EC). Surprisingly, it did not express Runx1. The second major cluster exhibited a hematopoietic gene-expression profile, and other minor clusters showed mixed or mature HC profiles.

Considering that the G2medCD31hi cKithi fraction contains a high frequency of multipotent HPC and HSC activity, it begs the question whether the EC and/or the other minor cell clusters possess some degree of hematopoietic identity but either are not yet functional or are the predecessors of HPCs and HSCs. Data from several laboratories have described embryonic HCs lacking HSC activity and have shown that these cells can exhibit HSC function after in vitro culturing (Ganuza et al., 2017a; Rybtsov et al., 2016) or neonatal transplantation (Yoder et al., 1997), perhaps by stabilization of a genetic program. While some of these cells could represent pre-HSCs, others could be products of EHT that never successfully achieve HSC function. They may be in a “neutral” or “naive” state awaiting establishment of another functional hematopoietic identity, perhaps to a multipotent progenitor that others have shown to maintain steady-state hematopoiesis (Busch et al., 2015; Pei et al., 2017; Rodriguez-Fraticelli et al., 2018; Schoedel et al., 2016; Sun et al., 2014).

Novel Combinations of Iterative Single-Cell Methods Identify the HSC Transcriptome

Iterative single-cell RNA-seq, index-sorting, functional analyses, and bioinformatics led us to progressively higher enrichments of HSCs. Several putative HSC markers were identified within the HC1 single-cell transcriptomes and included CD27, Selp/P-selectin, and Mpl. Expression level distribution of single cells in Louvain clusters indicated that CD27 and Selp were likely to be most useful. Previously, CD27 was reported to be expressed on bone marrow (Vazquez et al., 2015) and AGM HSCs (Li et al., 2017). Although CD27 is not exclusive to HSCs (and is expressed by some cells in the EC Louvain cluster), we discovered that cells expressing specific medium levels of CD27 yielded a further 6.5-fold enrichment on top of the HC1-only enrichment. While multipotent/bipotent HPCs were also enriched in this CD27med fraction, all unipotent HPCs were absent. As Selp is most prominently expressed by HC1 cells, it also could offer an opportunity for further enrichment of aortic HSCs (Sullivan et al., 2011). Finally, Mpl, the receptor for thrombopoietin is known to regulate the generation and self-renewal of AGM HSCs (Petit-Cocaut et al., 2007). Together, these markers offer additional opportunities for the isolation/identification of single functional HSCs as they are first generated in the embryo.

Stratifications of single-cell transcriptomic data revealed the expression of a combination of other genes expressed by functional HSCs. These include Ly6a, DLL4, Sox17, and eNOS, as well as others. Ly6aGFP has been shown to be expressed on all AGM HSCs (de Bruijn et al., 2002; Solaimani Kartalaei et al., 2015); DLL4 is involved in arterial endothelial identity and, most recently, has been shown to govern the dynamics of cluster formation and HSC production (Bonkhofer et al., 2019; Gama-Norton et al., 2015; Porcheri et al., 2020); Sox17 maintains stem cell potency (He et al., 2011); and the deficiency of eNOS affects the development of IAHCs and functional HSCs (North et al., 2009). The fact that the combined expression of these genes occurs in the 8 cells within the 119 sequenced G2medSSCloCD31hi cKithiCD27+ single cells (isolated from 6 embryos) has important implications for understanding the regulation of HSC fate/function establishment and provides a useful reference for the ex vivo generation and/or growth of HSCs.
Emerging Functional HSCs Are Localized in 1- to 2-Cell Ventral Aortic Clusters

Since HSC fate acquisition is rare, there has been much speculation on the specific localization of HSCs within the IAHs. Do they take on HSC fate immediately upon emergence from the hemogenic endothelium, or do they emerge as pre-HSCs? Do they emerge as a single cell and remain as a 1-cell cluster, or are they within larger clusters of proliferative cells?

Time-lapse imaging of the zebrafish and mouse embryonic aorta proved that the AGM region is a HC-generative niche (Bertrand et al., 2010; Boisset et al., 2010; Kissa and Herbomel, 2010). Single HCs emerge directly from ventral aortic endothelial cells, and results of genetic studies support that at least some of the cells taking on hematopoietic identity are functional HSCs. Here, by co-expression of CD27 and Ly6aGFP, we show the localization of HSCs to the small 1- to 2-cell aortic clusters. Additionally, our RNA-seq data show DLL4 expression within larger clusters of proliferative cells?

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
  - Lead Contact
  - Materials Availability
  - Data and Code Availability
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
  - Mouse models
- **METHOD DETAILS**
  - Mouse embryo generation
  - Preparation of AGM single cell suspensions for flow cytometry
  - Hematopoietic stem cell transplantation and (single-cell) CFU-C assay
  - Immunohistochemistry on cryosections
  - Single-cell RNA sequencing
  - Bioinformatic Analysis
- **QUANTIFICATION AND STATISTICAL ANALYSIS**

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.celrep.2020.107627.

ACKNOWLEDGMENTS

The authors thank all lab members for helpful comments, particularly Carmen Rodriguez-Seoane and Sarah J. Kinston for technical assistance and the QMRI Flow Cytometry and Cell Sorting Facility (S. Johnston, W. Ramsay, and M. Pattinson) for cell sorting. These studies were supported by ERC Advanced Grant 341096 (to E.D.), BBSRC research grant BB/S01845X/1 (to E.D.), and Bloodwise research grant 18010 (to E.D.). Work in the B.G. group was supported by Wellcome (206328/Z/17/Z), MRC (MR/M008975/1), CRUK (C1163/A21762), Bloodwise (18002), and NIH NIDDK (R24 DK106766-01A1), as well as by core infrastructure support to the Cambridge Stem Cell Institute by Wellcome and MRC (097922/Z/11/Z).

AUTHOR CONTRIBUTIONS

C.S.V. performed the embryo dissections, cell preparations, flow cytometry, in vitro/in vivo hematopoietic assays, and immunofluorescence staining and microscopy. A.M. helped with the collection of cells for RNA-seq and mouse irradiations. F.J.C.-N., W.J., and X.W. performed RNA-seq and bioinformatics analyses. S.A.M. performed dissection, cell preparation, and isolation for CD27 RNA-seq. C.S.V., B.G., and E.D. designed the study and interpreted data, and C.S.V. and E.D. wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: January 27, 2020
Revised: March 18, 2020
Accepted: April 18, 2020
Published: May 12, 2020

REFERENCES

Abranches, E., Guedes, A.M., Moravec, M., Maamar, H., Svoboda, P., Raj, A., and Henrique, D. (2014). Stochastic NANOG fluctuations allow mouse embryonic stem cells to explore pluripotency. Development 141, 2770–2779.
Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics 31, 166–169.
Baron, C.S., Kester, L., Klaus, A., Boisset, J.C., Thambryrajah, R., Yvennogeau, L., Kouskoff, V., Lacaud, G., van Oudenaarden, A., and Robin, C. (2018). Single-cell transcriptomics reveal the dynamic of haematopoietic stem cell production in the aorta. Nat. Commun. 9, 2517.
Becker, James, T., Huff, Jacob, Nebes, Robert, D., Holland, Audrey, Boller, Francois, et al. (1988). Neuropsychological Function in Alzheimer’s Disease: Pattern of Impairment and Rates of Progression. Arch Neurol. 45 (3), 263–268.
Bee, T., Liddiard, K., Swiers, G., Bickley, S.R., Vink, C.S., Jarratt, A., Hughes, J.R., Medvinsky, A., and de Brujin, M.F. (2008). Alternative Runx1 promoter usage in mouse developmental hematopoiesis. Blood Cells Mol. Dis. 43, 35–42.
Bertrand, J.Y., Chi, N.C., Santos, B., Teng, S., Stainier, D.Y., and Traver, D. (2010). Haematopoietic stem cells derive directly from aortic endothelium during development. Nature 464, 108–111.
Bigas, A., and Porcheri, C. (2018). Notch and Stem Cells. Adv. Exp. Med. Biol. 1066, 235–263.
Boisset, J.C., van Cappellen, W., Andrieu-Soler, C., Galjart, N., Dzierzak, E., and Robin, C. (2010). In vivo imaging of haematopoietic cells emerging from the mouse aortic endothelium. Nature 464, 116–120.
Bonkhofer, F., Riapoli, R., Pinheiro, P., Krecaamarik, M., Schneider-Swales, J., Tsang, I.H.C., de Brujin, M., Monteiro, R., Peterkin, T., and Patient, R. (2019).
Blood stem cell-forming haemogenic endothelium in zebrafish derives from arterial endothelium. Nat. Commun. 10, 3577.

Brennecke, P., Anders, S., Kim, J.K., Kolodziejczyk, A.A., Zhang, X., Proserpio, V., Baying, B., Benes, V., Teichmann, S.A., Marioni, J.C., and Heisler, M.G. (2013). Accounting for technical noise in single-cell RNA-seq experiments. Nat. Methods 10, 1093–1095.

Busch, K., Klapproth, K., Barile, M., Flossdorf, M., Holland-Lettz, T., Schlienger, S.M., Reth, M., Höfer, T., and Rodewald, H.R. (2015). Fundamental properties of unperturbed haematopoiesis from stem cells in vivo. Nature 518, 542–546.

Chang, H.H., Hemberg, M., Barahona, M., Ingber, D.E., and Huang, S. (2006). Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. Nature 453, 544–547.

Chen, M.J., Yokomizo, T., Zeigler, B.M., Dzierzak, E., and Speck, N.A. (2009). Runx1 is required for the endothelial to haematopoietic cell transition but not thereafter. Nature 457, 887–891.

Ciau-Ultz, A., and Patient, R. (2019). Gene Regulatory Networks Governing the Generation and Regeneration of Blood. J. Comput. Biol. 26, 719–725.

Cohen-Saidon, C., Cohen, A.A., Sigal, A., Liron, Y., and Alon, U. (2009). Dynamics and variability of ERK2 response to EGF in individual living cells. Cell 137, 815–825.

Cohen-Saidon, C., Cohen, A.A., Sigal, A., Liron, Y., and Alon, U. (2009). Understanding the Mechanisms of ERK2 Oscillations. Cell 139, 26–37.

Cohen, S., and Heisler, M.G. (2013). Accounting for technical noise in single-cell RNA-seq experiments. Nat. Methods 10, 887–891.

Cohyla, S., and Cedar, H. (2014). Gene Regulatory Networks Governing the Generation and Regeneration of Blood. J. Comput. Biol. 26, 719–725.

Cooper, J., Zhou, Z., and Career, M. (2015). Runx1 Is Required for the Endothelial to Haematopoietic Cell Transition but Not Thereafter. Nature 457, 887–891.

Cui, Y., and Shen, Q. (2019). Runx1 is required for the endothelial to haematopoietic cell transition but not thereafter. Nature 457, 887–891.

Cui, Y., and Shen, Q. (2019). Runx1 is required for the endothelial to haematopoietic cell transition but not thereafter. Nature 457, 887–891.
Moignard, V., Woodhouse, S., Haghverdi, L., Lilby, A.J., Tanaka, Y., Wilkinson, A.C., Buettner, F., Macaulay, I.C., Jawaid, W., Diamanti, E., et al. (2015). Decoding the regulatory network of early blood development from single-cell gene expression measurements. Nat. Biotechnol. 33, 269–276.

Müller, A.M., Medvinsky, A., Strouboulis, J., Grosfeld, F., and Dzierzak, E. (1994). Development of hematopoietic stem cell activity in the mouse embryo. Immunity 1, 291–301.

North, T.E., de Bruin, M.F., Stacy, T., Talebian, L., Lind, E., Robin, C., Binder, M., Dzierzak, E., and Speck, N.A. (2002). Runx1 expression marks long-term repopulating hematopoietic stem cells in the medidgestation mouse embryo. Immunity 15, 661–672.

North, T.E., Goessling, W., Peeters, M., Li, P., Ceoli, C., Lord, A.M., Weber, G.J., Harris, J., Cutting, C.C., Huang, P., et al. (2009). Hematopoietic stem cell development is dependent on blood flow. Cell 137, 736–748.

Olsson, A., Venkatadasabramaniam, M., Chaudhri, V.K., Aronow, B.J., Salomonis, N., Singh, H., and Grimes, H.L. (2016). Single-cell analysis of mixed-lineage states leading to a binary cell fate choice. Nature 537, 998–702.

Peeters, M., Ottersbach, K., Boillerot, K., Orelio, C., de Bruin, M., Wijgerde, M., and Dzierzak, E. (2009). Ventral embryonic tissues and Hedgehog proteins induce early AGM hematopoietic stem cell development. Development 136, 2613–2621.

Pei, W., Fraying, T.B., Rössler, J., Wang, X., Postrach, D., Busch, K., Rodé, I., Klapproth, K., Dietlein, N., Quedenau, C., et al. (2017). Polyxoy barcoding reveals hematopoietic stem cell fates realized in vivo. Nature 548, 456–460.

Petit-Cocault, L., Volle-Challier, C., Fleury, M., Pélault, B., and Souyri, M. (2007). Dual role of Mpl receptor during the establishment of definitive hematopoiesis. Development 134, 3031–3040.

Picelli, S., Faridani, O.R., Björklund, A.K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smartseq2. Nat. Protoc.

Pimanda, J.E., and Göttgens, B. (2010). Gene regulatory networks governing haematopoietic stem cell development and identity. Int. J. Dev. Biol. 54, 1201–1211.

Porcheril, C., Golan, O., Calero-Nieto, F.J., Thambyprajah, R., Ruiz-Herguido, C., Wang, X., Catto, F., Guillem, Y., Sinha, R., Gonza´ lez, J., et al. (2020). Notch ligand Dll4 impairs cell recruitment to aortic clusters and limits blood stem cell generation. EMBO J. 39, e104270.

Purvis, J.E., and Lahav, G. (2013). Encoding and decoding cellular information through signaling dynamics. Cell 152, 945–956.

Rodriguez-Fraticelli, A.E., Wolock, S.L., Weinreb, C.S., Panero, R., Patel, S.H., Janovick, M., Sun, J., Calogero, R.A., Klein, A.M., and Camargo, F.D. (2018). Clonal analysis of lineage fate in native hematopoiesis. Nature 553, 212–216.

Rybtsov, S., Ivanovs, A., Zhao, S., and Medvinsky, A. (2016). Concealed expansion of immature precursors underpins acute burst of adult HSC activity in foetal liver. Development 143, 1284–1289.

Ryu, H., Chung, M., Dobrynski, M., Fey, D., Blum, Y., Sik Lee, S., Peter, M., Khodolenko, B.N., Li Jeon, N., and Pertz, O. (2016). Frequency modulation of ERK activation dynamics rewrites cell fate. Mol. Syst. Biol. 12, 866.

Sasmono, R.T., Oceandy, D., Pollard, J.W., Wang, X., Pavli, P., Wainwright, C.J., Wang, Q., Stacy, T., Binder, M., Marin-Padilla, M., Sharpe, A.H., and Speck, N.A. (1996). Disruption of the Cbfa2 gene causes necrosis and hemorrhaging in the central nervous system and blocks definitive hematopoiesis. Proc. Natl. Acad. Sci. USA 93, 3444–3449.

Weinreb, C., Wolock, S., and Klein, A.M. (2018). SPRING: a kinetic interface for visualizing high dimensional single-cell expression data. Bioinformatics 34, 1246–1248.

Wilson, N.K., Foster, S.D., Wang, X., Knezevic, K., Schütte, J., Kaimakis, P., Chilarska, P.M., Kinston, S., Ouwehand, W.H., Dzierzak, E., et al. (2010). Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators. Cell Stem Cell 7, 532–544.

Wu, T.D., Reeder, J., Lawrence, M., Becker, G., and Brauer, M.J. (2016). GMAP and GSNAP for Genomic Sequence Alignment: Enhancements to Speed, Accuracy, and Functionality. Methods Mol. Biol. 1418, 283–334.

Yoder, M.C., Hiatt, K., and Mukherjee, P. (1997). In vivo repopulating hematopoietic stem cells are present in the murine yolk sac at day 9.0 postconitus. Proc. Natl. Acad. Sci. USA 94, 6776–6780.

Yokomizo, T., and Dzierzak, E. (2010). Three-dimensional cartographic of hematopoietic clusters in the vasculature of whole mouse embryos. Development 137, 3651–3661.

Zovein, A.C., Hofmann, J.J., Lynch, M., French, W.J., Turlo, K.A., Yang, Y., Becker, M.S., Zanetta, L., Dejana, E., Gasson, J.C., et al. (2008). Fate tracing reveals the endothelial origin of hematopoietic stem cells. Cell Stem Cell 3, 625–636.
## STAR METHODS

### KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---------------------|--------|------------|
| Antibodies          |        |            |
| Rat anti-Mouse CD31-BV605 (clone 390) | BioLegend | Cat# 102427; RRID: AB_2563982 |
| Rat anti-Mouse CD117-BV421 (c-Kit) (clone 2B8) | BD Biosciences | Cat#562609; RRID: AB_11154585 |
| Armenian hamster anti-Mouse CD27-APC (clone LG.7F9) | eBioscience/Thermo Fisher Scientific | Cat# 17-0271-82; RRID: AB_469370 |
| Rat anti-Mouse CD34-biotin (clone RAM34) | eBioscience/Thermo Fisher Scientific | Cat# 13-0341-82; RRID: AB_466425 |
| Rabbit anti-GFP     | MBL International | Cat# 598; RRID: AB_591819 |
| Armenian hamster anti-Mouse CD27 (clone LG.7F9) | eBioscience/Thermo Fisher Scientific | Cat# 14-0271-82; RRID: AB_467183 |
| Streptavidin-Alexa Fluor 555 | Invitrogen/Thermo Fisher Scientific | Cat# s21381; RRID: AB_2307336 |
| Goat anti-Rabbit IgG (H+L)-Alexa Fluor 488 | Invitrogen/Thermo Fisher Scientific | Cat# A-11008; RRID: AB_143165 |
| Goat anti-Armenian hamster IgG-Alexa Fluor 647 (clone Poly4055) | BioLegend | Cat# 405510, RRID: AB_2566695 |
| APC anti-Mouse CD45.1 (clone A20) | BD Biosciences | Cat#558701; RRID: AB_1645214 |
| PE anti-Mouse CD45.2 (clone 104) | BD Biosciences | Cat#5609695; RRID: AB_1727493 |
| Chemicals, Peptides, and Recombinant Proteins |        |            |
| Collagenase Type I from *Clostridium histolyticum* | Sigma-Aldrich | Cat#C0130; Lot#02M1425V |
| Foetal Bovine Serum Standard Quality | Brunschwig Chemie BV | Cat#BS01500; Lot#212-223650 |
| 10% Triton X-100 | Sigma-Aldrich | Cat#93443-100 ml |
| SUPERase-In RNase Inhibitor | Ambion | Cat#AM2694 |
| dNTP mix | ThermoFisher | Cat#10319879 |
| ERCC RNA Spike-In Mix | ThermoFisher | Cat#4456740 |
| Superscript II Reverse Transcriptase | ThermoFisher | Cat#18064-014 |
| Betaine | Sigma-Aldrich | Cat#B0300 |
| Magnesium chloride | ThermoFisher | Cat#AM9530G |
| KAPA HiFi Hotstart ReadyMix | KAPA Biosystems | Cat#KK2601 |
| Agencourt AMPure XP beads | Beckman Coulter | Cat#A63881 |
| EB solution | QIAGEN | Cat#19086 |
| Critical Commercial Assays |        |            |
| High-sensitivity DNA chip | Agilent | Cat#5067-4626 |
| Quant-IT Picogreen double stranded DNA assay kit | ThermoFisher | Cat#P7589 |
| Nextera XT DNA sample preparation kit 96 samples | Illumina | Cat#FC-131-1096 |
| Nextera XT 96-Index kit, 384 samples | Illumina | Cat#FC-131-1002 |
| KAPA qPCR quantification kit | KAPA Biosystems | Cat# KK4824 |
| Deposited Data |        |            |
| Single-cell RNaseq | This paper | Accession number: GSE143637 |
| Experimental Models: Organisms/Strains |        |            |
| Mouse: Gata2Venus | Laboratory of Elaine Dzierzak; Kaimakis et al., 2016, Blood | N/A |
| Mouse: B6.Cg-Tg(Ly6a-EGFP)G5Dzk/J | The Jackson Laboratory | https://www.jax.org/strain/012643 |
| Mouse: B6.Cg-Tg(Csf1r-EGFP)1Hume/J (also known as MacGreen) | The Jackson Laboratory | https://www.jax.org/strain/018549 |

(Continued on next page)
REAGENT or RESOURCE SOURCE IDENTIFIER

Mouse: B6.SJL-‐Ptprc<sup>+</sup>Pepc<sup>−</sup>/BoyCrl
(also known as Ly5.1) Charles River Strain code: 494 https://www.criver.com/products-services/find-model/ly51-mouse?region=3671

Mouse: C57BL/6JolaHsd Envigo Order code: 057 https://www.envigo.com/products-services/research-models-services/models/research-models/mice/inbred/c57bl-6-inbred-mice/c57bl-6olahsd/u.k.aspx

Oligonucleotides
TSO: AAGCAGTGGTATCAACG CAGAGTACAT<sup>rGrG</sup>+<sup>G</sup>-<sup>3</sup> Picelli et al., 2014 N/A
Oligo-dT<sub>30VN</sub>: AAGCAGTGGTATCAACGCAGAGTAC(T<sub>30</sub>)VN-<sup>3</sup> Picelli et al., 2014 N/A
ISPCR: AAGCAGTGGTATCAACG CAGAGT Picelli et al., 2014 N/A

Software and Algorithms
Prism Version 7 GraphPad http://www.graphpad.com/scientific-software/prism
BD Facs Diva version 8 BD Biosciences http://www.bdbiosciences.com/en-us
FlowJo Version 10.1r5 Tree Star, Inc https://www.flowjo.com/solutions/flowjo/downloads/previous-versions
GSNAP Wu et al., 2016 http://research-pub.gene.com/gmap/
HTSeq Anders et al., 2015 https://htseq.readthedocs.io/en/release_0.11.1/install.html
PCA Becker et al., 1988 pcomp in R stats package
Selection of Highly Variable Genes Brennecke et al., 2013 https://static-content.springer.com/esm/art%3A10.1038%2Fnmeth.2645/MediaObjects/41592_2013_BFnmeth2645_MOESM313_ESM.pdf
ICGS Olsson et al., 2016 https://altanalyze.readthedocs.io/en/latest/ICGS/
DESeq2 Love et al., 2014 http://bioconductor.org/packages/release/bioc/html/DESeq2.html
SPRING Weinreb et al., 2018 https://github.com/AlionKleinLab/SPRING
hscScore Hamey and Göttgens, 2019 https://github.com/fionahamey/hscScore
Other
Interactive website This paper https://gottgens-lab.stemcells.cam.ac.uk/DZIERZAK/
Single-cell RNaseq Zhou et al., 2016 Accession number: GSE67120

RESOURCE AVAILABILITY

Lead Contact
Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Elaine Dzierzak (Elaine.Dzierzak@ed.ac.uk).

Materials Availability
This study did not generate new unique reagents.

Data and Code Availability
The single-cell RNA sequencing data generated in this study have been deposited in GEO (https://www.ncbi.nlm.nih.gov/geo/) with accession number: GSE143637. The code used is available on request. An interactive website containing the single-cell RNASeq results is available at https://gottgens-lab.stemcells.cam.ac.uk/DZIERZAK/.

Our datasets were examined in the context of a published single cell dataset (Zhou et al., 2016) which can be found in GEO under accession number: GSE67120.
EXPERIMENTAL MODEL AND SUBJECT DETAILS

Mouse models
All mice were housed and bred in animal facilities at the University of Edinburgh, in compliance with UK Home Office regulations. All animal experiments were conducted according to a UK Home Office Project License (PPL 70/8076 and continuation PC6D479E0) and approved by the University of Edinburgh Ethical Review Committee.

Adult C57BL/6JOLAhsd (Ly5.2) wild-type females were obtained from Envigo (order code: 057) and used at 2-4 months old. Gata2+/+ (Kaimakis et al., 2016), Ly6aGFP+/− and Csf1rGFP+/− (MacGreen, (Sasmono et al., 2003) males (> 10 backcrosses onto C57BL/6JOLAhsd) were used at 2-12 months old.

Preparations of AGM single cell suspensions for flow cytometry
AGM (aorta-gonad-mesonephros) regions (including trunk portion of the vitelline and umbilical arteries) of E10/11 embryos were dissected as previously described (Medvinsky et al., 2008). Single cell suspensions were prepared by enzymatically digesting tissues at 37 °C in 0.125% type 1 collagenase (C0130, Sigma-Aldrich) for 45 min, after which tissues were further mechanically disrupted with a P1000 pipette. Cells were subsequently passed over a 0.35 μm strainer, collagenase washed away with PBS+10%FBS+1%P/S (PPF), spun down (234g) and resuspended in PPF for flow cytometric antibody staining. Cells were incubated at 4 °C for 30 min in PPF with directly conjugated antibodies against CD31 (CD31-BV605, 1:100, cat# 102427, Biolegend), cKit (cKit-BV421, 1:200, cat.# 562609, BD) and CD27 (CD27-APC, 1:75, cat.# 17-0271-82, ebioscience). Excess of antibodies was washed away with PPF before fluorescence activation buffer (FACS) on an Ariall (BD). Dead cell exclusion was by Hoechst 33342 and gates were set based on unstained WT and fluorescence-minus-one (FMO) controls. Sorted single cells were collected in 50% FBS/PBS for functional analyses (transplantation + bulk CFU-C assay). Both collagenase and FCS were batch tested for optimal AGM conditions.

Hematopoietic stem cell transplantation and (single-cell) CFU-C assay
For transplantation, sorted cell populations (Ly5.2/CD45.2 background) were resuspended in PBS together with 2x10^5 Ly5.1 spleen carrier cells and intravenously injected into irradiated Ly5.1 (CD45.1) recipients that had received two split-doses of 4.5 Gy (Cs-137 Y-irradiator). Peripheral blood was sampled for CD45.1- and CD45.2 flow cytometric analysis at 4 and 16 weeks post-transplantation. Mice were considered repopulated with long-term reconstituting cells at %50 donor chimerism. For population/bulk CFU-C assays, sorted cells were resuspended and cultured in Methylcellulose medium (M3434; Stem Cell Technologies) in a humidified chamber at 37 °C and 5% CO2 for 9-11 days after which colonies were counted. For single-cell CFU-C assays, cells were sorted directly into 96-well plates with one (P1000) drop of Methylcellulose medium deposited per well and cultured and counted under the same conditions as for population CFU-C assays.

Immunohistochemistry on cryosections
Gata2+/+ and Ly6aGFP+/− and MacGreen−/− embryos were staged and fixed in 2% PFA/PBS for 1 hr and quickly washed in PBS. Embryos were then equilibrated overnight (O/N) in 20% sucrose/PBS at 4 °C, after which they were washed in Tissue-Tek O.C.T. embedding matrix (Sakura). Tissue paper was used to remove remaining sucrose and each embryo was placed in an individual mold filled with embedding matrix. Forelimbs were orientated exactly above hind limbs before being frozen on dry ice in absolute ethanol. 10 μm transversal cryosections were cut through the AGM region and collected on Superfrost Plus slides (Thermo Fisher Scientific). For immunohistochemistry, sections were fixed for 10 mins with acetone, washed three times with 0.05% Tween20/PBS (PBST) and then blocked with a 1:1 mixture of Streptavidin solution (Vector Laboratories Inc) and PBSblock (0.05% Tween20, 1% bovine serum albumin/PBS) for 15 mins. After three PBST washes, slides were incubated with Biotin solution (Vector Laboratories Inc) for 15 mins and washed another three times in PBST. Sections were incubated with rat anti-mouse-CD34-biotin (1:100, 13-0341-82, eBioscience), rabbit anti-GFP (1:1000, 598, MBL) and Armenian hamster anti-mouse-CD27 (1:100, 14-0271-82, ebioscience) diluted in PBSblock for 1hr at RT. Unbound primary antibodies were washed away with PBST after which sections were incubated with Streptavidin-AlexaFluor555 (1:500, S21381, Invitrogen), anti-rabbit AlexaFluor488 (1:500, A11008, Invitrogen) and anti-Armenian hamster AlexaFluor647 (1:500, 405510, Biolegend) diluted in PBSblock for 30 mins at RT. Slides were washed three times in PBST and incubated in DAPI (1:1000) diluted in PBSblock for 10 mins at RT. Afterward, slides were washed once in PBST and
then mounted with SlowFade Diamond Anti-Fade Mountant (Invitrogen, Thermo Fisher Scientific). Sections were imaged with a Leica SP8 confocal microscope and images and videos generated with the LAS X software (Leica).

**Single-cell RNA sequencing**

Cells were sorted in 2.3 μl of lysis buffer containing 0.2% Triton X-100 (Sigma-Aldrich) and 1U of Superase-In RNase Inhibitor (Ambion) and processed following the Smart-Seq2 protocol (Picelli et al., 2014). Following preamplification, all samples were purified using Ampure XP beads (Beckman Coulter) at a ratio of 1:0.6. The cDNA was then quantified using the Quant-it PicoGreen dsDNA Assay Kit (Thermo Fisher) and size distributions were checked on high-sensitivity DNA chips (Agilent Bioanalyzer). Samples were used to construct Nextera XT libraries (Illumina) from 125 pg of preamplified cDNA. Libraries were purified and size selected (0.5X-0.7X) using Ampure XP beads. Then, libraries were quantified using KAPA qPCR quantification kit (KAPA Biosystems), pooled and sequenced in an Illumina HiSeq 4000 instrument.

**Bioinformatic Analysis**

Reads were mapped to a custom reference index that included the *Mus musculus* genome (EMSEMBL GRCm38.p4 Release 81) and the sequences of ERCC spike-ins (Life Technologies), IRES and Venus using GSNAP (version 2014-10-07) (Wu et al., 2016) with parameters: -A sam -B 5 -t 24 -n 1 -Q -N 1. HTseq-count was used to count reads mapped to each gene with parameters: -s no (Anders et al., 2015). For further analyses, we only retained samples that had: (1) more than 100,000 reads mapped to nuclear mRNAs; (2) a proportion of reads mapped to genes (nuclear + mitochondria) relative to total reads above 30%; and (3) less than 20% of mapped reads allocated to mitochondrial genes. Overall, 47 cells (4.1% of total) failed the QC under those criteria and 1087 cells were kept for further analysis distributed as follows: 87 E10 CD31+cKithiG2int cells, 409 E11 CD31+cKithiG2int cells, 472 E11 CD31+cKithiG2lo cells and 119 E11 CD31hiSSClocKithiG2medCD27+N cells. The mean of nuclear reads mapped per cell was 1,213,237. Data were normalized for sequencing depth and RNA quantity using size factors calculated on endogenous genes (Brennecke et al., 2013). Highly variable genes were identified as previously described (Brennecke et al., 2013), using a false discovery rate threshold equal to 0.1. Only highly variable genes were considered to perform PCA analysis, using the prcomp function in R. SPRING visualization (Weinreb et al., 2018) and was obtained using the top 90 principal components and the 4 nearest neighbors. Clusters were defined using the cluster_louvain function in igraph package. Violinplots were obtained using the ggplot2 package in R. hscScores were calculated as previously described (Hamey and Göttgens, 2019). Differential expression analysis was performed using DESeq2 (Love et al., 2014).

CD27+ cells (Protein level) was extracted from HSC clusters (clusters L2.7 and L2.8). In total, there are 85 cells, with 4444 highly variable genes selected. Highly variable genes were used for ICGS clustering (AltAnalyzer (Olsson et al., 2016)) with ‘stringent’ with cell cycle and all other parameters as default.

For integration with previously published dataset (Zhou et al., 2016), the data was downloaded from GEO repository (GSE67120) and processed as indicated above. Differential expression analysis was performed between either endothelial cells or pre-HSCs (including T1 and T2) or HSCs (including E12 and E14) and the remaining cells to obtain cell type specific genes using DESeq2. Only genes with log2 fold change > 2, adjusted p value (BH method) < 0.1 and base mean > 50 were used for further analysis. Only the top 30 upregulated genes were selected to use in the heatmap.

**QUANTIFICATION AND STATISTICAL ANALYSIS**

All graphs were generated using GraphPad Prism 7. We performed Bartlett’s test for homogeneity of variances to ensure variances for all samples were equal. With the assumption of normality, the data were compared for the same means using one-way ANOVA, followed by a pairwise comparison using a Student’s t test, and Bonferroni correction. Significance was determined as p < 0.05, **p < 0.05, ***p < 0.01, ****p < 0.001. FACS percentages, averages (mean; red horizontal lines in Figures 1C, 1D, 3E, and 4E) and error bars are ± standard error of the mean (SEM). Further statistical details of experiments can be found in the figure legends. The number of biological replicates is indicated with ‘n’.