Whole-transcriptome analysis of endothelial to hematopoietic stem cell transition reveals a requirement for Gpr56 in HSC generation

Citation for published version:

Digital Object Identifier (DOI):
10.1084/jem.20140767

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Journal of Experimental Medicine

Publisher Rights Statement:
This article is distributed under the terms of an Attribution–Noncommercial–Share Alike–No Mirror Sites license for the first six months after the publication date (see http://www.rupress.org/terms). After six months it is available under a Creative Commons License (Attribution–Noncommercial–Share Alike 3.0 Unported license, as described at http://creativecommons.org/licenses/by-nc-sa/3.0/).

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Hematopoietic stem cells (HSCs) are responsible for the life-long maintenance and regeneration of the adult vertebrate blood system. HSCs are generated through a natural transdifferentiation process occurring in specialized embryonic vascular cells, known as hemogenic endothelial cells (ECs [HECs]). In mice, the first adult HSCs are generated in the aorta-gonad-mesonephros (AGM) region at embryonic day (E) 10.5 (Müller et al., 1994; Medvinsky and Dzierzak, 1996). The emergence of the definitive hematopoietic system in the mouse embryo correlates with the temporal appearance of clusters of hematopoietic cells (HCs) associated with the aortic endothelium and the major arteries (Garcia-Porrero et al., 1995; North et al., 1999; de Bruijn et al., 2000). Chick embryo dye-marking studies were the first to show that aortic ECs give rise to HCs (Jaffredo et al., 1998). In mammalian embryos, the results of phenotypic and genetic studies, supported by stringent in vivo transplantation studies of enriched cell fractions, demonstrate that HSCs are derived from vascular ECs during a short window of developmental time (de Bruijn et al., 2002; North et al., 2002; Zovein et al., 2008; Chen et al., 2009). This developmental process is known as endothelial to hematopoietic cell transition (EHT).

To facilitate the study of HSC emergence in the mouse embryo, numerous markers have been used individually and/or in combination to identify HSCs and their direct precursors. Immunolocalization of these markers in the whole-transcriptome analysis of endothelial to hematopoietic stem cell transition reveals a requirement for Gpr56 in HSC generation

Parham Solaimani Kartalaei,1* Tomoko Yamada-Inagawa,1* Chris S. Vink,1 Emma de Pater,1 Reinier van der Linden,1 Jonathon Marks-Bluth,4 Anthon van der Sloot,2 Mirjam van den Hout,2 Tomomasa Yokomizo,5 M. Lucila van Schaick-Solernó,3 Ruud Delwel,3 John E. Pimanda,4 Wilfred F.J. van IJken,2 and Elaine Dzierzak1

1Erasmus MC Stem Cell Institute, Department of Cell Biology, 2Center for Biomics, and 3Department of Hematology, Erasmus University Medical Center, 3000 CA Rotterdam, Netherlands
4Lowy Cancer Research Centre and Prince of Wales Clinical School, University of New South Wales, Sydney, New South Wales 2052, Australia
5Cancer Science Institute of Singapore, National University of Singapore, Singapore 117599

Hematopoietic stem cells (HSCs) are generated via a natural transdifferentiation process known as endothelial to hematopoietic cell transition (EHT). Because of small numbers of embryonal arterial cells undergoing EHT and the paucity of markers to enrich for hemogenic endothelial cells (ECs [HECs]), the genetic program driving HSC emergence is largely unknown. Here, we use a highly sensitive RNAseq method to examine the whole transcriptome of small numbers of enriched aortic HSCs, HECs, and ECs. Gpr56, a G-coupled protein receptor, is one of the most highly up-regulated of the 530 differentially expressed genes. Also, highly up-regulated are hematopoietic transcription factors, including the “heptad” complex of factors. We show that Gpr56 (mouse and human) is a target of the heptad complex and is required for hematopoietic cluster formation during EHT. Our results identify the processes and regulators involved in EHT and reveal the surprising requirement for Gpr56 in generating the first HSCs.

© 2015 Solaimani Kartalaei et al. This article is distributed under the terms of an Attribution–Noncommercial–Share Alike–No Mirror Sites license for the first six months after the publication date (see http://www.rupress.org/terms). After six months it is available under a Creative Commons License Attribution–Noncommercial–Share Alike 3.0 Unported license, as described at http://creativecommons.org/licenses/by-nc-sa/3.0/.

1Erasmus MC Stem Cell Institute, Department of Cell Biology, 2Center for Biomics, and 3Department of Hematology, Erasmus University Medical Center, 3000 CA Rotterdam, Netherlands
4Lowy Cancer Research Centre and Prince of Wales Clinical School, University of New South Wales, Sydney, New South Wales 2052, Australia
5Cancer Science Institute of Singapore, National University of Singapore, Singapore 117599

Abbreviations used: AGM, aorta-gonad-mesonephros; BCV, biological coefficient of variation; ChIP, chromatin immunoprecipitation; DEG, differentially expressed gene; EC, endothelial cell; EHT, endothelial to hematopoietic cell transition; FDR, false discovery rate; HC, hematopoietic cell; HEC, hemogenic EC; hpf, hour post fertilization; HSC, hematopoietic stem cell; ISH, in situ hybridization; MO, morpholino oligo; qPCR, quantitative PCR; TF, transcription factor.

*P. Solaimani Kartalaei and T. Yamada-Inagawa contributed equally to this paper.

E. Dzierzak’s present address is University of Edinburgh/ Medical Research Council Centre for Inflammation Research, Queens Medical Research Institute, Edinburgh EH16 4TJ, Scotland, UK.

Correspondence
Elaine Dzierzak:
e.dzierzak@erasmusmc.nl
OR
Elaine.Dzierzak@ed.ac.uk

© 2015 Solaimani Kartalaei et al. This article is distributed under the terms of an Attribution–Noncommercial–Share Alike–No Mirror Sites license for the first six months after the publication date (see http://www.rupress.org/terms). After six months it is available under a Creative Commons License Attribution–Noncommercial–Share Alike 3.0 Unported license, as described at http://creativecommons.org/licenses/by-nc-sa/3.0/.

Cite by DOI: 10.1084/jem.20140767

Published December 29, 2014

The Rockefeller University Press $30.00
J. Exp. Med. 2015
www.jem.org/cgi/doi/10.1084/jem.20140767
AGM highlighted the heterogeneous nature of the cells in the hematopoietic clusters (Ody et al., 1999; Taoudi et al., 2005; Yokomizo and Dzierzak, 2010; Robin et al., 2011). Whereas combinations of these markers allow HSC enrichment, so far no combination of endothelial and/or hematopoietic markers has been able to distinguish hemogenic from nonhemogenic aortic ECs.

The Ly6aGFP (Sca1) mouse model, in which all HSCs throughout development are GFP+ (de Bruijn et al., 2002; Ma et al., 2002), has facilitated the study of EHT. Clear proof of EHT was obtained by real-time imaging of the mouse Ly6aGFP embryonic aorta (Boisset et al., 2010). In the E10.5 aorta, at the time when the number of hematopoietic clusters peak (Yokomizo and Dzierzak, 2010), flat endothelial GFP+ cells were observed to transition to morphologically round GFP+ cells that begin to express other HSC markers (Boisset et al., 2010). Real-time imaging of transgenic zebrafish embryos similarly revealed the transition of aortic ECs to HCs (Bertrand et al., 2010; Kissa and Herbomel, 2010), indicating that EHT is an evolutionarily conserved process by which the definitive hematopoietic system of vertebrates is generated.

To specifically understand the molecular program involved in EHT, we set out in this study to identify key genes and processes that are functionally relevant in mouse aortic HECs as they transit to HSCs. Based on the vital imaging of EHT, the Ly6aGFP reporter is currently the most tractable marker to distinguish and enrich the HSCs that are undergoing EHT from other aortic ECs, and also the emerging HSCs from other HCs. Here we present RNA sequencing data obtained from highly enriched small numbers of relevant EHT cells from Ly6aGFP embryos, aortic ECs, HECs, and emerging HSCs. Among the few (530) differentially expressed genes (DEGs) during EHT, Gpr56 is the highest up-regulated gene encoding a cell surface receptor. We show for the first time the functional involvement of Gpr56 in HSC emergence during EHT. In addition, the previously described “heptad” transcription factors (TFs; Wilson et al., 2010) are up-regulated during EHT; bind the Gpr56 enhancer, and regulate its expression. This unique dataset expands our understanding of EHT, identifying the gene networks and processes that are essential for HSC generation in the embryo.

RESULTS
Temporal-spatial and transcriptomic quantitation of aortic hemogenic endothelial and emerging HCs
Ly6aGFP expression marks HCs emerging from hemogenic endothelium at the time of HSC generation in the midgestation mouse aorta. To quantify and localize these cells, we performed confocal imaging of whole and sectioned immunostained E10 Ly6aGFP embryos (Fig. 1, A–D). CD31 marks all ECs and HCs, and cKit marks all HCs. However, Ly6aGFP marks only some ECs and some HCs. High-resolution imaging of transverse sections allowed quantitation of four different Ly6aGFP-expressing aortic cell types (Fig. 1 D): flat ECs, bulging cells in the single layer of endothelium, and two differently positioned round cells within the clusters distinguished by the close attachment to (juxtaposed) or a position distal from the endothelium. The total number of GFP+ cells increased from 287 at early E10 (32 sp [somite pairs]) to 1,592 at late E10 (37 sp; Fig. 1 D). From the small fraction of ECs that express GFP (range 13–19%), most aortic GFP+ cells are flat ECs with only 8% of GFP+ cells in hematopoietic clusters. Although by a random distribution more GFP+ cluster cells would be expected in distal positions (as compared with juxtaposed), we observed 70–88% of the GFP+ HCs localized in a juxtaposed position, most likely because GFP+ HCs are emerging from GFP+ ECs and/or are actively maintained at the juxtaposed position.

Because all HSCs are GFP+ and rare HCs have been observed to emerge from GFP+ ECs, Ly6aGFP is the best marker for high enrichment of HECs. Hence, we developed an enrichment method using the Ly6aGFP, cKit, and CD31 markers: ECs (CD31+cKit−GFP−), HECs (CD31+cKit−GFP+), HSCs (CD31+cKit−GFP+), and progenitor/differentiated HCs (CD31−cKit−GFP−; Fig. 1 B). The distinct cell types were sorted (Fig. 1 E) and hematopoietic function was assessed. Hematopoietic progenitors were found in the HC (64%) and HSC (33%) fractions as expected, with a majority of the immature progenitors (CFU−GM/E) in the HSC fraction (Fig. 1 F). In vivo transplantation assays revealed that only the CD31+cKit−GFP+ fraction contained HSCs (Fig. 1 G). These HSCs provided long-term high-level multilineage repopulation of adult irradiated recipients (Fig. 1 A; 6 engrafted of 10 injected with 1–5 ee [embryo equivalents]). Despite injection of high embryo equivalents of cells from the other fractions (4–9 ee), no repopulation was found with the ECs, HECs, or HCs.

RNA sequencing and validation
Sorted ECs, HECs, HSCs, and HCs from three independent biological replicates were used for RNA sequencing. As few as 4–14 E10.5 AGM equivalents (34–41 sp) of sorted cells per replicate were obtained, and cDNA was made from as few as 593 sorted cells (see Table S1 for details). The sequence reads of EHT cell fractions were mapped to the mouse genome (NCBI build 37/mm9), and the generated count table (with 7–57 million unique mapped reads to exons per sample) was normalized and analyzed by edgeR (Fig. 2 A; McCarthy et al., 2012).

To confirm that the transcriptome analysis was representative of the sorted EHT fractions, we measured the normalized number of fragments (in FPKMs [fragments per kilobase exon reads per million fragments mapped]) of CD31, cKit, and Ly6a (Fig. 1 H). As expected, CD31 transcripts were found in all four subsets (ECs, HECs, HSCs, and HCs), cKit transcripts were found only in HCs and HSCs, and Ly6a transcripts were found in HECs and HSCs.

Gene transcript Reads for endothelial genes Cdh5, Tek, Esam, Kdr, and Eng were highest in HECs as compared with ECs and were higher in HSCs than in ECs or HCs. When the four cell fractions were examined by FACS (Fig. 1 I), cell surface expression correlated significantly with transcript levels (r² = 0.54, P = 0.01). Thus, our datasets reflect a dynamic transcriptional program during EHT.
Global transcriptional differences between the EHT cell subsets

Biological coefficient of variation (BCV) analysis indicates (Fig. 2 B) that EC, HEC, and HSC fractions are closely related but distinct. EC, HEC, and HSC replicate 2 and 3 samples cluster together, whereas replicate 1 EC, HEC and HSC samples show a similar BCV pattern but are further dispersed in the plot. The tighter sample dispersion of replicates 2 and 3 is most likely the result of the higher sequencing depth (Table S1). Hence, distinct transcriptional variation between the EHT fractions is consistent for the three biological replicates.

Dataset comparisons showed a total of 530 DEGs (false discovery rate [FDR] < 0.05; Fig. 2 C and Table S2). The EC to HEC comparison shows 139 DEGs, whereas 340 genes were differentially expressed between HECs and HSCs. Moreover, comparison of ECs with HSCs identified 108 additional genes. MA plots of differential expression analysis show most genes being centered around zero, further confirming the correct normalization of datasets (Fig. S2 A). In the EC to HEC comparison, most DEGs are up-regulated, whereas a majority of the DEGs in HEC to HSC are down-regulated. Hierarchical clustering of DEGs grouped the three biological replicates of each fraction together, suggesting that ECs, HECs, and HSCs have recognizably distinct genetic programs (Fig. 2 D).

Transcriptome analysis reveals processes involved in EHT DEGs were grouped based on their relative expression levels (H, high; I, intermediate; and L, low expression) into representative patterns for EC genes (HIL, HLL, HHL, and HLI), HEC genes (LHL, IHL, IHI, and LHI), and HSC genes (LHH, LIH, LHIL, and LLH; Fig. 3 and Table S2). Each group was used as input for Gene Ontology (GO), KEGG, and Wiki-Pathways enrichment analysis, and Gene Set Enrichment Analysis (GSEA) was used to detect global shifts of gene sets during each transition.

EC genes show overrepresentation of “focal adhesion,” “ECM-receptor interaction,” “protein digestion and absorption,” “oxidative stress,” and “chemokine signaling” terms (Fig. 3 A and Table S3), consistent with EC function (Rajendran et al., 2013). Significant enrichment of “inflammatory response”...
HECs as compared with ECs, whereas no significant enrichment of these genes was found in HSCs as compared with HECs, indicating that the hematopoietic program is already activated in HECs (Fig. 4 A).

HSC genes showed clear overrepresentation of “hematopoietic processes,” “cell cycle,” and “histone methylation” related genes (Fig. 3 C). Significant enrichment of “cell cycle progression,” “DNA replication,” and “hematopoietic progenitor” sets was also detected by GSEA (Fig. 4 A and Fig. S2 C). “Hematopoietic progenitor” gene sets are enriched in HSCs as compared with HECs, and detection of “acute myeloid leukemia” from the KEGG database and “pluripotency network” from WikiPathways is in agreement with the acquisition of hematopoietic fate and self-renewal capacity in HSCs (Table S3). This is further supported by significant enrichment of gene sets characteristic of stem cells, such as “telomere lengthening” and “DNA repair,” in the HSC fraction by GSEA (Fig. 4 A; Yui et al., 1998; Rossi et al., 2005).

TF expression in cells undergoing EHT

The genetic program directing cell identity is coordinated by TFs, and thus we focused our attention on these genes in our EHT datasets. As compared with ECs, significant up-regulated expression of Mecom, Notch1 and 4, Gfi1, Sox17, Ets2, and Elk3 was found in HECs (Table 1), with Sox7, Sox18, Runx1, Hhex, and Lmo2 among the top up-regulated HEC TFs (Table S4).
In the HSC fraction, many TFs with known roles in HSC development, including Etv6, Gfi1, Gfi1b, Myb, Mlyc, Hlf, Meis1, Hhex, Runx1, Mpl, and Ikzf1 and 2 (Table 1 and Table S4), were found to be significantly up-regulated as compared with ECs or HECs. We identified several novel TFs not previously reported to be involved in embryonic HSC generation such as zinc-finger proteins Zfp106, Zfp445, Zfp748, and Zfp763; megakaryocyte factor Nfk2, transcriptional co-repressor Bior, and Chba2r3 (Eto2). Also present in the top hits were chromatin-remodeling factors Siz12, Paxip1, Kdm5a, Smarca4 (Brg1), Ezh1, Bpf1, and Hdac1 and de novo DNA methylation genes Dnmt3a/Dnmt3b and Dnmt1. The down-regulation of several Hox, Tbx, and Fox genes was observed in the EC to HEC and HEC to HSC transition, whereas only Hexa9, Hhex, and Foxe1 were up-regulated in HSCs as compared with ECs or HECs (Table S4).

It has been shown that a pivotal (heptad) group of TFs work together in transcriptional regulatory complexes to regulate the expression of downstream target genes in hematopoietic progenitor cell lines (Wilson et al., 2010). The heptad TFs could act as one of the transcriptional hubs for the regulation of EHT. Our RNAseq datasets reveal that all heptad TFs increase during EHT (Fig. 5 A). To identify genes encoding novel EHT and embryonic HSC surface markers, the 530 DEGs were compared with the 927 heptad TF targets identified by chromatin immunoprecipitation (ChIP)seq analyses in HPC7 cells. 58 DEGs were found to be targets of heptad TFs, with CD34 and Gpr56 as the top hits (Fig. 5, B and C). Interestingly, also in the whole transcriptome analysis of EHT, Gpr56 was identified as the top differentially expressed receptor gene in the HEC to HSC transition, followed by cKit (Fig. 4 B). Because both CD34 and cKit function has been studied in HSCs and these markers are used extensively for HSC isolation (Sánchez et al., 1996), we focused on Gpr56.

The Gpr56 heptad consensus region in the mouse is located 37 kb upstream of the translational start site. We identified this region as the Gpr56-37 enhancer. Enrichment of heptad factors at the Gpr56-37 element was found in mouse HPC7 cell line by quantitative PCR (qPCR; Fig. 5 D). Transactivation assays in hematopoietic progenitor cell lines showed significant activation of Gpr56-37 enhancer, whereas overexpression of three of the heptad factors (Gata2, Runx1, and Fli1)
showed synergistic activation of the \textit{Gpr56}-37 enhancer (Fig. 5, E and F). Moreover, we identified a homologous element 48 kb upstream of the human \textit{GPR56} gene. In human CD34+ HSC-enriched cells, we found binding of all seven heptad TFs to the human \textit{GPR56-48} element (Fig. 5 G; Chacon et al., 2014). These data suggest that the heptad TFs and their downstream target \textit{Gpr56} are important in HSC generation during EHT, as well as in healthy and leukemic human HCs. Because nothing is known concerning Gpr56 in embryonic hematopoietic development, we examined its regulation and role during EHT.

\textbf{Gpr56 is required during EHT for HSC generation}

To confirm localized expression of \textit{Gpr56} in cells undergoing EHT, we performed in situ hybridization (ISH) analysis of the E10.5 AGM. High-level \textit{Gpr56} expression was observed in some aortic HCs (Fig. 6 A), and no/low expression was observed in aortic ECs. An overlap of \textit{Gpr56} expression with some GFP+ HCs was found by ISH of \textit{Ly6aGFP} E10.5 AGM sections (Fig. 6 B). Thus, the localized expression of \textit{Gpr56} is consistent with FPKM values derived from RNAseq datasets and strongly suggests a role in HSC generation.

\textit{Gpr56} is highly conserved across different vertebrate species (mean multiple sequencing alignment score = 85%; Fig. 6 C). To validate the involvement of \textit{Gpr56} in HSC generation, we used a zebrafish morpholino oligo (MO) knockdown approach. At 30 and 48 h post fertilization (hpf), morphants were assayed by ISH (30 hpf) for \textit{cmyb}, a marker for emerging HSCs (Jing and Zon, 2011). WT embryos show \textit{cmyb}-expressing cells along the aorta (Fig. 6 D). In contrast, \textit{cmyb}-expressing cells are severely reduced in \textit{gpr56} MO–injected embryos. To validate that this was a defect in HSC generation, we injected the \textit{gpr56} MO into CD41-GFP transgenic embryos (CD41 marks HCs; Lin et al., 2005; Jing and Zon, 2011; Robin et al., 2011). The number of CD41-GFP+ cells in the caudal hematopoietic tissue at 48 hpf is significantly decreased from 36.7 ± 4.0 cells in WT to 12.5 ± 1.8 in \textit{gpr56} morphants (Fig. 6 E), suggesting that \textit{Gpr56} is important for the emergence of HSCs. No abnormalities in embryo growth or the structure of the vasculature/aorta were found by ISH for arterial endothelial marker \textit{gridlock} (grl; Fig. 6 F; Zhong et al., 2000). To test whether the \textit{gpr56} MO does not show an off-target effect, we performed rescue experiments by injecting \textit{gpr56} mRNA. \textit{Gpr56} morphants could be rescued with zebrafish
Table 1. Differentially expressed TFs

<table>
<thead>
<tr>
<th>Gene</th>
<th>HECs versus ECs</th>
<th>HSCs versus HECs</th>
<th>HSCs versus ECs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>logFC</td>
<td>FDR</td>
<td>logFC</td>
</tr>
<tr>
<td>(A) Top 25 up-regulated genes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EK3</td>
<td>2.42</td>
<td>1.1E-02</td>
<td>4.23</td>
</tr>
<tr>
<td>Mecon</td>
<td>3.21</td>
<td>1.4E-02</td>
<td>5.48</td>
</tr>
<tr>
<td>Notch4</td>
<td>5.41</td>
<td>2.9E-02</td>
<td>7.03</td>
</tr>
<tr>
<td>Notch1</td>
<td>2.16</td>
<td>4.2E-02</td>
<td>2.20</td>
</tr>
<tr>
<td>Rab11a</td>
<td>2.03</td>
<td>5.2E-02</td>
<td>2.43</td>
</tr>
<tr>
<td>Gfi1</td>
<td>7.69</td>
<td>6.8E-02</td>
<td>2.05</td>
</tr>
<tr>
<td>Vwtr1</td>
<td>3.20</td>
<td>6.8E-02</td>
<td>2.35</td>
</tr>
<tr>
<td>Junb</td>
<td>4.90</td>
<td>6.8E-02</td>
<td>2.51</td>
</tr>
<tr>
<td>Ets2</td>
<td>2.50</td>
<td>6.9E-02</td>
<td>2.05</td>
</tr>
<tr>
<td>Mapk3</td>
<td>1.62</td>
<td>7.3E-02</td>
<td>5.05</td>
</tr>
<tr>
<td>Nkx2-3</td>
<td>5.95</td>
<td>7.4E-02</td>
<td>4.29</td>
</tr>
<tr>
<td>Ldb2</td>
<td>2.30</td>
<td>7.5E-02</td>
<td>1.94</td>
</tr>
<tr>
<td>Hdac7</td>
<td>2.20</td>
<td>7.6E-02</td>
<td>1.55</td>
</tr>
<tr>
<td>Sox17</td>
<td>5.14</td>
<td>8.4E-02</td>
<td>2.22</td>
</tr>
<tr>
<td>Chmnb1</td>
<td>1.28</td>
<td>9.5E-02</td>
<td>2.64</td>
</tr>
<tr>
<td>Hey1</td>
<td>4.24</td>
<td>1.1E-01</td>
<td>1.99</td>
</tr>
<tr>
<td>Epo1</td>
<td>3.10</td>
<td>1.2E-01</td>
<td>8.87</td>
</tr>
<tr>
<td>Hey2</td>
<td>4.35</td>
<td>1.2E-01</td>
<td>1.93</td>
</tr>
<tr>
<td>Sox7</td>
<td>4.27</td>
<td>1.5E-01</td>
<td>1.28</td>
</tr>
<tr>
<td>Sox8</td>
<td>4.05</td>
<td>1.5E-01</td>
<td>2.68</td>
</tr>
<tr>
<td>Tsc2d1</td>
<td>1.46</td>
<td>1.6E-01</td>
<td>2.41</td>
</tr>
<tr>
<td>Nrap</td>
<td>3.76</td>
<td>1.7E-01</td>
<td>1.42</td>
</tr>
<tr>
<td>Nl6c</td>
<td>2.07</td>
<td>1.8E-01</td>
<td>1.29</td>
</tr>
<tr>
<td>Palim1</td>
<td>1.96</td>
<td>2.1E-01</td>
<td>1.84</td>
</tr>
<tr>
<td>Hmg20h</td>
<td>2.46</td>
<td>2.1E-01</td>
<td>1.45</td>
</tr>
<tr>
<td>(B) Top 25 down-regulated genes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pou5f1</td>
<td>-7.70</td>
<td>2.6E-04</td>
<td>-7.39</td>
</tr>
<tr>
<td>Utl1</td>
<td>-8.22</td>
<td>1.0E-02</td>
<td>-3.87</td>
</tr>
<tr>
<td>Gfi1b</td>
<td>-3.77</td>
<td>1.3E-02</td>
<td>-7.55</td>
</tr>
<tr>
<td>Foxd1</td>
<td>-10.54</td>
<td>6.6E-02</td>
<td>-6.49</td>
</tr>
<tr>
<td>Lmo1</td>
<td>-10.20</td>
<td>6.7E-02</td>
<td>-8.54</td>
</tr>
<tr>
<td>Hand1</td>
<td>-6.55</td>
<td>7.4E-02</td>
<td>-6.56</td>
</tr>
<tr>
<td>Zfp124</td>
<td>-7.77</td>
<td>1.0E-01</td>
<td>-2.60</td>
</tr>
<tr>
<td>Prx2</td>
<td>-7.19</td>
<td>1.0E-01</td>
<td>-4.08</td>
</tr>
<tr>
<td>Asb12</td>
<td>-9.04</td>
<td>1.1E-01</td>
<td>-7.43</td>
</tr>
<tr>
<td>Clec6</td>
<td>-1.75</td>
<td>1.3E-01</td>
<td>-5.39</td>
</tr>
<tr>
<td>Ncoa4</td>
<td>-1.48</td>
<td>1.4E-01</td>
<td>-4.69</td>
</tr>
<tr>
<td>Krf1</td>
<td>-1.47</td>
<td>1.4E-01</td>
<td>-3.42</td>
</tr>
<tr>
<td>Alx4</td>
<td>-5.93</td>
<td>1.5E-01</td>
<td>-3.81</td>
</tr>
<tr>
<td>Wt1</td>
<td>-4.76</td>
<td>2.0E-01</td>
<td>-7.63</td>
</tr>
<tr>
<td>Rhox2c</td>
<td>-3.29</td>
<td>2.0E-01</td>
<td>-5.00</td>
</tr>
<tr>
<td>Pox5</td>
<td>-8.85</td>
<td>2.1E-01</td>
<td>-4.59</td>
</tr>
<tr>
<td>Rhox6</td>
<td>-3.59</td>
<td>2.2E-01</td>
<td>-1.79</td>
</tr>
<tr>
<td>Neurod6</td>
<td>-6.02</td>
<td>2.2E-01</td>
<td>-2.93</td>
</tr>
<tr>
<td>Hoxc4</td>
<td>-8.73</td>
<td>2.3E-01</td>
<td>-4.21</td>
</tr>
<tr>
<td>Rhox9</td>
<td>-3.68</td>
<td>2.3E-01</td>
<td>-6.02</td>
</tr>
<tr>
<td>Six1</td>
<td>-7.30</td>
<td>2.3E-01</td>
<td>-6.86</td>
</tr>
</tbody>
</table>

(A and B) Top 25 up-regulated TFs (A) and top 25 down-regulated TFs (B) in HEC versus EC, HSC versus HEC, and HSC versus EC comparisons. FDR, FDR corrected p-value; logFC, log fold change. All genes with FDR < 0.05 except genes with underlining.
To further study the function of Gpr56 in HSCs, we used the 32D-CSF3R cell line, a unipotent mouse stem cell differentiation model in which colony-stimulating factor-3 stimulates their differentiation to neutrophils. (Fig. 6 H). When stimulated with CSF3, cells transduced with an empty vector or WT human GPR56 vector lost their blast characteristics and differentiated. However, cells transduced with constitutively active (MUT) human GPR56 resulted in an increase in blast-like cells and in significantly fewer differentiated gpr56 mRNA, as well as with a mouse Gpr56 mRNA to yield full restoration of aortic cmyb expression (Fig. 6 G). Interestingly, some ectopic expression of cmyb is observed in the region ventral to the aorta. These data indicate that gpr56 is an essential player in the HSC generation program and that its functional domains are maintained between mouse and zebrafish. Ectopic generation of phenotypic HSCs in zebrafish by Gpr56 overexpression further highlights the unexpected function of this molecule in induction of HSC generation.

### Table 1. (Continued)

<table>
<thead>
<tr>
<th>Gene</th>
<th>HECs versus ECs</th>
<th>HSCs versus HECs</th>
<th>HSCs versus ECs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tcf21</td>
<td>-5.63</td>
<td>2.3E-01</td>
<td>Ankrd1</td>
</tr>
<tr>
<td>Runx11</td>
<td>-1.79</td>
<td>2.4E-01</td>
<td>Onecut3</td>
</tr>
<tr>
<td>Klf1</td>
<td>-6.56</td>
<td>2.6E-01</td>
<td>Hivep3</td>
</tr>
<tr>
<td>Aff3</td>
<td>-3.18</td>
<td>3.1E-01</td>
<td>Zfp3611</td>
</tr>
<tr>
<td>Tcf21</td>
<td>-5.63</td>
<td>2.3E-01</td>
<td>Ankrd1</td>
</tr>
<tr>
<td>Runx11</td>
<td>-1.79</td>
<td>2.4E-01</td>
<td>Onecut3</td>
</tr>
<tr>
<td>Klf1</td>
<td>-6.56</td>
<td>2.6E-01</td>
<td>Hivep3</td>
</tr>
<tr>
<td>Aff3</td>
<td>-3.18</td>
<td>3.1E-01</td>
<td>Zfp3611</td>
</tr>
</tbody>
</table>

(A and B) Top 25 up-regulated TFs (A) and top 25 down-regulated TFs (B) in HEC versus EC, HSC versus HEC, and HSC versus EC comparisons. FDR, FDR corrected p-value; logFC, log fold change. All genes with FDR < 0.05 except genes with underlining.

![Figure 5. Gpr56 is a heptad target in mouse and human blood progenitors.](image-url)
cells (eight- to fourfold fewer), suggesting that GPR56 is also essential for the maintenance of an undifferentiated cell state.

DISCUSSION

RNA sequencing analyses of EHT and developing HSCs provide a new perspective on the molecules and processes. Whereas previous methods of transcriptome analysis have identified many of the obvious regulators of hematopoietic development, this method provides an accurate accounting of all expressed genes and also small gene expression level changes between the rare, relevant cell types. We have shown here that HECs (the precursors of the earliest emerging HSCs in the midgestation mouse aorta), as distinguished from other ECs by Ly6aGFP expression, have closely related but distinct transcriptional programs. Comparisons between ECs and HECs reveal a developing program indicative of cell migration and changing cell morphology in HECs, while they retain an angiogenic program. The up-regulation of the HSC hematopoiesis program begins in HECs, further distinguishing them from ECs. Of the 530 DEGs, important receptors and TF genes were identified, including Gpr56, which is required for HSC formation. Moreover, the heptad (hematopoietic) TFs were found to be up-regulated during EHT. These factors bind the Gpr56 enhancer and regulate its expression, thus providing a proof of principle for in silico bioinformatical predictions of the combinatory role of the heptad TFs in the emergence of HSCs during EHT. Thus, our datasets are predictive of functionally relevant EHT genes and processes.

RNAseq analysis of small numbers of physiologically relevant cells

The Ly6aGFP transgenic marker in combination with CD31 and c-Kit cell surface markers allowed the high enrichment of HSCs, HECs, and ECs isolated from the aorta at the developmental time when HSCs begin to emerge. Imaging experiments verified the correlation between these markers, the expected cell type, and localization within the AGM region.
Moreover, we confirmed that functional adult-repopulating HSCs are exclusively contained within the HSC fraction (0.002% of AGM cells) and that both endothelial fractions (ECs and HECs) do not contain hematopoietic progenitors or HSCs. Thus, the Ly6aGFP marker currently allows the highest level of enrichment for HECs (Fig. S1 B) that will undergo transition to HSCs, as compared with previously used markers.

Previous comparative HSC gene expression profiling (microarray) studies identified several new regulators of AGM HSCs, but the genetic program of HECs was not examined (Mascarenhas et al., 2009; McKinney-Freeman et al., 2012). During preparation of this manuscript, a new microarray study of EHT-relevant populations was performed based on cells expressing the Runx1+23-enhancer marker (Swiers et al., 2013). Runx1+23GFP marks 68% of VE-cadherin+ (endothelial and hematopoietic) cells at E8.5, marking many HECs that are not yet exhibiting EHT. However, Ly6aGFP expression marks only a small fraction (13–19%) of CD31+ aortic cells and is probably more specifically marking the active HECs at E10.5.

Given the limited number of cells in our enriched aortic EC, HEC, and emerging HSC fractions, RNA sequencing provides the most efficient and sensitive method for analysis of EHT-relevant cells. Only 4–14 embryos (aortas) were used per sequencing experiment to isolate sufficient quantities of total RNA from sorted cell populations. With as few as 593 sorted cells, we successfully applied RNAseq technology with the SMARTER protocol, recently shown to be the best RNAseq method for low numbers of cells (Bhargava et al., 2014). Additionally, RNAseq analysis has the great advantage over microarrays in not only providing the whole transcriptome, but also revealing isoform-specific transcripts in the sequenced samples. For example, Gpr56 expresses two transcript variants. We found variant 1 (GenBank accession no. NM_001198894) of Gpr56 to be expressed exclusively in HSCs, whereas variant 2 (GenBank accession no. NM_001198882) was expressed in ECs, HECs, and HSCs. Variant 2 was more highly represented in HSCs (FPKM = 323) as compared with variant 1 (FPKM = 52).

Identification of processes involved in EHT

For the first time, datasets from aortic ECs, HECs, and HSCs provide an overview of the general processes involved during EHT. Quantitative levels of gene expression between EHT-enriched cell fractions show only a small number of significant DEGs: 139 between HECs versus ECs and 340 between HSCs versus HECs. Not surprisingly, the genes with high expression in midgestation aortic ECs are mainly those involved in "general developmental processes." These and other GO categories related to cell migration and focal adhesion are highly represented in HECs, highlighting the fact that HECs must change their adhesive properties to bulge out of endothelial lining of the aorta, undergo morphological changes as they become HCs, adhere to other HCs within the clusters, and take on hematopoietic identity and function. GO analysis of DEGs with the highest expression in HSCs shows enrichment of "hematopoiesis" and "positive regulation of histone methylation" terms.

HECs are a transcriptionally dynamic cell type at the interface of EHT. Concurrent with the initiation of the hematopoietic program and HC formation, hematopoietic genes become activated in HECs, whereas endothelium-specific cell adhesion molecules and TFs are down-regulated in HSCs. Our RNAseq data are in agreement with the recent single-cell high-throughput qPCR analysis results for 18 known endothelial and hematopoietic genes during EHT (Fig. S2 D; Swiers et al., 2013). We also identified several genes involved in angiogenesis by selection for GO term "receptor activity" in HEC versus EC comparisons (Fig. 4 A). These include Plxn1, Eltd1, Calcr, Ramp2, and S1pr1; Plxn1 and Eltd1 are both induced by VEGF (Kim et al., 2011; Masiero et al., 2013). Calcr, a GPCR, induces angiogenesis upon association with Ramp2 and Kdr/Vegfr-2, both of which are significantly induced in HECs (Guidolin et al., 2008). Collectively, these findings suggest a role for angiogenesis-related receptors in activation of hematopoietic potential and generation of HECs.

GPR56: a novel EHT regulator

Gpr56, one of the top hits in our HSC versus HEC comparison (30-fold increase) and bound by all heptad TFs, is indeed a novel regulator for emerging HSCs in the embryonic vasculature. Contrary to expectations raised by the lack of HSC defects in mouse Gpr56 KO embryos (generated by deletion of the first two exons [Saito et al. 2013]), our RNAseq data suggested a strong role for Gpr56 in emergence of HSCs. In the E10.5 mouse aorta, we localized Gpr56 expression to a few HCs/HSCs (Ly6aGFP+). Upon gpr56 knockdown, zebrafish embryos showed severe reduction in HSCs (antib) and CD41+ hematopoietic stem/progenitor cells, revealing a requirement for Gpr56 in HSC generation. Our rescue experiments in gpr56 morphants show that both zebrafish and mouse Gpr56 RNA can restore aortic hematopoietic stem/progenitor generation. Moreover, Gpr56 overexpression resulted in ectopic hematopoietic progenitor/stem cell formation in the axial vein, suggesting that the Gpr56 signaling axis may be useful for inducing new HSCs.

We propose that the lack of embryonic lethality in Gpr56 KO embryos could be the result of redundancy by other GPCRs or residual Gpr56 activity in the mouse transgenic model. Our RNAseq and RT-qPCR validation (Fig. S2 E) data show an increase in the expression of Gpr114 (77 kb upstream of Gpr56) and Gpr97 (48 kb downstream) during EHT. The ligand binding N-terminal part of Gpr114 has 47% amino acid similarity (and 27% identity) with Gpr56. Gpr114 is present only in mammals. Also, assays testing a human Gpr56 variant missing a large part of the second exon and the complete third exon showed that it partially retains the ability to activate SRE, E2F, NFAT, and iNOS promoters (Kim et al., 2010). Thus, Gpr56 is an unexpected novel EHT regulator essential for HSC generation and maintenance, and its function is conserved between mouse and zebrafish.
How Gpr56 acts in HECs as they transdifferentiate to HSCs is unknown, but it could affect physical properties such as adherence, cluster formation, signal transduction, migration, and/or self-renewal. Some of these features are consistent with findings in neuronal stem cells, BM HSCs, and leukemic cells, in which it has been proposed that Gpr56 functions in cell adhesion, migration, and/or repression of apoptosis (Iguchi et al., 2008; Saito et al., 2013). We found that Gpr56 functions in the maintenance of the undifferentiated state of a unipotential HSC line. The conservation of Gpr56 across species will allow for future high-throughput study of the mechanism by which Gpr56 affects EHT and generation of HSCs.

Our results on heptad TF binding to the Gpr56 enhancer suggest that other heptad targets in the overlapping list are likely to be relevant in EHT. However, not all genes that we identified as highly up-regulated during EHT are targets of the heptad complex, for example cKit. Because EHT regulation is likely to be multilayered, we are using our whole transcriptome dataset as a resource to identify other candidate transcriptional hubs.

In summary, novel and known EC, HEC, and HSC genes were identified in our RNAseq datasets. These comparative quantitative data have high predictive value for identifying functionally important molecules that direct the cellular processes involved in EHT and could instruct methods for de novo HSC generation either by direct somatic cell conversion or pluripotent stem cell differentiation.

MATERIALS AND METHODS

Cell preparation and flow cytometry. Ly6aGFP and WT mouse embryos were dissected as described previously (Robin and Dzierzak 2010), and single cells were prepared by collagenase treatment (0.125%, 45 min, 37°C) and washed with PBS, 10% heat-inactivated FCS, and 1% penicillin/streptomycin (PS). Cells were stained with R–Alexa Fluor 488 (1:1,000; Invitrogen), and imaged on an SP5 confocal microscope (Leica). Cy5 (1:500; Jackson ImmunoResearch Laboratories, Inc.), and imaged for each gene across all samples were normalized by division with maximum FPKM observed for that gene. Patterns were generated based on normalized transcript abundance estimates in FPKMs (Trapnell et al., 2013). For DEGs, the FPKMs for each gene across all samples were normalized by division with maximum FPKM observed for that gene. Patterns were generated based on normalized FPKM, with expression levels lower than 1/3 assigned as low (L), between 1/3 and 2/3 as intermediate (I), and more than 2/3 as high (H). Patterns were then categorized as ECs, HECs, or HSCs. Genes corresponding to EC, HEC and HSC genes were directly sorted into PBS/10% FCS/1% PS and centrifuged, and supernatant was removed. Cells were resuspended in PBS and RNA was isolated using the mirVana miRNA Isolation kit (Ambion) according to the manufacturer’s protocol. RNA quality and quantity were measured by the 2100 Bioanalyzer (Agilent Technologies).

mRNA sequencing analysis. RNA samples (Table S1) were prepared by SMARTer protocol. Illumina TrueSeq v2 protocol was used on Hiseq2000 with single read of 36 bp + 7 bp index. Reads were aligned to the mouse genome (NCBI37/mm9) using TopHat/Bowtie and mapped to the mouse genome (NCBI37/mm9), and the generated count table was analyzed by R/Bioconductor package edgeR according to McCarthy et al. (2012). Counts were normalized for mRNA abundance, and differential expression analysis was performed using edgeR (Fig. 2A). B-H method was used for p-value correction with an FDR of 0.05 as statistically significant. Variance stabilized counts were calculated by R/Bioconductor package “DESeq” for all the genes (Anders and Huber, 2010). Heat maps were generated from the log-scaled variance stabilized counts of DEGs. GSEA was performed using the preranked option in combination with log fold change values of each comparison calculated by edgeR. Cufflinks was used to compute transcript abundance estimates in FPKMs (Trapnell et al., 2013). For DEGs, the FPKMs for each gene across all samples were normalized by division with maximum FPKM observed for that gene. Patterns were then categorized as ECs, HECs, or HSCs. Genes corresponding to EC, HEC and HSC patterns were separately used for GO, KEGG, and Phenotype ontology enrichment analysis using the webgestalt web application (Wang et al., 2013). GO terms were summarized using the REVIGO tool (Supek et al., 2011).
Cell lines. CHOK3 cells transfected with an expression vector for the mouse SCF gene were grown initially in DMEM (Gibco) until they became confluent. They were then grown in Stem Cell Pro media (Gibco) supplemented with 1% FBS, 5% FBS, HPC-7 mouse hematopoietic progenitor cells were grown in IMDM (Gibco) supplemented with 10% CHOK3 conditioned media, 1% FBS, 5% FBS, and 1.5 x 10^-4 M monothio-glycerol. See Knezevic et al. (2011) and Wilson et al. (2010) for details.

ChIP. ChIP assays were performed in HPC-7 cells. 2 x 10^6 cells per antibody were treated with 0.4% formaldehyde, and cross-linked chromatim was sonicated to fragments of 300-500 bp. Cross-linked, sonicated chromatim was distributed evenly for immunoprecipitaiton. SYBR Green RT-PCR was performed on a Stratagene Mx3000p and analyzed using the MxPro software. Relative enrichment levels were calculated by normalizing results to the IgG control. For details see Knezevic et al. (2011). CD34 and HPC7 ChIPseq data were downloaded from the BloodChIP database (Beck et al., 2013; Chacon et al., 2014) and from Wilson et al. (2016).

Luciferase and LacZ assays. Transfection was performed by electroporation of 5-10 x 10^5 cells with 10 µg vector DNA using a GenePuler Excell (Bio-Rad Laboratories). The luciferase assay was performed using a modified version of the Dual-Luciferase Reporter Assay System (Promega). Tran- sent transfections were cotransfected with the pEFBOS-LacZ vector, and luciferase data were normalized to the lacZ data. For stable transfection assays, cells were cotransfected with pGK Neo and resistant cells were used for luc-iferase assays as described in Knezevic et al. (2011).

Transactivation. Coxs7 cells were cultured in 6-well plates (5 x 10^5 cells/well) overnight. Runx1, pDNA3, CBFb, pDNA3, Flt1, pDNA3, and Gata2, pMSCV- PIG, or empty vectors were transfected along with the pEFBOS-LacZ control vector and the Gpr56-37-pGL2 p30 enhancer construct in varying combi-nations (0.5 µg DNA/well) using the ProFection Mammalian Transfection system (Promega). After 48 h, luciferase and lacZ assays were performed as detailed in Knezevic et al. (2011).

Zebrafish. Zebrafish (Danio rerio) embryos were raised at 28.5°C (Westenfeld, 1995). Heterozygous -6.09kbG-EFP embryos (CD41-GFP, Lin et al., 2005) were maintained by crosses with WT zebrafish. For ISH 0.003% formaldehyde, and cross-linked chromatin was fixed on a Stratagene Mx3000p and analyzed using the MxPro software. Relative enrichment levels were calculated by normalizing results to the IgG control. For details see Knezevic et al. (2011). CD34 and HPC7 ChIPseq data were downloaded from the BloodChIP database (Beck et al., 2013; Chacon et al., 2014) and from Wilson et al. (2016).

Online supplemental material. Fig. S1 provides FACS characterization of CD31^+ AGM cells and a representative multilineage repopulation analysis of transplanted mice. Further details of RNAseq analyses and a compari-son with the published single-cell qPCR analysis of EHT cells (Swiers et al., 2013) and adult BM HSCs (Riddell et al. 2014) is provided in Fig. S2. Table S1 contains details of the material used for RNAseq analysis. Table S2 contains the S30 DEGs and their expression patterns and groupings. Ontology enrichment analysis results for each group are provided in Table S3. Table S4 contains a list of up- and down-regulated TFs as found by differential expression analysis with edgeR. All sequencing data has been uploaded to GEO repository under accession number GSE63316. Online supplemental material is available at http://www.jem.org/cgi/content/full/jem.20140767/DC1.

We thank all laboratory members and N. Speck for stimulating discussions, P. Kaimakis for technical assistance, R. Patient, and the Experimental Animal Center of Erasmus MC. This study was supported by a National Institutes of Health grant (R37 DK54047), ZornMW (Netherlands Scientific Research Organization) equipment grant (91109030), ZornMW TOP Award (40-00812-98-11068), NIRM (Netherlands Institute for Regenerative Medicine) FES award, Erasmus MC PhD grant, and Landsteiner Society for Blood Research grant (0407). The authors declare no competing financial interests.

Submitted: 23 April 2014
Accepted: 5 December 2014

REFERENCES


