# Identification of Regulatory Networks in HSCs and Their Immediate Progeny via Integrated Proteome, Transcriptome, and DNA Methylome Analysis

Nina Cabezas-Wallscheid,[1,2,10] Daniel Klimmeck,[1,2,3,10] Jenny Hansson,[3,10] Daniel B. Lipka,[4,10] Alejandro Reyes,[3,10] Qi Wang,[5,9] Dieter Weichenhan,[4] Amelie Lier,[2,6] Lisa von Paleske,[1,2] Simon Renders,[1,2] Peer Wünsche,[1,2] Petra Zeisberger,[1,2] David Brocks,[4] Lei Gu,[4,5,9] Carl Herrmann,[5,9] Simon Haas,[2,7] Marieke A.G. Essers,[2,7] Benedikt Brors,[5,8,9] Roland Eils,[5,8,9] Wolfgang Huber,[3,11] Michael D. Milsom,[2,6,11] Christoph Plass,[4,8,11] Jeroen Krijgsveld,[3,11] and Andreas Trumpp[1,2,8,11,*]

[1]Division of Stem Cells and Cancer, Deutsches Krebsforschungszentrum (DKFZ), 69120 Heidelberg, Germany
[2]Heidelberg Institute for Stem Cell Technology and Experimental Medicine (HI-STEM gGmbH), 69120 Heidelberg, Germany
[3]European Molecular Biology Laboratory (EMBL), Genome Biology Unit, 69117 Heidelberg, Germany
[4]Division of Epigenomics and Cancer Risk Factors, DKFZ, 69120 Heidelberg, Germany
[5]Division of Theoretical Bioinformatics, Department of Bioinformatics and Functional Genomics, DKFZ, 69120 Heidelberg, Germany
[6]Junior Research Group Experimental Hematology, Division of Stem Cells and Cancer, DKFZ, 69120 Heidelberg, Germany
[7]Junior Research Group Stress-induced Activation of Hematopoietic Stem Cells, Division of Stem Cells and Cancer, DKFZ, 69120 Heidelberg, Germany
[8]German Cancer Consortium (DKTK), 69120 Heidelberg, Germany
[9]Institute for Pharmacy and Molecular Biotechnology (IPMB) and BioQuant, Heidelberg University, 69120 Heidelberg, Germany
[10]Co-first author
[11]Co-senior author
*Correspondence: a.trumpp@dkfz.de
http://dx.doi.org/10.1016/j.stem.2014.07.005

## SUMMARY

In this study, we present integrated quantitative proteome, transcriptome, and methylome analyses of hematopoietic stem cells (HSCs) and four multipotent progenitor (MPP) populations. From the characterization of more than 6,000 proteins, 27,000 transcripts, and 15,000 differentially methylated regions (DMRs), we identified coordinated changes associated with early differentiation steps. DMRs show continuous gain or loss of methylation during differentiation, and the overall change in DNA methylation correlates inversely with gene expression at key loci. Our data reveal the differential expression landscape of 493 transcription factors and 682 lncRNAs and highlight specific expression clusters operating in HSCs. We also found an unexpectedly dynamic pattern of transcript isoform regulation, suggesting a critical regulatory role during HSC differentiation, and a cell cycle/ DNA repair signature associated with multipotency in MPP2 cells. This study provides a comprehensive genome-wide resource for the functional exploration of molecular, cellular, and epigenetic regulation at the top of the hematopoietic hierarchy.
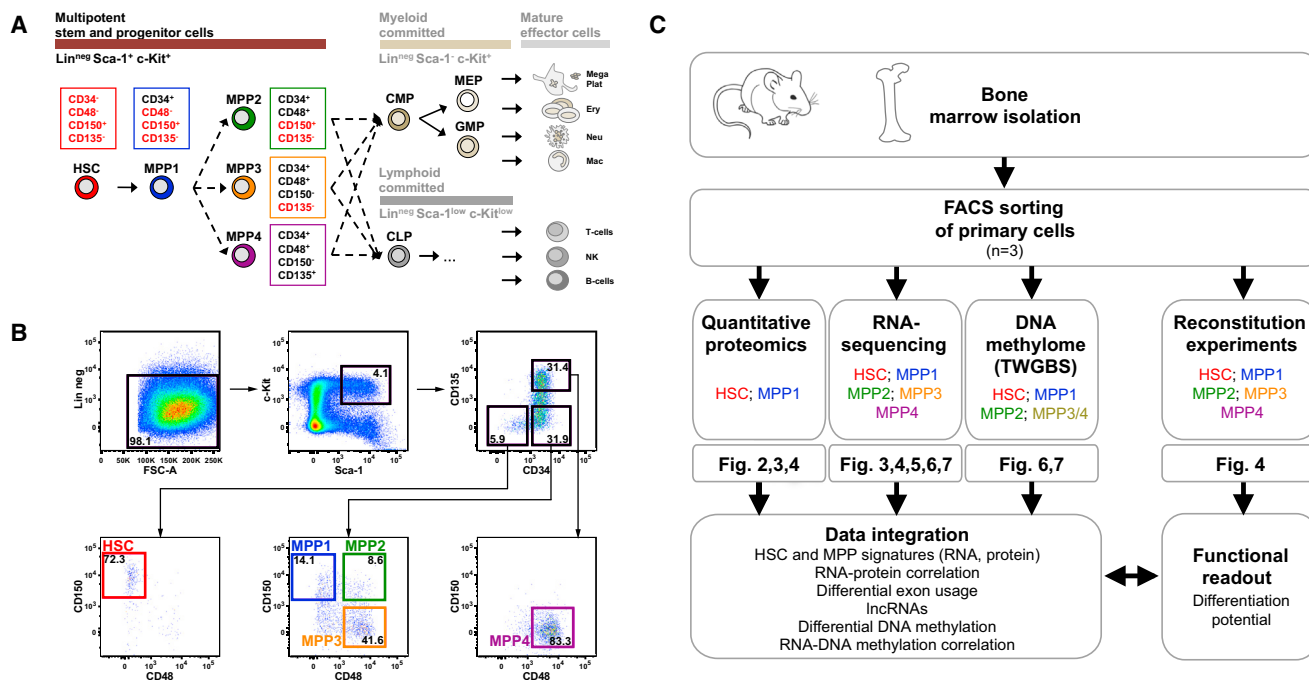
## INTRODUCTION

Hematopoietic stem cells (HSCs) are unique in their capacity to self-renew and replenish the entire blood system (Orkin and Zon, 2008; Purton and Scadden, 2007; Seita and Weissman, 2010; Wilson et al., 2009). They give rise to a series of multipotent progenitors (MPPs) with decreasing self-renewal potential, followed by differentiation toward committed progenitors and more mature cells (Adolfsson et al., 2005; Forsberg et al., 2006). MPPs have been subdivided immunophenotypically into MPP1, MPP2, MPP3, and MPP4 populations based on a stepwise gain of CD34, CD48, and CD135 as well as loss of CD150 expression (Wilson et al., 2008). However, despite recent efforts to characterize changes in gene expression and epigenome modifications that occur at distinct stages of differentiation (Gazit et al., 2013; Kent et al., 2009; McKinney-Freeman et al., 2012; Bock et al., 2012), the distinct functional characteristics and the molecular programs that maintain HSC self-renewal and drive progenitor differentiation are poorly characterized. We have taken advantage of recent technological advances enabling analysis of rare cell populations to establish comprehensive mass spectrometry-based proteome, transcriptome (RNA sequencing [RNA-seq]), and genome-wide DNA methylome (tagmentation-based whole genome bisulfite sequencing, TWGBS) data for HSCs and MPPs. We provide a comprehensive insight into the molecular mechanisms that are dynamically regulated during early HSC commitment through the MPP1–MPP4 populations. We uncovered molecular changes at the protein, RNA, and DNA levels as they occur in vivo in the context of physiologic commitment processes.

## RESULTS

The five stem/progenitor populations corresponding to HSC and MPP1–MPP4 (Wilson et al., 2008) were isolated by fluorescence-activated cell sorting (FACS) from the bone marrow of C57BL/6J mice (Figure 1; Figures S1A–S1C available online). These cells

**Figure 1. Global Analysis of Early Adult Hematopoiesis**

(A) The mouse hematopoietic system. HSCs give rise to MPPs, which commit to more mature progenitors with restricted cell fate. CMP, common myeloid progenitor; MEP, megakaryocyte erythroid progenitor; GMP, granulocyte macrophage progenitor; CLP, common lymphoid progenitor; Mega, megakaryocyte; Pla, platelet; Neu, neutrophil; Mac, macrophage; NK, natural killer cell.

(B) FACS scheme used to isolate primary mouse cells: HSC (Lin$^{neg}$ Sca-1$^+$ c-Kit$^+$, LSK, CD34- CD135- CD150+ CD48−), MPP1 (LSK CD34+ CD135− CD150+ CD48−), MPP2 (LSK CD34+ CD135− CD150+ CD48+), MPP3 (LSK CD34+ CD135− CD150− CD48+), and MPP4 (LSK CD34+ CD135+ CD150− CD48+). L, lineage-negative.

(C) Experimental study design. Shown are the different generated data sets and their respective figure numbers. See also Figures S1 and S7.
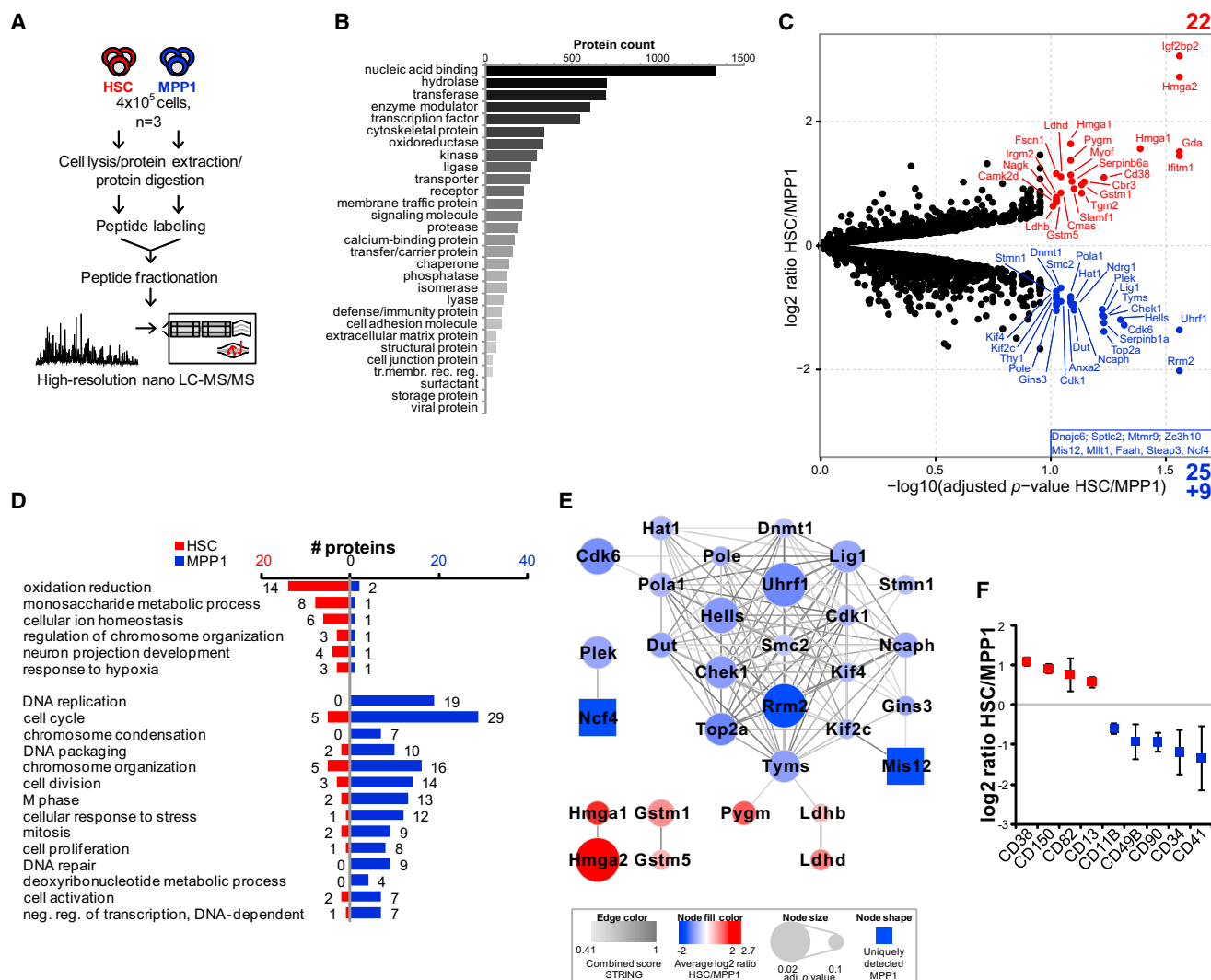
were used as a source for quantitative proteomics, RNA-seq, TWGBS, and functional reconstitution experiments.

## Proteome Differences between HSCs and MPP1 Cells Uncovered Key Molecular Players of Long-Term Reconstitution Potential

The transition from CD34− HSCs to CD34+ MPP1 is accompanied by a switch in the cells' reconstitution capabilities. We transplanted mice with 50 HSCs and observed that 100% of primary and 80% of secondary recipients showed multilineage repopulating activity (Figures S1D–S1F). In contrast, 56% of mice transplanted with MPP1 cells showed reconstitution of the primary recipient, and no engraftment was detected in secondary recipients. This is consistent with previous reports showing that the two populations are very similar but display a measurable difference in long-term self-renewal (Ema et al., 2006; Osawa et al., 1996). To investigate the molecular basis of this difference in self-renewal, we compared the proteomes of HSC and MPP1 cells in a quantitative mass spectrometry-based approach (Figure 2A; Figure S2A). From 400,000 HSCs and MPP1 purified in biological triplicate, 6,389 protein groups were identified (Table S1). These covered a broad range of protein classes, e.g. receptors (222) and transcription factors (549) (Figure 2B), as well as low-abundance proteins, as judged from the estimated protein levels that spanned more than seven orders of magnitude (Figure S2B).

In total, 4,037 proteins were quantified in all three replicates (Figure S2C). Of these, only 47 proteins were expressed differentially (false discovery rate [FDR] = 0.1), together with nine proteins detected exclusively in MPP1 (Figure 2C; Figure S2D; Table S1). This suggests that the differential self-renewal potential between HSCs and MPP1 is elicited by a relatively small number of proteins.

Gene ontology (GO) enrichment analysis revealed that cell cycle as well as related processes, such as DNA replication and cell proliferation, were strongly overrepresented in MPP1 (Figure 2D). Indeed, protein-protein interaction network analysis highlighted a large group of interconnected cell cycle proteins being expressed at elevated levels in MPP1 compared with HSCs. All major processes associated with the cell cycle machinery, including DNA polymerases (Pol1a, Pole), cell cycle checkpoint proteins (Chek1), DNA methylation maintenance and cell cycle progression (Cdk1, Cdk6), and others were represented in the network (Figure 2E). In contrast, HSCs were enriched in the monosaccharide metabolic process, including the glycolytic enzymes lactate dehydrogenase b and d (Ldhb, Ldhd) as well as Pygm. In addition, cellular ion homeostasis, including two iron transporters (Fth1, Ftl2), oxidation reduction (Rrm1, Rrm2), and response to hypoxia (Mecp2) were enriched in HSCs. Together, the data are consistent with an anaerobic metabolic program employed by quiescent HSCs, whereas MPP1 cells become primed for entry into the cell cycle and start proliferating.

**Figure 2. Proteomic Comparison of HSCs and MPP1**

(A) Quantitative proteomics workflow.

(B) Classification of identified proteins. Bars show the number of proteins within each functional class.

(C) Differential protein expression. Proteins expressed significantly higher (FDR = 0.1) in HSCs and MPP1 are shown in red and blue, respectively. The lower right corner shows proteins exclusively detected in MPP1.

(D) Overrepresented biological processes of differentially expressed proteins (202 with FDR = 0.15 and 9 exclusively detected in MPP1). Top, HSC-enriched. Bottom, MPP1-enriched. Neg. reg. of transcription, negative regulation of transcription.

(E) Protein network of differentially expressed proteins. The edges show known and predicted protein-protein interactions. STRING, Search Tool for the Retrieval of Interacting Genes/Proteins; adj., adjusted.

(F) CD cell surface proteins expressed differentially. The average log2 ratio ± SD is shown.

See also Figure S2.

Of the 22 proteins with higher expression in HSCs, we found three high mobility group AT hook proteins; namely, two isoforms of Hmga1 as well as Hmga2 (Figure 2C; Figure S2E). These chromatin-modulating transcriptional regulators have been reported recently to be potent rheostats of tumor progression (Morishita et al., 2013; Shah et al., 2013) and HSC self-renewal, proliferation, and lineage commitment checkpoints (Battista et al., 2003; Copley et al., 2013). In addition, the protein showing the highest fold change was the Hmga2 target Igf2bp2 (Figure 2C; Cleynen et al., 2007). Hmga1, Hmga2, and Igf2bp2 are downstream mediators of the Lin28-let7 pathway that link metabolism to prolifera-

tion and drive self-renewal (Shyh-Chang et al., 2013; Yaniv and Yisraeli, 2002). High expression of this pathway in HSCs suggests a role in self-renewal of HSCs, whereas its downregulation in MPP1 cells correlates with decreasing self-renewal activity.

Two glutathione S-transferases (GST), Gstm1 and Gstm5, were found to be expressed at higher levels in HSCs compared to MPP1 (Figure 2C). Moreover, elevated levels in HSCs were consistent for all 11 GSTs quantified (Figure S2F). This points to a requirement for this enzyme class in HSCs, which may relate to their ability to mediate the conjugation of xenobiotics for the purpose of detoxification and defense against environmental

stress and cellular damage (Tew and Townsend, 2012). This suggests that homeostatic HSCs have an array of mechanisms in place to protect their cellular integrity, extending an observation that we have made previously in hematopoietic progenitors (Klimmeck et al., 2012). Along these lines, two interferon-inducible proteins involved in host defense (Ifitm1 and Irgm2) were expressed at higher levels in HSCs compared with MPP1 (Figure 2C), suggesting that the type I interferon pathway is not only critical for the response to stress but also during homeostasis (Essers et al., 2009; Trumpp et al., 2010). Moreover, HSCs and MPP1 employ different intracellular serpins for protection against death during stress because Serpinb6a and Serpinb1a were expressed at higher levels in either HSC or MPP1, respectively (Figure 2C). Taken together, HSCs harbor proteins involved in immune defense and detoxification, indicative of an increased self-protecting repertoire compared with MPP1.

Among the 86 quantified cluster of differentiation (CD) surface proteins (Table S1), nine showed a strong differential expression between HSC and MPP1 (Figure 2F). These included CD34 as well as other membrane proteins described previously in the context of stem/progenitors, namely CD38, CD41, CD49b, and CD90 (Benveniste et al., 2010; Dumon et al., 2012; Weissman and Shizuru, 2008). Additionally, our analysis identified CD82 and CD13 to be expressed at higher levels in HSCs, whereas CD11b had a higher expression level in MPP1. These findings may be used to develop additional marker combinations to distinguish HSCs and MPP1 as well as to further refine the classification of stem/progenitor intermediates.

### The Transcriptome and Proteome Are Highly Coordinated upon HSC Differentiation

We next analyzed the transcriptome of HSCs and MPP1 using high-throughput RNA sequencing starting from 30,000 FACS-sorted primary cells (Figure 3A; Figure S3A). Robust and reproducible data were obtained for all samples, with more than $2 \times 10^8$ total readings per population (Figure S3B). In total, transcripts corresponding to 27,881 genes were identified (Table S2). Those genes were classified, according to their database annotation, into 21 RNA categories, and, as expected, protein-coding transcripts were highly represented (68.9%) (19,219; Figure 3B). In line with the proteome data, the protein-coding transcripts displayed a high diversity of functionalities (Figure 3C), including transcription factors (TF, 1,776 genes), receptors (1,796), and cell adhesion molecules (584). Additionally, the expression of 8,662 noncoding RNA species was identified, including pseudogenes (4,034), microRNAs (miRNAs) (642) and long noncoding RNAs (lncRNAs) (589).

We found 479 genes to be expressed differentially between HSCs and MPP1 (FDR = 0.1; Table S2; Figure 3D). Consistent with the GO terms enriched in the proteome data, we found an overrepresentation of processes related to metabolism (positive regulation of the phosphate metabolic process and sulfur compound metabolic process) and response to hypoxia (cellular response to oxygen-containing compounds) within the genes expressed at higher levels in HSCs (Figure 3E). Furthermore, TFs involved in different aspects of developmental signaling (cellular developmental process, cell differentiation *Mecom*, *Hoxb7*) were enriched in HSCs. In contrast, the DNA metabolic process, cell cycle, and nuclear division were enriched in
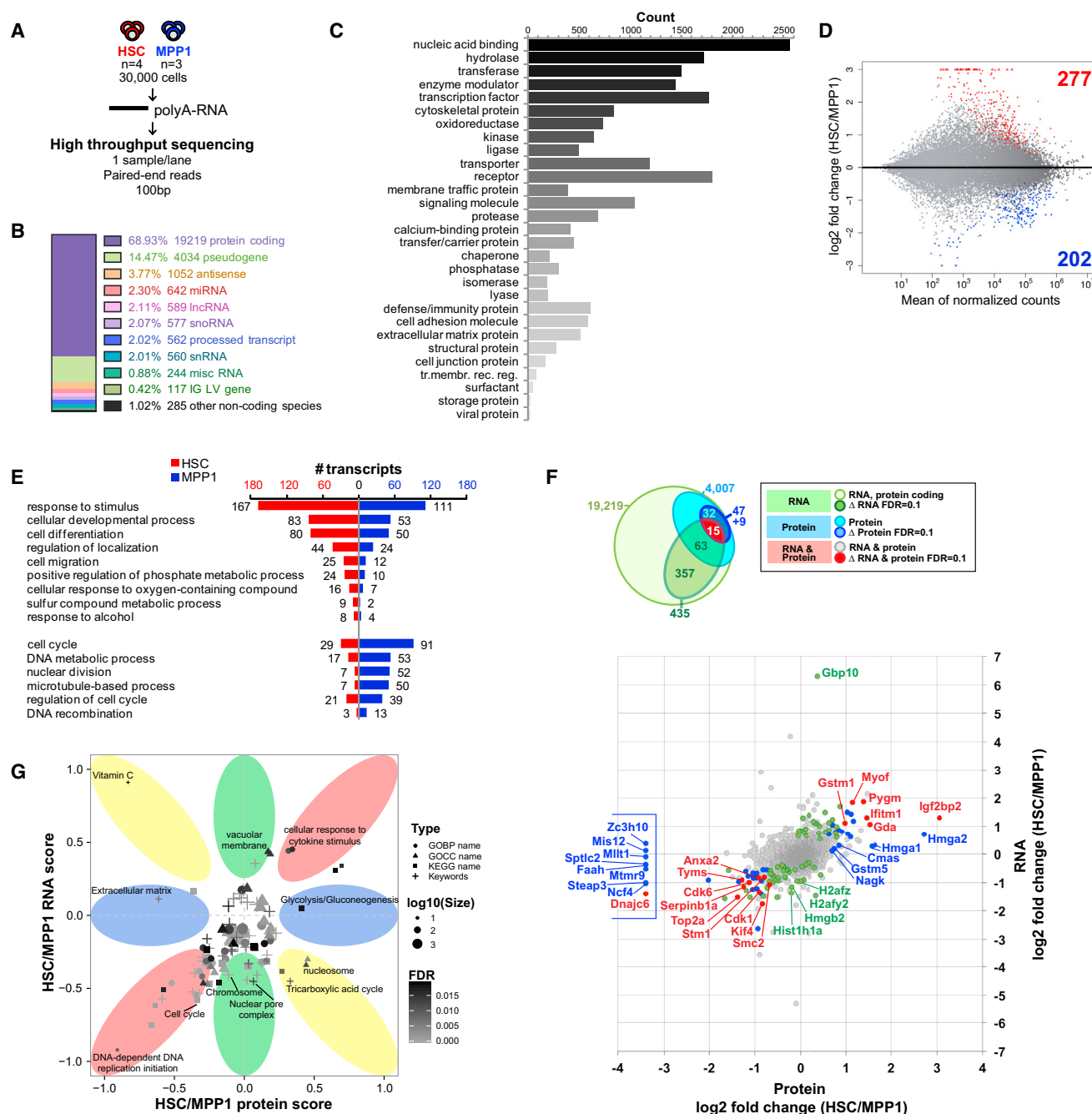
MPP1 in agreement with the proteome findings and increasing proliferative activity.

To systematically investigate potential posttranscriptional regulation at the transition from HSCs to MPP1, we correlated the transcriptome data with the proteome data (Figures 3F–3G; Table S3). Out of the quantified proteins, we were able to assign 99.3% (4,007 of 4,036) to their corresponding transcripts (Figure 3F; Figures S3D–S3E). Notably, in addition to a strong positive correlation between both data sets, all proteins that were expressed differentially were also consistently upregulated or downregulated at the transcript level (Figure 3F). For example, the response to cytokine stimulus was overrepresented at both the RNA and protein levels in HSCs, whereas genes belonging to the cell cycle and DNA replication were expressed strongly in MPP1 (Figure 3G). As a notable exception to this overall pattern, enzymes in glycolysis/gluconeogenesis (Cmas, Gstm5, and Nagk), as well as several downstream targets of the Lin28-let7 pathway (Hmga1, Hmga2, and Igf2bp2), showed only modest changes at the mRNA level (Figures 3F and 3G), in line with their suggested regulation at the posttranscriptional level (Shyh-Chang and Daley, 2013). In addition, we found the GO term chromosome enriched and several long-lived histones related to chromosome organization (e.g. *H2afz*, *H2afy2*) to be downregulated at the RNA level in MPP1. Because these were not detected as being expressed differentially at the protein level (Figures 3F and 3G), we hypothesize that the regulatory initiation of nuclear reorganization starts already in MPP1 but becomes only operational at subsequent stages. The overall high similarity between the transcriptome and proteome suggests that posttranscriptional regulation in HSC/progenitors might be used preferentially only for some pathways, including the Lin28-Hmga-Igf2bp2 axis.

### MPP2 Cells Are Multipotent, but MPP3 and MPP4 Show a Differentiation Lineage Bias

To investigate the functional attributes of MPP2–MPP4, we performed in vivo reconstitution assays using FACS-sorted cells (Figure 4A; Figures S4A–S4B). The contribution of transplanted MPP2 cells to the T cell, B cell, and myeloid progeny in peripheral blood (PB) after 4 months was 23%, 35%, and 32%, respectively, compared with the 56%, 70%, and 82% observed for transplanted HSCs (Figures 4B and 4C; Figure S4A). In contrast, MPP3 and MPP4 cells generated only a small number of differentiated cells (mostly below 2%), although the progeny were retained in the PB for at least 4 months. MPP3-transplanted mice showed a bias toward the production of myeloid cells, which was evident from 1 week posttransplantation but decreased over time. MPP4-transplanted animals preferentially generated lymphoid B cell progeny starting at 3 weeks posttransplantation, whereas myeloid progeny remained below 1% during the entire observation period. In line with this, T cells peaked after 3 weeks in the thymi of MPP4-transplanted mice (Figure S4B). In addition, colony-forming unit (CFU) assays showed that, like MPP2 and MPP3 progenitors, MPP4 cells also generate myeloid colonies (Figure S4C), in line with earlier studies performed with CD135+ lymphoid-primed multipotent progenitors (Adolfsson et al., 2005). In summary, MPP2 generates a large number of long-term myeloid, B cell, and T cell progeny upon transplantation. In contrast, the other two populations generate only a limited number of progeny in vivo, with a significant lineage

**Figure 3. Comparison of the Transcriptome and Proteome of HSCs and MPP1 Cells**

(A) RNA-seq workflow.

(B) RNA categories of 27,881 quantified genes. Shown are the percentage and number of RNA species within each RNA class.

(C) Classification of quantified protein-coding genes. Bars show the number of genes within each functional class, ranked based on the protein coverage (Figure 2B).
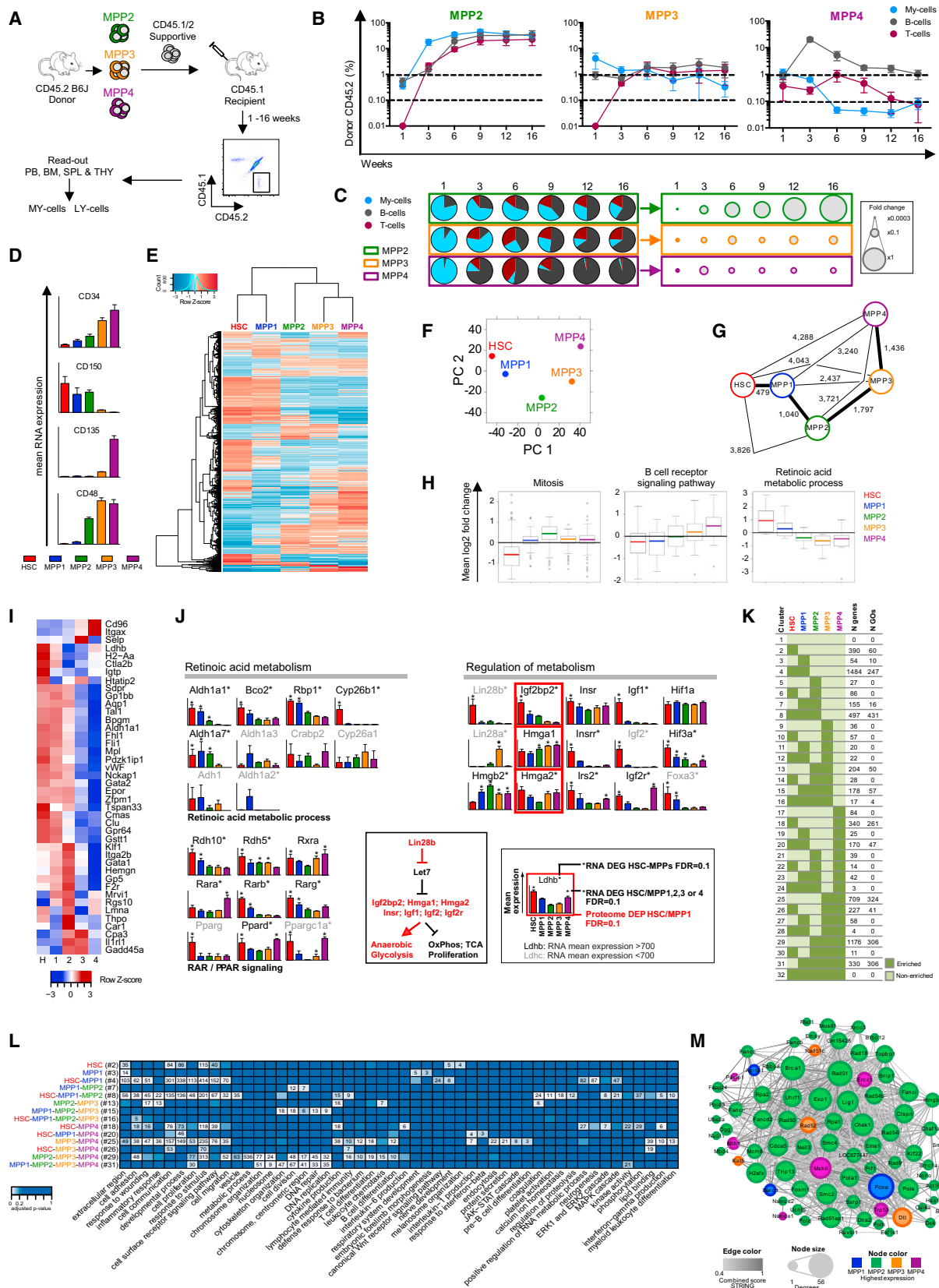
(D) Differential gene expression. Genes highly expressed (FDR = 0.1) in HSCs or MPP1 are shown in red and blue, respectively.

(E) Overrepresented biological processes of differentially expressed genes.

(F) Overlap and correlation between protein and RNA expression changes. Top panel, integration of proteome and transcriptome data sets. Quantified proteins (blue) were mapped to protein-coding transcripts (green). Differentially expressed transcripts (dark green, 78 mapped of 435) and differentially expressed proteins (dark blue, 47 significant plus 9 exclusively detected proteins) were assessed for overlap (red). Bottom panel, significant (FDR = 0.1) changes at RNA (green), protein (blue), and both levels (red, correlation coefficient R = 0.93), respectively. The box on the left indicates exclusively detected proteins.

(G) 2D GO enrichment analysis of protein and RNA expression changes. Red regions correspond to concordant enrichment or lower expression. Blue and green regions highlight terms that are enriched or lower in one direction but not in the other, whereas terms in yellow regions show anticorrelating behavior. GOBP, gene ontology biological process; GOCC, gene ontology cellular compartment; KEGG, Kyoto Encyclopedia of Genes and Genomes.

See also Figure S3.

bias toward mostly myeloid (MPP3) or lymphoid (MPP4) cell types.

## Clustering Analyses Identify Global Molecular Differences during HSC Differentiation

We sequenced the transcriptomes of the MPP2, MPP3, and MPP4 populations in biological triplicates (Figures 4D–4M; Figures S4D–S4G). The surface markers used for FACS showed consistency between transcript and protein levels (Figure 4D). We detected 6,487 genes to be expressed differentially across all five populations (FDR = 0.1; Table S2). The differential expression of a set of genes was confirmed by quantitative real-time PCR (Figure S3F), validating the robustness of the data. Unsupervised clustering of gene expression profiles grouped together HSC and MPP1 as well as MPP3 and MPP4 (Figure 4E). This was verified by principal component analysis (Figure 4F) and by the numbers of differentially expressed genes in each pairwise comparison (Figure 4G). In all analyses, MPP2 was placed between the HSC-MPP1 and MPP3-MPP4 clusters (Figures 4B and C; Figures S4A and S4B).

## Clustering of Gene Expression Profiles Suggests Processes Operating in HSC/MPP1-MPP4 Cells

The 6,487 genes with differential expression across the five cell populations were enriched in specific GO terms (http://www-huber.embl.de/SMHSC/HSCGO/goAnalysis.html; Supplemental Information), of which we highlight three selected examples (Figure 4H). Based on the expression data, MPP2 is the mitotically most active population, whereas HSC is the most quiescent one. In agreement with the functional lymphoid differentiation bias observed in MPP4-transplanted recipients (Figures 4B and 4C), the B cell receptor signaling pathway was enriched in this population. Consistent with this, mapping of the HSC-MPP data against granulocyte macrophage or lymphoid-primed progenitor signatures described previously revealed an enrichment for MPP3 and MPP4 populations, respectively (Må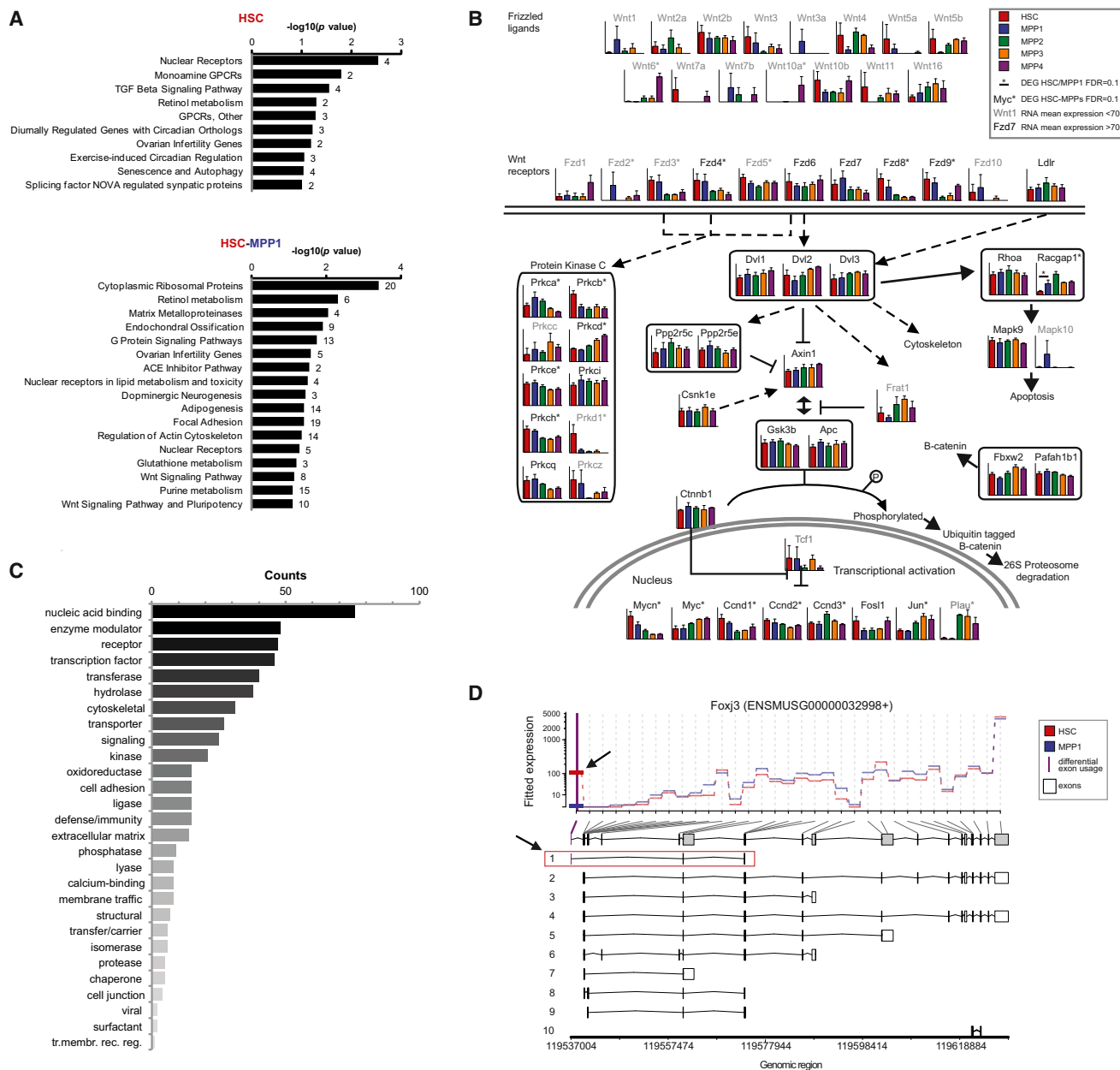nsson et al., 2007; Figure S4E). In addition, genes of a megakaryocytic erythroid signature described previously (Månsson et al., 2007; Gekas and Graf, 2013; Sanjuan-Pla et al., 2013) were expressed preferentially in HSCs, MPP1, and MPP2 compared with MPP3 and MPP4 (Figure 4I).

Consistent with the differential expression of metabolic enzymes between HSCs and MPP1 (Figures 2D and 3E), we found genes belonging to the retinoic acid metabolic process to be highly expressed in HSCs compared with all other populations (Figures 4H and 4J), suggesting a role in adult HSCs. Furthermore, we found evidence for stage-dependent and isoform-specific expression of essential glycolytic enzymes in HSCs (e.g. *Aldoc*, *Ldhb*; Figure S4D), extending recent studies demonstrating the relevance of anaerobic glucose metabolism for the maintenance of HSC self-renewal (Takubo et al., 2013). In contrast, the large majority of enzymes involved in the tricarboxylic acid cycle and oxidative phosphorylation showed MPP2-enriched expression (e.g. *Fh1*, *Aco2*). Interestingly, although *Lin28b* and its target genes, including *Hmga2* and *Igf2bp2*, were highly expressed in HSCs compared with MPP1, the family member *Lin28a* was specifically upregulated in MPP3 (Figure 4J). Therefore, our findings support the recently established role of the *Lin28-let7* axis in glucose metabolism in stem cells (Shyh-Chang and Daley, 2013) and suggest *Lin28b* as a candidate for an upstream posttranscriptional regulator of glycolytic enzymes in HSCs (Figures 3F and 3G).

Next we assigned each of the 6,487 differentially expressed genes to one of the 32 possible relative expression patterns (Figure 4K) and tested for overrepresented GO terms within each pattern (Table S4). Population-specific GO terms are displayed in Figure 4L. Consistent with the strong upregulation of cell cycle-associated genes in MPP2, we observed that 43% of the genes involved in DNA repair also showed the highest expression in MPP2 (Figure S4F). The protein-protein interaction network of the HSC-suppressed DNA repair genes comprised, e.g., *Brca1* and *Exo1* (Figure 4M), representing potential players for the switch in DNA repair mechanisms between HSCs and MPP2. Interestingly, the two common mechanisms for repair of DNA double strand breaks (homologous recombination and

---

**Figure 4. Lineage Potential and Whole Transcriptome Molecular Signatures of HSC-MPP1-4 Populations**

(A) Experimental workflow for investigation of the in vivo reconstitution potential. MPP2, MPP3, and MPP4 were transplanted into lethally irradiated mice, and the myeloid (MY) and lymphoid (LY) outcomes were monitored over time by flow cytometry. SPL, spleen; THY, thymus.

(B and C) Monitoring the myeloid and lymphoid outcomes in peripheral blood after transplantation of MPP2, MPP3, and MPP4. The relative percentage (B) and contribution (C) of donor myeloid (blue), B cells (gray), and T cells (purple) were monitored from the peripheral blood of MPP2-, MPP3-, and MPP4-transplanted animals at 1–16 weeks. n = 8-12 per group. The sizes of the circles in the right plot represent fold change relative to HSC contribution after 3 weeks.

(D) Differential RNA expression of the surface markers used for FACS. Average RNA expression ± SD is shown.

(E) Clustered heat map. The colors represent the normalized average read count in each of the five cell populations for all differentially expressed genes (FDR = 0.1).

(F) Principal component analysis.

(G) Relative distances are shown based on numbers of differentially expressed genes in pairwise comparisons.

(H) GO enrichment analysis of 6,387 genes with differential expression. Three representative examples of significantly enriched GO terms out of 1,001 terms are shown.

(I) Expression of megakaryocytic/erythrocytic genes.

(J) Transcriptome and proteome signatures of metabolism. Average RNA expression ± SD is shown (arbitrary units). RAR, retinoic acid receptor; PPAR, peroxisome proliferator-activated receptor; OxPhos, oxidative phosphorylation; DEG, differentially expressed genes.

(K) Gene expression clusters. 6,487 differentially expressed genes were grouped into 32 clusters based on higher expression (enriched) in one or several cell population(s) compared with mean expression level across all cell populations.

(L) Representative overrepresented GO terms of gene expression clusters. The number of genes within each cluster is shown for significantly enriched GO terms (FDR = 0.1).

(M) Network of HSC-suppressed DNA repair genes. Genes with low expression in HSCs (Z-score < −1.5) and with known and predicted protein-protein interactions are shown. The colors depict the MPP population with the highest expression. Note that most genes are most highly expressed in MPP2.

See also Figure S4.

**Figure 5. Gene Expression Signatures and Transcript Isoform Regulation in HSCs and MPP1–MPP4 Cells**

(A) Signaling pathway enrichment analysis. Shown is the overrepresentation of signaling pathways within HSC- and HSC-MPP1 clusters. NOVA, neuro-oncological ventral antigen; ACE, angiotensin-converting enzyme.

(B) Wnt signaling. Shown is a pathway analysis based on WikiPathways (WP403). Average RNA expression ± SD is shown (arbitrary units).

(C) Association of differentially spliced genes to protein classes.

(D) Representative example for a TF (*Foxj3*) showing differential exon use. The first *Foxj3* exon was detected at higher levels in HSC (red) compared with MPP1 (blue).

nonhomologous end joining) also showed a different expression pattern across the five cell populations. Both repair processes are low in HSCs and are most enriched in MPP2 (Figure S4G). In summary, our data indicate that the increased mitotic activity during HSC differentiation, starting in MPP1 and peaking in MPP2, is associated with a parallel increase in expression of DNA repair pathway genes.

To investigate the signaling pathway complexity of both the HSC and the HSC-MPP1 self-renewal clusters, we tested for overrepresentation of differentially expressed genes in pathways annotated in REACTOME within each group of genes. We found, among others, G protein-coupled receptor (*Gpr143*, *Pde1c*) and transforming growth factor β (*Inhba*, *Bmp4*) signaling to be enriched (Figure 5A; Table S4). In accordance with the GO

enrichments (Figure 4L), Wnt signaling was prominent in self-renewing HSCs and MPP1. Although the role of canonical β-catenin-mediated Wnt signaling in adult HSCs remains controversial (Koch et al., 2008; Luis et al., 2012), noncanonical signaling has recently been suggested to mediate critical interactions between HSCs and their niches (Sugimura et al., 2012). Therefore, we interrogated individual expression patterns of the entire Wnt signaling pathway during differentiation of HSCs toward MPP4. Although the pathway was highly overrepresented in HSC-MPP1, none of the Wnt ligands were expressed at high levels, arguing against autocrine production of Wnt ligands (Figure 5B). However, even low levels of Wnt ligands support stem cell self-renewal (Luis et al., 2012). In agreement with this, three frizzled receptors (*Fzd4*, *Fzd8*, and *Fzd9)* were highly expressed in HSC-MPP1, consistent with reported *Fzd4* expression in human CD34+ cells (Tickenbrock et al., 2008) and a role of noncanonical *Fzd8* in the maintenance of quiescent long-term HSCs (Sugimura et al., 2012). Taken together, our analysis underscores the relevance of Wnt signaling for HSCs but also outlines the complex isoform-specific enzyme utilization operational in HSCs and the different MPPs. This likely contributes to the dynamic regulation of this pathway, which is critical for most stem/progenitor cell types.

### Transcription Factor and Transcript Isoform Regulatory Landscape

Among the genes expressed differentially across the five populations, we identified 490 TFs, including members of the Fox, Gata, and Hox families (Table S2). TF splice isoforms can have stage- and tissue-specific expression patterns throughout development (López, 1995; Sebastian et al., 2013) and alternative splicing has been shown to trigger switches between activating and repressive TF isoforms (Taneri et al., 2004). We tested for differential exon use for detection of alternative transcription start sites, alternative splicing, and alternative terminations sites (Anders et al., 2012) in the HSC MPP1–MPP4 transcriptome data. Among the 497 genes expressing transcript isoform variants across HSC MPPs (FDR = 0.1), we identified 46 TFs (Figure 5C; http://www-huber.embl.de/SMHSC/HSCDEU/overall/testForDEU.html; Supplemental Information). We observed that the first exon of the TF forkhead box J3, *Foxj3*, was included more frequently in the HSC transcripts compared with MPP1, suggesting the expression of a specific *Foxj3* isoform in HSCs (Figure 5D). Although the role of this Foxj3 variant in hematopoiesis is uncharacterized, an alternative splicing switch for another forkhead family member, *Foxp1*, has recently been shown to regulate embryonic stem cell pluripotency and reprogramming by changing its DNA-binding preference (Gabut et al., 2011).

### Genome-wide DNA Methylation Analysis of HSCs/MPPs Identifies Candidate Regions for Epigenetic Regulation
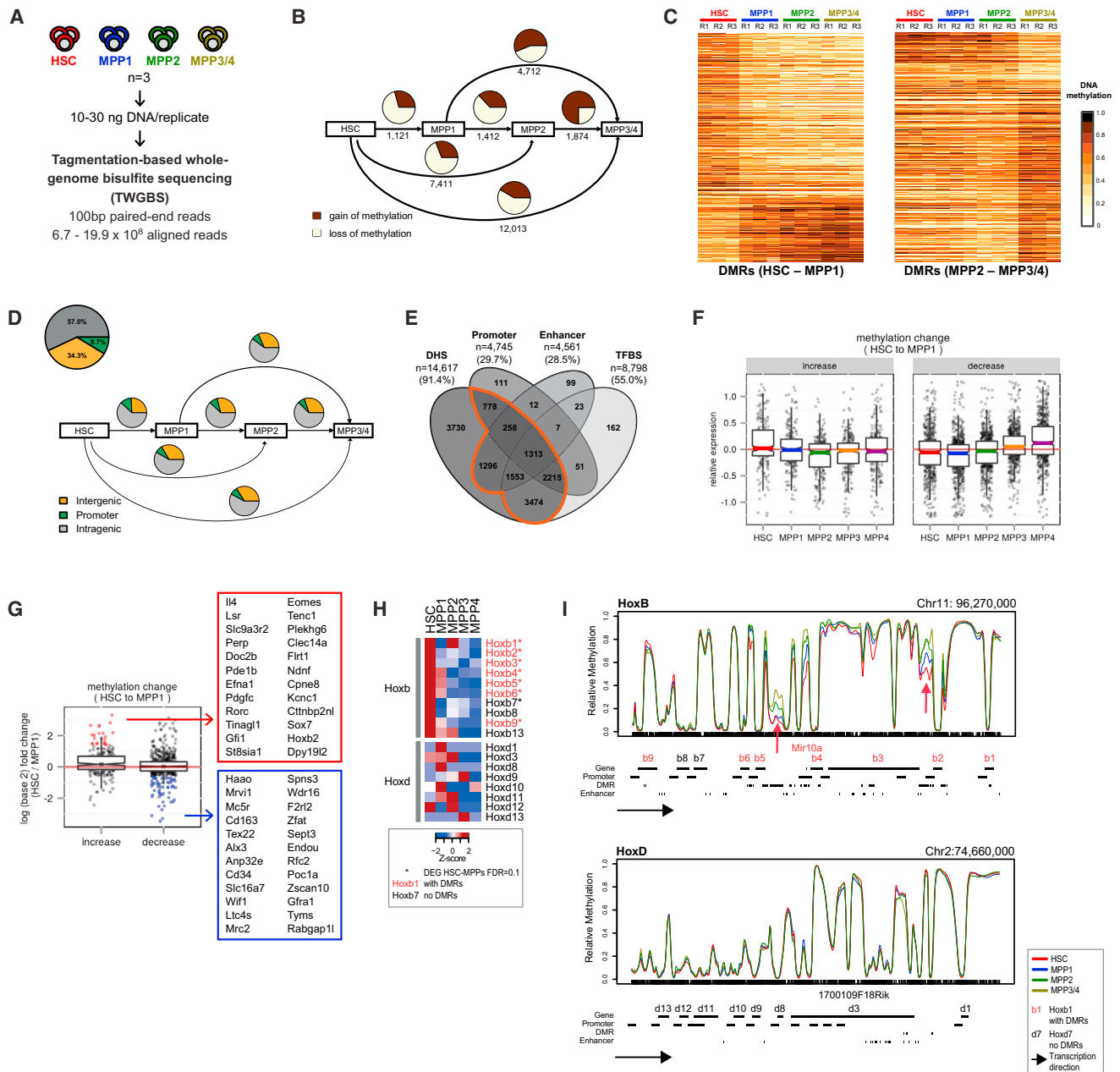
To investigate the methylation status of all cytosine residues within the genome of HSCs and MPPs, we subjected ≥10,000 FACS-sorted cells per biological replicate to TWGBS (Wang et al., 2013) (Figure 6A; Figure S5A). Robust data were obtained for all samples, with more than $6 \times 10^8$ reads and a combined genomic cytosine-phosphate-guanine (CpG) coverage of more than 33-fold per population across the three biological replicates (Figure S5B). Global levels of DNA methylation ranged between

81% and 83% and were not significantly different across populations, whereas pairwise comparisons identified a total of 15,887 distinct differentially methylated regions (DMRs) (Figure 6B; Table S5). Mapping these DMRs to previous DNA methylome data generated on HSC/MPPs using reduced representation bisulfite sequencing (RRBS) (Bock et al., 2012), we found that 85% of all DMRs were exclusively identified using the TWGBS analysis (e.g. *Mecom*; Figure S5C), demonstrating the additional coverage of our data set. Early commitment steps correlated with lower numbers of DMRs (1,121, HSC-MPP1), which likely reflects close ontogenic and functional relationships, and were mainly associated with loss of methylation (71%, HSC-MPP1). In contrast, transitions between more differentiated MPP populations showed higher numbers of DMRs (1,874, MPP2-MPP3/MPP4) and gain of methylation (75%, MPP2-MPP3/MPP4; Figure 6C; Figure S5D; Table S5). Globally, methylation changes showed continuous progressive behavior (gain or loss) through HSC differentiation (Figure 6C).

Next we investigated the overlap of DMRs with annotated genomic features derived from *Refseq* (Figure 6D; Pruitt et al., 2009). The majority of DMRs were located in intragenic regions (57%), whereas promoters and intergenic regions comprised 9% and 34%, respectively. Comparison of our data set with experimentally defined functional genomic elements (Stamatoyannopoulos et al., 2012; Shen et al., 2012) demonstrated a significant overrepresentation of DNaseI hypersensitivity sites (DHSs), promoters, enhancers, and/or transcription factor binding sites (TFBSs) across DMRs (Figure S5E; all p values < $10^{-16}$). Interestingly, 74% of all DMRs that overlapped with DHSs (10,887) mapped to regions annotated as TFBSs, promoters, and/or enhancers, suggesting that dynamic methylation upon hematopoietic commitment occurs predominantly at *cis*-regulatory elements (Figure 6E). Together, these data support the hypothesis that many DMRs represent *cis*-regulatory elements in HSC-MPPs.

### Gene Expression Anticorrelates Globally with Changes in DNA Methylation

Given the significant overlap between DMRs and functionally important regulatory regions of the genome, we postulated that the methylation changes could be associated with gene expression during HSC differentiation. Indeed, we found that DMRs coincided with previously identified *cis*-acting regulatory sites at the Sfpi1 (Rosenbauer et al., 2004) and Gata2 (Gao et al., 2013) loci (Figure S5F), which both encode TFs described as effectors of hematopoietic commitment. We generated gene sets based on DMRs that were stratified for methylation increase or decrease and interrogated the associated gene expression levels between HSCs and each of the MPP populations. DMRs detected at the early transition of HSC to MPP1 were significantly associated with anticorrelated gene expression along differentiation (Figure 6F). In addition, pairwise comparisons demonstrated a global anticorrelation between transcription and DNA methylation (Figure 6G; Supplemental Information). Among the top anticorrelating genes between HSC-MPP1 (80; Table S6), we found a number of well documented markers of early hematopoietic differentiation and HSC function (e.g. *Hoxb2*, *Rorc*, *Cd34*). For example, the increasing expression of *Cd34* and Wnt-inhibitory factor (*Wif1*) from HSC to MPP1 is

**Figure 6. Global DNA Methylation Analysis and Anticorrelation with Gene Expression**

(A) DNA methylome workflow.

(B) Gain or loss of methylation DMRs between HSC-MPPs. The numbers indicate total DMRs.

(C) Clustering of DMRs identified between HSC-MPP1 or MPP2-MPP3/MPP4. Each horizontal dash represents a DMR. R1–R3, replicates 1–3.

(D) Overlap of DMRs with gene-centric Refseq genomic regions.

(E) Percent overlap of all DMRs with experimentally defined functional genomic elements based on available data from the mouse ENCODE project.

(F) Overall comparison of DNA methylation to gene expression. The box plots represent relative gene expression associated with either gain (left) or loss (right) of methylation in the transition from HSC and MPP1.

(G) Pairwise comparison of DNA methylation to gene expression. The box plots represent log2 fold change of gene expression associated with either gain or loss of methylation in the transition from HSC to MPP1. The top 48 anticorrelated genes (24 highest/24 lowest expression in HSCs) are indicated.

(H) Gene expression levels of all members of the HoxB and HoxD clusters.

(I) Relative DNA methylation profile for the HoxB and HoxD clusters. Red arrows indicate examples of DMRs.

See also Figure S5.

associated with decreasing methylation at the DMRs of these loci. In contrast, the cyclic AMP-mediating cyclic nucleotide phosphodiesterase 1b (*Pde1b*), as well as the retinoic acid receptor orphan receptor γ(*Rorc*), become methylated and downregulated during differentiation toward MPP1.

Hox TF clusters are critical during normal and malignant hematopoiesis (Argiropoulos and Humphries, 2007), but little is known about their epigenetic regulation. We found that most members of the HoxB family showed the highest expression in HSCs and were associated with DMRs (seven out of nine; Figures 6H and 6I). Notably, HSC differentiation showed a continuous increase in DMR methylation and paralleled decrease of RNA expression (Figure 6I, see arrows for *Hoxb2* and *Hoxb4*). In contrast, the members within the HoxD cluster showed no specific expression pattern in HSCs and MPPs, and none of its genes were associated with a DMR (Figure 6I). Similar results were found for HoxA and HoxC clusters, respectively (Figures S5G and S5H). In conclusion, the integrated analysis of the methylome and transcriptome provides a resource of genes whose expression is, at least partially, regulated by DNA methylation and that are possibly involved in the hard wiring of the hierarchical organization of the HSC/progenitor populations.

## Differential lncRNA Expression and the Imprinted Gene Network in Stem/Progenitors

Long noncoding RNAs (lncRNAs) have been implicated in guiding chromatin remodeling complexes to specific target genes and mediating gene activation or silencing by recruiting the DNA demethylation machinery to the promoter (Arab et al., 2014) or recruiting repressive complexes such as PRC2, respectively (Rinn and Chang, 2012). However, little is known about their functional role or regulation in stem cells and hematopoiesis (Paralkar and Weiss, 2013). We identified 682 lncRNAs expressed in HSC-MPPs, 79 of which were differentially expressed across all five populations (FDR = 0.1; Figure 7A; Table S2). Unsupervised clustering (Figure 7B) and pairwise comparisons of differentially expressed lncRNAs revealed the same population relationship as using the entire transcriptome data set, again placing MPP2 between the HSC-MPP1 and the MPP3-MPP4 populations (compare Figures 4E–4G and S6A). Next we clustered the significantly changed lncRNAs based on their relative expression levels across all populations (Figure 7C; Table S4). As shown in cluster 2, 12 lncRNAs are strongly expressed in HSCs compared with the rest of the populations, and none of these has yet been functionally annotated or studied (e.g. *2410080I02Rik*, *Gm12474*). In addition, 14 lncRNAs are coexpressed in HSC-MPP1 (cluster 4), representing additional candidates for regulation of self-renewal (e.g. *H19*, *Malat1*, *Meg3*). In agreement, the imprinted *H19* lncRNA has recently been shown to mediate HSC quiescence by inhibiting insulin growth factor (IGF) signaling (Venkatraman et al., 2013). Thirteen lncRNAs were enriched in MPP3-MPP4, suggesting regulatory roles in these lineage-restricted populations (cluster 25; e.g. *Gm568*, *Neat1*). Although *Neat1* is essential for the integrity of the nuclear substructure, it has also been linked to the immune response after HIV-1 infection (Zhang et al., 2013) and might, therefore, be involved in the maintenance of immune regulatory circuits in MPP3-MPP4. Notably, most of the differentially expressed lncRNAs identified here have not been studied in hematopoiesis

and/or have unknown functions. The lncRNAs *H19* and *Meg3* that are highly expressed in HSCs compared with MPPs (Figure 7C, validated by quantitative RT-PCR; Figure S6B) are core members of the transcriptional imprinted gene network (IGN), which has been postulated to regulate embryonic growth (Varrault et al., 2006). In adult hematopoiesis, the genetic deletion of several members of this network affects HSC self-renewal integrity (e.g. *Cdkn1c/p57KIP2*) (Berg et al., 2011; Rossi et al., 2012). Therefore, we further investigated the expression of the IGN members and found an overall strong expression in HSCs but a steady decrease during the differentiation process (Figure 7D), suggesting a contribution of the IGN in the maintenance of self-renewal and/or quiescence of HSCs.
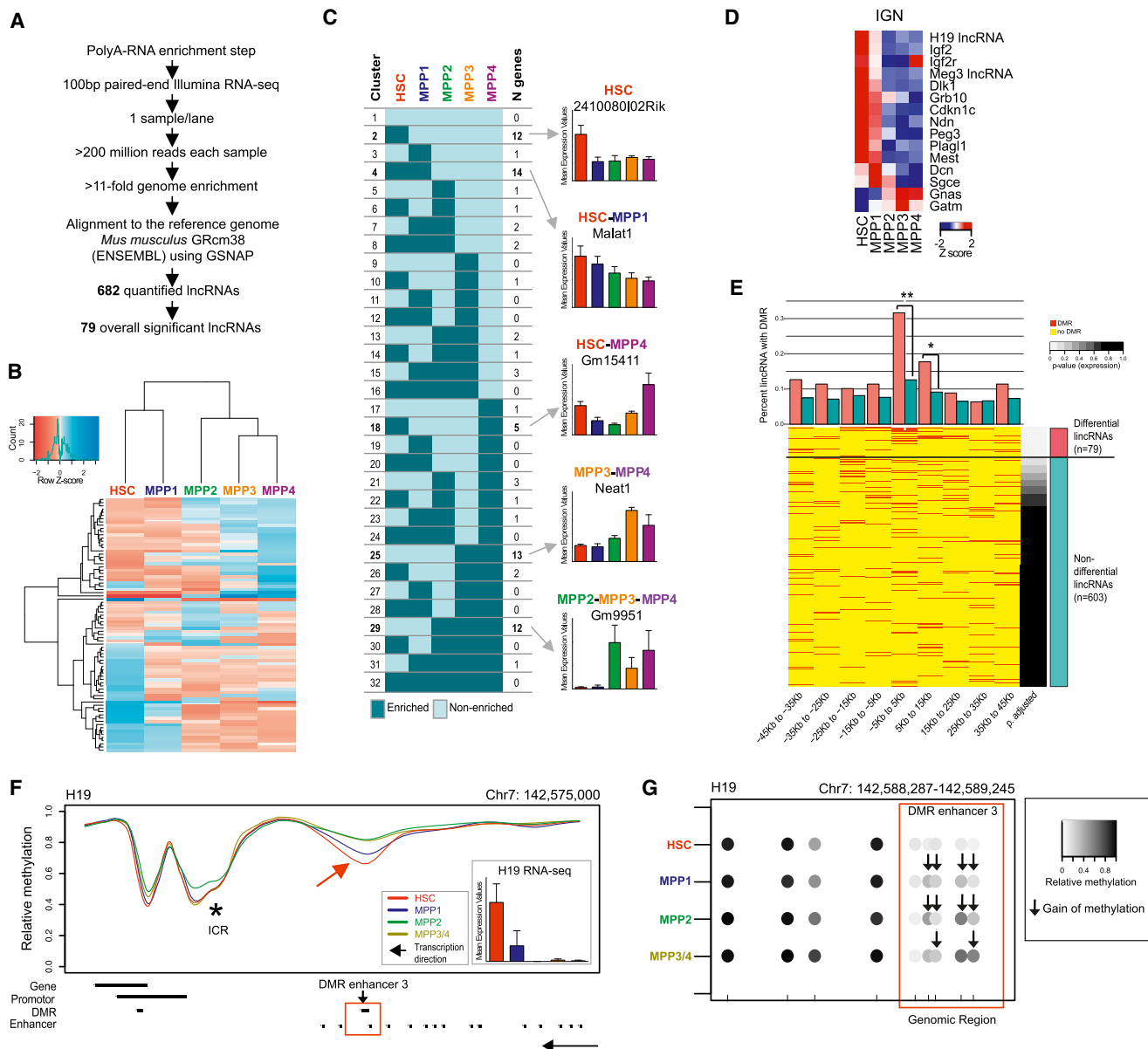
Finally, we interrogated the DMRs within the loci encoding the 682 quantified lncRNAs. This revealed a significant enrichment of DMRs in the differentially expressed lncRNAs. These DMRs were enriched within a 10 kb window centered on the transcription start site (Figure 7E). A notable example is the *H19* locus, which exhibits a DMR in an enhancer region outside of its imprinting control region (Figures 7F and 7G). The increasing level of methylation at this enhancer during the transition from HSCs to MPPs correlates with decreasing expression during differentiation (Figure 7F). Overall, these data suggest that differential DNA methylation of regulatory regions is a likely mechanism by which lncRNA expression is controlled in HSCs and their progeny.

## DISCUSSION

In this study, we describe a combined proteome, transcriptome, and DNA methylome analysis of highly purified primary HSCs and four downstream MPPs, which we characterized additionally using in vitro and in vivo functional assays (Figure S7). Our data sets uncover progressively changing cell type-specific methylation, gene, and protein expression landscapes starting with quiescent CD34-CD150+CD48-LSK HSCs that sit at the pinnacle of the hematopoietic hierarchy. These differentiate toward slowly cycling multipotent MPP1, followed by multipotently cycling MPP2. The steady increase in the activity of the cell cycle and proliferation machinery is paralleled by the robust upregulation of the entire DNA repair machinery. This raises the possibility that physiological DNA replication in proliferating early progenitors generates significant replicative stress that needs to be counteracted by the activation of the DNA repair machinery to ensure genome integrity (Bakker and Passegué, 2013).

We found that the majority of DMRs either progressively gain or lose DNA methylation through early HSC differentiation. Moreover, we observed a global anticorrelation between DNA methylation and gene expression. Because this was observed at a global level by the whole genome DNA methylome analysis (TWGBS) but not using previous approaches (array-based and RRBS) (Ji et al., 2010; Bock et al., 2012), our data suggest that many regulatory regions, including, e.g., distal enhancers critical for gene expression, are only covered by TWGBS analysis.

The observed high overall correlation between RNA and protein levels suggests that posttranscriptional regulation is not a predominant mechanism by which gene expression is regulated in homeostatic HSCs. However, a small number of specific functions in the stem/progenitor compartment may be regulated in

**Figure 7. Expression Landscape of lncRNAs and the Imprinted Gene Network**

(A) Workflow for lncRNA analysis.

(B) Clustering of 79 differentially expressed lncRNAs.

(C) LncRNA expression clusters. Differentially expressed lncRNAs were grouped into 32 clusters based on higher expression (enriched) in one or several cell population(s) compared to mean expression level across all cell populations. Right, an example diagram for each of the most enriched clusters. Average RNA expression ± SD is shown.

(D) Expression of imprinted gene network genes.

(E) Differential DNA methylation of quantified lncRNAs. Top panel, comparison of DMRs between differentially and nondifferentially expressed lncRNAs. Bottom panel, heatmap showing DMRs in red. LncRNAs are ranked by increasing adjusted p value (gray scale). Distances are relative to transcription start sites. *p = 0.01–0.001; **p < 0.001.

(F) Differential methylation of the lncRNA *H19* locus. The *H19* locus shows increasing methylation from HSC (red) to MPP1 (blue), MPP2 (green), and MPP3/4 (brown) in two DMRs. The inset shows a diagram depicting average *H19* RNA expression (mean ± SD).

(G) Differential methylation of *H19* locus at enhancer region 3. Black and white dots represent methylated and unmethylated CpGs, respectively. The red box indicates DMR.

See also Figure S6.

this way. For example, our results point to a potential posttranscriptional regulation of glycolytic metabolism in HSCs via Lin28b-Hmga-Igfbp2 signaling (Shyh-Chang and Daley, 2013), likely because of differences in relative protein synthesis rather than degradation rate (Kristensen et al., 2013; Signer et al., 2014). Moreover, our data suggest that HSC self-renewal and quiescence are regulated by an interplay between the Lin28b-let7-Hmga-Igfbp2 axis, the IGN regulatory network, and HSC-enriched pathways such as Wnt and retinoic acid (RA) signaling. Although some of these pathways are also implicated in embryonic HSC emergence from the hemogenic endothelium (Varrault et al., 2006; Chanda et al., 2013; Copley et al., 2013), the individual players and mechanisms governing HSC function remain to be explored in the embryo and in the adult. As an example, the *Lin28b* target *Igf2bp2* was found to be one of the most differentially expressed transcripts and proteins in HSCs. It is known to modulate expression of the lncRNA *H19*, leading to suppression of proproliferative IGF signaling as well as *Let7* miRNAs and has been suggested to mediate HSC quiescence (Runge et al., 2000; Kallen et al., 2013; Venkatraman et al., 2013). In agreement, our data show increasing H19 enhancer methylation during the transition from HSC to MPP1, providing an explanation for the release of HSCs out of quiescence associated with loss of self-renewal, which might be enhanced further by suppression of the IGN activity (i.e. *p57*; Zou et al., 2011; Tesio and Trumpp, 2011). RA signaling is not only known to be critical for embryonic HSC emergence (Chanda et al., 2013) but also for the regulation of Hox gene expression by chromatin reorganization in embryonic stem cells (Kashyap et al., 2011). Because we observed high expression and low DNA methylation of most members of the HoxA/B clusters in HSCs, it is plausible that RA signaling contributes to the control of the epigenetic landscape of HoxA/B transcription factors in HSCs. Moreover, because many Hox genes are mutated in leukemias, these mechanisms may also be relevant for leukemic stem cells (Alharbi et al., 2013).

An unexpected finding is the degree of alternatively spliced transcript isoforms present in HSCs and their progeny. To date, only rare cases of HSC regulation through alternative splicing have been reported (Bowman et al., 2006). In this study, we identified almost 500 genes with alternative transcript isoform regulation. Although the underlying regulatory mechanism remains unknown, the lncRNA *Malat1*, which is highly expressed in HSCs, has been suggested to be a regulator of alternative splicing (Tripathi et al., 2010). In line with this, *Malat1* has been implicated in multiple types of human cancer (Gutschner et al., 2013), and numerous genetic mutations encoding factors of the splicing machinery have been detected in patients with chronic lymphoid leukemia (Martín-Subero et al., 2013) and myelodysplastic syndrome, a disease derived directly from HSCs (Lindsley and Ebert, 2013; Medyouf et al., 2014). Our catalog of splicing variants, with *Foxj3* as an example of an HSC-specific splice isoform, will serve as a starting point to explore this largely uncharted area of regulation in HSCs and their immediate progeny. In addition to *Malat1*, we identified more than 70 differentially expressed lncRNAs, the vast majority of which have no documented role in hematopoiesis. The variety of molecular functions assigned to lncRNAs is expanding steadily, and their biological roles include regulation of genomic imprinting, differentiation, and self-renewal (Fatica and Bozzoni, 2014).

In summary, the global signatures for stemness and multipotency generated in this study represent not only a comprehensive reference but also suggest distinct areas of stem cell regulation (progressive DNA methylation, alternative splicing, and lncRNAs). This study significantly extends the current understanding of HSC progenitor biology at the global level and provides a solid basis for functional studies exploring the networks responsible for stem cell quiescence, self-renewal, and differentiation.

## EXPERIMENTAL PROCEDURES

### Animals
Eight- to twelve-week-old female C57BL/6 (CD45.2) mice purchased from Harlan Laboratories and B6.SJL-Ptprca Pepcb/BoyJ (CD45.1) animals purchased from Charles River Laboratories were used throughout the study. CD45.1/CD45.2 heterozygotes (F1) for transplant auxiliary bone marrow were bred in-house at the Deutsches Krebsforschungszentrum.

### Bone Marrow Reconstitution Experiments
Fifty (HSC, MPP1) or 2,000 cells (HSC, MPP2, MPP3, MPP4) were FACS-sorted and injected intravenously together with $2 \times 10^5$ supportive bone marrow (BM) cells (CD45.1/2) into lethally irradiated ($2 \times 450$ Gy) (CD45.1) recipient mice. CD45.2-donor cells were monitored at 1, 3–4, 6, 9, 12, and 16 weeks posttransplantation. For secondary transplantations, whole BM was isolated at 16 weeks posttransplant, and $1 \times 10^6$ cells were retransplanted into lethally irradiated CD45.1 recipient mice. Myeloid (CD11b$^+$Gr1$^+$) and lymphoid (T cells, CD4$^+$CD8$^+$; B cells, B220$^+$) lineages were addressed by FACS.

### Proteomic Analysis
FACS-sorted HSCs and MPP1 ($4 \times 10^5$) were lysed, and proteins were extracted, reduced/alkylated, and digested with trypsin. Peptides were labeled differentially with stable isotope dimethyl labeling on a column as described previously (Boersema et al., 2009) and fractionated by OFFGEL isoelectric focusing (Agilent Technologies). In technical duplicates, peptides were separated by nanoflow ultra-high-performance liquid chromatography on a 120 min gradient and analyzed by electrospray ionization-tandem mass spectrometry (ESI-MS/MS) on a linear trap quadrupole Orbitrap Velos or Orbitrap Velos Pro (Thermo Fisher Scientific). MS raw data files were processed with MaxQuant (version 1.3.0.5) (Cox and Mann, 2008). The derived peak list was searched using the built-in Andromeda search engine (version 1.3.0.5) in MaxQuant against the Uniprot mouse database (2013.02.20). A 1% FDR was required at both the protein level and the peptide level. Differential expression was assessed using the Limma package in R/Bioconductor (Gentleman et al., 2004; Smyth, 2004), and proteins with an adjusted p value of less than 0.1 were considered expressed differentially between HSC and MPP1.

### RNA-seq
Total RNA isolation was performed using an ARCTURUS PicoPure RNA isolation kit (Life Technologies, Invitrogen) according to the manufacturer's instructions. Total RNA was used for quality controls and for normalization of the starting material (Figure S3). cDNA libraries were generated with 10 ng of total RNA for HSC-MPP using the SMARTer Ultra Low RNA kit for Illumina sequencing (Clontech Laboratories) according to the manufacturer's indications. Sequencing was performed with a HiSeq2000 device (Illumina) and one sample per lane. Sequenced read fragments were mapped to the mouse reference genome GRCm38 (ENSEMBL release 69) using the Genomic Short-Read Nucleotide Alignment program (version 2012-07-20). DESeq2 (Love et al., 2014) and DEXSeq (Anders et al., 2012) were used to test for differential expression (FDR = 0.1) and differential exon use, respectively.

### TWGBS
DNA methylation analysis using TWGBS (Adey and Shendure, 2012) was performed as described previously (Wang et al., 2013). Genomic mouse DNA (10-30 ng) was used as input, and each sequencing library was subjected

to 101 bp paired-end sequencing on a single lane of a HiSeq2000 device (Illumina).

Animal procedures were performed according to protocols approved by the German authorities (Regierungspräsidium Karlsruhe [no. Z110/02, DKFZ nos. 261, G175-12, and G140-13]).

## ACCESSION NUMBERS

Proteome data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the Proteomics Identifications Database (PRIDE) partner repository (Vizcaíno et al., 2013) with the data set identifier PXD000572. RNA-seq data are available in the ArrayExpress database (http://www.ebi.ac.uk/arrayexpress) under accession number E-MTAB-2262. TWGBS data can be accessed under Gene Expression Omnibus under accession no. GSE52709.

## SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, six tables, one item and two online resources can be found with this article online at http://dx.doi.org/10.1016/j.stem.2014.07.005.

## REFERENCES

Adey, A., and Shendure, J. (2012). Ultra-low-input, tagmentation-based whole-genome bisulfite sequencing. Genome Res. 22, 1139–1143.

Adolfsson, J., Månsson, R., Buza-Vidas, N., Hultquist, A., Liuba, K., Jensen, C.T., Bryder, D., Yang, L., Borge, O.-J., Thoren, L.A., et al. (2005). Identification of Flt3+ lympho-myeloid stem cells lacking erythro-megakaryocytic potential a revised road map for adult blood lineage commitment. Cell 121, 295–306.

Alharbi, R.A., Pettengell, R., Pandha, H.S., and Morgan, R. (2013). The role of HOX genes in normal hematopoiesis and acute leukemia. Leukemia 27, 1000–1008.

Anders, S., Reyes, A., and Huber, W. (2012). Detecting differential usage of exons from RNA-seq data. Genome Res. 22, 2008–2017.

Arab, K., Park, Y.J., Lindroth, A.M., Schäfer, A., Oakes, C., Weichenhan, D., Lukanova, A., Lundin, E., Risch, A., Meister, M., et al. (2014). Long Noncoding RNA TARID Directs Demethylation and Activation of the Tumor Suppressor TCF21 via GADD45A. Mol. Cell, in press. Published online July 30, 2014. http://dx.doi.org/10.1016/j.molcel.2014.06.031.

Argiropoulos, B., and Humphries, R.K. (2007). Hox genes in hematopoiesis and leukemogenesis. Oncogene 26, 6766–6776.

Bakker, S.T., and Passegué, E. (2013). Resilient and resourceful: genome maintenance strategies in hematopoietic stem cells. Exp. Hematol. 41, 915–923.

Battista, S., Pentimalli, F., Baldassarre, G., Fedele, M., Fidanza, V., Croce, C.M., and Fusco, A. (2003). Loss of Hmga1 gene function affects embryonic stem cell lympho-hematopoietic differentiation. FASEB J. 17, 1496–1498.

Benveniste, P., Frelin, C., Janmohamed, S., Barbara, M., Herrington, R., Hyam, D., and Iscove, N.N. (2010). Intermediate-term hematopoietic stem cells with extended but time-limited reconstitution potential. Cell Stem Cell 6, 48–58.

Berg, J.S., Lin, K.K., Sonnet, C., Boles, N.C., Weksberg, D.C., Nguyen, H., Holt, L.J., Rickwood, D., Daly, R.J., and Goodell, M.A. (2011). Imprinted genes that regulate early mammalian growth are coexpressed in somatic stem cells. PLoS ONE 6, e26410.

Bock, C., Beerman, I., Lien, W.H., Smith, Z.D., Gu, H., Boyle, P., Gnirke, A., Fuchs, E., Rossi, D.J., and Meissner, A. (2012). DNA methylation dynamics during in vivo differentiation of blood and skin stem cells. Mol. Cell 47, 633–647.

Boersema, P.J., Raijmakers, R., Lemeer, S., Mohammed, S., and Heck, A.J. (2009). Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. Nat. Protoc. 4, 484–494.

Bowman, T.V., McCooey, A.J., Merchant, A.A., Ramos, C.A., Fonseca, P., Poindexter, A., Bradfute, S.B., Oliveira, D.M., Green, R., Zheng, Y., et al. (2006). Differential mRNA processing in hematopoietic stem cells. Stem Cells 24, 662–670.

Chanda, B., Ditadi, A., Iscove, N.N., and Keller, G. (2013). Retinoic acid signaling is essential for embryonic hematopoietic stem cell development. Cell 155, 215–227.

Cleynen, I., Brants, J.R., Peeters, K., Deckers, R., Debiec-Rychter, M., Sciot, R., Van de Ven, W.J.M., and Petit, M.M.R. (2007). HMGA2 regulates transcription of the Imp2 gene via an intronic regulatory element in cooperation with nuclear factor-kappaB. Mol. Cancer Res. 5, 363–372.

Copley, M.R., Babovic, S., Benz, C., Knapp, D.J.H.F., Beer, P.A., Kent, D.G., Wohrer, S., Treloar, D.Q., Day, C., Rowe, K., et al. (2013). The Lin28b-let-7-Hmga2 axis determines the higher self-renewal potential of fetal haematopoietic stem cells. Nat. Cell Biol. 15, 916–925.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. 26, 1367–1372.

Dumon, S., Walton, D.S., Volpe, G., Wilson, N., Dassé, E., Del Pozzo, W., Landry, J.-R., Turner, B., O'Neill, L.P., Göttgens, B., and Frampton, J. (2012). Itga2b regulation at the onset of definitive hematopoiesis and commitment to differentiation. PLoS ONE 7, e43300.

Ema, H., Morita, Y., Yamazaki, S., Matsubara, A., Seita, J., Tadokoro, Y., Kondo, H., Takano, H., and Nakauchi, H. (2006). Adult mouse hematopoietic stem cells: purification and single-cell assays. Nat. Protoc. 1, 2979–2987.

Essers, M.A., Offner, S., Blanco-Bose, W.E., Waibler, Z., Kalinke, U., Duchosal, M.A., and Trumpp, A. (2009). IFNalpha activates dormant haematopoietic stem cells in vivo. Nature 458, 904–908.

Fatica, A., and Bozzoni, I. (2014). Long non-coding RNAs: new players in cell differentiation and development. Nat. Rev. Genet. 15, 7–21.

Forsberg, E.C., Serwold, T., Kogan, S., Weissman, I.L., and Passegué, E. (2006). New evidence supporting megakaryocyte-erythrocyte potential of flk2/flt3+ multipotent hematopoietic progenitors. Cell 126, 415–426.

Gabut, M., Samavarchi-Tehrani, P., Wang, X., Slobodeniuc, V., O'Hanlon, D., Sung, H.K., Alvarez, M., Talukder, S., Pan, Q., Mazzoni, E.O., et al. (2011). An alternative splicing switch regulates embryonic stem cell pluripotency and reprogramming. Cell 147, 132–146.

Gao, X., Johnson, K.D., Chang, Y.I., Boyer, M.E., Dewey, C.N., Zhang, J., and Bresnick, E.H. (2013). Gata2cis-element is required for hematopoietic stem cell generation in the mammalian embryo. J. Exp. Med. 210, 2833–2842.

Gazit, R., Garrison, B.S., Rao, T.N., Shay, T., Costello, J., Ericson, J., Kim, F., Collins, J.J., Regev, A., Wagers, A.J., et al. (2013). Transcriptome analysis identifies regulators of hematopoietic stem and progenitor cells. Stem Cell Rev. 1, 1–15.

Gekas, C., and Graf, T. (2013). CD41 expression marks myeloid-biased adult hematopoietic stem cells and increases with age. Blood 121, 4463–4472.

Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., et al. (2004). Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 5, R80.

Gutschner, T., Hämmerle, M., and Diederichs, S. (2013). MALAT1 — a paradigm for long noncoding RNA function in cancer. J. Mol. Med. 91, 791–801.

Ji, H., Ehrlich, L.I., Seita, J., Murakami, P., Doi, A., Lindau, P., Lee, H., Aryee, M.J., Irizarry, R.A., Kim, K., et al. (2010). Comprehensive methylome map of lineage commitment from haematopoietic progenitors. Nature 467, 338–342.

Kallen, A.N., Zhou, X.B., Xu, J., Qiao, C., Ma, J., Yan, L., Lu, L., Liu, C., Yi, J.S., Zhang, H., et al. (2013). The imprinted H19 lncRNA antagonizes let-7 microRNAs. Mol. Cell 52, 101–112.

Kashyap, V., Gudas, L.J., Brenet, F., Funk, P., Viale, A., and Scandura, J.M. (2011). Epigenomic reorganization of the clustered Hox genes in embryonic stem cells induced by retinoic acid. J. Biol. Chem. 286, 3250–3260.

Kent, D.G., Copley, M.R., Benz, C., Wöhrer, S., Dykstra, B.J., Ma, E., Cheyne, J., Zhao, Y., Bowie, M.B., Zhao, Y., et al. (2009). Prospective isolation and molecular characterization of hematopoietic stem cells with durable self-renewal potential. Blood 113, 6342–6350.

Klimmeck, D., Hansson, J., Raffel, S., Vakhrushev, S.Y., Trumpp, A., and Krijgsveld, J. (2012). Proteomic cornerstones of hematopoietic stem cell differentiation: distinct signatures of multipotent progenitors and myeloid committed cells. Mol. Cell. Peoteomics 11, 1–54.

Koch, U., Wilson, A., Cobas, M., Kemler, R., Macdonald, H.R., and Radtke, F. (2008). Simultaneous loss of beta- and gamma-catenin does not perturb hematopoiesis or lymphopoiesis. Blood 111, 160–164.

Kristensen, A.R., Gsponer, J., and Foster, L.J. (2013). Protein synthesis rate is the predominant regulator of protein expression during differentiation. Mol. Syst. Biol. 9, 689.

Lindsley, R.C., and Ebert, B.L. (2013). Molecular pathophysiology of myelodysplastic syndromes. Annu. Rev. Pathol. 8, 21–47.

López, A.J. (1995). Developmental role of transcription factor isoforms generated by alternative splicing. Dev. Biol. 172, 396–411.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. BioRxiv. Preprint. Published online February 19, 2014. http://dx.doi.org/10.1101/002832.

Luis, T.C., Ichii, M., Brugman, M.H., Kincade, P., and Staal, F.J. (2012). Wnt signaling strength regulates normal hematopoiesis and its deregulation is involved in leukemia development. Leukemia 26, 414–421.

Månsson, R., Hultquist, A., Luc, S., Yang, L., Anderson, K., Kharazi, S., Al-Hashmi, S., Liuba, K., Thorén, L., Adolfsson, J., et al. (2007). Molecular evidence for hierarchical transcriptional lineage priming in fetal and adult stem cells and multipotent progenitors. Immunity 26, 407–419.

Martín-Subero, J.I., López-Otín, C., and Campo, E. (2013). Genetic and epigenetic basis of chronic lymphocytic leukemia. Curr. Opin. Hematol. 20, 362–368.

McKinney-Freeman, S., Cahan, P., Li, H., Lacadie, S.A., Huang, H.-T., Curran, M., Loewer, S., Naveiras, O., Kathrein, K.L., Konantz, M., et al. (2012). The transcriptional landscape of hematopoietic stem cell ontogeny. Cell Stem Cell 11, 701–714.

Medyouf, H., Mossner, M., Jann, J.C., Nolte, F., Raffel, S., Herrmann, C., Lier, A., Eisen, C., Nowak, V., Zens, B., et al. (2014). Myelodysplastic cells in patients reprogram mesenchymal stromal cells to establish a transplantable stem cell niche disease unit. Cell Stem Cell 14, 824–837.

Morishita, A., Zaidi, M.R., Mitoro, A., Sankarasharma, D., Szabolcs, M., Okada, Y., D'Armiento, J., and Chada, K. (2013). HMGA2 is a driver of tumor metastasis. Cancer Res. 73, 4289–4299.

Orkin, S.H., and Zon, L.I. (2008). Hematopoiesis: an evolving paradigm for stem cell biology. Cell 132, 631–644.

Osawa, M., Hanada, K., Hamada, H., and Nakauchi, H. (1996). Long-term lymphohematopoietic reconstitution by a single CD34-low/negative hematopoietic stem cell. Science 273, 242–245.

Paralkar, V.R., and Weiss, M.J. (2013). Long noncoding RNAs in biology and hematopoiesis. Blood 121, 4842–4846.

Pruitt, K.D., Tatusova, T., Klimke, W., and Maglott, D.R. (2009). NCBI Reference Sequences: current status, policy and new initiatives. Nucleic Acids Res. 37, D32–D36.

Purton, L.E., and Scadden, D.T. (2007). Limiting factors in murine hematopoietic stem cell assays. Cell Stem Cell 1, 263–270.

Rinn, J.L., and Chang, H.Y. (2012). Genome regulation by long noncoding RNAs. Annu. Rev. Biochem. 81, 145–166.

Rosenbauer, F., Wagner, K., Kutok, J.L., Iwasaki, H., Le Beau, M.M., Okuno, Y., Akashi, K., Fiering, S., and Tenen, D.G. (2004). Acute myeloid leukemia induced by graded reduction of a lineage-specific transcription factor, PU.1. Nat. Genet. 36, 624–630.

Rossi, L., Lin, K.K., Boles, N.C., Yang, L., King, K.Y., Jeong, M., Mayle, A., and Goodell, M.A. (2012). Less is more: unveiling the functional core of hematopoietic stem cells through knockout mice. Cell Stem Cell 11, 302–317.

Runge, S., Nielsen, F.C., Nielsen, J., Lykke-Andersen, J., Wewer, U.M., and Christiansen, J. (2000). H19 RNA binds four molecules of insulin-like growth factor II mRNA-binding protein. J. Biol. Chem. 275, 29562–29569.

Sanjuan-Pla, A., Macaulay, I.C., Jensen, C.T., Woll, P.S., Luis, T.C., Mead, A., Moore, S., Carella, C., Matsuoka, S., Bouriez Jones, T., et al. (2013). Platelet-biased stem cells reside at the apex of the haematopoietic stem-cell hierarchy. Nature 502, 232–236.

Sebastian, S., Faralli, H., Yao, Z., Rakopoulos, P., Palii, C., Cao, Y., Singh, K., Liu, Q.C., Chu, A., Aziz, A., et al. (2013). Tissue-specific splicing of a ubiquitously expressed transcription factor is essential for muscle differentiation. Genes Dev. 27, 1247–1259.

Seita, J., and Weissman, I.L. (2010). Hematopoietic stem cell: self-renewal versus differentiation. Wiley Interdiscip. Rev. Syst. Biol. Med. 2, 640–653.

Shah, S.N., Cope, L., Poh, W., Belton, A., Roy, S., Talbot, C.C., Jr., Sukumar, S., Huso, D.L., and Resar, L.M. (2013). HMGA1: a master regulator of tumor progression in triple-negative breast cancer cells. PLoS ONE 8, e63419.

Shen, Y., Yue, F., McCleary, D.F., Ye, Z., Edsall, L., Kuan, S., Wagner, U., Dixon, J., Lee, L., Lobanenkov, V.V., and Ren, B. (2012). A map of thecis-regulatory sequences in the mouse genome. Nature 488, 116–120.

Shyh-Chang, N., and Daley, G.Q. (2013). Lin28: primal regulator of growth and metabolism in stem cells. Cell Stem Cell 12, 395–406.

Shyh-Chang, N., Zhu, H., Yvanka de Soysa, T., Shinoda, G., Seligson, M.T., Tsanov, K.M., Nguyen, L., Asara, J.M., Cantley, L.C., and Daley, G.Q. (2013). Lin28 enhances tissue repair by reprogramming cellular metabolism. Cell 155, 778–792.

Signer, R.A., Magee, J.A., Salic, A., and Morrison, S.J. (2014). Haematopoietic stem cells require a highly regulated protein synthesis rate. Nature 509, 49–54.

Smyth, G.K. (2004). Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. Stat. Appl. Genet. Mol. Biol. *3*, 1–25.

Stamatoyannopoulos, J.A., Snyder, M., Hardison, R., Ren, B., Gingeras, T., Gilbert, D.M., Groudine, M., Bender, M., Kaul, R., Canfield, T., et al.; Mouse ENCODE Consortium (2012). An encyclopedia of mouse DNA elements (Mouse ENCODE). Genome Biol. *13*, 418.

Sugimura, R., He, X.C., Venkatraman, A., Arai, F., Box, A., Semerad, C., Haug, J.S., Peng, L., Zhong, X.B., Suda, T., and Li, L. (2012). Noncanonical Wnt signaling maintains hematopoietic stem cells in the niche. Cell *150*, 351–365.

Takubo, K., Nagamatsu, G., Kobayashi, C.I., Nakamura-Ishizu, A., Kobayashi, H., Ikeda, E., Goda, N., Rahimi, Y., Johnson, R.S., Soga, T., et al. (2013). Regulation of glycolysis by Pdk functions as a metabolic checkpoint for cell cycle quiescence in hematopoietic stem cells. Cell Stem Cell *12*, 49–61.

Taneri, B., Snyder, B., Novoradovsky, A., and Gaasterland, T. (2004). Alternative splicing of mouse transcription factors affects their DNA-binding domain architecture and is tissue specific. Genome Biol. *5*, R75.

Tesio, M., and Trumpp, A. (2011). Breaking the cell cycle of HSCs by p57 and friends. Cell Stem Cell *9*, 187–192.

Tew, K.D., and Townsend, D.M. (2012). Glutathione-s-transferases as determinants of cell survival and death. Antioxid. Redox Signal. *17*, 1728–1737.

Tickenbrock, L., Hehn, S., Sargin, B., Choudhary, C., Bäumer, N., Buerger, H., Schulte, B., Müller, O., Berdel, W.E., Müller-Tidow, C., and Serve, H. (2008). Activation of Wnt signalling in acute myeloid leukemia by induction of Frizzled-4. Int. J. Oncol. *33*, 1215–1221.

Tripathi, V., Ellis, J.D., Shen, Z., Song, D.Y., Pan, Q., Watt, A.T., Freier, S.M., Bennett, C.F., Sharma, A., Bubulya, P.A., et al. (2010). The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. Mol. Cell *39*, 925–938.

Trumpp, A., Essers, M., and Wilson, A. (2010). Awakening dormant haematopoietic stem cells. Nat. Rev. Immunol. *10*, 201–209.

Varrault, A., Gueydan, C., Delalbre, A., Bellmann, A., Houssami, S., Aknin, C., Severac, D., Chotard, L., Kahli, M., Le Digarcher, A., et al. (2006). Zac1 regulates an imprinted gene network critically involved in the control of embryonic growth. Dev. Cell *11*, 711–722.

Venkatraman, A., He, X.C., Thorvaldsen, J.L., Sugimura, R., Perry, J.M., Tao, F., Zhao, M., Christenson, M.K., Sanchez, R., Yu, J.Y., et al. (2013). Maternal imprinting at the H19-Igf2 locus maintains adult haematopoietic stem cell quiescence. Nature *500*, 345–349.

Vizcaíno, J.A., Côté, R.G., Csordas, A., Dianes, J.A., Fabregat, A., Foster, J.M., Griss, J., Alpi, E., Birim, M., Contell, J., et al. (2013). The PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013. Nucleic Acids Res. *41*, D1063–D1069.

Wang, Q., Gu, L., Adey, A., Radlwimmer, B., Wang, W., Hovestadt, V., Bähr, M., Wolf, S., Shendure, J., Eils, R., et al. (2013). Tagmentation-based whole-genome bisulfite sequencing. Nat. Protoc. *8*, 2022–2032.

Weissman, I.L., and Shizuru, J.A. (2008). The origins of the identification and isolation of hematopoietic stem cells, and their capability to induce donor-specific transplantation tolerance and treat autoimmune diseases. Blood *112*, 3543–3553.

Wilson, A., Laurenti, E., Oser, G., van der Wath, R.C., Blanco-Bose, W., Jaworski, M., Offner, S., Dunant, C.F., Eshkind, L., Bockamp, E., et al. (2008). Hematopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair. Cell *135*, 1118–1129.

Wilson, A., Laurenti, E., and Trumpp, A. (2009). Balancing dormant and self-renewing hematopoietic stem cells. Curr. Opin. Genet. Dev. *19*, 461–468.

Yaniv, K., and Yisraeli, J.K. (2002). The involvement of a conserved family of RNA binding proteins in embryonic development and carcinogenesis. Gene *287*, 49–54.

Zhang, Q., Chen, C.Y., Yedavalli, V.S., and Jeang, K.T. (2013). NEAT1 long noncoding RNA and paraspeckle bodies modulate HIV-1 posttranscriptional expression. mBio. *4*, e00596–00512.

Zou, P., Yoshihara, H., Hosokawa, K., Tai, I., Shinmyozu, K., Tsukahara, F., Maru, Y., Nakayama, K., Nakayama, K.I., and Suda, T. (2011). p57(Kip2) and p27(Kip1) cooperate to maintain hematopoietic stem cell quiescence through interactions with Hsc70. Cell Stem Cell *9*, 247–261.