Molecular landscapes of human haematopoietic stem cells in health and leukaemia.

Aditi Vedi [a,b,c], Antonella Santoro [a,b], Cyrille F. Dunant [d], John E. Dick [e-f] and Elisa Laurenti [a,b]

Affiliations:

[a] Wellcome Trust – Medical Research Council Cambridge Stem Cell Institute, University of Cambridge; [b] Department of Haematology, University of Cambridge, Cambridge, CB2 0AH; United Kingdom; [c] Paediatric Oncology, Royal Marsden Hospital, Sutton, London, United Kingdom; [d] Ecole Polytechnique Fédérale de Lausanne, LMC, Station 12, Lausanne, CH-1015, Switzerland; [e] Princess Margaret Cancer Centre, University Health Network, Toronto, Ontario, M5G 1L7, Canada, [f] Department of Molecular Genetics, University of Toronto, Toronto, Ontario, M5S 1A8, Canada.

Corresponding author:

Dr Elisa Laurenti

Wellcome Trust – Medical Research Council Cambridge Stem Cell Institute, University of Cambridge

The Clifford Allbutt building, Hills Road, Cambridge, CB2 0AH

Phone: +44 (0) 1223 762044;

el422@cam.ac.uk

Short title: Human haematopoietic stem cells molecular circuits

Key words: Haematopoietic stem cells; Acute Myeloid Leukemia; transcriptomics; epigenetics

Abstract

Blood cells are organized as a hierarchy with haematopoietic stem cells (HSCs) at its root. The advent of genomic technologies has opened the way for global characterization of the molecular landscape of haematopoietic stem cells and their progeny both in mouse and human models at the genetic, transcriptomic, epigenetic and proteomics level. Here we outline our current understanding of the molecular programmes that govern human HSCs and how dynamic changes occurring during HSC differentiation are necessary for well-regulated blood formation under homeostasis and upon injury. A large body of evidence is accumulating on how these programmes of normal haematopoiesis are modified in acute myeloid leukemia (AML), an aggressive adult malignancy, driven by leukemic stem cells (LSCs). Here we summarise these findings and their clinical implications.

Introduction

Haematopoietic Stem Cells (HSC) exist at the apex of a highly coordinated hierarchy of blood cells. They are the only blood cells that possess self-renewal, the ability to produce more of themselves, as well as multipotency, the capacity to differentiate into all cell types of the blood. Another important property of HSCs is their quiescence: they mostly reside outside of the cell cycle in G_0 and divide very infrequently[1–3]. Quiescence is actively regulated intrinsically and extrinsically [4], and is thought to protect HSC from exhaustion and prevent accumulation of mutations with malignant potential. HSC are also characterised by a unique metabolic programme [5] and their activity is fine-tuned by interaction with their microenvironment, the so-called HSC niche[6] .

To date, almost all of our understanding and characterisation of HSC, at the functional and phenotypic level, are derived from mouse models, but detailed studies of human HSC (hHSC) are starting to emerge. Because of the slow development of robust functional assays to test for hHSC and progenitor function, isolation strategies of these cells have lagged behind their mouse counterparts (reviewed in [7]). However the past decade has brought a number of novel tools that have allowed mapping the precise cellular roadmap of hHSC differentiation. The differentiation capacity of most human progenitor populations has now been described at single cell resolution[8–11]. Furthermore, hHSC can now be purified to a level sufficient to make single cell transplantation possible [12]. A number of studies have applied genome-wide techniques to start investigating the molecular circuits of human HSC and progenitor cells based on several purification strategies (Table 1). Here we summarise what has been discovered about the molecular landscape of human HSC and their progeny. Finally we discuss how the normal HSC and progenitor molecular programs are subverted in leukaemia and in particular in Leukaemic Stem Cells (LSC), and how this information can be exploited for therapeutic purposes.

Transcriptional control of human haematopoietic stem cells and their progeny

The transcriptional status of a cell is highly predictive of its cellular identity. A number of studies have thus aimed to derive transcriptional signatures of hHSC, isolated at different degrees of purity, by comparing their genome-wide transcriptional profiles to those of their early progeny and/or more differentiated cells, either by microarray technology[13–15] or more recently by RNA-seq[16]. All gene expression studies consistently found that hHSC isolated from cord blood have a distinct transcriptional programme compared to all mature haematopoietic cells, similar to mouse HSCs isolated from bone marrow[17,18]. Despite this clear dichotomy between HSC and progenitor transcriptional profiles, gene expression during haematopoiesis follows a modular architecture. Even though HSC programmes are drastically rearranged during differentiation, up to 80 common gene expression modules are used by distinct populations at different stages of differentiation[13].

Zooming in at high resolution on the very first steps of hHSC differentiation enables three main conclusions concerning their molecular make-up. First, it is clear that HSCs have a unique transcriptional programme even when compared to lineage-restricted progenitors (Figure 1). Almost 70% of the genes differentially expressed throughout early haematopoiesis are significantly changed upon exit from the HSC compartment[15]. Correspondingly, Gene Ontology analysis shows that almost all pathways and signalling cascades are significantly altered when HSCs differentiate. Functionally, this translates into a unique HSC metabolic programme[5], but also distinct responses of HSC and progenitors to a number of external stimuli, including responses to stresses such as DNA damage[19,20] and unfolded protein response stress [21,22].

Second, in contrast to other progenitor cells that are transcriptionally distinct from hHSC such as common myeloid progenitors (CMP), myeloerythroid progenitors (MEP) and granulocyte-macrophage progenitors (GMP), multilymphoid progenitors (MLP) share many transcriptional features of HSC. This similarity is particularly strong for transcription factor (TF) expression[15] suggesting that there is a common TF module that maintains self-renewal and

differentiation potential. Loss of function studies of EBF1, or of transcription factors acting on the self-sustaining EBF-1/PAX5 loop (IKZF, BCL11A, SOX4 or TEAD1) have shown the context-dependent effects of this circuitry: in MLP, these TF push differentiation towards the B cell lineage[15,23] whereas in HSCs they repress self-renewal [24,25].

A third important finding regarding the transcriptional differences between functionally distinct subsets of hHSC is that human LT- and ST-HSC/MPP differ in their repopulation capacity and in their cell cycle properties. hLT-HSC are capable of reconstituting a complete human haematopoietic system upon xenotransplantation, while hST-HSC only produce transient grafts. Surprisingly, very few genes distinguish hLT-HSC from hST-HSC[12,15,16], when these cells are isolated in a homeostatic quiescent state. One possibility may be that the distinct self-renewal capacity of these HSC subsets is regulated as discussed below by long non-coding RNA (lnc-RNA), usage of alternative splicing events or post-transcriptionally. While no proteomic analysis has been carried out at this level of resolution in human, Cabezas-Wallscheid et al. found that only 47 proteins were differentially expressed between mouse HSC and multipotent progenitors [26], suggesting that distinct hHSC subsets are also very likely to be extremely similar at the protein level. An alternative explanation may come from the quiescent state of these cells. Even though hLT- and hST-HSCs are equally quiescent under homeostatic conditions, their cell cycle behaviour is very different when they are challenged to divide[27]. Correspondingly hLT- and hST-HSCs diverge transcriptionally when cycling in vivo. When integrating the data at homeostasis and under activated cycling conditions, hLT- and hST-HSCs were found to differ in genes related to cell cycle, chromatin remodelling and immune response [27], categories very similar to those reported to differ between equivalent mouse HSC subsets [26]. These data also show that cell cycle status not only plays an important role in defining the identity of functionally distinct HSC subsets, but also determines their transcriptional output. In mouse models, several studies have recently used single cell RNA-seq to provide further insights into heterogeneity within the HSC pool [28–30]. In human, very little is known besides the characterization of HSC subsets based on CD90 [8], CD49f [12] or cKit

expression [31]. Future studies combining usage of novel cell surface markers, single cell transcriptomics and functional assays will help provide a more comprehensive understanding of hHSC heterogeneity.

Even though all findings at the gene level are highly consistent between studies independent of the technology used [15,16], the advent of RNA-seq has uncovered several novel levels of HSC regulation. hHSC are characterised by high levels of non-coding transcripts, in particular long non-coding RNAs (lncRNAs), which change significantly during their differentiation[16]. More than a hundred lncRNAs are also enriched in mouse HSC [26,32]. In addition, Chen et al.[16] identified a number of novel splice junctions that are differentially used in the haematopoietic tree, and catalogued 2301 alternative splicing events that occur across haematopoiesis. Many of these lead to gain or loss of functional domains, suggesting that modulation of alternative splicing may play an important role to determine protein function during differentiation of haematopoietic cells. Importantly, both lncRNAs and splicing events are highly species-specific and have been shown to fine tune expression and differentiation capacity in other stem cell systems [33–35]. Hence it will be interesting to verify to what extent this novel layer of regulation contributes to the subtle but important differences in HSC function observed between mice and humans[36] .

While the studies discussed above used hHSC isolated from cord blood, because of its ease of access and abundance compared to adult haematopoietic organs, adult bone marrow hHSC and progenitor fractions gene expression profiles have also been generated [10,14,37]. No direct bioinformatic comparison of cord blood and bone marrow HSC has been reported to date. However, Pang et al. compared the gene expression profiles of hHSC isolated from young and elderly healthy donors and found that aged hHSC downregulate genes linked to lymphoid specification such as FLT3 and SOX4, but also upregulate genes of the MAPK, G-CSF and DNA repair pathways [37], consistent with the myeloid-bias [37,38] and increased replication stress observed in elderly HSC [39]. In summary, hHSC are characterized by unique molecular programmes that are driven by specific gene,

splicing isoform and lncRNA expression, which globally determine a unique stemness pathway and signalling phenotype (Figure 1).

Epigenetic regulation of normal haematopoiesis

Epigenetic regulation, via histone modification and DNA methylation, is key to determining cell fate in many systems[16]. While the understanding of the epigenetic mechanisms underpinning HSC regulation comes principally from studies with transgenic mouse models [40,41], recently there have also been reports of how the epigenetic landscape changes when HSCs differentiate and upon aging. In mice, DNA methylation decreases on lineage-specific gene-regulatory elements as blood and skin stem cells differentiate [40–42]. In keeping with this finding, an array-based method found that CB-derived differentiated myeloid cells are globally hypomethylated at promoters compared to CD34+ human stem and progenitor cells (HSPCs), also derived from CB [43]. During aging, global hypo-methylation accompanied by de novo methylation at specific and restricted sites has been observed in many tissues. Accordingly, HSPCs isolated from mobilized peripheral blood of young adults, were less methylated than their CB counterparts, especially at differentiation genes, and also displayed some examples of de novo methylation. A much finer picture of changes in methylation with aging is derived from studies of mouse haematopoiesis. In contrast to other cell types, HSCs from old mice display higher levels of methylation at CpG islands[40,44]. However methylation changes are highly site-specific: TF binding sites associated with lineage restriction are generally hypermethylated compared to young animals, while loci associated with HSC maintenance are hypomethylated[44]. De novo methylation occurs predominantly in genes associated with lineage specification in aged HSCs, but only changes their expression in their downstream progeny[40]. In addition the proportion of 5-hydroxymethycytosine (5hmC) in HSCs decreases with age [44]. It will be important to verify whether similar changes are happening in elderly humans.

Chromatin marks are also dynamically changed in haematopoiesis. In both mice and humans, the HSC compartment displays high levels of histone methylation (H3K4me3) that is lost at many bivalent loci upon differentiation[44–46]. Using a chromatin immunoprecipitation system optimized to analyse a few hundred cells, Lara-Astasio et al. showed that during mouse haematopoiesis poised enhancers and open chromatin are established together mainly in progenitor cells to establish transcriptional programs that are only then executed in differentiated cells [47]. This makes progenitor cells, and not HSC, the most complex and dynamic populations in terms of chromatin modifications. As sequence changes in enhancers are among the major drivers of phenotypic evolution [48], similar studies in humans will prove important to understand human blood formation both in normal and malignant contexts. Chromatin marks in haematopoietic cells also change with age: many bivalent domains are also lost in aged mouse HSC, but overall there is a gain due to addition of H3K27me3 marks to regions uniquely marked by H3K4me3 in young HSC [44]. , While whole-genome DNA and histone methylation analysis has not been carried out on highly purified human HSC of any age group, increased incidence of mutations in epigenetic regulators (DNMT3A, TET2, IDH1, IDH2) is seen with age in humans [49–53]. It is thus very likely that epigenetic marks are also globally altered with time in humans and may contribute to the decline in HSC function.

Perturbation of HSC programmes in Acute Myeloid Leukemia

The cancer stem cell (CSC) model postulates that cancers are organised hierarchically with a subpopulation of CSCs at the apex, which have self-renewal capacity and can perpetuate cancer cell differentiation and progression to relapse [54]. It is now widely accepted that this model is applicable to most types of leukaemia and some solid malignancies [55,56]. Similarly to the normal haematopoietic system, Leukemic Stem Cells (LSCs) are considered to reside at the apex of a hierarchy of leukemic blood cells that gradually lose self-renewal capacity while becoming more differentiated and acquiring the abnormal blast phenotype. Relapsed disease is thought to be related to survival of some LSC

after induction chemotherapy, as LSC are thought to be quiescent and thus inherently resistant to most forms of chemotherapy [56–58]. In the last few years there has been a strong effort to identify the functional and molecular characteristics of LSCs that distinguish them from the bulk non-LSC leukemic cells as well as comparing LSC with normal HSCs and progenitor cells, with the goal to develop therapies targeted to LSC eradication in the clinical setting. Here we summarise the findings from studies at the genome-wide level in Acute Myeloid Leukaemia (AML).

The hierarchal structure of AML and the characterization of AML LSCs, both in mouse models and human primary samples, represent one of the best studied CSC systems [57,59]. Several studies have highlighted a hierarchy within AML, with LSCs enriched in the CD34+CD38- fraction, giving rise to CD34+CD38+ leukemia progenitor cells, which further differentiate into CD34- leukemic blasts [10,54,57]. This led to the idea that AML LSCs may be restricted to the CD34+ compartment, however with improved xenograft assays and broader patient sample analysis, it is now clear that there is no unique cell surface marker based definition of LSC [10,60]. While the majority of AML cases are CD34+, a significant proportion of primary AMLs are CD34-, with LSC activity shown to be present in both the CD34+ and CD34- compartments independently of CD34 status [61–63]. Hence, a functional definition is more reliable for the identification of LSC populations in individual AML patients. This is most commonly achieved by testing LSCs' capacity to re-establish disease in immunodeficient mice[57]. With this in mind, the relationship between normal hHSC and AML LSC has now been investigated at the transcriptional, proteomic and epigenetic level, with the goal being to capture a common LSC molecular state in AML independent of cell surface markers.

A number of transcriptional signatures of AML LSCs have been defined in the past few years, with different experimental strategies (Table 2). The first LSC signatures were established, comparing HSC and progenitor fractions isolated from healthy donors to the fractions with the same cell surface phenotype in AML patients [14,64–66]. In contrast, Eppert et al[60] defined LSC-containing fractions

in 16 primary human adult AML samples using xenotransplantation. They then used bioinformatics to compare functionally defined LSC-enriched fractions to fractions devoid of transplantation potential to derive an LSC-specific signature. This LSC-specific signature most closely resembles that of healthy HSCs rather than progenitor cells. In another study, Goardon et al. immunophenotypically defined the most expanded "normal" subpopulations and verified the presence of LSCs in 100 CD34+ AML samples, the most abundant type of AML. They found that most samples contain an MPP-like and GMP-like LSC. In most cases both fractions from the same patient can engraft in xenografts. They then compared the gene expression profiles of each of these fractions from 27 AML patients to age-matched normal HSC and progenitor compartments. They found that these LSC-enriched AML expanded populations most closely resemble normal progenitor cells rather than HSCs when compared by gene expression profiling[10]. However, self-renewal genes are enriched in these LSC fractions compared to normal progenitor cells, suggesting a strong HSC component. While these last 2 studies differ in the bioinformatics comparisons performed, the picture that emerges is one where LSC have hybrid molecular programs combining strong elements of HSC along with pathways typical of progenitors (Figure 1).

Interestingly, the similarity between HSC and LSC is also observed at the epigenetic level. In the MLL-AF10 induced AML mouse model, the LSC population has been well characterized phenotypically [67] and displays high H3K9me3 methylation [68], similar to HSC [44–46]. KDM5b-driven demethylation induces LSC differentiation also in primary human AMLs [68], demonstrating that dynamic changes in chromatin marks also regulate leukaemic hierarchies.

No single cell transcriptomics are yet available on leukaemic blasts, let alone in LSC. However a high dimensional single cell proteomics and signalling approach was used by Levine et al. [69] to provide characterisation of intra- and inter-tumoural heterogeneity from 31-parameter mass cytometry data from 16 pediatric AML samples and 5 healthy adult donors. They found that the AML phenotype landscape can be described by 14 metaclusters, each with a specific surface marker pattern, reminiscent of that of normal myeloid development. It is

possible that this limited repertoire of states within these AML patients is likely to represent the limited heterogeneity that generally marks paediatric malignancies [70]. However no patient had a unique phenotype, but rather had distinct proportions of each of the metaclusters, in a way partially correlated to genetic factors. To estimate the extent to which elements of a stem/progenitor cell state are maintained in each AML, the authors used signalling responses after perturbation. They found that similarly to functional definition of HSC, signalling definition is correlated with cell surface markers in healthy samples but not in AML. This highlights that despite the very high degree of inter-tumour heterogeneity, AML nonetheless maintains an architecture based on the reshuffling of the normal modules of normal myeloid development, and that a primitive stemness component is present in all AML to different degrees, regardless of surface marker expression and underlying genetic mutations.

Prognostic value of stem cell signatures in leukaemia

Age, cytogenetics and molecular mutations are currently the most used predictors of outcome for AML, but evidence is accumulating that HSC and LSC signatures may have additional prognostic value compared to these parameters. A gene expression signature derived from transcriptional data by comparing HSC to phenotypically defined LSC[14] is predictive of primary refractoriness to chemotherapy, overall (OS), event-free (EFS) and relapse-free (RFS) survival in cytogenetically normal patients, and with OS in patients with chromosomal abnormalities. When comparing LSC to non-LSC fractions defined functionally by their capacity to establish disease in the xenograft model, Eppert et al. established an LSC signature that can predict OS and EFS independently of other known prognostic factors such as age, low molecular risk and CEBPA mutational status [60]. An HSC signature derived by comparing normal HSC enriched fractions to progenitor fraction had similar prognostic value; gene set enrichment analysis shows a high enrichment of HSC genes within the LSC gene expression data explaining this result. In addition these signatures could further stratify low molecular risk patients according to their response to standard therapy, thus

allowing a priori identification of patients who could benefit from more aggressive clinical protocols. As multidimensional profiling of signalling responses can also unequivocally identify a stem cell or primitive state [69], Levine et al. also generated a gene expression signature that can serve as a proxy for the primitive signalling phenotype. Interestingly this signature also very strongly correlated with poor survival in adult AML patients[69].

The question remains of what is the molecular driver of these signatures. Diffner et al. investigated if the heptad of TFs (ERG, FLI1, GATA2, LMO2, LYL1, RUNX1 and SCL) forming a critical regulatory circuit in normal HSCs [71], also plays a role in determining primitive expression signatures in AML [72]. Indeed signatures derived from both the expression of these 7 TFs and their activity at promoters/enhancers independently correlated with poor prognosis in cytogenetically normal AMLs [72]. The exact pattern of regulatory expression changes depending on the mutational status of the AML sample (NPM1 wt or mutant), but these data indicate that the maintenance or re-activation of these regulatory circuits in AML is largely responsible for the similarities between HSC and LSC. Beyond control by TF networks, there are several other possible factors contributing to leukaemic signatures. These may include the cell of origin, age of the patient (pediatric vs adult leukaemias), the tumour environment, and combinations of molecular mutations in the context of the constitutional germline genome of the patient.

Interestingly, epigenetic signatures derived from AML genes with alteration of histone 3 lysine 9 trimethylation (H3K9me3) in combination with established clinical prognostic markers have also been shown to be more accurate prognostic factors than the clinical parameters alone [73]. Based on all of these findings, we speculate that any resemblance to a stem-cell like state, at the transcriptional, proteomic, epigenetic or signalling level drives aggressive AML behaviour and influences clinical outcome.

Finally, the advances in understanding the molecular programmes underlying healthy HSC differentiation have also proven useful to provide insights into the

intra- and inter-tumour heterogeneity of leukaemias other than AML. For example, a subpopulation with LSC characteristics has been shown to have gene signatures similar to that of HSCs in NPM-ALK induced anaplastic large cell lymphoma (ALCL) [74]. In addition, the gene expression profile of early T cell precursor acute lymphoblastic leukaemia (ETP-ALL), an aggressive paediatric and young adult malignancy, is highly correlated to that of normal human haematopoietic and myeloid stem cells, but also to the AML LSC signature [60]. These results suggest a potentially beneficial role of myeloid directed therapy in this disease [75].

Conclusions and clinical implications

Genome-wide technologies have provided the first insights into the molecular architecture underlying blood formation. The picture that emerges is one where programmes operating in HSC are distinct from any other healthy haematopoietic cell; not only at the gene expression level but also through unique usage of splicing isoforms, enhancers, regulatory RNAs. Epigenetic marks are also highly remodelled during HSC differentiation. One of the challenges of collecting all this genome-wide data is to understand its functional relevance on human HSC biology experimentally. Many of the hypotheses generated by these studies have allowed identification of novel molecular regulators of HSC function, when tested individually by gain or loss of function assays with in vitro or xenograft assays. There are certainly limitations to the current methodologies to test for hHSC function (rare populations, lack of a native microenvironment, relatively short-term assays compared to human lifespan, failure to expand hHSC in culture…). To overcome these, a number of novel humanized mouse models are being developed [76,77] as well as approaches to generate hHSPCs from pluripotent cells [78]. These coupled to fast progress in single cell technologies, tissue engineering, genome editing and more extensive information from human genetics, should allow a better understanding of the regulation of hHSC function.

In AML, LSC display a chimeric circuitry reflective of both HSC and progenitor cells, even though the exact HSC programme elements preserved vary between

patients but also within each tumour. Despite interpatient variation, there are several potential clinical applications deriving from a comprehensive understanding of LSC and their relationship with HSC. First, the current cytogenetics and molecular mutations classifications cannot be used reliably to determine risk of relapse following remission, which remains the main cause of mortality in AML. Detecting high levels of potentially therapy resistant and relapse-driving leukaemic populations, each with carefully developed molecular signatures can be used to better direct current therapies. High-risk patients can be treated with more intensive regimens, or conversely some patients at low risk of relapse can be spared treatment-related toxicity by de-escalating maintenance or consolidation protocols. Second, LSC signatures may also be useful to monitor patients in remission. Finally, since HSC and LSC share considerable properties it is important to identify therapies that target LSC without affecting HSC thereby reducing toxicity. Thus the identification of unique molecular and functional characteristics of LSCs will help the development of targeted therapies directed to LSC eradication.

ACKNOWLEDGMENTS:

The authors would like to thank Emily Calderbank for critical review of the manuscript. Research in EL laboratory is supported by a Wellcome Trust Sir Henry Dale Fellowship and core support grant from the Wellcome Trust and MRC to the Wellcome Trust – Medical Research Council Cambridge Stem Cell Institute.

FIGURE LEGEND

Figure 1: Schematic representation of the molecular programmes in the healthy and leukaemic haematopoietic hierarchies.

During healthy blood formation (left panel), the molecular programmes of HSC (top) are clearly distinct from those of progenitor (bottom) and differentiated cell types (not depicted). In contrast, in AML (right panel), where a malignant hierarchy is retained, LSCs are characterized by a chimeric reorganisation of HSC and progenitor programmes. Pathway here indicates a group of biological entities (mRNAs, miRNA, lnc-RNAs…) that interact together in a cascade or network to regulate a particular cellular function.

References:

1. Wilson, A., E. Laurenti, G. Oser, *et al.* 2008. Hematopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair. *Cell* **135:** 1118–1129
2. Foudi, A., K. Hochedlinger, D. Van Buren, *et al.* 2009. Analysis of histone 2B-GFP retention reveals slowly cycling hematopoietic stem cells. *Nat. Biotechnol.* **27:** 84–90
3. Takizawa, H., R. R. Regoes, C. S. Boddupalli, *et al.* 2011. Dynamic variation in cycling of hematopoietic stem cells in steady state and inflammation. *J. Exp. Med.* **208:** 273–284
4. Nakamura-Ishizu, A., H. Takizawa, & T. Suda. 2014. The analysis, roles and regulation of quiescence in hematopoietic stem cells. *Development* **141:** 4656–4666
5. Ito, K. & T. Suda. 2014. Metabolic requirements for the maintenance of self-renewing stem cells. *Nat. Rev. Mol. Cell Biol.* **15:** 243–256
6. Boulais, P. E. & P. S. Frenette. 2015. Making sense of hematopoietic stem cell niches. *Blood* **125:** 2621–2629
7. Doulatov, S., F. Notta, E. Laurenti, *et al.* 2012. Hematopoiesis: a human perspective. *Cell Stem Cell* **10:** 120–136
8. Majeti, R., C. Y. Park, & I. L. Weissman. 2007. Identification of a Hierarchy of Multipotent Hematopoietic Progenitors in Human Cord Blood. *Cell Stem Cell* **1:** 635–645
9. Doulatov, S., F. Notta, K. Eppert, *et al.* 2010. Revised map of the human progenitor hierarchy shows the origin of macrophages and dendritic cells in early lymphoid development. *Nat. Immunol.* **11:** 585–593

10. Goardon, N., E. Marchi, A. Atzberger, *et al.* 2011. Coexistence of LMPP-like and GMP-like leukemia stem cells in acute myeloid leukemia. *Cancer Cell* **19:** 138–152

11. Kohn, L. A., Q.-L. Hao, R. Sasidharan, *et al.* 2012. Lymphoid priming in human bone marrow begins before expression of CD10 with upregulation of L-selectin. *Nat. Immunol.* doi:10.1038/ni.2405

12. Notta, F., S. Doulatov, E. Laurenti, *et al.* 2011. Isolation of single human hematopoietic stem cells capable of long-term multilineage engraftment. *Science* **333:** 218–221

13. Novershtern, N., A. Subramanian, L. N. Lawton, *et al.* 2011. Densely Interconnected Transcriptional Circuits Control Cell States in Human Hematopoiesis. *Cell* **144:** 296–309

14. Gentles, A. J., S. K. Plevritis, R. Majeti, *et al.* December 22. Association of a Leukemic Stem Cell Gene Expression Signature With Clinical Outcomes in Acute Myeloid Leukemia. *JAMA J. Am. Med. Assoc.* **304:** 2706 –2715

15. Laurenti, E., S. Doulatov, S. Zandi, *et al.* 2013. The transcriptional architecture of early human hematopoiesis identifies multilevel control of lymphoid commitment. *Nat. Immunol.* **14:** 756–763

16. Chen, L., M. Kostadima, J. H. A. Martens, *et al.* 2014. Transcriptional diversity during lineage commitment of human blood progenitors. *Science* **345:** 1251033

17. Chambers, S. M., N. C. Boles, K.-Y. K. Lin, *et al.* 2007. Hematopoietic fingerprints: an expression database of stem cells and their progeny. *Cell Stem Cell* **1:** 578–591

18. Gazit, R., B. S. Garrison, T. N. Rao, *et al.* 2013. Transcriptome Analysis Identifies Regulators of Hematopoietic Stem and Progenitor Cells. *Stem Cell Rep.* **1:** 266–280

19. Milyavsky, M., O. I. Gan, M. Trottier, *et al.* 2010. A Distinctive DNA Damage Response in Human Hematopoietic Stem Cells Reveals an Apoptosis-Independent Role for p53 in Self-Renewal. *Cell Stem Cell* **7:** 186–197

20. Mohrin, M., E. Bourke, D. Alexander, *et al.* 2010. Hematopoietic Stem Cell Quiescence Promotes Error-Prone DNA Repair and Mutagenesis. *Cell Stem Cell* **7:** 174–185

21. van Galen, P., A. Kreso, N. Mbong, *et al.* 2014. The unfolded protein response governs integrity of the haematopoietic stem-cell pool during stress. *Nature* **510:** 268–272

22. Miharada, K., V. Sigurdsson, & S. Karlsson. 2014. Dppa5 Improves Hematopoietic Stem Cell Activity by Reducing Endoplasmic Reticulum Stress. *Cell Rep.* **7:** 1381–1392

23. Zandi, S., D. Bryder, & M. Sigvardsson. 2010. Load and lock: the molecular mechanisms of B-lymphocyte commitment. *Immunol. Rev.* **238:** 47–62

24. van Galen, P., A. Kreso, E. Wienholds, *et al.* 2014. Reduced Lymphoid Lineage Priming Promotes Human Hematopoietic Stem Cell Expansion. *Cell Stem Cell* **14:** 94–106

25. Beer, P. A., D. J. H. F. Knapp, N. Kannan, *et al.* A Dominant-Negative Isoform of IKAROS Expands Primitive Normal Human Hematopoietic Cells. *Stem Cell Rep.* doi:10.1016/j.stemcr.2014.09.006

26. Cabezas-Wallscheid, N., D. Klimmeck, J. Hansson, *et al.* 2014. Identification of regulatory networks in HSCs and their immediate progeny via integrated

proteome, transcriptome, and DNA methylome analysis. *Cell Stem Cell* **15:** 507–522

27. Laurenti, E., C. Frelin, S. Xie, *et al.* 2015. CDK6 Levels Regulate Quiescence Exit in Human Hematopoietic Stem Cells. *Cell Stem Cell* **16:** 302–313

28. Wilson, N. K., D. G. Kent, F. Buettner, *et al.* 2015. Combined Single-Cell Functional and Gene Expression Analysis Resolves Heterogeneity within Stem Cell Populations. *Cell Stem Cell* **16:** 712–724

29. Kowalczyk, M. S., I. Tirosh, D. Heckl, *et al.* 2015. Single cell RNA-seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells. *Genome Res.* gr.192237.115 doi:10.1101/gr.192237.115

30. Tsang, J. C., Y. Yu, S. Burke, *et al.* 2015. Single-cell transcriptomic reconstruction reveals cell cycle and multi-lineage differentiation defects in Bcl11a-deficient hematopoietic stem cells. *Genome Biol.* **16:** 178

31. Cosgun, K. N., S. Rahmig, N. Mende, *et al.* 2014. Kit Regulates HSC Engraftment across the Human-Mouse Species Barrier. *Cell Stem Cell* **15:** 227–238

32. Luo, M., M. Jeong, D. Sun, *et al.* 2015. Long Non-Coding RNAs Control Hematopoietic Stem Cell Function. *Cell Stem Cell* **16:** 426–438

33. Ramos, A. D., R. E. Andersen, S. J. Liu, *et al.* 2015. The Long Noncoding RNA Pnky Regulates Neuronal Differentiation of Embryonic and Postnatal Neural Stem Cells. *Cell Stem Cell* **16:** 439–447

34. Kim, D. H., G. K. Marinov, S. Pepke, *et al.* 2015. Single-Cell Transcriptome Analysis Reveals Dynamic Changes in lncRNA Expression during Reprogramming. *Cell Stem Cell* **16:** 88–101

35. Gabut, M., P. Samavarchi-Tehrani, X. Wang, *et al.* 2011. An Alternative Splicing Switch Regulates Embryonic Stem Cell Pluripotency and Reprogramming. *Cell* **147:** 132–146

36. Sykes, S. M. & D. T. Scadden. 2013. Modeling human hematopoietic stem cell biology in the mouse. *Semin. Hematol.* **50:**

37. Pang, W. W., J. V. Pluvinage, E. A. Price, *et al.* 2013. Hematopoietic stem cell and progenitor cell mechanisms in myelodysplastic syndromes. *Proc. Natl. Acad. Sci. U. S. A.* doi:10.1073/pnas.1222861110

38. Geiger, H., G. de Haan, & M. C. Florian. 2013. The ageing haematopoietic stem cell compartment. *Nat. Rev. Immunol.* **13:** 376–389

39. Flach, J., S. T. Bakker, M. Mohrin, *et al.* 2014. Replication stress is a potent driver of functional decline in ageing haematopoietic stem cells. *Nature* **512:** 198–202

40. Beerman, I., C. Bock, B. S. Garrison, *et al.* 2013. Proliferation-Dependent Alterations of the DNA Methylation Landscape Underlie Hematopoietic Stem Cell Aging. *Cell Stem Cell* **12:** 413–425

41. Rossi, D. J., D. Bryder, J. M. Zahn, *et al.* 2005. Cell intrinsic alterations underlie hematopoietic stem cell aging. *Proc. Natl. Acad. Sci. U. S. A.* **102:** 9194–9199

42. Bock, C., I. Beerman, W.-H. Lien, *et al.* 2012. DNA methylation dynamics during in vivo differentiation of blood and skin stem cells. *Mol. Cell* **47:** 633–647

43. Bocker, M. T., I. Hellwig, A. Breiling, *et al.* 2011. Genome-wide promoter DNA methylation dynamics of human hematopoietic progenitor cells during differentiation and aging. *Blood* **117:** e182–e189

44. Sun, D., M. Luo, M. Jeong, *et al.* 2014. Epigenomic Profiling of Young and Aged HSCs Reveals Concerted Changes during Aging that Reinforce Self-Renewal. *Cell Stem Cell* **14:** 673–688

45. Cui, K., C. Zang, T.-Y. Roh, *et al.* 2009. Chromatin Signatures in Multipotent Human Hematopoietic Stem Cells Indicate the Fate of Bivalent Genes during Differentiation. *Cell Stem Cell* **4:** 80–93

46. Weishaupt, H., M. Sigvardsson, & J. L. Attema. 2010. Epigenetic chromatin states uniquely define the developmental plasticity of murine hematopoietic stem cells. *Blood* **115:** 247–256

47. Lara-Astiaso, D., A. Weiner, E. Lorenzo-Vivas, *et al.* 2014. Chromatin state dynamics during blood formation. *Science* 1256271 doi:10.1126/science.1256271

48. Wittkopp, P. J. & G. Kalay. 2012. Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nat. Rev. Genet.* **13:** 59–69

49. Jaiswal, S., P. Fontanillas, J. Flannick, *et al.* 2014. Age-Related Clonal Hematopoiesis Associated with Adverse Outcomes. *N. Engl. J. Med.* **0:** null

50. Busque, L., J. P. Patel, M. E. Figueroa, *et al.* 2012. Recurrent somatic TET2 mutations in normal elderly individuals with clonal hematopoiesis. *Nat. Genet.* doi:10.1038/ng.2413

51. Genovese, G., A. K. Kähler, R. E. Handsaker, *et al.* 2014. Clonal Hematopoiesis and Blood-Cancer Risk Inferred from Blood DNA Sequence. *N. Engl. J. Med.* **0:** null

52. Xie, M., C. Lu, J. Wang, *et al.* 2014. Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nat. Med.* **advance online publication:**

53. McKerrell, T., N. Park, T. Moreno, *et al.* 2015. Leukemia-associated somatic mutations drive distinct patterns of age-related clonal hemopoiesis. *Cell Rep.* **10:** 1239–1245

54. Dick, J. E. 2008. Stem cell concepts renew cancer research. *Blood* **112:** 4793–4807

55. Visvader, J. E. 2011. Cells of origin in cancer. *Nature* **469:** 314–322

56. Kreso, A. & J. E. Dick. 2014. Evolution of the Cancer Stem Cell Model. *Cell Stem Cell* **14:** 275–291

57. Bonnet, D. & J. E. Dick. 1997. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nat. Med.* **3:** 730–737

58. Jordan, C. T., M. L. Guzman, & M. Noble. 2006. Cancer Stem Cells. *N. Engl. J. Med.* **355:** 1253–1261

59. Lapidot, T., C. Sirard, J. Vormoor, *et al.* 1994. A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. *Nature* **367:** 645–648

60. Eppert, K., K. Takenaka, E. R. Lechman, *et al.* 2011. Stem cell gene expression programs influence clinical outcome in human leukemia. *Nat. Med.* **17:** 1086–1093

61. Martelli, M. P., V. Pettirossi, C. Thiede, *et al.* 2010. CD34+ cells from AML with mutated NPM1 harbor cytoplasmic mutated nucleophosmin and generate leukemia in immunocompromised mice. *Blood* **116:** 3907–3922

62. Taussig, D. C., F. Miraki-Moud, F. Anjos-Afonso, *et al.* 2008. Anti-CD38 antibody-mediated clearance of human repopulating cells masks the heterogeneity of leukemia-initiating cells. *Blood* **112:** 568–575

63. Sarry, J.-E., K. Murphy, R. Perry, *et al.* 2011. Human acute myelogenous leukemia stem cells are rare and heterogeneous when assayed in NOD/SCID/IL2Rγc-deficient mice. *J. Clin. Invest.* **121:** 384–395

64. Majeti, R., M. W. Becker, Q. Tian, *et al.* 2009. Dysregulated gene expression networks in human acute myelogenous leukemia stem cells. *Proc. Natl. Acad. Sci. U. S. A.* **106:** 3396–3401

65. Saito, Y., H. Kitamura, A. Hijikata, *et al.* 2010. Identification of Therapeutic Targets for Quiescent, Chemotherapy-Resistant Human Leukemia Stem Cells. *Sci. Transl. Med.* **2:** 17ra9

66. Barreyro, L., B. Will, B. Bartholdy, *et al.* 2012. Overexpression of IL-1 receptor accessory protein in stem and progenitor cells and outcome correlation in AML and MDS. *Blood* **120:** 1290–1298

67. Somervaille, T. C. P. & M. L. Cleary. 2006. Identification and characterization of leukemia stem cells in murine MLL-AF9 acute myeloid leukemia. *Cancer Cell* **10:** 257–268

68. Wong, S. H. K., D. L. Goode, M. Iwasaki, *et al.* The H3K4-Methyl Epigenome Regulates Leukemia Stem Cell Oncogenic Potential. *Cancer Cell* doi:10.1016/j.ccell.2015.06.003

69. Levine, J. H., E. F. Simonds, S. C. Bendall, *et al.* 2015. Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis. *Cell* **162:** 184–197

70. Alexandrov, L. B., S. Nik-Zainal, D. C. Wedge, *et al.* 2013. Signatures of mutational processes in human cancer. *Nature* **500:** 415–421

71. Wilson, N. K., S. D. Foster, X. Wang, *et al.* 2010. Combinatorial Transcriptional Control In Blood Stem/Progenitor Cells: Genome-wide Analysis of Ten Major Transcriptional Regulators. *Cell Stem Cell* **7:** 532–544

72. Diffner, E., D. Beck, E. Gudgin, *et al.* 2013. Activity of a heptad of transcription factors is associated with stem cell programs and clinical outcome in acute myeloid leukemia. *Blood* **121:** 2289–2300

73. Müller-Tidow, C., H.-U. Klein, A. Hascher, *et al.* 2010. Profiling of histone H3 lysine 9 trimethylation levels predicts transcription factor activity and survival in acute myeloid leukemia. *Blood* **116:** 3564–3571

74. Moti, N., T. Malcolm, R. Hamoudi, *et al.* 2014. Anaplastic large cell lymphoma-propagating cells are detectable by side population analysis and possess an expression profile reflective of a primitive origin. *Oncogene* doi:10.1038/onc.2014.112

75. Zhang, J., L. Ding, L. Holmfeldt, *et al.* 2012. The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. *Nature* **481:** 157–163

76. Rongvaux, A., H. Takizawa, T. Strowig, *et al.* 2013. Human Hemato-Lymphoid System Mice: Current Use and Future Potential for Medicine. *Annu. Rev. Immunol.* **31:** 635–674

77. Goyama, S., M. Wunderlich, & J. C. Mulloy. 2015. Xenograft models for normal and malignant stem cells. *Blood* **125:** 2630–2640

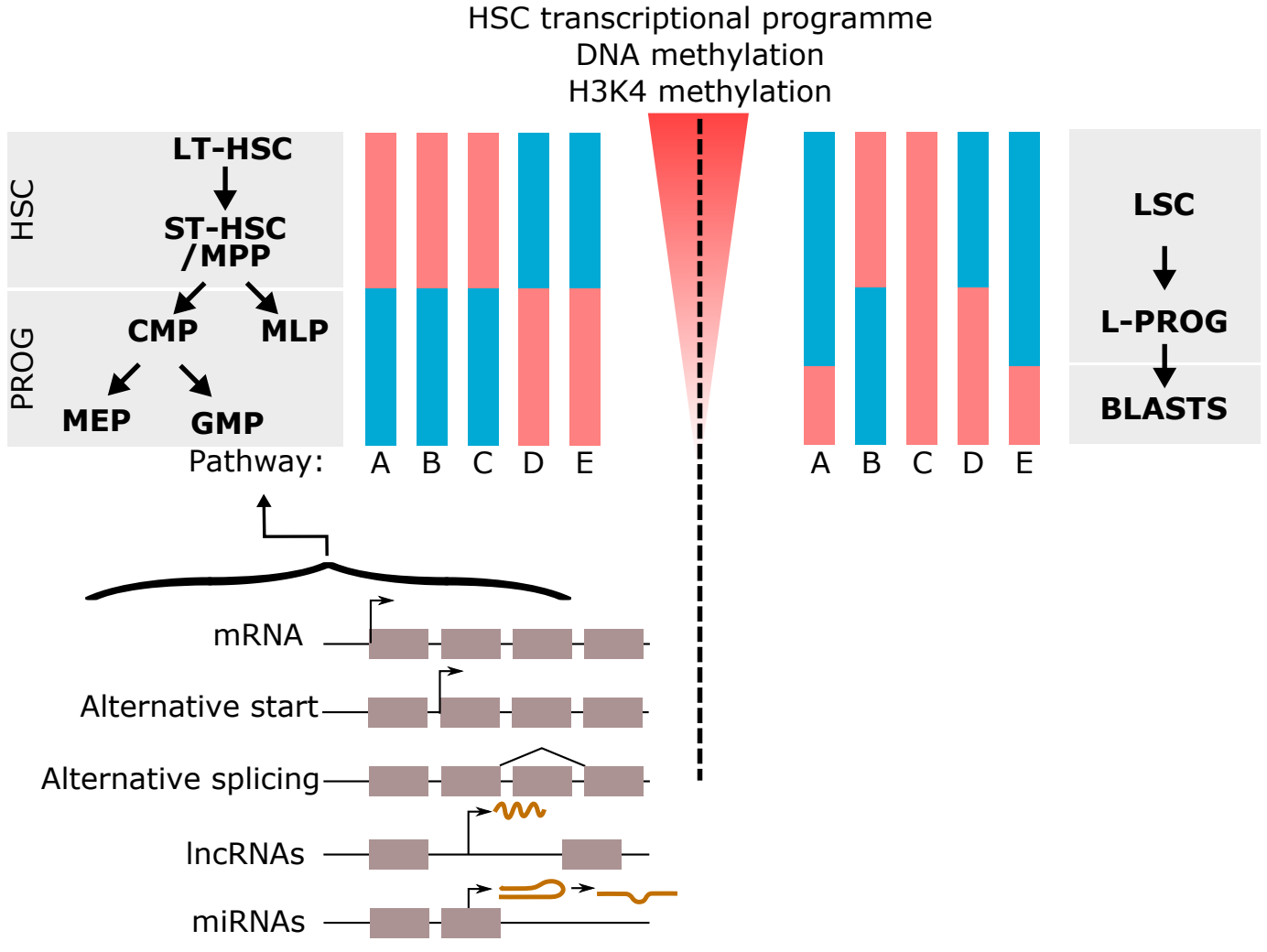
78. Vo, L. T. & G. Q. Daley. 2015. De novo generation of HSCs from somatic and pluripotent stem cell sources. *Blood* **125:** 2641–2648

Table 1: phenotypic definitions of HSC populations used in transcriptomics and epigenetic studies.

Population	Cell surface markers	Technique	Organ
HSPC	Lin- CD34+	Methylation array [39]	CB/mPB [39]
HSPC	CD133+	Histone methylation ChIP-seq [41]	mPB [41]
HSPC1	Lin- CD133+ CD34dim	Microarray [13]	CB [13]
HSPC2	Lin- CD34+ CD38-	Microarray [13]	CB [13]
HSC	Lin- CD34+ CD38- CD90+ CD45RA-	Microarray [10,14,33]	CB/BM [14] BM [10,33]
MPP	Lin- CD34+ CD38- CD90- CD45RA-	RNA-seq [16]	CB [16]
LT-HSCs	Lin- CD34+ CD38- CD90+ CD45RA- CD49f+	Microarray [12,15,27]	CB [12,15,27]
ST-HSCs	Lin- CD34+ CD38- CD90- CD45RA- CD49f-	Microarray [12,15,27]	CB [12,15,27]

Table 2: summary of studies reporting LSC signatures.

LSC definition	Bioinformatic comparison
Lin- CD34+ CD38-	LSC vs HSC [61]
Lin- CD34+ CD38- CD123+	LSC vs HSC [60]
Lin- CD34+ CD38+ CD90-	LSC vs HSC [62]
Lin- CD34+ CD38- CD90-CD45RA+	LSC vs HSC [14] LSC vs HSC/prog [10]
Functionally defined by xenograft transplantation	LSC vs non-LSC [56]


HSC transcriptional programme
DNA methylation
H3K4 methylation
HSC
PROG
LT-HSC
ST-HSC
/MPP
CMP
MLP
MEP
GMP
LSC
L-PROG
BLASTS
Pathway:
A B C D E
A B C D E
mRNA
Alternative start
Alternative splicing
lncRNAs
miRNAs