Published in final edited form as:

Exp Hematol. 2008 October ; 36(10): 1377–1389. doi:10.1016/j.exphem.2008.06.015.

Molecular Profiling Reveals Similarities and Differences Between Primitive Subsets of Hematopoietic Cells Generated *In Vitro* from Human Embryonic Stem Cells and *In Vivo* during Embryogenesis

Giorgia Salvagiotto¹, Yun Zhao², Maxim Vodyanik³, Victor Ruotti¹, Ronald Stewart¹, Marco Marra⁴, James Thomson^{5,6}, Connie Eaves^{2,†}, and Igor Slukvin^{1,3,7,†}

¹WiCell Research Institute, BC Cancer Agency, Vancouver, BC

²Terry Fox Laboratory, BC Cancer Agency, Vancouver, BC

³Wisconsin National Primate Research Center, University of Wisconsin, Madison, WI

⁴Genome Sciences Centre, BC Cancer Agency, Vancouver, BC

⁵Genome Center of Wisconsin, University of Wisconsin, Madison, WI

⁶Anatomy, University of Wisconsin, Madison, WI

⁷Pathology and Laboratory Medicine, University of Wisconsin, Madison, WI

Abstract

Objective—Cellular and molecular changes that occur during the genesis of the hematopoietic system and hematopoietic stem cells in the human embryo are mostly inaccessible to study and remain poorly understood. To address this gap we have exploited the human embryonic stem cell (hESC) system to molecularly characterize the global transcriptomes of the two functionally discrete and phenotypically separable populations of multipotent hematopoietic cells that first appear when hESCs are induced to differentiate on OP9 cells.

Methods—We prepared long serial analysis of gene expression (LongSAGE) libraries from $\text{lin}^- \text{CD34}^+ \text{CD43}^+ \text{CD45}^-$ and $\text{lin}^- \text{CD34}^+ \text{CD43}^+ \text{CD45}^+$ subsets of primitive hematopoietic cells derived *in vitro* from hESCs, sequenced them to a depth of 200,000 tags and compared their content with similar libraries prepared from highly purified populations of very primitive human fetal liver (FL) and cord blood (CB) hematopoietic cells.

Results—Comparison of libraries obtained from hESC-derived $\text{lin}^- \text{CD34}^+ \text{CD43}^+ \text{CD45}^-$ and $\text{lin}^- \text{CD34}^+ \text{CD43}^+ \text{CD45}^+$ revealed differences in their expression of genes associated with myeloid development, cellular biosynthetic processes, and cell cycle regulation. Further comparisons with analogous data for primitive hematopoietic cells isolated from first trimester human fetal liver and newborn cord blood showed an apparent similarity between the transcriptomes of the most primitive hESC- and *in vivo*-derived populations, with the main differences involving genes that regulate HSC

Corresponding author: Dr. Igor Slukvin, MD, PhD, Department of Pathology and Laboratory Medicine, Wisconsin National Primate Research Center, University of Wisconsin, 1220 Capitol Court, Madison, WI 53715, Phone: (608) 263 0058; Fax: (608) 265 8984; E-mail: E-mail: islukvin@wisc.edu.

[†]Co-senior authors

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

self-renewal and homing, chromatin remodeling, AP1 transcription complex genes, and non-coding RNAs.

Conclusion—These data suggest that primitive hematopoietic cells are generated from hESCs *in vitro* by processes similar to those operative during human embryogenesis *in vivo*, although some differences were also detected.

Keywords

embryonic stem cells; hematopoietic stem cells; fetal human liver; gene expression; hematopoietic development

INTRODUCTION

Human embryonic stem cells (hESCs) are capable of indefinite self-renewal but can also be induced to undergo a stepwise process of differentiation into a spectrum of recognizable mature blood cell types. These properties of hESCs have stimulated interest in their potential use as a scalable source of cells for a variety of blood cell replacement therapies and for more detailed investigations of the mechanisms of hematopoietic cell specification and stem cell generation. When undifferentiated hESCs are transferred to a coculture system containing mouse OP9 feeders, the cells undergo a series of changes leading to the formation of a full spectrum of myeloid progenitors, all of which express CD43 (Fig. 1) [1]. The first of these cells also express glycophorin A (CD235a) and CD41a (a marker of megakaryocytes) and appear restricted to the erythro-megakaryocytic lineages. Cells with a broader lympho-myeloid differentiation repertoire and a $\text{lin}^- \text{CD34}^+ \text{CD43}^+ \text{CD45}^-$ phenotype appear next. Finally, a $\text{lin}^- \text{CD34}^+ \text{CD43}^+ \text{CD45}^+$ population emerges. These latter cells are highly enriched in myeloid progenitors but also show some lymphoid differentiation potential [1,2].

A similar pattern of events in which sequential populations of hematopoietic cells with increasing differentiation potentialities appear is known to characterize the establishment of definitive hematopoiesis *in vivo* in mice [3]. Human embryos containing cells undergoing analogous early stages of hematopoietic development are difficult to access and hESCs therefore provide an important model to investigate this process. Nevertheless, in both species, a clear understanding of the molecular mechanism by which the first hematopoietic stem cells (HSCs) acquire their unique defining properties of self-renewal and repopulating potential is lacking [4–8]. As a first step towards obtaining the information needed to close this gap, we have undertaken a comparative gene expression analysis of different highly purified primitive human hematopoietic subpopulations generated either *in vitro* from hESCs or *in vivo* from embryonic mesodermal precursors. This involved preparing a long serial analysis of gene expression (LongSAGE) library from an RNA extracts of each prospectively isolated subpopulation and then sequencing each library to a depth of 200,000 tags. Comparison of these libraries revealed the greatest similarity between the most primitive subset of hESC-derived cells and their counterparts isolated from suspensions of human fetal liver, although distinctive features were also discerned between all libraries.

MATERIALS AND METHODS

Isolation of primitive hematopoietic subpopulations from hESCs

H1 hESCs (WiCell Research Institute, Madison, WI) were maintained as undifferentiated cells in co-cultures containing mouse embryonic fibroblasts (MEFs) [9]. Hematopoietic differentiation was induced by transferring the cells onto OP9 feeders [2] and 8 days later, $\geq 98\%$ pure subpopulations of $\text{CD43}^+ \text{CD235a}^+ \text{CD41a}^{+/-}$, $\text{lin}^- \text{CD34}^+ \text{CD43}^+ \text{CD45}^-$, and

lin⁻CD34⁺CD43⁺CD45⁺ cells were isolated by fluorescence activated cell sorting of antibody-stained harvests of these cells, as previously described in details [1,10].

Isolation of primitive hematopoietic subpopulations from fetal liver and cord blood

Human fetal liver (FL) cells (from 5 aborted fetuses of 14- to 21-weeks gestational age) and umbilical cord blood (CB) samples were obtained according to approved institutional procedures and the low-density (<1.077 g/cm³) cells cryopreserved. A population enriched in CD34⁺ cells was isolated immediately after thawing the cells using the EasySep kit (Stem Cell Technologies, Vancouver, BC, Canada) according to the manufacturer's instructions. The CD34⁺-enriched cells were then incubated overnight in Hanks' Balanced Salt Solution plus 50% fetal calf serum (FCS) at 4°C prior to being stained with allophycocyanin (APC)-conjugated anti-CD34 (8G12), phycoerythrin (PE)-conjugated anti-CD38 (HB7) from and fluorescein isothiocyanate (FITC)-conjugated anti-CD2, CD3, CD7, CD14, CD16, CD 19, CD20, CD24, CD36, CD45RA CD56, CD66b, CD71, and GlycophorinA (all antibodies from either BD Biosciences, San Jose, CA or StemCell Technologies). The CD38⁻ (PE⁻) and CD38⁺ (PE⁺) subsets of cells in the CD34⁺ (APC⁺) gate and negative for CD2, CD3, CD7, CD14, CD16, CD 19, CD20, CD24, CD36, CD45RA CD56, CD66b, CD71 and GlycophorinA (FITC⁻) were collected at ≥95% purity by double sorting using a FACSVantage SE (BD Biosciences). An aliquot from each population was then plated in methylcellulose medium (Methocult 4435, StemCell Technologies) to determine the content of (myeloid) colony-forming cells (CFCs). A second aliquot was used to determine the content of 6-week longterm culture-initiating cells (LTC-ICs) using mouse feeders engineered to produce human interleukin-3, human Steel factor and human granulocyte colony-stimulating factor [11]. The remainder was then used for RNA extraction.

LongSAGE library generation, sequencing and analysis

RNA was extracted from hESC-derived cells using the RiboPure™ Kit (Ambion, Austin, TX) and subsequently treated with DNaseI using DNA-free reagents (Ambion). RNA was isolated from FL and CB cells using the Pico Pure™ RNA extraction kit from Arcturus (Mountain View, CA), and then also treated with DNaseI (Invitrogen; Carlsbad, CA). After determining that the quantity and quality of total RNA obtained was as expected using a BioAnalyzer (Agilent, Foster City, CA), 10 ng of total RNA from the FL and CB subsets was first amplified as described [12]. Standard and "PCR"-LongSAGE libraries were then constructed from these hESC-derived and primary cDNA preparations, respectively, using kits supplied by Invitrogen. Each library was sequenced to a depth of ~200,000 tags (as a purchased service from the Genome Sciences Centre of the BC Cancer Agency). Selected transcripts were confirmed by quantitative reverse transcriptase (Q-RT)-PCR analysis using the Platinum® SYBR® Green qPCR SuperMix-UDG (Invitrogen). Housekeeping gene *RPL13A* was used as internal control for normalizing gene expression levels. All primers are shown in Table S1.

Bioinformatics and statistical methods

DiscoverySpace software (<http://www.bcgsc.ca/platmorm/bioinfo/software/ds> [13]) was used for tag-to-human gene mapping for all LongSAGE tags identified based on the RefSeq database (<http://www.ncbi.nlm.nih.gov/RefSeq>). Only tags that mapped to a single gene were considered further. To undertake the hierarchical clustering analysis, tag frequencies in each LongSAGE library were normalized to 100,000 tags prior to being analyzed using a MeV software algorithm [14] to calculate Euclidean distances and average linkage [15]. DiscoverySpace software was also used to determine differentially expressed genes in pairs of LongSAGE libraries using the Audic Claverie formula [16] and a 99.9% cut-off value. Classification, statistical and pathway analysis of overexpressed genes in comparisons of any

two LongSAGE libraries was performed using Gene Ontology (GO) and KEGG pathways analysis with FatiGO+ software. (<http://babelomics.bioinfo.cipf.es/index.html>).

RESULTS

Distribution of tag types in LongSAGE libraries prepared from primitive human hematopoietic cells originating from hESCs *in vitro* or the embryo *in vivo*

Three phenotypically defined subsets of primitive hematopoietic cells were isolated from the cells present after 8 days of co-culture of H1 hESCs on OP9 cells: CD43⁺CD235a⁺CD41a^{+/−} (erythro-megakaryopoietic; 10% CFC), lin[−]CD34⁺CD43⁺CD45[−] (multi-potent; 19% CFC) and lin[−]CD34⁺CD43⁺CD45⁺ (myeloid-skewed; 37% CFC) cells (Fig.1) [1]. At this day of differentiation more than 95–98% of lin[−]CD34⁺CD43⁺CD45^{+/−} cells are CD38-negative [2]. Another library was prepared from a subset of lin[−]CD34⁺CD38[−] cells (consisting of 31% CFCs and 58 % LTC-ICs, respectively) obtained from a sample of human FL and 2 more libraries were prepared from the same lin[−]CD34⁺CD38[−] subset as well as a matching lin[−]CD34⁺CD38⁺ subset (6% and 15% CFCs, and 14 % and 4% LTC-ICs, respectively) isolated from a large pool of human CB cells.

Library details are shown in Table 1 together with those for a LongSAGE library generated previously from undifferentiated H1 hESCs [17]. All libraries contained many tags for potentially novel transcripts - defined as tags for which no known transcript/gene could be identified in RefSeq. Interestingly, these tags for novel transcripts were 2 to 6-fold more numerous in the undifferentiated hESC library and in the FL and CB libraries as compared to the libraries prepared from any of the 3 hESC-derived hematopoietic populations.

Similarities in the transcriptomes of primitive hematopoietic cells isolated from different sources

The FL and CB cell libraries and the libraries prepared from both lin[−] subsets of hESC-derived hematopoietic cells all lacked transcripts for the lin markers used to isolate them thus validating the purification process used. The library prepared from the CD34⁺CD43⁺CD45[−] subset of hESC-derived cells contained tags for genes associated with lymphoid as well as myeloid differentiation consistent with the multi-potent functional activity of these cells (Fig. 2). In contrast, tags for myeloid transcripts were more common in the library prepared from the lin[−]CD34⁺CD43⁺CD45⁺ cells which are skewed towards myeloid differentiation [1]. Also as expected, the library prepared from the CD43⁺CD235a⁺CD41a^{+/−} hESC-derived cells showed a high representation of tags mapping to genes associated with their described erythroid and megakaryocytic differentiation activity [1].

To obtain a quantitative measure of the relatedness of the various libraries to one another, we employed a hierarchical clustering algorithm to the entire data set. This showed the FL lin[−]CD34⁺CD38[−] cell library to be more closely related to the hESC-derived lin[−]CD34⁺CD43⁺CD45[−] and lin[−]CD34⁺CD43⁺CD45⁺ cell libraries than to either of the CB libraries (Fig. 2). Therefore more detailed comparisons of the hESC-derived libraries were restricted to the FL library.

Differentially expressed genes in primitive human hematopoietic cells originating from hESCs *in vitro* or within the embryo *in vivo*

Using DiscoverySpace software, we identified 816 tags (748 genes) present at a significantly higher frequency and 1320 tags (1164 genes) present at a significantly lower frequency ($P < .001$) in the FL library as compared to the hESC-derived lin[−]CD34⁺CD43⁺CD45[−] cell library (Fig.3). Gene ontology (GO) assignments indicated that the first group contains an overrepresentation of genes involved in programmed cell death regulation of transcription and

DNA-dependent transcription, whereas the second group is enriched for transcripts in hexose metabolism (Fig. 4) and oxidoreductase activity (Fig. S1).

Differentially expressed genes in the FL and hESC-derived $\text{lin}^- \text{CD34}^+ \text{CD43}^+ \text{CD45}^+$ cell libraries categorized by GO analysis were shown to belong to a more diverse range of biological process and molecular function categories, with tags for genes associated with DNA-dependent transcription, regulation of transcription, mRNA processing, cell growth, and negative regulation of progression through the cell cycle being significantly enriched in the FL library and tags for genes involved in hexose metabolism, monosaccharide metabolism, oxidative phosphorylation, and oxidoreductase activity being enriched in the hESC-derived $\text{lin}^- \text{CD34}^+ \text{CD43}^+ \text{CD45}^+$ library (Fig. 4 and Fig. S1).

The top 25 most highly represented transcripts in libraries of primitive hematopoietic cells established from FL and hESCs are presented in Fig. 3, with a complete list of all differentially expressed genes presented in Table S2–S4. These findings suggest a common and/or important role in primitive hematopoietic cells of *AP-1* complex genes, which are involved in the regulation of cell proliferation, transformation, and death [18]. Whereas tags for *FOS*, *FOSB*, *JUN*, *JUNB*, and *JUND* were among the most prevalent in the FL library, relatively few tags for these genes were detected in the hESC-derived cell libraries (Fig. 5). The striking differences in expression of these genes were confirmed by Q-RT-PCR. Other highly and differentially expressed genes in the FL library include *IL8* (a chemokine), *VILL* and *VIL2* (members of the gelsolin superfamily), *NFKBIA*, *NFKBIZ*, and *RELB* (NF- κ B pathway elements) and *U2AF1*, *TCERG1*, and *LUC7L* (genes important in RNA splicing).

Genes found to be more highly expressed in the hESC-derived $\text{lin}^- \text{CD34}^+ \text{CD43}^+ \text{CD45}^+$ and CD45^- cells as compared to the FL library included *IGFBP2*, a modulator of insulin growth factor (IGF) activity [19] and cell proliferation, adhesion and migration [20], and an activator of CD24 expression [21]. CD24 is also highly expressed by hESC-derived hematopoietic cells which may thus be a consequence of their high expression of *IGFBP2*.

hESC-derived $\text{lin}^- \text{CD34}^+ \text{CD43}^+ \text{CD45}^+$ and CD45^- cells also expressed very high levels of transcripts for *PFN1*, *ARPC5*, *COTL* (genes involved in the regulation of the actin cytoskeleton) and *ARHGDIB* (a hematopoietic cell-specific RHO dissociation inhibitor [22–25]). Both of these libraries also contained high numbers of tags for *PTMA* and *SET* (genes involved in global chromatin decondensation and critical for activation of transcription [26] as well as tags for *HMGA1* (another gene involved in local and global changes of chromatin structure [27]) which was also highly expressed in undifferentiated hESCs. The tag count for genes encoding for PRC1 polycomb complex proteins, such as *CBX4* and *PHC2* genes, was higher in the CD45^- cell library as compared to FL $\text{lin}^- \text{CD34}^+ \text{CD38}^-$ cells. In contrast, tags for the PRC2 polycomb complex genes: *EED*, *EZH2*, and *SUZ12* were present at lower levels in both hESC-derived $\text{lin}^- \text{CD34}^+ \text{CD43}^+$ cell libraries as compared to the FL library (Fig. 6).

Substantial differences in expression of long non-coding RNAs were also noted. These included very high *H19* tag counts in the hESC-derived $\text{lin}^- \text{CD34}^+ \text{CD43}^+$ cell libraries as compared to the FL library with the converse situation for *MALAT1* tag counts. *H19* is a 2.3 kb paternally imprinted gene that is highly expressed through embryonic and fetal development especially in the endoderm and mesoderm [28]. *MALAT1* transcripts are highly expressed in a number of human carcinomas [29]. It has been reported that *H19* may function as a tumor suppressor gene [30] and regulator of gene expression at the *H19/Igf2* locus [31] but the actual function of *H19* remains poorly defined.

Genes important for HSC development, expansion, survival and engraftment

Gene-targeting experiments have identified several genes essential for the specification and development of HSCs in mice, including *CBFA2T3*, *CBFB*, both *HIPK1* and *HIPK2*, *LDB1*, *LMO2*, *RUNX1*, and *TAL1* [32–39]. As shown in Fig. 5, tags for the *TAL1*, *RUNX1*, *LMO2*, and *CBFA2T3* were present in both hESC-derived $\text{lin}^- \text{CD34}^+ \text{CD43}^+$ cell libraries. In addition, tag counts for *GATA3* gene known to be expressed in emerging HSCs [40] were high in both CD45^+ and CD45^- libraries (Fig. 2). Tags for *HIPK1* and *LDB1* were present in the CD45^+ cell library but not in the CD45^- cell library. Similarly, tags for several genes involved in HSC differentiation control like *EZH2*, *MEIS1*, and *MLL*, were not present in CD45^- library, but were detected in CD45^+ library, although at lower levels than in the FL $\text{lin}^- \text{CD34}^+ \text{CD38}^-$ cell library (Fig. 5). Interestingly, we found that both types of hESC-derived $\text{lin}^- \text{CD34}^+ \text{CD43}^+$ cells expressed *HOXB* cluster genes at a higher level than the FL $\text{lin}^- \text{CD34}^+ \text{CD38}^-$ cells, although the highest number of tags for *HOXB* genes was found in the more primitive hESC-derived $\text{lin}^- \text{CD34}^+ \text{CD43}^+ \text{CD45}^-$ cells with *HOXB3* tags predominating (Fig. 4).

Tags for *ARHGAP1*, *ETV6*, and *HLF* (3 genes essential for HSC survival) were not detected in the hESC-derived $\text{lin}^- \text{CD34}^+ \text{CD43}^+ \text{CD45}^-$ cell library, but were present at a low level in the hESC-derived $\text{lin}^- \text{CD34}^+ \text{CD43}^+ \text{CD45}^+$ cell library as compared to the FL library. Most of these differences indicated by comparison of the SAGE libraries were confirmed by Q-RT-PCR (Fig. 5, 6).

Recently, an essential role of *SOX17* in the maintenance of fetal and neonatal, but not adult HSCs has been demonstrated [41]. *SOX17*-specific tags were not identified in any of the hematopoietic cell SAGE libraries, although *SOX17* transcripts were detectable by Q-RT-PCR using specific primers to amplify its active form. These transcripts were present at the highest levels in the hESC-derived $\text{lin}^- \text{CD34}^+ \text{CD43}^+ \text{CD45}^-$ cells, at intermediate levels in the FL $\text{lin}^- \text{CD34}^+ \text{CD38}^-$ cells and at the lowest levels in hESC-derived $\text{lin}^- \text{CD34}^+ \text{CD43}^+ \text{CD45}^+$ cells (Fig. 5).

CD44, *ITGA4*, and *ITGB1* (genes implicated in the homing and engraftment of HSCs) were also found to be expressed in the FL $\text{lin}^- \text{CD34}^+ \text{CD38}^-$ and both hESC-derived $\text{lin}^- \text{CD34}^+ \text{CD43}^+$ cell libraries. However, tags for *CXCR4* were not detected in any of hESC-derived libraries. These patterns of adhesion molecule gene expression were confirmed by flow cytometry of cells stained with the corresponding antibodies (Fig. 7).

Expression of β -*CATENIN* (*CTNNB1*) and its positive regulator *DVL2* was higher in the FL library than in the hESC libraries, whereas the hESC-derived $\text{lin}^- \text{CD34}^+ \text{CD43}^+ \text{CD45}^-$ cells showed significantly higher expression of *CTNNBIP1*, *CSNK2B*, *RBX1*, and *BTRC* (genes involved in the inactivation and ubiquitin-mediated proteolysis of β -catenin; Fig. S2). Significant differences in expression of *HDAC1*, *APH1*, and *DTX* (genes involved in Notch signaling) were also seen, although no significant differences were noted in *NOTCH1*-specific tag counts. Interestingly, *KIT* and *MPL* (genes for receptors for growth factors active on fetal and adult HSCs) showed significantly higher expression in the hESC-derived $\text{lin}^- \text{CD34}^+ \text{CD43}^+ \text{CD45}^-$ cells than in the FL $\text{lin}^- \text{CD34}^+ \text{CD38}^-$ cells. The hESC-derived $\text{lin}^- \text{CD34}^+ \text{CD43}^+ \text{CD45}^-$ cells also showed greater expression of *STAT5B* and *PIM1* genes in the JAK-STAT signaling pathway as well *GRB2*, *AKT1*, *MYC*, and *BCL2L1* genes associated with this pathway.

Analysis of differentially expressed genes in the CD45^- and CD45^+ subsets of hESC-derived $\text{lin}^- \text{CD34}^+ \text{CD43}^+$ cells

Although the total number of tags that mapped to specific genes was lower in the hESC-derived CD45^- cell library as compared to the CD45^+ library, the total number of genes that were

expressed at a greater level in the CD45⁻ cells was higher (Table 1, Fig. 3). These were found in GO categories identified as negative progression through cell cycle, and generation of neurons (Fig. 4). Genes whose expression was higher in the CD45⁺ subset were overrepresented in GO categories associated with nucleoside metabolic processes, amino acid activation, oxidative phosphorylation, oxidoreductase activity, and structural constituent of ribosome (Fig. 4, Fig. S1).

The more highly expressed genes in the CD45⁻ cells included *MIB1* (implicated in the endocytosis of Notch ligands), *RNF6* (a potential tumor suppressor gene), *BAZ1B* (a chromatin-dependent regulator of transcription), *PARD6B* (a gene involved in asymmetrical cell division and cellular polarization), and *BTG1*, *PPP2R1A*, *PPP1R9B*, *SIPA1*, *ATM*, *CDC123*, *DAB2IP*, *LIN9*, *PKD2*, *PNN*, *PTEN*, *RASSF2*, *RB1CC1*, *TBRG4*, *TSC2*, *WT1* (negative regulators of cell growth and division). In addition, the CD45⁻ subset showed higher expression of *ETS1* transcription factor gene implicated in vascular and lymphoid development [42].

The most highly and differentially expressed genes in the CD45⁺ subpopulation included *U2AF1* (a gene involved in regulation RNA splicing), *PABPC1* and *EIF1* (genes involved in initiation of translation), *RPLP1* (protein synthesis), *TRAM1* and *KDELRI* (endoplasmic reticulum protein transport) and *CAPNS1* (posttranslational protein modification) indicative of activation of cellular protein biosynthetic processes in these cells. Surprisingly, the CD45⁺ cells also showed expression of keratin transcripts and transcripts for *ITGB4* (an epithelial-cell specific integrin).

Discussion

HSCs are defined as long-term repopulating cells with an ability both to generate all types of blood cells and to generate progeny with the same latent multipotent properties [43]. During embryogenesis, primary hematopoietic cells are generated in the yolk sac, para-aortic splanchnopleura/aorto-gential ridges-mesonephros (AGM) and placenta. However, at least in the mouse model, pluripotent cells with long-term repopulation ability appear only in the AGM region [44] and placenta [45]. It is generally accepted that HSCs generated in the AGM region colonize the fetal liver and finally the bone marrow [3,46]. In the fetal liver, HSCs rapidly expand through symmetrical divisions and the entire population of FL-HSCs remain in cycle, whereas throughout adulthood, the rate of turnover of the HSC population is much slower [47–49]. There is also growing evidence that the genetic program controlling engraftment and expansion of fetal HSCs is different from that operative in adult HSCs [50,51]. We previously characterized the first two subpopulations of multipotent hematopoietic cells to appear in OP9 cocultures of hESCs on the basis of their common expression of CD43 and differential expression of CD45. Interestingly, the cells acquiring expression of CD45 are highly enriched in more committed myeloid progenitors [1]. Therefore, it was of interest to examine the transcriptomes of these cells and compare them with each other as well as with the most primitive subpopulations of hematopoietic cells that could be isolated from human FL and CB samples.

Both of these hESC-derived hematopoietic subpopulations were more similar to the most primitive lin⁻CD34⁺CD38⁻ subset of FL cells than to the same subset in CB or the earlier erythro-megakaryopoietic hESC-derived cells and both also expressed at very high levels several genes characteristic of HSCs emerging in the AGM (i.e., *GATA2*, *GATA3*, *RUNX1* and *TAL1*) [37,40,52–54]. The multipotent hESC-derived cells also expressed high levels of *SOX17*, a transcription factor required for maintenance of fetal, but not adult murine HSCs [41]. However, the transcriptome of the multipotent hESC-derived cells also differed from that of the HSC-enriched FL subset. This included a lack of expression by the hESC-derived cells of genes associated with HSC self-renewal (*EZH2*, *MEIS1*, *MLL*) and survival (*ARHGAP1*,

ETV6, and *HLF*), as well as *AP-1* genes, and a higher expression of certain noncoding RNAs. Additional experiments will now be required to discriminate whether these differences are due to persisting heterogeneity within the populations being compared or whether they reflect overriding programmatic differences that are developmentally determined.

Interestingly, by comparison to the CD45[−] subset of hESC-derived hematopoietic cells, the later emerging CD45⁺ cells showed fewer differentially expressed genes when both were compared to the FL lin[−]CD34⁺CD38[−] cells. This was explained in part by higher common expression of genes associated with myeloid commitment in CD45⁺ cells but also others associated with HSC self-renewal and survival that were not detected in the CD45[−] subset of hESC-derived hematopoietic cells. The two hESC derived subsets, when directly compared, also showed differences in expression of genes regulating cellular biosynthetic process, including protein synthesis (higher in the CD45⁺ subset), and genes negatively regulating cell growth and division (higher in the CD45[−] subset). The latter are features reported to be characteristic of the most primitive quiescent hematopoietic cells [55,56].

Cell adhesion molecules play a critical role in the homing of HSCs and hence in their detection using standard transplantation protocols. Interestingly, we did not detect significant expression of *CXCR4* at either the transcript or protein level in any of the hESC-derived hematopoietic cells analyzed here. Given the importance of *CXCR4* associated with HSC homing [57], the very low *CXCR4* expression could contribute to an inability of HSC detection simply because they had not yet acquired a competent homing capability. Previously we and others [2,7] demonstrated high expression of *CXCR4* on hESC-derived CD34⁺ cells, but more recent analyses indicate that this is restricted to CD34⁺ endothelial cells that arise concurrently.

In conclusion, our current findings demonstrated interesting similarities in the transcriptional programs of primitive hematopoietic cells generated from hESC *in vitro* when these were compared to analogous populations that arise endogenously in the human embryo. In addition, unique features of each subset were revealed. These findings set the stage for future experiments directed at the identification of genes essential for establishing multipotent hematopoietic cells with long-term engraftment potential and obtaining hESC-derived cells with transplantable *in vivo* repopulating activity.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

This work was supported by NIH grants HL081962, and HD044067 (to IS), NIH grant P51 RR000167 to the Wisconsin National Primate Research Center, University of Wisconsin-Madison, and grants funded by Genome British Columbia/Canada (to MM and CE), the Stem Cell Network, the Canadian Institutes of Health Research and the National Cancer Institute of Canada (with funds from the Terry Fox Run) (to CE). YZ was supported by a Leukemia Research Fund of Canada Fellowship. We thank members of the Stem Cell Assay Service and the Flow Cytometry Core of the Terry Fox Laboratory for assistance in the processing and functional characterization of the FL and CB samples and Joan Larson for editorial assistance.

References

1. Vodyanik MA, Thomson JA, Slukvin II. Leukosialin (CD43) defines hematopoietic progenitors in human embryonic stem cell differentiation cultures. *Blood* 2006;108:2095–2105. [PubMed: 16757688]
2. Vodyanik MA, Bork JA, Thomson JA, et al. Human embryonic stem cell-derived CD34⁺ cells: efficient production in the coculture with OP9 stromal cells and analysis of lymphohematopoietic potential. *Blood* 2005;105:617–626. [PubMed: 15374881]

3. Mikkola HK, Orkin SH. The journey of developing hematopoietic stem cells. *Development* 2006;133:3733–3744. [PubMed: 16968814]
4. Keller G. Embryonic stem cell differentiation: emergence of a new era in biology and medicine. *Genes Dev* 2005;19:1129–1155. [PubMed: 15905405]
5. Lensch MW, Daley GQ. Scientific and clinical opportunities for modeling blood disorders with embryonic stem cells. *Blood* 2006;107:2605–2612. [PubMed: 16332966]
6. Olsen AL, Stachura DL, Weiss MJ. Designer blood: creating hematopoietic lineages from embryonic stem cells. *Blood* 2006;107:1265–1275. [PubMed: 16254136]
7. Tian X, Woll PS, Morris JK, et al. Hematopoietic engraftment of human embryonic stem cell-derived cells is regulated by recipient innate immunity. *Stem Cells* 2006;24:1370–1380. [PubMed: 16456127] Epub 2006 Feb 1372
8. Wang L, Menendez P, Shojaei F, et al. Generation of hematopoietic repopulating cells from human embryonic stem cells independent of ectopic HOXB4 expression. *J Exp Med* 2005;201:1603–1614. [PubMed: 15883170] Epub 2005 May 1609
9. Amit M, Carpenter MK, Inokuma MS, et al. Clonally derived human embryonic stem cell lines maintain pluripotency and proliferative potential for prolonged periods of culture. *Developmental Biology* 2000;227:271–278. [PubMed: 11071754]
10. Vodyanik MA, Slukvin II. Hematoendothelial differentiation of human embryonic stem cells. *Curr Protoc Cell Biol.* 2007Chapter: Unit 23.26
11. Hogge DE, Lansdorp PM, Reid D, et al. Enhanced detection, maintenance, and differentiation of primitive human hematopoietic cells in cultures containing murine fibroblasts engineered to produce human steel factor, interleukin-3, and granulocyte colony-stimulating factor. *Blood* 1996;88:3765–3773. [PubMed: 8916940]
12. Zhao Y, Raouf A, Kent D, et al. A modified polymerase chain reaction-long serial analysis of gene expression protocol identifies novel transcripts in human CD34+ bone marrow cells. *Stem Cells* 2007;25:1681–1689. [PubMed: 17412892] Epub 2007 Apr 1685
13. Robertson N, Oveisi-Fordorei M, Zuyderduyn SD, et al. DiscoverySpace: an interactive data analysis application. *Genome Biol* 2007;8:R6. [PubMed: 17210078]
14. Saeed AI, Sharov V, White J, et al. TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* 2003;34:374–378. [PubMed: 12613259]
15. Eisen MB, Spellman PT, Brown PO, et al. Cluster analysis and display of genomewide expression patterns. *Proc Natl Acad Sci U S A* 1998;95:14863–14868. [PubMed: 9843981]
16. Audic S, Claverie JM. The significance of digital gene expression profiles. *Genome Res* 1997;7:986–995. [PubMed: 9331369]
17. Hirst M, Delaney A, Rogers SA, et al. LongSAGE profiling of nine human embryonic stem cell lines. *Genome Biol* 2007;8:R113. [PubMed: 17570852]
18. Shaulian E, Karin M. AP-1 as a regulator of cell life and death. *Nat Cell Biol* 2002;4:E131–E136. [PubMed: 11988758]
19. Hoeflich A, Reisinger R, Lahm H, et al. Insulin-like growth factor-binding protein 2 in tumorigenesis: protector or promoter? *Cancer Res* 2001;61:8601–8610. [PubMed: 11751371]
20. Schutt BS, Langkamp M, Rauschnabel U, et al. Integrin-mediated action of insulin-like growth factor binding protein-2 in tumor cells. *J Mol Endocrinol* 2004;32:859–868. [PubMed: 15171717]
21. Fukushima T, Tezuka T, Shimomura T, et al. Silencing of insulin-like growth factor-binding protein-2 in human glioblastoma cells reduces both invasiveness and expression of progression-associated gene CD24. *J Biol Chem* 2007;282:18634–18644. [PubMed: 17475624] Epub 12007 May 18631
22. Lelias JM, Adra CN, Wulf GM, et al. cDNA cloning of a human mRNA preferentially expressed in hematopoietic cells and with homology to a GDP-dissociation inhibitor for the rho GTP-binding proteins. *Proc Natl Acad Sci U S A* 1993;90:1479–1483. [PubMed: 8434008]
23. Witke W. The role of profilin complexes in cell motility and other cellular processes. *Trends Cell Biol* 2004;14:461–469. [PubMed: 15308213]
24. Millard TH, Behrendt B, Launay S, et al. Identification and characterisation of a novel human isoform of Arp2/3 complex subunit p16-ARC/ARPC5. *Cell Motil Cytoskeleton* 2003;54:81–90. [PubMed: 12451597]

25. Dai H, Huang W, Xu J, et al. Binding model of human coactosin-like protein with filament actin revealed by mutagenesis. *Biochim Biophys Acta* 2006;1764:1688–1700. [PubMed: 17070122]Epub 2006 Sep 1623
26. Karetsou Z, Martic G, Tavoulari S, et al. Prothymosin alpha associates with the oncoprotein SET and is involved in chromatin decondensation. *FEBS Lett* 2004;577:496–500. [PubMed: 15556635]
27. Harrer M, Luhrs H, Bustin M, et al. Dynamic interaction of HMGA1a proteins with chromatin. *J Cell Sci* 2004;117:3459–3471. [PubMed: 15213251]Epub 2004 Jun 3422
28. Gabory A, Ripoche MA, Yoshimizu T, et al. The H19 gene: regulation and function of a non-coding RNA. *Cytogenet Genome Res* 2006;113:188–193. [PubMed: 16575179]
29. Ji P, Diederichs S, Wang W, et al. MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer. *Oncogene* 2003;22:8031–8041. [PubMed: 12970751]
30. Hao Y, Crenshaw T, Moulton T, et al. Tumour-suppressor activity of H19 RNA. *Nature* 1993;365:764–767. [PubMed: 7692308]
31. Schoenfelder S, Smits G, Fraser P, et al. Non-coding transcripts in the H19 imprinting control region mediate gene silencing in transgenic *Drosophila*. *EMBO Rep* 2007;8:1068–1073. [PubMed: 17948025]Epub 2007 Oct 1019
32. Meier N, Krpic S, Rodriguez P, et al. Novel binding partners of Ldb1 are required for haematopoietic development. *Development* 2006;133:4913–4923. [PubMed: 17108004]Epub 2006 Nov 4915
33. Wang Q, Stacy T, Miller JD, et al. The CBFbeta subunit is essential for CBFalpha2 (AML1) function in vivo. *Cell* 1996;87:697–708. [PubMed: 8929538]
34. Aikawa Y, Nguyen LA, Isono K, et al. Roles of HIPK1 and HIPK2 in AML1- and p300-dependent transcription, hematopoiesis and blood vessel formation. *EMBO J* 2006;25:3955–3965. [PubMed: 16917507]Epub 2006 Aug 3917
35. Mukhopadhyay M, Teufel A, Yamashita T, et al. Functional ablation of the mouse Ldb1 gene results in severe patterning defects during gastrulation. *Development* 2003;130:495–505. [PubMed: 12490556]
36. Wang Q, Stacy T, Binder M, et al. Disruption of the Cbfa2 gene causes necrosis and hemorrhaging in the central nervous system and blocks definitive hematopoiesis. *Proc Natl Acad Sci U S A* 1996;93:3444–3449. [PubMed: 8622955]
37. North T, Gu TL, Stacy T, et al. Cbfa2 is required for the formation of intra-aortic hematopoietic clusters. *Development* 1999;126:2563–2575. [PubMed: 10226014]
38. Robertson SM, Kennedy M, Shannon JM, et al. A transitional stage in the commitment of mesoderm to hematopoiesis requiring the transcription factor SCL/tal-1. *Development* 2000;127:2447–2459. [PubMed: 10804185]
39. Shivdasani RA, Mayer EL, Orkin SH. Absence of blood formation in mice lacking the T-cell leukaemia oncoprotein tal-1/SCL. *Nature* 1995;373:432–434. [PubMed: 7830794]
40. Manaia A, Lemarchandel V, Klaine M, et al. Lmo2 and GATA-3 associated expression in intraembryonic hemogenic sites. *Development* 2000;127:643–653. [PubMed: 10631184]
41. Kim I, Saunders TL, Morrison SJ. Sox17 dependence distinguishes the transcriptional regulation of fetal from adult hematopoietic stem cells. *Cell* 2007;130:470–483. [PubMed: 17655922]Epub 2007 Jul 2026
42. Sato Y. Role of ETS family transcription factors in vascular development and angiogenesis. *Cell Struct Funct* 2001;26:19–24. [PubMed: 11345500]
43. Kondo M, Wagers AJ, Manz MG, et al. Biology of hematopoietic stem cells and progenitors: implications for clinical application. *Annu Rev Immunol* 2003;21:759–806. [PubMed: 12615892]Epub 2002 Dec 2017
44. Medvinsky AL, Samoylina NL, Muller AM, et al. An early pre-liver intraembryonic source of CFU-S in the developing mouse. *Nature* 1993;364:64–67. [PubMed: 8316298]
45. Gekas C, Dieterlen-Lievre F, Orkin SH, et al. The placenta is a niche for hematopoietic stem cells. *Dev Cell* 2005;8:365–375. [PubMed: 15737932]
46. Cumano A, Godin I. Ontogeny of the Hematopoietic System. *Annual Review Immunology* 2007;2007:745–785.

47. Rebel VI, Miller CL, Thornbury GR, et al. A comparison of long-term repopulating hematopoietic stem cells in fetal liver and adult bone marrow from the mouse. *Exp Hematol* 1996;24:638–648. [PubMed: 8605969]
48. Pawliuk R, Eaves C, Humphries RK. Evidence of both ontogeny and transplant dose-regulated expansion of hematopoietic stem cells in vivo. *Blood* 1996;88:2852–2858. [PubMed: 8874181]
49. Bowie MB, McKnight KD, Kent DG, et al. Hematopoietic stem cells proliferate until after birth and show a reversible phase-specific engraftment defect. *J Clin Invest* 2006;116:2808–2816. [PubMed: 17016561]
50. Lessard J, Faubert A, Sauvageau G. Genetic programs regulating HSC specification, maintenance and expansion. *Oncogene* 2004;23:7199–7209. [PubMed: 15378080]
51. Pina C, Enver T. Differential contributions of haematopoietic stem cells to foetal and adult haematopoiesis: insights from functional analysis of transcriptional regulators. *Oncogene* 2007;26:6750–6765. [PubMed: 17934483]
52. Marshall CJ, Moore RL, Thorogood P, et al. Detailed characterization of the human aorta-gonad-mesonephros region reveals morphological polarity resembling a hematopoietic stromal layer. *Dev Dyn* 1999;215:139–147. [PubMed: 10373018]
53. Debacker C, Catala M, Labastie MC. Embryonic expression of the human GATA-3 gene. *Mech Dev* 1999;85:183–187. [PubMed: 10415360]
54. North TE, de Bruijn MF, Stacy T, et al. Runx1 expression marks long-term repopulating hematopoietic stem cells in the midgestation mouse embryo. *Immunity* 2002;16:661–672. [PubMed: 12049718]
55. Forsberg EC, Prohaska SS, Katzman S, et al. Differential expression of novel potential regulators in hematopoietic stem cells. *PLoS Genet* 2005;1:e28. [PubMed: 16151515]
56. Venezia TA, Merchant AA, Ramos CA, et al. Molecular signatures of proliferation and quiescence in hematopoietic stem cells. *PLoS Biol* 2004;2:e301. [PubMed: 15459755] Epub 2004 Sep 2028
57. Peled A, Petit I, Kollet O, et al. Dependence of human stem cell engraftment and repopulation of NOD/SCID mice on CXCR4. *Science* 1999;283:845–848. [PubMed: 9933168]

Figure 1.

Major subsets of CD34⁺ cells identified in hESC-OP9 cocultures after 8 days of incubation. (A) hESC-derived CD34⁺ cells represent a heterogeneous population which includes CD31⁺CD43⁻ endothelial cells (EC), CD31⁻CD43⁻ mesenchymal cells (MC), and CD43⁺ hematopoietic progenitor (HP)-enriched cells. Within the CD43⁺ population, the expression of Glycophorin A (CD235a) and CD41a defines erythro-megakaryocytic progenitor cells (E-MkPs). Two distinct populations of multipotent cells can be recognized based on their expression of CD45: lin⁻CD43⁺CD45⁻ (mHP1) cells and lin⁻CD43⁺CD45⁺ (mHP2) cells. > 95–98% of mHP1 and mHP2 cells are CD38⁻ (2). (B) Representative FACS analysis of major subsets of CD43⁺ HPs generated in hESC-OP9 coculture.

Figure 2.

Cluster analysis of seven LongSAGE libraries prepared from primitive hESC-derived and FL and CB hematopoietic cells. (A) Library tree generated using the hierarchical clustering algorithm of the MeV software. Both hESC-derived hematopoietic progenitor (lin⁻CD34⁺CD43⁺CD45⁻ and CD45⁺) libraries were found to be more closely related to the FL library than to either of the CB libraries. (B) Clusters reflecting the expression of lineage-specific genes and transcriptional regulators of lineage commitment in the same libraries. MeV software was used to compare gene expression profiles of all 7 libraries. A representation of gene expression is shown in a color scale ranging from black (no expression) to bright red (high

expression levels). The gradient upper limit was set arbitrarily to 30 (or 20 for lymphoid genes) tag counts.

Figure 3.

Analysis of differentially expressed genes in the FL and hESC-derived lin⁻ CD34⁺CD43⁺ cell libraries. Top 25 overexpressed tags and their non-normalized counts in each of the 6 pairs of comparisons are presented. In the middle, similarity scatter plots from a Poisson model generated using Discovery software are shown. The number of differentially expressed tags/genes using the 99.9% statistical difference cut off is indicated along each axis.

Figure 4.

Distribution of differentially expressed transcripts according to Gene Ontologies (GO) in biological processes categories. Data generated by DiscoverySpace software were analyzed using FatiGO+ software. Each bar shows the number of significantly overexpressed genes ($p < 0.05$) in a given GO category (level 7) expressed as a percent of the total number of genes expressed in that category.

Figure 5.

Expression of specific genes involved in HSC development, self-renewal, and survival, and AP-1 complex genes in FL $\text{lin}^{-}\text{CD34}^{+}\text{CD38}^{-}$ cells and CD45^{-} and CD45^{+} subsets of hESC-derived $\text{lin}^{-}\text{CD34}^{+}\text{CD43}^{+}$ cells. Normalized tag counts for each gene analyzed are shown in the horizontal black bars. Statistically significant differences in tag counts ($p < 0.001$) are indicated by (*) for differences between the FL library and the hESC-derived $\text{lin}^{-}\text{CD34}^{+}\text{CD43}^{+}\text{CD45}^{-}$ cell library, and by (+) for differences between the FL library and the hESC-derived $\text{lin}^{-}\text{CD34}^{+}\text{CD43}^{+}\text{CD45}^{+}$ cell library. Graphs on the right show values (mean $\pm$ SE for at least two experiments) on a logarithmic scale for gene expression levels in the hESC-derived $\text{lin}^{-}\text{CD34}^{+}\text{CD43}^{+}\text{CD45}^{-}$ cells (grey bars) and in the hESC-derived

lin⁻CD34⁺CD43⁺CD45⁺ cells (white bars) relative to FL lin⁻CD34⁺CD38⁻ cells (set to 1) as determined by Q-RT-PCR. The literature references related to the functional significance of each listed gene for HSCs are provided in Data Supplements, S Additional References.

Figure 6.

Comparison of levels of expression of HOX, polycomb, and trithorax group genes in primitive hematopoietic cells derived from hESCs and isolated from FL. Normalized tag counts for each gene analyzed are shown in the horizontal black bars. Statistically significant differences in gene expression ($p < 0.001$) between FL and hESC-derived cells are indicated by a symbol next to the bar: (*) for differences between FL and hESC-derived $\text{lin}^{-}\text{CD34}^{+}\text{CD43}^{+}\text{CD45}^{-}$ cells and (+) for differences between FL and hESC-derived $\text{lin}^{-}\text{CD34}^{+}\text{CD43}^{+}\text{CD45}^{+}$ cells.

Expression of some genes from each group was validated by qRT-PCR (graphs on the right). The graphs on the right show transcript levels on a logarithmic scale (mean $\pm$ SE for at least two experiments) in hESC-derived $\text{lin}^{-}\text{CD34}^{+}\text{CD43}^{+}\text{CD45}^{-}$ cells (grey bars) and in hESC-

derived $\text{lin}^- \text{CD34}^+ \text{CD43}^+ \text{CD45}^+$ cells (white bars) relative to FL $\text{lin}^- \text{CD34}^+ \text{CD38}^-$ cells (set to 1) as determined by Q-RTPCR.

Figure 7.

Flow cytometric analysis of cell surface proteins implicated in HSC migration and homing. The black line represents hESC-derived $\text{lin}^- \text{CD34}^+ \text{CD43}^+ \text{CD45}^-$ cells and the dotted line hESC-derived $\text{lin}^- \text{CD34}^+ \text{CD43}^+ \text{CD45}^+$ cells, compared to the Ig isotype control in grey.

Table 1
Distribution of SAGE tags in hESC, FL, and CB-derived primitive hematopoietic cell libraries

	hESC CD43 ⁺ CD235a ⁺ CD41a ^{+/−}	hESC lin [−] CD34 ⁺ CD43 ⁺ CD45 [−]	hESC lin [−] CD34 ⁺ CD43 ⁺ CD45 ⁺	FL lin [−] CD34 ⁺ CD38 [−]	CB lin [−] CD34 ⁺ CD38 [−]	CB lin [−] CD34 ⁺ CD38 [−]	hESCs Undifferentiated
Total Tags	153,554	159,535	185,982	225,194	191,931	201,337	276,203
Unique tag sequences	20,051	14,936	40,415	66,722	53,283	59,989	71,104
Tags mapping to a single gene	9,386	6,382	16,977	13,230	12,550	16,350	23,004
Tags mapping to a multiple genes	620	778	1,671	1,755	1,754	980	1,994
Unique mapped transcripts	5,860	4,702	8,586	8,307	6,895	8,094	10,145
Tags with no RefSeq mapping	10,045	7,776	21,767	51,737	38,979	42,659	46,106