Differential gene expression profiling of adult murine hematopoietic stem cells

In-Kyung Park, Yaqin He, Fangming Lin, Ole D. Laerum, Qiang Tian, Roger Bumgarner, Christopher A. Klug, Kaijun Li, Christian Kuhr, Michelle J. Doyle, Tao Xie, Michèl Schummer, Yu Sun, Adam Goldsmith, Michael F. Clarke, Irving L. Weissman, Leroy Hood, and Linheng Li

Hematopoietic stem cells (HSCs) have self-renewal capacity and multilineage developmental potentials. The molecular mechanisms that control the self-renewal of HSCs are still largely unknown. Here, a systematic approach using bioinformatics and array hybridization techniques to analyze gene expression profiles in HSCs is described. To enrich mRNAs predominantly expressed in uncommitted cell lineages, 54 000 cDNA clones generated from a highly enriched population of HSCs and a mixed population of stem and early multipotent progenitor (MPP) cells were arrayed on nylon membranes (macroarray or high-density array), and subtracted with cDNA probes derived from mature lineage cells including spleen, thymus, and bone marrow. Five thousand cDNA clones with very low hybridization signals were selected for sequencing and further analysis using microarrays on glass slides. Two populations of cells, HSCs and MPP cells, were compared for differential gene expression using microarray analysis. HSCs have the ability to self-renew, while MPP cells have lost the capacity for self-renewal. A large number of genes that were differentially expressed by enriched populations of HSCs and MPP cells were identified. These included transcription factors, signaling molecules, and previously unknown genes. (Blood. 2002;99:488-498)

Introduction

Hematopoiesis is a dynamic process with significant complexity in which a subset of hematopoietic stem cells (HSCs) gives rise to cells of both the myeloid and lymphoid lineages.[1] In addition, HSCs have the ability to self-renew to produce more HSCs. This property allows HSCs to repopulate the bone marrow of lethally irradiated congenic hosts.

Mouse HSCs isolated with Thy1.1^{lo}c-kit^{+}Sca-1^{hi}Lin$^{-/lo}$ markers using fluorescence activated cell sorting (FACS) represent only approximately 0.05% of mouse bone marrow cells and these cells can fully reconstitute all blood cell elements.[2-4] This population of cells has been further divided into 3 subpopulations[4,5]: long-term HSCs (LT-HSCs), short-term HSCs (ST-HSCs), and multipotent progenitor (MPP) cells, according to their abilities to support hematopoiesis and self-renewal. These 3 populations of cells can be arranged in a lineage according to a progressive loss of the ability to self-renew.[6] The LT-HSC population has extensive self-renewal ability and supports long-term reconstituting ability ($>$ 6 months), representing approximately 0.005% to 0.01% of bone marrow cells.[4] ST-HSCs with limited self-renewal ability briefly contribute to hematopoiesis (6-8 weeks). MPP cells cannot self-renew but can reconstitute bone marrow for less than 4 weeks.[5] Efflux of dyes such as rhodamine-123 (Rh) can be used to separate early hematopoietic cells into HSC and early progenitor subpopulations by flow cytometry (see Pohlmann et al[7] and Weissman[8] for detailed description). Rh is a mitochondria-binding fluorescent dye, and can be effluxed from the cell by the ABC transporter including the multidrug-resistance (MDR) gene product, P glycoprotein.[12] LT-HSCs, which are relatively quiescent,[13] have high ABC transporter activity, thus the population of cells that stain most weakly with this dye is highly enriched for LT-HSCs. Alternatively, LT-HSCs, ST-HSCs, and MPP cells can be identified according to the expression levels of lineage-associated antigens such as Mac-1 and CD4.[4,5]

The rarity of the HSCs, coupled with an inability to maintain these cells in culture for a significant period, has greatly hindered the biochemical and molecular characterization of these cells. Although various growth factors can induce stem cells to proliferate in vitro, no known combination of growth factors has been shown to promote sustained self-renewal of the HSCs. Rather, stem cells induced to proliferate in vitro invariably undergo concomitant differentiation. HSC functions are thought to be regulated by interactions of stem cells with stromal cells.[14] The dynamic process of HSC development, including self-renewal, expansion, and maturation, involves interactions between the intrinsic genetic program and extrinsic signals from stromal cells. This dynamic process is accompanied by global changes in gene expression profiles involving molecular events such as signal transduction, transcriptional and translational regulation, and chromatin modification. In the present study, we intend to systematically look into these molecular events by analyzing the gene expression profiles of self-renewing HSCs and non–self-renewing MPP cells. Therefore, any changes in the components of these aforementioned molecular events may provide insight into understanding the molecular mechanisms in controlling self-renewal of HSCs.

From the University of Michigan, Department of Internal Medicine, Ann Arbor; the Stowers Institute for Medical Research, Kansas City, MO; the University of Washington, Departments of Pediatrics, Microbiology, Surgery, and Molecular Biotechnology, Seattle; the University of Bergen, Department of Pathology, Norway; the University of Alabama at Birmingham, Department of Microbiology; Stanford University, Department of Pathology, Stanford, CA; and the Institute for Systems Biology, Seattle, WA.

Submitted June 8, 2001; accepted September 14, 2001.

Supported in part by National Institutes of Health grant 1P01 DK53074-02 and Stowers Institute for Medical Research.

I.-K.P. and Y.H. contributed equally to this article.

Reprints: Linheng Li, Stowers Institute for Medical Research, 1000 E 50th St, Kansas City, MO 64110; e-mail: lil@stowers-institute.org.

The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 U.S.C. section 1734.

Materials and methods

Isolation of HSCs and generation of cDNA libraries

For cDNA library A, 4.7×10^9 bone marrow cells were collected from the femurs and tibias of 60 C57BL/6J mice (6-8 weeks old). These bone marrow cells were incubated with rat monoclonal antibodies against lineage-positive markers (Pharmingen, San Jose, CA) including CD4 (RM4-4), CD2 (RM2-5), CD45R/B220 (RA3-6B2), Gr-1 (RB6-8C5), Mac-1 (M1/70), TER-119 (TER-119), and IL-7Rα chain (B12-1); lineage negative ($Lin^{-/lo}$) cells were enriched by twice depleting lineage-positive cells through incubation with antibody-coated (sheep antirat IgG) Dynabeads M-450 (Dynal). The remaining $Lin^{-/lo}$ cells were stained with Sca-1 fluorescein isothiocyanate (FITC) for isolation of $Lin^{-/lo}Sca\text{-}1^+$ cells.

The cDNA library A was constructed from 1.3×10^6 $Lin^{-/lo}Sca\text{-}1^+$ cells described above using the ZAP expression cDNA synthesis kit following the manufacturer's procedure (Stratagene). Briefly, total RNA was isolated from the $Lin^{-/lo}Sca\text{-}1^+$ cells using Trizol Reagent (BRL). The cDNA was synthesized first by reverse transcription using Superscript II (Invitrogen, Carlsbad, CA) and then by DNA synthesis using Klenow DNA polymerase (Invitrogen). The cDNA inserts were cut with *Eco*RI and *Xho*I restriction enzymes and cloned into the *Eco*RI/*Xho*I sites of λZAPExp vector. The plasmids (pBK; Stratagene, La Jolla, CA) bearing the cDNA inserts were excised from their parent vector λZAPExp using helper phage according to the manufacturer's protocol. The primary cDNA library contained 36 000 clones.

The cDNA library B was made from 2.8×10^4 twice-sorted $Lin^{-/lo}Sca\text{-}1^+Thy\text{-}1.1^{lo}c\text{-}kit^+$ HSCs. Isolation of a population of $Lin^{-/lo}Thy\text{-}1.1^{lo}c\text{-}kit^+Sca\text{-}1^+$ HSCs, RNA extraction, cDNA synthesis, and polymerase chain reaction (PCR) amplification have been described previously.[15] There were 7.5×10^6 clones in the original library.

Probe preparation for macroarray and microarray analyses

For preparation of cDNA probes used for macroarray hybridization, total RNA isolated from lineage-positive cells derived from thymus, spleen, and bone marrow was used as template to generate the first strand of cDNAs using SuperScript II (Invitrogen), incorporating ^{33}P-dCTP into the cDNA probes during the reverse transcription.

For preparation of cDNA probes used for microarray hybridization, c-kit-phycoerythrin (PE) (2B8) and Sca-1-biotin (D7)/Streptavidin-Cy-Chrome (Pharmingen) were used to stain $Lin^{-/lo}$ cells prior to separation using a Vantage fluorescence-activated cell sorter (Becton Dickinson, San Jose, CA). In order to isolate HSCs from MPP cells, Rh (0.1 μg/mL; Sigma, St Louis, MO) was used to separate $Lin^{-/lo}c\text{-}kit^+Sca\text{-}1^+$ cells into Rh^{lo} (lowest 15%) and Rh^{hi} (highest 15%) populations. We were able to obtain 1.4×10^4 $Rh^{lo}Lin^{-/lo}c\text{-}kit^+Sca\text{-}1^+$ and 1.5×10^4 $Rh^{hi}Lin^{-/lo}c\text{-}kit^+Sca\text{-}1^+$ cells from 15 mice. The postsorting analysis showed that the purity of the sorted $Lin^{-/lo}c\text{-}kit^+Sca\text{-}1^+$ cells was 98%. Total RNA was isolated from equal numbers of $Rh^{lo}Lin^{-/lo}c\text{-}kit^+Sca\text{-}1^+$ and $Rh^{hi}Lin^{-/lo}c\text{-}kit^+Sca\text{-}1^+$ cells (1×10^4). In order to amplify the cDNA for microarray hybridization, PCR amplification was performed using pSMART cDNA synthesis strategy following the manufacturer's procedure (ClonTech, San Jose, CA). Briefly, $VNdT^{20}$ (5′-GACTCTAGAGCGGCCGC-CTTTTTTTTTTTTTTTTTTTTVN-3′, V = A, G, C; n = T, A, G, C) was used for the first-strand cDNA synthesis. 5′-cap oligo (5′-TACGGCTGCGAGAA-GACGACAGAAGGG-3′) was used for the second-strand cDNA synthesis. Both the VNdT and 5′-cap primer (5′-TACGGCTGCGA GAAGACGACAGAA-3′) were used for the PCR amplification with limited amplification cycles (20 cycles were used in this study). Cy3-dCTP (for Rh^{lo} cells) and Cy5-dCTP (for Rh^{hi} cells) fluorescence were incorporated during the PCR reaction. The unincorporated nucleotides were removed through a Millipore column (Millipore, Bedford, MA). The optical density $(OD)_{200\text{-}700nm}$ of the probes was measured using a spectrophotometer. The incorporation rate was calculated subsequently according to the ratio of the absorbance of Cy3 (523 nm) and Cy5 (633 nm) over the DNA absorbance at 260 nm.

Macroarray preparation, hybridization, and analysis

To prepare for macroarray, the 2 HSC cDNA libraries, A and B, described above were plated onto LB agarose/ampicillin plates in 24 × 24-cm Q-trays (Genetix, Christchurch, England) with densities of 500 to 2000 colonies per tray, and incubated overnight at 37°C. The bacterial colonies were then picked up by a Q-bot robotic system (Genetix) and inoculated into 384-well plates. Thus, each cDNA clone was addressed by its position in the plate. The macroarrays were generated by spotting bacterial colonies from 384-well plates onto 24 × 24-cm nylon membranes using Q-bot (Genetix). Thirty-six thousand cDNA clones from library A and 18 000 clones from library B were spotted in duplicate, and the position of each clone on the membrane was recorded according to its address in the 384-well plate. These macroarray membranes were incubated on the top of LB-agarose plus antibiotics at 30°C overnight in Q-trays. DNAs were denatured by soaking the membrane in 0.5 N NaOH/1.5 M NaCl for 10 minutes and neutralized with 0.5 M Tris-Cl, pH 8.0/1.5 M NaCl for 5 minutes. DNAs were fixed by ultraviolet (UV) crosslinking (UVP, Upland, CA). For hybridization, membranes were prehybridized for 2 hours at 60°C in 100 mL solution (12.5 mM PEG8000; 250 mM NaCl; 85 mM Na_2HPO_4; 7.5 mM H_3PO_4 [86%]; 243 mM sodium dodecyl sulfate [SDS]; 10 mM ethylenediaminetetraacetic acid [EDTA]; 32 mM NaOH, pH 7.2) containing 100 μg/mL of denatured salmon sperm DNA and 100 μg/mL denatured mouse Cot-1 DNA. Denatured ^{33}P-labeled probes from mature cell lineages were added to the prehybridized solution and incubated overnight in roller bottle or shaking bath at 60°C. Membranes were washed twice with 2x sodium chloride sodium citrate (SSC)/1% SDS at room temperature for 15 minutes, and then once in 0.2x SSC/0.1% SDS at 60°C for 30 minutes. Membranes were exposed to a phosphor imager screen and scanned using a phosphoimage scanner (Molecular Dynamics, Sunnyvale, CA). The hybridization result was analyzed using Spotfinder for image processing (H. Hammersmark and R.B., University of Washington, Seattle). Since each cDNA clone was spotted in duplicate, an average signal intensity of the duplicate was used. Local background hybridization signals surrounding each spot were subtracted prior to comparing spot intensity. Quantitative comparison of signal intensities was analyzed using Microsoft Excel. In order to avoid false-negative signals generated from slow or nongrowing bacterial colonies, a probe derived from the vector sequence was hybridized to the macroarray in parallel. The negative selection process was performed by selecting cDNA clones with positive signals for vector hybridization and negative or very weak signals for the cDNA probes derived from lineage-positive cells. Using this process, 3000 of 36 000 cDNA clones and 2000 of 18 000 cDNA clones were selected from library A and B, respectively, for microarray analysis.

Microarray preparation, hybridization, and analysis

cDNA inserts from 5000 clones selected using macroarray analysis described above were PCR-amplified using primers derived from the vector. For library A, ZAP-F (5′-AGTGGATCCAAAGAATTC-3′) and ZAP-R (5′-CTCTAGAAGTACTCTCGAG-3′) were used, and for library B, SubA1 (5′-CTTCGAACCGCGGATATCAGATC-3′) and SubS2 (5′-AAGGTTCCTTCACAAAGATCCCTCGAG-3′) were used. Amplified PCR products were purified using Sephacryl S500 (Pharmacia, Peapack, NJ), mixed 1 to 1 with dimethyl sulfoxide (DMSO) (Amersham, Piscataway, NJ), and spotted in duplicate onto each 75 × 25-mm coated type 7 microarray slide (Amersham) using Genespoter II (Molecular Dynamics). The microarray slides were air-dried and DNA was fixed by UV crosslinking at 500 mJ. The microarray slides were hybridized with Cy3- and Cy5-labeled probes described above. Briefly, the purified probes were concentrated with a speed-vacuum in the dark and resuspended in 20 μL hybridization buffer (5x SSC, 5x Denhardt, 0.1% SDS, 50% formamide, 100 μg/mL denatured salmon sperm DNA, 20 μg/mL polyA60 RNA, 100 μg/mL mouse Cot I DNA). An equal amount of Cy3- and Cy5-labeled probes were combined, denatured, and added onto the microarray slides, and hybridized at 42°C in a hybridization oven or in a humidified chamber for 16 hours to 18 hours. Microarray glass slides were washed twice with 2x SSC/0.1% SDS, and 0.2x SSC/0.1% SDS at 55°C for 5 minutes, and then with 0.1x SSC at room temperature for 1 minute. Glass slides were

immersed in water for 10 seconds and dried immediately with N_2. The hybridized microarray was scanned with a confocal dual-laser scanner at 523 nm and 633 nm (Molecular Dynamics) and analyzed using the custom array analysis software developed at the University of Washington (H. Hammersmark and R.B.). This array analysis software includes image processing, data normalization, and error analysis as described previously.[16]

Reverse transcriptase–polymerase chain reaction

Reverse-transcription reactions were carried out using SuperScriptII following the manufacturer's manual (Invitrogen). The following primers were used for PCR: Activin βC, 5′-GACACCTTACTCTGGAGCTG-3′ and 5′-GGGAGGCAGA GTAGATTACA-3′; BA_RA6B66, 5′-TCATGGTATGCCCTCGTGTA-3′ and 5′-AAATGTGTGGGCTTTTCAGG-3′; BA_RA5A82, 5′-AACATGGCTTGGAGAC AACC-3′ and 5′-AAGCCTGGGATTCACTGCTA-3′; BA_RA4B58, 5′-ATGAGGGC CATTGTTACACC-3′ and 5′-TTATGGCCAGCTTGGTTCAC-3′; MSC_RA15C76, 5′-TGATACCCTTGGCTCGAAAC-3′ and 5′-CCAAGTGCTGGGATTAAAGG-3′; MSC_RA16D13, 5′-GGCTCGAAATTAACCCTCAC-3′ and 5′-CCAGATCTCGT TACGGATGG-3′. The PCR condition was 35 cycles consisting of 94°C for 30 seconds, 55°C for 40 seconds, and 72°C for 1 minute.

Competitive repopulation assay

Two hundred fifty cells (either $Rh^{lo}Lin^{-/lo}Sca\text{-}1^{+}c\text{-}kit^{+}$ or $Rh^{hi}Lin^{-/lo}Sca\text{-}1^{+}c\text{-}kit^{+}$) from Ly5.2 mice were mixed with 4×10^4 $Lin^{-/lo}$ cells from congenic host Ly5.1 mice[17] and were then transplanted into lethally irradiated (11 Gy [1100 rads]) Ly5.1 mice by intravenous (IV) injection. Each group included 8 mice. Peripheral blood cells were harvested at different time points after transplantation, as indicated in Figure 2C, and flow cytometric analyses were performed using different specific cell surface markers (CD3, B220, Mac-1, and Gr-1) together with anti-Ly5.2 antibody.

Results

Construction and normalization of HSC cDNA libraries

In order to enrich for genes that are predominantly expressed in the uncommitted, or lineage-negative cell populations, and remove highly redundant and housekeeping genes for further analysis, high-density array[18] (macroarray) analysis was used and cDNA clones that hybridized to the lineage-positive probes were removed. cDNA clones that had no or very weak hybridization signals to the lineage-positive probes were selected. These negatively selected cDNA clones were subsequently subject to analysis of differential gene expression using a microarray system.[19] The strategy for using macroarrays and microarrays to analyze the gene expression patterns of the HSC is illustrated and described in Figure 1A. First, 2 populations of HSCs were isolated from mouse bone marrow cells by FACS based on their cell surface antigen expression: the $Lin^{-/lo}Sca\text{-}1^{+}$ population (A) containing MPP cells, LT-HSCs, and ST-HSCs[20]; and the $Lin^{-/lo}c\text{-}kit^{+}Thy\text{-}1^{lo}Sca\text{-}1^{+}$ population (B) enriched with LT-HSCs.[5] Correspondingly, cDNA libraries A and B were generated from each of these 2 complementary populations of HSCs. Sequence analysis of 200 randomly chosen clones from each library indicated that they are very diverse. Bacterial colonies derived from these HSC cDNA libraries were arrayed onto nylon membranes using the Q-bot robotic system to generate the macroarrays, or high-density cDNA arrays. These macroarrays were hybridized with the cDNA probes generated from committed or mature hematopoietic cell lineages of the thymus, bone marrow, and spleen (lineage-positive probes). Clones that hybridized to the lineage-positive probes were eliminated through the process of

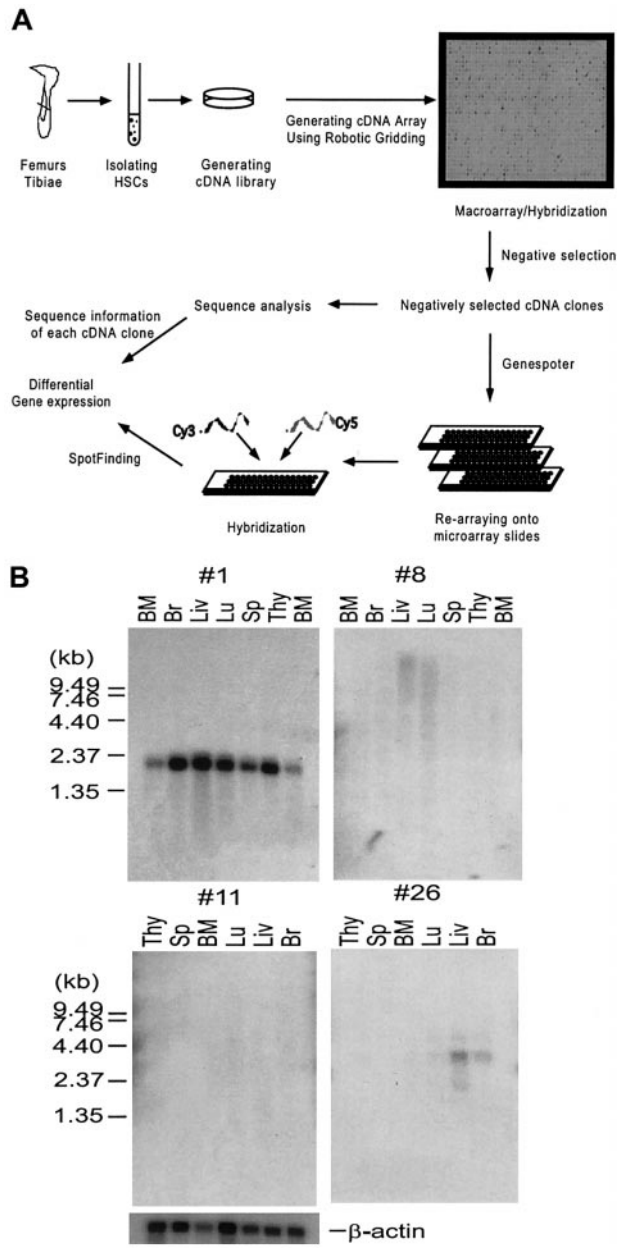

Figure 1. Enrichment of genes that are predominantly expressed in hematopoietic stem cells. (A) A schematic drawing illustrating the strategy of systematic analyses of gene expression in hematopoietic stem cells (HSCs). Bacterial colonies from the 2 HSC cDNA libraries A and B described in the text were spotted in duplicates onto 24 × 24 cm nylon membranes using a Q-bot robotic system (Genetix), generating macroarrays. The macroarray membranes were hybridized with ^{33}P-labeled cDNA probes derived from mature cell lineages described in the text in parallel with the vector-derived probe. Hybridization signals were analyzed using Spotfinder software (see "Materials and methods" for detailed description) and converted to numeric numbers using Microsoft Excel. cDNA clones, showing negative or very weak hybridization signals to the cDNA probes derived from lineage-positive cells but positive to the vector-derived probe, were selected. This process was called negative selection. The negatively selected cDNA clones were analyzed by DNA sequencing, and spotted in duplicate onto microarray slides. Total RNA isolated from $Rh^{lo}Lin^{-/lo}c\text{-}kit^{+}Sca\text{-}1^{+}$ or $Rh^{hi}Lin^{-/lo}c\text{-}kit^{+}Sca\text{-}1^{+}$ HSCs was used to generate Cy3- and Cy5-labeled cDNA probes, respectively, by RT-PCR. The probes were hybridized to the microarray slides and the data were analyzed using array analysis software. Details of the experiment are described in "Materials and methods." (B) Northern analysis of unknown genes from cDNA library B. Two micrograms of poly A^{+} RNA purified from indicated tissues was separated on a 1% agarose-formaldehyde gel, transferred to nylon membrane, and then hybridized with a ^{32}P-labeled probe prepared from each unknown clone. Four representative blots are shown.

Table 1. A partial list of genes analyzed after negative selection of the cDNA library generated from Lin$^{-/lo}$Sca-1^{+}Thy-1.1^{lo}c-kit^{+} cells

Name	Clone code	Accession number	Description
Transcription factors			
Bmi-1	BA_HSC3_18	M64279	Polycomb-group transcriptional repressor; involved in cell proliferation and senescence
Nfe2	BA_RA1B_64	L09600	AP-1–related DNA-binding protein, leucine zipper protein, hematopoietic specific
TBX2	BA_RA3A_94	AF172088	Limb morphogenesis; immortalizing gene with tumorigenic potential
Ash1	BA_RA2A_52	NM_018489	A SET domain, a PHD finger, 4 AT hooks, and a region with homology to the bromodomain; may be involved in adhesion-mediated transcription
Maid	BA_RA2B_83	NM_010761	Negative regulator of bHLH proteins in the mouse egg and zygote
LEF-1S	BA_RA5D_78	D16503	A lymphoid-specific protein with an HMG domain, regulates TCRα enhancer function
Elf1	BA_RA6B_30	NM_007920	Ets-family transcription factor; highest in lymphocytes
ZF216	BA_RA4C_60	AF062071	1 N-terminal A20 (an inhibitor of cell death)-like zinc fingers, mediating IL-1–induced NF-kappaB activation
Evi-1	BA_RA3A_50	M21829	2 sets of ZF; overexpression causes abnormalities in megakaryopoiesis
Evi-9	BA_RA1D_75	NM_016707	A ZF protein; interacts with BCL6, a known human B-cell proto-oncogene product
Tcf4	BA_RA1A_07	NM_013685	A splicing variant of the ITF-2 transcript encoding a transcription factor that inhibits MyoD activity
Chx10	BA_RA2B_69	NM_007701	C elegans ceh-10 homeo-domain–containing homologue; developmentally regulated
Putative myelin regulatory factor 1	BA_RA4A_40	U14648	A brain-derived DNA-binding protein; regulates expression of myelin basic protein gene
KIAA0697	BA_RA4B_88	AB014597	1 TM and sequence-specific RNA binding by a novel KH domain; implicated in paraneoplastic disease and the fragile X syndrome
Membrane proteins			
Notch1	BA_8-3	NM_008714	Controls cell fate choices by inhibiting certain differentiation pathways
Type III multi-pass TM protein	BA_RA1A_50	AF082535	4 TM; a gene similar to Homo sapiens GMP-17/NKG7/GIG-1; expressed by NK, CD8^{+}, activated T, and macrophage R2 cells
Mouse homologue of Nessy	BA_RA1A_60	NM_005768	Putative transmembrane protein Nessy, an evolutionarily conserved gene controlled by Hox proteins during Drosophila embryogenesis
CD34	BA_RA1C_19	S69295	Hematopoietic stem cell antigen, present in embryonic brain and adult brain, lung, and skeletal muscle
CGI-69	BA_RA4D_91	AF151827	Similar to human CGI-69 protein
HSA-A	BA_RA4D_26	X56469	A mouse hematopoietic differentiation marker from B and T cells; signal peptide, 1 C-terminal TM
Human sperm surface Ag	BA_RA4B_33	NM_003971	A novel human testis mRNA specifically expressed in testicular haploid germ cells; carrying leucine zipper dimerization motif
Secreted molecules			
BMP4	BA_RA6B_87	S65032	A signal peptide; secreted protein; TGF-beta family; exhibits mesoderm-inducing activity
L2G25B	BA_RA4B_55	J04491	Signal peptide, CXC chemokine
SDF4	BA_RA5B_24	D50461	A signal peptide with Ca^{2+} binding motif
Flt3 ligand isoform E6 Δ16	BA_RA4A_50	S76464	Signal peptide; flt3 ligand isoform E6 Δ16, missing 1 TM, soluble form
Apoptosis			
Mcl 1	BA_RA3A_85	NM_008562	Myeloid leukemia sequence 1; a bcl-2–related gene; expressed in the early stage of embryonic stem cells
Heterochromatin protein 1α	BA_RA2A_37	AF216290	Apoptosis-associated proteins in a human Burkitt lymphoma cell line
Signaling molecules			
WNK1	BA_RA1A_15	AF227741	9 TM, 3 coil-coil structures; similar to MEKK; unknown function
ARK2	BA_RA1A_91	U69107	Aurora-related kinase 2; involved in chromosomal segregation; expressed during S and G2/M phases; localized to the midbody
STK-1	BA_RA1A_91	D21099	Expressed in a large number of proliferating cells, but not in other tissues; Xenopus lavas XLP46 protein kinase; involved in cell growth
LynB	Ba_RA1D_84	M57697	Tyrosine kinase; expressed in myeloid and B-lymphoid lineage cells
Truncated form of c-kit	BA_RA6B_92	X65997	Kinase domain
GAP	BA_RA5B_81	L13151	GTPase activating protein; inhibits ras-mediated signaling
RGS2	BA_RA6B_56	NM_00906	RGS box; GAP for trimeric G-protein; inhibits signaling from Gq
Nbr1	BA_RA1B_56	AF008676	B box and zinc finger domain–containing protein, unknown function; next to *BRCA1* gene
DMX-like 1	BA_RA6A_84	NM_005509	10 WD and 2 TM; highly conserved throughout revolution; involved in regulatory function
Sop2p-like	BA_RA3B_29	Y08999	6 WD40; interacts with Arp3p and modulates profilin function
Cell cycle			
Cyclin H	BA_RA4D_55	AF154914	Associates with cyclin-dependent kinase 7 and Mat1 to form the mammalian Cdk-activating kinase, CAK; suppressed by p53
Clk4	BA_RA1C_80	NM_007714	CDK-like kinase 4; phosphorylate spliceosome
PCM-1	BA_RA5A_71	L27841	6 coils; a 228-kd centrosome autoantigen with a distinct cell cycle distribution; from fetal liver
PACT	BA_RA5B_02	U28789	p53 and pRB binding protein; binding to p53 and inhibits p53 DNA binding
Tousled-like kinase	BA_RA4C_16	NM_011903	S/T kinase; 3 coils, NLS; cell cycle–dependent activity with highest activity during S phase and is regulated by phosphorylation
Miscellaneous			
p13MTCP1	BA_RA4D_10	U32332	Found in t(X;14) translocations
Meis 1	BA_RA6D_67	NM_010789	Myeloid ecotropic viral integration site 1; a PBX1-related homeobox gene involved in myeloid leukemia

Table 1. A partial list of genes analyzed after negative selection of the cDNA library generated from $Lin^{-/lo}Sca\text{-}1^{+}Thy\text{-}1.1^{lo}c\text{-}kit^{+}$ cells (continued)

Name	Clone code	Accession number	Description
RA-inducible protein E3	BA_RA6D_70	U29539	Present in the myeloid, B-lymphoid, and erythroid lineages, absent in nonhematopoietic cells
DJ-1	BA_RA2A_33	AB015652	A novel oncogene that transforms mouse NIH3T3 cells in cooperation with ras
HMG-1–related protein	BA_RA3D_37	S50213	DNA-binding components of the V-(D)-J recombinase
LSP1	BA_RA6D_01	S74179	Lymphocyte-specific protein; binds F-actin; a negative regulator of neutrophil chemotaxis
Mea	BA_RA2A_25	M27938	Male-enhanced antigen; involved in spermatogenesis

In clone codes, RA refers to the clones that have been negatively selected by subtraction with mature hematopoietic cells. Note that Bmi-1 and Notch 1 were identified during the sequencing of the random clones. ZF indicates zinc finger; TM, transmembrane; KH, K homology RNA-binding domain; HMG, high-mobility group; TCR, T-cell receptor; RRM, RNA recognition motif.

The EST (expressed sequence tag) sequences that are predominantly expressed in hematopoietic stem cells are available at the web-accessible database at www.systemsbiology.org/research/stemcells. This EST database also applies to Tables 2 and 3.

negative selection (see "Materials and methods") and the remaining clones were selected and spotted onto microarray type glass slides. These microarrays were used to analyze the gene expression profiles of HSCs and MPP cells by hybridizing with fluorescent-labeled (Cy3 and Cy5) cDNA probes derived from the sorted cell populations of HSCs and MPP cells. This allowed the identification of genes preferentially expressed in the HSCs and MPP cells, respectively.

To generate the macroarrays of cDNA libraries A and B, 36 000 clones from library A (made from $Lin^{-/lo}Sca\text{-}1^{+}$ cells) and 18 000 of 7.5×10^6 cDNA clones from library B (made from $Lin^{-/lo}$ $c\text{-}kit^{+}Thy\text{-}1^{lo}Sca\text{-}1^{+}$ cells) were arrayed onto nylon membranes (Figure 1), and hybridized with ^{33}P-labeled cDNA probes derived from the hematopoietic lineage-positive populations of cells. The image data were analyzed using Spot-Finding software,[16] which converted the image intensity of each spot into numerical data. After analyses, 3000 cDNA clones from macroarray A and 2000 clones from macroarray B that were negative or weakly positive were chosen to further construct the microarray. We have analyzed these negatively selected cDNA clones. A sequence comparison of 384 randomly chosen clones after the negative selection indicated that: (1) the percentage of housekeeping gene sequences represented by sequences encoding for mitochondrial and ribosomal proteins was greatly reduced from 25% to 5%; (2) the percentage of unique genes was increased to 60%; and (3) the percentage of novel sequences, which have no significant homologies with known sequences, was increased to 10% at the time the present manuscript was submitted, albeit the percentage decreased with the rapid progress of human and mouse genomic projects (data not shown).

The sequence analysis of the 2000 negatively selected clones from library B provides a glimpse of the gene expression profile of adult bone marrow enriched with HSCs. A partial annotated list of genes identified in cDNA library B is shown in Table 1. Numerous gene encoding transcription factors, chromatin modifiers, transmembrane proteins, and signaling molecules were identified. Based on their molecular and cellular roles in other tissues, contexts, or species, these genes might be likely candidates for playing an essential role in stem cell functions. For example, the *bmi-1* gene is a Polycomb-group gene that represses expression of *Hox* genes during development, and regulates cell proliferation and senescence by down-regulating the *ink4a* locus.[21] The *ink4a* locus expresses 2 genes by alternative splicing: p16 is a G1 cyclin inhibitor and $p19^{ARF}$ stabilizes expression of p53 by binding to Mdm2. Constitutive expression of both *bmi-1* and telomerase results in immortalization of epithelial cells.[22] Telomerase is activated in most immortal cell lines and cancers, and possibly is involved in the regulation of HSC self-renewal.[23] The Notch family has been implicated in the determination of HSC fate.[24-26] Recently, constitutive expression of the activated form of Notch1 was shown to be able to immortalize pluripotent, cytokine-dependent HSCs.[27]

HSCs express genes that are shared by other tissues such as brain, muscle, and germ line cells (Table 1). For example, male-enhanced antigen (Mea) specifically expressed in the testis is involved in spermatogenesis.[28] Some of the genes we identified are specific to hematopoietic tissues; for example, RGS18 is expressed in LT-HSCs and ST-HSCs as well as monocytes.[15] Among unknown genes, there are genes that code for a variety of important functional domains and several genes with no obvious homology to any known functional domains. Some of the other unknown genes were analyzed by Northern hybridization. Of 10 unknown clones tested, 2 were expressed in all tissues (no. 1), 2 were expressed in some tissues (no. 26), and 6 had no detectable signals in any of the tissues examined (no. 8, no. 11). These 6 clones are candidates for HSC-specific genes (Figure 1B).

Previously, Philips and his coworkers described the gene expression profile of $Sca\text{-}1^{+}AA4.1^{+}c\text{-}kit^{+}Lin^{-/lo}$ cells isolated from fetal liver HSCs.[29] There are a number of genes commonly expressed in both the adult bone marrow HSCs ($Lin^{-/lo}c\text{-}kit^{+}$ $Thy\text{-}1^{lo}Sca\text{-}1^{hi}$) generated by this study and the fetal liver HSCs ($Sca\text{-}1^{+}AA4.1^{+}c\text{-}kit^{+}Lin^{-/lo}$) generated by Philips et al,[29] such as Evi-1, TSA-1, Ramp1, RGS2, ZF216, macroH2A1.2, CD34, Notch1, vascular endothelial ZF1, and calmodulin. (Tables 1-3). In comparison to gene expression profiles of neural progenitor cells,[30] a number of genes were also found to be expressed in both neural progenitor and hemaptopoietic stem cells. These included histone H2A, interferon-induced factors, guanine triphosphate (GTP)–nucleotide binding protein, nuclear poly(a)-binding protein, GCN5 histone acetyltransferase, and so on, indicating a shared, at least partially, genetic program among HSCs and neural progenitor cells.

Differential gene expression between HSCs and MPP cells

Since Rh staining allows for separation of HSCs into populations enriched for ST-HSC/early progenitor cells and LT-HSCs,[9-11] the $Lin^{-/lo}Sca\text{-}1^{+}c\text{-}kit^{+}$ cells were further separated into Rh^{lo} and Rh^{hi} populations using flow cytometry (Figure 2A-B). In order to avoid a potential interference from the intermediate staining of Rh (Rh^{itm}) cells, we chose a symmetrical portion of cells containing either the 15% highest staining or 15% lowest staining for Rh. A competitive repopulation assay was performed to confirm the functional differences between these 2 populations of cells. As Figure 2C shows, the engraftment rate using Rh^{lo} cells is much higher than that using Rh^{hi} cells after transplantation. In addition, the Rh^{lo} cells could support hematopoiesis for up to 6 months after transplantation (Figure 2C) and were able to reconstitute the bone marrow in the secondary transplantation (F. L. and L. L., data not shown). The Rh^{hi} cells, which gave rise to both myeloid and lymphoid lineages,

Table 2. List of differentially expressed genes between $Rh^{lo}Lin^{-/lo}Sca\text{-}1^{+}c\text{-}kit^{+}$ and $Rh^{hi}Lin^{-/lo}Sca\text{-}1^{+}c\text{-}kit^{+}$ HSCs

Name	Clone code	Accession number	Background	Rh^{lo}	Rh^{hi}	Library
Transcription factor				Fold change	Fold change	
Nfix	BA-RA6D21	NM_010906	Nuclear factor I family; role in development	3.5		B
Hoxa9	BA-RA1C45	M28449	Homologous to the Drosophila Hoxa9		2.4	B
Zfp207	BA-RA1D05	NM_011751	Zinc finger protein Zep; 2 C2H2 ZF, NLS	2.1		B
Enx-1	BA-RA1C35	NM_007971	Polycomb family, regulate Hox gene expression		2.5	B
Kinase						
PI3 kinase	BA-RA5A82	BG354681	Phosphoinositide 3-kinase	2		B
PKA-like	BA-RA4C80	BG354682	cAMP-dependent protein kinase	3.1		B
Phosphatase						
Tyr phosphatase	BA-RA1D25	X63440	p19 protein-tyrosine-phosphatase		2.3	B
RNA binding						
RBM3	MSC-RA1A25	NM_016809	RNA binding motif 3	4.1		A
Napor-3	BA-RA6B66	AF090697	Apoptosis-related RNA-binding protein	2.7		B
Poly(a) binding	BA-RA6B35	X65553	RNA-binding protein	2.3		B
Protease						
Calpain 4	MSC-RA17B72	NM_009795	Calcium-dependent protease	6.2		A
Cell growth						
Activin beta C	MSC-RA15B58	X82540	TGF-beta family		4.3	A
TAX responsive	MSC-RA15B86	X81987	Malignancy-related protein 107		3.5	A
Inhibitor of CDC42	MSC-RA16A75	L07918	Hematopoietic-specific, GDP dissociation inhibitor		21	A
Protein modification						
CCT eta	MSC-RA12A61	Z31399	Chaperonin containing TCP-1	4.7		A
NEDD-4	MSC-RA16D13	P46935	Containing WW domain	2.1		A
Membrane associated						
pRGR1	BA-RA6C77	AF041429	4 TM; unknown function		3.9	B
TM-protein	BA-RA4C93	AF082835	Type III transmembrane protein		3.7	B
Ly-6C.2	MSC-RA11C15	M18466	u-PAR domain proteins	6.1		A
Nuclear protein						
PWP1-like	BA-RA5C56	L07758	WD-40, nuclear phosphoprotein		2.3	B
Npm1	BA-RA4A70	NM_008722	Nucleophosmin 1–related protein		3	B
Karyopherin	BA-RA1C82	NM_010655	Importin		5.9	B
Calcium binding						
PW29	BA-RA4C42	D49429	Calcium-binding protein		3.9	B
Cam I	BA-RA6C79	M19381	Calmodulin-dependent kinase and phosphatase		5.9	B
Chromosome modeling						
HAT1	BA-RA6D08	NM_003642	Histone acetyltransferase 1	2		A/B
Histone H2A	BA-RA5A12	AF171080	Chromatin structure	3.1		A/B
Set beta isoform	MSC-RA15B10	S68987	Inhibitor of HAT and PP2a		3.7	A
HMG1-related	MSC-RA17A09	U00431	Chromatin organization	2		A/B
Miscellaneous						
Ant2	BA-RA2D75	AF240003	Adenine nucleotide translocase 2		4.1	B
Prg2	MSC-RA14C84	NM_008920	Major basic protein in eosinophils of bone marrow		4.9	A
Calponin 2	MSC-RA16D77	NM_007725	Actin binding protein		6.5	A
Transport-like proteins	BA-RA1D09	AY008297	RPGR-interacting protein	2		A/B
CYLN2	BA-RA4A42	AAF99333	Cytoplasmic linker proteins	2		B
Novel						
KIAA0780	BA-RA1D26	BG354685	Unknown function		2.4	B
RP42	BA-RA2C82	BG354683	Unknown function	2		B
PTD011	BA-RA5D01	BG354684	Unknown function	2.9		B
Unknown	BA-RA3D39	BG354687	Unknown function		2.1	B
Unknown	MSC-RA16B18	BG354690	Unknown function	5.1		A
Unknown	BA-RA4A52	BG354679	Unknown function		2.3	B
Unknown	BA-RA4B58	BG354678	Unknown function	2		B
Unknown	MSC-RA15C76	BG354689	Unknown function	2		A

Libraries A and B were made from $Lin^{-/lo}Sca\text{-}1^{+}$ and $Lin^{-/lo}Sca\text{-}1^{+}Thy\text{-}1.1^{lo}c\text{-}kit^{+}$ cells, respectively. The changes in expression levels (fold) between Rh^{lo} and Rh^{hi} cells are indicated in the corresponding column of Rh^{lo} and Rh^{hi}. For abbreviations see Table 1.

only engrafted the bone marrow for less than 4 weeks (Figure 2C). This result demonstrated that we had obtained 2 distinct cell populations: the Rh^{lo} population of cells were enriched for HSCs and the Rh^{hi} population of cells were enriched for MPP cells.[5]

To study differential gene expression among Rh^{lo} and Rh^{hi} cells, we used the microarray glass slides on which the 5000 negatively selected cDNA clones from both libraries A and B had been arrayed in duplicate. These microarrays were hybridized to Cy3- and Cy5-labeled probes generated from Rh^{lo} and Rh^{hi} cells, respectively. Figure 3A shows a typical hybridization result showing differentially expressed genes between the Rh^{lo} and the Rh^{hi} cells. Rh^{lo}-Cy3 (green) represents genes that are expressed higher in the

Table 3. A partial list of genes commonly expressed in both $Rh^{lo}Lin^{-/lo}Sca\text{-}1^{+}c\text{-}kit^{+}$ and $Rh^{hi}Lin^{-/lo}Sca\text{-}1^{+}c\text{-}kit^{+}$ cells

Name	Clone code	Accession number	Description
Transcription factor			
GCN5	BA_RA5B01	AF254441	Mouse histone acetyltransferases
Vezf1	BA_RA1B23	NM_016686	Vascular endothelial zinc finger, high homology with DB1
Ctcf	BA_RA1D45	NM_007794	CCCTC-binding factor, zinc-finger protein
Pit-1	BA_RA2A02	S77417	Pituitary-specific POU-domain transcription factor
Zfp148	BA_RA2B29	NM_011749	Zinc finger protein 148
LIM only 2 (Lmo2)	BA_RA2B65	NM_008505	Hemotopoietic-related transcription factors
MEF2C	BA_RA2C38	NM_002397	MADS/MEF2-family transcription factor
BRX	BA_RA1A21	Y11896	Breast cancer nuclear receptor-binding auxiliary protein
Scmh1	BA_RA2A80	AB03090	Sex comb on midleg homologue protein
Fiz1	BA_RA1D58	NM_011813	Flt3 interacting zinc finger protein 1
Ciz1	BA_RA3D83	NM_012127	p21 (Cip1/Waf1)-interacting zinc finger protein
p53	MSC_RA14D53	AF161020	Cell cycle regulator
TOK-1	BA_RA1D59	XM_005841	Cdk inhibitor; p21 binding protein
Kinase			
p38	BA_RA5D64	NM_011951	MAP kinase
Phosphatase			
P19	BA_RA1D25	NM_009878	Nonreceptor protein tyrosine phosphatase (PTP)
Nucleic acid binding protein			
Translin	BA_RA5C46	X81464	RNA, DNA-binding protein
Cnbp	BA_RA6C69	NM_013493	Cellular nucleic acid binding protein
Srp20	BA_RA1C83	X91656	RNA-binding protein
Nabp	MSC_RA12A01	L12693	Nucleic acid binding protein
Cell growth			
P40	BA_RA6C60	NM_011029	A subunit of IL-12
Cyclophilin	MSC_RA13B90	X52803	Secreted growth factor, cyclosporin A–binding protein
Stathmin	BA_RA1C77	J04797	Also called p19, cytosolic phosphoprotein, developmentally regulated
Ptma	MSC_RA14D10	NM_008972	Prothymosin alpha, proliferation-related protein
Protein modification			
Cct8	BA_RA6B12	NM_009840	Chaperonin subunit 8 (theta)
Skp1	BA_RA2C21	AF083214	Skp1/Cull 1/F-box protein (SCF) complex protein Skp1
RC7-l	MSC_RA12B73	D21799	Beta-type subgroup of proteasomes
Adhesion			
Vinexin	BA_RA5C01	AF064806	Adhesion associated protein, a vinculin-binding protein with multiple SH3 domains
E25B	BA_RA5C53	U76253	Integral membrane protein 2B
Membrane associated			
Ramp 2	BA_RA2B94	NM_019444	Receptor-activity modifying protein 2
Ramp 1	BA_RA1C19	NM_016894	Receptor-activity modifying protein 1
TSA-1	MSC_RA12B43	U47737	Thymic shared antigen-1
Vamp5	BA_RA2C71	NM_016872	Vesicle-associated membrane protein 5
Nuclear protein			
BRCA2	BA_RA6C10	U89652	Breast cancer susceptibility genes
RING finger protein	MSC_RA15D38	AF022081	Function as a bridging factor and regulate steroid receptor-dependent transcription
PTAC97	BA_RA3B79	D45836	Nuclear pore–targeting complex component of 97kDa
Apoptotic-related protein			
Survivin	BA_RA5C63	AF115517	Inhibitor of apoptosis
PCAR	BA_RA6B65	AF174394	Apoptotic-related protein
Naip7	BA_RA6C21	AF242433	Neuronal apoptosis inhibitory protein
Histone related			
Nap1l4	BA_RA6C49	NM_008672	A histone chaperonin
H3f3b	BA_RA6C53	NM_008211	H3b histone
Centrosome associated			
IB3/5	BA_RA5B13	X79131	Polypeptide attached to centrosomes
Miscellaneous			
MNCb-3848	BA_RA6B62	AB041599	Brain cDNA, clone MNCb-3848
Gdi3	BA_RA1A63	NM_008112	GDP dissociation inhibitor 3 (Gdi3)
Septin6	BA_RA1D47	AB023622	A novel member of the structural GTPase subfamily associated with cytokinesis, KIAA0128 is the human counterpart of mouse Septin6; 1 coil
Mtprd	BA_RA2A46	AB008516	Developmentally regulated gene, involved in Down syndrome
Zinc finger protein	BA_RA2D28	D10627	Zinc finger protein
Unr	BA_RA4B25	U40223	Rat unr mRNA for unr protein with unknown function
SNX5	MSC_RA11A92	NM_014426	Sorting nexin 5 (SNX5)
Transgelin	MSC_RA13A39	AF149291	1 calponin homology domain; in thymus
ATPase	MSC_RA15D42	NM_003945	Lysosomal proton pump
BAF57	MSC_RA18A78	AF035263	Mammalian SWI/SNF subunits, critical for chromatin remodeling

Table 3. A partial list of genes commonly expressed in both $Rh^{lo}Lin^{-/lo}Sca\text{-}1^{+}c\text{-}kit^{+}$ and $Rh^{hi}Lin^{-/lo}Sca\text{-}1^{+}c\text{-}kit^{+}$ cells (continued)

Name	Clone code	Accession number	Description
Unknown			
RP11	BA_RA5D56	AC022766	Unknown function
4F5rel	BA_RA3B25	AF073297	A candidate phenotypic modifier for SMA
EST	MSC_RA16B50	BG354688	Unknown function
Interferon induced	BA_RA5D20	X61381	R. rattus interferon–induced mRNA

enriched HSCs. Rh^{hi}-Cy5 (red) represents genes that are expressed higher in the enriched MPP cells. Yellow represents genes that are expressed in a relatively similar level in both cell populations. Figure 3B shows the results of statistical analysis based on 2 experiments with 4 points (2 points on each slide) representing each cDNA clone, and selection using SpotUnite and SpotSelection software (H. Hammersmark and R.B., see "Materials and methods"). The vertical axis represents the Cy3/Cy5 signal intensity ratio and the horizontal axis was sorted by mean ± standard deviation (sd) (mean indicates average signal intensity of 4 spots for each cDNA insert). We chose cutoff lines that were either 2-fold higher or 2-fold lower for the Cy3/Cy5 intensity ratio. Thus, genes highly expressed in the Rh^{lo} population fell into the upper (positive) region and genes highly expressed in the Rh^{hi} population fell into the lower (negative) region. The spots that fell very close to the left axis represent cDNAs with very weak hybridization signals and, along with spots of significant variation, were ignored. Using this strategy we have identified approximately 30 genes that were highly expressed in the LT-HSCs and around 30 genes that were highly expressed in the MPP cells. These clones were marked with error bars (yellow, Figure 3B). Of these, some were not found in the nonredundant database and therefore were defined as novel or functionally unknown genes (Table 2). To confirm the differential expression of selected genes using microarray analysis, RT-PCR was performed and 15% of the clones were confirmed to be expressed preferentially in HSCs. Activin βC was the only clone from the Rh^{hi} cell group tested that showed predominant expression in MPP cells (Figure 3C).

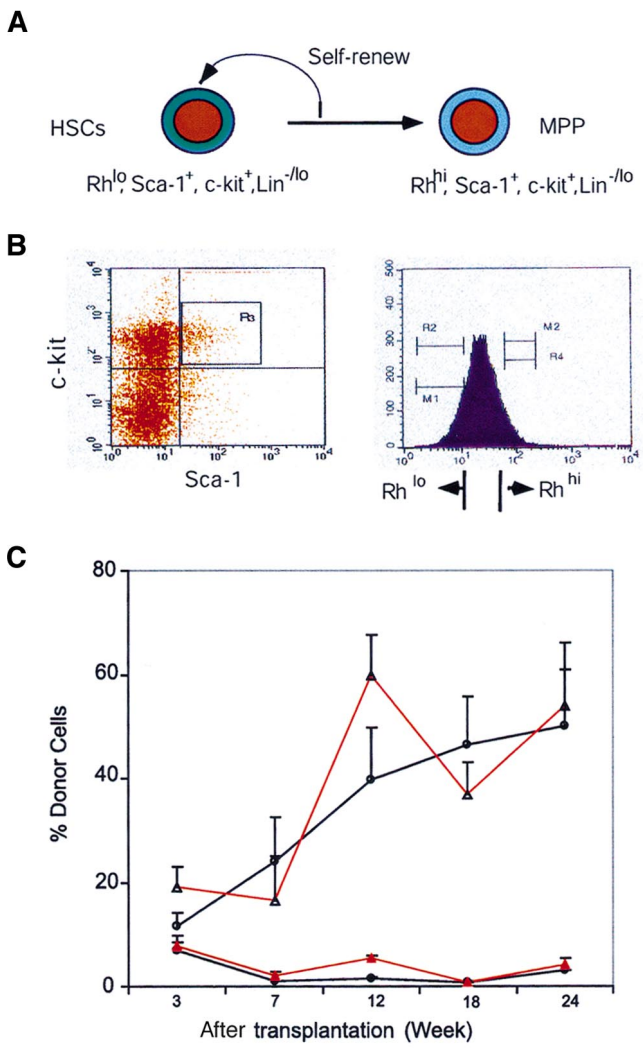

Figure 2. Isolation and functional characterization of Rh^{lo} and Rh^{hi} cell populations. (A) A cartoon showing 2 important populations of cells: LT-HSCs and multipotent progenitors (MPPs). The LT-HSCs can support long-term (up to 6 months) hematopoiesis in competitive repopulation assay. MPP cells can only briefly support (< 4 weeks) hematopoiesis in the same assay. Rhodamine-123 (Rh) was used to separate these 2 populations of cells. (B) Separation of $Rh^{lo}Lin^{-/lo}Sca\text{-}1^{+}c\text{-}kit^{+}$ and $Rh^{hi}Lin^{-/lo}Sca\text{-}1^{+}c\text{-}kit^{+}$ cells using flow cytometry. R3: $Lin^{-/lo}Sca\text{-}1^{+}c\text{-}kit^{+}$ cells; R2: Rh^{lo} (15% of cells with the lowest staining); R4: Rh^{hi} (15% of cells with the highest staining). (C) Competitive repopulation assay. Either $Rh^{lo}Lin^{-/lo}Sca\text{-}1^{+}c\text{-}kit^{+}$ or $Rh^{hi}Lin^{-/lo}Sca\text{-}1^{+}c\text{-}kit^{+}$ cells (250 cells) derived from the donor Ly5.2 mice were mixed with 4×10^4 $Lin^{-/lo}$ cells from congenic host strain Ly5.1,[17] and injected to lethally irradiated (11 Gy [1100 rads]) Ly5.1 mice. Eight mice were used per group. Peripheral blood cells were harvested at different time points after transplantation as indicated. Flow cytometric analyses were performed using different specific cell surface markers (CD3, B220, Mac-1, and Gr-1) together with anti-Ly5.2 antibody. ○ indicates Rh^{lo} lymphoid lineage; △, Rh^{lo} myeloid lineage; ●, Rh^{hi} lymphoid lineage; ▲, Rh^{hi} myeloid lineage.

The differentially expressed genes fall into a variety of gene families (Table 2). Among the genes that are highly expressed in HSCs are several clones identified as transcription factors. Nfix belongs to the conserved nuclear factor I (NFI) family of transcription/replication proteins. Loss of function of Nfia, one of 4 Nfix family members, causes severe developmental defects, suggesting this family may have distinct roles in vertebrate development.[31] Two novel genes that encode a PI3 kinase and a PKA kinase were also identified. The homologue of PKA-like kinase in *Schizosaccharomyces pombe*, Kin, was found to be important for growth polarity.[32] PI3 kinase can be activated by c-kit/stem cell factor receptor signaling and is essential for male fertility.[33] Furthermore, PI3 kinase has been known to be essential for cell survival.[34] RBM3, an RNA-binding protein, is up-regulated by granulocyte macrophage–colony-stimulating factor (GM-CSF)[35] and is closely related to the Y chromosome ribonucleic acid recognition motif (YRRM).[36] YRRM has been implicated in azoospermia, a male infertility disease, and spermatogenesis.[36] This suggests that RBM3 might be involved in the regulation of HSC development. Histone acetyltransferase 1 (HAT1) is also highly expressed in the Rh^{lo} cells. HAT1 belongs to the GCN5-related N-acetyltansferase superfamily and functions to acetylate histone H4, thereby participating in chromatin-related transcriptional regulation.[37] Recently HAT1 was found to be involved with telomeric silencing by the transcriptional repression of telomere-proximal genes.[38] In contrast, Set oncoprotein, a component of a recently identified protein complex, INHAT, that inhibits the activity of HAT,[39] was highly expressed in the Rh^{hi} cell population.

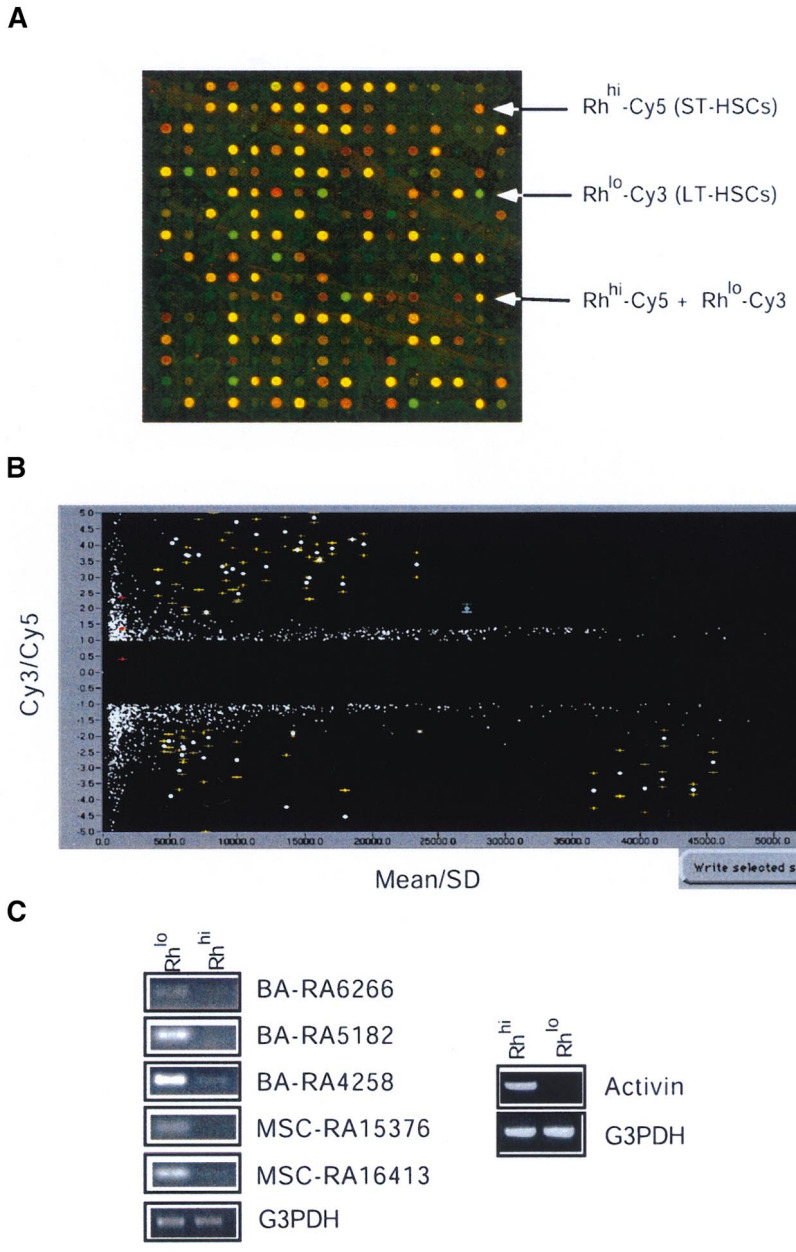

Figure 3. Microarray analysis of HSC gene expression. (A) Microarray hybridization and spotfinding. Microarray slides described in Figure 1 were hybridized with cDNA probes prepared from cells. Each cDNA probe was labeled with Cy3 and Cy5 fluorophores separately. A combination of these cDNA probes was hybridized with microarrays A (3000 × 2) and B (2000 × 2). A typical hybridization result was shown using the Cy3-labeled cDNA probe from $Rh^{lo}Lin^{-/lo}Sca\text{-}1^{+}c\text{-}kit^{+}$ and the Cy5-labeled probe from $Rh^{hi}Lin^{-/lo}Sca\text{-}1^{+}c\text{-}kit^{+}$ cells. (B) Statistical analysis and selection. A typical example showing the microarray hybridization results obtained from 2 experiments scanned by Genescanner (Molecular Dynamics) and analyzed by SpotUnite and SpotSelection programs (see text for details). (C) RT-PCR assays. To confirm the differential expression of genes between the Rh^{lo} and Rh^{hi} cells, 15% candidate genes were analyzed by RT-PCR as described in "Materials and methods."

Among the genes that are highly expressed in Rh^{hi} cells is the growth factor Activin βC. Activin βC belongs to the transforming growth factor β (TGF-β) family, which functions in the determination of cell fate and pattern formation during embryogenesis.[40] However, the function of Activin βC remains unknown. Other members of the Activin family (A and B) are known to function as dimeric proteins with diverse biologic activities in vertebrate reproduction.[41] Expression of Activin βC in the MPP cells suggests that it may be involved in the regulation of HSC differentiation and proliferation. Another gene differentially expressed in the MPP population is *Enx-1*, a mouse homologue of *Drosophila enhancer of zeste* and member of the Polycomb family of transcriptional regulators.[42] It controls the expression of several homeobox genes. The *HoxA9* gene was also identified in the Rh^{hi} cells. HoxA9 was found to be involved in translocation t(7;11)(p15;p15), present in acute myeloid leukemia (AML) French-American-British (FAB) classification M2-M4.[43] Inactivation of *HoxA9* was shown to have defects in myeloid, erythroid, and lymphoid development in knockout mice.[44] Overexpression of *HoxA9* and Meis1 has been shown to induce murine leukemia[45] and overexpression of *HoxA9* alone can immortalize myeloid progenitors.[46]

We also identified approximately 70 genes that are expressed in both LT-HSCs and MPP cells (Table 3). Among these is the transcription factor GCN5, a histone acetyltransferase, which interacts with Notch 1 and plays an important role in the recombination recognition sequence binding protein J (RBP-J)–mediated transactivation by Notch 1.[47] Loss of GCN5 leads to increased apoptosis and mesodermal defects during mouse development.[48] Fiz1, a zinc finger protein with 11 C2H2-type zinc fingers, interacts with receptor tyrosine kinase Flt3.[49] Flt3 has been shown to play a role in proliferation and survival of hematopoietic progenitor cells as well as differentiation of early B-lymphoid progenitors, dendritic cells, and natural killer cells.[50] TOK-1, a POU-domain–containing transcription factor, is a p21 C-terminal–binding protein and preferentially binds to an active form of cyclin-dependent kinase 2 (CDK2) via p21.[51] p21 is a negative

cell-cycle regulator and is involved in maintaining the quiescent state of HSCs.[52] One of the identified apoptosis-related molecules is survivin. Survivin is an inhibitor of apoptosis and is overexpressed in various human cancers but undetectable in normal tissues.[53] Overexpression of survivin resulted in an accelerated S phase shift, resistance to G1 arrest, and activated Cdk2/Cyclin E complex leading to Rb phosphorylation.[54]

Discussion

The mechanisms that regulate the self-renewal of HSCs are largely unknown. The deterministic model of hematopoiesis suggests that HSC self-renewal is autonomous; thus, expression of specific sets of genes could determine the stem cell fate. Therefore, a systematic approach was used to study differential gene expression profiles in adult HSCs. The ability to isolate a cell population enriched with HSCs allowed us to study the gene expression profile of HSCs. Using macroarray and microarray techniques, we were able to obtain a snapshot of genes expressed by enriched populations of HSCs as well as MPP cells. The identification of multiple genes that regulate proliferation, cell survival, immortalization, and differentiation gives new insights into HSC functions. This is important not only to study their biologic behaviors but also for potential clinical applications. For example, the efficacy of treatment modalities for cancer, such as radiation therapy and many chemotherapy agents, is constrained by dose-limiting bone marrow toxicity. Therefore, the ability to isolate HSCs with self-renewal and repopulating potentials, in vitro amplification of these cells, and transplantation without tumor cells would greatly increase the success rate of current cancer therapies.

Ex vivo–expanded stem cells have been extensively tested for possible use in transplantation. Soluble jagged-1, a Notch ligand, induced the survival and expansion of human stem cells with multipotent repopulating capacity.[55] The presence of Notch 1 in the HSC library suggests that these proteins may play a role in self-renewal or maintaining identity of HSCs. Although various growth factors can induce stem cells to proliferate in vitro, they result in concomitant differentiation. The function of stem cell factor (SCF) has been known as a hematopoietic stem cell survival factor.[56] Recently, it has also been shown that the in vitro culture of the HSCs, in the presence of both SCF and thrombopoietin (TPO), induced self-renewal cell division in which only one of the daughter cells had self-renewal potential.[57] Asymmetric cell division requires unequal segregation of cell-fate determinants, such as mRNAs and proteins, which are important to maintain self-renewal and repopulating potentials, during mitosis. This type of asymmetric cell division is also observed in germline cell division. During *Drosophila* oogenesis, a germline stem cell divides asymmetrically to produce a daughter stem cell and a cytoblast, which further divides to eventually produce oocytes. This process is controlled by several factors including decapentaplegic (DPP)[58] as well as by intrinsic mechanisms involving pumilio, nanos, arrest, and bag-of-marbles. Expression of BMP4, the mammalian homologue of DPP, in HSCs (Table 1) suggests that BMP4 may also play a similar role.

The differential gene expression pattern between Rh^{lo} and Rh^{hi} indicated that some of the genes were either HSC- or MPP-specific (Table 2), and many of them were expressed in both populations of cells (Table 3). Some of the MPP highly expressed genes included Activin βC, nuclear molecules (PWP-1, Npm1, and karyopherin), TAX-responsive factor, and inhibitor of CDC42. CDC42 was reported to play a role in controlling HSC shape, adhesion, migration, and mobilization.[59] Genes that are preferentially expressed in HSCs include transcription factors, RNA-binding proteins, chromatin modifiers, and protein kinases, many of which are involved in developmental regulation. The HSC preferentially expressed genes might be candidate genes that play a role in stem cell self-renewal.

In summary, cDNA libraries were generated from mouse HSCs with long-term reconstituting potential and from progenitor cells, and the differential gene expression patterns were studied using bioinformatics and array technologies. Although the microarray system has been known for its tendency to lose less abundant mRNAs due to limited resolution, our approach overcomes this tendency by removing cDNAs that show high hybridization signals using macroarray and thereby enriching for cDNAs derived from less-abundant mRNAs. On the other hand, our subtraction technique using cDNA probes derived from mature populations of cells may lead to a loss of molecules that are expressed in multiple developmental stages. The systematic approach used in this study provides an effective and efficient method for identification of genes that are specifically expressed in HSCs with self-renewal capability. This study will therefore provide a fundamental tool for identifying important candidate genes involved in the regulation of self-renewal and expansion of HSCs.

Acknowledgments

We thank G. van den Engh, K. Allen, and D. Corden for the assistance of cell sorting. We thank E. Hammersmark for the assistance on microarray analysis software. We appreciate V. Ng and X. Tao for their assistance on analysis of Est sequences. We thank A. Banta for the involvement of macroarray work. We are grateful to Drs G. Vassilopoulous, D. Russel, and N. Wolf for the consultation on the bone marrow transplantation and repopulation assay. We appreciate the help of R. Krumlauf, B. Steenhard, and D. Stenger on the manuscript proofreading. We are particularly grateful to X. He, J. Zhang, and C. Niu for their work on the reanalysis of cDNA sequences listed in the tables.

References

1. Morrison SJ, Uchida N, Weissman IL. The biology of hematopoietic stem cells. Annu Rev Cell Dev Biol. 1995;11:35-71.
2. Uchida N, Weissman IL. Searching for hematopoietic stem cells: evidence that Thy-1.1lo Lin- Sca-1+ cells are the only stem cells in C57BL/Ka-Thy-1.1 bone marrow. J Exp Med. 1992;175:175-184.
3. Spangrude GJ, Heimfeld S, Weissman IL. Purification and characterization of mouse hematopoietic stem cells. Science. 1988;241:58-62.
4. Lagasse E, Shizuru JA, Uchida N, Tsukamoto A, Weissman IL. Toward regenerative medicine. Cell. 2001;14:425-436.
5. Morrison SJ, Weissman IL. The long-term repopulating subset of hematopoietic stem cells is deterministic and isolatable by phenotype. Immunity. 1994;1:661-673.
6. Kondo M, Weissman IL, Akashi K. Identification of clonogenic common lymphoid progenitors in mouse bone marrow. Cell. 1997;91:661-672.
7. Pohlmann SJ, Slayton WB, Spangrude GJ. Stem cell populations: purification and behavior. In: Zon LI, ed. Hematopoiesis—A Developmental Approach. Oxford, United Kingdom: Oxford University Press; 2001: 35-47.
8. Weissman IL. Translating stem and progenitor cell biology to the clinic: barriers and opportunities. Science. 2000;287:1442-1446.
9. Spangrude GJ, Johnson GR. Resting and activated subsets of mouse multipotent hematopoietic stem cells. Proc Natl Acad Sci U S A. 1990; 87:7433-7437.
10. Wolf NS, Kone A, Priestley GV, Bartelmez SH. In vivo and in vitro characterization of long-term repopulating primitive hematopoietic cells isolated by sequential Hoechst 33342-rhodamine 123 FACS selection. Exp Hematol. 1993;21: 614-622.
11. Zijlmans JM, Visse JW, Kleiverda K, Kluin PM, Willemze R, Fibbe WE. Modification of rhodamine staining allows identification of hematopoietic stem cells with preferential short-term or

long-term bone marrow-repopulating ability. Proc Natl Acad Sci U S A. 1995;92:8901-8905.

12. Chaudhary PM, Roninson IB. Expression and activity of P-glycoprotein, a multidrug efflux pump, in human hematopoietic stem cells. Cell. 1991;66:85-94.
13. Cheshier SH, Morrison SJ, Liao X, Weissman IL. In vivo proliferation and cell cycle kinetics of long-term self-renewing hematopoietic stem cells. Proc Natl Acad Sci U S A. 1999;96:3120-3125.
14. Dexter TM, Moore MA, Sheridan AP. Maintenance of hemopoietic stem cells and production of differentiated progeny in allogeneic and semi-allogeneic bone marrow chimeras in vitro. J Exp Med. 1977;145:1612-1616.
15. Park IK, Klug CA, Li K, et al. Molecular cloning and characterization of a novel regulator of G-protein signaling from mouse hematopoietic stem cells. J Biol Chem. 2001;276:915-923.
16. Geiss GK, Bumgarner RE, An MC, et al. Large-scale monitoring of host cell gene expression during HIV-1 infection using cDNA microarrays. Virology. 2000;266:8-16.
17. Osawa M, Hanada K, Hamada H, Nakauchi H. Long-term lymphohematopoietic reconstitution by a single CD34-low/negative hematopoietic stem cell. Science. 1996;273:242-245.
18. Maier E, Meier-Ewert S, Ahmadi AR, Curtis J, Lehrach H. Application of robotic technology to automated sequence fingerprint analysis by oligonucleotide hybridisation. J Biotechnol. 1994;35:191-203.
19. Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science. 1995;270:467-470.
20. Spangrude GJ, Brooks DM, Tumas DB. Long-term repopulation of irradiated mice with limiting numbers of purified hematopoietic stem cells: in vivo expansion of stem cell phenotype but not function. Blood. 1995;85:1006-1016.
21. Jacobs JJ, Kieboom K, Marino S, DePinho RA, van Lohuizen M. The oncogene and Polycomb-group gene bmi-1 regulates cell proliferation and senescence through the ink4a locus. Nature. 1999;397:164-168.
22. Kiyono T, Foster SA, Koop JI, McDougall JK, Galloway DA, Klingelhutz AJ. Both Rb/p16INK4a inactivation and telomerase activity are required to immortalize human epithelial cells. Nature. 1998;396:84-88.
23. Morrison SJ, Prowse KR, Ho P, Weissman IL. Telomerase activity in hematopoietic cells is associated with self-renewal potential. Immunity. 1996;5:207-216.
24. Li L, Milner L, Deng Y, et al. Human homolog of rat Jagged 1 expressed by marrow stroma inhibits differentiation of 32D cells through interaction with Notch1. Immunity. 1998;8:43-55.
25. Jones P, May G, Healy L, et al. Stromal expression of Jagged 1 promotes colony formation by fetal hematopoietic progenitor cells. Blood. 1998; 92:1505-1511.
26. Varnum-Finney B, Purton LE, Yu M, et al. The Notch ligand, Jagged 1, influences the development of primitive hematopoietic precursor cells. Blood. 1998;91:4084-4091.
27. Varnum-Finney B, Xu L, Brashem-Stein C, et al. Pluripotent, cytokine-dependent, hematopoietic stem cells are immortalized by constitutive notch1 signaling. Nat Med. 2000;6:1278-1281.
28. Lau YF, Chan KM, Sparkes R. Male-enhanced antigen gene is phylogenetically conserved and expressed at late stages of spermatogenesis. Proc Natl Acad Sci U S A. 1989;86:8462-8466.
29. Phillips RL, Ernst RE, Brunk B, et al. The genetic program of hematopoietic stem cells. Science. 2000;288:1635-1640.
30. Geschwind DH, Ou J, Easterday MC, et al. A genetic analysis of neural progenitor differentiation. Neuron. 2001;29:325-339.
31. das Neves L, Duchala CS, Godinho F, et al. Disruption of the murine nuclear factor I-A gene (Nfia) results in perinatal lethality, hydrocephalus, and agenesis of the corpus callosum. Proc Natl Acad Sci U S A. 1999;96:11946-11951.
32. Levin DE, Bishop JM. A putative protein kinase gene (kin1+) is important for growth polarity in Schizosaccharomyces pombe. Proc Natl Acad Sci U S A. 1990;87:8272-8276.
33. Jiang G, Hyman R, Lee KF, Gorman SO, Blume-Jensen P, Hunter T. Kit/stem cell factor receptor-induced activation of phosphatidylinositol 3′-kinase is essential for male fertility. Nature Genet. 2000;24:157-162.
34. Blume-Jensen P, Janknecht R, Hunter T. The kit receptor promotes cell survival via activation of PI 3-kinase and subsequent Akt-mediated phosphorylation of Bad on Ser136. Curr Biol. 1998;8:779-782.
35. Baghdoyan S, Dubreuil P, Eberle F, Gomez S. Capture of cytokine-responsive genes (NACA and RBM3) using a gene trap approach. Blood. 2000;95:3750-3757.
36. Derry JM, Kerns JA, Francke U. RBM3, a novel human gene in Xp11.23 with a putative RNA-binding domain. Hum Mol Genet. 1995;4:2307-2311.
37. Dutnall RN, Tafrov ST, Sternglanz R, Ramakrishnan V. Structure of the histone acetyltransferase Hat1: a paradigm for the GCN5-related N-acetyltransferase superfamily. Cell. 1998;94:427-438.
38. Kelly TJ, Qin S, Gottschling DE, Parthun MR. Type B histone acetyltransferase Hat1p participates in telomeric silencing. Mol Cell Biol. 2000; 20:7051-7058.
39. Sang-beom S, McNamara P, Heo S, Turner A, Lane WS, Chakravarti D. Regulation of histone acetylation and transcription by INHAT, a human cellular complex containing the Set oncoprotein. Cell. 2001;104:119-130.
40. Padgett RW. TGFbeta signaling pathways and human diseases. Cancer Metastasis Rev. 1999; 18:247-259.
41. Ethier JF, Findlay JK. Roles of activin and its signal transduction mechanisms in reproductive tissues. Reproduction. 2001;121:667-675.
42. Hobert O, Sures I, Ciossek T, Fuchs M, Ullrich A. Isolation and developmental expression analysis of Enx-1, a novel mouse Polycomb group gene. Mech Dev. 1996;55:171-184.
43. Borrow J, Stanton VP, Andresen JM, et al. The t(7;11)(p15;p15) translocation in acute myeloid leukaemia fuses the genes for nucleoporin NUP98 and class I homeoprotein HOXA9. Nat Genet. 1996;12:159-167.
44. Lawrence HJ, Helgason CD, Sauvageau G, et al. Mice bearing a targeted interruption of the homeobox gene HOXA9 have defects in myeloid, erythroid, and lymphoid hematopoiesis. Blood. 1997;89:1922-1930.
45. Lawrence HJ, Rozenfeld S, Cruz C, et al. Frequent co-expression of the HOXA9 and MEIS1 homeobox genes in human myeloid leukemias. Leukemia. 1999;13:1993-1999.
46. Schnabel CA, Jacobs Y, Cleary ML. HoxA9-mediated immortalization of myeloid progenitors requires functional interactions with TALE cofactors Pbx and Meis. Oncogene. 2000;19:608-616.
47. Kurooka H, Honjo T. Functional interaction between the mouse notch1 intracellular region and histone acetyltransferases PCAF and GCN5. J Biol Chem. 2000;275:17211-17220.
48. Xu W, Edmondson DG, Evrard YA, Wakamiya M, Behringer RR, Roth SY. Loss of Gcn5l2 leads to increased apoptosis and mesodermal defects during mouse development. Nat Genet. 2000;26:229-232.
49. Wolf I, Rohrschneider LR. Fiz1, a novel zinc finger protein interacting with the receptor tyrosine kinase Flt3. J Biol Chem. 1999;274:21478-21484.
50. McKenna HJ, Stocking KL, Miller RE, et al. Mice lacking flt3 ligand have deficient hematopoiesis affecting hematopoietic progenitor cells, dendritic cells, and natural killer cells. Blood. 2000;95:3489-3497.
51. Ono T, Kitaura H, Ugai H, et al. TOK-1, a novel p21Cip1-binding protein that cooperatively enhances p21-dependent inhibitory activity toward CDK2 kinase. J Biol Chem. 2000;275:31145-31154.
52. Cheng T, Rodrigues N, Shen H, et al. Hematopoietic stem cell quiescence maintained by p21cip1/waf1. Science. 2000;287:1804-1808.
53. Adida C, Recher C, Raffoux E, Daniel MT, et al. Expression and prognostic significance of survivin in de novo acute myeloid leukaemia. Br J Haematol. 2000;111:196-203.
54. Li F, Ambrosini G, Chu EY, et al. Control of apoptosis and mitotic spindle checkpoint by survivin. Nature. 1998;396:580-584.
55. Karanu FN, Murdoch B, Gallacher L, et al. The notch ligand jagged-1 represents a novel growth factor of human hematopoietic stem cells. J Exp Med. 2000;192:1365-1372.
56. Brandt JE, Bhalla K, Hoffman R. Effects of interleukin-3 and c-kit ligand on the survival of various classes of human hematopoietic progenitor cells. Blood. 1994;83:1507-1514.
57. Ema H, Takano H, Sudo K, Nakauchi H. In vitro self-renewal division of hematopoietic stem cells. J Exp Med. 2000;192:1281-1288.
58. Xie T, Spradling AC. Decapentaplegic is essential for the maintenance and division of germline stem cells in the *Drosophila* ovary. Cell. 1998;94: 251-260.
59. Yang FC, Atkinson SJ, Gu Y, et al. Rac and Cdc42 GTPases control hematopoietic stem cell shape, adhesion, migration, and mobilization. Proc Natl Acad Sci U S A. 2001;98:5614-5618.