Unique Biological Properties and Application Potentials of CD34$^+$ CD38$^-$ Stem Cells From Various Sources

Tao-Yeuan Wang[1,2], Shing-Jyh Chang[3,4†], Margaret Dah-Tsyr Chang[4], Hsei-Wei Wang[5,6,7*]

[1]Department of Pathology, Mackay Memorial Hospital, [2]Mackay Medicine, Nursing and Management College, Taipei, [3]Department of Obstetrics and Gynecology, Mackay Memorial Hospital, Hsinchu, [4]National Tsing Hua University, Hsinchu, [5]Institute of Microbiology and Immunology, National Yang-Ming University, [6]VGH-YM Genome Center, National Yang-Ming University, and [7]Department of Education and Research, Taipei City Hospital, Taipei, Taiwan.

SUMMARY

Objective: Somatic CD34$^+$ CD38$^-$ stem cells can differentiate into cells of hematopoietic and endothelial lineages and have been clinically used to treat diseases. These stem cells can be obtained from cord blood (CB), bone marrow or granulocyte-macrophage colony-stimulating factor–mobilized peripheral blood. Unmasking genes differentially expressed in hematopoietic stem cells (HSCs) from different anatomic locations can improve our understanding of their basic biological features and help in clinical decision making when applying different HSCs.

Materials and Methods: We performed microarray analysis on human CD34$^+$ CD38$^-$ HSCs isolated from CB, bone marrow and peripheral blood. Systems biology and advanced bioinformatics tools were used to better understand the biological modules and genetic networks accompanying each HSC subtype.

Results: We identified HSC genes differentially expressed in various HSCs and found them to be involved in critical biological processes such as cell cycle regulation, cell motility, and endogenous antigen presentation. Among these three HSC types, HSCs from CB expressed the fewest rejection and immune response-associated genes, thereby showing the best potential as a transplantation source. Analysis of HSC-enriched genes using systems biology tools revealed a complex genetic network functioning in different CD34$^+$ CD38$^-$ cells, in which several genes act as hubs, such as MYC in CB HSCs and hepatic growth factor in bone marrow HSCs, to maintain the stability or connectivity of the whole network.

Conclusion: This study provides the foundation for a more detailed understanding of CD34$^+$ CD38$^-$ HSCs from different sources, and reveals the potentials of different HSCs for different clinical applications. [*Taiwan J Obstet Gynecol* 2009;48(4):356–369]

Key Words: CD34, hematopoietic stem cells, systems biology, transplantation

Introduction

Stem cells have attracted much attention recently because of their unique biological behaviors and attractive clinical usages. CD34$^+$ hematopoietic stem cells (HSCs), for example, have been used for clinical stem cell transplantation [1]. Clinically, HSCs can be obtained from three different anatomic locations: bone marrow (BM), peripheral blood (PB), and umbilical cord blood (CB). To improve our ability to utilize various HSCs, it is necessary to understand the differences between these stem cells and to decipher their underlying mechanisms.

In humans, most HSCs and precursors express CD34 antigen while being negative for the CD38 hematopoietic lineage marker [2,3]. The CD34$^+$ CD38$^-$ population can be divided into long-term and short-term functional HSCs with surface makers CD34$^+$ CD33$^-$ CD38$^-$ Rho^{low} and CD34$^+$ CD33$^-$ CD38$^-$ Rho^{high},

Correspondence to: Dr Hsei-Wei Wang, Institute of Microbiology and Immunology, National Yang-Ming University, 155, Section 2, Li-Nong Street, Taipei, Taiwan.
E-mail: hwwang@ym.edu.tw
Accepted: March 9, 2009
†*Co-first author*

respectively [4]. Recently, the CD133 antigen has also been used for HSC isolation [5,6]. The gene expression patterns of different human CD34+ HSCs isolated from various anatomic locations have been characterized by genomic analyses, which revealed genes involved in self-renewal, differentiation and lineage choice [2,7]. However, the assignment of biological significance to the filtered genes or even to each HSC type remains difficult. In this post-genomics era, modern bioinformatics and systems biology tools can help to predict biological behaviors or interactions of gene products. Deductions of *de novo* gene relationships by exploratory computational tools and systems biology algorithms can provide a systematic approach to discovering novel molecular events and relationships [8–10]. The combination of microarray and computational results enables biologists to speed up their research and to analyze data.

The aims of this study were to identify the genes responsible for stem cell properties in various CD34+ CD38− HSCs and to reveal the biological differences between them using systems biology approaches. We compared the gene expression profiles of HSCs from various anatomic locations. Our analysis revealed that CD34+ CD38− HSCs from different sources possess different biological properties and hence hold different application potentials.

Materials and Methods

CD34+ CD38− cells and primary microvascular endothelial cells (MVECs)

Human CD34+ CD38− cells were isolated from BM, PB or CB of healthy individuals (Poietics; Lonza Group Ltd., Basel, Switzerland). Human MVECs (Clonetics; Lonza Group Ltd.) were cultured in EGM-2 MV BulletKit medium (Lonza Group Ltd.).

RNA isolation and real-time polymerase chain reaction (PCR)

Total messenger RNA was extracted using an RNeasy mini kit (catalog no. 74106; Qiagen, Hilden, Germany) and 100 ng to 1 μg of total RNA was reverse transcribed using a First cDNA Synthesis Kit (catalog no. K1612; Fermentas, Glen Burnie, MD, USA). For quantitative real-time PCR analysis, the human pre-messenger RNA sequence was obtained from the National Center for Biotechnology Information AceView program (http://www.ncbi.nlm.nih.gov/AceView/). All primers were designed across introns in the Primer3 Web site (http://frodo.wi.mit.edu/primer3/) or Primer Express software (Applied Biosystems, Foster City, CA, USA). Thermodynamics and primer specificity analyses were performed using the Vector NTI suite (Invitrogen, Carlsbad, CA, USA) and the National Center for Biotechnology Information reverse e-PCR program (http://www.ncbi.nlm.nih.gov/sutils/e-pcr/reverse.cgi/). Real-time PCR reactions were performed using Maxima SYBR Green qPCR Master Mix (catalog no. K0222; Fermentas), and specific products were detected and analyzed using a StepOne sequence detector (Applied Biosystems). The expression level of each gene was normalized to the expression level of glyceraldehyde 3-phosphate dehydrogenase.

Array probe preparation, data analysis, and functional network analyses

Total RNA collection, complementary RNA probe preparation, array hybridization and data analysis were performed as previously described [10,11]. Expression profiles of CD34+ CD38− HSCs, MVECs and PB mononuclear cells were implemented using the Affymetrix HG-U133A chip (Affymetrix, Santa Clara, CA, USA). Gene annotation enrichment analysis was performed using the DAVID 2008 tool (http://david.abcc.ncifcrf.gov/) [12]. EASE score, a Fisher's exact test-based scoring system, was used to calculate *p* values to determine the probability of the number of genes being specifically associated (enriched) with a given gene ontology (GO) term or occurring by random chance. The Ingenuity Pathway Analysis (IPA) Web tool developed by the Ingenuity Co. (http://www.ingenuity.com/) was used to construct functional regulatory networks of gene profiles. IPA uses the Ingenuity Pathways Knowledge Base to identify known interactions between focus genes as well as other genes that are not present in the gene list. IPA then determines a statistical score for each network according to the fit of the network to the set of focus genes. The score is the negative log of *p* and denotes the likelihood of the focus genes in the network being found together by chance.

Results

Molecular signatures of various CD34+ CD38− HSCs

To access molecular signature genes for HSCs, we obtained human CD34+ CD38− HSCs from umbilical CB. Abundant expression of known "stemness" or precursor genes (including *GATA2*, *GATA1* and *RUNX1*) in CD34+ CD38− cells were verified by quantitative PCR (Figure 1A).

To compare the gene expression profiles and biological variations between HSCs from different anatomic

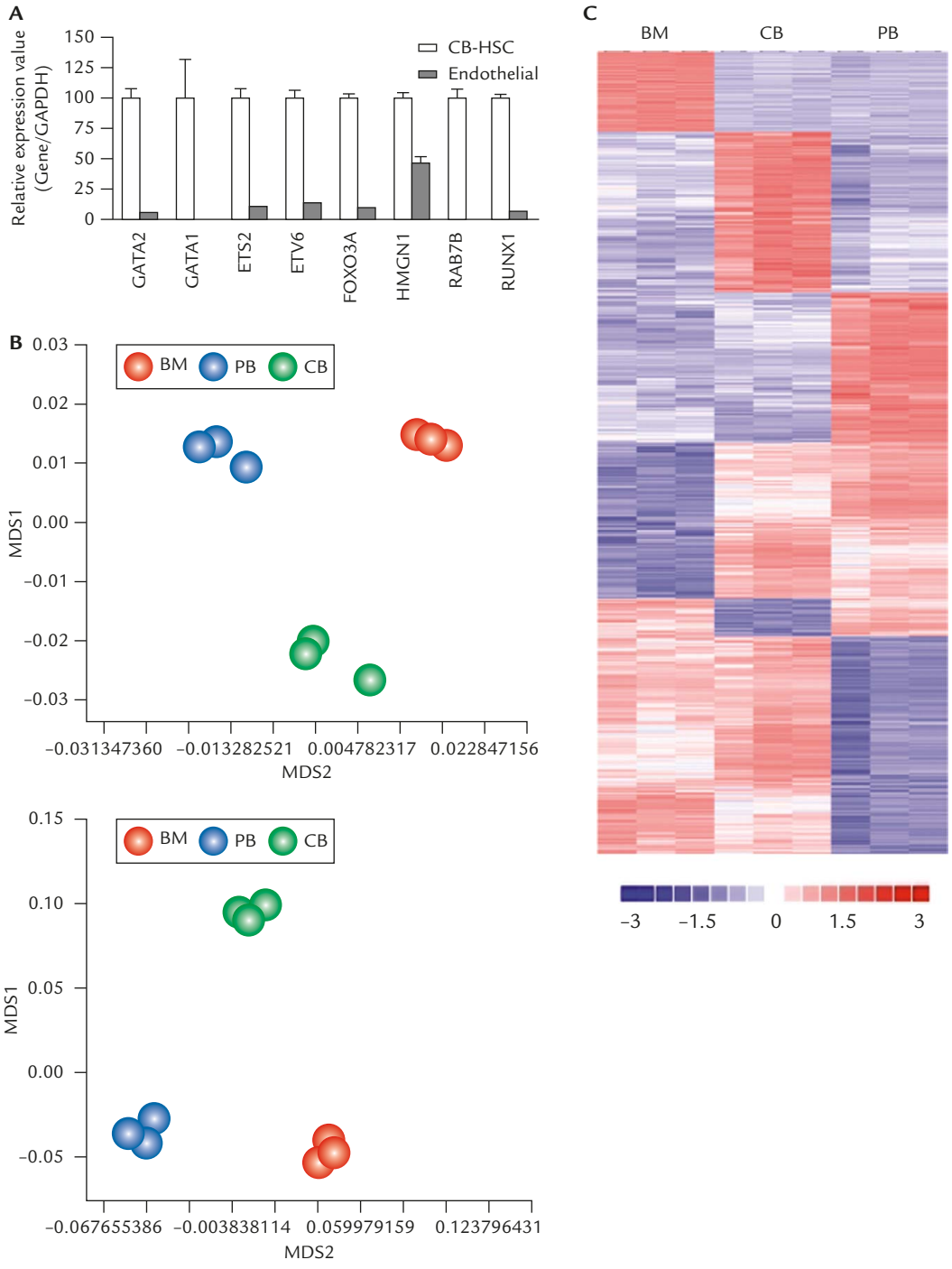

Figure 1. Transcriptome analysis of hematopoietic stem cells (HSCs) isolated from different anatomic locations. (A) Enriched expression of stemness and precursor genes in isolated HSCs. $CD34^+$ $CD38^-$ HSCs isolated by magnetic beads were subjected to total RNA isolation and quantitative polymerase chain reaction analysis. Primary microvascular endothelial cells were used as a mature progeny control. Mean expression levels of target genes were compared with that of glyceraldehyde-3-phosphate dehydrogenase (GAPDH) control. (B) Multidimensional scaling (MDS) plots using the whole transcriptome (upper) or filtered genes (lower) show the discrimination ability of the obtained molecular signature for cell groups. Each spot represents a single array sample. BM = bone marrow; CB = cord blood; PB = peripheral blood. (C) A heat map shows genes differentially expressed in different $CD34^+$ $CD38^-$ HSCs. Genes in red, increased expression; in blue, decreased.

locations, we also isolated $CD34^+$ $CD38^-$ HSCs from BM and granulocyte-macrophage colony-stimulating factor–mobilized PB from healthy individuals. The gene expression profiles of various HSCs were determined and compared. The gene expression profiles of these three cell types were implemented at least in triplicate using the Affymetrix HG-U133A chip. According to the statistical pipeline used [10,11,13], genes differentially

expressed between each cell type (the molecular signature) were identified. A total of 910 probe sets were uniquely upregulated in CB HSCs, 450 in BM HSCs, and 912 in PB HSCs (with a positive false discovery rate threshold, $q < 0.005$). In contrast, 220 probe sets were uniquely downregulated in CB HSCs, 851 in BM HSCs, and 1287 in PB HSCs ($q < 0.005$).The discrimination ability of these genes was assessed using multidimensional scaling (Figure 1B). A gene expression heat map for these genes indicated their unique expression in different $CD34^+$ $CD38^-$ HSCs (Figure 1C).

The top 50 genes most strongly expressed or repressed in $CD34^+$ $CD38^-$ CB HSCs are listed in Tables 1 and 2, respectively. The top 50 genes in BM HSCs and PB HSCs are listed in Tables 3 and 4, respectively. In CB HSCs, genes involved in transplant rejection (such as *HLA-A*, *HLA-B*, *HLA-C*, *HLA-DPA1*, *HLA-DQA1* and *HLA-DQB1*) or immune response (such as interferon-induced proteins and CXCR4) were strongly downregulated (labeled with asterisks, Table 2).

In BM HSCs, genes involved in cell survival (such as the baculoviral IAP repeat-containing 5, also known as survivin [14]) and cell cycle progression (such as cyclin A2) were most active (Table 3). In contrast, in PB HSCs, cell cycle inhibitors such as cyclin-dependent kinase inhibitor 1C (CDKN1C; also known as $p57^{KIP2}$) were overexpressed (Table 4), suggesting a less active cell cycle or cell proliferating condition in HSCs from PB. PB HSCs expressed the highest levels of *PROM1* (*CD133*) gene, a marker of HSCs and other stem cell types, such as colon epithelial stem cells [13], indicating a higher stemness status of this HSC subtype. PB HSCs also expressed more genes involved in endothelial (vascular endothelial growth factor A) and megakaryocytic (RUNX1 [15]) differentiation, indicating their multipotency. Ezrin, a cell membrane–cytoskeleton cross-linker that can increase cellular motility [16,17] was also strongly expressed in PB HSCs (Table 4).

Functional grouping reveals unique biological properties in each HSC subtype

Gene signatures provided indications of the functional differences among these three HSC subtypes. To provide quantitative evidence and to gain more insights into the functional consequences of differential gene expression patterns, probe sets enriched or repressed in each HSC subtype were subjected to the GO database search [18] to find statistically overrepresented functional groups within each gene list. The DAVID 2008 Web-based tool, a graph theory evidence-based method to agglomerate gene/protein identifiers [12], was used for this task.

The GO categories of biological processes statistically overrepresented ($p < 0.05$) are shown in Figure 2. The predominant processes in CB HSCs included those pertaining to DNA repair, nuclear messenger RNA splicing and export, and ribosomal RNA or transfer RNA processing (Figure 2A). In contrast, response to stress and endogenous antigen processing and presentation pathways were less active in CB-derived HSCs (Figure 2A). Of the three HSC subtypes, HSCs from BM possessed the largest number of genes involved in cell cycle progression (especially those involved in mitosis and spindle biogenesis) and transplant rejection (i.e. major histocompatibility complex class I endogenous antigen processing and presentation). BM HSCs also possessed the fewest genes involved in cell cycle arrest and cellular migration, indicating an active proliferation yet stationary property of HSCs inside BM (Figure 2B). HSCs from PB expressed the most genes involved in cell motility, reflecting the fact that they were collected after granulocyte-macrophage colony-stimulating factor mobilization (Figure 2C). PB HSCs may also replicate less and participate more in immune response or angiogenesis, because the genes involved in these biological processes were the most differentially expressed in PB HSCs (Figure 2C).

Coordinated changes in functional networks of CB HSC upregulated genes

The above genomics data suggest different biological potentials of HSC subtypes isolated from various anatomic locations. Increasing evidence shows that genes do not act individually but collaborate in genetic networks. To better understand how genes enriched in each population of $CD34^+$ $CD38^-$ cells are related and how they relate to cellular function, we performed genetic network analysis for signature genes using the IPA Web tool and the Ingenuity knowledge database. The knowledge base behind IPA summarizes known molecular interactions evidenced in the published literature. The term *network* in IPA does not refer to a biological or canonical pathway with a distinct function (i.e. angiogenesis) but is, rather, a reflection of all the interactions of a given protein as defined in the literature (see "Materials and Methods").

In CB HSCs, a major network consisting of 258 genes was identified (Figure 3A). This network revealed genes whose roles may be more crucial than others in CB HSCs. *GATA1*, *MYC*, *NMYC*, *JUN* and *YWHAZ* are the "hub" genes (genes involved in the greatest number of interactions with other components) that either preserve the stability of the whole network or link submodules in the same network to each other (Figure 3A). These hub genes are all involved in the regulation

Table 1. Top 50 genes enriched in cord blood $CD34^+$ $CD38^-$ hematopoietic stem cells

Probe set ID	UniGene ID	Gene title	Gene symbol	Location
203196_at	Hs.508423	ATP-binding cassette, sub-family C (CFTR/MRP, member 4)	*ABCC4*	chr13q32
212224_at	Hs.76392	Aldehyde dehydrogenase 1 family, member A1	*ALDH1A1*	chr9q21.13
204446_s_at	Hs.89499	Arachidonate 5-lipoxygenase	*ALOX5*	chr10q11.2
204174_at	Hs.507658	Arachidonate 5-lipoxygenase-activating protein	*ALOX5AP*	chr13q12
203304_at	Hs.533336	BMP and activin membrane-bound inhibitor homolog (*Xenopus laevis*)	*BAMBI*	chr10p12.3-p11.2
202391_at	Hs.201641	Brain abundant, membrane attached signal protein 1	*BASP1*	chr5p15.1-p14
215440_s_at	Hs.184736	BEX family member 4	*BEX4*	chrXq22.1-q22.3
218829_s_at	Hs.709651	Chromodomain helicase DNA binding protein 7	*CHD7*	chr8q12.2
201735_s_at	Hs.481186	Chloride channel 3	*CLCN3*	chr4q33
205624_at	Hs.646	Carboxypeptidase A3 (mast cell)	*CPA3*	chr3q21-q25
205159_at	Hs.592192	Colony stimulating factor 2 receptor, beta, low-affinity (granulocyte-macrophage)	*CSF2RB*	chr22q13.1
218858_at	Hs.112981	DEP domain containing 6	*DEPDC6*	chr8q24.12
219799_s_at	Hs.179608	Dehydrogenase/reductase (SDR family) member 9	*DHRS9*	chr2q31.1
209560_s_at	Hs.533717	Delta-like 1 homolog (*Drosophila*)	*DLK1*	chr14q32
204720_s_at	Hs.647643	DnaJ (Hsp40) homolog, subfamily C, member 6	*DNAJC6*	chr1pter-q31.3
201123_s_at	Hs.534314	Eukaryotic translation initiation factor 5A	*EIF5A*	chr17p13-p12
220643_s_at	Hs.173438	Fas apoptotic inhibitory molecule	*FAIM*	chr3q22.3
204417_at	Hs.513439	Galactosylceramidase	*GALC*	chr14q31
204115_at	Hs.83381	Guanine nucleotide binding protein (G protein), gamma 11	*GNG11*	chr7q21
215780_s_at	Hs.436687	SET nuclear oncogene	*SET*	chr9q34
207067_s_at	Hs.1481	Histidine decarboxylase	*HDC*	chr15q21-q22
203394_s_at	Hs.250666	Hairy and enhancer of split 1, (*Drosophila*)	*HES1*	chr3q28-q29
208025_s_at	Hs.505924	High mobility group AT-hook 2	*HMGA2*	chr12q15
200679_x_at	Hs.712556	High mobility group box 1	*HMGB1*	chr13q12
207194_s_at	Hs.706750	Intercellular adhesion molecule 4 (Landsteiner-Wiener blood group)	*ICAM4*	chr19p13.2-cen
211953_s_at	Hs.643743	Importin 5	*IPO5*	chr13q32.2
218170_at	Hs.483296	Isochorismatase domain containing 1	*ISOC1*	chr5q22.1-q33.3
205885_s_at	Hs.694732	Integrin, alpha 4 (antigen CD49D, alpha 4 subunit of VLA-4 receptor)	*ITGA4*	chr2q31.3
221558_s_at	Hs.555947	Lymphoid enhancer-binding factor 1	*LEF1*	chr4q23-q25
209204_at	Hs.436792	LIM domain only 4	*LMO4*	chr1p22.3
201669_s_at	Hs.519909	Myristoylated alanine-rich protein kinase C substrate	*MARCKS*	chr6q22.2
212473_s_at	Hs.501928	Microtubule associated monooxygenase, calponin and LIM domain containing 2	*MICAL2*	chr11p15.3
212604_at	Hs.154655	Mitochondrial ribosomal protein S31	*MRPS31*	chr13q14.11
212372_at	Hs.16355	Myosin, heavy chain 10, non-muscle	*MYH10*	chr17p13
205006_s_at	Hs.60339	*N*-myristoyltransferase 2	*NMT2*	chr10p13
208857_s_at	Hs.279257	Protein-L-isoaspartate (D-aspartate) O-methyltransferase	*PCMT1*	chr6q24-q25
218676_s_at	Hs.285218	Phosphatidylcholine transfer protein	*PCTP*	chr17q21-q24
206726_at	Hs.128433	Prostaglandin D_2 synthase, hematopoietic	*PGDS*	chr4q22.3
209318_x_at	Hs.444975	Pleiomorphic adenoma gene-like 1	*PLAGL1*	chr6q24-q25
213241_at	Hs.584845	Plexin C1	*PLXNC1*	chr12q23.3
219485_s_at	Hs.522752	Proteasome (prosome, macropain) 26S subunit, non-ATPase, 10	*PSMD10*	chrXq22.3
206157_at	Hs.591286	Pentraxin-related gene, rapidly induced by IL-1 beta	*PTX3*	chr3q25
212012_at	Hs.332197	Peroxidasin homolog (*Drosophila*)	*PXDN*	chr2p25
218668_s_at	Hs.119889	RAP2C, member of RAS oncogene family	*RAP2C*	chrXq25
202988_s_at	Hs.75256	Regulator of G-protein signaling 1	*RGS1*	chr1q31
205352_at	Hs.478153	Serpin peptidase inhibitor, clade I (neuroserpin), member 1	*SERPINI1*	chr3q26.1
203775_at	Hs.489190	Solute carrier family 25, member 13 (citrin)	*SLC25A13*	chr7q21.3
220059_at	Hs.435579	Signal transducing adaptor family member 1	*STAP1*	chr4q13.2
201147_s_at	Hs.701968	TIMP metallopeptidase inhibitor 3	*TIMP3*	chr22q12.1
219478_at	Hs.36688	WAP four-disulfide core domain 1	*WFDC1*	chr16q24.3

Table 2. Top 50 genes downregulated in cord blood CD34+ CD38− hematopoietic stem cells

Probe set ID	UniGene ID	Gene title	Gene symbol	Location
209160_at	Hs.78183	Aldo-keto reductase family 1, member C3	*AKR1C3*	chr10p15-p14
207848_at	Hs.89648	Arginine vasopressin	*AVP*	chr20p13
205681_at	Hs.227817	BCL2-related protein A1	*BCL2A1*	chr15q24.3
201641_at	Hs.118110	Bone marrow stromal cell antigen 2	*BST2*	chr19p13.2
208683_at	Hs.350899	Calpain 2, (m/II) large subunit	*CAPN2*	chr1q41-q42
209732_at	Hs.85201	C-type lectin domain family 2, member B	*CLEC2B*	chr12p13-p12
213415_at	Hs.655445	Chloride intracellular channel 2	*CLIC2*	chrXq28
217028_at	Hs.593413	Chemokine (C-X-C motif) receptor 4	*CXCR4**	chr2q21
218094_s_at	Hs.655055	Dysbindin domain containing 2	*DBNDD2*	chr20q13.12
209383_at	Hs.505777	DNA-damage-inducible transcript 3	*DDIT3*	chr12q13.1-q13.2
201324_at	Hs.436298	Epithelial membrane protein 1	*EMP1*	chr12p12.3
205767_at	Hs.115263	Epiregulin	*EREG*	chr4q13.3
212998_x_at	Hs.728	Major histocompatibility complex, class II, DQ beta 1	*HLA-DQB1**	chr14q24-q31
206082_at	Hs.654480	HLA complex P5	*HCP5**	chr6p21.3
213932_x_at	Hs.181244	Major histocompatibility complex, class I, A	*HLA-A**	chr6p21.3
209140_x_at	Hs.549053	Major histocompatibility complex, class I, B	*HLA-B**	chr6p21.3
208812_x_at	Hs.654404	Major histocompatibility complex, class I, C	*HLA-C**	chr6p21.3
213537_at	Hs.347270	Major histocompatibility complex, class II, DP alpha 1	*HLA-DPA1**	chr6p21.3
213831_at	Hs.387679	Major histocompatibility complex, class II, DQ alpha 1	*HLA-DQA1**	chr6p21.3
211654_x_at	Hs.409934	Major histocompatibility complex, class II, DQ beta 1	*HLA-DQB1**	chr6p21.3
200599_s_at	Hs.192374	Heat shock protein 90kDa beta (Grp94), member 1	*HSP90B1**	chr12q24.2-q24.3
214453_s_at	Hs.82316	Interferon-induced protein 44	*IFI44**	chr1p31.1
204439_at	Hs.389724	Interferon-induced protein 44-like	*IFI44L**	chr1p31.1
204415_at	Hs.511731	Interferon, alpha-inducible protein 6	*IFI6**	chr1p35
201601_x_at	Hs.458414	Interferon induced transmembrane protein 1 (9-27)	*IFITM1**	chr11p15.5
205376_at	Hs.658245	Inositol polyphosphate-4-phosphatase, type II, 105kDa	*INPP4B*	chr4q31.21
204698_at	Hs.459265	Interferon stimulated exonuclease gene 20 kDa	*ISG20**	chr15q26
213281_at	Hs.525704	Jun oncogene	*JUN*	chr1p32-p31
203542_s_at	Hs.150557	Kruppel-like factor 9	*KLF9*	chr9q13
214056_at	Hs.632486	Myeloid cell leukemia sequence 1 (BCL2-related)	*MCL1*	chr1q21
212509_s_at	Hs.250723	Matrix-remodelling associated 7	*MXRA7*	chr17q25.1-q25.2
218966_at	Hs.487036	Myosin VC	*MYO5C*	chr15q21
202074_s_at	Hs.332706	Optineurin	*OPTN*	chr10p13
202464_s_at	Hs.195471	6-Phosphofructo-2-kinase/fructose-2, 6-biphosphatase 3	*PFKFB3*	chr10p14-p15
204279_at	Hs.654585	Proteasome (prosome, macropain) subunit, beta type, 9	*PSMB9*	chr6p21.3
204897_at	Hs.199248	Prostaglandin E receptor 4 (subtype EP4)	*PTGER4*	chr5p13.1
204070_at	Hs.17466	Retinoic acid receptor responder (tazarotene induced) 3	*RARRES3*	chr11q23
212044_s_at	Hs.523463	Ribosomal protein L27a	*RPL27A*	chr11p15
214041_x_at	Hs.433701	Ribosomal protein L37a	*RPL37A*	chr2q35
202028_s_at	Hs.380953	Ribosomal protein L38	*RPL38*	chr17q23-q25
221943_x_at	Hs.380953	Ribosomal protein L38	*RPL38*	chr17q23-q25
200908_s_at	Hs.437594	Ribosomal protein, large, P2	*RPLP2*	chr11p15.5-p15.4
204030_s_at	Hs.134665	Schwannomin interacting protein 1	*SCHIP1*	chr3q25.33
200665_s_at	Hs.111779	Secreted protein, acidic, cysteine-rich (osteonectin)	*SPARC*	chr5q31.3-q32
217995_at	Hs.511251	Sulfide quinone reductase-like (yeast)	*SQRDL*	chr15q15
208763_s_at	Hs.522074	TSC22 domain family, member 3	*TSC22D3*	chrXq22.3
202096_s_at	Hs.202	Translocator protein (18 kDa)	*TSPO*	chr22q13.31
202112_at	Hs.440848	von Willebrand factor	*VWF**	chr12p13.3
201368_at	Hs.503093	Zinc finger protein 36, C3H type-like 2	*ZFP36L2*	chr2p22.3-p21
201369_s_at	Hs.503093	Zinc finger protein 36, C3H type-like 2	*ZFP36L2*	chr2p22.3-p21

**Discussed in the text.*

Table 3. Top 50 genes enriched in bone marrow $CD34^+$ $CD38^-$ hematopoietic stem cells

Probe set ID	UniGene ID	Gene title	Gene symbol	Location
205239_at	Hs.645475	Amphiregulin (schwannoma-derived growth factor)	*AREG*	chr4q13-q21
214575_s_at	Hs.72885	Azurocidin 1 (cationic antimicrobial protein 37)	*AZU1*	chr19p13.3
218332_at	Hs.334370	Brain expressed, X-linked 1	*BEX1*	chrXq21-q23|Xq22
202095_s_at	Hs.514527	Baculoviral IAP repeat-containing 5 (survivin)	*BIRC5**	chr17q25
205950_s_at	Hs.23118	Carbonic anhydrase I	*CA1*	chr8q13-q22.1
214315_x_at	Hs.515162	Calreticulin	*CALR*	chr19p13.3-p13.2
213226_at	Hs.58974	Cyclin A2	*CCNA2**	chr4q25-q31
206488_s_at	Hs.120949	CD36 molecule (thrombospondin receptor)	*CD36*	chr7q11.2
218157_x_at	Hs.22065	CDC42 small effector 1	*CDC42SE1*	chr1q21.2
209057_x_at	Hs.485471	CDC5 cell division cycle 5-like (*S. pombe*)	*CDC5L*	chr6p21
201131_s_at	Hs.461086	Cadherin 1, type 1, E-cadherin (epithelial)	*CDH1**	chr16q22.1
209714_s_at	Hs.84113	Cyclin-dependent kinase inhibitor 3	*CDKN3*	chr14q22
219555_s_at	Hs.55028	Centromere protein N	*CENPN*	chr16q23.2
205382_s_at	Hs.155597	Complement factor D (adipsin)	*CFD*	chr19p13.3
214665_s_at	Hs.406234	Calcium binding protein P22	*CHP*	chr15q13.3
210140_at	Hs.143212	Cystatin F (leukocystatin)	*CST7*	chr20p11.21
204971_at	Hs.518198	Cystatin A (stefin A)	*CSTA*	chr3q21
201533_at	Hs.476018	Catenin (cadherin-associated protein), beta 1, 88kDa	*CTNNB1**	chr3p21
205653_at	Hs.421724	Cathepsin G	*CTSG*	chr14q11.2
220179_at	Hs.302028	Dipeptidase 3	*DPEP3*	chr16q22.1
214313_s_at	Hs.158688	Eukaryotic translation initiation factor 5 B	*EIF5B*	chr2q11.2
206871_at	Hs.99863	Elastase 2, neutrophil	*ELA2*	chr19p13.3
219672_at	Hs.274309	Erythroid associated factor	*ERAF*	chr16p11.2
209480_at	Hs.409934	Major histocompatibility complex, class II, DQ beta 1	*HLA-DQB1*	chr6p21.3
207165_at	Hs.72550	Hyaluronan-mediated motility receptor (RHAMM)	*HMMR*	chr5q33.2-qter
201163_s_at	Hs.479808	Insulin-like growth factor binding protein 7	*IGFBP7*	chr4q12
213792_s_at	Hs.465744	Insulin receptor	*INSR*	chr19p13.3-p13.2
219306_at	Hs.658939	Kinesin family member 15	*KIF15*	chr3p21.31
202209_at	Hs.111632	LSM3 homolog, U6 small nuclear RNA associated	*LSM3*	chr3p25.1
213975_s_at	Hs.706744	Lysozyme (renal amyloidosis)	*LYZ*	chr12q15
217692_at	Hs.674373	Mago-nashi homolog 2, proliferation-associated	*MAGOH2*	chr17p13.1
203948_s_at	Hs.458272	Myeloperoxidase	*MPO**	chr17q23.1
206023_at	Hs.418367	Neuromedin U	*NMU*	chr4q12
218039_at	Hs.615092	Nucleolar and spindle associated protein 1	*NUSAP1*	chr15q15.1
212239_at	Hs.132225	Phosphoinositide-3-kinase, regulatory subunit 1 (alpha)	*PIK3R1*	chr5q13.1
201407_s_at	Hs.705383	Protein phosphatase 1, catalytic subunit, beta isoform	*PPP1CB*	chr2p23
203401_at	Hs.654581	Phosphoribosyl pyrophosphate synthetase 2	*PRPS2*	chrXp22.3-p22.2
220051_at	Hs.72026	Protease, serine, 21 (testisin)	*PRSS21*	chr16p13.3
207341_at	Hs.928	Proteinase 3	*PRTN3*	chr19p13.3
222077_s_at	Hs.708122	Rac GTPase activating protein 1	*RACGAP1*	chr12q13.13
206851_at	Hs.73839	Ribonuclease, RNase A family, 3	*RNASE3*	chr14q24-q31
214097_at	Hs.190968	Ribosomal protein S21	*RPS21*	chr20q13.3
202917_s_at	Hs.416073	S100 calcium binding protein A8	*S100A8*	chr1q21
220232_at	Hs.379191	Stearoyl-CoA desaturase 5	*SCD5*	chr4q21.22
202083_s_at	Hs.464184	SEC14-like 1 (*S. cerevisiae*)	*SEC14L1*	chr17q25.1-q25.2
204496_at	Hs.401843	Striatin, calmodulin binding protein 3	*STRN3*	chr14q13-q21
206438_x_at	Hs.167165	Tectonic family member 2	*TCTN2*	chr12q24.31
202704_at	Hs.653129	Transducer of ERBB2, 1	*TOB1*	chr17q21
201690_s_at	Hs.368433	Tumor protein D52	*TPD52*	chr8q21
202954_at	Hs.93002	Ubiquitin-conjugating enzyme E2C	*UBE2C*	chr20q13.12

**Discussed in the text.*

Table 4. Top 50 genes enriched in granulocyte-macrophage colony-stimulating factor–mobilized peripheral blood CD34+ CD38− hematopoietic stem cells

Probe set ID	UniGene ID	Gene title	Gene symbol	Location
216356_x_at	Hs.458427	BAI1-associated protein 3	*BAIAP3*	chr16p13.3
212757_s_at	Hs.699281	Calcium/calmodulin-dependent protein kinase (CaM kinase) II gamma	*CAMK2G*	chr10q22
213183_s_at	Hs.106070	Cyclin-dependent kinase inhibitor 1C ($p57^{KIP2}$)	*CDKN1C**	chr11p15.5
219400_at	Hs.408730	Contactin associated protein 1	*CNTNAP1*	chr17q21
209432_s_at	Hs.522110	cAMP responsive element binding protein 3	*CREB3*	chr9pter-p22.1
217200_x_at	Hs.355264	Cytochrome b-561	*CYB561*	chr17q11-qter
209178_at	Hs.570079	DEAH (Asp-Glu-Ala-His) box polypeptide 38	*DHX38*	chr16q21-q22.3
208810_at	Hs.490745	DnaJ (Hsp40) homolog, subfamily B, member 6	*DNAJB6*	chr7q36.3
219759_at	Hs.591249	Endoplasmic reticulum aminopeptidase 2	*ERAP2*	chr5q15
211742_s_at	Hs.5509	Ecotropic viral integration site 2B	*EVI2B*	chr17q11.2
208621_s_at	Hs.712548	Ezrin	*EZR**	chr6q25.2-q26
205021_s_at	Hs.434286	Forkhead box N3	*FOXN3*	chr14q31.3
210428_s_at	Hs.514590	Hepatocyte growth factor-regulated tyrosine kinase substrate	*HGS*	chr17q25
208569_at	Hs.248174	Histone cluster 1, H2ab	*HIST1H2AB*	chr6p21.3
221435_x_at	Hs.709864	Hydroxypyruvate isomerase homolog (*E. coli*)	*HYI*	chr1p34.2
202439_s_at	Hs.460960	Iduronate 2-sulfatase (Hunter syndrome)	*IDS*	chrXq28
209342_s_at	Hs.656458	Inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase beta	*IKBKB*	chr8p11.2
201626_at	Hs.520819	Insulin induced gene 1	*INSIG1*	chr7q36
209099_x_at	Hs.224012	Jagged 1 (Alagille syndrome)	*JAG1*	chr20p12.1
220777_at	Hs.189915	Kinesin family member 13A	*KIF13A*	chr6p23
203068_at	Hs.7764	Kelch-like 21 (*Drosophila*)	*KLHL21*	chr1p36.31
210671_x_at	Hs.138211	Mitogen-activated protein kinase 8	*MAPK8*	chr10q11.22
202617_s_at	Hs.200716	Methyl CpG binding protein 2 (Rett syndrome)	*MECP2*	chrXq28
209636_at	Hs.73090	Nuclear factor of kappa light polypeptide gene enhancer in B-cells 2 (p49/p100)	*NFKB2*	chr10q24
203859_s_at	Hs.631841	Paralemmin	*PALM*	chr19p13.3
207838_x_at	Hs.505806	Pre-B-cell leukemia homeobox interacting protein 1	*PBXIP1*	chr1q21.3
215175_at	Hs.446559	Pecanex homolog (*Drosophila*)	*PCNX*	chr14q24.2
222125_s_at	Hs.654944	Hypoxia-inducible factor prolyl 4-hydroxylase	*PH-4*	chr3p21.31
211668_s_at	Hs.77274	Plasminogen activator, urokinase	*PLAU*	chr10q24
41577_at	Hs.45719	Protein phosphatase 1, regulatory (inhibitor) subunit 16B	*PPP1R16B*	chr20q11.23
204304_s_at	Hs.614734	Prominin 1	*PROM1**	chr4p15.32
203149_at	Hs.655455	Poliovirus receptor-related 2 (herpesvirus entry mediator B)	*PVRL2*	chr19q13.2
202963_at	Hs.709229	Regulatory factor X, 5 (influences HLA class II expression)	*RFX5*	chr1q21
211181_x_at	Hs.149261	Runt-related transcription factor 1 (acute myeloid leukemia 1; aml1 oncogene)	*RUNX1**	chr21q22.3
201250_s_at	Hs.706748	Solute carrier family 2 (facilitated glucose transporter), member 1	*SLC2A1*	chr1p35-p31.3
210993_s_at	Hs.656534	SMAD family member 1	*SMAD1*	chr4q31
212625_at	Hs.43812	Syntaxin 10	*STX10*	chr19p13.13
203753_at	Hs.644653	Transcription factor 4	*TCF4*	chr18q21.1
202085_at	Hs.50382	Tight junction protein 2 (zona occludens 2)	*TJP2*	chr9q13-q21
206472_s_at	Hs.709205	Transducin-like enhancer of split 3 [E(sp1) homolog, *Drosophila*]	*TLE3*	chr15q22
221702_s_at	Hs.288912	TM2 domain containing 3	*TM2D3*	chr15q26.3
219253_at	Hs.376722	Transmembrane protein 185B (pseudogene)	*TMEM185B*	chr2q14.2
202510_s_at	Hs.525607	Tumor necrosis factor, alpha-induced protein 2	*TNFAIP2*	chr14q32
209295_at	Hs.521456	Tumor necrosis factor receptor superfamily, member 10b	*TNFRSF10B*	chr8p22-p21
221962_s_at	Hs.700368	Ubiquitin-conjugating enzyme E2H (UBC8 homolog, yeast)	*UBE2H*	chr7q32
214792_x_at	Hs.25348	Vesicle-associated membrane protein 2 (synaptobrevin 2)	*VAMP2*	chr17p13.1
212171_x_at	Hs.73793	Vascular endothelial growth factor A	*VEGFA**	chr6p12
202932_at	Hs.194148	V-yes-1 Yamaguchi sarcoma viral oncogene homolog 1	*YES1*	chr18p11.31-.21
219571_s_at	Hs.431471	Zinc finger protein 12	*ZNF12*	chr7p22.1
219540_at	Hs.460645	Zinc finger protein 267	*ZNF267*	chr16p11.2

**Discussed in the text.*

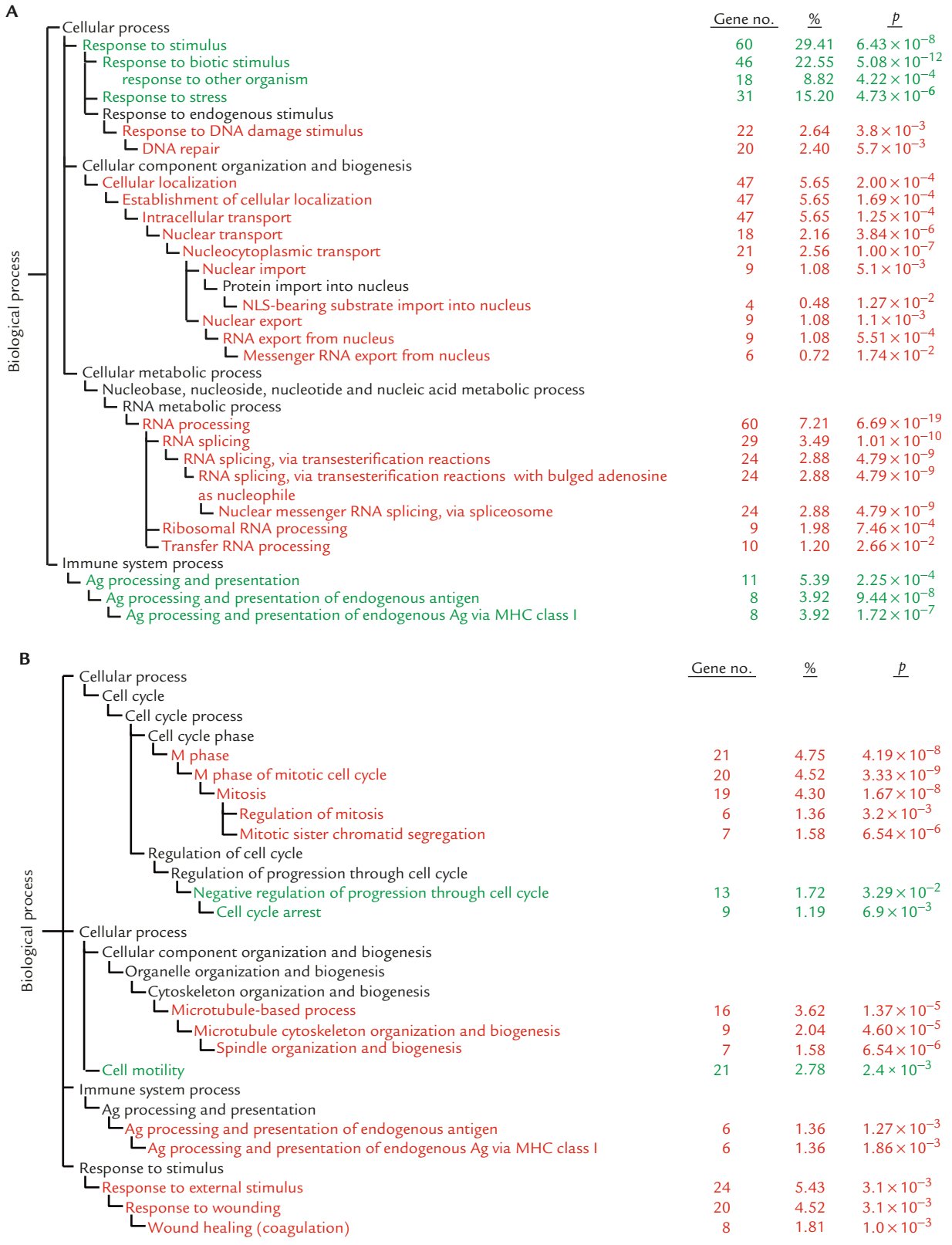

Figure 2. Unique biological processes in hematopoietic stem cells from different anatomic locations. Selected biological processes overrepresented by genes enriched or downregulated in $CD34^+$ $CD38^-$ cells isolated from (A) cord blood, (B) bone marrow or (C) peripheral blood. Probe sets differentially expressed (either up or down) in each hematopoietic stem cell subtype were subjected to gene set enrichment analysis using the DAVID 2008 Web tool. These categories were selected from the Biological Process organizing principle in the Gene Ontology project. Categories in red indicate increased expression; those in green indicate decreased expression. The number of genes, their percentage in cell type-specific probe sets, and *p* values for each category significantly ($p < 0.05$) overrepresented are listed. NLS = nuclear localization sequence; Ag = antigen.

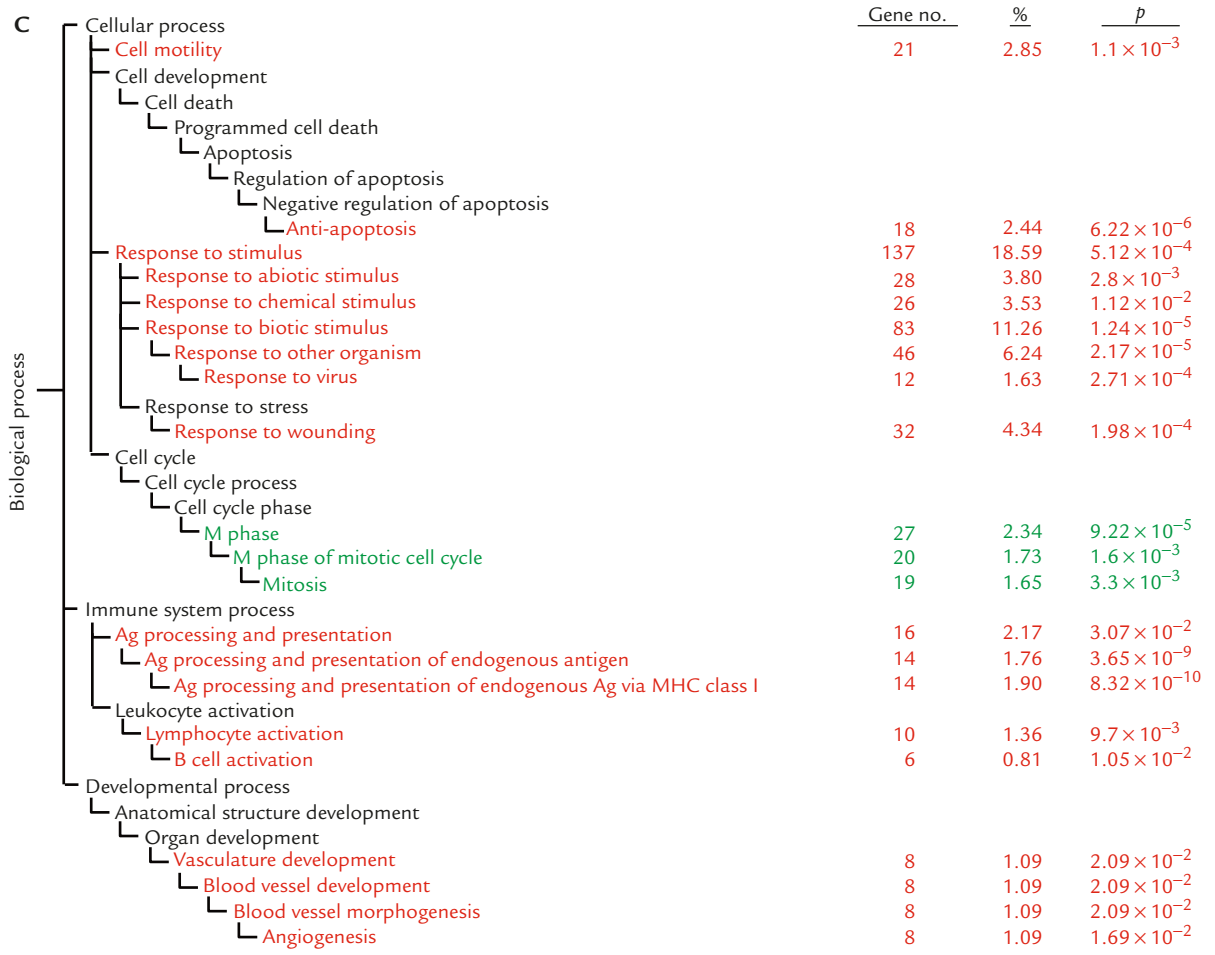

Figure 2. *(continued)*

of a specific sub-network in CB HSCs (Figures 3B–D), suggesting their critical roles in maintaining the basic properties and biological function of CB HSCs.

GATA1 regulates a specific sub-network in CB HSCs (Figure 3D). This transcription factor was identified as an erythroid-specific gene essential for erythrocytic differentiation at relatively early stages [19]. KLF1 is also found in this GATA1 sub-network. GATA1 and KLF1 (EKLF) are expressed in precursors and are definitive differentiation-associated genes [20], while GATA1 also cooperates with RUNX1 in megakaryocytic differentiation [15]. The presence of GATA1 indicates that HSCs from CB may contain relatively more lineage-specific precursors than those isolated from BM or PB.

Genetic networks of BM HSCs

Hepatic growth factor, *KRAS*, *CTNNB1* (β-catenin), *CDH1* (also known as E-cadherin), *PAX3* and *CDC2* are hub genes in BM HSCs (Figure 4). Hepatic growth factor and myeloperoxidase, which can both support self-renewal and keep HSCs in an undifferentiated state [13], are highly expressed in this subtype of HSC (Table 3 and Figure 4).

Discussion

Somatic and embryonic stem cells maintain the ability to self-renew and differentiate and, therefore, possess great potentials in regenerative medicine for healing tissues that are unable to repair themselves. In this study, we performed an extensive comparative transcriptome and gene network analyses of CD34+ HSCs from different sources. The results of these analyses will help to unravel the riddle of HSCs and contribute to developments in cell-based therapy.

In addition to identifying specific genes, we also applied refined computational methods to highlight key functional networks. There is increasing recognition that a systematic approach is necessary to view the overall molecular events responsible for a given biological process [21,22]. In this study, we applied systems biology tools to reveal the functional influences of differential gene expression. Knowledge of the enriched or repressed GO categories in each HSC subtype will improve understanding of their basic biological properties and aid their development as potential clinical applications.

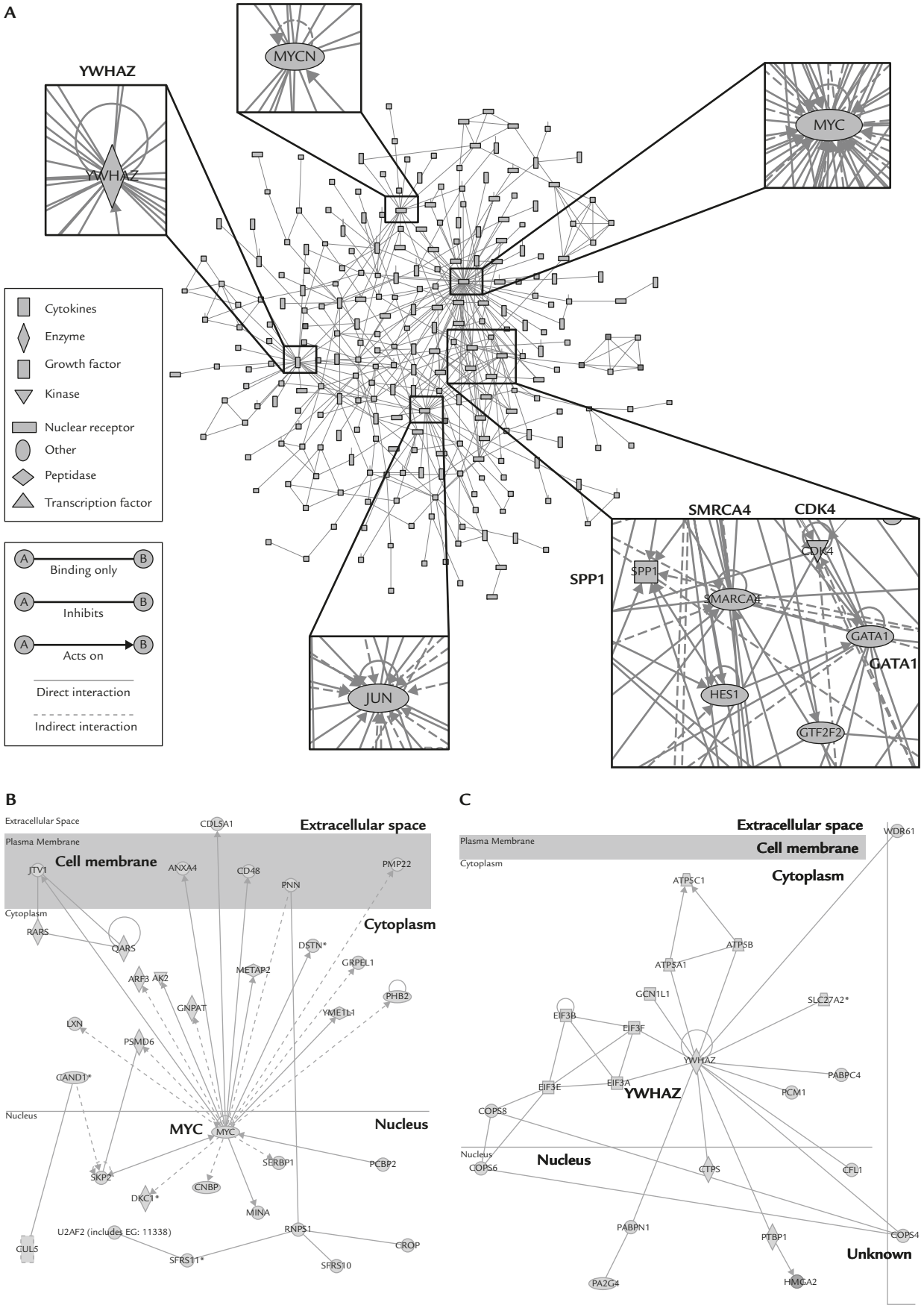

Figure 3. Genetic networks as a framework for the interpretation of cord blood CD34$^+$ CD38$^-$ stem cell biology. (A) Functional networks composed of multiple genes, some of which (such as *MYC*, *NMYC*, *GATA1*, *YWHAZ* and *JUN*) are hub genes with the

Cell cycle regulators were upregulated in all three HSC types, indicating a restricted cell cycle regulation. However, in BM HSCs, genes involved in mitosis, chromosome condensation and mitotic sister chromatid segregation were more abundant, indicating a relatively active cell cycle and replication. The downregulation of cell cycle arrest genes in BM HSCs was consistent with this theme. Genes involved in cell motility and small GTPase (Rho) signal transduction were downregulated, indicating an inert and immobile property of HSCs inside BM.

In PB HSCs, the regulation of transcription from RNA polymerase II promoter and the NF-κB and MAPKKK cascades were active, indicating the physiological activities of these stem cells. For example, angiogenesis and vascular development-related genes were highly expressed in PB HSCs, which are more motile, suggesting that they should aid vascular morphogenesis and could be useful for the treatment of vascular diseases.

Among the three HSC subtypes analyzed, HSCs from CB appear to provide an excellent source for transplantation. Genes involved in immunogenic responses, such as major histocompatibility complex class I-mediated endogenous antigen processing/presentation and cell defense response, are least active in this subtype, suggesting a lower risk of transplant rejection or graft-versus-host disease. HSCs from PB or BM, especially those from PB, expressed more genes associated with active antigen presentation, response to biotic stimuli, and lymphocyte (B cells) activation, and would, therefore, be less suitable than CB HSCs in clinical applications.

Our analyses had certain limitations. The functional networks were mapped based on predetermined knowledge databases (GO and IPA), in which molecular interactions were established under various physiological or pathologic conditions. It is known that molecular functions vary according to the cellular and tissue contexts, and the stemness genes may, therefore, not interact in a similar way in human CD34+ cells. Functional analyses are needed for the rigorous evaluation of individual gene interactions deduced by *in silico* modeling.

Acknowledgments

The authors acknowledge the technical services provided by the Microarray and Gene Expression Analysis Core Facility of the National Yang-Ming University

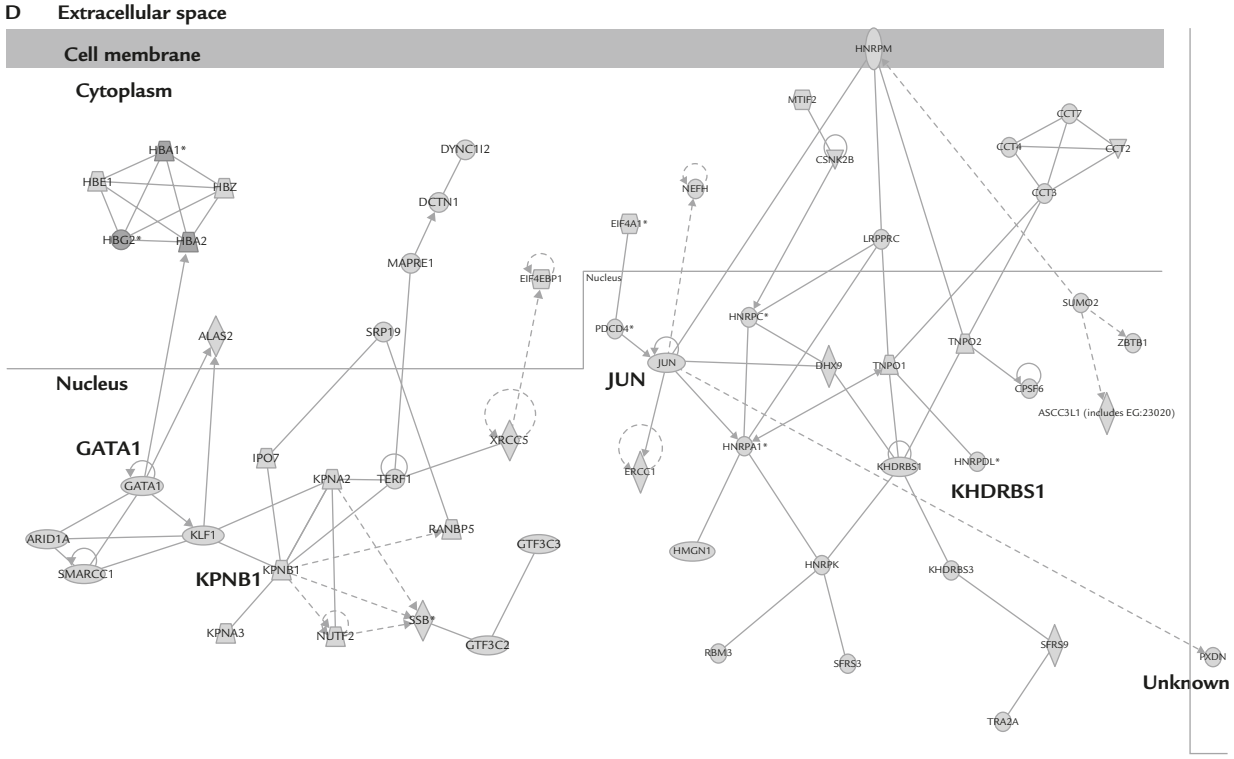

Figure 3. *(continued)*
most connectivity to other network genes. This network is displayed graphically as nodes (gene products) and edges (biological relationships between nodes) mapped by the Ingenuity Pathway Analysis tool. The intensity of the node color indicates the degree of upregulation. (B–C) A subnetwork regulated by (B) *MYC* or (C) *YWHAZ*. Genes are arranged according to their positions in a cell. (D) Two subnetworks regulated by *GATA1* and *KPNB1* (left) or *JUN* and *KHDRBS1* (right), respectively.

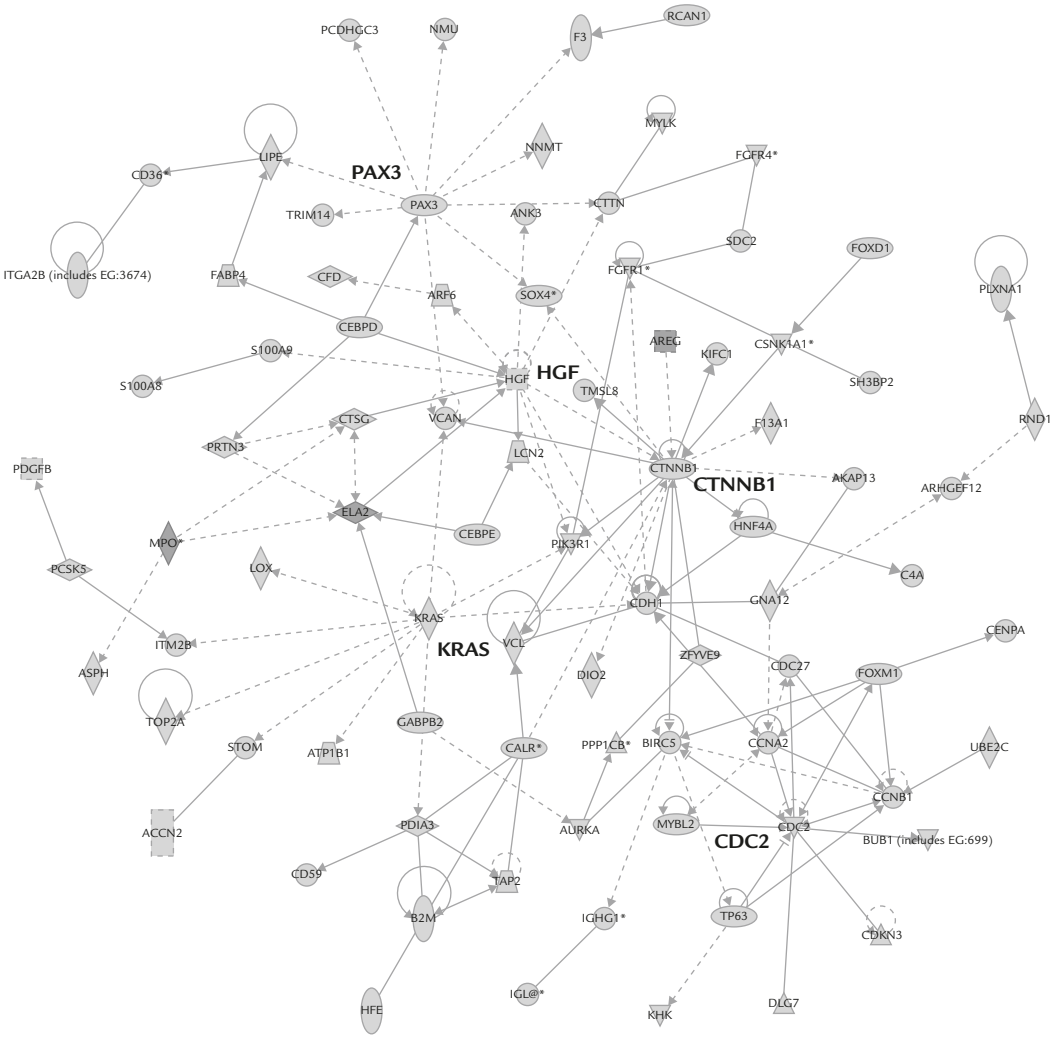

Figure 4. The major genetic network in bone marrow hematopoietic stem cells. Hub genes are labeled.

VGH Genome Research Center (VYMGC). This work was supported by the Mackay Memorial Hospital (MMH-HB-97-05) and the National Science Council (NSC97-3111-B-010-004).

References

1. Brunstein CG, Baker KS, Wagner JE. Umbilical cord blood transplantation for myeloid malignancies. *Curr Opin Hematol* 2007;14:162–9.
2. Ng YY, van Kessel B, Lokhorst HM, Baert MR, van den Burg CM, Bloem AC, Staal FJ. Gene-expression profiling of $CD34^+$ cells from various hematopoietic stem-cell sources reveals functional differences in stem-cell activity. *J Leukoc Biol* 2004; 75:314–23.
3. Oswald J, Steudel C, Salchert K, et al. Gene-expression profiling of $CD34^+$ hematopoietic cells expanded in a collagen I matrix. *Stem Cells* 2006;24:494–500.
4. Ivanova NB, Dimos JT, Schaniel C, Hackney JA, Moore KA, Lemischka IR. A stem cell molecular signature. *Science* 2002; 298:601–4.
5. Loges S, Fehse B, Brockmann MA, et al. Identification of the adult human hemangioblast. *Stem Cells Dev* 2004;13:229–42.
6. Jaatinen T, Hemmoranta H, Hautaniemi S, et al. Global gene expression profile of human cord blood-derived $CD133^+$ cells. *Stem Cells* 2006;24:631–41.
7. Georgantas RW 3rd, Tanadve V, Malehorn M, et al. Microarray and serial analysis of gene expression analyses identify known and novel transcripts overexpressed in hematopoietic stem cells. *Cancer Res* 2004;64:4434–41.
8. Featherstone DE, Broadie K. Wrestling with pleiotropy: genomic and topological analysis of the yeast gene expression network. *Bioessays* 2002;24:267–74.
9. Jen CH, Manfield IW, Michalopoulos I, Pinney JW, Willats WG, Gilmartin PM, Westhead DR. The Arabidopsis co-expression tool (ACT): a WWW-based tool and database for microarray-based gene expression analysis. *Plant J* 2006; 46:336–48.
10. Chang SJ, Huang TS, Wang KL, et al. Genetic network analysis of human $CD34^+$ hematopoietic stem/precursor cells. *Taiwan J Obstet Gynecol* 2008;47:422–30.
11. Wang HW, Trotter MW, Lagos D, et al. Kaposi sarcoma herpesvirus-induced cellular reprogramming contributes to the lymphatic endothelial gene expression in Kaposi sarcoma. *Nat Genet* 2004;36:687–93.

12. Dennis G Jr, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* 2003; 4:P3.
13. Huang TS, Hsieh JY, Wu YH, et al. Functional network reconstruction reveals somatic stemness genetic maps and dedifferentiation-like transcriptome reprogramming induced by GATA2. *Stem Cells* 2008;26:1186–201.
14. Malcles MH, Wang HW, Koumi A, Tsai YH, Yu M, Godfrey A, Boshoff C. Characterisation of the anti-apoptotic function of survivin-Δ Ex3 during TNFα-mediated cell death. *Br J Cancer* 2007;96:1659–66.
15. Elagib KE, Racke FK, Mogass M, Khetawat R, Delehanty LL, Goldfarb AN. RUNX1 and GATA-1 coexpression and cooperation in megakaryocytic differentiation. *Blood* 2003; 101:4333–41.
16. Khanna C, Wan X, Bose S, et al. The membrane-cytoskeleton linker ezrin is necessary for osteosarcoma metastasis. *Nat Med* 2004;10:182–6.
17. Yu Y, Khan J, Khanna C, Helman L, Meltzer PS, Merlino G. Expression profiling identifies the cytoskeletal organizer ezrin and the developmental homeoprotein Six-1 as key metastatic regulators. *Nat Med* 2004;10:175–81.
18. Harris MA, Clark J, Ireland A, et al. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res* 2004;32:D258–61.
19. Kawabata H, Germain RS, Ikezoe T, Tong X, Green EM, Gombart AF, Koeffler HP. Regulation of expression of murine transferrin receptor 2. *Blood* 2001;98:1949–54.
20. Crossley M, Tsang AP, Bieker JJ, Orkin SH. Regulation of the erythroid Kruppel-like factor (EKLF) gene promoter by the erythroid transcription factor GATA-1. *J Biol Chem* 1994;269:15440–4.
21. Barabasi AL, Oltvai ZN. Network biology: understanding the cell's functional organization. *Nat Rev Genet* 2004;5: 101–13.
22. Janes KA, Yaffe MB. Data-driven modelling of signal-transduction networks. *Nat Rev Mol Cell Biol* 2006;7:820–8.