

Developmental Fate Determination and Marker Discovery in Hematopoietic Stem Cell Biology Using Proteomic Fingerprinting*[§]

Elaine Sponcer‡, Nathalie Brouard§¶, Susie K. Nilsson§||, Brenda Williams§||, Mira C. Liu§, Richard D. Unwin‡, David Blinco‡, Ewa Jaworska‡, Paul J. Simmons§¶, and Anthony D. Whetton‡**

In hematopoiesis, co-expression of Sca-1 and c-Kit defines cells (LS⁺K) with long term reconstituting potential. In contrast, poorly characterized LS⁻K cells fail to reconstitute lethally irradiated recipients. Relative quantification mass spectrometry and transcriptional profiling were used to characterize LS⁺K and LS⁻K cells. This approach yielded data on >1200 proteins. Only 32% of protein changes correlated to mRNA modulation demonstrating post-translational protein regulation in early hematopoietic development. LS⁺K cells had lower expression of protein synthesis proteins but did express proteins associated with mature cell function. Major increases in erythroid development proteins were observed in LS⁻K cells; based on this assessment of erythroid potential we showed them to be principally erythroid progenitors, demonstrating effective use of discovery proteomics for definition of primitive cells. *Molecular & Cellular Proteomics* 7:573–581, 2008.

Hematopoiesis generates mature cells of many different lineages via a process of differentiation and development from a common multipotent hematopoietic stem cell (HSC).¹ Although the mechanisms that govern these processes remain to be fully elucidated, two experimental approaches, namely the establishment of rigorous clonal assays to measure HSC function in animal models and the development of methodologies for the prospective isolation of these rare cells, have contributed to HSC being among the best understood population of tissue stem cells in adult vertebrate physiology.

From the ‡Stem Cell and Leukaemia Proteomics Laboratory, Faculty of Medical and Human Sciences, University of Manchester, Kinnaid House, Kinnaid Road, Manchester M20 4QL, United Kingdom and §Stem Cell Laboratory, Peter MacCallum Cancer Centre, St. Andrews Place, East Melbourne, Victoria 3002, Australia

Received, June 22, 2007, and in revised form, December 14, 2007
Published, MCP Papers in Press, December 14, 2007, DOI 10.1074/mcp.M700292-MCP200

¹ The abbreviations used are: HSC, hematopoietic stem cell; LS⁺K, Lin⁻Sca-1⁺Kit⁺; LS⁻K, Lin⁻Sca-1⁻Kit⁺; iTRAQ, isobaric tagging for relative and absolute quantitation; HPP-CFC, high proliferative potential colony-forming cell; LPP-CFC, low proliferative potential colony-forming cells; CFU-e, colony-forming unit-erythroid; hnRNP, heterogeneous nuclear ribonucleoprotein.

Over the past 25 years, a wide variety of strategies have been developed for the prospective isolation of HSC. Such is the state of the art of HSC purification that several groups have reported reconstitution of mice with a *single* prospectively isolated HSC (1–7). The expression of a range of cell surface markers has been shown to enable enrichment of cells with long term reconstituting capacity (1–7). In the mouse, the most commonly utilized methodology for isolation of HSC was developed by Weissman and co-workers (7). According to this strategy, HSCs are restricted to a subpopulation of bone marrow cells that lack expression of mature hematopoietic cell lineage markers such as Mac1 (macrophage marker), Gr-1 (granulocytic marker), CD4/CD8 and B220 (T and B lymphoid markers, respectively), and Ter119 (erythroid marker) (6, 8). Within this lineage-negative (Lin⁻) fraction, primitive hematopoietic stem and progenitor cells are further identified by their expression of Sca-1 (Ly-6E) and the receptor tyrosine kinase c-Kit (stem cell factor receptor). c-Kit is present on both primitive hematopoietic cells and their immediate progeny (3). c-Kit mutant mice are sterile, lack melanocytes, and have defects in their hematopoiesis associated with a macrocytic anemia (9, 10).

Sca-1 is a glycosylphosphatidylinositol-linked cell surface protein with a potential role in membrane protein organization for signaling via lipid raft structures (11). It has also been implicated in Src protein-tyrosine kinase family signaling and in the regulation of integrin function in stem cells (12–14). Sca-1 expression is a marker enabling enrichment of stem cells (3, 7) as demonstrated by the capacity of Lin⁻Sca-1⁺Kit⁺ (LS⁺K) cells to reconstitute multilineage hematopoiesis in lethally irradiated mice, whereas the corresponding Lin⁻Sca-1⁻Kit⁺ (LS⁻K) population fail to do so (3). In Sca-1-null murine bone marrow there was a decreased number of hematopoietic progenitor cells and a reduced ability of Sca-1-null cells to repopulate hematopoiesis in lethally irradiated mice. HSC from Sca-1-null mice with a Kit mutation at a single allele display a profound anemia with reduced progenitor cell numbers (15), far greater than observed in Sca-1-null animals, thus demonstrating the critical and linked role for both these proteins in hematopoiesis. Indeed there is evidence that Sca-1 expression can modulate the expression of c-Kit (13).

Although the nature of the LS⁺K cells is defined by their ability to promote long term reconstitution in lethally irradiated recipient mice, the LS⁻K cells have not been characterized beyond their lack of ability to rescue myeloablated mice. Subsequent studies have defined a discrete subpopulation of cells with progenitor cell activity within the LS⁻K compartment such as common myeloid progenitor cells (IL-7R α ⁻, Fc γ R^{low}, CD34⁺, Lin⁻, Sca⁻, Kit⁺ cells) (16). However, the characteristics of the majority of the cells await definition. We have shown previously that systematic proteomics analysis of primary hematopoietic stem cells is feasible on small populations of cells (~1 million) using an isobaric tagging procedure for relative quantification coupled to two-dimensional LC plus MS/MS (17). This approach revealed that post-translational regulation of protein expression is critically important in hematopoietic cell development, and therefore many transcriptomic changes observed do not affect the proteome (17, 18). Similarly alterations in the proteome occur in the absence of significant changes in transcript expression. To define the Lin⁻Sca⁻Kit⁺ population we compared its proteome with that of Lin⁻Sca⁺Kit⁺ cells and demonstrate that systematic analysis of enriched hematopoietic populations can give a direct insight into their biological characteristics.

EXPERIMENTAL PROCEDURES

Enrichment of Hematopoietic Cells—C57Bl/6J mice were purchased from Animal Resources Centre (Perth, Western Australia, Australia) and housed for at least a week prior to experimental use.

Murine bone marrow was collected from femora, tibiae, and iliac crests. Lineage marker-depleted cells were enriched as described previously (19). Briefly low density (<1.077 g/cm³) cells were labeled with a mixture of biotinylated rat anti-mouse antibodies: B220, macrophage antigen-1 (Mac-1), Gr-1, CD4, CD8, CD3, CD5, and Ter119. Lineage marker-positive cells were removed by immunomagnetic selection using magnetic activated cell separation (Miltenyi Biotec, Bergisch, Gladbach, Germany). Lineage marker-negative cells were labeled with Sca-1-FITC (1 μ g/5 \times 10⁶ cells; Pharmingen), c-Kit-phycoerythrin (1 μ g/5 \times 10⁶ cells; Pharmingen), and streptavidin-Red 670 (1:160 final concentration; Invitrogen). LS⁻K and LS⁺K cells were isolated by flow cytometry as described previously (19) and stored as dry pellets at -80 °C prior to mass spectrometric analysis.

Liquid Chromatography and Mass Spectrometry—Methods used for hematopoietic cell analyses were as described previously (18), and the experiment was run in triplicate. Briefly cells were solubilized in 0.5 M triethylammonium bicarbonate (Sigma) + 0.05% (w/v) SDS on ice for 20 min. Protein (90 μ g in 100 μ l) was reduced by addition of 5 μ l of 50 mM tris(2-carboxyethyl)phosphine, and reduced cysteine residues were then blocked by addition of 2.5 μ l of 200 mM methylmethanethiosulfate in isopropanol. Protein was digested by addition of 18 μ l of trypsin at 0.5 μ g/ μ l and incubated at 37 °C overnight. Peptides were dried, reconstituted in 20 μ l 0.5 M triethylammonium bicarbonate, and labeled with iTRAQ reagent (Applied Biosystems Inc., Framingham, MA) as described previously (17).

For separation prior to reversed phase LC-MS/MS peptides were fractionated into ~50 fractions off line using a strong cation exchange column (10 \times 2.1-cm PolyLC Polysulfoethyl A column, 5- μ m beads, 200-Å pore size; Hichrom Ltd., Reading, Berks, UK) on a Dionex LC system using the gradient described previously (18). Dried peptide fractions were resuspended in 180 μ l of 2% (v/v) acetonitrile, 0.1% (v/v) formic acid, and a 60- μ l aliquot was loaded onto a 75- μ m-

inner diameter \times 15-cm column packed with C₁₈ PepMap100 (3 μ m, 100 Å) using an LC Packings (Amsterdam, Netherlands) UltiMate™ pump and separated as described previously (17). Data were acquired using an independent data acquisition protocol where an MS scan was taken and then the two highest ions were selected for fragmentation followed by dynamic exclusion for 1 min. Data were processed using ProQUANT software version 1.1 (Applied Biosystems) as described previously (18). Briefly data were searched against the murine International Protein Index (IPI) database version 3.13 (50,489 entries), allowing for iTRAQ labeling on lysines and N termini and cysteine modification with methylmethanethiosulfate with mass tolerances of 0.15 Da for MS and 0.1 Da for MS/MS, allowing for one missed (trypsin) cleavage, and including peptides with a minimum confidence score of 70. Quantitation was performed using the default software settings. ProGROUP (Applied Biosystems Inc.) software was used to assign a single accession number for each protein. These outputs are available as supplemental data and provide alternative names and accession numbers for proteins that match to the identified peptides. For a protein to be considered significantly different between LS⁺K and LS⁻K cells, at least three (non-unique) peptides had to be identified and quantified, the -fold change had to be >1.2, that protein had to have a $p < 0.05$ from a Student's t test in two of three analyses, and the LS⁻K1 versus LS⁻K2 and LS⁺K1 versus LS⁺K2 ratios had to be between 0.92 and 1.10 in all replicates where data were obtained for that protein.

Transcriptome and Proteome Analysis—Total cellular RNA was prepared from 1 \times 10⁴ freshly isolated cells of each cell population using RNeasy minicolumns (Qiagen) according to the manufacturer's instructions. RNA was amplified and labeled using the Two-cycle Target Labeling and Control Reagents (Affymetrix, Santa Clara, CA) according to the manufacturer's instructions. 15 μ g of biotinylated amplified cRNA were hybridized onto Affymetrix GeneChip arrays (MOE430 Plus-2.0, Affymetrix). Arrays were scanned using Gene Array Scanner (Affymetrix).

Affymetrix GCOS 1.2 was used for data acquisition and analysis. The comparative analysis of results obtained for the different cell populations was performed using the Silicon Genetics GeneSpring GX 7.3.1 software (Agilent Technologies, Palo Alto, CA). Affymetrix CEL files were normalized using the GC-RMA method. A transcript was determined to be "significantly differentially expressed" if it was called "present" in all three replicate arrays and had a -fold change of ± 1.8 that was deemed to be different ($p < 0.05$) by Student's t test. To compare transcriptome and proteome data, IPI protein accession numbers were used to identify the relevant probe set identifiers from the microarray data set using the NetAffx tool (Affymetrix).

Transplantation and Colony-forming Cell Analysis—Transplantation analysis was performed using C57Bl/6J female recipient mice and male donor LS⁺K or LS⁻K cells. Six weeks after engraftment, bone marrow was harvested, DNA was isolated, and the proportion of male donor was determined by using real time PCR targeting the Y6 amplicon on the Y chromosome as described previously (20) and modified by Nilsson *et al.* (21). High proliferative potential colony-forming cells (HPP-CFCs) were assayed in a double layer nutrient agar culture system as described previously except that stem cell factor (100 ng/ml; Amgen, Thousand Oaks, CA) was added to CSF-1, IL-1, and IL-3 (21, 22). Two doses of 50 mg/kg busulfan were given subcutaneously 1 and 3 days prior to transplanting limiting numbers of LS⁺K and LS⁻K cells intravenously (10,000, 3000, 1000, and 300/mouse) (21). Colony forming assays were performed essentially as described previously (21).

For committed erythroid and megakaryocytic colony formation, cells were plated at 1, 3, 10, and 30 cells/well in 100 μ l of CellGro Good Manufacturing Practice serum-free medium (CellGenix, Antioch, IL) supplemented with 4 units/ml erythropoietin (Amgen). After 7

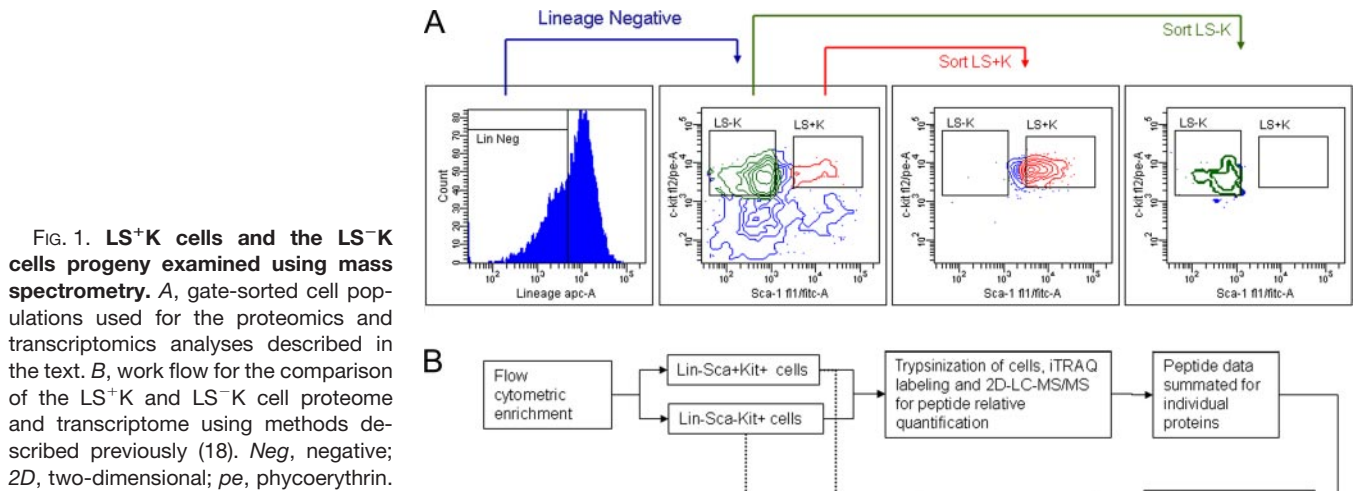


FIG. 1. LS⁺K cells and the LS⁻K cells progeny examined using mass spectrometry. *A*, gate-sorted cell populations used for the proteomics and transcriptomics analyses described in the text. *B*, work flow for the comparison of the LS⁺K and LS⁻K cell proteome and transcriptome using methods described previously (18). *Neg*, negative; *2D*, two-dimensional; *pe*, phycoerythrin.

days the wells were scored for the presence of erythroid and megakaryocytic cells. A cluster of eight or more erythroid cells was defined as a colony-forming unit-erythroid (CFU-e). The incidence of progenitor cells was determined from the number of colonies formed using L-Calc software (StemSoft Software Inc.). The erythroid nature of cells in this assay was confirmed by flow cytometric analysis after Ter119 staining.

RESULTS

LS⁻K Cells Have No Long Term Reconstitution Potential—Populations of LS⁺K and LS⁻K cells from murine bone marrow were isolated by means of flow cytometry according to the gates shown in Fig. 1A. LS⁺K cells constituted $0.017 \pm 0.003\%$ (mean \pm S.E., $n = 3$) of the nucleated bone marrow cells, and LS⁻K cells constituted $0.44 \pm 0.07\%$ (mean \pm S.E., $n = 3$). To confirm their long or short term repopulation potential, these populations were assessed for the ability to reconstitute busulfan-myeloablated animals (21). After 6 weeks, no LS⁻K cell-derived cells were detected in the bone marrow of recipient animals transplanted with 30,000, 15,000, 3000, or 300 LS-K cells plus 300 LS⁺K cells. Similarly 30,000 LS⁻K cells transplanted in the absence of LS⁺K cells gave less than 0.5% donor cells in recipient animals after 6 weeks. As shown previously, transplantation of 300–10,000 LS⁺K cells was effective in re-establishing hematopoiesis and enabling recipient mice to survive busulfan myeloablation (21). The lack of the Sca-1 marker therefore indicated that this is a more mature cell population as expected (3). However, this observation provides no insights into the composition, cellular identity, or differentiation potential of the LS⁻K population. Thus, the functional and phenotypic significance of the loss of Sca-1 expression was explored.

Proteome Versus Transcriptome Analysis of LS⁺K Cells and

LS⁻K Cells—The ability to perform relative quantification proteomics analyses on flow cytometrically enriched primitive hematopoietic cell populations offers opportunities for analysis of transcription changes *versus* proteome changes. LS⁺K cells have a more primitive phenotype than LS⁻K cells using the above assays yet are poorly defined in terms of developmental potential. Transcriptome and proteome analyses were performed to determine the potential cellular fate of LS⁻K cells and identify whether post-translational regulation of protein expression is observed in LS⁺K to LS⁻K transition.

Isobaric tagging of tryptic digests from cell populations followed by two-dimensional liquid chromatography and tandem mass spectrometry was applied to LS⁺K and LS⁻K cells (see Fig. 1B for workflow). In the first instance a high level of reproducibility in this technique was demonstrated by comparing two separate preparations of LS⁻K cells with each other (Fig. 2A). Different preparations of LS⁻K cells showed -fold differences in expression between 0.8 and 1.2 for the vast majority of proteins. The same experiment performed for two separate preparations of LS⁺K cells also showed marked similarity between two preparations of these cells (Fig. 2B). Thus, the reproducibility of this approach for biological replicates of sorted primitive hematopoietic cell populations was confirmed, and any differences lying outside this 0.8–1.2-fold difference boundary in comparisons of two distinct populations can be considered potentially significant. Three such comparisons were then performed between LS⁻K and LS⁺K cells. A representative experiment is shown in Fig. 2C. It can be seen that there are profound differences occurring that the control experiments discussed above demonstrated are not

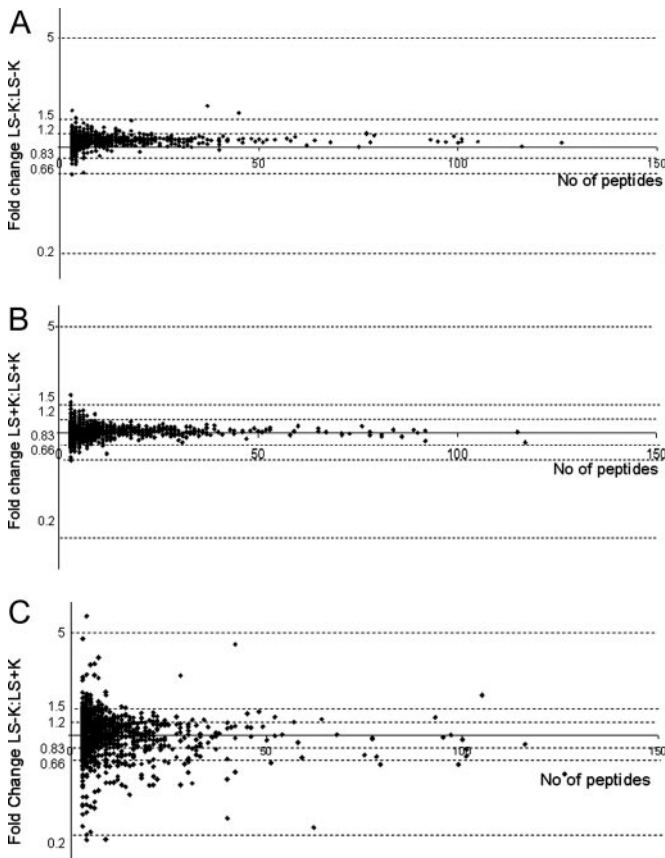


FIG. 2. Proteomics relative quantification of biological replicates for LS^+K and LS^-K populations and a comparison of LS^+K cells versus LS^-K cells. A and B, representative plotted ratios from proteins using the internal duplicates LS^-K versus LS^-K and LS^+K versus LS^+K cells, respectively, from one fourplex experiment. C, protein ratios for one of the $LS^-K:LS^+K$ comparisons, demonstrating differences in the protein expression profiles between the two cell types. The x axis is the number of peptides identified for a single protein. Each diamond shown is one protein.

due to biological or experimental variability. Where significant differences were seen, MS/MS spectra were manually checked to validate sequence assignment and reporter group relative quantification. An example of the raw MS data from some key protein changes are provided in supplemental Table 2 with appropriate example spectra in supplemental Fig. 1.

LS^+K and LS^-K cells have profoundly different proteomes even though the populations are separated only on this single marker. In total relative quantification on over 1263 proteins (with >3 peptides quantified) was achieved with starting material of the order of 90 μ g of protein. There were 96 proteins showing a 1.2-fold or greater increase and 121 proteins with a 0.83-fold or greater decrease in expression. The total number of proteins found in a single analysis only yet meeting the above criteria was 20 up-regulated and nine down-regulated proteins. A full list of proteins detected including those showing a significant change is shown in supplemental Table 1.

The next question we addressed was whether the LS^+K to

TABLE I
Correlation of transcriptomic and proteomic comparisons of LS^+K and LS^-K cells

Shown are relative changes in mRNA levels for all of the proteins shown to be differentially expressed by iTRAQ labeling and for proteins shown to have no change between the two cell types where more than three peptides were identified.

Protein (iTRAQ)	mRNA (Affymetrix array)		
	Up	No change	Down
Up	32	40	0
No change	38	833	15
Down	1	62	17

LS^-K transition at the proteome level is congruent with changes seen at the transcriptome level. We performed a comparison of changes in the transcriptome and the proteome. For this analysis an mRNA probe set had to be called present in all replicates to be included in the analysis. Furthermore a match between the accession numbers for the mRNA and protein data sets had to be found. Where no mRNA data were available, the protein was omitted from this comparison. This limited the set analyzed to 1038 gene products. Of these gene products in the study there was a 14.6% change at the protein level and 9.9% change at the transcriptome level (Table I). Of the protein changes, 32% correlated to a change at the mRNA level. Similarly of all the mRNA changes in expression observed only 48% correlated to a change in protein level. Clearly the sensitivity of the two approaches has some role to play with iTRAQ detecting 1.2-fold changes and mRNA arrays detecting 1.8-fold, and a small number of positive matches may be lost due to the methods used for quantification. To examine the picture of relative intensities in the change of protein levels to transcriptome levels and vice versa, a ranked order heat map of changes in protein level was plotted for each of the genes where probe sets had been called present, and the concomitant relative change at mRNA level is shown on an aligned heat map for each gene product. This demonstrated that the relative change in protein level also bears no correlation to -fold change in transcriptome level. High level changes in protein do not correlate more markedly to high level changes in mRNA level. Thus there is no qualitative similarity in protein/mRNA changes in the two populations as no specific trend can be observed in mRNA level changes compared with the proteome data set when all proteins with matching mRNA data are compared (Fig. 3A). Where protein or mRNA changes are ranked and compared with mRNA changes again no clear trend can be observed between the enriched stem cell population and the population believed to be enriched in progenitor cells (Fig. 3B).

Previously we examined the proteome of LS^+Kit^+ cells and LS^+Kit^- cells. The latter population has no colony forming potential in soft gel assays and as such probably does not represent a progenitor of LS^+K^+ cells in development. How-

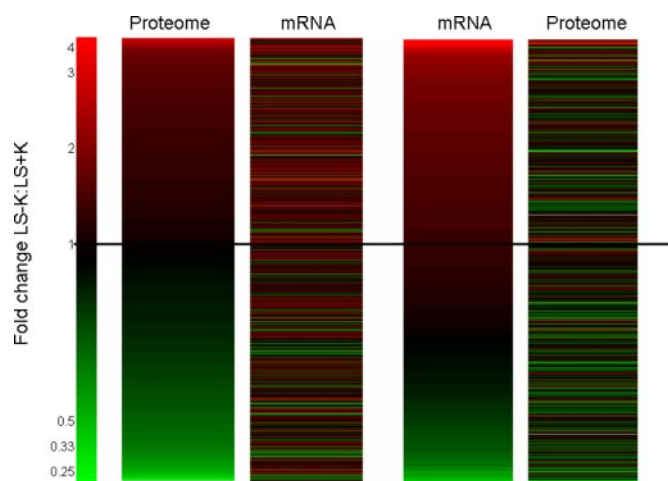


FIG. 3. Expression profiles of $LS^{-}K$ cells versus $LS^{+}K$ cells. Data for all proteins with a matching mRNA data set are shown as the -fold change in each parameter. The horizontal line shown indicates where an identical change in mRNA level and protein level would fall. Data are shown as color-coded -fold changes in protein levels in rank order between $LS^{+}K^{-}$ and $LS^{-}K^{-}$ cells alongside corresponding expression ratios for transcripts of these proteins obtained by microarray analysis. The right-hand panel shows the same expression data sorted in rank order of -fold change in the mRNA level alongside corresponding protein expression ratios.

ever, the $LS^{-}K$ cells do have colony forming potential, and it is more likely that the $LS^{-}K$ cells are in hierarchical terms downstream of the $LS^{+}K$ with functional potential to develop into mature cells. Nonetheless the generation of these two data sets has value in that it allows identification of some of the proteins that are significantly altered in expression levels between $LS^{+}K^{+}$ cells when compared with both $LS^{-}K$ and $LS^{+}K^{-}$ cells, respectively (Fig. 4). Using the full proteome data set, the common decreases total 22; 29% of the decreases associated with the $LS^{+}K/LS^{-}K$ transition were seen in the LSK^{+} to LSK^{-} transition. Only five (11%) of increases seen in the $LS^{+}K/LS^{-}K$ transition were seen in the differences between LSK/LSK^{-} populations. The common changes may give indications of proteomic changes associated with loss of stem cell phenotype.

We found a common decrease in aldehyde dehydrogenase, a known stem cell marker (23). The other proteins in this category are shown on Fig. 4 and include 14-3-3 signaling proteins and metabolic enzymes such as glyceraldehyde-3-phosphate dehydrogenase, isocitrate dehydrogenase, and malate dehydrogenase. Given their relevance in transcription factor-mediated regulatory processes, the observed decrease in heterogeneous nuclear ribonucleoproteins (hnRNPs) may have some significance in the development of hematopoietic cells (see below).

Cellular Proteome Analysis Reveals Post-translational Regulation in the $Sca-1^{+}$ to $Sca-1^{-}$ Cell Transition—Given the profound differences in the proteome and transcriptome changes, respectively, we concentrated on analysis of proteome changes to discern the potential cellular phenotype and developmental fate of $LS^{-}K$ cells.

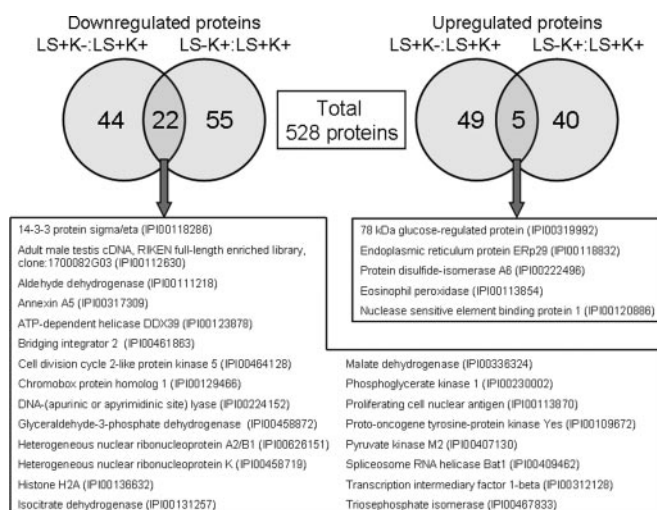


FIG. 4. Common changes in the expression of proteins between $LS^{+}K^{+}$ cells and $LS^{-}K^{+}$ cells and $LS^{+}K^{-}$ cells and $LS^{-}K^{-}$ cells, respectively. Two relative quantification studies on proteomic changes in primitive hematopoietic cells as they change their surface expression markers were performed. Approximately 600 proteins with >3 peptides quantified in each study could be matched between the two experiments. Common changes are shown.

The penetrative power of the technique described was demonstrated by the relative quantification of c-Kit. This was expressed at an $LS^{-}K$ to $LS^{+}K$ ratio of 0.58 and 0.82 in two experiments at the single peptide level, representing the relative total protein ratio in the Sca^{-} and Sca^{+} cell populations (whereas flow cytometric analysis is based on surface expression).

The third highest increase we saw in $LS^{-}K$ cells was in a protein similar to carboxylesterase 1. This protein has a role in drug metabolism but also in retention and traffic control of proteins within the endoplasmic reticulum (24). Thus, there may be increased flux through the protein synthetic machinery in $LS^{-}K$ cells. Fig. 5A shows that ribosomal protein expression was increased as was expression of several eukaryotic translation initiation factors. Once again there were several proteins that showed a marked increase in protein level where there was little change in the transcriptome expression. Thus, there was an element of post-translational regulation observed with respect to changes in the translational machinery in early hematopoietic cells. Many of the changes observed are among the largest seen (ribosomal proteins L1, L37a, L35, and 11-kDa protein). The high molar expression level of ribosomal protein means that this represents a major difference in the proteins synthesized between $LS^{-}K$ and $LS^{+}K$ cells.

The evidence above suggests there is a strong element of post-translational regulation as $LS^{+}K$ cells progress to $LS^{-}K$ cells. hnRNPs have a role in pre-mRNA binding, mRNA formation, export, and degradation (25). In hematopoietic cells dysregulation of hnRNPs results in altered transcription factor expression (26, 27). $LS^{-}K$ cells displayed decreased expression of four hnRNPs or associated proteins as compared with

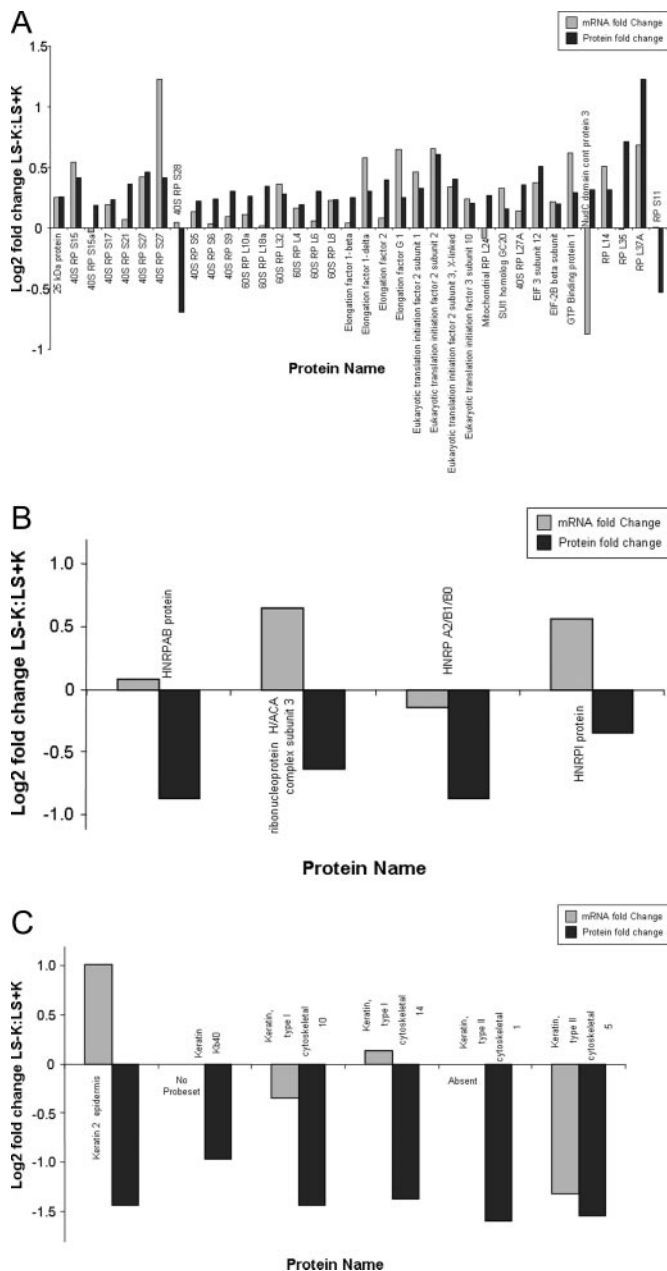


FIG. 5. Differential protein and mRNA expression between LS⁺K stem cells and their LS⁻K progeny. A, protein synthesis: differences observed in protein synthetic machinery gene products at the protein and mRNA level are shown. *eIF*, eukaryotic initiation factor; *RP*, ribosomal protein. Details of protein product accession number are provided in supplemental Table 1. Mitochondrial ribosomal proteins L24 and L48, elongation factor G1, and eIF2B δ subunit are included here for completeness but not included in Table 1 as only two peptides/protein were found (Table 1 includes only proteins where >3 peptides/protein were found). B, hnRNP differences observed in hnRNP gene products at the protein and mRNA levels are shown.

LS⁺K cells (Fig. 5B), and hnRNPs were seen to be differentially expressed at the protein level but not transcriptome level in LSK cells compared with LSK⁻ cells (see Fig. 4). The regulation of hnRNP protein expression is at least in part

post-translational in the LSK to LS⁻K transition and may be of significance in development of primitive hematopoietic cells.

The Population Enriched for Hematopoietic (LS⁺K) Stem Cells Expresses Mature Cell Marker Proteins—PCR analysis has shown that both erythroid and myeloid gene expression occurs in the same progenitor cell prior to commitment to a specific developmental fate (28). Whether the proteins are also expressed has not been addressed. Our studies showed that LS⁺K cells express proteins associated with a mature cell phenotype, such as neutrophil elastase. Other mature myeloid cell markers found in LS⁺K cells that increased in LS⁻K cells include myeloblastin (proteinase 3) and myeloperoxidase (Fig. 6A). Thus, cells enriched for their long term reconstitution potential are indeed able to express the protein associated with mature cell function. Lymphocyte-specific protein 1 (expressed in macrophages, lymphocytes, and neutrophils (29)) was also found to be expressed in LS⁺K and LS⁻K cells (a single peptide was detected at a ratio of 2.1:1), indicating a potential expression of proteins believed to be associated with lymphoid cells in LS⁺K cells at the protein as well as mRNA level.

Noteworthy in the case of these proteins associated with metabolism and expression of mature cell function is the concordance between microarray changes between LS⁺K and LS⁻K cells. This set of proteins does appear to be under transcriptional regulation. The exception to this is the hemoglobin β subunit, a potentially toxic protein that decreased in expression between LS⁺K cells and LS⁻K cells.

The quantitative protein changes that were most marked in LS⁻K cells compared with LS⁺K cells were assessed to give an indication of the cellular fate that LS⁻K may undergo (Fig. 6A). Proteins associated with the erythroid lineage and synthesis of hemoglobin and porphyrins were the major set to be elevated in the LS⁺K to LS⁻K transition: flavin reductase is an enzyme that has been proposed to store heme (30) and reduce methemoglobin and was increased 2.9-fold in Sca⁻ cells as compared with Sca⁺. Coproporphyrinogen oxidase is one of the last three enzymes in the heme synthesis pathway (31, 32), and it showed a 1.7-fold increase in expression. Knockdown of this enzyme in zebrafish leads to a suppression of hemoglobin production without affecting blood cell production (32).

α -Hemoglobin-stabilizing protein binds to hemoglobin, limiting its intracellular toxicity (33). A marked increase in this protein was observed, 4.5-fold, one of the biggest changes (Fig. 5A). This may be related to the decrease in hemoglobin β subunit protein levels despite increased mRNA levels for this gene product. Carbonic anhydrase 1 is a marker for early erythroid cells and is involved in erythrocyte function (34). It too showed a marked elevation in the LS⁻K cells compared with LS⁺K cells (4.3-fold for carbonic anhydrase 1 and 1.2-fold for carbonic anhydrase 2). These proteins represent the major quantitative proteomic changes observed. As we have shown, LS⁻K cells express proteins to engage in increased protein synthesis and express some proteins associated with red blood cell development in

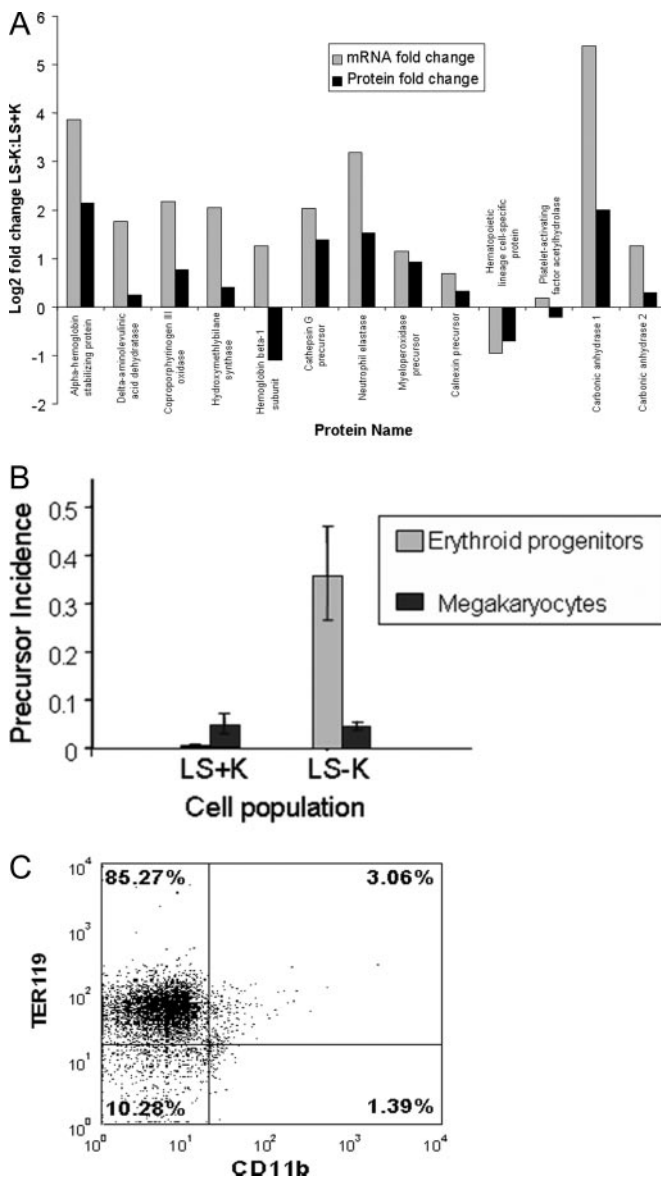


FIG. 6. The hematopoietic phenotype of LS⁺K cells and LS⁻K cells. Changes in selected proteins associated with hematopoietic cell fate and determination of cellular potential are shown. *A*, changes observed in proteins associated with maturation of hematopoietic cells. Full details (statistical analysis and accession number) are provided in supplemental Table 2. *B*, the erythroid and megakaryocytic potential of LS⁺K and LS⁻K cells assessed by colony formation as described under “Experimental Procedures.” Results shown are the mean of three experiments, and *error bars* are S.E. *C*, expression of a myeloid marker (CD11b) and an erythroid marker (Ter119) in LSK cells cultured for 5 days in a combination of growth factors that can promote erythroid and myeloid cell growth.

abundance. Therefore, we examined the erythroid potential of the LS⁻K cell population.

Proteomic Indication of Biological Potential of LS⁻K Cells—LS⁺K cells have long term reconstitution potential, whereas LS⁻K cells do not. To further assess differences between the two populations we used the HPP-CFC and LPP-CFC assays.

The LS⁺K cells showed a $9.3 \pm 2.2\%$ plating efficiency in the HPP-CFC assay, whereas the LS⁻K cells exhibited an efficiency of $2.6 \pm 0.2\%$ (mean \pm S.E., $n = 3$). The LPP-CFC assay (for more mature progenitors) yielded different data in that LS⁺K cells had a lower plating efficiency (4.0 ± 0.6) compared with LS⁻K (17.2 ± 1.2 , mean \pm S.E.) cells. To determine the myeloid potential of the cell populations they were grown in soft gel assays with granulocyte/macrophage colony-stimulating factor. LS⁻K cells produced $65 \pm 9\%$ (mean \pm S.E., $n = 3$) of the myeloid progenitors seen with LS⁺K cells. This implies that cells with neutrophil/macrophage potential do not represent an increased proportion of the LS⁻K population compared with LS⁺K cells.

The proteomics data give an indication that cells with erythroid potential are present in greater numbers in LS⁻K and not LS⁺K cells. We therefore assessed the erythroid potential of these populations. Limiting dilution analysis of the LS⁺K and LS⁻K cell populations showed that LS⁻K cells form clonogenic erythroid colonies at an incidence of 1 in 2. The LS⁺K population in the same assay procedures had a 1 in 164 incidence (0.6%) of cells with erythroid colony forming potential (Fig. 6B). This suggests that the LS⁻K cell population is enriched for erythroid progenitors, a hypothesis that is supported by the protein expression profile of these cells. When cultured under conditions where myeloid and erythroid progenitor cell proliferation and development were supported, the majority of cells adopted an erythroid fate as determined by expression on CD11b and Ter119 markers for myeloid and erythroid development, respectively (Fig. 6C).

DISCUSSION

We have demonstrated using proteomics and microarray analysis that there is a major role for post-translational regulation in developing hematopoietic cells. LSK cells are highly enriched for long term reconstituting cells, and as such we developed a proteomic signature of a stem cell-containing population. This systematic analysis of the very small quantities of primitive hematopoietic stem cells available is feasible because of the deployment of isobaric tag-mediated relative quantification by tandem mass spectrometry on highly fractionated tryptic lysates of flow cytometrically enriched cells. The use of subcellular fractionation and the increased sensitivity of new mass spectrometry instrumentation means that stem cell proteomic signatures are now a realistic proposition. The loss of the well defined stem cell marker aldehyde dehydrogenase (23) was validated using this approach. A decrease in expression in LS⁻K cells (Sca⁻:Sca⁺ ratio of 0.49 ± 0.04 ; mean \pm S.E.; data derived from 52 spectra) was observed. This gives confidence that changes in other proteins such as the early development regulator protein (EDR2 or homolog of polyhomeotic), seen as a single peptide level in two experiments only with a LS⁻K:LS⁺K expression ratio of 3.5 and 4.7, one of the major changes seen, is significant. EDR2 associates with BMI1 (35). BMI1 is required for the self-renewal of

hematopoietic stem cells (36). Thus changes in the level of BMI1 binding partners may have a role in the differential long term reconstitution abilities of LS⁺K and LS⁻K cells.

Evidence from the systematic proteomics analysis suggests an LS⁺K to LS⁻K transition that is driven by a post-translationally regulated gearing up for increased protein synthesis (Fig. 5). Eukaryotic IF5 protein was elevated 1.9-fold in LS⁻K cells; this protein binds to eIF2 (whose subunits were also increased in expression; see Fig. 5). These proteins are part of a complex with eIF3 (also increased) ensuring stringent selection of the initiation codon during the scanning process (37). Ribosomal protein levels were also increased. We conclude that LS⁻K cells have enhanced protein biosynthetic capacity. Such changes have been observed previously in proteomics studies on developing B lymphocytes preparing to engage in immunoglobulin production (38).

hnRNPs are post-translationally regulated in the switch from LSK cells to LS⁻K cells. hnRNPs are required for normal myeloid cell development (39). The developmental changes in hnRNPs A/B that have been observed during erythropoiesis can affect physiologically important processes (40). These proteins were decreased at the proteome level in the LSK to LS⁻K transition but not at the transcriptome level. This led to the speculation that hnRNPs and their post-translational regulation are a key element in early hematopoietic development. Given the parallel studies performed using microarray analyses, which indicated no such change in hnRNP levels, it is clear that proteomics analysis has value for systematic studies leading to development of hypotheses concerning stem cell regulation.

The question of what kind of cells the LS⁺K to LS⁻K transition forms in developmental terms has also been answered using proteomics approaches. The major increases in proteins expressed were associated with erythroid development. This led to the hypothesis that erythroid progenitor cells predominate in the population of LS⁻K cells, and this was proven with further experimentation. Paradoxically the β -hemoglobin protein fell in expression when LS⁺K cells were compared with LS⁻K cells (showing about a 2-fold decrease). Given the potentially toxic nature of hemoglobin this likely represents the inception of a control mechanism in the early progenitors to limit potential damage as cells commit to major hemoglobin synthesis. This is in keeping with the major elevation in α -hemoglobin-stabilizing protein. Thus, the potentially cytotoxic effects of $\alpha\beta$ -hemoglobin in early progenitor cells are dealt with by both generation of a binding protein and decreased protein expression.

One of the critical features of our study is that Ter119 erythroid marker-negative LS⁻K cells had a prolific ability to form erythroid cells. In adult mice, Ter119 reacted with 20–25% of bone marrow cells, and in fetal hematopoietic tissues, Ter119 reacted with 80–90% of day 14 fetal liver cells (41). Ter119 recognizes erythroid cells at differentiation stages from early proerythroblast to mature erythrocyte. That the LS⁻K cells are negative for Ter119 indicates that this population lacks late erythroid cells, yet they express intracellular

markers and enzymes associated with erythroid development. We compared the reported marrow cellularity of populations enriched for common myeloid progenitor cells and CFU-e (16, 42). Lin⁻Kit⁺Sca⁻IL-7R α ⁻IL-3R α ⁻CD41⁻CD71⁺ cells (a population enriched in CFU-e) constitute 0.41% of nucleated bone marrow cells (42), whereas the LS⁻K population we sorted comprised 0.44% of bone marrow cells. The common myeloid progenitor population (IL-7R α ⁻, Fc γ R^{low}, CD34⁺, Lin⁻, Sca⁻, Kit⁺ cells) (16) represents 0.1% of total bone marrow. However interlaboratory variation in these figures depending on among other things flow cytometry settings can be anticipated, and this is exemplified in the fact that Lin⁻IL-7R α ⁻Sca⁺Kit⁺ cells constituted 0.04% of whole bone marrow cells (16) and represented a highly enriched stem cell population, whereas our LS⁺K cells (not sorted for IL-7R α negativity) were only a 0.017% fraction of whole bone marrow.

Taken together these data demonstrate that comparative proteome analysis of hematopoietic cell subpopulations can be used to derive a cell type-specific “fingerprint.” This fingerprint enables generation of hypotheses, and further tests can then be performed to define the population directly using cell biology assays. Given the sensitivity of mass spectrometric techniques (which is ever increasing), small populations of cells (~0.5–1 million presently) can be used for progenitor cell/flow cytometry-sorted cell fingerprinting. Proteomics defines the LS⁻K population compared with the LS⁺K cells as committing to erythroid progenitors plus some myeloid progenitors; this was confirmed with cell biology assays and entirely in line with the marrow cellularity data above. The importance of this advance may in part lie in the ability to define the progeny of embryonic stem cells before they reach a fully differentiated phenotype.

* This work was supported by the Leukemia Research Fund (UK), The Biological Sciences and Biotechnology Research Council (UK), and the Australian Stem Cell Centre. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ The on-line version of this article (available at <http://www.mcponline.org>) contains supplemental material.

¶ Present address: Stem Cell Research Center, Brown Foundation Inst. of Molecular Medicine, University of Texas Health Science Center at Houston, 1825 Pressler St., Houston, TX 77030.

|| Present address: Adult Stem Cell Program, Australian Stem Cell Centre, P. O. Box 8002, Monash University L. P. O., Victoria 3168, Australia.

** To whom correspondence should be addressed. Tel.: 44-161-446-8247; Fax: 44-161-446-8203; E-mail: anthony.whetton@manchester.ac.uk.

REFERENCES

- Christensen, J. L., and Weissman, I. L. (2001) Flk-2 is a marker in hematopoietic stem cell differentiation: a simple method to isolate long-term stem cells. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 14541–14546
- Goodell, M. A., Brose, K., Paradis, G., Conner, A. S., and Mulligan, R. C. (1996) Isolation and functional properties of murine hematopoietic stem cells that are replicating in vivo. *J. Exp. Med.* **183**, 1797–1806
- Okada, S., Nakauchi, H., Nagayoshi, K., Nishikawa, S. I., Miura, Y., and

- Suda, T. (1992) In vivo and in vitro stem-cell function of c-Kit-positive and Sca-1-positive murine hematopoietic-cells. *Blood* **80**, 3044–3050
4. Osawa, M., Hanada, K., Hamada, H., and Nakauchi, H. (1996) Long-term lymphohematopoietic reconstitution by a single CD34⁺ low/negative hematopoietic stem cell. *Science* **273**, 242–245
 5. Osawa, M., Nakamura, K., Nishi, N., Takahashi, N., Tokuomoto, Y., Inoue, H., and Nakauchi, H. (1996) In vivo self-renewal of c-Kit(+) Sca-1(+) Lin(low/-) hemopoietic stem cells. *J. Immunol.* **156**, 3207–3214
 6. Randall, T. D., Lund, F. E., Howard, M. C., and Weissman, I. L. (1996) Expression of murine CD38 defines a population of long-term reconstituting hematopoietic stem cells. *Blood* **87**, 4057–4067
 7. Spangrude, G. J., Heimfeld, S., and Weissman, I. L. (1988) Purification and characterization of mouse hematopoietic stem-cells. *Science* **241**, 58–62
 8. Bertoncello, I., Bradley, T. R., and Watt, S. M. (1991) An improved negative immunomagnetic selection strategy for the purification of primitive hematopoietic-cells from normal bone-marrow. *Exp. Hematol.* **19**, 95–100
 9. Huang, E., Nocka, K., Beier, D. R., Chu, T. Y., Buck, J., Lahm, H. W., Wellner, D., Leder, P., and Besmer, P. (1990) The hematopoietic growth factor-KL is encoded by the Si-locus and is the ligand of the c-Kit receptor, the gene-product of the W-locus. *Cell* **63**, 225–233
 10. Zsebo, K. M., Williams, D. A., Geissler, E. N., Broudy, V. C., Martin, F. H., Atkins, H. L., Hsu, R. Y., Birkett, N. C., Okino, K. H., Murdock, D. C., Jacobsen, F. W., Langley, K. E., Smith, K. A., Takeishi, T., Cattaneach, B. M., Galli, S. J., and Suggs, S. V. (1990) Stem-cell factor is encoded at the Si-Locus of the mouse and is the ligand for the c-Kit tyrosine kinase receptor. *Cell* **63**, 213–224
 11. Horejsi, V., Drbal, K., Cebeceauer, M., Cerny, J., Brdicka, T., Angelisova, P., and Stockinger, H. (1999) GPI-microdomains: a role in signalling via immunoreceptors. *Immunol. Today* **20**, 356–361
 12. Epting, C. L., Lopez, J. E., Shen, X., Liu, L. S., Bristow, J., and Bernstein, H. S. (2004) Stem cell antigen-1 is necessary for cell-cycle withdrawal and myoblast differentiation in C2C12 cells. *J. Cell Sci.* **117**, 6185–6195
 13. Bradfute, S. B., Graubert, T. A., and Goodell, M. A. (2005) Roles of Sca-1 in hematopoietic stem/progenitor cell function. *Exp. Hematol.* **33**, 836–843
 14. Stefanova, I., Horejsi, V., Anstegui, I. J., Knapp, W., and Stockinger, H. (1991) GPI-anchored cell-surface molecules complexed to protein tyrosine kinases. *Science* **254**, 1016–1019
 15. Ito, C. Y., Li, C. Y. J., Bernstein, A., Dick, J. E., and Stanford, W. L. (2003) Hematopoietic stem cell and progenitor defects in Sca-1/Ly-6A-null mice. *Blood* **101**, 517–523
 16. Akashi, K., Traver, D., Miyamoto, T., and Weissman, I. L. (2000) A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. *Nature* **404**, 193–197
 17. Unwin, R. D., Pierce, A., Watson, R. B., Sternberg, D. W., and Whetton, A. D. (2005) Quantitative proteomic analysis using isobaric protein tags enables rapid comparison of changes in transcript and protein levels in transformed cells. *Mol. Cell. Proteomics* **4**, 924–935
 18. Unwin, R. D., Smith, D. L., Blinco, D., Wilson, C. L., Miller, C. J., Evans, C. A., Jaworska, E., Baldwin, S. A., Barnes, K., Pierce, A., Spooncer, E., and Whetton, A. D. (2006) Quantitative proteomics reveals posttranslational control as a regulatory factor in primary hematopoietic stem cells. *Blood* **107**, 4687–4694
 19. Nilsson, S. K., Johnston, H. M., and Coverdale, J. A. (2001) Spatial localization of transplanted hemopoietic stem cells: inferences for the localization of stem cell niches. *Blood* **97**, 2293–2299
 20. Peters, S. O., Bauermeister, K., Simon, J. P., Branke, B., and Wagner, T. (2002) Quantitative polymerase chain reaction-based assay with fluorogenic Y-chromosome specific probes to measure bone marrow chimerism in mice. *J. Immunol. Methods* **260**, 109–116
 21. Nilsson, S. K., Johnston, H. M., Whitty, G. A., Williams, B., Webb, R. J., Denhardt, D. T., Bertoncello, I., Bendall, L. J., Simmons, P. J., and Haylock, D. N. (2005) Osteopontin, a key component of the hematopoietic stem cell niche and regulator of primitive hematopoietic progenitor cells. *Blood* **106**, 1232–1239
 22. Bartelmez, S. H., Bradley, T. R., Bertoncello, I., Mochizuki, D. Y., Tushinski, R. J., Stanley, E. R., Hapel, A. J., Young, I. G., Kriegl, A. B., and Hodgson, G. S. (1989) Interleukin-1 plus Interleukin-3 plus colony-stimulating factor-1 are essential for clonal proliferation of primitive myeloid bone-marrow cells. *Exp. Hematol.* **17**, 240–245
 23. Storms, R. W., Trujillo, A. P., Springer, J. B., Shah, L., Colvin, O. M., Ludeman, S. M., and Smith, C. (1999) Isolation of primitive human hematopoietic progenitors on the basis of aldehyde dehydrogenase activity. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 9118–9123
 24. Redinbo, M. R., Bencharit, S., and Potter, P. M. (2003) Human carboxylesterase 1: from drug metabolism to drug discovery. *Biochem. Soc. Trans.* **31**, 620–624
 25. Fahling, M., Mrowka, R., Steege, A., Martinka, P., Persson, P. B., and Thiele, B. J. (2006) Heterogeneous nuclear ribonucleoprotein-A2/B1 modulate collagen prolyl 4-hydroxylase, α (I) mRNA stability. *J. Biol. Chem.* **281**, 9279–9286
 26. Notari, M., Neviani, P., Santhanam, R., Blaser, B. W., Chang, J. S., Galletta, A., Willis, A. E., Roy, D. C., Caligiuri, M. A., Marcucci, G., and Perrotti, D. (2006) A MAPK/HNRPK pathway controls BCR/ABL oncogenic potential by regulating MYC mRNA translation. *Blood* **107**, 2507–2516
 27. Perrotti, D., Cesi, V., Trotta, R., Guerzoni, C., Santilli, G., Campbell, K., Iervolino, A., Condorelli, F., Gambacorti-Passerini, C., Caligiuri, M. A., and Calabretta, B. (2002) BCR-ABL suppresses C/EBP α expression through inhibitory action of hnRNP E2. *Nat. Genet.* **30**, 48–58
 28. Hu, M., Krause, D., Greaves, M., Sharkis, S., Dexter, M., Heyworth, C., and Enver, T. Multilineage gene expression precedes commitment in the hemopoietic system. *Genes Dev.* **11**, 774–785, 1997
 29. Jongstra, J., Ittel, M. E., Iscove, N. N., and Brady, G. (1994) The Lsp1 gene is expressed in cultured normal and transformed mouse macrophages. *Mol. Immunol.* **31**, 1125–1131
 30. Chikuba, K., Yubisui, T., Shirabe, K., and Takeshita, M. (1994) Cloning and nucleotide-sequence of a cDNA of the human erythrocyte NADPH-flavin reductase. *Biochem. Biophys. Res. Commun.* **198**, 1170–1176
 31. Jones, M. A., Hamilton, M. L., and Lash, T. D. (1997) Effect of covalent modification on coproporphyrinogen oxidase from chicken red blood cells. *Prep. Biochem. Biotechnol.* **27**, 47–57
 32. Hanaoka, R., Katayama, S., Dawid, I. B., and Kawahara, A. (2006) Characterization of the heme synthesis enzyme coproporphyrinogen oxidase (CPO) in zebrafish erythropoiesis. *Genes Cells* **11**, 293–303
 33. Weiss, M. J., Zhou, S. P., Feng, L., Gell, D. A., MacKay, J. P., Shi, Y. G., and Gow, A. J. (2005) Role of α hemoglobin-stabilizing protein in normal erythropoiesis and β -thalassemia. *Ann. NY Acad. Sci.* **1054**, 103–117
 34. Villeval, J. L., Testa, U., Vinci, G., Tonthat, H., Betteieb, A., Titeux, M., Cramer, P., Edelman, L., Rochant, H., Bretongorius, J., and Vainchenker, W. (1985) Carbonic anhydrase-I is an early specific marker of normal human erythroid-differentiation. *Blood* **66**, 1162–1170
 35. Gunster, M. J., Satijn, D. P. E., Hamer, K. M., denBlaauwen, J. L., deBruijn, D., Alkema, M. J., vanLohuizen, M., vanDriel, R., and Otte, A. P. (1997) Identification and characterization of interactions between the vertebrate polycomb-group protein BMI1 and human homologs of polyhomeotic. *Mol. Cell. Biol.* **17**, 2326–2335
 36. Park, I. K., Qian, D. L., Kiel, M., Becker, M. W., Pihajla, M., Weissman, I. L., Morrison, S. J., and Clarke, M. F. (2003) Bmi-1 is required for maintenance of adult self-renewing haematopoietic stem cells. *Nature* **423**, 302–305
 37. Huang, H. K., Yoon, H., Hannig, E. M., and Donahue, T. F. (1997) GTP hydrolysis controls stringent selection of the AUG start codon during translation initiation in *Saccharomyces cerevisiae*. *Genes Dev.* **11**, 2396–2413
 38. van Anken, E., Romijn, E.P., Maggioni, C., Mezghrani, A., Sitia, R., Braakman, I., and Heck, A. J. R. (2003) Sequential waves of functionally related proteins are expressed when B cells prepare for antibody secretion. *Immunity* **18**, 243–253
 39. Iervolino, A., Santilli, G., Trotta, R., Guerzoni, C., Cesi, V., Bergamaschi, A., Gambacorti-Passerini, C., Calabretta, B., and Perrotti, D. (2002) hnRNP A1 nucleocytoplasmic shuttling activity is required for normal myelopoiesis and BCR/ABL leukemogenesis. *Mol. Cell. Biol.* **22**, 2255–2266
 40. Hou, V. C., Lersch, R., Gee, S. L., Pontthier, J. L., Lo, A. J., Wu, M., Turck, C. W., Koury, M., Krainer, A. R., Mayeda, A., and Conboy, J. G. (2002) Decrease in hnRNP A/B expression during erythropoiesis mediates a pre-mRNA splicing switch. *EMBO J.* **21**, 6195–6204
 41. Kina, T., Ikuta, K., Takayama, E., Wada, K., Majumdar, A. S., Weissman, I. L., and Katsura, Y. (2000) The monoclonal antibody TER-119 recognizes a molecule associated with glycophorin A and specifically marks the late stages of murine erythroid lineage. *Br. J. Haematol.* **109**, 280–287
 42. Terszowski, G., Waskow, C., Conradt, P., Lenze, D., Koenigsmann, J., Carstanjen, D., Horak, I., and Rodewald, H.-R. (2005) Prospective isolation and global gene expression analysis of the erythrocyte colony-forming unit (CFU-E). *Blood* **105**, 1937–1945