Physiol Genomics 28: 114–128, 2006.
First published August 29, 2006; doi:10.1152/physiolgenomics.00055.2006.

Transcriptional regulatory network analysis of developing human erythroid progenitors reveals patterns of coregulation and potential transcriptional regulators

M. A. Keller,[1] S. Addya,[1] R. Vadigepalli,[2] B. Banini,[1] K. Delgrosso,[1] H. Huang,[1] and S. Surrey[1]

[1]*Cardeza Foundation of Hematologic Research and* [2]*Daniel Baugh Institute for Functional Genomics and Computational Biology, Department of Pathology, Jefferson Medical College, Philadelphia, Pennsylvania*

Submitted 31 March 2006; accepted in final form 8 August 2006

Keller MA, Addya S, Vadigepalli R, Banini B, Delgrosso K, Huang H, Surrey S. Transcriptional regulatory network analysis of developing human erythroid progenitors reveals patterns of coregulation and potential transcriptional regulators. *Physiol Genomics* 28: 114–128, 2006. First published August 29, 2006; doi:10.1152/physiolgenomics.00055.2006.—Deciphering the molecular basis for human erythropoiesis should yield information benefiting studies of the hemoglobinopathies and other erythroid disorders. We used an in vitro erythroid differentiation system to study the developing red blood cell transcriptome derived from adult CD34+ hematopoietic progenitor cells. mRNA expression profiling was used to characterize developing erythroid cells at six time points during differentiation (*days 1*, *3*, *5*, *7*, *9*, and *11*). Eleven thousand seven hundred sixty-three genes (20,963 Affymetrix probe sets) were expressed on *day 1*, and 1,504 genes, represented by 1,953 probe sets, were differentially expressed (DE) with 537 upregulated and 969 downregulated. A subset of the DE genes was validated using real-time RT-PCR. The DE probe sets were subjected to a cluster metric and could be divided into two, three, four, five, or six clusters of genes with different expression patterns in each cluster. Genes in these clusters were examined for shared transcription factor binding sites (TFBS) in their promoters by comparing enrichment of each TFBS relative to a reference set using transcriptional regulatory network analysis. The sets of TFBS enriched in genes up- and downregulated during erythropoiesis were distinct. This analysis identified transcriptional regulators critical to erythroid development, factors recently found to play a role, as well as a new list of potential candidates, including Evi-1, a potential silencer of genes upregulated during erythropoiesis. Thus this transcriptional regulatory network analysis has yielded a focused set of factors and their target genes whose role in differentiation of the hematopoietic stem cell into distinct blood cell lineages can be elucidated.

erythropoiesis; hematopoietic progenitor; transcriptional regulation; expression profiling

THE HUMAN ERYTHROID CELL develops from a hematopoietic progenitor cell (HPC) in the adult bone marrow and can be used as a model to identify critical steps in cell fate determination. Circulating peripheral blood contains a small number of HPCs that can be isolated and differentiated in vitro. The erythroid cell, once mature, transports oxygen via hemoglobin, which contains two α- and two β-like globin chains coordinated with a heme moiety for oxygen binding and delivery. In humans, expression of globin genes is under developmental control, with different α- and β-like globin chains expressed in a temporal fashion (for review, see Ref. 49a). Studies of the erythroid transcriptome in the context of hemoglobinopathies, a common class of diseases involving these genes, will offer insight into the regulation of globin gene expression. Creation of a model system that recapitulates erythropoiesis in the normal adult will allow dissection of the red cell and globin gene expression program. One consequence of these studies may be insights regarding reactivation of the fetal globin (Hb F) program, which may be of clinical benefit for patients with sickle cell disease and β-thalassemia.

Human HPCs, characterized by surface expression of CD34, have been used to study erythroid differentiation in vitro. CD34+ cells can be differentiated from bone marrow, fetal liver, cord and/or peripheral blood samples in semisolid and/or liquid cultures. Liquid culture systems yield relatively pure, reasonably synchronized erythroid cells recapitulating the developmental program in vivo (29, 37, 39, 47, 51). The early erythroid progenitors (burst-forming units, erythroid; or BFUe) proliferate and develop into colony-forming units, erythroid (CFUe), which divide and give rise to proerythroblasts, orthochromatic normoblasts, and enucleated erythrocytes. These systems generate sufficient material for molecular analysis, having been used to study accumulation of a limited number of mRNAs including globins, transcription factors, and cytokine receptors (47). Earlier studies showed fetal bovine serum (FBS)-induced Hb F in cultures from adults (37); thus we implemented a single-phase, serum-free, liquid culture system in which to differentiate CD34+ HPCs into erythroid progenitors. This system allows adult-derived HPCs to mature into erythroid progenitors that have downregulated fetal hemoglobin expression, thus recapitulating all the initial steps in erythroid development.

Genome-wide expression profiling allows insights into development of normal cell types. Recent studies focused on defining transcriptomic profiles of various human stem cell sources [bone marrow, cord blood, granulocyte-colony stimulating factor (G-CSF)-mobilized peripheral blood CD34+ cells] and subsets (CD34+Lin− and CD34+Lin+) (3, 13, 17, 33, 41, 50, 56) to better understand distinctions between these cell populations as well as examine the effects of stroma (57), both of which are important and relevant to studies of human transplantation. Examinations of the erythroid transcriptome have involved use of embryonic stem cell-derived erythroid colonies (62), long-term cultured murine FDCP-mix cells (8), the G1E-ER4 GATA-1-null cell line (58), retrovirally transduced CD34+ cells (10), and our studies and those of others

Article published online before print. See web site for date of publication (http://physiolgenomics.physiology.org).

Address for reprint requests and other correspondence: S. Surrey, Cardeza Foundation of Hematologic Research and the Division of Hematology, Jefferson Medical College, 703 Curtis Bldg., 1015 Walnut St., Philadelphia, PA 19107 (e-mail: saul.surrey@jefferson.edu).

using the K562 chronic myelogenous leukemic cell line to model erythroid differentiation (1, 29, 39, 64). More recently, studies of in vitro erythroid differentiation of human bone marrow-derived HPCs were performed (27); however, the in vitro culture medium contained FBS, such that the regulation of the globin program in this setting may not be reflective of the in vivo environment. If erythroid differentiation and globin gene regulation are to be understood at the transcriptional level, transcriptome analysis must be performed sequentially during erythroid differentiation under conditions that recapitulate the in vivo globin gene expression program. In the present study, the changing transcriptome of in vitro differentiated erythroid progenitors from adult peripheral blood was examined using a culture system optimized to yield adult-derived erythroid progenitor cells expressing low Hb F. In this setting, we performed a bioinformatic characterization of the transcriptome at six time points during red cell development, allowing us to examine the expression patterns of known regulators of developmentally controlled globin programs and to trace their role during early erythroid development.

In the present study, we analyzed expression profiles in timed samples during erythroid differentiation, following an unbiased approach to define the minimal number of statistically significantly distinct clusters of coregulated genes across this time course. To overcome limitations of available clustering algorithms such as K-means clustering (52) and self-organizing maps (53), in which data sets are arbitrarily fit to a user-defined number of clusters, we used the silhouette coefficient (SC) metric to separate the differentially expressed (DE) set into clusters distinct from randomly permuted expression data (45). Then, we analyzed the promoters of these 1,504 DE genes, unclustered or divided into two, three, four, five, or six clusters, for enrichment or overrepresentation of transcription factor binding sites (TFBS) compared with a reference gene set using transcriptional regulatory network analysis (TRNA). We used our promoter analysis and interaction network toolset (PAINT) software (Ref. 54; **http://www.dbi.tju.edu/dbi/tools/paint**), which automates promoter retrieval and mining of existing databases for known regulatory information for a large number of genes identified in a particular biological experiment. Using this analysis, we identified enriched TFBS in the DE gene set as a whole as well as at the level of two, three, four, five, or six clusters of DE genes. The list of candidate regulators includes GATA, whose role in erythroid development has been well documented, PITX and ATF, whose functions in red cell development were identified only recently, and Nkx2.5 and Ecotropic virus integration site-1 (Evi-1), which have not been implicated previously in this process.

EXPERIMENTAL PROCEDURES

Isolation of Mononuclear Cells from Peripheral Blood

Use of buffy coat specimens from the Thomas Jefferson Blood Center was approved by the Thomas Jefferson University (TJU) Institutional Review Board. Buffy coat (~40 ml) from 500 ml of peripheral blood was diluted 1:1 (vol/vol) with phosphate-buffered saline (PBS), pH 7.4, and gently layered on Ficoll-Hypaque (density 1.077 g/ml). Tubes were centrifuged (600 *g*, 15 min) at room temperature without applying the brake. The interphase was isolated, diluted with PBS, and centrifuged (~300 *g*, 5 min), and the mononuclear cell pellet was washed twice with PBS. The yield of mononuclear cells was generally 2–5 $\times$ 10^8 cells/buffy coat.

Isolation of CD34+ Cells, Erythroid Culture, and Cell Staining

Washed mononuclear cells (2 $\times$ 10^8 to 5 $\times$ 10^8 cells) were resuspended in 1 ml of PBS containing 2% (vol/vol) FBS and 1 mM EDTA in 12 $\times$ 75-mm polystyrene round-bottom tubes (BD, Franklin Lakes, NJ). EasySep CD34+ Positive Selection kit was used, as per the manufacturer's instructions (Stem Cell Technologies, Vancouver, BC, Canada). Purified CD34+ cells were cultured for 1 day in DMEM containing 20% (vol/vol) serum substitute (Stem Cell Technologies), 2 mM glutamine, 100 μg/ml each penicillin and streptomycin, 10^{-5} M β-mercaptoethanol, 0.3 mg/ml holo-transferrin, 10 ng/ml IL-3, 10 ng/ml stem cell factor, and 4 U/ml erythropoietin. At *day 1*, nonadherent cells (1 $\times$ 10^4 cells /ml) were transferred to new flasks and cultured for up to 14 consecutive days at 37°C in 5% CO_2. At *days 7* and *10*, cultures were supplemented with 2 ml of complete medium. The cells were harvested at the following time points: days *1*, *3*, *5*, *7*, *9*, and *11*. Each of three samples was pooled from three cultures for *days 1* and *3* (from a total of 9 donors), while HPCs from three different donors were used to generate timed cell harvests at *days 5*, *7*, *9*, and *11*. Cytospins were performed on *days 1*, *3*, *5*, *7*, *9*, and *11*, and cell morphology assessed by Giemsa staining (36).

Benzidine Staining

PBS-washed cells were stained in a benzidine solution containing 0.6% (wt/vol) benzidine base, 2% (vol/vol) hydrogen peroxide, and 12% (vol/vol) acetic acid. At least 500 cells were counted at each time point using a light microscope to assess the percentage of cells that appeared blue because of staining of heme-containing globin tetramers.

Protein Analysis and Enzyme-Linked Immunosorbent Assays

Protein was estimated by the bicinchoninic acid (BCA) protein assay kit (Pierce, Rockford, IL). Erythroid progenitors were washed with PBS and lysed in RBC Lysis Buffer (Gentra Systems, Minneapolis, MN). Fetal hemoglobin concentration in cultured cells was determined using a colorimetric enzyme-linked immunosorbent assay (ELISA). The Hb F ELISA (Bethyl Lab, Montgomery, TX) uses a two-antibody sandwich to detect Hb F. The percent Hb F was determined by dividing micrograms of Hb F by micrograms of total hemoglobin, measured spectrophotometrically at 415 nm (NanoDrop ND-1000; NanoDrop Technologies, Rockland, DE) and calculated using 125 as the millimolar extinction coefficient for human hemoglobin (e.g., 128 μg/ml has an optical density of 1.0 at 415 nm) (55).

Hemoglobin Analysis via HPLC

Cord blood hemolysates were examined for Hb F via HPLC and used as standards in the ELISA assay. After centrifugation of hemolysates, the supernatant was filtered through Ultrafree-MC devices (Millipore, Bedford, MA) before cation exchange chromatography. Hemoglobins were separated on a 100 $\times$ 4-mm POLYCATA column (PolyLC, Columbia, MD) fitted to a Waters automatic HPLC system using a gradient of *Buffer A* (50 mM Na_3PO_4, pH 5.5, 2 mM KCN) and *Buffer B* (50 mM Na_3PO_4, 500 mM NaCl, 2 mM KCN). Hemoglobin was detected by absorbance at 540 nm. Ratios of Hb F to Hb A were calculated by peak integration using the Dynamax R Data Reprocessing Program (Rainin Instrument, Oakland, CA). Purified Hb standards (FASC) were used for reference (Helena Laboratories, Beaumont, TX). Modeling experiments in which mixtures of adult and cord hemolysates were analyzed by HPLC indicated that <2% Hb F in a 100-μg total hemoglobin sample is detectable.

Total RNA Isolation and cDNA Synthesis

DNA-free total RNA of cultured cells was isolated with RNeasy microkit (Qiagen, Valencia, CA), according to the manufacturer's instructions. In brief, 1 $\times$ 10^6 cells from triplicate cultures (*days 1*, *3*, *5*, *7*, *9*, and *11*) were pelleted, lysed in RLT buffer containing 1%

(vol/vol) β-mercaptoethanol. DNase-treated RNA was ethanol precipitated and quantified on a NanoDrop ND-1000 spectrophotometer, followed by RNA quality assessment by analysis on an Agilent 2100 bioanalyzer (Agilent, Palo Alto, California). First-strand cDNA was synthesized using oligo(dT) and Superscript II RT (Invitrogen, Grand Island, NY). Alternatively, cDNA was prepared using OVATION RNA Amplification System (NuGen Technologies, San Carlos, CA).

Real-Time RT-PCR

Validation of DE genes. cDNA was assayed in quadruplicate reactions using 2× SYBR Green JumpStart TAQ Ready Mix (Sigma-Aldrich, St. Louis, MO) according to the SYBR Green protocol at the following input RNA concentrations: 2 ng, 0.2 ng, and 0.02 ng. Relative steady-state levels of mRNA expression of each gene and a reference gene (GAPDH) were assayed on an ABI 7900 (Applied Biosystems, Foster City, CA) using the relative quantification or "$2^{-\Delta\Delta CT}$" method, essentially as described previously (1).

Determination of γ/γ+β globin mRNA levels. For globin gene mRNA expression, real-time PCR assays were performed in quadruplicate using degenerate primers (forward 5′-GTC TAC CCW TGG ACC CAG AGG TTC-3′, reverse 5′-GGC AAA GGT GCC CTT GAG R-3′) (Integrated DNA Technologies, Coralville, IA) that simultaneously amplify both γ- and β-globin gene regions. Gene-specific probes (G 5′-[FAM] AGA TGC CAT AAA GCA CC-3′, B 5′-[TET] GGC CTG GCT CAC C-3′) (Applied Biosystems) were used in separate reactions to detect γ- and β-globin-amplified products. To document specificity of detection, plasmids containing cDNA for γ- or β-globin were used in individual real-time PCR assays, with detection of γ-globin template limited to reactions containing the γ-globin-specific probe and visa versa. A standard curve generated using increasing amounts of cultured erythroid progenitor cDNA (*day 9*) was used to confirm that the amplification efficiency of γ- and β-globin transcripts was equivalent. The difference in cycle thresholds (CT) for γ- and β-globin amplification (ΔCT) defines the fold increase in mRNA using the formula $2e^{\Delta CT}$, where e is the efficiency of amplification, derived from cDNA titration experiments. A titration of cDNA input showed a constant ratio of γ-globin to β-globin mRNA over a range of input amounts and facilitated determination of the percentage of γ-globin mRNA (γ/γ+β) present in erythroid progenitors during development, using the following two equations:

1) $\gamma + \beta = 100$, where the percentages of γ-globin and β-globin mRNA = 100%; and

2) $\beta/\gamma = 2e\Delta^{CT}$, where e is the average of the average efficiency of amplification for γ- and β-globin mRNA, CT is the cycle threshold, and Δ^{CT} represents the difference in cycle thresholds for β- and γ-globin signals.

Microarray Methods

Linear amplification. Ribo-SPIA-based RNA amplifications and target preparations were performed according to the manufacturer's instructions (Ovation Biotin System, NuGen). Briefly, first-strand cDNA was synthesized from 50 ng of total RNA using RT with a unique oligo(dT)/RNA chimeric primer. RNA was degraded by heating, and fragments served as primers for second-strand synthesis, yielding double-stranded cDNAs with RNA/DNA hetero-duplexes at one end. RNA in the hetero-duplexes was digested using RNase H added to the reaction with DNA polymerase and a second chimeric cDNA/cRNA primer (SPIA amplification primer). Amplification was continued using primer extension product hybridization to the target to reveal part of the priming site for subsequent primer hybridization and extension by strand displacement DNA synthesis. Amplified cDNA products were purified by Zymo Research DNA clean and concentrator (Zymo Research, Orange, CA).

Fragmentation and biotin labeling. cDNA amplification products were fragmented and chemically labeled with biotin to generate biotinylated cDNA targets. Finally, biotin-labeled product was purified on a DyeEx 2.0 spin column (Qiagen, Germantown, MD).

Hybridization. Fragmented and biotin-labeled target (2.5 μg) in 200 μl of hybridization cocktail was used for each Affymetrix HG U133 Plus 2.0 array (Affymetrix, Santa Clara, CA). Target denaturation was done at 99°C for 2 min, and hybridization was performed for 18 h. Arrays were washed and stained using GeneChip Fluidic Station 450, and hybridization signals were amplified using antibody amplification with goat IgG (Sigma-Aldrich) and anti-streptavidin biotinylated antibody (Vector Laboratories, Burlingame, CA). GeneChips were scanned using a GeneArray scanner 3000 (Affymetrix).

Bioinformatic analysis of mRNA expression profiling. Fragmented biotin-labeled cDNA was hybridized to the Human Genome 133 Plus 2.0 oligonucleotide array chip (Affymetrix) containing 56,000 probe sets representing 34,000 well-characterized human genes. Chips were scanned with Affymetrix GeneChip Scanner 3000, and the data were scaled from each array to a target intensity value of 500 using GeneChip Operating Software (GCOS) v3.0. GeneSpring v7.2 (Silicon Genetics, Redwood City, CA) was utilized to set intensities less than zero to zero, normalize signal per chip to the 50th percentile, and normalize signal per gene to the median of each gene. The raw microarray data set (series no. GSE4655) can be accessed at the Gene Expression Omnibus (GEO) website (**http://www.ncbi.nlm.nih.gov/geo/**). The complete list of *day 1* Present (P) calls is available in the Supplemental Materials (the online version of this article contains the supplemental data) including raw intensities of probe sets and gene identification information (Supplemental Table 1). Statistically significantly DE genes were identified from the normalized data using one-way analysis of variance (ANOVA), with analysis of the local false discovery rate employing a sliding window of 50 probe set *P* values (2, 14); the size of the window was chosen based on our group's experience with multiple data sets (11, 61). In contrast to overall false discovery rate (FDR) estimates, the local false discovery rate (*fdr*) estimates the false positive rate within a neighborhood of genes (chosen as 50 here). Often, as is the case in our data set, the choice of FDR threshold is not clear (e.g., why 10 and not 12%, or 14%, etc.?; with less-restrictive threshold resulting in an increasing no. of DE probe sets). In contrast, within a certain *fdr* range, the number of DE genes is relatively insensitive to the choice of a particular *fdr* threshold (2). This allows us to derive a differential gene expression list with a particular *fdr* threshold. A total of 1,953 probe sets were chosen as DE at a 10% local *fdr* threshold, which corresponds to allowing ~7% overall false positives.

Gene annotation. The expressed and DE lists were linked to the genome database NetAffx at the NetAffx Analysis Center (**http://www.affymetrix.com**) using Microsoft Excel and Access, and discrepancies identified by cross-reference to **http://mriweb.moffitt.usf.edu/mpv/** (19) were manually corrected in Supplemental Table 2. Gene Ontology (GO) functions were assigned using Database for Annotation, Visualization and Integrated Discovery (DAVID v2.0; **http://david.abcc.ncifcrf.gov/**).

Clustering of DE Probe Sets

Clusters of different sizes ranging from 2 to 10 were obtained using partitioning around mediods (PAM) (23). The quality of the clusters was evaluated using SC metric (49). SC is a combined measure of cohesion within a cluster and separation between clusters utilizing both inter- and intracluster distances as a ratio. Following the computational negative control approach of Pearson et al. (46), the SC of different PAM clusters of actual DE data were compared with the SC of the randomly permuted DE data. Differences between SC of the actual and randomized data allowed determination of the number of nonrandom clusters that were well distinguished from clusters of randomized data. On the basis of this difference, we chose to analyze PAM results based on two to six clusters. We manually constructed a

hierarchical schematic to illustrate the dominant movement of probe sets from the DE set into increasing numbers of clusters.

Transcriptional Regulatory Network Analysis

With the use of PAINT, TFBS were analyzed within 1,000 base pairs (bp) upstream of the transcriptional start sites (TSS) in the DE gene set. Several array probes correspond to the same gene, and this redundancy must be removed before further analysis. The UniGene cluster identification number (ID) corresponding to each probe set was obtained through cross-reference with GenBank accession number. From the UniGene database, the corresponding Entrez gene ID was obtained and used to cross-reference with Ensembl IDs. The 1,953 DE probe sets mapped to 1,853 genomic locations (Ensembl annotation), of which 1,397 were unique. The TFBS were identified using TransfacPro 9.2 database (35) and associated MATCH software (25) with option for minimizing the sum of false positives and false negatives in the TFBS results. The enrichment analysis was performed for different cluster sets (2–6, from above), and the DE list was compared with different background reference sets. In the enrichment analysis, the "Factor" corresponding to the TFBS (as indicated by MATCH/TRANSFACPro) was considered instead of the individual TFBS. This mapping allows us to account for the correlations in the position-weight matrices (PWMs) to a certain extent. A total of 264 TFBS families were considered. The enrichment *P* values based on hypergeometric distribution were adjusted for multiple testing using a FDR estimate (5). At each of the clustering levels, we do account for multiple testing using FDR correction for different TFBS enrichment *P* values. It should be noted that this FDR estimate may be conservative, as it does not account for the correlations among the TFBS present on different promoters. The correlations could arise because of the biological nature of coordinate action of different transcription factors (TFs) and because of similarity of PWMs of several TFs in the TRANSFAC database. The FDR method used might identify a higher number of TFs as being significant than if all correlations were taken into account. However, this is not fatal, given the limited number of predictions from our discovery approach. In the case of the hypothetical factor X family, all binding sites corresponding to the X family of TFs (FACTORX-1/V$FACTORX-1_02, FACTORX-1/V$FACTORX-1_03, FACTORX-2/V$FACTORX-2_02, FACTORX-2/V$FACTORX2_03) would be grouped into one set. Those binding sites present in ≥10% of promoters in the particular set are presented.

Isolation of Nuclear Extracts and Assessment of TF:DNA Binding Activity

Cells were harvested, and nuclear extracts were prepared from erythroid cultures on *day 1* and *day 8* using Panomics Nuclear Extraction kit (Panomics, Fremont, CA). Briefly, 1×10^6 cells were incubated in hypotonic buffer and dispersed, and nuclei were lysed in high-salt buffer. Protein was quantified using the BCA protein assay kit and NanoDrop spectrophotometer. Approximately 75 μg of protein were isolated and stored in the presence of protease inhibitor cocktail. Nuclear extracts from *day 1* and *day 8* (~7.5 μg) were incubated with the biotin-labeled probe mix from the Panomics Protein/DNA TranSignal comboarray. Protein-bound probes were isolated via streptavidin, protein was removed, and the probes were hybridized to the array and visualized using chemiluminescence detection on X-ray film.

RESULTS

In Vitro Erythroid Differentiation of Adult-Derived HPCs

We implemented a single-phase liquid culture method for expansion and differentiation of HPC from adult blood samples, and we isolated RNA for expression profiling, protein for hemoglobin analysis, and cells for cytocentrifugation from cultures after *days 1*, *3*, *5*, *7*, *9*, and *11* in culture in serum-free medium containing erythropoietin, IL-3, and stem cell factor. Because before *day 5*, cell numbers are insufficient for all necessary analyses, for *day 1* and *day 3*, a pooling strategy was used (see Fig. 1). In this way, microarray analysis at all time points was performed in triplicate, with *days 1* and *3* each involving three arrays of pooled samples containing equal amounts of three RNA samples. That is, three different donors were used to generate one RNA sample for *day 1* and one for *day 3*. This was repeated two more times so that nine donors in all were used for triplicate pools of RNA for *days 1* and *3*.

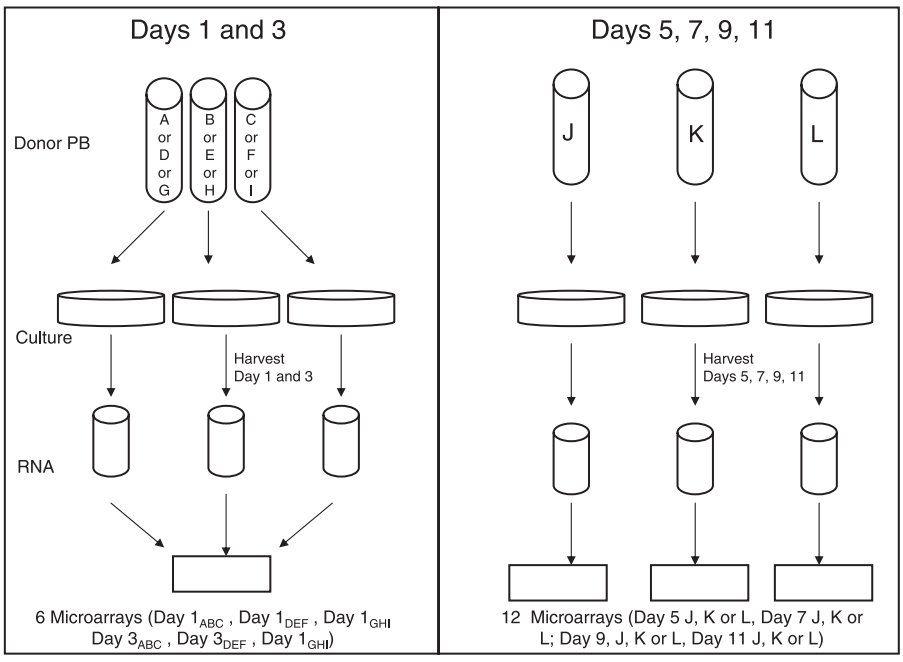

Fig. 1. Schematic of experimental design. For examination of *days 1* and *3* of erythroid differentiation, CD34+ cells from 9 adult donors (*donors A–I*) were cultured and harvested on *days 1* and *3*; cells from 3 cultures were combined to generate hemolysate and RNA samples for triplicate analysis (*ABC*, *DEF*, and *GHI*). One microarray was generated with the pooled sample for each of *days 1* and *3*. For the later time points, 3 donors (*donors J*, *K*, and *L*) were used, and samples were removed from the culture at *days 5*, *7*, *9*, and *11*. Three microarrays were performed for each time point, as depicted.

Expression profiling at the later time points involved three donors representing biological replicates. Specifically, for *days 5*, *7*, *9*, and *11*, three donors were used to generate RNA at each time point. In total, 18 array hybridizations were analyzed (3 at each time point). Cell number rose from ~2 million on *day 5* to ~30 million on *day 13* (Fig. 2*A*), and the number of cells containing hemoglobin peaked on *day 9* at 80–90% (Fig. 2*B*). Cells contained ~4 pg/cell of hemoglobin on *day 5*, peaking at 5–8.5 pg/cell between *days 9* and *11* (M. A. Keller, unpublished observations). Hb F determination showed that one of the cultures expressed <2% Hb F at all time points, while another expressed 3.2% Hb F at *day 9* before declining to 2.7% at *day 11* (Fig. 2*D*). Microscopic examination of stained, cytospun samples showed that the majority of cells were erythroid with condensed nuclei as well as enucleating red blood cells after 11 days (Fig. 2*C*).

Analysis of γ/γ+β Globin mRNA Levels

We developed a real-time RT-PCR assay with custom-designed, gene-specific, fluorescently labeled, minor-groove binder probes to assess changes in globin gene mRNA over time. A titration of cDNA input showed a constant ratio of β/γ-globin mRNA over a range of input amounts and facilitated determination of the percentage of γ-globin mRNA (γ/γ+β) present in erythroid progenitors during development (data not shown). This approach eliminates the need to calculate amounts of each mRNA relative to an internal control (GAPDH), greatly facilitating determination of γ/γ+β mRNA ratios in cultured cells. Transcript analysis from *day 5* to *day 11* demonstrated that the ratio of γ/γ+β is <3% at all times examined in two of the cultures, and in one culture, it is highest on *day 5* (12.8%), followed by a decrease to 3% on *day 11* (Fig. 3). Similar percentages of Hb F mRNA have been seen by others using in vitro erythroid differentiation of adult-derived HPC (38). These findings demonstrate individual variation in γ-globin gene expression that correlates with Hb F expression in this in vitro differentiation system. Thus we have cultured progenitors from adult blood under conditions that minimize expression of Hb F and recapitulate the in vivo state using a single-phase, totally defined, serum-free medium.

Analysis of DE Genes During Erythroid Differentiation

Total RNA from *days 1*, *3*, *5*, *7*, *9*, and *11* was used to examine mRNA expression via Affymetrix U133 Plus 2.0 arrays. One-way ANOVA was used with a (local) *fdr* analysis to identify DE genes (Fig. 4). While the (global) FDR (5)

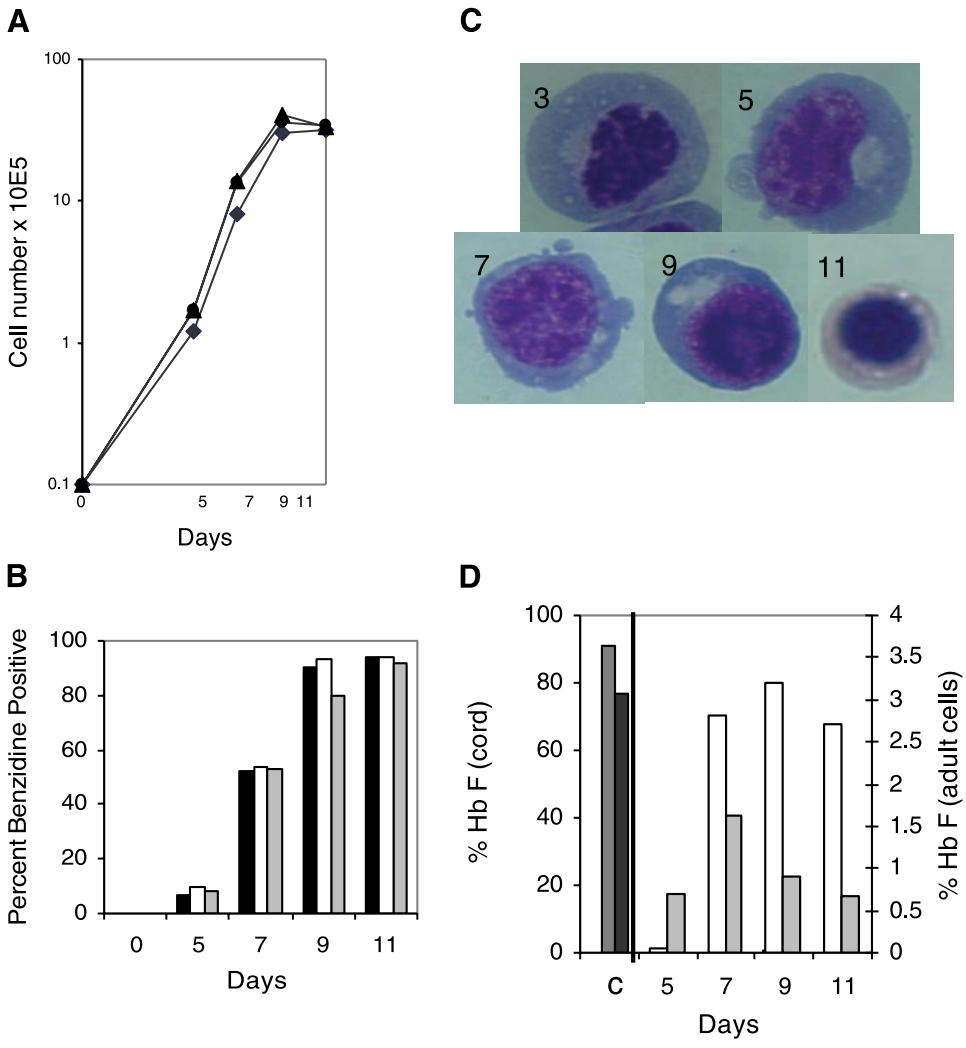

Fig. 2. Cell proliferation of CD34+ cells during in vitro erythroid differentiation. Cell numbers (*y*-axis, log scale) are plotted as a function of time in culture (*x*-axis) for 3 independent donor cultures (*A*). Hemoglobinization in these 3 cultures was measured by benzidine staining, and percent benzidine-positive cells (*y*-axis) as a function of time in days (*x*-axis) is shown (*B*). Cell morphology was examined using Giemsa staining, and representative microscope fields are shown at ×40 magnification on *days 3*, *5*, *7*, *9*, and *11* (*C*). *D*: percent fetal globin (Hb F; *y*-axis) was measured in each culture and plotted as a function of days in culture (*x*-axis); for comparison, percent Hb F in 2 representative cord blood hemolysates (containing 77 and 91% Hb F) is shown at *left*.

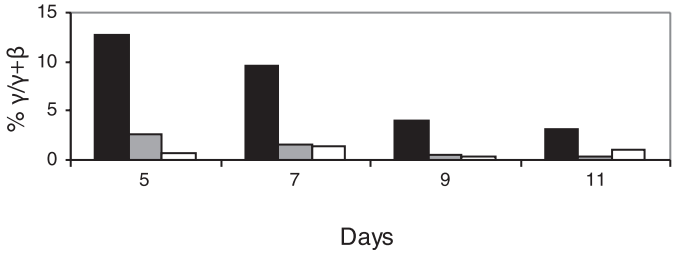

Fig. 3. Percent γ-globin mRNA in RNA from cultured adult-derived hematopoietic progenitor cells (HPCs). Real-time RT-PCR with a primer pair that amplifies both γ- and β-globin mRNA regions, followed by detection with gene-specific probes, was performed on cDNA from total RNA extracted from cultured cells.

estimates overall false positive rate, the *fdr* (2, 14) estimates as a function of gene index (sorted in ascending order of *P* values). When identifying DE genes, those with *P* values close to the chosen threshold have a significantly higher fraction of false positives than the genes with *P* values far below the threshold. A threshold is chosen based on the opportunity cost (in specificity) of predicting more genes as DE. The local *fdr* represents that opportunity cost and provides a robust metric for choosing the appropriate false positive rate threshold. For this study, we estimated the *fdr* within a neighborhood of 50 genes. A total of 1,953 probe sets were chosen as DE. By examination of the 1,953 DE probe sets for known functions, erythroid-specific genes of several classes were found among the upregulated genes. These included red cell surface antigens such as GLYA, Kell and Duffy; heme biosynthesis enzymes such as FECH and ALAS2; and members of the globin gene family. Erythroid-associated factor (ERAF), the protein that regulates α-globin chain stability (16), is not expressed on *day 1*, is upregulated >10-fold on *day 3*, and is highly expressed at later time points as well. The downregulated gene set includes TFs (ELK3, ETS1, EVI2A, GATA-2, KLF12, NMYC, RUNX3 and ZNF521), signaling molecules (FLT3, JAK3, PRKCB1), and structural proteins [CD34, CD109, HLA-A, -B and -C, platelet endothelial cell adhesion molecule-1 (PECAM1)] whose expression has been seen in human peripheral blood CD34+ cells (17). The complete list of DE probe sets with intensities at each time point is available as Supplemental Material (Supplemental Table 2).

Several genes were chosen for validation by real-time RT-PCR. When mRNA levels at *days 3*, *5*, *7*, *9*, and *11* were compared with those at *day 1*, the five genes examined were shown to be upregulated, confirming their classification as DE (Fig. 5). Three of the genes were chosen for follow-up because of their roles as TFs involved in globin gene regulation (BACH1, KLF1, GATA-1). Interestingly, three genes encoding markers for platelet development, SELP, platelet factor-4 (PF4), and proplatelet basic protein, showed 5-, 15-, and 22-fold upregulation with peak signal intensities of 400, 3,600, and 7,900 on *day 11*, respectively. Two other genes, ALAS2 and FECH, are enzymes in the heme biosynthesis pathway. Although the microarray analysis confirmed that ALAS2 is upregulated during erythroid development (>250-fold), the magnitude of the change in expression is much higher when assessed using real-time RT-PCR (>1,000-fold). Also worthy of note, the microarray data show expression from the γ- and β-globin genes increasing ~15-fold and 10-fold, respectively, over the 11-day time course. Real-time RT-PCR using SYBR Green detection showed larger fold changes for γ- (>100-fold) and β-globin (>1,000-fold) mRNA (data not shown). Differences in the fold change comparing microarray and real-time analysis are not unexpected, given that real-time RT-PCR analysis has a wider dynamic range compared with microarray analysis.

The DE gene set was compared with those identified in our prior study of hemin-induced erythroid differentiation of the erythroleukemic cell line, K562. When the 899 probe sets DE in the K562 study were compared with the 1,953 probe sets in the present study, ~100 genes were DE in both data sets (see Supplemental Table 3). Aside from erythroid-specific genes such as blood group antigens, heme biosynthesis enzymes and globins, the list showed several TFs (ATF3, CEBPβ, v-maf F, and several zinc-finger proteins) whose expression pattern increased during differentiation in both data sets and a transcriptional regulator (Wilm's tumor suppressor Wt1) whose expression decreased in both data sets. Interestingly, some probe sets that are DE in both data sets show divergent

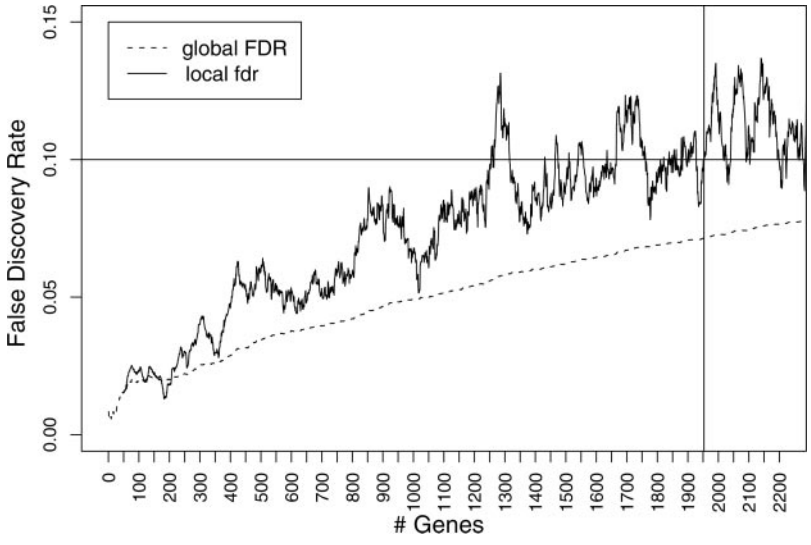

Fig. 4. Determination of statistically significant, differentially expressed (DE) genes from ANOVA using local false discovery rate approach. The global false discovery rate (FDR, dotted line) and local false discovery rate (*fdr*, solid line) are plotted as a function of gene index (sorted in ascending order of *P* values). A total of 1,953 probe sets were chosen as DE (vertical line) at a 10% local *fdr* threshold. This corresponds to an ~7% global FDR.

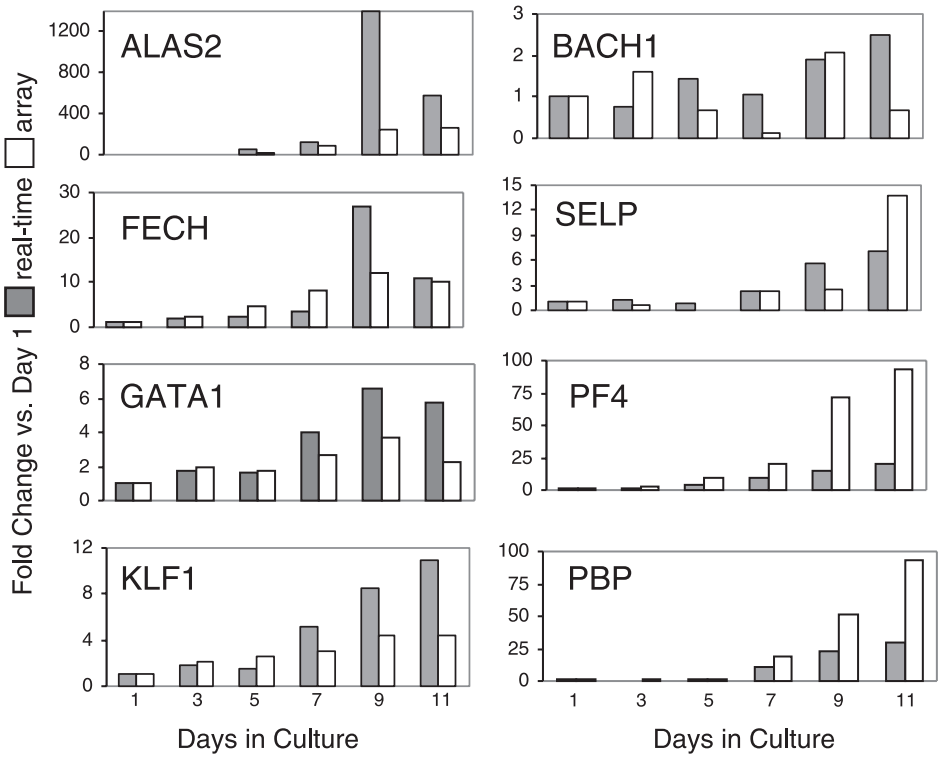

Fig. 5. Comparison of fold change from microarray expression profiling (open bars) to real-time RT-PCR analysis (gray bars) of a subset of genes identified as DE.

expression patterns, (e.g., upregulated in the BFUe data set and downregulated in the K562 data set), reinforcing well-known distinctions between primary cells and cancer cell lines.

Identification of Coregulated Clusters of DE Genes

After identifying the 1,953 DE probe sets, we analyzed them for shared expression patterns. A clustering metric (silhouette coefficient) (18) was used to examine the divergence from randomly permuted DE data when the DE list was segregated into increasing numbers of clusters, from 2 to 10. The DE probe sets can be divided into a number of clusters that are distinct from randomly permuted DE data. As shown in Fig. 6*A*, two to six clusters are distinct from randomly permuted DE data. We present a graphic representation of cluster assignments with temporal gene expression patterns in Fig. 6*B*. Cluster memberships are included in Supplemental Table 2. Although the greatest distance from random, and therefore the greatest confidence, is seen when the DE gene set is divided in two clusters, one up- and one downregulated, it is likely that regulatory information can be gained by further clustering into additional expression patterns. Therefore, transcriptional regulatory network analysis was done at all levels of clustering found to be distinct from randomly permuted data (e.g., 2, 3, 4, 5, and 6). Mean expression data from representative probe sets from each cluster are shown in Fig. 7. At the most basic level of two clusters, the upregulated genes in cluster C2-2 include many genes necessary to make a red blood cell, including the globins, heme biosynthesis enzymes, and membrane and surface proteins. On further clustering of the upregulated genes, they are distributed into three clusters, with C6-2 containing α- and β-globin, β-spectrin, and glycophorins-A, -B, and -E; C6-4 containing Duffy blood group antigen and NFE2; and C6-6 containing ankryin, the erythropoietin receptor, and γ-globin genes.

Functional Classification of DE Gene Set

The 1,953 DE probe sets, representing 1,504 unique UniGene IDs, were examined for GO functions. Those classified functions assigned to ≥2% of the genes were graphed, either in the unclustered set (DE1953) or at the six-cluster level (Fig. 8). In cluster C6-2, which contains several globin genes, blood group antigens, and glycophorins, there is an increased proportion of carrier and metal ion, protein, and receptor binding functions. It is the only cluster with <2% of gene functions made up of TFs. Cluster C6-3 contains genes with enzyme inhibitor functions, while C6-4 contains both transcription cofactor and ligase activity subsets, and C6-5 contains genes with oxidoreductase function.

Analysis of Gene Clusters for Transcriptional Coregulation

The potential coregulation of these genes in the unclustered set or in two, three, four, five, or six clusters was examined by identifying statistically enriched TFBS in the 1,397 promoters of genes in the DE gene set and after each level of clustering compared with a reference set. The enrichment of TF families was determined compared with three different reference lists, yielding different contextual results on TFs associated with distinct expression clusters. The first reference list, termed "array," encompasses 17,741 available promoters of all genes represented on the Affymetrix U133 Plus 2.0 array. This comparison would be expected to uncover regulators involved in hematopoietic gene expression, as this reference list includes promoters representing the entire human genome. The second

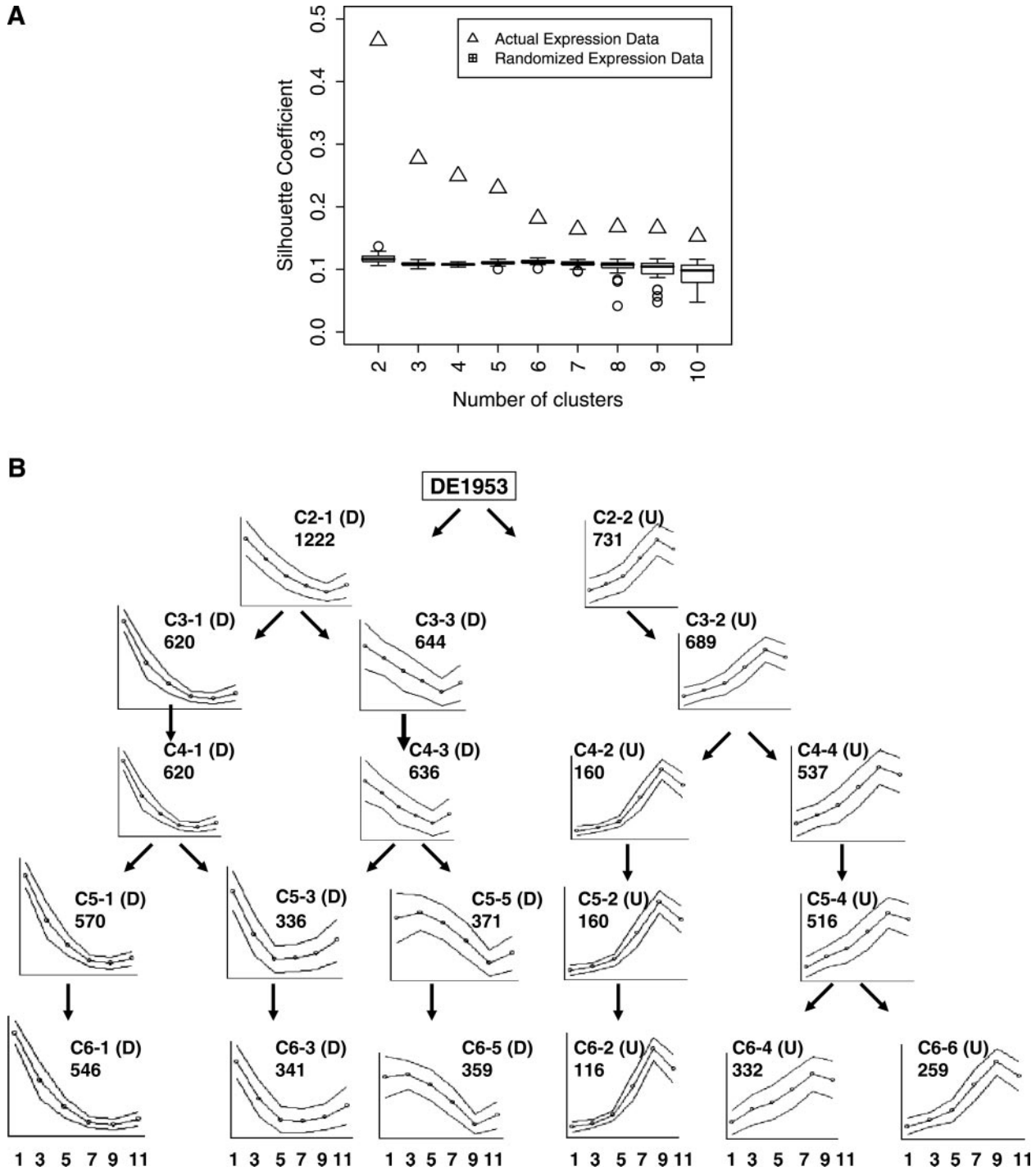

Fig. 6. Cluster analysis. A comparison of silhouette coefficients (SC) between partitioning around mediods (PAM) clusters of actual expression data (triangles) and randomly permuted expression data (box plots) is shown in *A*. Temporal expression patterns after varying levels of clustering from 2 to 6 are shown in *B*. Median intensity (*y*-axis) with standard deviation around the median is shown as a function of culture time in days (*x*-axis), with no. of DE probe sets in that specific cluster listed below cluster nos. Letter in parentheses next to each cluster designation indicates its classification of up- (U) or downregulated (D). Clusters, from 2 to 6, are graphically represented to illustrate paths of cluster emergence, with arrows representing the most dominant paths. In most cases, as clusters are further divided, a small fraction of probe sets are assigned to a cluster other than the indicated dominant paths. Cluster memberships of all DE probe sets are listed in Supplemental Table 2.

reference list of 9,846 promoters, termed "expressed," is composed of promoters of all genes expressed in the developing erythroid cultures (e.g., the total set of genes expressed on *day 1* plus the DE set). The comparison of DE genes to this reference list may identify factors that are involved in erythroid differentiation specifically, since the reference list includes the DE genes as well as genes expressed on *day 1* (before erythroid commitment); the *day 1* gene set includes both genes involved in self-renewal and genes involved in other hematopoietic lineages. The third comparison, "cluster to list," uses the DE

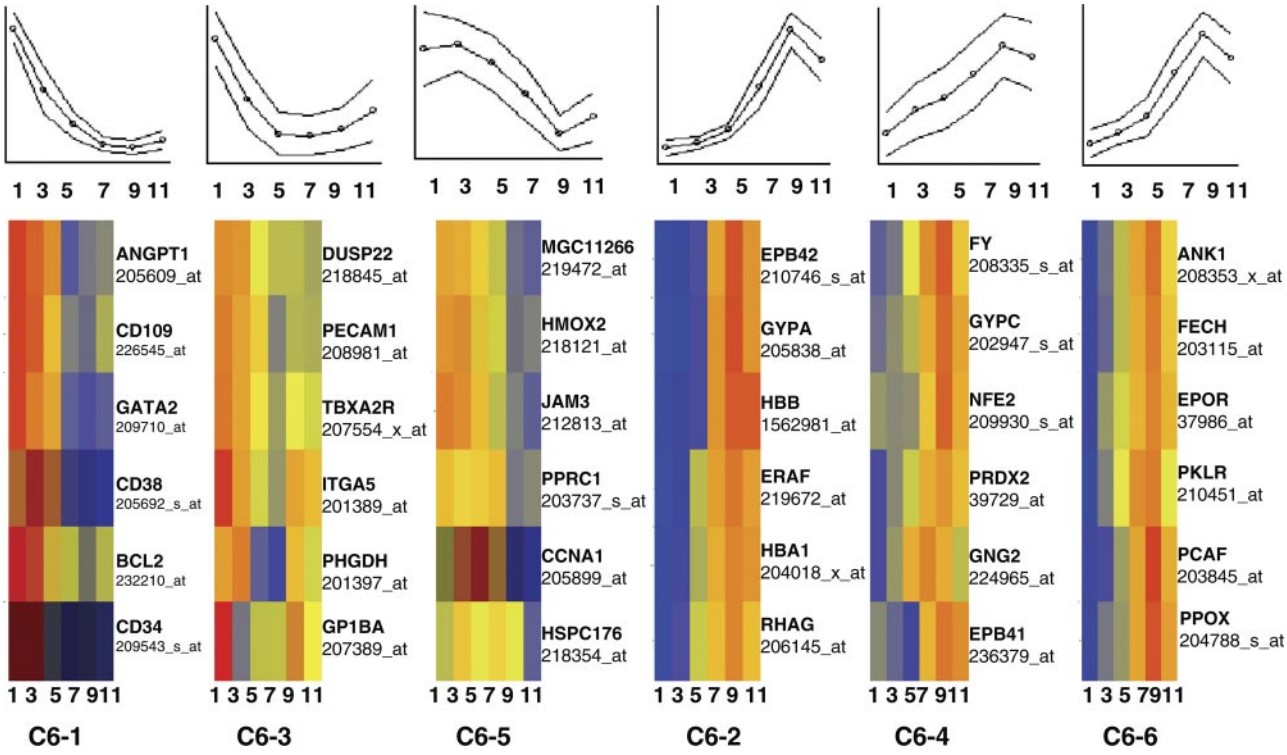

Fig. 7. Heat map of representative DE probe sets at the level of 6 clusters. Expression levels of 6 different genes in each of the clusters at the 6 time points during erythroid differentiation are depicted as heat maps (blue indicates low, while red indicates high expression). Gene symbols are shown to the *right*, with Affymetrix probe set designation below, and the corresponding graphs of temporal expression (from Fig. 6*B*) are shown at *top*.

gene set as the reference list, encompassing the promoters of all the DE genes. Because each of the reference sets corresponds to a different biological question being answered, the results from three analyses were not collated into one combined multiple testing correction but were analyzed separately. Each cluster is compared with the DE list, such that enriched TFBS may identify factors whose role is specific to that cluster of genes compared with sites involved in regulating the larger set of DE genes.

The unclustered DE gene set was examined for evidence of coregulated gene expression. Seventeen binding sites identified as statistically significantly overrepresented (FDR <35%) in

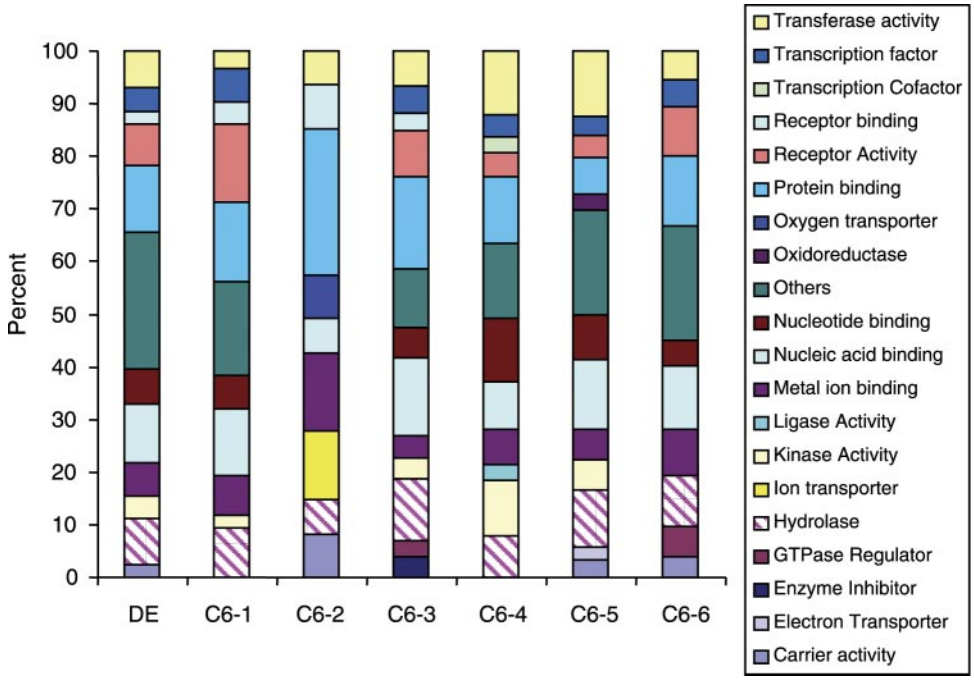

Fig. 8. Bar graph of gene ontology functions in the DE set, unclustered, and at the 6-cluster level. The various categories of functions are listed in the key and indicated in the graph with different colors and fill patterns, with approximate percentages of the various functions listed for each cluster at the 6-cluster level. Functions represented by <2% of all functions at any cluster are combined in the "other" classification.

Table 1. *Schematic of overrepresented factor-binding sites in the DE gene promoters after varying levels of clustering (0–6)*

		2 Clusters		3 Clusters			4 Clusters				5 Clusters					6 Clusters						
Factor-Binding Site	DE	2-1	2-2	3-1	3-2	3-3	4-1	4-2	4-3	4-4	5-1	5-2	5-3	5-4	5-5	6-1	6-2	6-3	6-4	6-5	6-6	RefS
ATF			↑		↑					↑				↑					↑			A, C
ATF4			↑							↑				↑					↑		↑	C
CP2	x	↓		↓			↓				↓					↓						A, E
CREB										↑				↑					↑			A, C
CREBATF																			↑			C
DEAF1																			↑			C
E12	x	↓		↓			↓				↓					↓						A, E
E2F						↓			↓						↓				↑			C
E2F-1										↑				↑					↑			A, C
E2F-1:DP-1																			↑			A, C
ELF-1	x			↓			↓				↓					↓			↑			A, E
Elk-1						↓			↓						↓							A, C
Ets (c-Ets-2)				↓			↓				↓					↓						A, E
Evi-1																			↑			A, E, C
FOX homolog XFD2																↓						E
FOXA2 (HNF-3beta)	x																		↑			A, E
FOXL1 (Freac-7)																↓						E
GABP						↓			↓						↓							C
GATA	x			↓			↓				↓					↓						A, E
HLF	x																					A
HNF-1	x															↓						A, E
IRF				↓			↓									↓						A, E
KROX	x																					A
Lmo2 complex	x			↓			↓															A, E
MAZ	x																					E
Muscle TATA box	x																		↑			A, E, C
Myb (c-Myb)																			↑			A, C
Myb (v-Myb)						↓			↓										↑			A, C
MYEF-2 (MEF2)																			↑			C
NF-kappaB	x	↓		↓			↓				↓					↓						A, E
NKX2.5				↓			↓				↓					↓						A, E, C
NRF2						↓			↓						↓					↓		C
PAX																			↑			C
PAX3										↑				↑					↑			C
PAX6										↑				↑							↑	C
PITX2										↑				↑					↑			C
PolyA binding	x	↓		↓			↓				↓					↓						E
POUF1 (Pit-1)	x																					E
RUNX1 (AML)	x																					E
RUNX2	x																					E
SMAD3				↓			↓				↓					↓						E
Sox-5	x	↓									↓					↓						A, E
Sp3				↓			↓				↓					↓						E
TCF-4																			↑			A, E, C

Factor-binding sites, listed alphabetically, enriched in the unclustered (differentially expressed; DE) gene promoters (x), or after different clustering levels (2–6) are indicated by the arrows, with upward arrows (↑) for transcription factor-binding sites in upregulated clusters and downward arrows (↓) in downregulated clusters. Overrepresented sites were determined relative to 3 different reference sets (RefS): the array (A), the expressed gene set (E), or the DE list (C); and, when enrichment is found in one set and/or another, it is listed in the far-*right* column.

the DE group are summarized in the first column of Table 1 ("x") with *P* values and binding site enumerations listed in Supplemental Table 4. In clusters where adjusted *P* values were >35%, the top several binding sites for TFs known to play a role in hematopoiesis are included in this group. Among these are GATA-1, Lmo2 complex, MAZ, and RUNX1 (AML1) as well as cell cycle-specific TFs E2F and CREB-ATF.

Next, statistically significant TFBS enrichments were determined in the DE set after each level of clustering from two to six compared with the three reference sets. Of the TFBS identified in the unclustered DE set, enrichment of six [hepatic leukemia factor (HLF), RUNX1, RUNX2, POUF1, KROX, and MAZ] binding sites is no longer significant once the list is divided into two or more clusters. This may indicate that these sites are important in regulation of genes that are both up- and downregulated during erythroid differentiation. In fact, the percentage of promoters that contain one or more copies of these TFBS is not significantly different in the down- and upregulated clusters (C2-1 and C2-2, respectively). For example, the percentage of HLF TFBS-containing promoters in C2-1 and C2-2 is 11% in both.

Of the 17 TFBS identified as enriched in the unclustered list, 5 (E12, CP2, polyA binding, NF-κB, and Sox-5) are enriched in the downregulated clusters (Table 1, downward arrows) starting with the 1,221 promoters in C2-1 through to the 546 promoters in C6-1. Sp3, Nkx2.5, c-Ets-2, SMAD3 IRF, and NRF are newly identified in the downregulated cluster C3-1. The promoters of the other downregulated cluster, C3-3, are enriched for the TFBS of NRF2, GABP, Elk-1, E2F, and

Table 2. *Candidate regulators of up- and downregulated genes identified by PAINT analysis evaluating overrepresentation of factor-binding sites in DE gene promoters*

Enriched in downregulated clusters
c-Ets-2, CP2, E12, ELF-1, Elk-1, FOXL1 (Freac-7), GABP, GATA, HNF-1, IRF, Lmo2 complex, NF-κB, Nkx2.5, NRF-2, polyA binding protein, SMAD-3, Sox-5, Sp3, XFD-2
Enriched in upregulated clusters
ATF, ATF4, c-myb, CREB, CREBATF, DEAF-1, E2F, E2F-1, E2F-1:DP-1, Evi-1, FOX2A (HNF-3β), MYEF-2, Muscle TATA, Pax, Pax-3, Pax-6, PITX, TCF-4, v-myb
Enriched in unclustered only
HLF, KROX, MAZ, POUF-1 (Pit-1), RUNX (AML1), RUNX2 (core-binding factor)

Overrepresented factor-binding sites present in one or more downregulated or upregulated gene clusters are listed. Downregulated clusters include C2-1, C3-1 and C3-3, C4-1, C4-3, C5-1, C5-3, C5-5, C6-1, C6-3, and C6-5, indicated by "(D)" next to expression patterns in Fig. 6*B*. Upregulated clusters include C2-2, C3-2, C4-2, C4-4, C5-2, C5-4, C6-2, C6-4, and C6-6, indicated by "(U)" next to expression patterns in Fig. 6*B*. PAINT, promoter analysis and interaction network toolset.

v-myb. At the level of six clusters, two TF binding sites (FOXL1 and FOX homolog XFD-2) are newly identified as enriched in the downregulated cluster C6-1.

In examining TFBS identified in upregulated clusters (those indicated by upward arrows, Table 1), we found fewer sites are enriched at the lower levels of clustering. At the two-cluster level, the TFBS for ATF and ATF4 are overrepresented in C2-2, with both sites persisting in the upregulated clusters through to C6-4 (ATF) and both C6-4 and C6-6 (ATF4). At the four-cluster level, upregulated cluster C4-4 shows enrichment for five newly identified factor binding sites: CREB, E2F1, Pax3, Pax6, and PITX2. All continue to be enriched in C5-4. When the clustering reaches six, enrichment of Pax3 sites segregates to C6-4, while Pax6 enrichment is seen in C6-6. Eight TFBS are newly identified at the six-cluster level, overrepresented in C6-4, including Evi-1 and TCF4.

Reference Set Differences

In total, 44 TFBS were identified as enriched (Table 1). While the majority of TFBS enriched in promoters of genes in downregulated clusters (downward arrows in Table 1) were identified in both the array and expressed reference set comparisons ("A, E" in "RefS" column in Table 1), the majority of the TFBS enriched in promoters of genes in upregulated clusters (upward arrows in Table 1) were identified when the reference set was the array or the DE set, not the expressed set. The cluster-to-list analysis, examining enrichment in each cluster compared with the DE set, resulted in 11 TFBS not identified with the other reference sets (those with "C" only in RefS column, Table 1).

Compilation of Candidate Regulators of Up- and Downregulated Genes

A listing of candidate transcriptional regulators whose TFBS were found to be enriched in one or more clusters of up- or downregulated sets is summarized in Table 2. Closely related TFs are found in the same list, for example, several CREB and E2F family members in the upregulated genes, GABP and NRF2 as putative regulators of the downregulated genes, and the RUNX family members as putative regulators of both up- and downregulated genes. This list contains TF families already known to play a role in hematopoietic gene regulation (GATA, Evi-1, RUNX, myb) along with many factors whose role in hematopoietic lineage commitment and differentiation has heretofore been unknown.

Examination of TF Binding Activity in Day 1 and Day 8 Erythroid Progenitors

We examined binding to a set of consensus TFBS using nuclear extracts from early (*day 1*) and late (*day 8*) stages of erythroid differentiation. Binding of nuclear proteins to biotin-tagged consensus oligonucleotides followed by hybridization to membrane and chemiluminescent detection was performed using Panomics TransSignal arrays (21). We chose to examine Evi-1, whose TFBS was enriched in cluster C6-4 compared with all three reference sets. Steady-state mRNA for Evi-1 decreases during erythroid differentiation (Fig. 9), and this is inversely correlated with the differential expression of genes in C6-4, which increase over time. This would be consistent with Evi-1 acting as a transcriptional repressor for these target genes, such that when it is expressed and functional, the genes are repressed, and when it is absent or nonfunctional, the target genes are expressed. The TransSignal array data support this model, showing considerable binding activity at *day 1* but not at *day 8* (Fig. 9, *inset*); both experiments use equivalent amounts of protein (~7.5 μg) and are exposed for equivalent amounts of time (20 min). Binding activity to Evi-1 TFBS is not detectable on *day 8*, even with a longer exposure to X-ray film (2 h, data not shown). The ability to use nuclear extracts from primary human HPCs to examine TF function will be useful in examining gene regulation during erythroid differentiation. Thus both steady-state mRNA levels and DNA binding activity to Evi-1 TFBS suggest a transcriptional silencer role for this factor in upregulation of some genes upregulated late in culture.

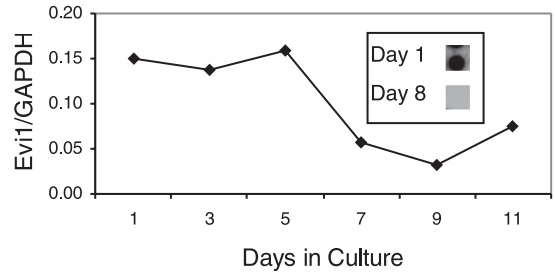

Fig. 9. Evi-1 mRNA levels decrease with erythroid differentiation and correlate with DNA binding activity. Plot of Evi-1 mRNA expression as assessed by ratio of Evi-1 probe set 226420at_ to GAPDH probe set on Affymetrix arrays. *Inset* shows DNA binding activity in nuclear extracts derived from *day 1* and *day 8* progenitor cells. Equivalent amounts of extract (~7.5 μg protein each) were used to examine transcription factor binding activity on Panomics Combo TranSignal array and exposed to X-ray film for 20 min each. The portion of the array containing the Evi-1 consensus transcription factor binding sites is shown for *day 1* (*top*) and *day 8* (*bottom*) arrays.

DISCUSSION

Establishment of In Vitro Erythroid Differentiation System

This study employed an in vitro culture system to examine early transcriptional events leading to erythroid commitment of human HPCs in the absence of FBS. Hemoglobin analysis, at both the RNA and protein levels, indicated that erythroid progenitor cells produced using this method express predominantly adult globins. Thus, the in vitro culture system permits the study of transcriptional regulation under conditions that mimic development. Microarray expression profiling of triplicate samples at six time points during an 11-day differentiation regimen allowed for statistical analysis to identify DE genes.

Identification of DE Genes During Erythroid Differentiation

The expression profile study design, using three replicates at each time point, allowed for statistical determination of DE without using a raw signal intensity cutoff to determine differential expression, such that low-abundance mRNAs including those encoding transcriptional regulators and signaling molecules were included in the DE list. Upregulated genes included those known to be critical to red cell development including structural proteins and enzymes specific to the erythroid lineage, such as the globins, blood group antigens and heme biosynthesis enzymes. Downregulated genes included cell cycle regulators and TFs regulating hematopoietic progenitors.

Three genes, P-selectin, PF4, and platelet basic protein, heretofore thought to be platelet specific, were found to be upregulated by microarray, and this upregulation was validated by real-time RT-PCR analysis. There are at least two explanations for this finding. It is possible that a minor fraction of megakaryocyte precursors survive under erythroid culturing conditions. Alternatively, these genes may have an unrecognized role in normal erythropoiesis. Recently, integrin-αIIb (15) and PECAM (4) were found to be expressed in erythroid progenitors. Since the erythroid and megakaryocytic lineages share a common progenitor, these findings may not be unexpected.

Clustering Metric Identifies Levels of Coexpression

Given that clustering of DE genes can facilitate identification of genes sharing common temporal expression patterns, we hypothesized that examination of promoters of genes represented in a given cluster that share TFBS would uncover coregulation by transcriptional regulators. While many bioinformatics packages, including GeneSpring, allow for clustering of DE genes into an arbitrary, user-defined number of clusters, we used a clustering metric to examine the potential number of clusters whose members share an expression pattern distinct from randomly permuted data. In this way, we determined that our DE gene set can be divided into clusters ranging from two to six with different temporal patterns of expression. Further clustering was not distinct from such randomly permuted data. Next, we used a family analysis, examining the enrichment of TFBS, made up of all transcriptional regulatory elements that are bound by a given factor. Enrichment analysis is based on the principle that, if the site or sites for the factor were present in a greater number compared with a reference set, this site may play a role in the regulation of the genes in that cluster containing the site. Obvious caveats include the fact that enrichment of one site may not be sufficient, since multiple interactions may be required involving multiple TFBS. Furthermore, enrichment and functional utilization are not necessarily equivalent, and further experimentation involving gel shifts with TFBS-containing oligonucleotides and knockdown, knockout, and/or chromatin occupancy studies are required to validate functionally the role of enriched TFBS. Enrichment for a TFBS could signify binding of a different factor than the one initially described to bind to this site. Finally, a TFBS highly prevalent in the genome could play a critical role in the process being studied but be missed in this analysis, since its frequency already may be high in the reference lists used for comparison.

Comparison of DE List with Other Studies of Erythroid Development

In addition, the DE gene list was compared with that of Komor et al. (27) in which bone marrow-derived CD34+ cells were differentiated toward the erythroid, granulocyte, or megakaryocyte lineage over an 11-day period. Of the DE probe sets in the erythroid lineage in their study, there were 60 unique UniGene IDs, and of these, 31 are shared with the present study and include many erythroid-specific genes such as ALAS2, GYPA, RHD, and ANK1. The "signature" of the erythroid lineage described in their data set was highlighted by upregulation of two GTPases, RAP1GA1 and ARHGAP8, both upregulated also in our data set, and downregulation of Mina53, GLMN (FAP48), and MAX gene-associated protein (MGA). GLMN is in our downregulated DE set. Probe sets for MGA and MINA show decreased expression over time, but neither were classified as DE after ANOVA. Although there are significant commonalities between the two studies, our study of erythroid differentiation identifies a considerably larger set of DE genes, and, importantly, the culturing conditions used by Komor et al. included FBS, and no characterization of globin expression was presented.

We examined our data set for genes whose gene products recently have been implicated in erythroid development. Annexin 1 (ANXA1) was reported to be critical to erythroid differentiation after showing that its knockdown in K562 cells decreased hemin-induced erythroid differentiation (20). However, in our study, ANXA1 is downregulated 20-fold during erythroid differentiation, suggesting its role may be limited to hemin-induced hemoglobinization of K562 cells. Another study in K562 cells suggested that expression of transglutaminase 2 (TGM2) was important in erythroid differentiation (22), and, in our study, expression of TGM2 is upregulated on arrays 63-fold during erythroid differentiation. Growth arrest and DNA damage-inducible transcript 34 (GADD34), also known as protein phosphatase 1, regulatory subunit 15A (PPP1R15A), was shown to be important to hemoglobin synthesis in a Gadd34-null mouse model (44). Competition was proposed between GADD34 and eukaryotic translation initiation factor-2α (eIF2α) kinase 1, also known as HRI, for regulation of the phosphorylation state of eIF2α, such that GADD34-null mice demonstrate a thalassemic phenotype. In our system, GADD34 is upregulated more than fivefold during erythroid commitment.

Transcriptional Regulatory Network Analysis Yields a List of Candidate Regulators

Using PAINT software, we performed TRNA to identify enriched TFBS in the promoters of DE genes, with or without clustering, and using three different reference sets. A total of 44 TFBS were identified with similar numbers of candidates identified as potential regulators of up- and downregulated genes (see Table 1). Interestingly, 10 factors were identified only at the six-cluster level. The use of different reference sets allowed us to identify regulators of hematopoietic gene expression, such as HLF and KROX, identified when the DE set was compared with the array, as well as regulators of erythroid differentiation, such as MYEF-2 and PITX, identified when clusters were compared with the DE set.

Furthermore, we explored the literature for validation of erythroid gene regulation by some of the candidate factors identified by our enrichment analyses. First, the binding site for the homeobox protein Pitx2 was enriched in upregulated clusters C4-4, C5-4, and C6-4 in the cluster-to-list analysis. Two recent studies of Pitx2-deficient mice show that Pitx2-null fetal livers have a reduced erythroid component, and that there is partial rescue of the hematopoietic potential when stromal cells are rescued with Pitx2 +/+ cells (26, 63). Our finding that Pitx2 binding sites are enriched in the promoters of upregulated genes during erythroid maturation of human hematopoietic progenitors supports an emerging role for Pitx2 and/or Pitx2-like factors.

Second, TFBS for Evi-1 were enriched compared with all three reference sets, consistently overrepresented in cluster C6-4. Evi-1, first identified as a key regulator in myeloid leukemias and myelodysplastic syndromes (9, 40, 43), is also critical to early erythroid differentiation. Aberrant expression of Evi-1 in patients is associated with abnormal erythroid development (6), and overexpression of Evi-1 during in vitro erythroid differentiation results in reduced CFUe (28). Evi-1 is a GATA binding factor, and the GATA-1 consensus site (WGATAR) is contained in the Evi-1 consensus site. Recently, Evi-1 was shown to regulate GATA-2 expression (59, 60). The 76 promoters within C6-4 that contain these sites included CD36, CPOX, and NFE2. A role for Evi-1 or related factors in regulation of expression of these genes is supported by binding assays performed with nuclear extracts from *day 1* and *day 8* erythroid progenitor cells (Fig. 9, *inset*). This sets the stage for further examination of the role of Evi-1 in transcriptional silencing of target genes identified in this study as well as its role in erythroid differentiation using this model system.

Third, ATF binding sites were enriched in the upregulated gene clusters compared with the array, and ATF4 binding sites were enriched when the clusters were compared with the DE list. The ATF/CREB TFs are basic leucine-zipper transcriptional regulators. Interestingly, ATF4 deficiency results in severe fetal anemia due to a defect in proliferation of hematopoietic progenitors (34), and, as mentioned earlier, it is known to bind to the promoter and regulate expression of GADD34 (32, 42). As mentioned earlier, GADD34 was upregulated, and, at the six-cluster level, was found in cluster C6-6 where TFBS for ATF4 were overrepresented.

Fourth, TFBS for Nkx2.5 were overrepresented in the downregulated gene clusters, from C2-1 to C6-1, independent of the reference list used. Nkx2.5 is a member of the homeobox family, with a critical role in cardiac development (30). Nkx2.5 is known to interact with a GATA family member, GATA-4, and recently GATA-6 has been shown to be necessary for upregulation of Nkx2.5 (7). A potential role for Nkx2.5 or related factors in hematopoietic development is interesting to postulate. The genes represented in cluster C6-1 showed a significant decline in mRNA expression from *day 1* to *day 3*, suggesting that transcriptional repression must occur early. At *day 1*, expression of Nkx family members was low or absent in our microarray data set. Nkx transcription may be transient, perhaps measurable at *day 0* (CD34+ cells), a population not examined in this analysis. Alternatively, enrichment of Nkx2.5 sites in the downregulated clusters may be indicative of another factor that binds a similar sequence. Functional follow-up will be necessary to address this.

Some Critical Regulators Not Identified Using TRNA

Just as intriguing as the list of candidate regulators (Table 2) are those not identified in our analyses, including a family of regulators critical to hemoglobin gene regulation. Kruppel-like factors, such as erythroid KLF (KLF1), are known to play a role in expression of erythroid genes (12), many of which (HBA1, HBA2, HBB) are expressed in C6-2. We performed a post hoc analysis of binding site enrichment for these factors, which are classified as CACCC binding factors. Although these sites did not emerge as significantly enriched after multiple testing, the CACCC binding site is enriched in C6-2 (unadjusted P value = 0.15) in the cluster-to-list analysis, with 10% of the promoters in this cluster containing one or more of these sites compared with 6% in the unclustered promoter list. When we examined our data set for expression of EKLF target genes identified through the use of an EKLF knockout mouse model (12), heme biosynthesis enzyme ALAS2, α-globin stabilizing protein (ERAF), peroxiredoxin 2 (PRDX2), and the erythroid-specific membrane protein Band 4.9 (EPB49) were contained in our DE list. EKLF target genes EIF2AK1 (HRI) and transferrin receptor (CD71 or TFRC) were upregulated more than twofold in our system but were not statistically significantly DE, as assessed by ANOVA. Another EKLF target gene, Gardos channel (KCNN4), showed very little change in expression in our system but is represented by a single probe set on the U133 Plus 2.0 array.

The TFBS for GATA family members was enriched in the downregulated gene set but absent from the upregulated set (Table 2). The enrichment for GATA sites is complicated by several factors. First, the GATA consensus is common, being present on 57% of the promoters of the genome, 56% of promoters of the expressed set, and 59% of the DE set. Second, GATA is known to play a role in both up- and downregulation of gene expression in hematopoiesis, through interaction with various corepressors and coactivators (48), and therefore would be expected to play a role in both up- and downregulated genes. GATA was found to be enriched in the downregulated clusters in both the comparison to the array and to the expressed gene set. At the six-cluster level, GATA binding sites are enriched in C6-1, present on 62% of promoters in that cluster. Further examination of enrichment of TFBS for GATA in conjunction with binding sites for coregulators, using predictive tools for composite binding site analysis (24, 31), may help elucidate the role of this complex family of regulators in erythroid development. This type of composite TFBS analysis also may be necessary to determine whether the

segregation of the adult and fetal globin genes into different clusters at the six-cluster level (adult in C6-2, fetal in C6-6) is due to factors critical to developmental-specific regulation of globin expression.

Improved annotation is a continuous process, and, in that context, our analysis has the same inherent limitations that affect all microarray studies. To estimate annotation discrepancies, we performed a comparison of the NetAffx annotation we employed at the time of original submission with alternative annotation available at **http://mriweb.moffitt.usf.edu/mpv/**. We found ~1.2% discordance at the level of UniGene cluster ID. Considering the FDR threshold used in our analysis of gene clusters for transcriptional coregulation, we do not expect this to substantially affect the enrichment results.

In conclusion, this study of erythroid differentiation in primary, nontransformed cells provides a detailed characterization of the developing erythroid transcriptome. Our bioinformatics tools allowed us to use this transcriptome information to define gene clusters and to examine the promoters of the genes in these clusters for shared transcriptional regulation. We have identified candidate transcriptional regulators, some established, some novel, and some of which are gathering interest based on multiple studies for a potential role in erythroid development. Thus, we have demonstrated that a computational analysis of transcriptomic information during erythroid development is a valid approach for identification of candidate regulators in hematopoietic lineage commitment and red cell development, and have generated a focused list of candidates that will be the subject of future functional studies. Furthermore, we have demonstrated that candidate transcriptional regulators identified by this analysis can be studied through TF binding analysis in primary human HPCs.

ACKNOWLEDGMENTS

We thank C. Ponte for technical assistance, K. Adachi for performing HPLC analysis, C. Pratt and P. Chakravarthula for help with TRNA as well as P. Fortina and the TJU Center for Translational Medicine Genomics Core Facility staff for contributions to microarray experiments. We also acknowledge W. K. Hofman and colleagues for sharing microarray expression data files from their recent publication (27) and S. Dessain and S. E. McKenzie for critical reading of the manuscript.

Present address of M. A. Keller: Coriell Institute for Medical Research, 403 Haddon Ave., Camden, NJ 08103-1505.

GRANTS

This work was supported in part by National Heart, Lung, and Blood Institute Grant HL-69256 (S. Surrey), the Commonwealth of Pennsylvania (M. A. Keller), The Cardeza Foundation for Hematologic Research (S. Surrey, M. A. Keller), and the Defense Advanced Research Projects Agency BioComputation program (R. Vadigepalli).

DISCLOSURES

This project is funded, in part, under a grant with the Pennsylvania Department of Health. The Department specifically disclaims responsibility for any analyses, interpretations, or conclusions.

REFERENCES

1. **Addya S, Keller MA, Delgrosso K, Ponte CM, Vadigepalli R, Gonye GE, Surrey S.** Erythroid-induced commitment of K562 cells results in clusters of differentially expressed genes enriched for specific transcription regulatory elements. *Physiol Genomics* 19: 117–130, 2004.
2. **Aubert J, Bar-Hen A, Daudin JJ, Robin S.** Determination of the differentially expressed genes in microarray experiments using local FDR. *BMC Bioinformatics* 5: 125, 2004.
3. **Bartolovic K, Balabanov S, Berner B, Buhring HJ, Komor M, Becker S, Hoelzer D, Kanz L, Hofmann WK, Brummendorf TH.** Clonal heterogeneity in growth kinetics of CD34+CD38− human cord blood cells in vitro is correlated with gene expression pattern and telomere length. *Stem Cells* 23: 946–957, 2005.
4. **Baumann CI, Bailey AS, Li W, Ferkowicz MJ, Yoder MC, Fleming WH.** PECAM-1 is expressed on hematopoietic stem cells throughout ontogeny and identifies a population of erythroid progenitors. *Blood* 104: 1010–1016, 2004.
5. **Benjamini YHY.** Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Royal Stat Soc Ser B*: 289–300, 1995.
6. **Brada S, de Wolf J, Hendriks D, Esselink M, Ruiters M, Vellenga E.** The supportive effects of erythropoietin and mast cell growth factor on CD34+/CD36− sorted bone marrow cells of myelodysplasia patients. *Blood* 88: 505–510, 1996.
7. **Brewer AC, Alexandrovich A, Mjaatvedt CH, Shah AM, Patient RK, Pizzey JA.** GATA factors lie upstream of Nkx 2.5 in the transcriptional regulatory cascade that effects cardiogenesis. *Stem Cells Dev* 14: 425–439, 2005.
8. **Bruno L, Hoffmann R, McBlane F, Brown J, Gupta R, Joshi C, Pearson S, Seidl T, Heyworth C, Enver T.** Molecular signatures of self-renewal, differentiation, and lineage choice in multipotential hemopoietic progenitor cells in vitro. *Mol Cell Biol* 24: 741–756, 2004.
9. **Buonamici S, Li D, Chi Y, Zhao R, Wang X, Brace L, Ni H, Saunthararajah Y, Nucifora G.** EVI1 induces myelodysplastic syndrome in mice. *J Clin Invest* 114: 713–719, 2004.
10. **Cammenga J, Mulloy JC, Berguido FJ, MacGrogan D, Viale A, Nimer SD.** Induction of C/EBPalpha activity alters gene expression and differentiation of human CD34+ cells. *Blood* 101: 2206–2214, 2003.
11. **Covarrubias MY, Khan RL, Vadigepalli R, Hoek JB, Schwaber JS.** Chronic alcohol exposure alters transcription broadly in a key integrative brain nucleus for homeostasis: the nucleus tractus solitarius. *Physiol Genomics* 24: 45–58, 2005.
12. **Drissen R, von Lindern M, Kolbus A, Driegen S, Steinlein P, Beug H, Grosveld F, Philipsen S.** The erythroid phenotype of EKLF-null mice: defects in hemoglobin metabolism and membrane stability. *Mol Cell Biol* 25: 5205–5214, 2005.
13. **Eckfeldt CE, Mendenhall EM, Flynn CM, Wang TF, Pickart MA, Grindle SM, Ekker SC, Verfaillie CM.** Functional analysis of human hematopoietic stem cell gene expression using zebrafish. *PLoS Biol* 3: e254, 2005.
14. **Efron B, Tibshirani R, Storey JD, Tusher V.** Empirical Bayes analysis of a microarray experiment. *J Am Stat Assoc* 96: 1151–1160, 2001.
15. **Ferkowicz MJ, Starr M, Xie X, Li W, Johnson SA, Shelley WC, Morrison PR, Yoder MC.** CD41 expression defines the onset of primitive and definitive hematopoiesis in the murine embryo. *Development* 130: 4393–4403, 2003.
16. **Gell D, Kong Y, Eaton SA, Weiss MJ, Mackay JP.** Biophysical characterization of the alpha-globin binding protein alpha-hemoglobin stabilizing protein. *J Biol Chem* 277: 40602–40609, 2002.
17. **Georgantas RW 3rd, Tanadve V, Malehorn M, Heimfeld S, Chen C, Carr L, Martinez-Murillo F, Riggins G, Kowalski J, Civin CI.** Microarray and serial analysis of gene expression analyses identify known and novel transcripts overexpressed in hematopoietic stem cells. *Cancer Res* 64: 4434–4441, 2004.
18. **Gordon AD.** *Classification*. London: Chapman and Hall, 1999.
19. **Harbig J, Sprinkle R, Enkemann SA.** A sequence-based identification of the genes detected by probesets on the Affymetrix U133 plus 2.0 array. *Nucleic Acids Res* 33: e31, 2005.
20. **Huo XF, Zhang JW.** Annexin1 regulates the erythroid differentiation through ERK signaling pathway. *Biochem Biophys Res Commun* 331: 1346–1352, 2005.
21. **Jiang X, Norman M, Roth L, Li X.** Protein-DNA array-based identification of transcription factor activities regulated by interaction with the glucocorticoid receptor. *J Biol Chem* 279: 38480–38485, 2004.
22. **Kang SK, Lee JY, Chung TW, Kim CH.** Overexpression of transglutaminase 2 accelerates the erythroid differentiation of human chronic myelogenous leukemia K562 cell line through PI3K/Akt signaling pathway. *FEBS Lett* 577: 361–366, 2004.
23. **Kaufman L, Rousseeuw PJ.** Finding groups in data: an introduction to cluster analysis. In: *Wiley Series in Probability and Mathematical Statistics*. New York: John Wiley and Sons, 1989, p. 68–122.
24. **Kel A, Konovalova T, Waleev T, Cheremushkin E, Kel-Margoulis O, Wingender E.** Composite Module Analyst: a fitness-based tool for iden-

tification of transcription factor binding site combinations. *Bioinformatics* 22: 1190–1197, 2006.
25. **Kel AE, Gossling E, Reuter I, Cheremushkin E, Kel-Margoulis OV, Wingender E.** MATCH: a tool for searching transcription factor binding sites in DNA sequences. *Nucleic Acids Res* 31: 3576–3579, 2003.
26. **Kieusseian A, Chagraoui J, Kerdudo C, Mangeot PE, Gage PJ, Navarro N, Izac B, Uzan G, Forget BG, and Dubart-Kupperschmitt A.** Expression of Pitx2 in stromal cells is required for normal hematopoiesis. *Blood* 107: 492–500, 2006.
27. **Komor M, Guller S, Baldus CD, de Vos S, Hoelzer D, Ottmann OG, Hofmann WK.** Transcriptional profiling of human hematopoiesis during in vitro lineage-specific differentiation. *Stem Cells* 23: 1154–1169, 2005.
28. **Kreider BL, Orkin SH, Ihle JN.** Loss of erythropoietin responsiveness in erythroid progenitors due to expression of the Evi-1 myeloid-transforming gene. *Proc Natl Acad Sci USA* 90: 6454–6458, 1993.
29. **Lee YT, Miller LD, Gubin AN, Makhlouf F, Wojda U, Barrett AJ, Liu ET, Miller JL.** Transcription patterning of uncoupled proliferation and differentiation in myelodysplastic bone marrow with erythroid-focused arrays. *Blood* 98: 1914–1921, 2001.
30. **Lints TJ, Parsons LM, Hartley L, Lyons I, Harvey RP.** Nkx-2.5: a novel murine homeobox gene expressed in early heart progenitor cells and their myogenic descendants. *Development* 119: 419–431, 1993.
31. **Liu X, Brutlag DL, Liu JS.** BioProspector: discovering conserved DNA motifs in upstream regulatory regions of co-expressed genes. *Pac Symp Biocomput*: 127–138, 2001.
32. **Ma Y, Hendershot LM.** Delineation of a negative feedback regulatory loop that controls protein translation during endoplasmic reticulum stress. *J Biol Chem* 278: 34864–34873, 2003.
33. **Manfredini R, Zini R, Salati S, Siena M, Tenedini E, Tagliafico E, Montanari M, Zanocco-Marani T, Gemelli C, Vignudelli T, Grande A, Fogli M, Rossi L, Fagioli ME, Catani L, Lemoli RM, Ferrari S.** The kinetic status of hematopoietic stem cell subpopulations underlies a differential expression of genes involved in self-renewal, commitment, and engraftment. *Stem Cells* 23: 496–506, 2005.
34. **Masuoka HC, Townes TM.** Targeted disruption of the activating transcription factor 4 gene results in severe fetal anemia in mice. *Blood* 99: 736–745, 2002.
35. **Matys V, Fricke E, Geffers R, Gossling E, Haubrock M, Hehl R, Hornischer K, Karas D, Kel AE, Kel-Margoulis OV, Kloos DU, Land S, Lewicki-Potapov B, Michael H, Munch R, Reuter I, Rotert S, Saxel H, Scheer M, Thiele S, Wingender E.** TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res* 31: 374–378, 2003.
36. **McDonald MJ.** Modification of the Wright-Giemsa stain for rapid staining. *Med Lab Sci* 44: 88, 1987.
37. **Migliaccio A, Migliaccio G, Brice M, Constantoulakis P, Stamatoyannopoulos G, Papayannapoulou T.** Influence of recombinant hematopoietins and of fetal bovine serum on the globin synthetic pattern of human BFUe. *Blood* 76: 1150–1157, 1990.
38. **Miller JS, McCullar V, Punzel M, Lemischka IR, Moore KA.** Single adult human CD34(+)/Lin−/CD38(−) progenitors give rise to natural killer cells, B-lineage cells, dendritic cells, and myeloid cells. *Blood* 93: 96–106, 1999.
39. **Mitchell T, Plonczynski M, McCollum A, Hardy CL, Safaya S, Steinberg MH.** Gene expression profiling during erythroid differentiation of K562 cells. *Blood Cells Mol Dis* 27: 309–319, 2001.
40. **Morishita K, Parganas E, William CL, Whittaker MH, Drabkin H, Oval J, Taetle R, Valentine MB, Ihle JN.** Activation of EVI1 gene expression in human acute myelogenous leukemias by translocations spanning 300–400 kilobases on chromosome band 3q26. *Proc Natl Acad Sci USA* 89: 3937–3941, 1992.
41. **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* 75: 314–323, 2004.
42. **Novoa I, Zeng H, Harding HP, Ron D.** Feedback inhibition of the unfolded protein response by GADD34-mediated dephosphorylation of eIF2alpha. *J Cell Biol* 153: 1011–1022, 2001.
43. **Ogawa S, Kurokawa M, Tanaka T, Mitani K, Inazawa J, Hangaishi A, Tanaka K, Matsuo Y, Minowada J, Tsubota T, Yazaki Y, Hirai H.** Structurally altered Evi-1 protein generated in the 3q21q26 syndrome. *Oncogene* 13: 183–191, 1996.
44. **Patterson AD, Hollander MC, Miller GF, Fornace AJ Jr.** Gadd34 requirement for normal hemoglobin synthesis. *Mol Cell Biol* 26: 1644–1653, 2006.
45. **Pearson BM, Pin C, Wright J, I'Anson K, Humphrey T, Wells JM.** Comparative genome analysis of Campylobacter jejuni using whole genome DNA microarrays. *FEBS Lett* 554: 224–230, 2003.
46. **Pearson RK, Zylkin T, Schwaber JS, Gonye GE.** Analytical evaluation of clustering results using computational negative controls. *Proc 4th Soc Indust Appl Math Int Conf Data Mining*, 2004, p. 188–199.
47. **Pope S, Gibach E, Sun J, Chin K, Rodgers G.** Two-phase liquid culture sytem models normal human adult erythropoiesis at themolecular level. *Eur J Haematol* 64: 292–303, 2000.
48. **Rodriguez P, Bonte E, Krijgsveld J, Kolodziej KE, Guyot B, Heck AJ, Vyas P, de Boer E, Grosveld F, Strouboulis J.** GATA-1 forms distinct activating and repressive complexes in erythroid cells. *EMBO J* 24: 2354–2366, 2005.
49. **Rousseeuw PJ.** Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J Comp Appl Math* 20: 53–65, 1987.
49a. **Stamatoyannopoulos G, Grosveld F.** Hemoglobin switching. In: *The Molecular Basis of Blood Diseases*, edited by Stamatoyannopoulos G, Majerusk P, Perlmutter R, Varmua H. Philadelphia, PA: W. B. Saunders, p. 135–182, 2001.
50. **Steidl U, Kronenwett R, Haas R.** Differential gene expression underlying the functional distinctions of primary human CD34+ hematopoietic stem and progenitor cells from peripheral blood and bone marrow. *Ann NY Acad Sci* 996: 89–100, 2003.
51. **Su AWT, Batalov S, Lapp H, Ching KA, Block D, Zhang J, Soden R, Hayakawa M, Kreiman G, Cooke MP, Walker JR, Hogenesch JB.** A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc Natl Acad Sci USA* 101: 6062–6067, 2004.
52. **Tavazoie S, Hughes JD, Campbell MJ, Cho RJ, Church GM.** Systematic determination of genetic network architecture. *Nat Genet* 22: 281–285, 1999.
53. **Toronen P, Kolehmainen M, Wong G, Castren E.** Analysis of gene expression data using self-organizing maps. *FEBS Lett* 451: 142–146, 1999.
54. **Vadigepalli R, Chakravarthula P, Zak DE, Schwaber JS, Gonye GE.** PAINT: a promoter analysis and interaction network generation tool for gene regulatory network identification. *Omics* 7: 235–252, 2003.
55. **Vinogradov SN, Sharma PKS.** Preparation and characterization of invertebrate globin complexes. In: *Methods in Enzymology*, edited by Abelson JN and Simon MI. San Diego, CA: Academic, 1994, p. 112–124.
56. **Wagner W, Ansorge A, Wirkner U, Eckstein V, Schwager C, Blake J, Miesala K, Selig J, Saffrich R, Ansorge W, Ho AD.** Molecular evidence for stem cell function of the slow-dividing fraction among human hematopoietic progenitor cells by genome-wide analysis. *Blood* 104: 675–686, 2004.
57. **Wagner W, Saffrich R, Wirkner U, Eckstein V, Blake J, Ansorge A, Schwager C, Wein F, Miesala K, Ansorge W, Ho AD.** Hematopoietic progenitor cells and cellular microenvironment: behavioral and molecular changes upon interaction. *Stem Cells* 23: 1180–1191, 2005.
58. **Welch JJ, Watts JA, Vakoc CR, Yao Y, Wang H, Hardison RC, Blobel GA, Chodosh LA, Weiss MJ.** Global regulation of erythroid gene expression by transcription factor GATA-1. *Blood* 104: 3136–3147, 2004.
59. **Yatsula B, Lin S, Read AJ, Poholek A, Yates K, Yue D, Hui P, Perkins AS.** Identification of binding sites of EVI1 in mammalian cells. *J Biol Chem* 280: 30712–30722, 2005.
60. **Yuasa H, Oike Y, Iwama A, Nishikata I, Sugiyama D, Perkins A, Mucenski ML, Suda T, Morishita K.** Oncogenic transcription factor Evi1 regulates hematopoietic stem cell proliferation through GATA-2 expression. *EMBO J* 24: 1976–1987, 2005.
61. **Zak DE, Hao H, Vadigepalli R, Miller GM, Ogunnaike BA, Schwaber JS.** Systems analysis of circadian time-dependent neuronal epidermal growth factor receptor signaling. *Genome Biol* 7: R48, 2006.
62. **Zambidis ET, Peault B, Park TS, Bunz F, Civin CI.** Hematopoietic differentiation of human embryonic stem cells progresses through sequential hematoendothelial, primitive, and definitive stages resembling human yolk sac development. *Blood* 106: 860–870, 2005.
63. **Zhang HZ, Degar BA, Rogoulina S, Resor C, Booth CJ, Sinning J, Gage PJ, Forget BG.** Hematopoiesis following disruption of the Pitx2 homeodomain gene. *Exp Hematol* 34: 167–178, 2006.
64. **Zhu Y, Lee HC, Zhang L.** An examination of heme action in gene expression: heme and heme deficiency affect the expression of diverse genes in erythroid k562 and neuronal PC12 cells. *DNA Cell Biol* 21: 333–346, 2002.