Genomics **59,** 168–177 (1999)
Article ID geno.1999.5855, available online at http://www.idealibrary.com on IDEAL®

Gene Expression in Proliferating Human Erythroid Cells

Alexander N. Gubin, J. Muthoni Njoroge, Gerard G. Bouffard,* and Jeffery L. Miller[1]

*Laboratory of Chemical Biology, National Institute of Diabetes and Digestive and Kidney Diseases, and *NIH Intramural Sequencing Center and the National Human Genome Research Institute, National Institutes of Health, 9000 Rockville Pike, Bethesda, Maryland 20892*

Received February 22, 1999; accepted April 13, 1999

A complete understanding of human erythropoiesis will require a robust description of transcriptional activity in hematopoietic cells that proliferate and differentiate in response to erythropoietin (EPO). For this purpose, we cultured peripheral blood mononuclear cells in the presence or in the absence of EPO and examined the transcriptional profile of those cells arising only in response to EPO. A distinct population of CD71^{++++} cells that demonstrated an average of six additional doublings in suspension culture and erythroid colony formation in methylcellulose was isolated. Suppression subtractive hybridization of mRNA isolated from those cells permitted the identification of transcribed genes. A summary of 719 expressed sequence tags (ESTs) describing 505 independent transcripts is provided here with a full analysis of each EST available at http://hembase.niddk.nih.gov. Several transcripts that matched genes previously reported in the context of erythroid differentiation including 4 cell surface proteins were expressed at this developmental stage. Active chromatin remodeling was suggested by the identification of 4 histone proteins, 4 high-mobility group proteins, 13 transcription factors, and 6 genes involved in DNA recombination and repair. Numerous genes associated with leukemic translocations were also recognized including topoisomerases I and II, nucleophosmin, Translin, EGR1, dek, pim-1, TFG, and MLL. In addition to known transcripts, 44 novel EST were discovered. This transcriptional profile provides the first genomic-scale description of gene activity in erythroid progenitor cells.

INTRODUCTION

In addition to the genes involved directly in hemoglobin synthesis, the human genome must encode an array of genes required to maintain the orderly balance between growth and differentiation that results in the production of 10^{11} erythrocytes during each day of adult life. Erythropoietin (EPO)-dependent proliferation occurs primarily in cells referred to as progenitors, and transcriptional cascades initiated in those cells are thought to govern the terminal differentiation of more mature erythroid cells (Orkin and Zon, 1997). Erythropoietin is produced by the kidney in response to a local reduction in oxygen (Jacobson *et al.,* 1957). The hormone initiates its biologic effects by binding to surface protein receptors that transduce signals to the interior of the cell (Damen and Krystal, 1996; Spivak *et al.,* 1991). The crucial role for EPO in murine erythropoiesis has been defined *in vivo* by the introduction of null mutations to the murine EPO gene (Wu *et al.,* 1995). Whereas primitive erythropoiesis was not affected, no definitive erythropoiesis was detected in EPO^{-} animals beyond the late progenitor cell stage.

CD71, the cellular receptor for diferric transferrin, is also expressed on erythroid progenitor cells (Huebers, 1987). Unlike the low level of EPO receptor expression estimated at 50–1000 receptors per erythroid progenitor (Wognum *et al.,* 1990), 300,000–800,000 transferrin receptors per cell are present on actively proliferating erythroid progenitor cells (Iacopetta *et al.,* 1982). The extremely high level of transferrin receptor expression in erythroid progenitors has been associated with their rapid proliferation and iron requirements for hemoglobin synthesis (Shintani *et al.,* 1994). Within the erythroid lineage, CD71 levels are highest on late progenitor cells (Loken *et al.,* 1987). As erythroid cells differentiate into precursors cells, CD71 levels diminish, and the receptor is not detectable on mature erythrocytes. The CD71 receptor is expressed on other proliferating hematopoietic progenitors, but expression at levels three logs above the isotypic control is regularly found only among erythroid progenitors (Gross *et al.,* 1997; Mayani and Lansdorp, 1995).

Here we describe a transcriptional profile of erythroid cells cultured from donated peripheral blood that share the dual criteria of EPO-dependent proliferation and high-level CD71 expression. The cells exhibited an average of six additional cell divisions when left in liquid culture and erythroid colony formation in

[1] To whom correspondence should be addressed at Laboratory of Chemical Biology, Building 10, Room 9N308, National Institutes of Health, Bethesda, MD 20892. Telephone: (301) 402-2373. Fax: (301) 402-0101. E-mail: jm7f@nih.gov.

0888-7543/99

methylcellulose. Sequencing of a subtracted, normalized library was performed to provide a general description of genes transcribed in those cells. Seven hundred nineteen expressed sequenced tags (ESTs) are provided here as a transcriptional profile of proliferation and lineage-specific differentiation that underlie the proliferative phase of human erythropoiesis.

MATERIALS AND METHODS

Cell culture. Buffy coats were obtained from blood samples from normal adult human donors with informed consent. Mononuclear cells were isolated after Ficoll-Hypaque gradient separation and cultured using a two-phase model described previously (Fibach *et al.*, 1989; Miller *et al.*, 1998). Briefly, peripheral blood mononuclear cells were cultured at 37°C in minimal medium (Sigma, St. Louis, MO), supplemented with 10% fetal bovine serum (Intergen, Purchase, NY). After 7 days in the first culture phase, the cells were washed and placed in phase II medium, supplemented with 30% fetal calf serum, 10% deionized bovine serum albumin, 10^{-5} M β-mercaptoethanol, 10 mg/ml holotransferrin, and 10^{-6} M dexamethasone (Sigma). All buffy coats in phase II were grown in matching flasks with and without supplemental 1 U/ml EPO (Ortho Pharmaceuticals, Raritan, NJ). Sorted populations were plated in Methocult GFH4434 according to the manufacturer's instructions (Stem Cell Technologies, Vancouver, British Columbia, Canada).

Antibody staining and flow cytometry. Antibody staining was carried out by adding 10 μl of the antibody to approximately 10^5–10^6 cells in 100 μl of culture medium and incubating at 4°C for 30 min. The cells were washed in PBS and then fixed with 2% paraformaldehyde prior to analysis. Positive staining was defined as fluorescence at levels greater than 2 standard deviations above the isotypic control. The monoclonal antibodies used for phenotyping were as follows: CD7I-FITC cloneYDJ1.2.2 (Immunotech, Miami, FL), Glycophorin A-PE clone 11E4B-7-6 (Immunotech), CD13 clone SJ1D1 (Immunotech), CD34-PE clone S81 (Immunotech), CD10-PE clone HI10A (Pharmingen, San Diego, CA), CD36-PE clone SMO (Sigma), and CD48 clone MEM102 (RDI, Flanders, NJ). All flow cytometry was performed with an Epics Elite flow cytometer (Coulter, Hialeah, FL) equipped with an air-cooled argon laser at 488 nm. Cell gating and two-color analyses of CD71 and CD48 were chosen to select $CD71^{++++}$ populations for fluorescence-activated cell sorting. Sorting parameters were selected by High Purity Logic software provided by the manufacturer. In addition, the manufacturer's software was used for all statistical analyses of the flow cytometric data.

Construction of subtraction library. Total RNA from cultured blood cells (phase II, day 5) was purified using TRIzol reagent (Life Technologies, Gaithersburg, MD) according to the manufacturer's specifications. Approximately 60,000 sorted $CD71^{++++}$ cells grown in the presence of erythropoietin (the tester) and approximately 900,000 cells grown in the absence of erythropoietin (the driver) were used for total RNA isolation. Total RNAs were converted into cDNAs using MMLV reverse transcriptase and a CapFinder PCR Synthesis Kit (Clontech, Palo Alto, CA) using 19 cycles of PCR for the tester and 23 cycles for the driver under recommended conditions. The subsequent steps were performed according to the user manuals for the CapFinder PCR Synthesis Kit and PCR-Select cDNA Subtraction Kits (Clontech, Palo Alto, CA), respectively. The subtracted cDNA was cloned directly into pCR2.1 TA cloning vector (Invitrogen, Carlsbad, CA), and the ligation mixture was used to transform *Escherichia coli* DH5 alpha maximum efficiency competent cells (Life Technologies). The library was plated onto 10-cm ampicillin-containing agar plates supplemented with X-gal and IPTG, and bacteria were grown overnight at 37°C in an air incubator. White colonies were picked and grown individually at 37°C overnight in a shaker incubator in 3 ml of LB medium containing 50 μg/ml of carbenicillin.

Sequencing. Plasmid DNAs were prepared from picked clones using Qiawell 8 Plus Plasmid Kit (Qiagen, Valencia, CA). The clones were sequenced with an ABI Prism Cycle Sequencing Kit (PE Applied Biosystems, Foster City, CA) using AmpliTaq FS Polymerase and a primer, 5′ TCA CAC AGG AAA CAG CTA TGA C 3′. Sequencing reactions were analyzed using an ABI 377 or 373 automated sequencers kit (PE Applied Biosystems) following the manufacturer's protocols.

Virtual Northern blot. Total RNA was purified using TRIzol reagent (Life Technologies) according to the manufacturer's protocol. Approximately 10^7 cells were used for total RNA isolation from cells grown in the presence or in the absence of 1 U/ml EPO. RNA was converted into cDNA using the CapFinder PCR Synthesis Kit (Clontech) employing 19 cycles of PCR and the manufacturer's recommended conditions. Two hundred fifty nanograms of each amplified cDNA from EPO^+ and EPO^- cells was run on a 1% agarose gel and transferred to a nylon membrane using the Turboblotter equipment and supplied protocol (Schleicher & Schuel, Keene, NH). The DNA was fixed to the membrane using a UV crosslinker (Stratagene, La Jolla, CA). The DIG-labeled probes were denatured for 10 min in boiling water and rapidly chilled on ice. The prehybridization and hybridization steps and chemiluminescence detection were performed as suggested by the manufacturer (Boehringer Mannheim, Indianopolis, IN).

RESULTS

Identification of EPO-Responsive Cells

A two-phase liquid culture was utilized to establish a method for identifying EPO-responsive populations of cells among peripheral blood mononuclear cells. The culture method has been used extensively for culturing erythroid cells in part because globin expression patterns mimic those found *in vivo* (Dalyot *et al.*, 1992). Multiple confounding variables inherent to this model including the use of fetal calf serum and conditioned medium in the first culture phase were controlled by comparing EPO supplemented cultures and those not supplemented with EPO. In this way, EPO became the single variable parameter in the culture model. EPO-responsive populations were identified and isolated on the basis of differential CD71 expression between matching EPO^+ and EPO^- cultures. After several days in medium supplemented with 1 U/ml EPO, a distinct population of cells expressing CD71 at levels above 100 fluorescence units was detected by flow cytometry (Fig. 1A). A comparable population of cells expressing CD71 at that level, designated $CD71^{++++}$ cells, was not detected in EPO^- cultures (Fig. 1B).

Figure 2A demonstrates the remarkable specificity gained by this strategy. The addition of EPO resulted in the detection of $CD71^{++++}$ cells after 4 days at a frequency of 0.4%. The frequency continued to rise during the subsequent 72 h, peaking at 2.0%. After several additional days in EPO, the percentage of $CD71^{++++}$ cells dropped as precursor populations having a lower level of CD71 expression were generated. In the absence of erythropoietin, an average of only 1 per 20,000 cells appeared in the gated $CD71^{++++}$ cytometric window throughout the culture period. Our data suggest that $CD71^{++++}$ populations were generated specifically in response to EPO during the culture period, and this strategy was used in all subsequent experiments to identify EPO-responsive cells with a

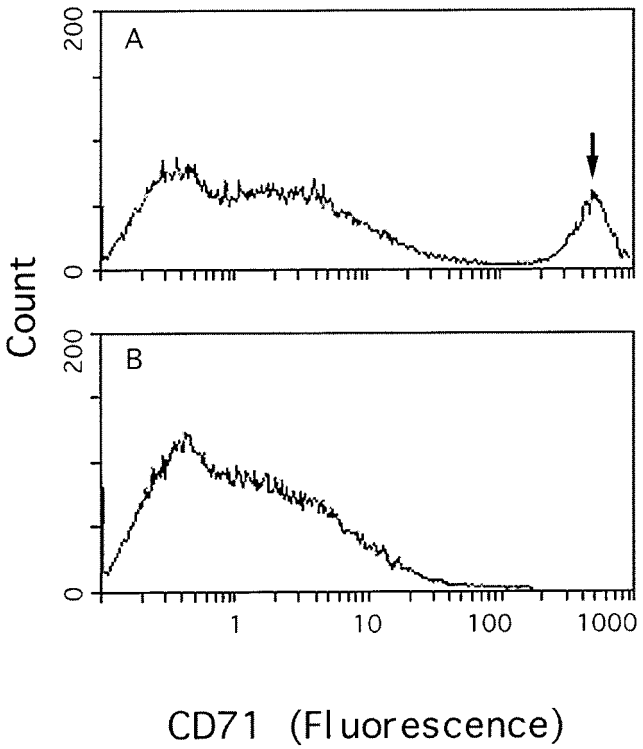

FIG. 1. Detection of EPO-stimulated, $CD71^{++++}$ cells. Flow cytometric histogram of suspension cultured cells stained with anti-CD71-FITC antibody. Cells cultured in the presence (**A**) or in the absence (**B**) of 1 U/ml EPO in the culture supernatants are shown. Histograms depict the CD71 fluorescence distribution (*x*-axis) and cell count (*y*-axis) appearing in the blast gate 5 days after the addition of EPO to the culture medium. 100,000 total events were analyzed for each sample. Arrow points to the population of cells referred to as $CD71^{++++}$.

predicted detection rate of false-positive events of approximately 0.005%.

Rapid proliferation among the EPO-responsive cells population occurred between days 4 and 7 in suspension culture as shown in Fig. 2A. Phenotyping of those populations by two-color flow cytometry is summarized in Fig. 2B. CD34 and glycophorin A immunophenotyping of the $CD71^{++++}$ population was performed for comparison with the results of others (Nakahata and Okumura, 1994). CD34 expression is reported lost early, and glycophorin A is not expressed on cells before the proerythroblast stage of development. CD34 expression in the EPO-responsive cells studied here decreased from 48 ± 12% on day 4 to 15 ± 11% by day 7. The portion of cells expressing detectable glycophorin A rose rapidly from 8 ± 7% on day 4 to 61 ± 20% by day 7. The reciprocal patterns of CD34 and glycophorin A expression mark the transition from progenitor to precursor erythroid populations detected here and by others (Loken *et al.*, 1987). The expression pattern of CD36 (thrombospondin receptor for GPIV) provided further evidence that the $CD71^{++++}$ population of cells exhibited the phenotype of erythroid progenitor cells. CD36 expression in erythroid progenitor cells is thought to increase as CD34 is lost. The detection of CD36 on erythroid progenitors precedes that of glycophorin A (Okumura *et al.*, 1992). As predicted, CD36 was present on 67 ± 27% of the $CD71^{++++}$ cells on day 4 and greater than 90% of the population thereafter. Other markers used for flow cytometry to identify nonerythroid cells (CD10, CD13, and CD48) were not detected at levels significantly higher than those for isotypic controls.

Morphology and Proliferation of EPO-Responsive, $CD71^{++++}$ Cells

Immunophenotyping indicated that a broad range of erythroid progenitors balanced between the $CD34^{+}$ and glycophorin A^{+} populations existed among the $CD71^{++++}$ cell populations after 5 days in 1 U/ml EPO (Fig. 2B). Those cells were sorted to explore their morphologic, proliferative, and genetic characteristics. A uniform $CD71^{++++}$ phenotype was confirmed by postsort analysis (data not shown). Giemsa staining of the sorted cells revealed a light-microscopic appearance (Fig. 3) much like that described previously for human erythroid progenitors including prominent Golgi and basophilic agranular cytoplasm (Sawada *et al.*, 1987). Chromatin obscured the nucleoli of many blasts, and nuclear indentations were common at the Golgi pole. Nuclear size was variable with occasional cells having large nuclei over 20 μm in diameter (Fig. 3A). Cytoplasmic blebbing and membrane vesiculations characterized many of the blasts, with occasional cells exhibiting gross nuclear herniation on cytospun preparations (Fig. 3B). Proerythroblasts with more condensed chromatin and less membrane blebbing or vesiculation were also observed (Fig. 3D).

Quantitation of the proliferative potential of the $CD71^{++++}$ cells was determined by analysis of proliferation in suspension culture and colony assays in methylcellulose. Resuspension of the sorted $CD71^{++++}$ cells in liquid culture medium was performed to determine the percentage of cells that differentiate into glycophorin A-positive, erythropoietic cells. After 1 additional week in suspension culture, an increase in glycophorin A expression associated with lower CD71 expression was detected on 80–90% of the sorted cells. An average 72 ± 10-fold increase in the number of glycophorin A-expressing cells occurred over that period. Colony growth and morphology were also measured after the sorted $CD71^{++++}$ cells were plated in methylcellulose (Fig. 4). Approximately 2–10% of the sorted $CD71^{++++}$ cells formed large colonies (estimated as >50 cells) after 5–7 says. A doubling time of less than 24 h was easily demonstrated by the counting the cells on progressive days (Figs. 4A–4F). The number of cells generated per colony varied widely from several cells to several hundred cells (Figs. 4G and 4H). Little additional growth was detected in the majority of colonies after 1 week, and a clustered pattern of proliferation in methylcellulose was observed in over 90% of the colonies. The surface phenotype, morphology, and the pattern of growth in suspension as well as methyl-

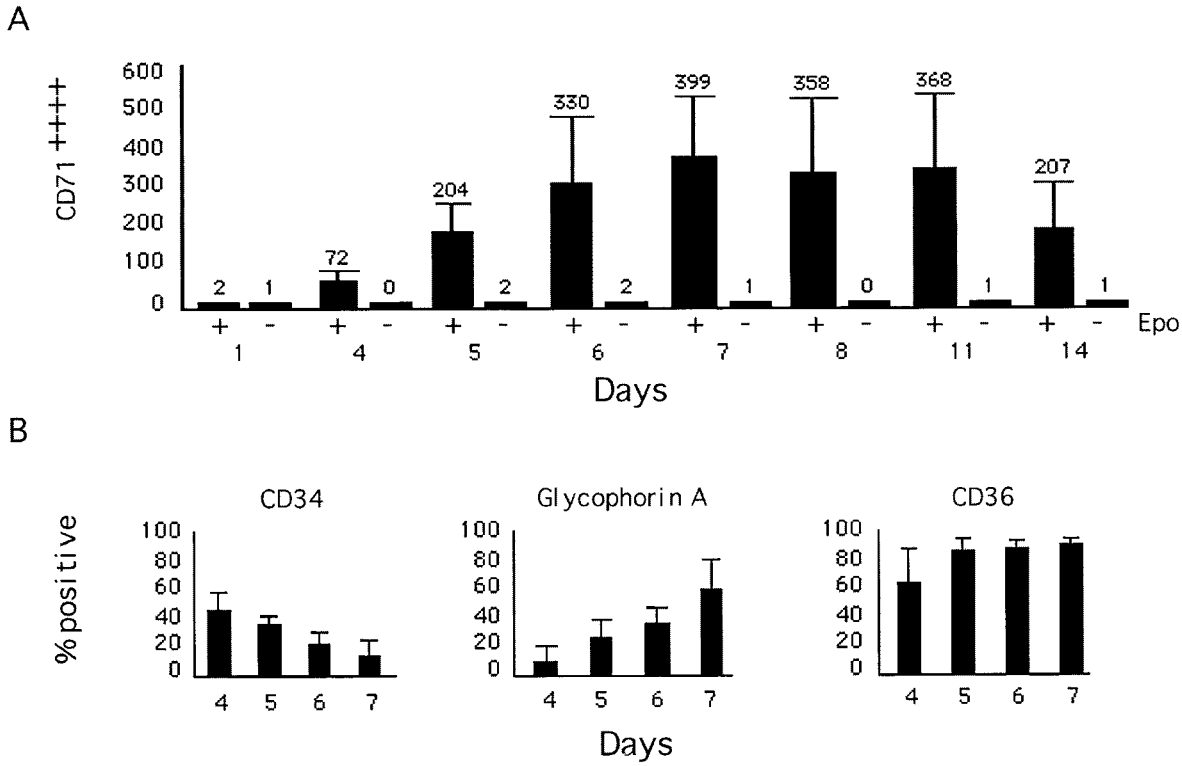

FIG. 2. EPO-dependent generation and phenotype of $CD71^{++++}$ cells. (**A**) The average number of $CD71^{++++}$ cells appearing in the blast gate among 20,000 total events is shown above each bar. Samples from three separate buffy coats were analyzed on days 1 through 14 in phase II culture medium. Matched populations cultured in the presence (+) or in the absence (−) of 1 U/ml supplemental EPO in the culture medium are shown. (**B**). Immunophenotyping of the $CD71^{++++}$ cells detected on days 4–7 in phase II medium supplemented with 1 U/ml EPO. The mean percentage of the cells expressing CD34, Glycophorin A, and CD36 in the $CD71^{++++}$ population described in (**A**). The mean percentages ± standard deviation for triplicate experiments are shown.

cellulose suggested that $CD71^{++++}$ EPO-responsive cell populations represent a highly proliferative population of erythroid progenitor cells.

Gene Expression Profile

The strategy used to identify genes with upregulated transcription in the EPO-responsive cells involved suppression subtractive hybridization (Diatchenko *et al.*, 1996). The method permitted the subtraction and normalization of transcripts present in the proliferating erythroid cells. A total of 60,000 EPO-stimulated cells generated after 5 days in EPO (Figs. 3 and 4) were sorted, and mRNA was isolated. Approximately 10^6 cells cultured in the absence of EPO were used to provide an excess of mRNA to reduce the number and redundancy of commonly expressed genes. An estimate of subtraction efficiency (manufacturer's protocol) suggested a significant reduction in abundant transcripts after subtraction as has been reported elsewhere (Hoon and Ryba, 1997). Novel sequences were less redundant (4%) than those matching characterized genes (25%). A complete database describing the sequence and homology analyses of the 719 sequenced EST is provided on the Internet at http://hembase.niddk.nih.gov. Five hundred five independent clones were identified including 291 characterized genes, 170 uncharacterized genes that match genomic sequence or other public ESTs, and 44 novel transcripts not reported elsewhere. Those sequences with statistical probabilities of $\geq 10^{-5}$ homology after BLAST analyses were defined here as novel.

All categories of cellular function were represented among the 219 of 505 ESTs matching previously characterized genes (Table 1). Elongation factor 1-α was the most common gene matched in this and other EST expression profiles (Nelson *et al.*, 1998). It has also been suggested that increased expression of this ubiquitous gene correlates with cytoskeletal disruptions and membrane blebbing among erythroid blasts like those seen in Fig. 3 (Kato *et al.*, 1997). Several genes previously reported within the context of erythroid disease or differentiation were identified by the ESTs in Table 2. As expected, α-globin, β-globin and δ-globin genes were present, but globin transcripts were relatively rare (1% of the sequenced clones), suggesting adequate normalization of the library. Interestingly, the surface proteins CD36, CD71, and glycophorin A used to phenotype these cells were detected in the transcriptional profile. Expression of the fibronectin receptor has also been reported to have a role in erythroid progenitor cell migration (Goltry and Patel, 1997). The lack of ESTs matching cell surface receptors that are expressed on other peripheral blood cells is consistent with a lack of nonerythroid cells among the

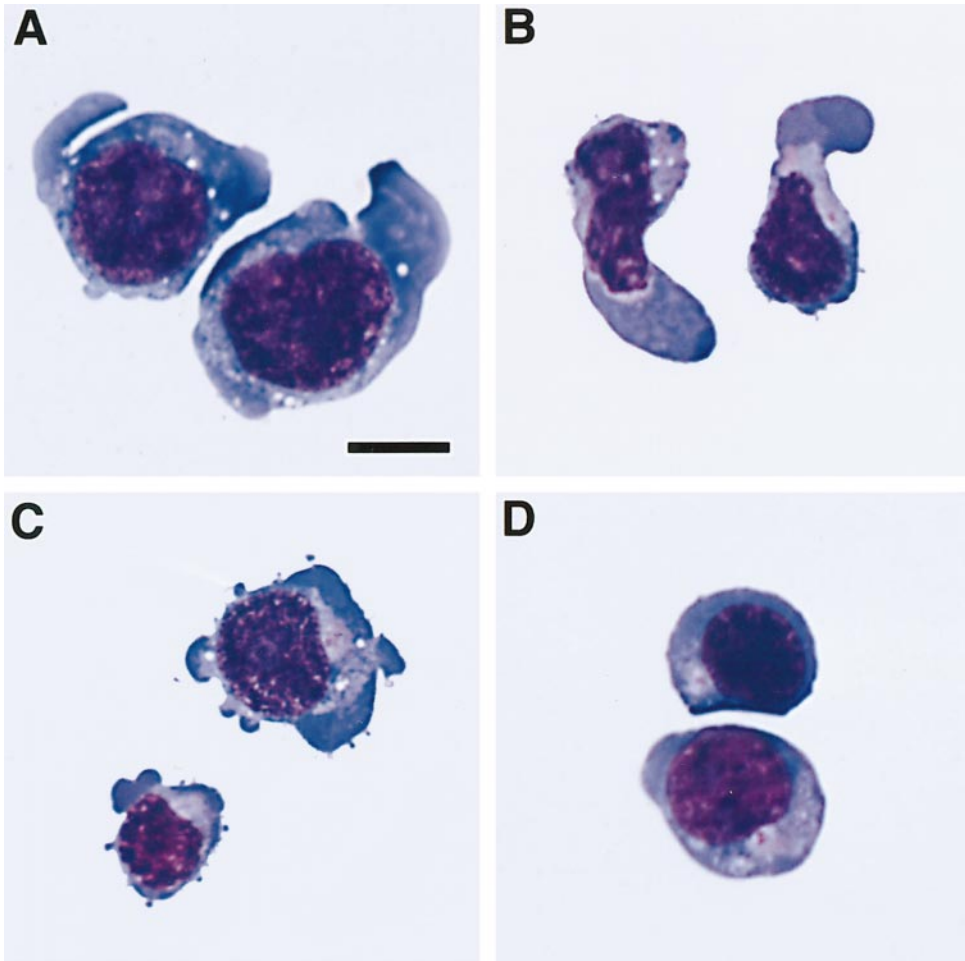

FIG. 3. Morphology of EPO-stimulated CD71^{++++} Cells. (**A–D**) Morphologies of CD71^{++++} cells sorted 5 days after the addition of 1 U/ml EPO to the culture medium. (Bar = 10 μm.)

CD71^{++++} EPO-responsive population. Five clones matched KIAA genes transcribed in the myeloblasts of a patient with erythroleukemia (Koeffler and Golde, 1978).

Approximately one-third of the ESTs matching characterized genes are involved in replication and transcription. Among that group are several genes involved in all aspects of chromatin remodeling (Table 2). Included are several histone and high-mobility group (HMG) proteins. Surprisingly, the histone H4 transcripts in the profile (ad09a12 and ad12f12) were closely homologous to genomic sequence near the hemochromatosis gene locus. Several genes involved in DNA recombination and repair genes were also identified. Thirteen ESTs share significant homology with known transcription factors including multiple growth-associated factors and members of the trithorax (HRX) and polycomb (M96) groups. Many of these factors as well as the HMG genes have been previously associated with globin gene regulation. However, the relevance of the other transcription factors and the DNA repair genes for globin gene regulation versus other aspects of erythroid cell differentiation is currently not known.

We produced virtual Northern blots (Fig. 5) (Larrick, 1998) to validate our strategy for identifying gene products induced by erythropoietin. As expected, β-globin expression was detected only in the EPO^{+} cDNA sample (Fig. 5, lane 2). Illegitimate transcription from the globin gene locus reported elsewhere (Humphries *et al.*, 1976) was not detected here in the absence of EPO. The increased hybridization of the transferrin receptor transcript (Fig. 5, lane 3) was consistent with the flow cytometric and transcriptional data described above. The similar intensities from β-actin expression in the presence or in the absence of EPO (Fig. 5, lane 1) suggest incomplete subtraction of some highly expressed housekeeping genes by this method since β-actin was identified in the database. The expression patterns of five clones containing novel sequences are also shown. Expression of two novel clones was detected

FIG. 4. Proliferation of sorted cells in methylcellulose. **A–F** were photographed on successive days after sorted EPO-responsive, CD71^{++++} cells were plated in methylcellulose. (**A**) 18 h; (**B**) 1.5 days; (**C**) 2.5 days; (**D**) 3.5 days; (**E**) 7.5 days; (**F**) 10.5 days. **G** (7.5 days) and **H** (10.5 days) were photographed at 5× magnification for a comparison of multiple colonies.

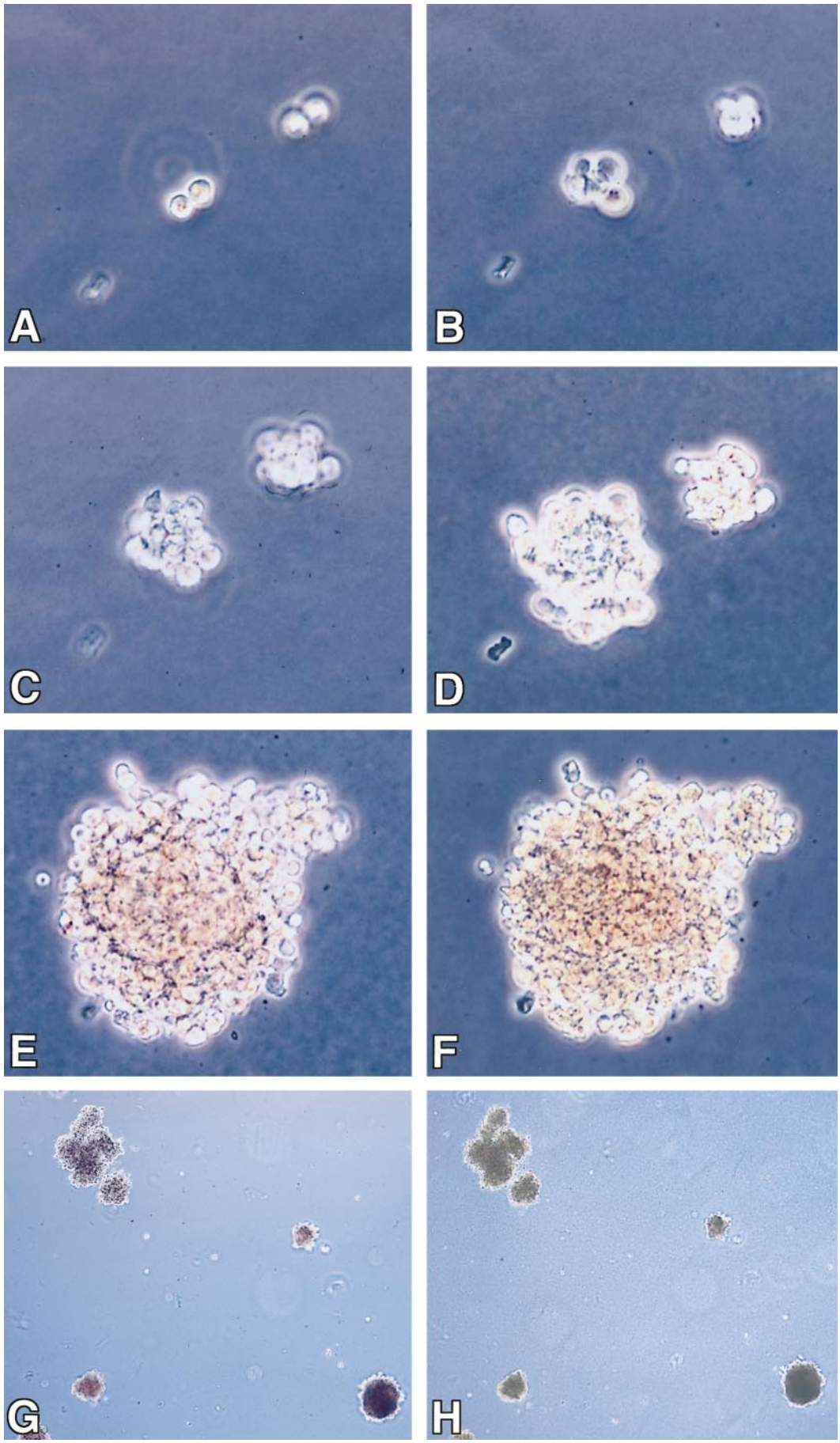
A
B
C
D
E
F
G
H

TABLE 1

Summary of ESTs Identified in This Study That Match Gene Products with Known Function

Category	EST matches
Replication/transcription	85 (29%)
Signaling	35 (12%)
Translation	64 (22%)
Metabolism/enzymes	80 (28%)
Other	27 (9%)
Total	291 (100%)

only in the EPO^+ sample (Fig. 5, lanes 4 and 7). Upregulated expression of two other novel clones in the presence of erythropoietin was demonstrated (Fig. 5, lanes 5 and 6). Expression of a fifth novel clone (Fig. 5, lane 8) was below the detection limits, suggesting low abundance of the transcript. Multiple hybridized bands like those seen in lanes 4 and 7 could be attributed to several causes including cDNA from a multigene family or alternatively spliced or truncated gene products. Six of eight gene products examined by this method were detectable and appeared to be induced by erythropoietin.

DISCUSSION

We have begun investigating the overall expression of genes during normal human erythropoiesis with the eventual goal of developing novel genetic-based therapies for diseases involving erythroid cells. Unlike whole tissue or sorted cell transcriptional analyses (Adams *et al.*, 1993; Zhang *et al.*, 1997), we have demonstrated a simple strategy for generating a transcriptional profile on the basis of a lineage-specific biological response (EPO-dependent proliferation) in addition to surface phenotype. The rapid proliferation of this population in both suspension culture (average $\geq$6 doublings in suspension culture) and clustered erythroid colonies formed in methylcellulose strongly suggests that the uniform population of $CD71^{++++}$ EPO-responsive cells studied here is made up of actively proliferating erythroid progenitor cells. Our data confirm the expression of several genes in primary human erythroid cells that have previously been reported in transformed cell lines or nonhuman models. In addition, we provide an Internet-linked database that describes several hundred additional genes for further study. The transcriptional profile reported here represents a first step toward translating a knowledge of human erythroid cell biology into the language of molecular genomics.

All functional aspects of human erythropoiesis are described within the 719 EST profile. The successful identification of erythroid disease-related genes is demonstrated by the identification of transcripts previously reported in the context of thalassemias, HEMPAS (hereditary erythroblastic multinuclearity with positive acidified serum lysis test), porphyria, and hemolytic anemias (Table 2). Our use of EPO responsiveness in addition to cellular phenotype resulted in a lineage- and stage-specific gene profile quite distinct from that reported for $CD34^+$ cord blood cells (Mao *et al.*, 1998). The absence of genes with expression limited to other nonerythroid hematopoietic cells further supports our stringent cell purification strategy. Proteins involved in all aspects of chromatin remodeling were prevalent within the profile. The likeness of histone H4 transcripts to a particular genomic locus suggests that at least one member of the H4 multigene family is differentially regulated during erythroid development. Interestingly, knockout of the DNA helicase associated with Bloom syndrome has recently been shown to cause profound anemia in mice (Chester *et al.*, 1998). Four separate ESTs matched a gene encoding the dbpB-like protein, a CCAAT binding factor associated with repression of γ-globin transcription (Horwitz *et al.*, 1994; Tang *et al.*, 1997). Also identified was a human homologue resembling the murine cerebellar enriched transcriptional factor NFI-B. The NFI-B sequence (ad00179) shares no homology to the human NFI gene family. Expression of an erythroid-restricted member of the NFI transcription factor family may be relevant to globin gene expression since NFI binding to the γ-globin locus may interfere with interactions between GATA-1 and Sp1 (Fischer *et al.*, 1993). Other transcription factors specifically associated with erythropoiesis, like GATA-1 (Zon *et al.*, 1990) and EKLF (Bieker, 1996), were not identified among the 719 ESTs.

In addition to many of the genes identified in Table 2, several genes involved in malignant translocations were matched, including nucleophosmin (ad09h09), Translin (ad12e04), dek (ad13g06), pim-1 (ad11d05), and TFG (ad03c06). Coexpression of several of these growth-related genes has been confirmed by virtual Northern blotting of transcripts from EPO-responsive cells cultured without conditioned medium (unpublished data). Since erythropoiesis represents one of the most durable and robust models of nonmalignant cell proliferation in adult animals, it is not entirely surprising that a large group of cancer-related genes was identified. Surprisingly, a matching group of growth-limiting genes was not obvious. Those genes may be transcribed in more mature cells that undergo terminal erythroid differentiation. The genomic-scale silencing of transcription that occurs during terminal erythroid differentiation may explain the low incidence of human erythroleukemia, as well as the extreme difficulty in transforming β-globin-expressing cells in the laboratory.

While a detailed summary of several hundred genes is inherently difficult within a single article, this report and the corresponding Internet-linked database should significantly broaden the genetic description of human erythropoiesis. In addition to several EST-matching genes expected to be expressed in these cells, we have identified and provided a detailed description of sev-

TABLE 2

Summary of ESTs Matching Gene Products Specifically Involved in Erythropoiesis or Chromatin Remodeling

EST	Category	Probability
	Erythroid-specific	
ad08b02	Human mRNA for α globin (thalassemia)	00
ad09b07	Human mRNA for β globin (thalassemia)	00
ad13c05	Human mRNA for ferrochelatase (protoporphyria)	00
ad02f03	Human triosephosphate isomerase (hemolytic anemia)	00
ad11d11	Human mRNA for α mannosidase II (HEMPAS)	00
ad01e06	Human mRNA for transferrin receptor	00
ad10h09	Human mRNA for carbonic anhydrase I	00
ad09b04	Human mRNA for CD36	00
ad06b09	Human mRNA for glycophorin A, MN types	00
ad09d05	Human ankyrin (ANK1) gene (hereditary spherocytosis)	−83
ad00047	Human mRNA for fibronectin receptor subunit β	−86
	Chromatin remodeling	
Histone proteins		
ad09a12	Human H4/g gene for H4 histone	−171
ad03f07	Human mRNA for histone (H2A.Z)	−127
ad00023	Human H3.3 gene	−120
ad08f02	Human mRNA for histone macroH2A1.2	−103
High-mobility group proteins		
ad14e05	Human mRNA for high-mobility group (HMG) box (SSRP1)	−158
ad11e12	Human mRNA for chromosomal protein HMG-17	−138
ad04f07	Human mRNA for high-mobility group-1 protein (HMG-1)	−126
ad14g10	Human mRNA for chromosomal protein HMG-14	−66
Transcription factors		
ad12h09	Human mRNA for germline HRX	00
ad12g08	Human mRNA for α NAC	00
ad08f04	Human mRNA for fos proto-oncogene (c-fos)	00
ad03a06	Human mRNA for dbpB-like protein	−172
ad00086	Human mRNA for G0S3	−151
ad06c03	Human mRNA for nuclear factor NF45	−139
ad05e09	Human mRNA for znf6	−136
ad00084	Human mRNA for c-myb	−128
ad12b05	Human mRNA for ERF-1	−126
ad01g07	Human mRNA for early growth response protein 1	−116
ad00179	Mouse mRNA for NFI-B protein	−113
ad09h06	Human mRNA for SON protein	−87
ad06g08	Human mRNA for PHD finger DNA binding protein M96	−18
	DNA recombination and repair proteins	
ad12e04	Human mRNA for translin	00
ad02a10	Human mRNA for DNA mismatch repair (hmlh1)	00
ad14h12	Human mRNA for XP-C repair complementing protein	00
ad13f02	Human mRNA for DNA topoisomerase II	00
ad11g11	Human mRNA for Bloom syndrome protein (BLM)	−156
ad11c01	Human mRNA for DNA topoisomerase I	−97

Note. Probability is the lowest *P* value identified by BLAST search of public databases. The transcripts are shown in order of probable homology between the ESTs and the corresponding gene product. Erythroid disease association is shown in parentheses. (Probabilities are expressed as *e* values. 00 identifies a perfect match.)

eral hundred additional genes active in the cells. As demonstrated in Fig. 5, the use of suppression subtraction methods resulted in the identification of differentially expressed genes. Particularly interesting are 44 novel transcripts that have an erythroid-restricted expression when compared to the 2,000,000 other ESTs currently deposited in public databases. However, based on the low frequency of exclusive gene expression among other tissues and low-level hybridization of two clones with cDNA from cells cultured in the absence of EPO (Fig. 5, lanes 5 and 6), we predict that many of the novel ESTs identified here are expressed in a differential rather than an exclusive pattern in erythroid cells. A full map of the human genome will soon provide the means to correlate these ESTs with linkage analyses involving patients diagnosed with inherited and acquired erythroid diseases. The correlation of ESTs in this profile with genomic loci linked to increased γ-globin gene expression should be useful for understanding polygeneic influence on the clinical pre-

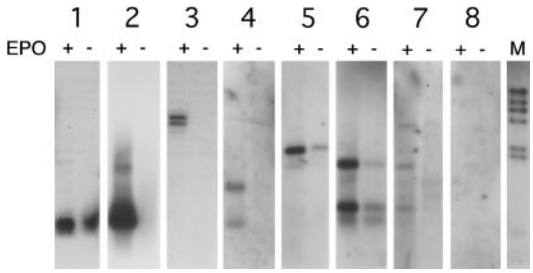

FIG. 5. Virtual Northern blot analysis of expressed cDNA fragments. Virtual Northern blots of three controls (lane **1,** β-actin; lane **2,** β-globin; lane **3,** transferrin receptor) and five novel transcripts (lanes **4–8**) are shown. **M,** DIG-labeled λ-DNA, cleaved with *Hin*dIII molecular weight markers.

sentation of patients with hemoglobinopathies (Chang *et al.*, 1995). Once the full repertoire of erythroid genes is identified, genomic-based strategies for the diagnosis and treatment of erythroid diseases may be conceivable.

ACKNOWLEDGMENTS

We thank Sudha Koduru, Lindsey Mitchell, and Terry Lee for assisting with data collection and analyses, and Dr. Susan Leitman in the Department of Transfusion Medicine for providing the buffy coats. We thank N. Dietrich, J. Touchman, M. Fergusson, E. Sorbello, A. Gupta, R. Torkzadeh, C. Varner, M. Walker, and S. Beckstrom-Sternberg of the NIH Intramural Sequencing Center for technical and computational assistance. We also thank Dr. Alan Schechter for many helpful discussions and critically reading the manuscript.

REFERENCES

Adams, M. D., Kerlavage, A. R., Fields, C., and Venter, J. C. (1993). 3,400 new expressed sequence tags identify diversity of transcripts in human brain. *Nat. Genet.* **4:** 256–267.

Bieker, J. J. (1996). Isolation, genomic structure, and expression of human erythroid Kruppel-like factor (EKLF). *DNA Cell Biol.* **15:** 347–352.

Chang, Y. C., Smith, K. D., Moore, R. D., Serjeant, G. R., and Dover, G. J. (1995). An analysis of fetal hemoglobin variation in sickle cell disease: The relative contributions of the X-linked factor, beta-globin haplotypes, alpha-globin gene number, gender, and age. *Blood* **85:** 1111–1117.

Chester, N., Kuo, F., Kozak, C., O'Hara, C. D., and Leder, P. (1998). Stage-specific apoptosis, developmental delay, and embryonic lethality in mice homozygous for a targeted disruption in the murine Bloom's syndrome gene. *Genes Dev.* **12:** 3382–3393.

Dalyot, N., Fibach, E., Rachmilewitz, E. A., and Oppenheim, A. (1992). Adult and neonatal patterns of human globin gene expression are recapitulated in liquid cultures. *Exp. Hematol.* **9:** 1141–1145.

Damen, J. E., and Krystal, G. (1996). Early events in erythropoietin-induced signaling. *Exp. Hematol.* **24:** 1455–1459.

Diatchenko, L., Lau, Y. F., Campbell, A. P., Chenchik, A., Moqadam, F., Huang, B., Lukyanov, S., Lukyanov, K., Gurskaya, N., Sverdlov, E. D., and Siebert, P. D. (1996). Suppression subtractive hybridization: A method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc. Natl. Acad. Sci. USA* **93:** 6025–6030.

Fibach, E., Manor, D., Oppenheim, A., and Rachmilewitz, E. A. (1989). Proliferation and maturation of human erythroid progenitors in liquid culture. *Blood* **73:** 100–103.

Fischer, K. D., Haese, A., and Nowock, J. (1993). Cooperation of GATA-1 and Sp1 can result in synergistic transcriptional activation or interference. *J. Biol. Chem.* **268:** 23915–23923.

Goltry, K. L., and Patel, V. P. (1997). Specific domains of fibronectin mediate adhesion and migration of early murine erythroid progenitors. *Blood* **90:** 138–147.

Gross, S., Helm, K., Gruntmeir, J. J., Stillman, W. S., Pyatt, D. W., and Irons, R. D. (1997). Characterization and phenotypic analysis of differentiating $CD34^+$ human bone marrow cells in liquid culture. *Eur. J. Haematol.* **59:** 318–326.

Hoon, M. A., and Ryba, N. J. (1997). Analysis and comparison of partial sequences of clones from a taste-bud-enriched cDNA library. *J. Dent. Res.* **76:** 831–838.

Horwitz, E. M., Maloney, K. A., and Ley, T. J. (1994). A human protein containing a "cold shock" domain binds specifically to H-DNA upstream from the human gamma-globin genes. *J. Biol. Chem.* **269:** 14130–14139.

Huebers, H. A. (1987). The physiology of transferrin and transferrin receptors. *Physiol. Rev.* **67:** 520–582.

Humphries, S., Windass, J., and Williamson, R. (1976). Mouse globin gene expression in erythroid and non-erythroid tissues. *Cell* **7:** 267–277.

Iacopetta, B. J., Morgan, E. H., and Yeoh, G. C. (1982). Transferrin receptors and iron uptake during erythroid cell development. *Biochim. Biophys. Acta* **687:** 204–210.

Jacobson, L. O., Goldwasser, E., Fried, W., and Plzak, L. (1957). Role of the kidney in erythropoiesis. *Nature* **179:** 633–634.

Kato, M. V., Sato, H., Nagayoshi, M., and Ikawa, Y. (1997). Up-regulation of the elongation factor-1alpha gene by p53 in association with death of an erythroleukemic cell line. *Blood* **90:** 1373–1378.

Koeffler, H. P., and Golde, D. W. (1978). Acute myelogenous leukemia: A human cell line responsive to colony-stimulating activity. *Science* **200:** 1153–1154.

Larrick, J. W. (1998). Gene Cloning and Analysis By RT-PCR (P. Siebert, Ed.), BioTechniques Molecular Laboratory Methods Series, No. 1.

Loken, M. R., Shah, V. O., Dattilio, K. L., and Civin, C. I. (1987). Flow cytometric analysis of human bone marrow. I. Normal erythroid development. *Blood* **69:** 255–263.

Mao, M., Fu, G., Wu, J. S., Zhang, Q. H., Zhou, J., Kan, L. X., Huang, Q. H., He, K. L., Gu, B. W., Han, Z. G., Shen, Y., Gu, J., Yu, Y. P., Xu, S. H., Wang, Y. X., Chen, S. J., and Chen, Z. (1998). Identification of genes expressed in human CD34(+) hematopoietic stem/progenitor cells by expressed sequence tags and efficient full-length cDNA cloning. *Proc. Natl. Acad. Sci. USA* **95:** 8175–8180.

Mayani, H., and Lansdorp, P. M. (1995). Proliferation of individual hematopoietic progenitors purified from umbilical cord blood. *Exp. Hematol.* **14:** 1453–1462.

Miller, J. L., Njoroge, J. M., Gubin, A. N., and Rodgers, G. P. (1998). Prospective identification of erythroid elements in cultured peripheral blood. *Exp. Hematol.* **27:** 624–629.

Nakahata, T., and Okumura, N. (1994). Cell surface antigen expression in human erythroid progenitors: Erythroid and megakaryocytic markers. *Leuk. Lymphoma* **13:** 401–409.

Nelson, P. S., Ng, W. L., Schummer, M., True, L. D., Liu, A. Y., Bumgarner, R. E., Ferguson, C., Dimak, A., and Hood, L. (1998). An expressed-sequence-tag database of the human prostate: Sequence analysis of 1168 cDNA clones. *Genomics* **47:** 12–25.

Okumura, N., Tsuji, K., and Nakahata, T. (1992). Changes in cell surface antigen expressions during proliferation and differentiation of human erythroid progenitors. *Blood* **80:** 642–650.

Orkin, S. H., and Zon, L. I. (1997). Genetics of erythropoiesis: Induced mutations in mice and zebrafish. *Annu. Rev. Genet.* **31:** 33–60.

Sawada, K., Krantz, S. B., Kans, J. S., Dessypris, E. N., Sawyer, S., Glick, A. D., and Civin, C. I. (1987). Purification of human erythroid colony-forming units and demonstration of specific binding of erythropoietin. *J. Clin. Invest.* **2:** 357–366.

Shintani, N., Kohgo, Y., Kato, J., Kondo, H., Fujikawa, K., Miyazaki, E., and Niitsu, Y. (1994). Expression and extracellular release of transferrin receptors during peripheral erythroid progenitor cell differentiation in liquid culture. *Blood* **83:** 1209–1215.

Spivak, J. L., Pham, T., Isaacs, M., and Hankins, W. D. (1991). Erythropoietin is both a mitogen and a survival factor. *Blood* **77:** 1228–1233.

Tang, D. C., Ebb, D., Hardison, R. C., and Rodgers, G. P. (1997). Restoration of the CCAAT box or insertion of the CACCC motif activates delta-globin gene expression. *Blood* **90:** 421–427.

Wognum, A. W., Lansdorp, P. M., Humphries, R. K., and Krystal, G. (1990). Detection and isolation of the erythropoietin receptor using biotinylated erythropoietin. *Blood* **76:** 697–705.

Wu, H., Liu, X., Jaenisch, R., and Lodish, H. F. (1995). Generation of committed erythroid BFU-E and CFU-E progenitors does not require erythropoietin or the erythropoietin receptor. *Cell* **83:** 59–67.

Zhang, L., Zhou, W., Velculescu, V. E., Kern, S. E., Hruban, R. H., Hamilton, S. R., Vogelstein, B., and Kinzler, K. W. (1997). Gene expression profiles in normal and cancer cells. *Science* **276:** 1268–1272.

Zon, L. I., Tsai, S. F., Burgess, S., Matsudaira, P., Bruns, G. A., and Orkin, S. H. (1990). The major human erythroid DNA-binding protein (GF-1): Primary sequence and localization of the gene to the X chromosome. *Proc. Natl. Acad. Sci. USA* **87:** 668–672.