Development of a murine hematopoietic progenitor complementary DNA microarray using a subtracted complementary DNA library

Xianyong Ma, Tupur Husain, Hui Peng, Sharon Lin, Olga Mironenko, Noel Maun, Steven Johnson, David Tuck, Nancy Berliner, Diane S. Krause, and Archibald S. Perkins

With the goal of creating a resource for in-depth study of myelopoiesis, we have executed a 2-pronged strategy to obtain a complementary DNA (cDNA) clone set enriched in hematopoietic genes. One aspect is a library subtraction to enrich for underrepresented transcripts present at early stages of hematopoiesis. For this, a hematopoietic cDNA library from primary murine bone marrow cells enriched for primitive progenitors was used as tester. The subtraction used 10,000 known genes and expressed sequence tags (ESTs) as driver. The 2304 randomly picked clones from the subtracted cDNA libraries represent 1255 distinct genes, of which 622 (50%) are named genes, 386 (30%) match uncharacterized ESTs, and 247 (20%) are novel. The second aspect of our strategy was to complement this subtracted library with genes known to be involved in myeloid cell differentiation and function. The resulting cDNAs were arrayed on polylysine-coated glass slides. The microarrays were used to analyze gene expression in primary and cultured murine bone marrow–derived progenitors. We found expression of various types of genes, including regulatory cytokines and their receptors, signal transduction genes, and transcription factors. To assess gene expression during myeloid differentiation, we examined patterns of change during induced differentiation of EML cells. Several hundred of the genes underwent fluctuations in expression level during myeloid cell differentiation. The complete database, accessible on the World Wide Web at http:// yale130132115135.med.yale.edu/, allows for retrieval of information regarding these genes. Our microarray allows for genomewide expression analysis of myeloid stem cells, which will help in defining the regulatory mechanisms of stem cell differentiation. (Blood. 2002; 100:833-844)

© 2002 by The American Society of Hematology

Introduction

Acute myeloid leukemia (AML) remains a highly lethal malignancy requiring novel therapeutic strategies.1,2 An integral component of the AML phenotype is the loss of the capacity to differentiate into mature myeloid cells. Consequently, a major focus of research in this area has been on the molecular mechanisms controlling normal myeloid differentiation. One conclusion of this work is that differential expression of key regulatory genes in hematopoietic stem cells controls their differentiation into mature cell types, including erythrocytes, platelets, neutrophils, monocytes, eosinophils, and basophils. A detailed understanding of the gene expression patterns throughout hematopoietic differentiation obtained by means of messenger RNA (mRNA)–expression profiling and bioinformatics can provide valuable insights into this complex process and will perhaps lead to novel treatment approaches for AML. We are interested in the patterns of gene expression at early stages of myeloid commitment and differentiation. Previous studies have identified a small but diverse group of genes that are down-regulated during this process, including CD34,3 ckit,4 Jagged2,5 mpl,6 sca-1,7 SCL,8 GATA-1 and GATA-2,8 Flt-1,9 Notch,10 Ap-1,11 Mzf-1,12 C/ebp,13 and STATs.14 Up-regulated genes include Pu.115 and others.16 However, there are likely to be many more genes, some known and some yet to be identified, involved in the molecular events of differentiation.17,18

To better understand the interacting pathways and networks involved in hematopoiesis, we decided to employ the genomewide strategy of gene expression profiling using complementary DNA (cDNA) microarrays.19 A requisite component of this technology is inclusion of potentially critical genes on the array. With the aim of developing a cDNA microarray for use in the study of gene expression during early myelopoiesis, we constructed a subtracted cDNA library derived from sorted hematopoietic progenitor cells, and we complemented this set of genes with a set of available clones known to be important to myelopoiesis. This clone set was sequenced, characterized, and then spotted onto glass slides to create a microarray for analyzing the profile of gene expression during early steps in myelopoiesis. We have employed this microarray to assess expression in primary hematopoietic precursors and to analyze changes in gene expression during induced differentiation of the myeloid progenitor cell line EML.

Materials and methods

Reagents

The α-[32P]deoxyctydine 5’triphosphate (α-[32P]dCTP) (3000 Ci/mmol [111 TBq/mmol]) was purchased from Amersham Pharmacia Biotech.
Buckinghamshire, United Kingdom), and restriction enzymes were purchased from New England Biolab (Beverly, MA). Iscoves modified Dulbecco medium (IMDM) (Life Technologies, Rockville, MD) supplemented with 20% horse serum was the culture medium used throughout. Recombinant human stem cell factor (SCF) was purchased from Peprotech (Rocky Hills, NJ). Rhodamine123 and Hoechst 33342 dyes were from Molecular Probes (Eugene, OR). Plasmid vectors λ ZipLox and pSport were from Life Technologies (Rockville, MD).

Cell lines

The EML and EPRO cell lines20 were generously contributed by S. Tsai. EML cells were maintained in IMDM supplemented with 20% horse serum, 15% BHK/MKL-conditioned medium (containing SCF), 1% l-glutamine, 1% penicillin/streptomycin/amphotericin (P/S/A), and 1% nonessential amino acids. EPRO cells were maintained in IMDM supplemented with 20% horse serum, 10% HM-5-conditioned medium (containing granulocyte-macrophage colony-stimulating factor), 1% l-glutamine, 1% P/S/A. Cell lines were cultured at 37°C in 5% CO2. EML cells were induced to differentiate into myeloid cells with 10−4 M all-trans-retinoic acid (ATRA) and 5% WEHI-conditioned medium as a source of interleukin-3 (IL-3).

cDNA library for subtraction

The lineage-rhodamine123-Hoescht134 (Lin-rhodamine123-Hoescht134) (LRH) library was derived by means of previously described techniques,21,22 and its construction has been described.23 Briefly, primary bone marrow cells were depleted of lineage-committed cells and then further enriched for primitive cells by fluorescent-activated cell sorting for cells with low-level staining with rhodamine123 and Hoechst 33342 dyes. From 30 mice, 5000 cells were obtained, from which a directionally cloned cDNA library was created in the lambda vector λ ZipLox (Gibco BRL) as SalI-Eag1 fragments, in such a way that the 5′ end of the cDNA was adjacent to the SalI site and the 3′ end was adjacent to the Eag1 site. The original library had an initial plating efficiency of 1.44 × 107 clones.23

The LRH library was converted to single-stranded DNA (ssDNA) by in vivo excision via cre-mediated excision and filamentous phage rescue as described.24,25 Briefly, we electroplated 50 μg library DNA into competent Escherichia coli DH5α F− bacteria, F−/λdλ lacZΔM15 lacZΔYα-argF Lys− thi-1 gyrA96 relA1. We incubated the transformed bacteria in 100 mL 2 × YT broth at 30°C on an orbital shaker for 1 hour and then added 150 μM ampicillin (50 mg/mL) and 1 mL 20% glucose. Bacteria were grown overnight until OD600 = 0.1 (OD indicates optical density), followed by incubation at 37°C for 1 hour until OD600 = 0.2. The culture was superinfected with M13KO7 helper phage and then cultured an additional 2 hours. After eliminating bacteria by centrifugation filamentous phage particles were precipitated with the addition of 4 g polyethylene glycol and 2.92 g NaCl into 100 mL solution and incubation at 4°C for 16 hours. Particles were collected by centrifugation; the phage DNA was purified with phenol-chloroform extraction; and the final product was dissolved in TE (10 mM Tris-HCl pH 7.9, 1 mM EDTA). The ssDNA was confirmed by digestion with mung-bean nuclease. This yielded approximately 25 mg single-stranded phage DNA composing the entire LRH library of cDNAs.

Prior to hybridization, the single-stranded library DNA was purified by means of hydroxyapatite (HAP) column chromatography to eliminate double-stranded DNAs (dsDNAs). First, 10 μg library ssDNA was digested with PstI, and then it was applied to a 10-μL HAP column. The flowthrough (7 to 8 μL), representing the ssDNA, was collected and concentrated by means of Qiagen (Valencia, CA) spin columns following the manufacturer’s protocol. DNA eluted from the Qiagen spin column was precipitated and resuspended in 5 μL double-distilled water (ddW). The ssDNA was confirmed by digestion with mung-bean nuclease.

Preparation of driver DNA

The DNA driver pool was prepared with 10,000 mouse cDNA clones that were a gift from Research Genetics (Huntsville, AL). These clones were derived from mouse testes, kidney, diaphragm, skin, lung, brain, heart, and whole embryonic fetus; mouse melanoma; embryonic carcinoma; and mouse macrophages. The inserts were amplified by polymerase chain reaction (PCR) with Expand high-fidelity PCR system (Invitrogen, Carlsbad, CA) under the following conditions: 94°C, 7 min for 1 cycle; 20 cycles at 94°C for 1 minute, 55°C for 2 minutes, and 72°C for 3 minutes; and a final extension of 7 minutes at 72°C. The PCR products were purified by phenol-chloroform extraction and checked by ethidium bromide–stained agarose gel electrophoresis. All inserts were combined to make the driver pool by transfer of 2 μL from each PCR product.

Subtraction of cDNA library

Subtraction of the LRH library was performed essentially as described by Bonaldo et al.24 with minor modifications. The hybridizations of liver ssDNA and pooled driver DNA were performed in 20 μL volume hybridization buffer (50% formamide, 0.12M NaCl, and 1% sodium dodecyl sulfate (SDS)) with 2.5 μg driver DNA and 50 ng tracer ssDNA from cDNA library at 30°C for 110.4 hours (Ct = 50; here, Ct is substrate concentration of total DNA in solution, and t is hybridization time at 30°C). To block hybridization via the vector and poly(adenylic acid) (poly(A)) tail sequences, blocking oligonucleotides were designed (Table 1) and were included in the hybridization at a concentration of 2 μg/μL for blocking vector homology sequence and 0.5 μg/μL for blocking poly(A) tail sequence. Following hybridization, DNA molecules remaining single stranded were purified by HAP chromatography and were concentrated to 11 μL volume as described above. The ssDNA was converted into dsDNA in vitro by transferring the ssDNA into premixed reaction solution, (5 μL sequenase buffer [5X] and 1 μL M13 forward primer [1 μg/μL]), heating at 65°C for 5 minutes, then 37°C for 3 minutes, adding 2 μL deoxynucleoside-5′-triphosphates (10 mM each), 1 μL dithiothreitol (DTT) (0.1 M), and 1 μL sequenase (5 U/μL) into reaction solution, incubating at 37°C for 30 minutes, and then purifying the dsDNA with phenol-chloroform extraction. The resultant dsDNA was transformed into E. coli DH10β, which was then plated on Luria-Bertani broth/ampicillin agar plates. The total number of clones were calculated. An aliquot of the subtracted LRH library was submitted to Lawrence Livermore National Laboratory (Livermore, CA) for transformation, plating, and robotic picking of colonies into 96-well plates as part of the Cancer Genome Anatomy Project (National Institutes of Health, Bethesda, MD). A separate aliquot of the libraries was amplified as a population and used to prepare DNA.

DNA preparation and sequencing

Plasmid DNA was prepared in 96-well plates. The clones derived from the subtracted cDNA libraries were grown in 96-well plates, and plasmids were isolated by the alkaline-lysing method. The final plasmid was dissolved in 100 μL ddW. (Note: for sequencing purposes, we used 20 μL plasmid directly, and for arraying, we further purified the plasmid DNA with a 96-well filter plate. The picked clones were sequenced with single-pass automated sequences by the W. M. Keck Facility at Yale University (New Haven, CT).

Table 1. Blocker sequences for library subtraction for LRH library

| Vector | Blocker sequences |
|--------|-------------------|
| pSport | 5′ gtc gtc act gac aac aac cag tac 3′ |
| 3′ end | 5′ ggt agt cac gaa taa taa cta ctc cta gga gaa aac acc tac 3′ |
| 5′ end | 5′ gag cgc tct aga gga tcc ggg ctt aac tgc tgc aag 3′ |
| NA | 5′ ggc tgt tac 3′ |
| NA | 5′ tgt ttc ttc ttc ttc 3′ |

NA indicates not applicable.
Haven, CT) and/or the Genome Sequencing Center at Washington University Medical School (St Louis, MO) with the use of an M13AEK forward primer (5’/H11032 CAA AAG GGT CAG TGC TG 3’/H11032), which primes synthesis at the 3’-end of clones. Some clones were also sequenced from the 5’-end with the use of the T7 promoter primer (5’/H11032 TAA TAC GAC TCA CTA TAG GG 3’/H11032). The M13/pUC reverse primer (AGC GGA TAA CAA TTT CAC ACA GGA) for 5’-termini was used to confirm LRH novel sequences.

Sequence editing and analysis

Because some sequencing primers contained common vector sequence, we first removed vector sequences from the sequences with CodonCode-Cross_Match software (http://www.codoncode.com). FASTA formatted DNA sequences were compared with known nucleotide sequences with the use of the Blast algorithm in batches of 3228 sequences and the use of the blastall program (BLASTN and BLASTX programs) installed in a Dell Workstation with a Linux operating system. Three publicly accessible databases were searched: Genbank nonredundant (nr) nucleotide, database for expressed sequence tags (dbESTs), and Genbank nr protein. Internal redundancy within our clone set was determined by comparison of each sequence against our own database. Categorization of sequence homology was based on the following criteria: exact match to known named mouse genes (threshold score exceeding 200) or protein, or near-identity to a known gene or protein from a species other than mouse (usually either human or rat); EST only (no extensive homology to any nucleotide or protein sequence in these public databases). Sequence data from 5’ and 3’-sequence reads were assembled with the use of the PHRAP software package (http://www.phrap.org/) kindly provided by Phil Green (Washington University). Protein motifs within the assembled sequences were identified by converting the DNA sequence to open reading frame using the ORF analysis program (http://curagen.com/) (CuraGen, New Haven, CT) and then performing domain searches with Pfam, ProDom, Prosite, and Prints software programs (http://curagen.com/) (CuraGen). Cutoff parameters for match selection were P < .05; identities exceeded 40%, and positives exceeded 50%.

Southern hybridization

Five micrograms of library DNA was double-digested with restriction enzymes BamHI and EcoRI, fractionated on 0.8% agarose gel, and transferred to nylon membranes. Hybridization probe DNAs were cut with restriction enzyme, gel-purified, and labeled with random primer DNA labeling. The labeled probes were purified with Sephadex G-50 Quick Spin Column (Boehringer Mannheim, Germany), and Southern blot analysis was performed according to standard methods.

Table 2. Summary of subtracted cDNA library from Blast search of 2304 LRH clones

| Type of clone      | No. |
|--------------------|-----|
| Nonredundant clones* |     |
| Novel              | 247  |
| ESTs               | 386  |
| Known              | 622  |
| Redundant          | 599  |
| Ribosomal          | 10   |
| Bad sequences      | 80   |
| Vector             | 309  |
| Mitochondrial      | 35   |
| Bacterial          | 16   |

*There were 1255 clones in this category.
Table 3. Examples of known genes in subtracted cDNA library placed into functional categories

| Category/gene name | Description | Log2 (EML24/EML0) | Log2 ([LHLow R Bright/LHLow R Low]) |
|--------------------|-------------|-------------------|-----------------------------------|
| CBIIb              | Core binding factor β | — | 0.9 ± 0.2 |
| Cebpα-rs1          | Transcription regulator related to C/ebp α | — | 0.9 ± 0.3 |
| Cebpβ              | Transcription regulator C/ebp β | — | — |
| Cebpγ              | Controls the activity of IL-4 promoter activity | — | 1.9 ± 0.68 |
| Crem               | Transcriptional modulator of cAMP responsive genes | — | 1.5 ± 0.7 |
| Elk4                | Member of ETS oncogene family | — | 1.2 ± 0.5 |
| Fox                | Part of AP1; propagation of mitogenic signals | — | 1.2 ± 0.4 |
| Hmgα1              | High-mobility group AT-hook 1 | — | 2.0 ± 0.05 |
| Homeo5             | Homeobox A5 | — | — |
| Jun                | Jun oncogene, part of AP1 transcriptional complex with Fos | — | — |
| Klf9               | Erythroid-specific transcription | 1.8 ± 0.53 | 2.5 ± 0.15 |
| Lrf                | BTB-POZ transcription factor, dimerizes with Bcl6 | — | 2.7 ± 0.1 |
| Mint               | Organizes transcriptional complexes | 2.9 ± 0.3 | 2.5 ± 0.7 |
| Mid2               | Related to Mid1, a partner in t(3;5) with Npm in leukemia | — | 2.9 ± 0.45 |
| Nkx2-3             | Regulates emigration of circulating leukocytes | — | 1.6 ± 1.0 |
| Sox4               | Important for very early B-cell differentiation | — | 2.1 ± 1.1 |
| Sox5               | SRY-box containing gene 5 | — | 3.1 ± 0.22 |
| Tal1               | T-cell acute lymphocytic leukemia 1 | — | 1.6 ± 0.4 |
| Xbp1               | Essential for plasma cell differentiation, liver development | — | 1.1 ± 0.5 |
| Arhgef1            | Rho guanine nucleotide exchange factor (GEF) 1 | — | 0.9 ± 0.58 |
| Calm               | Calmodulin, component of calcium signaling pathway | — | 1.2 ± 0.13 |
| Calr               | Calreticulin, major ER calcium-binding protein | — | 2.2 ± 0.3 |
| Camkk2             | CaM-kinase kinase β, regulates calcium signaling | — | 3.1 ± 0.2 |
| Ccnd1              | Cyclin D1, cell cycle regulation | — | 1.7 ± 1.0 |
| Csnk1a             | Casein kinase i, α isoform | — | 2.9 ± 0.45 |
| Husl               | Cell cycle checkpoint protein | — | 1.4 ± 0.07 |
| Iqgap              | IQ motif containing GTPase-activating protein 1 | — | — |
| Itp3               | Inositol 1, 4, 5-triphosphate receptor 3 | — | — |
| Lats1              | Large tumor suppressor, modulates CDC2 activity | — | — |
| Map2k3             | Mitogen-activated protein kinase isoenzyme 3 | — | — |
| Nsp3               | Novel SH2-protein; couples Eph receptors to R-Ras | — | 2.1 ± 0.06 |
| Parg1              | PTPL1-associated RhoGAP 1; activates Rho | ND | 2.7 ± 0.55 |
| Pde4b              | Phosphodiesterase 4B, cAMP specific | — | — |
| Pde8a              | Phosphodiesterase 8A, cAMP specific | — | — |
| Rala               | Ras-related GTPase | — | 0.74 ± 0.22 |
| Rangap1            | RAN GTPase-activating protein 1 | — | — |
| Tax1bp1            | Antiapoptotic; involved in IL-1 signaling pathway | 0.6 ± 0.35 | — |
| Tpd52              | Regulator of cell proliferation, gene transcription | — | 2.4 ± 0.2 |
| Traf1              | TNF receptor-associated factor 1 | — | — |
| Traf6              | TNF receptor-associated factor 6 | — | — |
| Tyk2               | Plays a restricted role in IFNγ signaling | — | 2.1 ± 0.2 |
| Vav3               | Activates Rho GTP-binding proteins | — | — |
| Wnk1               | Serine/threonine protein kinase | — | 1.9 ± 0.5 |
| Growth factors     | | | |
| Efrb1              | Ephrin B1, cek5 receptor ligand | — | — |
| Hdgf               | Hepatoma-derived growth factor; a mitogen for fibroblasts | — | 1.4 ± 0.04 |
| Hegfl              | Heparin-binding EGF-like, potent mitogen | — | 3.2 ± 0.12 |
| Sgfr               | Controls the proliferation of hematopoietic stem cells | — | 1.4 ± 0.17 |
| Surface molecules/membrane proteins | | | |
| Fgfr1              | Basic fibroblast growth factor receptor 1 precursor | — | 3.1 ± 0.1 |
| Gpr21              | G protein-coupled receptor 21 (GPR21) | — | 0.65 ± 0.1 |
| H2-K               | Class I antigen of MHc, mediates transplant rejection | — | 1.5 ± 0.08 |
| Hegfl              | Heparin-binding EGF-like growth factor | — | — |
| Itm1               | Integral membrane protein 1; unknown function | — | 1.5 ± 0.08 |
| Ncam2              | Neural cell adhesion molecule 2 | — | — |
| Oprs1              | Opioid receptor, δ1 | — | — |
| Trhstf1b           | TNF receptor superfamily, member 1b | — | — |

Preparation of DNA samples for arraying

Bacterial cultures were grown overnight in 96-well culture plates (Qiagen), and plasmid DNA was prepared as described above. The cDNA inserts were amplified by means of PCR (96-Well GeneAmp PCR System 9700) (Perkin-Elmer–Applied Biosystems, Foster City, CA) in 96-well plates with the use of M13 AEK forward and reverse primers (1 μM) for amplification. The PCR reaction was carried out in 100 μL solution of 1 mM deoxyadenosine 5'-triphosphate (dATP), dCTP, deoxyguanosine 5'-triphosphate (dGTP), dTTP, and deoxythymidine 5'-triphosphate (dTTP), and 1× PCR buffer.
Table 3. Examples of known genes in subtracted cDNA library placed into functional categories (continued)

| Related to blood cell function | Description | Log2 (EML24/EML0) | Log2 (LHLow R Bright/LRLow H Low) |
|-------------------------------|-------------|-------------------|----------------------------------|
| Gpx1 Glutathione peroxidase; red cell antioxidant defense | 1.2 ± 1.0 | — |
| H2-Aa Histocompatibility 2, class II antigen A, α | — | — |
| Ncf1 NADPH oxidase subunit (47 kd) | — | 0.14 ± 0.1 |
| Prg2 Eosinophil major basic protein | — | 2.8 ± 0.3 |

The Log2 ratios of EML24/EML0 and LHLow R Bright/LRLow H Low were the average values of 3 experiments and normalized the signal intensity of each hybridization to internal control glyceraldehyde-3-phosphate dehydrogenase. A minus sign in front of a Log2 ratio indicates downregulation (and the absence of a minus sign indicates upregulation) of (1) gene expression when EML cells were treated with all ATRA/IL-3 for 24 hours, or (2) in primary sorted bone marrow cells, LHLow R Low relative to LHLow R Bright. A data cell containing only a minus sign indicates no change.

LR Low H Low indicates lineage-minus (Lin−) rhodamineLow HoechstLow; L HLow R Bright, Lin− HoechstLow rhodamineBright; cAMP, cyclic adenosine monophosphate; GTPase, guanosine 5’-triphosphatase; Eph, ephrin; TNF, tumor necrosis factor; IFN, interferon; GTP, guanosine 5’-triphosphate; EGF, epidermal growth factor; MHC, major histocompatibility complex; and NADPH, nicotinamide adenine dinucleotide phosphate.

Table 4. List of domains identified in novel genes by means of multiple domain search tools

| Clone ID | Gene Ontology category | Description | Searching tool |
|----------|------------------------|-------------|---------------|
| 2405 | Cell cycle regulator | Cell division control protein S4 | ProDom |
| 2073 | Chaperone | Heat shock 70-kd protein. ATP-binding; heat shock; | ProDom |
| 2247 | Enzyme | Prelyltransferase and squalene oxidase | Pfam |
| 2656 | Enzyme | E-class p450 group i signature (motif source owl:s22339) | Prints |
| 2100 | Enzyme | Acetyl esterase | ProDom |
| 2923 | Enzyme | Cyclase adenylate type ATP | ProDom |
| 2632 | Enzyme | Sulite reductase (NADPH) flavoprotein | ProDom |
| 3121 | Enzyme | NOS-type oxidoreductase | ProDom |
| 3332 | Enzyme | Nitrate reductase (NADP|H) | ProDom |
| 2363 | Enzyme | Nitrate reductase | ProDom |
| 2941 | Enzyme | Acyl-coA oxidase | Pfam |
| 2255 | Enzyme | Dehydrogenase γ-glutamyl | ProDom |
| 1860 | Enzyme | 3-isopropylmalate dehydrogenase precursor | ProDom |
| 2247 | Enzyme | Farnesyltransferase β subunit | ProDom |
| 1667 | Enzyme | DNA-directed DNA polymerase | ProDom |
| 2148 | Enzyme | Preprotein translocase SECA subunit | ProDom |
| 2821 | Enzyme | Glutathione S-transferases | Pfam |
| 3063 | Enzyme inhibitor | Trypsin inhibitor, serine protease inhibitor, 41 aa | ProDom |
| 2940 | Ligand binding | Phosphorylubiquinase/reC protein | Pfam |
| 2639 | Ligand binding | Pyridine nucleotide disulfide reductase class-II signature | Prints |
| 2199 | Nucleic acid binding | Ribosomal protein 40S S29 S33 repeat acetylation | ProDom |
| 1248 | Nucleic acid binding | U1 small nuclear ribonucleoprotein | ProDom |
| 2097 | Nucleic acid binding | Chromosome segregation protein SM2 | ProDom |
| 2097 | Nucleic acid binding | λ and other repressor helix-turn-helix signature | Prints |
| 2625 | Nucleic acid binding | Protein ATP-binding | ProDom |
| 2526 | Nucleic acid binding | Transmembrane protein HWLF3 late glycoprotein | ProDom |
| 2524 | Nucleic acid binding | Elongation factor Tu mitochondrial precursor protein | ProDom |
| 1998 | Nucleic acid binding | Ribosomal protein L20 signature | ProDom |
| 1497 | Nucleic acid binding | Protein homeobox DNA-binding | ProDom |
| 3103 | Nucleic acid binding | Transcriptional regulator YCF30 protein transcription | ProDom |
| 3132 | Nucleic acid binding | Double-stranded RNA-specific editase 1 | ProDom |
| 2952 | Nucleic acid binding | Homeobox protein GSH-1 | ProDom |
| 2840 | Nucleic acid binding | Exonuclease ABC subunit a SOS response excision | ProDom |
| 2704 | Signal transducer | Dopamine receptor. G-protein coupled | ProDom |
| 1984 | Signal transducer | Receptor polypeptide gastric inhibitory precursor (GIP-R) | ProDom |
| 3023 | Signal transducer | Receptor precursor repeat glycoprotein | ProDom |
| 1981 | Signal transducer | SL cytokine precursor (FLT3 ligand) | ProDom |
| 3001 | Signal transducer | Tyrosine-protein kinase JAK3 | ProDom |
| 2995 | Signal transducer | mapk/erk kinase kinase 1 | ProDom |
| 2366 | Signal transducer | Gap junction cx32.7 | ProDom |
| 2316 | Signal transducer | Chemotaxis protein transducer | ProDom |
| 2316 | Signal transducer | HAMP domain/methyl-accepting chemotaxis protein (mcp) | Pfam |
| 1925 | Signal transducer | Insulinlike receptor precursor | ProDom |
| 2858 | Signal transducer | Tyrosine-protein kinase JAK3 | ProDom |
| 3091 | Transporter | Transport protein precursor signal periplasmic | ProDom |
| 1939 | Transporter | Calcium calcium-transporting ATPase plasma membrane | ProDom |

NOS indicates nitric-oxide synthase.
and deoxyribothymidine 5'-triphosphate (dTTP); 1.5 mM MgCl2; and 2.5 U Taq polymerase in 96-well plate with the following cycles: 5 cycles of 94°C for 30 seconds, 55°C for 1 minute, and 72°C for 1.5 minutes; followed by 30 cycles of 94°C for 30 seconds, 56°C for 1 minute, and 72°C for 1.5 minutes; and then 1 cycle of 72°C for 10 minutes. Resulting PCR products were purified with the use of a 96-well glass-fiber filter (MAFB NOB) (Millipore, Bedford, MA) according to the manufacturer’s user manual. The purity and yield were approximated by running the purified PCR products on a 0.8% agarose gel. The DNAs were prepared for arraying by transferring 5 μL to 384-well plates and adding SSC to a final concentration of 3×. Glass slides were prepared for printing and arrayed by the Yale Microarray Facility (http://info.med.yale.edu/wmkeck/dna_arrays.htm) with the use of a GeneMachines (San Carlos, CA) Ommigene Arrayer. After printing, the slides were postprocessed as described by P. Brown and J. DeRisi (http://www.microarrays.orgprotocols.html).

RNA preparation, probe labeling, and hybridization

Total RNA was prepared by means of Trizol reagent (Life Technologies). Microarray slide hybridization was performed as follows. The cDNA probes were synthesized by reverse transcription with oligo deoxothyminidne (dT) as primer, incorporating allyl amine-deoxyuridine triphosphate (aa-dUTP) (Sigma, St Louis, MO) into synthesized cDNA. Reactions were performed in 100 μL reaction with the following final concentrations: 85.8 μg/mL oligo dT primer; 0.5 mM each dATP, dCTP, and dGTP; 0.2 mM aa-dUTP; 0.3 mM dTTP; 10 mM DTT; and 1280 U MLV reverse transcriptase (Superscript II) per milliliter. Coupling of cyanine-3 (cy-3) or cy-5 dyes to aa-dUTP by incubation at 25°C for 1 hour in subdued light. Hybridization of fluorescently labeled probes to glass slides was performed with hybridization buffer (50% deionized formamide, 12.5% SSPE, 0.625% SDS, 1.5 × Denhardt reagent with blockers [0.5 μg/μL mouse Cot 1 DNA, 0.1 μg/μL poly(A) (15A), and 0.2 μg/μL yeast transfer RNA] at 42°C for 18 to 24 hours. After hybridization, the probe was washed first with 1 × SSC, 0.1% SDS, at 25°C for 15 minutes, then with 0.2 × SSC, 0.1% SDS, and finally with 0.2 × SSC. The slides were scanned with a GSI Lumonics (Packard, Billerica, MA) or Axon (Axon Instruments, Union City, CA) laser scanner. Analysis of the fluorescent hybridization signal of microarray slide was performed with Quantarray (Packard) or Genepix software (Axon Instruments, Union City, CA), and the data were analyzed by means of Microsoft Excel. Further clustering of data was performed with the use of Genespring software (Silicon Genetics, Redwood City, CA).

Northern blot analysis

Northern hybridization was carried out using standard methods. Total RNA (10 μg) was electrophoresed on a 1% agarose/formaldehyde gel and was blotted onto Hybond-N nylon membranes (Amersham Pharmacia Biotech) followed by UV cross-linking. DNA probes were labeled with random primers, and the hybridization was performed at 65°C for 16 hours. Signals on the washed filter were visualized by autoradiography.

Results

Creation of a subtracted myeloid cDNA library that is enriched for low-abundance transcripts

With the long-term goal of fully characterizing changes in gene expression during the early stages of myelopoiesis, we wanted to develop a cDNA microarray that was enriched in genes expressed in primitive hematopoietic cells and early committed myeloid cells. We took a 2-pronged approach to achieve this: one prong was to create a subtracted cDNA library from an early hematopoietic library; the second was to complement this set with available genes known to be involved in myelopoiesis. As starting points for the library subtractions, we used cDNA library LRH, derived from primary bone marrow samples that were sorted for early progenitors by flow cytometry (Degan et al23). The initial complexity of the LRH cDNA library was 1.44 × 107 clones. Subtraction of the libraries was performed by using as driver a pool of 10 000 mouse Integrated Molecular Analysis of Genomes and Their Expression (IMAGE) Consortium cDNA clones that were derived from several different mouse organ cDNA libraries (Figure 1). Following subtraction, we obtained 1 × 106 total clones; the complexity within this population is not known.

To assess the efficacy of the subtraction process, we performed Southern blot analysis of library-derived cDNA populations derived before and after normalization. As hybridization probes, we used 3 different sequences known to be present in both the driver and the tracer populations. The results (Figure 2) clearly indicate that the subtraction

Figure 3. Alignments of important motifs for 9 proteins identified in hematopoietic stem cells. (A) Clone 1317 with 3 closely related WEEl kinase protein family members, which play an important role in mitosis. (B) Clone 1508 with 5 closely related members of the Zinc-finger protein CBBP DNA-binding family, which play a role in gene transcription. (C) Clone 2265 with 6 GLI protein family members, all of which are Zinc-finger DNA-binding proteins. (D) Clone 1405 with 2 C-ETA type protein kinases. (E) Clone 1797 with 5 closely related insulin/IGF/relaxin family members, which play a role in insulin expression and regulation. (F) Clone 2131 with 2 GPR1/FUN34/yaaH family members. (G) Clone 2995 with a region of homology to MEK1 proteins. (H) (I) Regions of similarity between 2 clones (3001 and 2858) and the Jak3 protein.
was effective in greatly reducing the abundance of these clones, but the degree of reduction for these genes was variable. For example, clone ID9063 (IMAGE: 421622) was reduced around 3-fold with subtraction, but superoxide dismutase precursor was reduced more than 20-fold (Figure 2). We also tested for enrichment of genes present only in the tracer population, not in the driver. In this instance, some low-copy genes were enriched more than 1.5- to 5-fold through hybridization (eg, Mel18; Figure 2C).

Table 6. Cytokines, lineage-specific receptors, and transcription factors on the myeloid cDNA microarray

| Clone ID | UniGene No. | Locus link | Symbol | Name |
|----------|-------------|------------|--------|------|
| Cytokines |             |            |        |      |
| 4 345    | 795         | 12 977     | Csfl   | CSF 1 (M-CSF) |
| 4 154    | 35 814      | 16 156     | Il11   | IL-11 |
| 4 284    | 371         | 16 189     | Il4    | IL-4  |
| 22       | 28 830      | 59 027     | Pbef   | Pre-B-cell colony-enhancing factor |
| Cytochrome receptors | | | | |
| 4 594    | 22 574      | 12 978     | Csflr  | Colony stimulating factor 1 receptor |
| 1 105    | 156 264     | 12 982     | Csflra | GM-CSF receptor |
| 4 400    | 1 940       | 12 984     | Csflrb2| CSF 2 receptor, beta 2, low-affinity (GM) |
| 4 377    | 22 673      | 14 127     | Fcrlg  | Fc receptor, IgE, high affinity 1, gamma polypeptide |
| 4 269    | 10 809      | 14 130     | Fcgr2b | Fc receptor, IgG, low affinity IIb |
| 4 165    | 3 303       | 14 132     | Fcgrt  | Fc receptor, IgG, alpha chain transporter |
| 4 307    | 1 54       | 16 155     | Il10rb | IL-10 receptor, beta |
| 4 233    | 731         | 16 161     | Il12rb1| IL-12 receptor, beta 1 |
| 3 963    | 896         | 16 177     | Il1r1  | IL-1 receptor, type I (Il1r1) |
| 4 158    | 1 349       | 16 178     | Il1r2  | IL-1 receptor, type II |
| 4 157    | 24 771      | 16 180     | Il1rap | IL-1 receptor accessory protein |
| 4 142    | 4 944       | 16 188     | Il3ra  | IL-3 receptor, alpha chain |
| 4 224    | 389         | 16 197     | Il7r   | IL-7 receptor alpha chain precursor |
| Transcription factors | | | | |
| 3 604    | 1 318       | 11 350     | Abf1   | v-abl Abelson murine leukemia oncogene 1 |
| 1 113    | 34 537      | 12 606     | Cebpa  | C/EBP, alpha |
| 1 067    | 10 794      | 13 599     | Cebpe  | C/EBP, epsilon |
| 4 521    | 131         | 15 399     | Hoxa2  | Homeobox A2 |
| 4 557    | 173         | 15 402     | Hoxa5  | Homeobox A5 |
| 4 565    | 3 546       | 15 412     | Hoxb4  | Homeobox B4 |
| 4 553    | 207         | 15 413     | Hoxb5  | Homeobox B5 |
| 4 578    | 215         | 15 414     | Hoxb6  | Homeobox B6 |
| 4 432    | 255         | 15 417     | Hoxb9  | Homeobox B9 |
| 4 331    | 1 351       | 15 423     | Hoxc4  | Homeobox C4 |
| 4 134    | 20 305      | 15 424     | Hoxc5  | Homeobox C5 |
| 4 556    | 4 444       | 15 425     | Hoxc6  | Homeobox C6 |
| 4 602    | 4 765       | 15 427     | Hoxc9  | Homeobox C9 |
| 1 057    | 2 444       | 17 869     | Myc    | Myeloid differentiation oncogene |
| 1 037    | 88 061      | 20 185     | Ncor1  | Nuclear co-repressor 1 |
| 4 197    | 3 107       | 18 021     | Nfatc3 | Nuclear factor of activated T-cells, cytoplasmic 3 |
| 4 486    | 31 255      | 18 128     | Notch1 | Notch gene homolog 1 |
| 4 187    | 4 173       | 18 130     | Notch2 | Notch2-like |
| 4 494    | 4 945       | 18 131     | Notch3 | Notch gene homolog 3 |
| 4 146    | 173 813     | 18 132     | Notch4 | Notch gene homolog 4 |
| 3 865    | 22 675      | 18 514     | Pbx1   | Pre-B-cell leukemia transcription factor 1 |
| 4 588    | 4 788       | 19 401     | Rara   | Retinoic acid receptor, alpha |
| 4 585    | 1 273       | 19 411     | Rarg   | Retinoic acid receptor, gamma |
| 4 496    | 4 869       | 19 696     | Rel    | Reticuloendotheliosis oncogene |
| 4 203    | 455         | 21 405     | Tcf1   | Transcription factor 1 |

CSF, colony-stimulating factor; M, macrophage; GM, granulocyte-macrophage; C/EBP, CCAAT/enhancer binding protein.
The subtracted LRH cDNA library contains a high percentage of novel sequences

To determine the identity of the cDNA clones derived from the library subtraction, we subjected 2304 LRH clones to partial sequence determination and analysis (Table 2). Of the clones, 54% (1255) were nonredundant cDNAs, representing protein-encoding mRNAs. Of these, 247 (20%) were novel sequences; 386 (31%) ESTs; and 622 (50%) known genes. Of the LRH sequences, 46% were not useful, the majority because of being redundant, ribosomal, or empty vector. Sequence data for all of the novel genes have been submitted to GenBank (dbEST).

To facilitate the analysis, retrieval, and further accrual of information concerning these genes, we created a database that is accessible via the World Wide Web (http://yale130132115135.med.yale.edu/).

Examples of the known genes derived from the subtracted cDNA library are shown in Table 3, categorized by the functional criteria. Of interest is the presence of 3 members of the C/ebp family of transcription factors, as well as Cbfb, Lsf, Sox4, Tal1, and Xbp1, each of which is an important regulator of cell differentiation of blood cell lineages and/or other organs. It is also notable that while genes for growth factors (eg, Hdgf, Hagfl, Efnb1) and growth factor receptors (eg, Fgfr1, Tnfrsf1b, Oprs1) are present, none of the classical hematopoietic-specific cytokines or their receptors is present in our subtracted library. Components of apoptotic pathways are represented by Tnfrsf1b, Traf1, Traf6, Prg2, Taxlbp1, Bnip3, and Casp6. Calcium signaling transducers are included, eg, Cimp, Calr, Cnkk2, and Ipr3. Also present are regulators of the cell cycle, including Ccnd1, Hus1, and Lats.

Protein structure analysis of novel genes

We identified 247 novel sequences among the subtracted LRH clones. These clones are considered novel because of our inability to find any matching sequence in available databases. For each of the potentially novel genes, we subjected clones to additional sequencing from both the 5' and 3' ends of the clones. After compiling the 5' and 3' sequence data, we derived potential open reading frames from these sequences and analyzed them for domains and/or functional motifs (Table 4). This revealed that our novel sequences contained 13 potential nucleic acid–binding proteins, including 4 transcription factors, 11 signal transducers including 2 with similarity to Jak3, 1 with homology to Fli3 ligand, 1 bearing resemblance to the insulin receptor, and 1 mapk/erk kinase kinase–like protein. Sixteen proteins with similarities to known enzymes or enzyme inhibitors were identified, including some potential drug targets (eg, farnesyltransferase, prenyltransferase, and adenylate cyclase). What is notable is the relative paucity of more structural proteins (Table 4). In Figure 3, we show detailed analyses of 9 potentially important novel genes and their homology to known proteins.

Development of a cDNA microarray for analysis of early hematopoiesis

A major goal of this endeavor was to create a cDNA microarray for evaluating gene expression changes during hematopoietic differentiation with specific interest in the myeloid lineage. Thus, the second prong of our approach was to supplement the subtracted library with genes known to be expressed in myeloid cells as well as genes encoding proteins that regulate cell cycle, apoptosis, differentiation, and cell signaling. Thus, we added 587 cDNAs for known genes from an IMAGE Consortium clone set, 310 genes from EML cells isolated by 2 separate subtractive cloning procedures, 96 putative Evi-1 target genes, and 576 T-cell–expressed genes (B. Lu, S. Kim, and R. A. Flavell, unpublished data, 2001) (Table 5). A significant number of cytokines, hematopoietic transcription factors, growth factors, and growth factor receptors are also on the array (Table 6). A detailed description of these cDNAs and their sources can be found on the Web-accessible database mentioned earlier (http://yale130132115135.med.yale.edu/). Purified PCR-amplified cDNA inserts from this collection of plasmids were robotically spotted on polylysine-coated glass slides.

Myeloid cell differentiation is accompanied by abundant fluctuations in gene expression

We tested the cDNA microarray by employing it to analyze patterns of gene expression during induced differentiation of EML cells, a myeloid progenitor cell line. We compared the spectrum of gene expression in uninduced EML cells to that of EML cells induced to differentiate for 6, 24, and 72 hours, as...
well as to that of EPRO cells, which represent a promyelocyte-like stage derived from the EML cells.24 In each experiment, a competitive hybridization was performed between labeled cDNA from uninduced EML cell and from induced EML or EPRO cells, except for the 24-hour time point samples, for which some hybridizations were not competitive. Fluorescently labeled cDNA samples (with either cy-3 or cy-5) (“Materials and methods”) were hybridized. Following washing of the slide, the amount of hybridized probe was quantitated as pixel intensity of fluorescence; low-intensity signals were discarded; and the normalized data were expressed as a log2 of the ratio of signal from induced RNA to uninduced RNA cells. These values for various ribosomal proteins (Table 7) and also induced, suggesting down-regulation of the NF-κB. The major class of genes expressed at higher level in RA-induced EML cells and EPRO cells relative to EML cells was the class encoding ribosomal proteins (Table 7) and included proteins in both the large- and small-ribosomal subunit. These data, together with the increase observed for elongation factors 1α and Tu-binding and polyA–binding protein, are consistent with a generalized increase in protein synthesis. The calcium signaling pathway also appeared up-regulated: calmodulin, calreticulin, and annexin A1 were all increased, suggesting down-regulation of the NF-κB signaling pathway. In addition, there were 74 uncharacterized genes (ESTs) and 37 novel genes that increased in expression during EML cell differentiation. The latter were derived from our library subtractions, and this demonstration that they are

| Symbol | Name | Gene Ontology |
|--------|------|---------------|
| Gzmb   | C11 encoding T-cell–specific protein CCP1 CTLA-1, granzyme B | Apoptosis |
| Png2   | Proteoglycan 2, esoinophil major basic protein | Apoptosis |
| Dad1   | Defender against cell death 1 | Apoptosis |
| Bnip3I | BCL2 adenovirus E1B 19-kd interacting protein 3-like | Apoptosis |
| Tia1   | Tia1 cytotoxic granule-associated RNA-binding protein-like 1 | Apoptosis |
| Calm   | Calmodulin | Calcium binding |
| Calr   | Calreticulin | Calcium binding |
| Cdkn1a | Cyclin-dependent kinase inhibitor 1A (P21) | Cell cycle |
| Ppm1g  | Protein phosphatase 1G, magnesium-dependent, γ isoform | Cell cycle |
| Lats1  | Large tumor suppressor homolog 1 | Cell cycle control |
| Ran    | GTP-binding nuclear protein RAN, member RAS oncogene family | Cell cycle regulator |
| Mdm2   | Transformed mouse 3T3 cell double minute 2 | Cell growth and maintenance |
| Ct5    | Chaperonin subunit 5 ξ | Chaperone |
| C1qa   | Complement component 1, q subcomponent, α polypeptide | Complement component |
| Myo1c  | Myosin Ic | Cytoskeleton organization and biogenesis |
| Ncl    | Nucleolin | DNA, RNA binding |
| Madh5  | MAD homolog 5 | Embryogenesis and morphogenesis |
| Mt1    | Metallothionein1 | Heavy metal binding |
| Elav3  | ELAV-like 3 (Hu antigen C) | Histogenesis and organogenesis |
| Igj    | Immunoglobulin joining chain | Humoral defense mechanism |
| Ppla   | Peptidylprolyl isomerase A | Immune response |
| Scya19 | Small inducible cytokine A19 | Inflammatory response |
| Itgb7  | Integrin β 7 | Integrin receptor signal signaling pathway |
| Emp3   | Epithelial membrane protein 3 | Membrane fraction |
| Vasp   | Vasodilator-stimulated phosphoprotein | Mineralocorticoid metabolism |
| Ada    | Adenosine deaminase | Nucleic acid metabolism |
| UbB    | Ubiquitin B | Nucleus |
| Gnas   | Guanine nucleotide binding protein, α stimulating | Protein ADP-ribosylation |
| DHP5   | Deoxyhypusyn synthase | Protein biosynthesis |
| Ef1α1  | Elongation factor 1-α (EF 1-α) | Protein biosynthesis |
| Pp1n11 | Protein tyrosine phosphatase, nonreceptor type 11 | Protein tyrosine phosphatase |
| Impdh2 | IMP dehydrogenase | Purine nucleotide biosynthesis |
| Various | Ribosomal proteins L3, L4, L5, L6, L7, L10, L11, L13, L23L3A, L29 | Ribosome biogenesis |
| Various | Ribosomal proteins S3, S4, S6, S7, S12, S14, S16, S17, S19, S20 | Ribosome biogenesis |
| Fgfl4  | Fibroblast growth factor receptor 4 | Signal transduction |
| Nsp3   | SH2-containing protein 3 | Signal transduction |
| Rgl2   | Raf guanine nucleotide dissociation stimulator–like 2 | Signal transduction |
| Mint   | Mxs2–interacting nuclear target protein | Transcription factor |
| Mit2   | Metal response element binding transcription factor 2 | Transcription factor |
| Bleb3  | Basic transcription element binding protein 3 | Transcription from Pol II promoter |
| Klf9   | Kruppel-like factor 9 | Transcription regulation |
| Creg   | Cellular repressor of E1A–stimulated genes | Transcription regulation |
| Rb1    | Retinoblastoma 1 | Transcription regulation |

ELAV indicates embryonic lethal, abnormal vision; IMP, inosine 5′-monophosphate. Other abbreviations are explained in the Table 3 footnote.
differentially expressed and thus likely to be of interest, shows the utility of this undertaking for investigating myelopoiesis.

Genes that were down-regulated in this differentiation pathway were more varied (Table 8), but also included 13 novel genes derived from our cloning effort, as well as 68 uncharacterized genes. Several key transcription factors were down during ATRA/IL-3-induced differentiation of EML cells (Table 8). Some of these, such as Klf1, Hoxb4, and Xbp1, have known regulatory roles in hematopoietic cells. These data provide an important starting point for further analyses aimed at understanding myelopoiesis at the molecular level, studies that are ongoing in the laboratory.

Table 8. Genes down-regulated during ATRA/IL-3–induced EML cell differentiation

| Symbol | Name | Gene Ontology |
|--------|------|---------------|
| Gzma   | Granzyme A | Apoptosis     |
| Pde1b  | Phosphodiesterase 1B, Ca2⁺-calmodulin dependent, 63 kd | Calmodulin binding |
| Vcam1  | Vascular cell adhesion molecule 1 | Cell adhesion |
| Ppm1g  | Fibroblast growth factor inducible 13 (FIN13) | Cell cycle |
| Ccn6   | Cyclin G | Cell cycle control |
| Yes    | Yamaguchi sarcoma viral (v-yes) oncogene homolog | Cell cycle regulator |
| Vav2   | Vav2 oncogene | Cell cycle regulator |
| Fos    | FBJ osteosarcoma oncogene | Cell cycle regulator |
| Hsp70-2| Heat shock protein, 70 kd 2 | Chaperone |
| FosL1  | fos-like antigen 2 | DNA binding |
| Hox11l1| Homeobox 11-like 1 | DNA binding |
| Rbpsuh | Recombining binding protein suppressor of hairless (Drosophila) | DNA recombination |
| Madh4  | MAD homolog 4 | Morphogenesis |
| Gna13  | Guanine nucleotide binding protein, α 13 | GTP binding |
| Gnao   | Guanine nucleotide binding protein, α 0 | GTP binding |
| Sh3bp1 | SH3 domain-binding protein 1 | GTPase activator |
| Notch1 | Notch gene homolog 1 | Organogenesis |
| Eif2ak2| Eukaryotic translation initiation factor 2 α kinase 2 | Immune response |
| Sca2   | Small inducible cytokine A2 | Inflammatory response |
| Pia2g7 | Phospholipase A2 group VII (platelet-activating factor acetylhydrolase) | Inflammatory response |
| Apba2  | Amyloid β (Aβ) precursor protein-binding, family A, member 2 | Intracellular protein traffic |
| Hef1   | Heparin-binding epidermal growth factor-like growth factor | Membrane fraction |
| Klf1b  | Kinesin heavy chain member 1B | Microtubule motor |
| Citel2 | Cbp/p300-interacting transactivator, with Glu/Asp-rich C-terminal domain, 2 | Nucleus |
| Psmc4  | Proteasome (prosome, macropain) 26S subunit, ATPase, 4 | Nucleus |
| Crat   | Carnitine acetyltransferase | Peroxisome |
| Acox   | Acyl-Coenzyme A oxidase | Peroxisome |
| Cd63   | CD63 antigen | Plasma membrane |
| Cd152  | CD152 antigen | Plasma membrane |
| Cd14   | CD14 | Posttranslational targeting |
| Ppip8  | Protein tyrosine phosphatase, nonreceptor type 8 | Protein tyrosine phosphatase |
| Expi   | Extracellular proteinase inhibitor | Proteinase inhibitor |
| Es1    | Esterase 1 | Serine esterase |
| Serpinb2| Serine (or cysteine) proteinase inhibitor, clade B (ovalbumin), member 2 | Serine protease inhibitor |
| Foxd3  | Forkheadbox D3 | Transcription regulation |
| Hnf4   | Hepatic nuclear factor 4 | Transcription regulation |
| Hoxb4  | Homeobox B4 | Transcription regulation |
| Kif1   | Kruppel-like factor 1 (erythroid) | Transcription regulation |
| Mid1   | Midline 1 | Transcription regulation |
| Nfatc3 | Nuclear factor of activated T cells, cytoplasmic 3 | Transcription regulation |
| Stat4  | Signal transducer and activator of transcription 4 | Transcription regulation |
| Tcf1   | Transcription factor 1 | Transcription regulation |
| Xbp1   | X-box binding protein | Transcription regulation |
| Ybx1   | Y-box protein 1 | Transcription regulation |
| Zfp308 | Zinc finger protein 38 | Transcription regulation |
| Zfp100 | Zinc finger protein 100 | Transcription regulation |
| Zfp216 | Zinc finger protein 216 | Transcription regulation |
| ZNF183 | Zinc finger protein 183 (RING finger, C3HC4 type) | Transcription regulation |
| Hoxd12 | Homeobox D12 | Transcription regulation |
| LATS2  | Serine/threonine kinase KPM | Tumor suppressor |

Some abbreviations are explained in the Table 3 footnote.

Northern blot analysis was performed as confirmation of the microarray results obtained for several novel genes (ID1567, ID2131, ID1199, and ID1457). For example, ID2131 (which contains a GTW motif of G-protein receptor [GPR1]/FUN34/yaah family proteins) is dramatically down-regulated during differentiation, and ID1567 (containing a KGR motif) and ID1457 are up-regulated during EML cell differentiation (Figure 5). Novel gene ID1199 showed similar expression levels before and after induction of differentiation.

While the EML cell culture system has proved useful in identifying changes in gene expression during hematopoietic differentiation, it is nonetheless an immortalized cell line and...
netted 1255 different gene sequences from the subtraction effort. Enriched in regulatory genes and novel genes. In aggregate, we commonly expressed genes, leaving a residual that was relatively undertook a library subtraction step that successfully removed most on the primary sorted mouse bone marrow cells. The samples in each lane are as indicated on the figure. The probes used for hybridization are listed to the right of each panel. Hybridization with β-actin serves as a control for loading.

Figure 5. (A) Graphic representation of hybridization intensity in pixels of 21 genes to the cDNA microarray. Values for LRH are depicted by bars to the left; for Lin* Hoechst^high rhodamine^bright (LRB), by bars to the right. Gene name or novel gene ID number is indicated to the left. (B) Southern blot of amplified cDNAs derived by reverse-transcription PCR of entire mRNA populations from 2 cell preparations from primary sorted mouse bone marrow cells. The samples in each lane are as indicated on the figure. The probes used for hybridization are listed to the right of each panel. Hybridization with β-actin serves as a control for loading.

thus may not accurately represent normal hematopoietic cells. To compare the changes in gene expression observed in EML cells with normal hematopoiesis, we analyzed RNA from sorted primary bone marrow cells. We obtained 2 pools of cDNAs from sorted primary mouse bone marrow cells: Lin* Hoechst^low rhodamine^bright (LRB), representing late-stage progenitor cells; and LRH, representing more primitive progenitors. These cDNAs were amplified by PCR with the use of primers specific to adaptor sequences and concomitantly labeled with fluorescent dyes. The LRH and LRB pools were competitively hybridized to the cDNA microarray. Hybridizations were performed in triplicate, and data were normalized to internal control (GAPD). This analysis revealed differences in expression between the LRH and LRB preparations of a number of key regulatory transcription factors, including Hox, Klf, and Sox family genes, Evi-1, Tal-1, GATA-1, and Kars (Figure 5A). In addition, a number of novel genes (designated by ID number) were differentially expressed, including ID2131 and ID1457, which were up-regulated and down-regulated, respectively, during EML cell differentiation (Figure 4). To confirm these microarray results, samples of amplified cDNA from LRH and LRB cells were fractionated by gel electrophoresis and then subjected to Southern blot analysis with specific cDNAs used as probes. These data (Figure 5B) support the microarray data in that they demonstrate differential expression between LRH and LRB preparations.

Discussion

We have described the creation of a resource for the in-depth study of gene expression in early hematopoietic cells that should be useful in the study of the molecular regulation of myeloid cell differentiation. Several features of this work are notable and essentially novel. First, the library that we exploited for the creation of the subtracted library represents an early stage of hematopoietic differentiation distinct from that used previously. Second, we undertook a library subtraction step that successfully removed most commonly expressed genes, leaving a residual that was relatively enriched in regulatory genes and novel genes. In aggregate, we netted 1255 different gene sequences from the subtraction effort. Given that more than 50% of the clones picked were nonredundant, it is likely that further sequencing of clones from this subtracted library will allow isolation of additional interesting genes.

Third, we have been successful in creating a glass slide–based microarray from the gene sequences we have isolated. To complement the clones from the subtracted library, we have added genes from a variety of other sources (Table 5). We have tested the utility of this array in 2 initial hybridization experiments. The first identifies genes that are up-regulated or down-regulated during ATRA/IL-3–induced differentiation of EML cells. The second documents transcriptional differences between sorted primary bone marrow cells. This investigation is continuing. However, some of our initial results with the microarray have been confirmed by Northern blot analysis (Figure 4D) or by Southern blot analysis of cDNA populations (Figure 5B), which attests to the validity of the microarray-based quantitation of mRNA or cDNA copies.

Fourth, we have created a Web-accessible database for the genes on the microarray. Using this database, one can download a list of genes present on the array and can query to obtain information regarding specific genes. This Web site represents the starting point for a variety of features, including posting of downloadable microarray data and accrual of information on genes important to hematopoietic progenitor and myeloid cell biology.

A remarkable feature of our sequence analysis was the high number of novel gene sequences present in the subtracted library. This should prove to be an important resource for the isolation of genes that play regulatory roles in early hematopoiesis. Initial protein motif analysis reveals the presence of numerous interesting motifs (Table 4; Figure 3) within these genes. Also remarkable is the paucity of growth factor receptors or cytokines among the known genes in the subtracted library. This is likely to be due to their being present in the driver or to their lack of expression in the LRH library. Our finding of Ephrin-B1 in LRH is novel. Previous studies have shown that a related transmembrane ligand, Ephrin B2, is expressed in certain leukemias and lymphomas. It has also been shown that the receptor for Ephrin B2, EphB4 (hepatoma transmembrane kinase), is expressed on human erythroid progenitors cord blood cells and that it was regulated by SCF.34,35 However, no report of expression of EphB1, the receptor for Ephrin B1, in hematopoietic cells has been made. The role of this signaling system in hematopoietic cells is unknown. Interestingly, in the subtracted library, we also identified Nsp3, which encodes a protein that couples Eph receptors to Ras, further suggesting that this is an important pathway in early hematopoietic cells.

Our studies complement and extend data reported by Phillips et al.,32 who reported on 2119 nonredundant gene products and the creation of a Stem Cell Database as a repository for these sequences. In aggregate, our effort, combined with theirs, provide an abundance of cloned sequences from early hematopoietic progenitors that allow for investigation into the molecular control of hematopoiesis.

Acknowledgments

We thank Troy Moore of Research Genetics for the contribution of the cDNA clone set used for the subtractive hybridization and arraying; Sherman Weissman for use of the LRH library; Nathan Lawson for construction of the EML and derivative libraries; Barbara Degar for contributing the LRH, LRB, and Lin* cDNA; Bernard Forget for helpful comments and suggestions; and Janet Hager and the Yale Microarray Facility for the arraying of cDNAs.
References

1. Randolph T. Acute promyelocytic leukemia (AML-M3), part 1: pathophysiology, clinical diagnosis, and differentiation therapy. Clin Lab Sci. 2000;13:98-105.

2. Randolph T. Acute promyelocytic leukemia (AML-M3), part 2: molecular defect, DNA diagnosis, and proposed models of leukemogenesis and differentiation therapy. Clin Lab Sci. 2000;13:106-116.

3. Krause D, Kapadia S, Raj N, May W. Regulation of CD34 expression in differentiating M1 cells. Exp Hematol. 1997;25:1051-1061.

4. Hino M, Nishizawa Y, Tatsuni N, Tojo A, Morii H. Down-modulation of c-kit mRNA and protein expression by erythroid differentiation factor/activin A. FEBS Lett. 1995;374:69-71.

5. Tsai S, Fero J, Bartelmez S. Mouse Jagged2 is differentially expressed in hematopoietic progenitors and endothelial cells and promotes the survival and proliferation of hematopoietic progenitors by direct cell-to-cell contact. Blood. 2000;96:950-957.

6. Matsumura I, Kanakura Y, Kato T, et al. Growth response of acute myeloblastic leukemia cells to recombinant human thrombopoietin. Blood. 1995;86:703-709.

7. Spangrude G, Aihara Y, Weissman I, Klein J. The stem cell antigens Sca-1 and Sca-2 subdivide thymic and peripheral T lymphocytes into unique subsets. J Immunol. 1988;141:3697-3707.

8. Zinovyeva M, Zijlmans J, Fibbe W, Visser J, Bel-yavskiy A. Analysis of gene expression in subpopulations of murine hematopoietic stem and progenitor cells. Exp Hematol. 2000;28:318-334.

9. Kaipainen A, Korhonen J, Pajusola K, et al. The related FLT4, FLT1, and KDR receptor tyrosine kinases show distinct expression patterns in human fetal endothelial cells. J Exp Med. 1993;178:2077-2088.

10. Artavasidou S, Rossen M, Lake R. Notch signaling: cell fate control and signal integration in development. Science. 1999;284:770-776.

11. Jacobs-Hebler S, Weisman S, and proposed models of leukemogenesis and differentiation therapy. Clin Lab Sci. 2000;13:106-116.

12. Chaturvedi P, Reddy M, Reddy E. Src kinases and not Jak5 activate STAT5a in IL-3-induced myeloid proliferation. Oncogene. 1998;17:1749-1758.

13. Zhang D, Hohaus S, Voso M, et al. Function of PU.1 (Sp1), C/EBP, and AML1 in early myelopoiesis: regulation of multiple myeloid CSF receptor promoters. Curr Top Microbiol Immunol. 1996;211:137-147.

14. Clarke S, Gordon S. Myeloid-specific gene expression. J Leukoc Biol. 1998;63:153-168.

15. Domen J, Weissman I. Self-renewal, differentiation or death: regulation and manipulation of hematopoietic stem cell fate. Mol Med Today. 1999;5:201-208.

16. Lieberman D, Hoffman-Liebermann B. Genetic programs of myeloid cell differentiation. Curr Opin Hematol. 1994;1:24-32.

17. DeRisi J, Iyer V, Brown P. Exploring the metabolic and genetic control of gene expression on a genomic scale. Science. 1997;278:680-686.

18. Tsai S, Bartelmez S, Silitnicka E, Collins S. Lymphohematopoietic progenitors immortalized by a retroviral vector harboring a dominant-negative retinoic acid receptor can recapitulate lymphoid, myeloid, and erythroid development. Genes Dev. 1994;8:2831-2841.

19. Subrahmanyan Y, Baskaran N, Newburger P, Weissman S. A modified method for the display of 3′-end restriction fragments of cDNAs: molecular profiling of gene expression in neutrophils. Methods Enzymol. 1999;303:272-297.

20. Leemhuis T, Yoder M, Gribskov S, Aguero B, Eder P, Sour E. Isolation of primitive human bone marrow hematopoietic progenitor cells using Hoechst 33342 and Rhodamine 123. Exp Hematol. 1996;24:1215-1224.

21. Dega R, Baskaran N, Hulspsas R, Quesenberry P, Weissman S, Forget B. The homeodomain gene Pitx2 is expressed in primitive hematopoietic stem/progenitor cells but not in their differentiated progeny. Exp Hematol. 2001;29:994-992.

22. Bonaldo M, Lennon G, Soares M. Normalization and subtraction: two approaches to facilitate gene discovery. Genome Res. 1996;6:791-806.

23. Altschul S, Gish W, Miller W, Myers E, Lipman D. Basic local alignment search tool. J Mol Biol. 1990;215:402-410.

24. Bonaldo M, Lennon G, Soares M. Normalization and subtraction: two approaches to facilitate gene discovery. Genome Res. 1996;6:791-806.

25. Altschul S, Gish W, Miller W, Myers E, Lipman D. Basic local alignment search tool. J Mol Biol. 1990;215:402-410.

26. Weissman S, Perkins A. Stem cell transcription. In: Quesenberry P, Stein G, Forget B, Weissman S, eds. Stem Cell Biology and Gene Therapy. New York, NY: John Wiley and Sons, 1998.

27. Kim J, Hui P, Yue D, et al. Identification of candidate target genes for EVII, a zinc finger oncoprotein, using novel selection strategies. Oncogene. 1998;17:1527-1538.

28. Tsai S, Collins S. A dominant negative retinoic acid receptor blocks neutrophil differentiation at the promyelocyte stage. Proc Natl Acad Sci U S A. 1993;90:7153-7157.

29. Bleker J. Isolation, genomic structure, and expression of human erythroid Kruppel-like factor. DNA Cell Biol. 1996;15:347-352.

30. Sauvageau G, Thorsteindottir U, Eaves C, et al. Overexpression of HOXB4 in hematopoietic cells causes the selective expansion of more primitive populations in vitro and in vivo. Genes Dev. 1995;9:1735-1765.

31. Reimold A, Iwakoshi N, Manis J, Vallabahajosula P, Szomolanyi-Tsuda E. Plasma cell differentiation requires the transcription factor XBP-1. Nature. 2001;412:300-307.

32. Phillips R, Ernst R, Bruk B, et al. The genetic program of hematopoietic stem cells. Science. 2000;288:1635-1637.

33. Steube K, Meyer C, Habig S, Uphoff C, Dreher X. Expression of receptor tyrosine kinase HTK (hepatoma transmembrane kinase) and HTK ligand by human leukemia-lymphoma cell lines. Leuk Lymphoma. 1999:33:371-376.

34. Sakano S, Serizawa R, Inada T, et al. Characterization of a ligand for receptor protein-tyrosine kinase HTK expressed in immature hematopoietic cells. Oncogene. 1996;13:813-822.

35. Inada T, Iwama A, Sakano S, Ohno M, Sawada K, Suda T. Selective expression of the receptor tyrosine kinase HTK on human erythroid progenitor cells. Blood. 1997;90:2757-2765.

36. Maun NA, Lawson ND, Berliner N. Immediate-early gene expression during myeloid differentiation of EML cells: evidence for a synergistic induction by IL-3 and ATRA [abstract]. Blood. 1998;92(suppl 1):194a.
Development of a murine hematopoietic progenitor complementary DNA microarray using a subtracted complementary DNA library

Xianyong Ma, Tupur Husain, Hui Peng, Sharon Lin, Olya Mironenko, Noel Maun, Steven Johnson, David Tuck, Nancy Berliner, Diane S. Krause and Archibald S. Perkins