Genomic and proteomic analysis of the myeloid differentiation program

Zheng Lian, Le Wang, Shigeru Yamaga, Wesley Bonds, Y. Beazer-Barclay, Yuval Kluger, Mark Gerstein, Peter E. Newburger, Nancy Berliner, and Sherman M. Weissman

Although the mature neutrophil is one of the better characterized mammalian cell types, the mechanisms of myeloid differentiation are incompletely understood at the molecular level. A mouse promyelocytic cell line (MPRO), derived from murine bone marrow cells and arrested developmentally by a dominant-negative retinoic acid receptor, morphologically differentiates to mature neutrophils in the presence of 10 μM retinoic acid. An extensive catalog was prepared of the gene expression changes that occur during morphologic maturation. To do this, 3′-end differential display, oligonucleotide chip array hybridization, and 2-dimensional protein electrophoresis were used. A large number of genes whose mRNA levels are modulated during differentiation of MPRO cells were identified. The results suggest the involvement of several transcription regulatory factors not previously implicated in this process, but they also emphasize the importance of events other than the production of new transcription factors. Furthermore, gene expression patterns were compared at the level of mRNA and protein, and the correlation between 2 parameters was studied. (Blood. 2001;98:513-524)

Introduction

Studies of normal myeloid maturation from many laboratories have identified genes that may play critical roles in myeloid differentiation.[1-4] Current studies suggest that these events are dependent on a cascade of molecular changes that involve complex modulation of mRNA transcription. Furthermore, studies of acute leukemia have suggested that the disease arises from the accumulation of myeloid precursors arrested at early stages of differentiation and associated, in many cases, with chromosomal rearrangements that alter the structure of specific transcription factors.[5] Nevertheless, the molecular events underlying the production of mature myeloid cells are not well understood and appear to use interacting pathways and networks, the elucidation of which requires an extensive description of the molecular components available to the myeloid cell.

An extensive body of information is accumulating with respect to gene expression profiles of mammalian cells. However, much of the information available in public databases has been accumulated by the use of techniques such as single oligonucleotide chips or cDNA arrays that measure fewer than 6000 of potentially 30 000 to 120 000 transcripts. The more limited range of analyses reported by the serial analysis of gene expression (SAGE)[6,7] technique accurately estimates changes in levels of the more abundant mRNAs but requires extensive redundant analyses to measure changes in the patterns of expression of scarce mRNAs. We have used a modified polymerase chain reaction (PCR)–based cDNA differential display (DD) method in which single restriction fragments derived from the 3′ end of cDNAs are separated on a sequencing gel.[8,9] Bands from the gel can be identified initially by sequencing, but then comparison of patterns from different samples can be made without further sequencing. This sensitive and reproducible method detects, in principle, most cDNAs regardless of whether they are represented in existing databases.

Systematic analysis of the function of genes can also be performed at the protein level. This approach has the advantage of being closest to function, because proteins perform most of the reactions necessary for the cell. The most common method of proteome analysis is the combination of 2-dimensional gel electrophoresis (2DE) to separate and visualize protein and mass spectrometry (MS) for protein identification.[10] Several such analyses of yeast and of normal or malignant mammalian cells have been performed. To date, however, there have been few studies in which both mRNA and protein have been compared by applying analyses to the same samples. The studies of Anderson[11] and Gygi[12] showed that there is not a good correlation between mRNA and protein levels, in yeast or human liver cells. However, other analyses disagree with this conclusion (Greenbaum et al, manuscript submitted, and Futcher et al[14]). Furthermore, global correlations between changes in mRNA and protein levels have not been examined during the execution of any developmental program.

The MPRO cell line was derived by transduction of a dominant-negative retinoic acid receptor construct into normal mouse bone marrow cells. It is a granulocyte-macrophage colony-stimulating factor (GM-CSF)–dependent line arrested at a promyelocytic stage of development.[15,16] After treatment with all-*trans* retinoic acid (ATRA) most of the cells acquire the morphology of mature

From the Department of Genetics, Boyer Center for Molecular Medicine, the Section of Hematology, Department of Internal Medicine, and the Department of Molecular Biophysics and Biochemistry, Yale University School of Medicine, New Haven, CT; the Department of Pediatrics, University of Massachusetts Medical School, Worcester, MA; and Gene Logic, Gaithersburg, MD.

Submitted December 4, 2001; accepted March 28, 2001.

Supported by grants from the National Institutes of Health (NIH) (CA42556) and Gene Logic (A143558, DK54369, and HL63357). Z.L. is supported by NIH grant HL 63357. P.E.N. is supported by NIH grant DK 54369, grants from the Arthritis Foundation and the Charles H. Hood Foundation, and the Pierce Family Cancer Research Fund. M.G. is supported by the Keck Foundation and by NIH grant GM54160-04.

L.W. and S.Y. contributed equally to this research.

Reprints: Sherman M. Weissman, Department of Genetics, Boyer Center for Molecular Medicine, Yale University School of Medicine, Rm 336, 295 Congress Ave, New Haven, CT 06536-0812; e-mail: sherman.weissman@yale.edu.

The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 U.S.C. section 1734.

neutrophils and begin to produce neutrophil lactoferrin and gelatinase, 2 proteins characteristic of neutrophil secondary granules.[17] As such, it offers a valuable model for studying neutrophil differentiation in vitro.

We now report the analysis of mRNA expression changes during the process of MPRO cell maturation to neutrophils and compare the results with a limited analysis of cellular protein composition. mRNA expression changes were studied by combining the use of oligonucleotide arrays and DD. A database (dbMC) with comprehensive genomic information for myeloid differentiation program was constructed (accessible at http://www.bioinfo.mbb.yale.edu/expression/neutrophil). We have grouped the changes in mRNA levels of a large number of genes into 6 patterns, with implications for the genetic program of myeloid differentiation.

We also compared 2-dimensional high-resolution gel electrophoretograms from control cells and cells differentiated for 72 hours in the presence of ATRA. Fifty protein spots whose relative intensity changed prominently during differentiation were examined by mass spectrometry. The results suggest a poor correlation between mRNA expression and protein abundance, indicating that it may be difficult to extrapolate directly from individual mRNA changes to corresponding ones in protein levels (as estimated from 2DE).

Materials and methods

Cell lines

MPRO cells and HM-5 cells provided by Dr Schickwann Tsai (Fred Hutchinson Cancer Research Center, Seattle, WA)[15] were used throughout the study. The cells proliferated continuously as a GM-CSF–dependent cell line at 37°C in Iscoves modified Dulbecco medium (Gibco BRL, Grand Island, NY) supplemented with 5% to 10% fetal calf serum (Gibco BRL) and 10% HM-5–conditioned medium as a source of GM-CSF. Morphologic differentiation of the blocked MPRO promyelocytes was induced by treatment with 10 μM ATRA (Sigma, St Louis, MO). Controls were cultured in the absence of ATRA but with the same volume of vehicle (ethanol).

RNA isolation and differential display

After exposure to 10 μM ATRA for 0, 24, 48, or 72 hours, total cellular RNA was isolated from MPRO cells using TRIzol reagent (Life Technologies, Gaithersburg, MD). cDNA was then synthesized using a T-7 Sal-Oligo d(T) 32 primer as described previously.[8,18] The double-stranded cDNA was digested with 1 of 9 different restriction enzymes (*Apa*I, *Bgl*II, *Bam*HI, *Eag*I, *Eco*RI, *Hin*dIII, *Xba*I, *Kpn*I, and *Sph*I) and ligated to Y-shaped adaptors with a complementary overhang. DNA fragments were then amplified by PCR as described previously.[8,18] PCR products were separated on a sequencing gel of 6% polyacrylamide with 7 M urea. The gel was dried and exposed to x-ray film. Genes from differential display gels, whose maximum intensity changes equaled 2+ on a scale of 1+ to 8+, were recorded as significantly changed.[19] Individual DNA bands were recovered from the gels, amplified by PCR, and sequenced.

Oligonucleotide chip analysis of RNA samples

Ten micrograms total RNA from each sample (0, 24, 48, or 72 hours) was used to prepare cDNA. This cDNA was transcribed with T7 RNA polymerase to prepare a fluorescently labeled probe.[20,21] Each sample was hybridized to mouse array chip (Mu11K Array; Affymetrix, Santa Clara, CA) containing oligonucleotide probe sets corresponding to approximately 7000 known genes or ESTs represented by UniGene clusters.[22] cDNAs were considered present if their probe set results were rated as such by the GeneChip software (Affymetrix) and if the average difference (AD) between perfect match and mismatch probe pairs was not less 100 U. If a gene was represented by more than one array probe set, the average of all probe sets for the gene was taken. Genes with AD values between 100 and 200 were considered unchanged because of their low expression levels. Those genes with AD values equal to or more than 200 U at one time point were further studied by rescaling, threshold, and normalization methods described in the MIT Center for Genome Research Web site.[13] A value of 20 was assigned to any gene with an AD below 20 at some time point.

Bioinformatics and database development

All the sequences or gene fragments were searched using Blast against GenBank and TIGR gene indices. A database of genes or ESTs whose expression levels changed during myeloid differentiation was constructed containing information for each band or gene. This included GenBank matches, Locus Link or Unigene clusters, expression patterns, tissue distribution, synonym(s) protein name, gene name(s), notations of possible functions, poly A signal and sequence quality, and hyperlinks to the database searches, sequence trace files, and related references. All gene data were then gathered into a cluster file. Supplementary information is available at http://bioinfo.mbb.yale.edu/expression/neutrophil.

Classification and analysis of DNA fragments

Sequences from differential display analyses were classified as representing known genes, ESTs, genomic sequences, or novel genes as described.[19,23] Known genes from both differential display and arrays were clustered into 27 functional categories and searched against SWISS-PROT (http://www.expasy.cbr.nrc.ca/cgi-bin/sprot-search-ful) or PIR (http://www.pir.georgetown.edu/). Information such as function, subcellular location, family and superfamily classification, map position, similarity, synonym(s) protein name, gene name(s), and so on was recorded in a variety of databases.

Northern blot analysis

Thirty micrograms total cellular RNA per lane from time-course MPRO cells were loaded onto 1.2% formaldehyde-agarose gels, then transferred to Hybond-N+ membranes (Amersham Pharmacia Biotech, Uppsala, Sweden). After standard prehybridization, membranes were hybridized overnight at 65°C with radiolabeled cDNA probes (ordered from Research Genetics according to their dbEST Image ID). Membranes were washed at a final stringency of 60°C in 0.1 × SSC.

Immobilized pH gradient 2-dimensional gel electrophoresis and mass spectrometry

Induced MPRO cells collected at 0 and 72 hours were lysed with lysis buffer (540 mg urea, 20 mg dithiothreitol, 20 μL Pharmalyte [3-10], 1.4 mg phenylmethylsulfonyl fluoride, 1 μg each aprotinin, leupeptin, pepstatin A, and antipain 50 μg TLCK, and 100 μg TPCK/1 mL). We applied 100 μL each MPRO cell lysate (2.5×10^6cells/100 μL) to immobilized pH gradient (IPG) strips (pH 3-10 L; Amersham Pharmacia Biotech), and IPG electrophoresis was conducted for 16 hours (20 100 Vh) using an Immobiline Drystrip Kit (Amersham Pharmacia Biotech). Electrophoresis in the second dimension was carried out in a 12% sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) gel with the Laemmli-SDS continuous system in a Protean II xi 2-D cell (Bio-Rad) run at 40 mA constant current for 4.5 hours. Proteins were detected by Brilliant Blue G–colloidal staining.[24] Protein spots were excised from the gel and digested with trypsin. ACTH clip (average [M+H] 2466.70) and bradykinin (average [M+H] 1061.23) were used for calibration of peptide masses. One microliter sample digest was mixed with 1.0 μL α-cyano-4-hydroxy cinnamic acid (4.5 mg/mL in 50% CH_3CN, 0.05% TFA) matrix solution and 1 μL calibrants (100 fmol) each. The spectra of the peptides were acquired in reflector/delayed extraction mode on a Voyager-DE STR mass spectrometer (Perseptive Biosystems, Foster City, CA). Peptides were identified using the ProFound search engine.[39]

Results

Differentiation of MPRO cells

Figure 1 illustrates the morphologic changes in an MPRO cell population representative of those used for RNA expression analysis. Undifferentiated MPRO cells resembled promyelocytes under the light microscope (Figure 1A). After induction with ATRA for 24 hours, the cells morphologically differentiated into metamyelocytes (Figure 1B). At 48 hours, the cells further developed into metamyelocytes and band neutrophils (Figure 1C). At 72 hours, nearly 100% of MPRO cells became mature neutrophils (Figure 1D).

Identification of mRNAs by differential display assay

MPRO cellular mRNA was analyzed at 0, 24, 48, and 72 hours after ATRA treatment. Nine restriction enzymes were used in a 3′-end DD approach. During MPRO differentiation, 1109 fragments corresponding to 837 transcripts were found to change substantially in expression levels (Figure 2). These represented approximately 279 known genes, 112 ESTs, and 59 putative new genes, each with a perfect or fair polyadenylation signal at an appropriate distance from the oligo-dT priming site. The gene information detected by DD was collected in database dbMCd.

Identification of mRNAs by oligonucleotide chip assay

We used an oligonucleotide chip containing 13 179 probe sets corresponding to approximately 7000 murine genes to analyze patterns of mRNA expression in the same RNA samples used for DD. The information obtained by oligonucleotide arrays was collected in the database dbMCa.

We clustered the genes by their similarity to idealized expression patterns. For instance, the expression pattern of an ideal gene that is overexpressed (high) at time 0 and underexpressed (low) at 24, 48, and 72 hours, would be high-low-low-low (HLLL). Overall we have (2^4-2) idealized patterns excluding HHHH and LLLL. Pearson correlation was used as the measure of similarity of each gene expression pattern, $x = (x_1,x_2,x_3,x_4)$ to each of the 14 idealized patterns $y = (y_1,y_2,y_3,y_4)$. The 4 entries of x and y corresponded to the 4-dimensional gene expression levels at 0, 24, 48, and 72 hours, respectively. Each gene was assigned to a cluster labeled by the idealized pattern that had the maximal correlation with that gene. We selected only genes that hybridized well compared with the background (considered "present" by GeneChip software) and had maximal AD amplitude greater than 200 U in at least 1 of the 4 stages. We further tabulated the 14 patterns according to whether the gene expression changed at early (0-hour), intermediate (24- and 48-hour), and late (72-hour) time points and whether gene expression monotonically increased (up-regulated), monotonically decreased (down-regulated), or was not monotonic (transient). Table 1 shows 8 clusters of 104 genes that had significant changes of mRNA levels, arranged according to the temporal stage and the monotonic/transient changes of expression levels.

Principal component analysis determined whether we could comprehensively present multidimensional data (4-dimensional in our case) in a simple 2-dimensional graph. First, we found the 4 principal components, which were the axes of the most compact 4-dimensional ellipsoid that encompassed the 4-dimensional cloud of data. Each axis was a different linear combination of the original 4 variables. Then we verified that the first 2 principal components (the first 2 largest axes of the ellipsoid) captured most (95.2%) of the variation of the data. Therefore, the data could be faithfully projected (with a minor loss of information) into a 2-dimensional graph, with the 2 largest principal components as the x- and y-axes. As shown in Figure 3, genes tend to coalesce in clusters, according to their labels determined by their similarity to an ideal expression pattern. In summary, a genomic (global) picture of the distribution of genes according to their similarity to predetermined idealized multidimensional expression patterns is concisely displayed in a 2-dimensional graph.

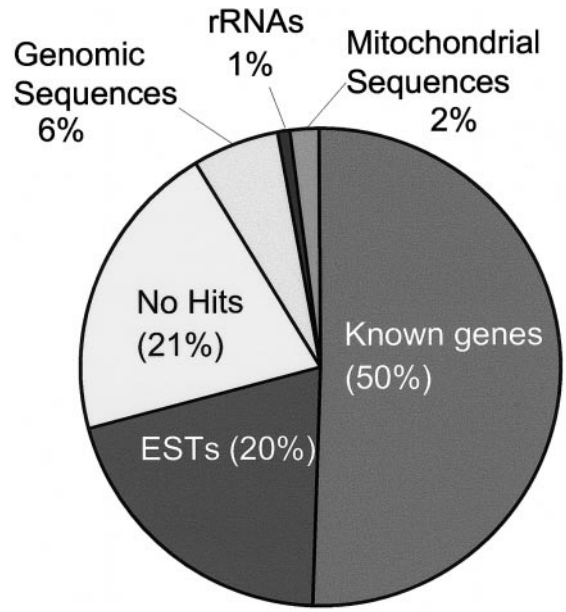

Figure 2. Distribution of genes obtained by DD assay. MPRO cell mRNA was analyzed at 0, 24, 48, and 72 hours after ATRA treatment; 1109 fragments corresponding to 837 transcripts were found to change substantially in expression levels. The total 837 transcripts were classified into 6 categories according to the bioinformatic analysis. Percentages show the gene distributions in these 6 categories. Information for each transcript was collected in database dbMCd.

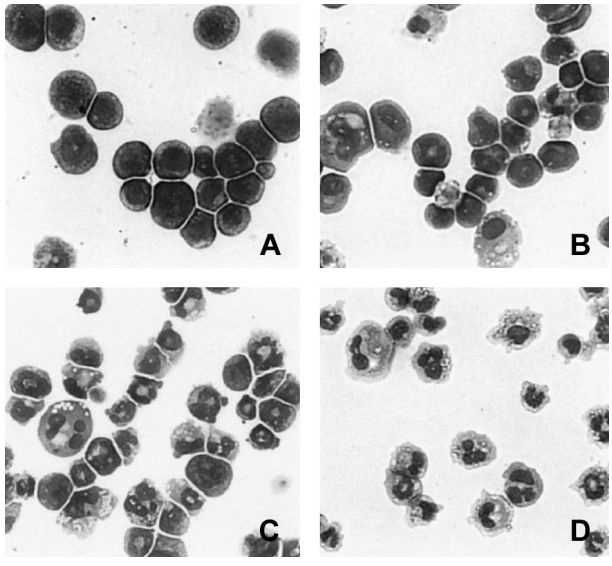

Figure 1. Morphology of MPRO cells during differentiation. MPRO cells were induced as described in "Materials and methods," concentrated by cytospin, and Wright-Giemsa stained. (A) Uninduced MPRO cells. (B) MPRO cells induced with ATRA for 24 hours. (C) MPRO cells induced with ATRA for 48 hours. (D) MPRO cells induced with ATRA for 72 hours.

Table 1. Genes differently regulated during the different stages of mouse promyelocytic cell line differentiation process

Category	Timing		
	Early	Middle	Late
Up-regulation	LHHH (n = 10) *Mad* P2rx1 **Itgb2 Il1r2 Lcn2** *Itpr5* Cebpb H2-D Etohi6 Zyx	LLHH (n = 6) *Piral Cybb* Pfc Pira5 Cd53 *Ifngr2*	LLLH (n = 13) *Il1a Csflr* Ii *Ctsl* **S100a8** L-CCR *Ctss* **Aldo1 Rac2** Fpr1 **Ctsd Ubb Ptmb4**
Down-regulation	**HLLL** (n = 11) Tcrg-V4 Ly64 Ctsg Spi2-1 Mcpt8 Myc Myb Tlr4 Npm1 Erh Hsp60	**HHLL** (n = 1) **Mpo**	**HHHL** (n = 37) Actx *Irf2* EL2 **Rpl19 Actb Ly6e** *Atf1* Hist2 Psma2 **Gnas** *Zfp36 Il4ra Ltbr* Shfdg1 Max **Rps8** Csf2rbl Slpi Tctex1 **Tpi** Btf3 *Cntf* Gys3 Slc10a1 Ctsb Sepp1 Rtn3 Ccnb2 S100a9 Cf11 Hist5-2ax Rela Copa Gstm1 **Gnb2-rs1** Grn RPL8
Transient		**LLHL** (n = 9) Sell *Klf2* Pira6 Pirb Lst1 Ltf Sema4d *Stat6* Mmp9 **LHHL** (n = 17) *Cebpa* **Lyzs Fcgr3** Arf5 Lamp1 Stat3 Csf2ra Osi **Actg** Sfpi1 Gpx3 Ptprc Prtn3 Irf1 Rps6ka1 Ltb4r Myln	

Arrays of Affymetrix Mu11k containing 13103 probe sets corresponding to 12002 GenBank accessions were used for hybridization. Arrays were hybridized with streptavidin-phycoerythrin (Molecular Probes) biotin-labeled RNA and scanned. Intensity for each feature of the array was captured using Genechip software (Affymetrix), and a single raw expression level for each gene was derived from the 20 probe pairs representing each gene using a trimmed mean algorithm. For each gene, an AD of 24-, 48-, and 72-hour samples was calibrated by dividing the slope of the linear regression line for a graph with the x-axis the AD of 0-hour probe sets and the y-axis the AD of the respective time point (24, 48, or 72 hours). A threshold of 20 U was assigned to any gene with a calculated expression level below 20 because discrimination of expression below this level could not be performed with confidence.[38] Each gene expression profile was categorized as described in Tables 3, 4, and 5. For the 4 time points, the minimum AD of the relatively higher group (MIN-H) was divided by the maximum AD of the relatively low group (MAX-L), and those genes whose MIN-H/MAX-L greater than 2 were selected as meaningfully regulated. Genes were sorted in descending order based on the MIN-H/MAX-L. Genes in boldface are those whose expression level was in the top 20% (ie, maximum AD of 4 time points greater than 3000), and genes in italics are those in the bottom 20% (ie, maximum AD of 4 time points less than 300). The differentiation period was grouped into 3 stages: early (0-hour), middle (24-hour and 48-hour), and late (72-hour) stages.

AD indicates average difference; gene symbols are expanded in an Appendix at the end of this article.

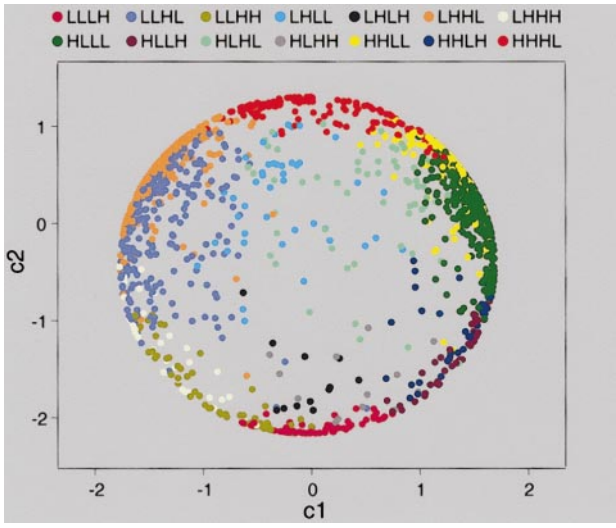

Figure 3. Gene clusters in the first 2 principal component spaces. Principal component analysis allowed us to present the multidimensional data (in this case, 4-dimensional data of each gene expression pattern) in a simple 2-dimensional graph. We derived the 4 principal components, which are a linear combination of the standardized expression intensities (zero mean and unit variance) at 0, 24, 48, and 72 hours. The first 2 principal components captured most of the variation of the data (approximately 85%). Therefore, the data can be displayed (with a minor loss of information) in a 2-dimensional graph. The first and second principal components, c1 and c2, are given by the linear combinations $c_1 = 0.747 \cdot n1 - 0.11 \cdot n2 - 0.656 \cdot n3 + 0 \cdot n4$ and $c_2 = 0.278 \cdot n1 + 0.353 \cdot n2 + 0.233 \cdot n3 - 0.863 \cdot n4$, where $n1$, $n2$, $n3$, and $n4$ are the rescaled and standardized expression levels at 0, 24, 48, and 72 hours, respectively. The axes legends c1 and c2 stand for the first 2 principal components. In this paper we used the Pearson correlation to measure the similarity of each gene with the idealized expression patterns, as opposed to the Euclidean distance we used in a previous work,[19] because clusters were better separated using this measure. In both cases, we presented the data in the 2-dimensional space of the lowest principal components. The data had a tendency to be circularly distributed when we used the Pearson correlation as a distance measure.

Correlation between array and DD analyses

We have previously demonstrated a correlation coefficient of 0.93 between visual estimates of changes in band intensity on DD and Phosphorimager System (Molecular Dynamics, Sunnyvale, CA) estimates of band intensity and a correlation coefficient of 0.88 between hybridization intensity changes of mRNA on Northern blot analyses and changes in band intensity on DD.[19] In a few cases there were clear discrepancies in the pattern of expression of a gene, as estimated by DD and by oligonucleotide chip analysis. We chose the 6 most extreme cases and examined the levels of mRNA change for these genes by Northern blot analysis (Figure 4). In 5 cases, the Northern blot results agreed with the results of the DD analysis, whereas the results of Gnb2-rs1 disagreed with the oligonucleotide array but duplicate bands from DD showed a relatively high level of expression in the 0 time sample that did not correlate with the Northern blot (Table 2). One possible explanation for these findings was the change in the relative use of different polyadenylation sites after the addition of ATRA to the MPRO cells.

Constructing a database for mRNA level changes during myeloid differentiation

Based on the data obtained above, an in-house database (dbMC) was constructed that included 2 subdatabases, dbMCd and dbMCa, for collecting gene information from DD or oligonucleotide arrays, respectively. Each entry in dbMC is accompanied by a so-called executive summary. The linkage between dbMCd and dbMCa was established by UniGene ID and cluster ID. dbMC contains the temporal expression patterns of genes during the MPRO cell differentiation process, including not only products represented in public databases but also novel transcripts.

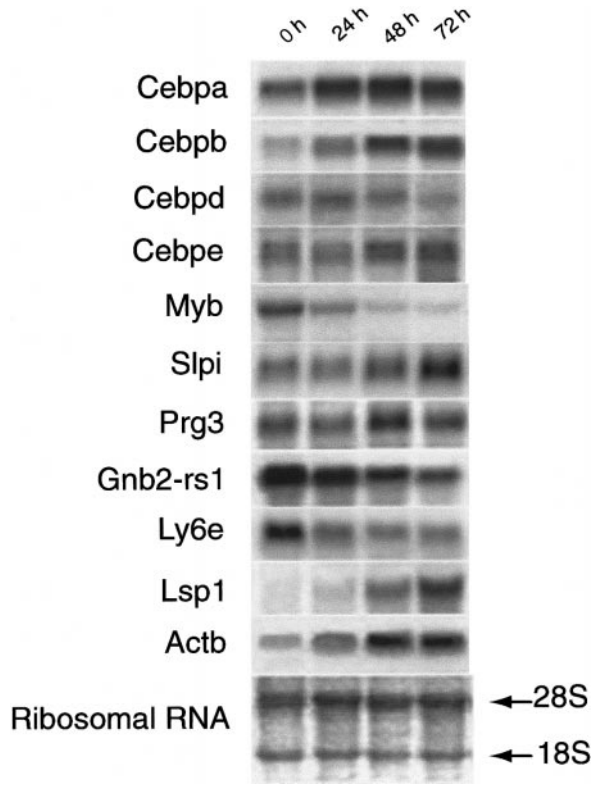

Figure 4. Northern blot analysis of selected mRNAs. Equivalent amounts of RNA from MPRO cells induced by ATRA at different time points (0 hour, 24 hours, 48 hours, and 72 hours) were resolved by formaldehyde-agarose gel electrophoresis, stained to verify the amount of loading. Eleven genes were separately probed on the RNA filters. The gene symbol of each probe was listed at the left of a related Northern blot result. Detailed information on these 11 probes was listed in Table 5. One of the RNA-blotted membrane photographs is shown with methylene blue–stained 28S and 18S RNA subunits demonstrating the quality and quantity of RNA loaded in individual lanes.

Analysis of gene expression patterns during MPRO differentiation

Many of the genes identified in this study were found in myeloid cells or were implicated in myeloid development for the first time. We detected 8 cytokines[25] and chemokines whose mRNA levels changed more than 5-fold by arrays and 2-fold by DD during the maturation of MPRO cells (see our Web site, http://bioinfo.mbb.yale.edu/expression/neutrophil). Among these were 2 members of the CC chemokine family. Interleukin-1α (IL-1α) was up-regulated at the late stage of differentiation (LLLH pattern, Table 1).

mRNA for approximately 52 receptors was detected by one or the other method. A number of the receptors known to be present on mature neutrophils showed late induction of mRNA, and their levels of induction were high, indicating that the expression of these products is a prominent event late in neutrophil maturation (Table 3). Rarely was mRNA for receptors down-regulated, consistent with myeloid maturation being accompanied by increasing responsiveness of the cell to a variety of external stimuli.

Expression of mRNA for granule proteins

Neutrophils contain several types of granules that develop at different stages of myeloid maturation.[3,17,26] Levels of mRNAs encoding secondary granule proteins, such as lactoferrin, increased as the cells matured (Table 4). The level of mRNA for Mmp9, reported as a tertiary granule protein, increased markedly between 24 and 48 hours after the induction of differentiation, whereas mRNAs for secondary granule proteins either increased less markedly or showed a maximum increase by 24 hours. mRNAs for several primary granule constituents, such as myeloperoxidase and cathepsin G, were present in unstimulated cells and decreased as the cells matured. There was a discrepancy in the measurements of proteoglycan mRNA by DD and oligonucleotide chips, but Northern blots showed that it reached a peak at 48 hours and then declined (Figure 4). Cathepsin D is reported as a primary granule protein, but its pattern of mRNA expression more closely resembled that of secondary granule constituents. In addition to known granule components, mRNAs for several other cathepsins were up-regulated during myeloid differentiation, in parallel with or later than the tertiary granule protein mRNAs.

mRNAs for transcription factors

Transcription factor genes, including several identified at the sites of consistent chromosome rearrangements in acute myeloid leukemia, have been implicated in normal myeloid differentiation and in the expression of neutrophil proteins.[2,5,27] However comprehensive information concerning the expression of these transcription factors during myeloid development is not readily available. Therefore, we compared gene names and identifiers in our databases to those of the transcription factor database Transfac (http://

Table 2. Expression patterns of genes detected by Northern blot analysis

Gene symbol	Gene accession	AD value by array				Intensity by DD			
		0 h	24 h	48 h	72 h	0 h	24 h	48 h	72 h
Cebpa	M62362	33	212	182	44	—	—	—	—
Cebpb	X62600	390	1248	1380	1903	—	—	—	—
Cebpd	X61800	157	262	168	430	—	—	—	—
Cebpe	—	—	—	—	—	—	—	—	—
Myb	M12848	892	356	230	435	—	—	—	—
Slpi	U73004	617	501	783	402	**1**	**2**	**3**	**3**
Prg3	W45834	153	259	339	345	**5**	**1**	**1**	**2**
Gnb2-rs1	X75313	4231	3623	3215	3403	**4**	**4**	**1**	**1**
Ly6e	U04268	3061	5391	2844	1282	**3**	**2**	**1**	**1**
Lsp1	M90316	65	376	840	28	**2**	**3**	**5**	**6**
Actb	X03765	3095	3588	3976	2434	**1**	**2**	**3**	**2**

Gene symbol and gene accession refer to National Center for Biotechnology Information databases and, in particular, to Locus Link. AD value is the average difference in the value of hybridization intensity between the set of perfectly matched oligonucleotides and the set of mismatched oligonucleotide in the oligonucleotide array. Band intensities from DD were semiquantified on a scale from 1 (+) to 8 (++++++++). These estimates are shown as boldface numbers in this table.[19] Both AD value and intensity of genes were studied at 4 time points corresponding to MPRO cells induced for the indicated times.

DD indicates differential display; MPRO, mouse promyelocytic cell line; for gene symbols, see the Appendix at the end of this article.

Table 3. Receptors expressed during myeloid differentiation process

Maximal fold change	Gene symbol	Gene accession	AD value by array			
			0 h	24 h	48 h	72 h
Less than 2						
	Bzrp	D21207	641	658	881	887
	Cmkar4	X99581	508	447	378	684
	Crry	M34173	433	384	506	506
	Csf2rb1	M34397	318	345	410	241
	Htr5a	Z18278	188	272	273	339
	M6pr	X64068	536	409	408	649
	MPPIR	AA116789	232	84	63	381
	TCRGB	M26053	165	212	244	299
	Tnfrsfla	M59377	**0**	**1**	**1**	**1**
2 or more, less than 3						
	Cmkbr1	U28404	221	244	504	638
	Crhr	X72305	121	200	250	355
	Csf2ra	M85078	171	372	402	254
	Ebi3	AF013114	187	270	428	148
	Grid1	D10171	128	164	150	257
	Ifngr	J05265	141	263	327	251
	Il2rg	U21795	205	184	231	477
	Ldlr	X64414	1399	1653	1665	3968
	P40-8	J02870	849	677	381	640
	Plaur	X62701	312	443	476	734
	Rarg	M34476	102	113	114	218
	Srb1	U37799	126	232	132	258
3 or more, less than 4						
	Cr2	M29281	83	138	243	77
	Csf2rb2	M29855	209	249	437	111
	Fcer1g	J05020	2398	2766	3365	8751
	Fcgr2b	X04648	1703	1652	1431	4605
	Ifngr2	U69599	**1**	**2**	**2**	**3**
4 or more, less than 5						
	Nr4a1	X16995	96	188	202	401
5 or more						
	I11r2	X59769	482	1796	2872	3818
	C5r1	L05630	185	434	808	1078
	Drd2	X55674	0	0	0	219
	Fcgr3	M14215	**1**	**1**	**1**	**2**
	Fpr1	L22181	0	89	141	671
	GCR	AA240711	**2**	**0**	**0**	**0**
	L-CCR	AA034646	48	175	314	2056
	NMDARGB	AAB20211	**2**	**2**	**0**	**0**
	P2rx1	X84896	79	346	530	744
	Pira1	U96682	0	43	172	378
	Pira5	U96686	274	391	954	1874
	Pira6	U96687	122	635	2014	1716
	Pirb	U96689	191	445	966	747
	Sell	M25324	46	104	570	20
	Tcrg-V4	M54996	1650	78	65	315

Receptors are identified as present whose maximal AD values were more than or equal to 200 U in this study. Genes were sorted by their expression patterns as follows: first by the average difference value, then by the difference between minimum and maximum AD for the 4 time points, and last by the alphabetical order of gene symbols. Genes were ordered according to the maximal fold change of AD values. Abbreviations of gene names are taken from gene symbols listed in the Locus Link portion of the National Center for Biotechnology Information database where available. Numbers in bold denote those gene expression patterns obtained by differential display rather than by oligonucleotide array assays. The other information is presented as in the legend to Table 2.

AD indicates average difference; gene symbols are expanded in an Appendix at the end of this article.

www.transfac.gbf-braunschweig.de/TRANSFAC) and determined which factors contained in this database were present at detectable levels in MPRO cell mRNA, using Affymetrix software for the criteria for inclusion of mRNAs from approximately 200 murine transcription factors probe sets on the oligonucleotide chip. Of these, 54 were expressed and 13 showed changes of 3-fold or more in chip signal (Table 5).

The changes in certain transcription factors, such as the moderate down-regulation of *myb* and *myc* and the up-regulation of the Max dimerization protein MAD, were consistent with the shift of the cells from a proliferative to a differentiated state.[28] Some changes are more difficult to explain, such as the up-regulation of DP1, a partner for E2f factors in the regulation of S-phase genes, and the mild up-regulation of the *Id* genes, commonly associated with an inhibition of differentiation by competition with bHLH transcriptional activators.[29]

The C/EBP family has been extensively studied with respect to myeloid differentiation.[2,30] Absolute levels of the C/EBP α and δ mRNAs were low, probably at the borderline of significance for the oligonucleotide chip assay, whereas the level of C/EBP β appeared higher. In addition, there were discrepancies between the chip

Table 4. Granule constituents expressed during mouse promyelocytic cell line cell differentiation

Granule constituent	Gene symbol	Gene accession	AD value by array			
			0 h	24 h	48 h	72 h
Azurophil (primary) granules						
	Man2c1	AA161860	178	134	99	164
	Ctsb	M65270	442	480	595	389
	Ctsd	X52886	214	1087	1828	2784
	Ctsg	M96801	1509	405	46	286
	E12	U04962	658	1273	843	157
	E1a2	AA689016	47	159	134	163
	Gus-s	M63836	544	226	266	254
	Lyzs	M21050	0	1	1	3
	Mcpt8	X78545	831	268	66	491
	Mpo	X15378	3788	3009	776	692
	Prg	X16133	2621	2653	2920	9859
Possible granule proteins						
	Ctsc	AA144887	252	194	342	576
	Ctse	X97399	**1**	**3**	**4**	**5**
	Ctsh	U06119	45	124	195	156
	Ctsl	X06086	16	11	31	237
	Ctss	AA089333	12	9	88	463
Specific secondary granules						
	Cpa3	J05118	621	270	90	801
	Cd36l2	AB008553	113	93	157	187
	Cnlp	X94353	80	479	704	626
	Cybb	U43384	8	24	91	128
	Ear2	—	**0**	**1**	**1**	**2**
	Fpr1	L22181	178	220	235	846
	Itgb2	X14951	**0**	**2**	**4**	**2**
	Lcn2	W13166	916	3513	3931	6036
	Ltf	J03298	19	162	333	138
	MBP	W45834	**5**	**1**	**1**	**2**
	Mmp13	X66473	44	43	72	178
	Ngp	L37297	2661	4782	2311	6912
Tertiary granules						
	Mmp9	Z27231	**0**	**1**	**2**	**2**

Shown are the possible granule protein cDNAs represented on the oligionucleotide arrays, sorted by their expression patterns as follows: first by the average difference AD value, then by the granule types, and last by the alphabetical order of gene symbols. Data are presented as described in the legend to Table 3.

AD indicates average difference; gene symbols are expanded in an Appendix at the end of this article.

estimates and the mRNA levels observed by Northern blotting with specific probes for these genes. In particular, the latter method, more sensitive and specific, showed that C/EBP α began to decline in the most mature cells, whereas C/EBP δ mRNA declined progressively beginning at 24 hours after the onset of differentiation.

C/EBP ϵ is a more recently cloned C/EBP family member. Previous studies indicated it is expressed in a large array of human leukemia cell lines blocked at various stages of differentiation and that it is up-regulated during granulocytic differentiation.[31] A C/EBP ϵ probe was not included in the oligonucleotide chips, and this mRNA was not detected by DD. Therefore, we examined the C/EBP ϵ expression patterns by quantitative PCR and Northern blot analysis (Figure 4). C/EBP ϵ exon 1 was PCR amplified from MPRO RNAs using primers RY48 (AGCCCCCGACACCCTTGATGA) and RY49 (TGGCACACT-GCGGGCAGACAG).[32] The results showed that C/EBP ϵ is expressed throughout myeloid differentiation, with expression levels increased moderately in the later stages.

We detected a number of other transcription factors that are broadly expressed or that have been reported in other studies of hematopoiesis (Table 5). Some of the factors that were most strongly induced during differentiation have been studied in other contexts but not previously implicated in hematopoiesis, such as a mammalian homologue to the *Drosophila* enhancer of split gene, a transcriptional silencer. The mammalian gene is expressed at relatively high levels as measured by the oligonucleotide chip and is a candidate for mediation of the silencing of growth-related genes in the maturing neutrophil. Another candidate transcriptional silencer, Tif1b, may serve as a corepressor for the KRAB domain family of zinc finger transcription factors and also may mediate binding of the heterochromatin protein HP1 to DNA.[33]

There were 26 transcription factors whose mRNAs showed no significant changes by oligonucleotide chip analysis and were not identified as differentially regulated genes by differential display assays. PU.1, a factor necessary for the production of neutrophils and the expression of several neutrophil genes,[34] showed less than a 3-fold increase in mRNA, below the threshold for a significant change. Other candidate hematopoietic transcription factors, such as PEBP1aB2 (AML1), GATA-1, and SP-2, were represented on the oligonucleotide chips, but their mRNA levels were so low that they were reported as absent in this study. The possibility that small changes in the levels or ratios of some transcription factors could produce marked changes in transcription potentially limits the ability of data generated by present methods to explain transcriptional changes during differentiation.

Protein expression patterns of MPRO cells during ATRA induction

We visually compared the 2DE patterns from MPRO cells at the same time points used for mRNA analysis. In most cases the

Table 5. Transcription modulators presented during myeloid differentiation

Maximal fold change	Gene symbol	Gene accession	AD value by array			
			0 h	24 h	48 h	72 h
Less than 2-fold						
	Zfp11-6	AB020542	2630	2989	2795	2515
	Btf3	W13502	3	3	2	1
	Gata2	AB000096	562	770	472	730
	Hmgi	J04179	337	348	177	232
	Idb1	M31885	455	787	721	637
	Max	M63903	256	224	312	172
	Nfatc2	AA560093	2313	3218	2396	2542
	Pm1	U33626	173	281	329	306
	Rarg	M34476	102	113	114	218
	Rela	M61909	297	260	304	244
	Sox15	W53527	419	461	484	837
	Ybx1	M62867	643	489	472	496
	Zfp162	Y12838	671	734	720	992
2 or more, less than 3						
	Cebpd	X61800	157	262	168	430
	Idb2	M69293	244	210	310	604
	Jund1	W29356	1274	2002	1434	3085
	Lyl1	X57687	399	342	347	891
	Nfe2	L09600	458	743	1042	505
	Nfkb1	L28117	953	2044	1876	2034
	Pbx1	AF020196	611	303	345	212
	sfpi1	A34693	375	784	991	529
	Tif1b	U67303	673	659	420	863
	Trp53	P10361	259	149	125	361
	Usf2	U12283	129	185	285	192
	Ybx3	L35549	96	169	210	119
	Zfp216	AA510137	82	151	204	106
3 or more, less than 4						
	Irf1	M21065	85	207	278	198
	Klf2	U25096	62	86	246	77
	Myb	M12848	892	356	230	435
	Stat3	AA396029	484	1057	1012	290
	Tfdp1	Q08639	307	560	505	1093
4 or more, less than 5						
	Cebpb	X62600	390	1248	1380	1903
	Stra14	Y07836	223	383	510	936
5 or more						
	Cebpa	M62362	33	212	182	44
	Grg	X73359	99	565	916	1005
	Mad	X83106	0	111	167	327
	Myc	L00039	314	112	62	173
	Etohi6	W89667	169	386	313	1003
	TBX1	AA542220	0	0	1	2

Shown are the transcription factors identified as present by the oligonucleotide array analysis whose maximal AD between perfect match and mismatch oligonucleotide sets was greater than or equal to 200 U in this study. Data are presented as described in the legend to Table 3.

AD indicates average difference; gene symbols are expanded in an Appendix at the end of this article.

peptides identified for a given protein were derived from regions along the entire length of the protein, indicating the observed products were not the result of proteolytic degradation. These data must be considered with several caveats: membrane and other hydrophobic proteins and very basic proteins are not well displayed by the standard 2DE approach, and proteins present at low levels will be missed.[35] In addition, to simplify MS analysis, we used a Coomassie dye stain rather than silver to visualize proteins, and this decreased the sensitivity of detection of minor proteins. The MS method we used was sufficiently sensitive to identify proteins that could barely be visualized by colloidal blue staining. However, a limitation of the method for the mouse is that the current database lacks predicted amino acid sequences for a substantial fraction of murine genes. In addition, very small proteins give only a few peptides, making statistically confident identification difficult.

Figure 5 shows the analytical colloidal blue–stained 2DE IPG reference maps of differentiated MPRO cells. Expression patterns of more than 500 protein spots were detected and observed through the entire series of gels. Protein spots could easily be cross-matched to each other, indicating the reproducibility of the method. As marked on the gel pictures (Figure 5), 50 proteins with a wide range of molecular weights (1 to 200 kd), isoelectric points (4 to 9), and abundances were subjected to MS protein identification. The results are presented in Table 6.

Comparing the theoretical value of the molecular weight and *pI* of each protein to that of the observed value, we confidently identified 28 proteins in the expected position on the gels (spots 1 to 28). Some of the other proteins with strong matches to the murine databases migrated to a somewhat unexpected *pI* position. Nine spots gave clear peptide peaks on mass spectroscopy but did not match any known gene. Their identification will require amino acid

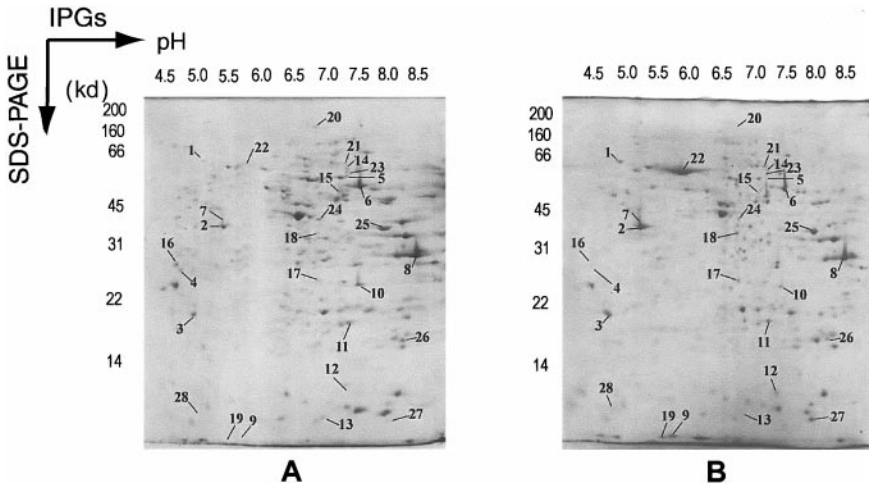

Figure 5. 2DE electrophoretograms of MPRO cells. MPRO cell lysate (2.5×10^6 cell/sample) was loaded for 2DE analysis. Gels were stained with brilliant blue G–colloidal dye. (A) 2DE map of uninduced MPRO cell (0 hour). (B) 2DE map of matured MPRO cells (72 hours). Protein spots marked in the maps were considered differentially expressed and were subjected to MS analysis. The resultant protein information is listed in Table 6.

sequence analysis or availability of more extensive murine databases. We searched for the expression patterns of the genes cognate to the expressed proteins in dbMC (Table 6). Nineteen genes were found in dbMC, the mRNA for 5 genes was reported as absent, and 13 genes were present during MPRO differentiation. Comparison of the expression patterns showed only 4 genes of 18 present on the oligonucleotide chips whose expression was consistent at the RNA level and protein level. None of these was on the list of the genes that were differentially expressed significantly (5-fold or greater change by array or 2-fold or greater change by DD).

Discussion

We explored the temporal patterns of gene expression during myeloid development. A database has been developed to provide a

Table 6. Correlation of expression patterns between mRNA level and protein level

Spot	Protein definition	Gi number	Predicted value		Percentage	2DE pattern		cDNA expression pattern		Ag
			kd	*pI*	(%)	0 h	72 h	0 h	72 h	
1	GRP 78	2506545	72.4	5.1		1	3	1321	1043.3	N
2	Actin, gamma, cytoplasmic	6752954	41.77	5.3	40	3	6	**0**	**2**	Y
3	RHO GDI 2	2494703	22.83	4.9	33	3	3	341	441.6	Y
4	Proliferating cell nuclear antigen	7242171	28.77	4.7	42	1	0	544	430.9	Y
5	APS kinase	4038346	69.8	7.1	24	2	1	43	50.7	N
6	Pyruvate kinase 3	6755074	57.9	7.2	48	6	4	3047	5880.3	N
7	Melanoma X-actin	6671509	41.72	5.3	39	1	3	2539	341.3	N
8	Glyceraldehyde-3-phosphate dehydrogenase	6679937	35.79	8.7	39	8	7	3073	5742.3	N
9	Stefin 3	461911	10.99	5.9	48	0	4	N/A	N/A	—
10	Guanine nucleotide binding protein, beta-2, related sequence1	6680047	35.06	7.9	21	4	2	139	303.1	N
11	Triosephosphate isomerase	6678413	26.69	6.9	26	3	3	3312	2660.1	Y
12	Testis-derived c-abl protein	1196524	17.19	7	51	2	3	152	126.9	N
13	RNA binding motif protein 3	7949121	16.59	6.8	25	1	0	628	812.4	N
14	Collapsin response mediator	6681019	62.16	6.4	36	2	0	Absent	Absent	N
15	Lamin A	220474	47.52	6.6	35	2	0	Absent	Absent	N
16	47-kd keratin	52783	35.82	4.8	29	3	0	Absent	Absent	N
17	sid478p	5931565	31.3	6.7	30	1	2	Absent	Absent	N
18	MHC class II H2-IA-beta-5	3169662	28.6	7.1	39	1	2	N/A	N/A	—
19	Androgen-binding protein: subunit alpha	739346	8.04	6.4	68	0	2	Absent	Absent	N
20	Neuronal apoptosis inhibitory protein	5932010	158.7	6	17	1	0	N/A	N/A	—
21	PAD type IV	6755018	74.46	7.2	21	1	3	N/A	N/A	—
22	Human serum albumin homologoue	3212625	66.45	5.7	24	0	6	N/A	N/A	—
23	syncrip	6576815	62.53	7.2	33	2	1	N/A	N/A	—
24	Transamidinase	1730203	48.22	7.2	31	3	1	N/A	N/A	—
25	PGK crigr phosphoglycerate	1730519	44.54	8.3	47	5	4	1088	1402.3	N
26	Proliferation-associated gene A	6754976	22.16	8.6	53	3	1	N/A	N/A	—
27	Putatuve peroxisomal antioxidant enzyme	3913065	17	7.8	55	0	3	N/A	N/A	—
28	IgE chain C2 region	2137430	12.1	5.2	38	0	1	N/A	N/A	—

The proteins listed here are represented by the spots marked in the electrophoretograms shown in Figure 5.

Protein definition, Gi number, and predicted value refer to the protein name, accession number, and properties derived from the National Center for Biotechnology Information protein database. The column labeled % shows the percentage of peptides predicted from the protein sequence that were detected by mass spectroscopy. The expression level of protein spots expressed in mouse promyelocytic cell line cell induced by all-*trans* retinoic acid for 0 hours and 72 hours (Figure 5) were scored on a scale of 1 (+) to 8 (++++++++) in the 2DE pattern column. The cDNA expression patterns of the cognate mRNAs are listed in the cDNA expression pattern column abstracted from the dbMC database. The genes not represented on the oligonucleotide arrays were marked as N/A. Ag showed the correlation of gene patterns at mRNA level or protein level.

Y indicates agreement and N discrepancy between changes in cDNA and protein spot intensity. The numbers in bold were obtained with DD. 2DE indicates 2-dimensional gel electrophoresis; IgE, immunoglobulin E; DD, differential display.

reference for later research on the molecular mechanisms underlying normal myeloid development.

The MPRO cell system morphologically mimics normal myeloid differentiation and biochemically proceeds further toward mature neutrophils than most other in vitro systems. Because the arrest in differentiation of MPRO cells growing in the absence of ATRA is not physiologic, there is a theoretical risk that gene expression in these cells is not coordinated in the way that it is in normal differentiation. It is encouraging that, for the most part, the timing of expression of genes for proteins of the various neutrophil granules is consistent with the timing of the morphologic and biochemical appearance of these granule components during normal myeloid differentiation.

The DD technique provides certain advantages for detecting and comparing mRNA levels in different samples. First, the method is, in principle, similar to competitive RT-PCR, and, with the use of stringent PCR conditions, is expected to be about as reliable. Second, display patterns are reproducible. Third, the method detects the levels not only of RNAs already represented in the database but also of unknown RNA species that may represent "new" genes. Fourth, closely related genes can be distinguished regardless of cross-hybridization, provided there are some single nucleotide differences in the 3′ end sequence. Limitations associated with this technique are that numerous gels are necessary to get complete information and that comparison of the levels of different mRNAs is only approximate because of the differential amplification of bands of different size or sequence.

Oligonucleotide chip analysis is a fast and effective means of accessing mRNA expression patterns.[20] Cluster analysis of groups of samples by this approach is effective. However, the present results indicate that alternative methods of verification are desirable before the data on an unexpected change in a particular gene are definitively accepted.

To obtain the broadest range of information from the myeloid differentiation process, both differential display and oligonucleotide chip techniques were applied in the current study. As a result, 65.3% of the observed changes in mRNA levels came from the differential display method and 41.5% came from oligonucleotide chip assays.

Our data showed in general that changes in expression pattern by the 2 methods agreed qualitatively but that there was some quantitative variation. Our results indicate that DD may be a more accurate way to detect changes in levels of gene expression than the oligonucleotide chip assay. However, improvements in the types of oligonucleotides used in arrays may close this gap in the future.

The mRNAs for a limited number of transcription factors vary in a pattern correlating with that of the mRNAs for primary or secondary granule proteins. However, more detailed information is needed, and the underlying mechanisms of granule gene regulation remain unclear. The number of potential positive and negative regulatory factors found here is sufficiently small as to make it feasible to perform in vivo studies, such as chromatin immunoprecipitation.

The oligonucleotide chip used in this study focused on known genes, whereas the DD method samples all polyadenylated transcripts. The latter method generated a large number of products not associated with known genes, in part because the mouse genome is not as well represented in the database as the human genome. However, our experience with DD and human mRNAs indicates that substantial fractions of the products represented as ESTs or not represented at all in the public databases are cDNA copies from introns, hnRNA, or other RNA with internal A runs.

Approximately 59 sequences obtained from gel-display bands had significant changes in the level of expression and a sequence that did not match that for any named gene in the public databases. Of these, 38 had plausible or excellent polyA signals. This is only an approximate estimate of the number of new genes found[36] because a fraction of the mRNAs for known genes still had poor polyA signals. In addition, the full 3′ untranslated region is often not known for characterized genes, and in some cases these new genes may prove to be identical to products identified by the oligonucleotide chips when more complete sequences are obtained. At the least, their presence indicates that a substantial fraction of the regulatory or functional circuitry of maturing myeloid cells remains unexplored and that valuable tools for their investigation will emerge from a combination of RNA expression studies and analysis of emerging genomic sequences.

The desired end point for the description of gene expression in a biologic system is not only the analysis of mRNA transcript levels but also the accurate measurement of protein abundance. The developments in 2DE and new MS instrumentation make it possible to accomplish this work rapidly and efficiently. In this study, we attempted to identify a number of the proteins differentially expressed between uninduced and ATRA-differentiated MPRO cells and to examine the relation between mRNA and protein expression levels for these genes representing the same state.

For protein levels based on estimated intensity of Coomassie dye staining in 2DE, there was poor correlation between changes in mRNA levels and estimated protein levels. Other groups have studied the correlation between mRNA and protein levels in yeast and liver cells.[11,12,14] In the liver cell experiments,[11,12] correlation coefficients of 0.4 to less than 0.5 were observed. In an extensive study in yeast,[11,12] the correlation coefficient was high if the most abundant mRNAs and proteins were considered. If a handful of these products was omitted, the remaining correlation coefficient was 0.4 or less. However, one could restore some of the correlation by averaging individual data points into broad proteomic categories.[37]

The discrepancies between mRNA and protein levels in MPRO cells appear to be substantially larger than those observed for yeast. Possible causes for the discrepancies include translational regulation, differential expression of certain mRNAs at various stages of cell growth in vitro, post-translational protein modification that varies with the stage of maturation of the cells, and selective degradation or excretion of proteins in vivo. Furthermore, here we are focusing on a developmental time-course, whereas the yeast study concentrated on the organism in vegetative growth. New techniques, equipment, and bioinformatic analysis tools must be developed to make such systematic, global, and quantitative analyses feasible.

The initial studies of protein expression presented here provide a cautionary note for efforts to interpret cell composition and function in relation to mRNA levels. Discrepancies we observed between gene expression and protein abundance suggest that selective post-transcriptional controls may be at least as important as changes in mRNA levels in determining the protein composition of neutrophils and that they are phenomena less well explored than transcriptional control. Analysis of mRNA expression patterns is itself only a small beginning toward a genome-wide description of cellular components.

Acknowledgments

We thank Dr S. Tsai (Fred Hutchinson Cancer Research Center) for his kind gift of the MPRO cell line, Dr Fuki M. Hisama (Yale University School of Medicine) for helpful advice, and the staff at Gene Logic Inc for data and support.

References

1. Lawson ND, Berliner N. Neutrophil maturation and the role of retinoic acid. Exp Hematol. 1999; 27:1355-1367.
2. Tenen DG, Hromas R, Licht JD, Zhang DE. Transcription factors, normal myeloid development, and leukemia. Blood. 1997;90:489-519.
3. Sigurdsson F, Khanna-Gupta A, Lawson N, Berliner N. Control of late neutrophil-specific gene expression: insights into regulation of myeloid differentiation. Semin Hematol. 1997;34:303-310.
4. Lenny N, Westendorf JJ, Hiebert SW. Transcriptional regulation during myelopoiesis. Mol Biol Rep. 1997;24:157-168.
5. Yunis JJ, Tanzer J. Molecular mechanisms of hematologic malignancies. Crit Rev Oncog. 1993;4: 161-190.
6. Velculescu VE, Zhang L, Vogelstein B, Kinzler KW. Serial analysis of gene expression. Science. 1995;270:484-487.
7. Stollberg J, Urschitz J, Urban Z, Boyd CD. A quantitative evaluation of SAGE. Genome Res. 2000;10:1241-1248.
8. Subrahmanyam YV, Baskaran N, Newburger PE, Weissman SM. 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.
9. Subrahmanyam YVBK, Yamaga S, Newburger PE, Weissman SM. A modified approach for the efficient display of 3′ end restriction fragments of cDNAs. In: Leslie RA, Robertson HA, eds. Differential Display: A Practical Approach. Practical Approach Series. Oxford, UK: Oxford University Press. 2000;101-129.
10. Appella E, Arnott D, Sakaguchi K, Wirth PJ. Proteome mapping by two-dimensional polyacrylamide gel electrophoresis in combination with mass spectrometric protein sequence analysis. EXS. 2000;88:1-27.
11. Anderson NL, Anderson NG. Proteome and proteomics: new technologies, new concepts, and new words. Electrophoresis. 1998;19:1853-1861.
12. Gygi SP, Rochon Y, Franza BR, Aebersold R. Correlation between protein and mRNA abundance in yeast. Mol Cell Biol. 1999;19:1720-1730.
13. The Whitehead Institute for Biomedical Research/MIT Center for Genome Research. Molecular Pattern Recognition Web site. Available at: www.genome.wi.mit.edu/MPR/analysis.html#RS. Accessed May 4, 2001.
14. Futcher B, Latter GI, Monardo P, McLaughlin CS, Garrels JI. A sampling of the yeast proteome. Mol Cell Biol. 1999;19:7357-7368.
15. Tsai S, Collins SJ. A dominant negative retinoic acid receptor blocks neutrophil differentiation at the promyelocyte stage. Proc Natl Acad Sci U S A. 1993;90:7153-7157.
16. Johnson M, Caiazzo T, Molina JM, Donahue R, Groopman J. Inhibition of bone marrow myelopoiesis and erythropoiesis in vitro by anti-retroviral nucleoside derivatives. Br J Haematol. 1988;70:137-141.
17. Lawson ND, Krause DS, Berliner N. Normal neutrophil differentiation and secondary granule gene expression in the EML and MPRO cell lines. Exp Hematol. 1998;26:1178-1185.
18. Prashar Y, Weissman SM. Analysis of differential gene expression by display of 3′ end restriction fragments of cDNAs. Proc Natl Acad Sci U S A. 1996;93:659-663.
19. Subrahmanyam YVBK, Yamaga S, Prashar Y, et al. RNA expression patterns change dramatically in human neutrophils exposed to bacteria. Blood. 2001;97:2456-2468.
20. Chee M, Yang R, Hubbell E, et al. Accessing genetic information with high-density DNA arrays. Science. 1996;274:610-614.
21. Lipshutz RJ, Chee M, Hubbell E, et al. Using oligonucleotide probe arrays to access genetic diversity. Biotechniques. 1995;19:442-447.
22. Lipshutz RJ, Fodor SP, Gingeras TR, Lockhart DJ. High-density synthetic oligonucleotide arrays. Nat Genet. 1999;21:20-24.
23. Mao M, Fu G, Wu JS, et al. 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 U S A. 1998;95:8175-8180.
24. Neuhoff V, Arold N, Taube D, Ehrhardt W. Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250. Electrophoresis. 1988;9:255-262.
25. Wang Q, Miyakawa Y, Fox N, Kaushansky K. Interferon-alpha directly represses megakaryopoiesis by inhibiting thrombopoietin-induced signaling through induction of SOCS-1. Blood. 2000;96:2093-2099.
26. Gullberg U, Bengtsson N, Bulow E, et al. Processing and targeting of granule proteins in human neutrophils. J Immunol Methods. 1999;232:201-210.
27. Nichols J, Nimer SD. Transcription factors, translocations, and leukemia. Blood. 1992;80:2953-2963.
28. Amati B, Land H. Myc-Max-Mad: a transcription factor network controlling cell cycle progression, differentiation and death. Curr Opin Genet Dev. 1994;4:102-108.
29. Pagliuca A, Gallo P, De Luca P, Lania L. Class A helix-loop-helix proteins are positive regulators of several cyclin-dependent kinase inhibitors' promoter activity and negatively affect cell growth. Cancer Res. 2000;60:1376-1382.
30. Yamanaka R, Lekstrom-Himes J, Barlow C, Wynshaw-Boris A, Xanthopoulos KG. CCAAT/enhancer binding proteins are critical components of the transcriptional regulation of hematopoiesis (review). Int J Mol Med. 1998;1:213-221.
31. Morosetti R, Park DJ, Chumakov AM, et al. A novel, myeloid transcription actor, C/EBP epsilon, is up-regulated during granulocytic, but not monocytic, differentiation. Blood. 1997;90:2591-2600.
32. Yamanaka R, Barlow C, Lekstrom-Himes J, et al. Impaired granulopoiesis, myelodysplasia, and early lethality in CCAAT/enhancer binding protein epsilon-deficient mice. Proc Natl Acad Sci U S A. 1997;94:13187-13192.
33. Nielsen AL, Ortiz JA, You J, et al. Interaction with members of the heterochromatin protein 1 (HP1) family and histone deacetylation are differentially involved in transcriptional silencing by members of the TIF1 family. EMBO J. 1999;18:6385-6395.
34. Anderson KL, Smith KA, Perkin H, et al. PU.1 and the granulocyte- and macrophage colony-stimulating factor receptors play distinct roles in late-stage myeloid cell differentiation. Blood. 1999;94:2310-2318.
35. Gorg A, Obermaier C, Boguth G, Weiss W. Recent developments in two-dimensional gel electrophoresis with immobilized pH gradients: wide pH gradients up to pH 12, longer separation distances and simplified procedures. Electrophoresis. 1999;20:712-717.
36. Wilusz J, Pettine SM, Shenk T. Functional analysis of point mutations in the AAUAAA motif of the SV40 late polyadenylation signal. Nucleic Acids Res. 1989;17:3899-3908.
37. Jansen R, Gerstein M. Analysis of the yeast transcriptome with structural and functional categories: characterizing highly expressed proteins. Nucleic Acids Res. 2000;28:1481-1488.
38. Tamayo P, Slonim D, Mesirov J, et al. Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. Proc Natl Acad Sci U S A. 1999; 96:2907-2912.
39. ProteoMetrics. ProFound search engine Web site. Available at: http://www.proteometrics.com/profound_bin/WebProFound.exe. Accessed May 4, 2001.

Appendix

Gene symbols used in tables: Actb: actin, beta, cytoplasmic; Actg: actin, gamma, cytoplasmic; Actx: melanoma X-actin; Aldo1: aldolase 1, A isoform; Arf5: ADP-ribosylation factor 5; Atf1: activating transcription factor 1; Atf2: activating transcription factor 2; Btf3: basic transcription factor 3a; Bzrp: peripheral-type benzodiazepine receptor; C5r1: complement component 5, receptor 1/G protein-coupled receptor (C5a); Ccnb2: cyclin B2; Cd36l2: CD36 antigen (collagen type I receptor, thrombospondin receptor)-like 2; Cd53: CD53 antigen; Cebpa: CCAAT/enhancer binding protein C/EBP, alpha; Cebpb: CCAAT/enhancer binding protein (C/EBP), beta; Cebpd: CCAAT/enhancer binding protein (C/EBP), delta; Cebpe: CCAAT/enhancer binding protein (C/EBP), epsilon; Cfl1: cofilin 1, nonmuscle; Cmkar4: chemokine (C-X-C) receptor 4; Cmkbr1: chemokine (C-C) receptor 1/Mip1a receptor; Cnlp: cathelin-like protein; Cntf: ciliary neurotropic factor/zinc finger protein PZF; Copa: coatomer protein complex subunit alpha; Cpa3: carboxypeptidase A3, mast cell; Cr2: complement receptor 2; Crhr: corticotropin releasing hormone receptor; Crry: complement receptor–related protein; Csf1r: CSF 1 (M-CSF) receptor/c-fms/CD115; Csf2ra: CSF 2 (GM-CSF) receptor, alpha, low-affinity/CD116; Csf2rb1: CSF 2 (GM-CSF) receptor, beta 2, low-affinity/IL 3 receptor-like protein (AIC2B)/CDw131; Csf2rb2: CSF 2 (GM-CSF) receptor, beta 2, low-affinity/IL-3 receptor (AIC2A); Ctsb: cathepsin B; Ctsc: cathepsin C; Ctsd: cathepsin D; Ctse: cathepsin E; Ctsg: cathepsin G; Ctsh: cathepsin H; Ctsl: cathepsin L; Ctss: cathepsin S; Cybb: cytochrome b-245, beta; Drd2: dopamine receptor 2; E2f1: E2F transcription factor 1; Ear2: eosinophil-associated ribonuclease 2; Ebi3: Epstein-Barr virus–induced gene 3/cytokine receptor–like molecule (EBI3); El2: Balb/c neutrophil elastase; Ela2: elastase 2; Erh: enhancer of rudimentary homolog (Drosophila); Etohi6: ethanol induced 6/sterol regulatory element binding transcription factor 1 (SREBF1) homolog; F2rl2: coagulation factor II (thrombin) receptor–like 2; Fcer1g: Fc receptor, IgE, high affinity I, gamma polypeptide; Fcgr2b: Fc receptor, IgG, low affinity IIb; Fcgr3: Fc receptor, IgG, low affinity III; Fpr1: formyl peptide receptor 1/fMLP receptor; Gabpb1: GA repeat binding protein (GABP-beta1 subunit); Gata2: GATA-binding protein 2; Gnas: guanine nucleotide binding protein, alpha stimulating; Gnb2-rs1: guanine nucleotide binding protein, beta-2, related sequence 1; Gpx3: glutathione peroxidase 3; Grg: related to Drosophila groucho gene; Grid1: glutamate receptor channel subunit delta 1; Grn: granulin; Gstm1: glutathione-S-transferase, mu 1; Gus-s: beta-glucuronidase structural; Gys3: glycogen synthase 3, brain; H2-D: histocompatibility 2, D

region locus 1; Hist2: histone gene complex 2; Hist5–2ax: H2A histone family, member X; Hmgi: high mobility group protein I; Hsp60: heat shock protein, 60 kDa; Htr5a: 5-hydroxytryptamine (serotonin) receptor 5A; Idb1: inhibitor of DNA binding 1/helix-loop-helix DNA binding protein regulator (Id); Idb2: inhibitor of DNA binding 2; Ifngr: interferon gamma receptor; Ifngr2: interferon gamma receptor 2; Ii: Ia-associated invariant chain; Il1a: IL1 alpha; Il1r2: IL1 receptor, type II; Il2rg: IL2 receptor, gamma chain; Il4ra: IL4 receptor, alpha; Il10rb: IL10 receptor, beta; Il17r: IL17 receptor; Irf1: interferon regulatory factor 1; Irf2: interferon regulatory factor-2; Itgb2: integrin beta 2 (Cd18); Itpr5: inositol 1,4,5-trisphosphate receptor (type 2); Jund1: Jun proto-oncogene–related gene d1/transcription factor JUN-D; Klf2: Kruppel-like factor LKLF; L-CCR: lipopolysaccharide inducible C-C chemokine receptor–related; Lcn2: lipocalin 2; Ldlr: low density lipoprotein receptor; Lsp1: Lymphocyte–specific 1/S37/pp52; Lst1: leucocyte–specific transcript 1; Ltb4r: leukotriene B4 receptor; Ltbr: lymphotoxin-beta receptor; Ltf: lactotransferrin; Ly64: lymphocyte antigen 64; Ly6e: lymphocyte antigen 6 complex, locus E; Lyl1: lymphoblastomic leukemia/bHLH factor; Lyzs: lysozyme; M6pr: mannose-6-phosphate receptor, cation dependent; Mad: Max dimerization protein; Man2c1: mannosidase, alpha, class 2C, member 1; Max: Max protein; Maz: MYC-associated zinc finger protein (purine-binding transcription factor); MBP: eosinophil granule major basic protein precursor; Mcpt8: mast cell protease 8; Mll: myeloid/lymphoid or mixed-lineage leukemia; Mmp13: matrix metalloproteinase 13/collagenase; Mmp9: matrix metalloproteinase 9/gelatinase B; Mpo: myeloperoxidase; Myb: myeloblastosis oncogene; Mybl2: myeloblastosis oncogene-like 2; Myc: myelocytomatosis oncogene; Myln: myosin light chain, alkali, nonmuscle; Nfatc2: nuclear factor of activated T cells, cytoplasmic 2; Nfe2: nuclear factor, erythroid-derived 2, 45 kDa; Nfkb1: NF-kappa-B (p105); Ngp: neutrophilic granule protein; NMDRGB: N-methyl-D-aspartate receptor glutamate-binding chain homolog; Npm1: nucleophosmin 1; Nr4a1: nuclear receptor subfamily 4, group A, member 1; Osi: oxidative stress induced; P2rx1: purinergic receptor P2X, ligand-gated ion channel, 1; P2ry2: purinergic receptor P2Y, G-protein–coupled 2; P40–8: P40–8, functional/laminin receptor; Pbx1: pre B-cell leukemia transcription factor 1; Pfc: properdin factor, complement; Pira1: paired-Ig–like receptor A1; Pira5: paired-Ig–like receptor A5; Pira6: paired-Ig–like receptor A6; Pirb: paired-Ig–like receptor B; Plaur: urokinase plasminogen activator receptor; PMI: putative receptor protein (SP: P17152); Pml: promyelocytic leukemia; Prg: proteoglycan, secretory granule; Prg3: proteoglycan 3/eosinophil major basic protein 2; Prtn3: proteinase 3; Psma2: proteasome (prosome, macropain) subunit, alpha type 2; Ptmb4: prothymosin beta 4; Ptprc: protein tyrosine phosphatase, receptor type, C; Rac2: RAS-related C3 botulinum substrate 2; Rarg: retinoic acid receptor, gamma; Rela: avian reticuloendotheliosis viral (v-rel) oncogene homolog A/NF-kappa-B p65; Rpl19: ribosomal protein L19; RPL8: ribosomal protein L8; Rps6ka1: ribosomal protein S6 kinase polypeptide 1; Rps8: ribosomal protein S8; Rtn3: reticulon 3; S100a8: S100 calcium binding protein A8 (calgranulin A); S100a9: S100 calcium-binding protein A9 (calgranulin B); Sdfr2: stromal cell–derived factor receptor 2; Sell: selectin L (lymphocyte adhesion molecule 1); Sema4d: semaphorin 4D; Sepp1: selenoprotein P, plasma, 1; Sfpi1: SFFV proviral integration 1; Shfdg1: split hand/foot deleted gene 1; Slc10a1: solute carrier family 10 (sodium/bile acid cotransporter family), member 1; Slpi: secretory leukocyte protease inhibitor; Sox15: SRY-box containing gene 15; Spi2–1: serine protease inhibitor 2-1; Srb1: scavenger receptor class B1; Stat3: signal transducer and activator of transcription 3; Stat5a: signal transducer and activator of transcription 5A; Stat6: signal transducer and activator of transcription 6; Stra14: basic-helix-loop-helix protein-retinoic acid induced; Tbx1: TBX1 protein/LPS-induced TNF-alpha factor homolog; Tcrgb: T-cell–receptor germline beta-chain gene constant region; Tcrg-V4: T-cell–receptor gamma, variable 4; Tctex1: t-complex testis expressed 1; Tfdp1: transcription factor Dp 1; Tif1b: transcriptional intermediary factor 1, beta; Tlr4: toll-like receptor 4; Tnfrsf1a: TNF receptor superfamily, member 1a; Tnfrsf1b: TNF superfamily, member 1b; Tomm70a: translocase of outer mitochondrial membrane 70 (yeast) homolog A; Tpi: triosephosphate isomerase; Trp53: transformation-related protein 53; Ubb: ubiquitin B; Usf2: upstream transcription factor 2; Ybx1: Y box transcription factor; Ybx3: Y box binding protein; Zfp11–6: zinc finger protein s11–6; Zfp18: zinc finger protein 18 homolog; Zfp36: zinc finger protein 36; Zfp162: zinc finger protein 162; Zfp216: zinc finger protein 216; Zfpm1: zinc finger protein, multitype 1; Znfn1a1: zinc finger protein, subfamily 1A, 1 (Ikaros); Zyx: zyxin.