Analysis of gene expression in a complex differentiation hierarchy by global amplification of cDNA from single cells

Gerard Brady*†, Filio Billia*, Jennifer Knox*, Trang Hoang‡, Ilan R. Kirsch§, Evelyn B. Voura¶, Robert G. Hawley¶, Rob Cumming#, Manuel Buchwald#, Kathy Siminovitch¥ Neil Miyamoto*, Guido Boehmelt* and Norman N. Iscove*

*The Ontario Cancer Institute and Department of Medical Biophysics, University of Toronto, 500 Sherbourne Street, Toronto M4X 1K9, Canada. ‡Clinical Research Institute of Montreal, Department of Pharmacology, University of Montreal, Montreal H2W 1 R7, Canada. §National Cancer Institute, Navy Medical Oncology Branch, Bethesda, Maryland 20892, USA. ¶Sunnybrook Health Science Centre, Department of Medical Biophysics, University of Toronto, 2075 Bayview Avenue, Toronto M4N 3M5, Canada. #Department of Molecular and Medical Genetics, University of Toronto, Research Institute, Hospital for Sick Children, 555 University Avenue, Toronto M5G 1 X8, Canada. ¥Samuel Lunenfeld Research Institute, Department of Medicine, University of Toronto, 600 University Avenue, Toronto M5G 1X5, Canada.

Background: Many differentiating tissues contain progenitor cells that differ in their commitment states but cannot be readily distinguished or segregated. Molecular analysis is therefore restricted to mixed populations or cell lines which may also be heterogeneous, and the critical differences in gene expression that might determine divergent development are obscured. In this study, we combined global amplification of mRNA transcripts in single cells with identification of the developmental potential of processed cells on the basis of the fates of their sibling cells from clonal starts.

Results: We analyzed clones of from four to eight hemopoietic precursor cells which had a variety of differentiative potentials; sibling cells generally each formed clones of identical composition in secondary culture. Globally amplified cDNA was prepared from individual precursors whose developmental potential was identified by tracking sibling fates. Further cDNA samples were prepared from terminally maturing, homogeneous hemopoietic cell populations. Together, the samples represented 16 positions in the hemopoietic developmental hierarchy. Expression patterns in the sample set were determined for 29 genes known to be involved in hemopoietic cell growth, differentiation or function. The cDNAs from a bipotent erythroid/megakaryocyte precursor and a bipotent neutrophil/macrophage precursor were subtractively hybridized, yielding numerous differentially expressed cDNA clones. Hybridization of such clones to the entire precursor sample set identified transcripts with consistent patterns of differential expression in the precursor hierarchy.

Conclusions: Tracking of sibling fates reliably identifies the differentiative potential of a single cell taken for PCR analysis, and demonstrates the existence of a variety of distinct and stable states of differentiative commitment. Global amplification of cDNA from single precursor cells, identified by sibling fates, yields a true representation of lineage- and stage-specific gene expression, as confirmed by hybridization to a broad panel of probes. The results provide the first expression mapping of these genes that distinguishes between progenitors in different commitment states, generate new insights and predictions relevant to mechanism, and introduce a powerful set of tools for unraveling the genetic basis of lineage divergence.

Current Biology 1995, **5**:909-922

Background

The progression of cells along developmental pathways is defined and driven by changes in gene expression. Identification of these changes would provide the keys to understanding the mechanisms of differentiative progression. But the quest for such information is frequently complicated by the heterogeneous composition of differentiating tissues. Hemopoietic tissues are typical of developmental and cell-renewal systems in which differentiative decisions take place within rare precursor cells that are structurally indistinguishable and greatly outnumbered amongst their maturing progeny. Nevertheless, it is possible to detect hemopoietic precursor cells in functional assays, and to distinguish precursors that differ in their differentiative and growth potentials on the basis of the composition of the clones that they can generate. Such assays define a hierarchy of hemopoietic precursors extending from multipotential stern cells through pluripotential and oligopotential intermediates, to cells that are committed uniquely to forming one of a dozen alternative blood and tissue cell types.

Methods of cell separation can purify precursor cells from more differentiated progeny, but have not been particularly effective in separating early precursor cells from others that differ in their commitment states [1]. Molecular analysis of this system has therefore been limited to the study of precursor populations of mixed composition, or of cell lines which often comprise a mixture of cells in varying stages of the maturation process.

We recently introduced a method for the general amplification of all polyadenylated (polyA$^+$) RNA transcripts in samples as small as a single cell (Fig. 1) [2]. The polymerase chain reaction (PCR) applied to polyA$^+$ RNA

†Present address: School of Biological Sciences, University of Manchester, Manchester M13 9PT.
Correspondence to: Norman N. Iscove. E-mail address: iscove@oci.utoronto.ca

yields amplified cDNA that is several hundred bases in length and is representative of the extreme 3' untranslated ends of the original transcripts. As a result of the length limitation, the abundance relationships in the original sample are preserved in the final amplified product. Analysis of gene expression at the level of single cells would guarantee absolute homogeneity of each source of cDNA, to a level that could never be attainable by fractionation of cell populations. However, the advance comes at a cost: each cell is consumed by the procedure, and its developmental potential can no longer be determined by tracking its eventual fate.

For the hemopoietic system, an answer to the dilemma arises from the natural synchrony of development that is characteristic of growing clones in culture. When precursor cells are cultured with the required cytokines, they generate daughter cells. The commitment status of the daughter cells can be assessed by culturing them in isolation and observing their ultimate outcomes. Most growth-competent precursors in hemopoietic tissues are committed to development within a single lineage. Not surprisingly, daughter cells generated by the first two or three divisions are typically identical to each other in differentiative, as well as growth, potential [3]. They also retain close cell-cycle synchrony through several divisions, generally dividing within minutes of each other. If members of a clone can be established to be identical to each other in potential, then it should be possible to use some cells from a clone as reporters of the probable potential of siblings that are taken for RNA analysis.

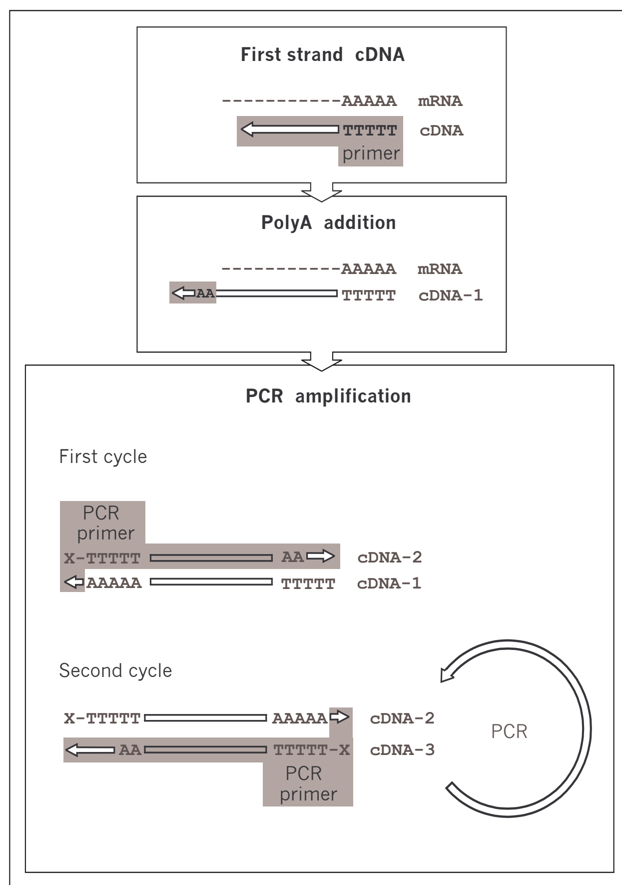

Fig. 1. General amplification of polyadenylated RNA transcripts.

Less predictable results arise when clone formation is initiated by the smaller number of more primitive precursors in hemopoietic tissues that have the capacity to develop into multiple lineages. In the course of initial cell divisions, such cells can generate daughter cells that are individually committed to differing lineages [3]. If sibling cells in some clones prove to differ from one another in their differentiative potential, they cannot serve as reliable indicators of the potential of processed siblings. But clones whose tested siblings retain pluripotentiality and which display the same set of differentiative potentials when subcultured would be useful as indicators of the likely potential of siblings analyzed for RNA expression.

In this study, we exploited the characteristic synchrony of clones developing in culture and derived a set of cDNA samples from individual hemopoietic precursor cells mapped by sibling analysis to a variety of positions in the hemopoietic hierarchy. Using probes for genes with well - understood specificity of expression in the system, we show the expected hybridization patterns that validate our assignments. We also present results from a panel of probes whose RNA-level expression patterns were less predictable, providing a unique look at stage-by-stage expression in the system. Finally, we show the results of pilot efforts at cDNA- and PCR-based subtraction that identify numerous sequences expressed in one individual cell but not another. Collectively, these experiments introduce a new and powerful investigative approach that will allow the mapping of differentiation in heterogeneous systems at unprecedented resolution.

Results

Similarity of sibling cells in colony 'starts'

This study was based on a strategy involving the induction of nascent colony 'starts' from marrow cells in semisolid cultures, selection of single cells from a start for cDNA amplification, and separate secondary culture of each of the remaining siblings in conditions that support full development of cells in all myeloid lineages (Fig. 2). For sibling outcomes to be useful as indicators of the potential of the processed cells, siblings would have to have identical outcomes in a high frequency of colony starts. One set of four- to eight-cell colony starts was generated from unpurified marrow cells cultured with either interleukin-3 (IL-3) or granulocyte colony-stimulating factor (G-CSF). Sibling cells grew in the secondary cultures into erythroid, monocytic, granulocytic or megakaryocytic colonies (Table 1). Most starts generated secondary colonies that represented only a single lineage. Two lineages were represented in the progeny of only 11 % of starts.

A second, larger set of colony starts was generated from cells that were enriched for primitive precursors. Most were derived from marrow cells that had been cultured first with IL-1 and IL-3 under thymidine selection to eliminate actively cycling cells; this treatment enriches marrow for more-primitive precursors that are known to cycle less frequently [4]. A small subset came from day 12 fetal liver cells that were sorted for expression of the AA4.1 surface antigen. Among starts from these populations, 45 % yielded progeny that represented two lineages and 17 % yielded subclones of three or more lineages (Table 1). Overall, siblings generated colonies of identical composition in 81 % of colony starts. Even among colony starts from cells with two or more potentials, siblings yielded identical outcomes in 63 % of cases.

These proportions did not differ significantly when four- to five-cell and six- to eight-cell starts were analyzed separately (data not shown). Among the instances of non-uniform outcome from starts of six to eight cells, half had only a single discrepant daughter outcome. From the frequencies of disparate results in the secondary cultures, the numbers of cells per start, and the proportions of differing secondary clones in each disparate start, it is possible to calculate the probability of mistaken identification of a cell sampled for cDNA amplification, given that all of its siblings display identical outcomes in secondary cultures. Even in the worst case of multilineage starts, that probability did not exceed 0.13 in our series - meaning that less than one in seven multipotential assignments would be incorrect.

In these experiments, sibling cells from nascent clones were plated individually into culture conditions that were designed to support growth and terminal maturation in all myeloid lineages. Although cells from individual starts typically generated clones that were identical to one another in composition, the composition of the secondary clones differed from one start to another. As culture

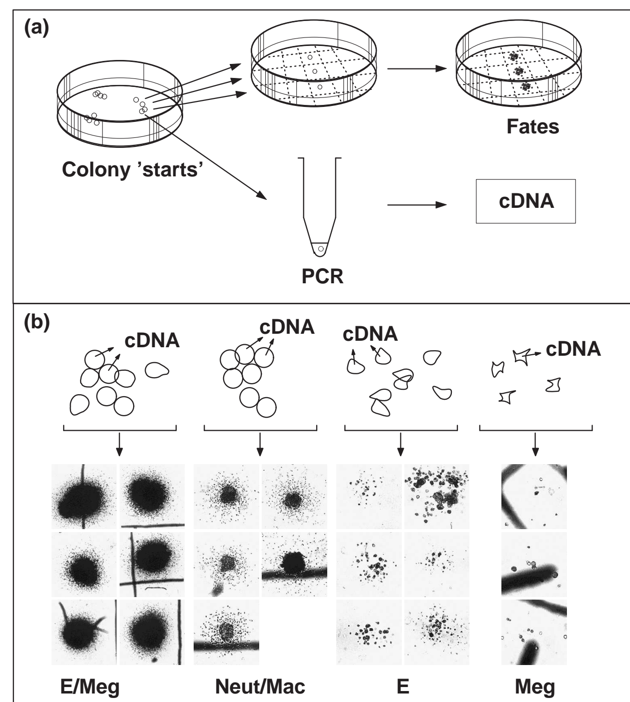

Fig. 2. Determination of the differentiative potential of hemopoietic precursor cells by analysis of the fates of their sibling cells. **(a)** Marrow cells are plated in primary cultures and precursor cells allowed to divide two or three times. The culture plate shown on the left contains three colony starts each consisting of four cells. From one of the starts, one cell is withdrawn and lysed for cDNA amplification. The remaining sibling cells are plated individually in a gridded secondary culture plate, where each develops with time into a secondary colony, Analysis of the lineage content of the secondary colonies establishes the differentiative potential of each of the cultured sibling cells. If each forms the same kind of colony, it is likely that the cell processed for cDNA would have had a similar potential. **(b)** Identical outcomes of sibling cells from colony starts. The appearance of the cells in each colony start is shown at the top. The micrographs show the appearance of the resulting secondary colonies, with the lineage content indicated at the bottom (abbreviations as in Fig. 3).

Table 1. Differentiative outcomes of sibling cells in four- to eight-cell colony starts.

Cell source		Proportion of starts				
		Number of lineages		Uniform outcome		Single
	Number of starts analyzed	Two	≥ Three	(Overall)[1]	(Pluripotent)[2]	exception[3]
*Marrow	19	0.11	0	0.89	(too few)	
†TdR/FLAA4.1+	75	0.45	0.17	0.79	0.65	
Overall	94			0.81	0.63	0.47

Cells from the indicated sources were cultured for 24-48 h in methyl cellulose. Colony starts of four to eight non-adhesive and agranular cells were selected, as these morphological features were predictive of significant further growth potential. Individual cells were replated in isolation into gridded secondary methyl cellulose plates, generally three cells from four-cell starts, and up to six cells from eight-cell starts. Secondary colonies were sampled at 6-14 days, spread on glass slides, fixed and stained for morphological and histochemical identification of erythroid, megakaryocytic, monocytic and granulocytic cells. Mast cell potential was assessed by culture of cells from secondary colonies for at least four weeks in medium containing IL-3, followed by specific staining of mast cell granules with Alcian blue. [1]Of all starts, those yielding secondary clones that were identical to each other in lineage content. [2]Of starts yielding two or more lineages, those generating identical secondary clones. [3]Of starts yielding divergent secondary clones, those with only one discrepant daughter clone. *Unseparated marrow cells were cultured for 48 h in methyl cellulose containing IL-3 or G-CSF. †Eight starts were derived from day 12 fetal liver cells that had been pan-purified for expression of the AA4.1 antigen; the remainder from marrow cells selected in liquid cultures with 100 μg ml^{-1} thymidine for 20 h in the presence of IL-1 and IL-3. The resulting cells were plated in methyl cellulose containing IL-1, IL-3 and c-kit ligand, and starts were subdivided after 27 h.

conditions in the secondary cultures were invariant, the uniform outcomes within clones provide compelling evidence for the existence of stable differentiative commitment, whereas the varying outcomes from start to start must reflect the existence of a variety of distinct commitment states.

Collection of cDNA samples from cells identified by sibling analysis

Amplified cDNA was prepared from precursor cells that were individually drawn from colony starts, and their likely potential was read from the outcomes of their sibling cells in secondary cultures. If a resulting cDNA sample hybridized successfully to probes for the 'housekeeping' genes *L32* (encoding a ribosomal protein) and *gapdh* (encoding a metabolic enzyme), and if its cultured sibling cells had uniform outcomes, the sample was included in the study without any additional selection criteria. Terminally maturing cells were sampled from homogeneous populations growing in bulk cultures, or as colonies in semi-solid medium, and cDNA was usually prepared from individual samples of 50-200 cells. Among the various samples of mature cells, only those from megakaryocytes were prepared from single cells. As shown in Figure 3, multiple samples were obtained from each of 16 distinct positions in the illustrated hierarchy.

By the nature of our approach, all sampled precursor cells were active in the cell cycle, and information relevant to cycle phase was often available. Colony starts were visualized after 26 hours of culture, and after recording their locations in the cultures they were returned to the incubator. After an additional 1-2 hours, the plates were removed from the incubator and cells from the mapped starts were micromanipulated. The cultured siblings were then monitored at intervals until they divided. Frequently, starts of four cells when mapped had become eight cells by the time they were sampled. In other examples of four-cell starts, the three monitored siblings divided within 1-2 hours of sampling. Such observations placed some of our sampled precursor cells in late G2 and others in early G1 phase of the cell cycle (Table 2).

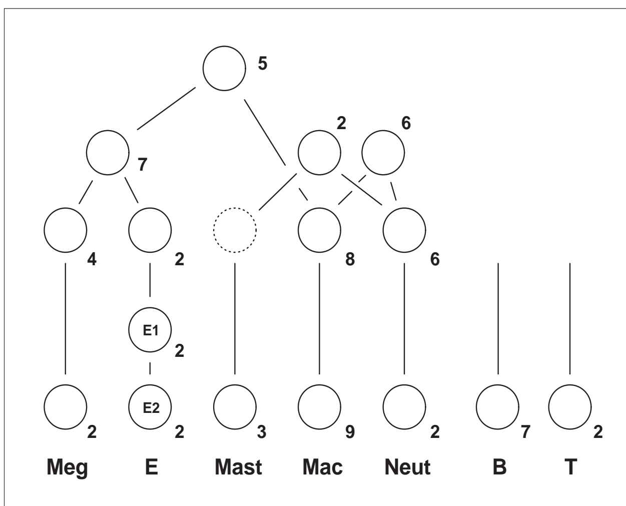

Fig. 3. Hierarchical relationships of the hemopoietic cells sampled for cDNA. Each circle represents a stage in the hierarchy. The ghosted circle represents a stage assumed to exist *but* not actually sampled. The numbers indicate how many samples were obtained from each stage. Along the bottom are the samples taken from terminally maturing cells in each indicated lineage. The topmost circle represents 5 instances of pluripotential precursor cells that were able to differentiate into erythroid, megakaryocytic and macrophage lineages. The stage marked E1 represents hemoglobin-negative erythroid cells at the proerythroblast/CFUE stage, and that marked E2 represents hemoglobin-positive normoblasts and reticulocytes. Meg, megakaryacytes; E, erythroid cells; Neut, neutrophils; B, B lymphocytes; T, T lymphocytes.

Table 2. cDNA samples from single precursors.

Potential[a]	Source[b]	1° culture[c]	Start size[d]	Cycle phase[e]
E/Meg/Mac				
1-5	BM/TdR	I	4	
E/Meg				
1	FL/AA 4.1$^+$	I	8	
2-4	BM/TdR	I	8	
5	BM/TdR	I	4	
6	BM/TdR	I	6	
7	BM/TdR	I	4	
Neut/Mac				
I,2	BM/TdR	I	11	
3-6	BM/TdR	I	8	
Neut/Mast				
1	BM/TdR	I	6	
2	BM/TdR	I	8	
E				
1,2	BM/TdR	I	8	
Meg				
1	BM/TdR	I	5	G1
2	BM/TdR	I	6	M
3	BM/TdR	I	4	
4	BM/TdR	I	6	
Mac				
1	BM	G-CSF	7	
3	BM	G-CSF	10	
3-6	BM	IL-3	8	
7,8	FL/AA4.1 +	I	8	
Neut				
1	BM/TdR	I	4	G1
2	BM	IL-3	10	
3	BM	IL-.3	4	
4	BM/TdR	I	8	
5	BM/TdR	I	7	G1
6	BM/TdR	I	7	G2/M

[a]Assignment of differentiative potential based on fates of cultured siblings.
[b]Primary cultures were initiated with unseparated bone marrow (BM), marrow that had been subjected to thymidine selection for 27 h (BM/TdR), or AA4.1+ cells from day 12 fetal liver (FL/AA4.1+).
[c]Conditions in primary cultures. I: IL-1, IL-3. c-Kit ligand and erythropoietin, with or without conditioned medium from human 5637 cells; starts sampled at 26 h. G-CSF/IL-3: single cytokine present during incubation, starts sampled a! 36 h
[d]Number of cells in colony start when micromanipulated.
[e]Phase of the cell cycle. Where indicated, cycle phase was established by observation of mitosis within 2 h before, or after, micromanipulation of the colony start. Where not indicated, the likeliest phase for cells from 4-cell starts was G2, and from 8-11-cell starts was G1.

Confirmation of stage- and lineage-fidelity of the sample set using lineage-specific probes

The cDNA samples from the cells summarized in Figure 3 were electrophoresed in agarose gels, Southern blotted, and hybridized to a set of lineage-specific probes. Because the polyA-PCR procedure generates a spectrum of 3' cDNA fragments ranging between 200-600 bases, hybridization signals appeared as correspondingly disperse patterns rather than as single bands. Most samples hybridized to probes for the *L32* and *gapdh* transcripts.

Lineage-specific probes showed the expected hybridization specificities (Fig. 4). Transcripts encoding lysozyme (*lys*) and, most notably, myeloperoxidase (*mpo*), were expressed very early in granulocytic and monocytic lineages, before lineage commitment was complete (in agreement with earlier descriptions [5]). Myeloperoxidase activity is not detected; by staining, until later in maturation in these lineages, stressing the distinction between RNA, protein and mature protein levels of expression. The expression of α-*globin* was strong in samples from terminally maturing erythroblasts, but a low level of hybridization was also observed in a bipotent erythroid/megakaryocytic precursor. Two samples of lymphocytes, grown from fetal liver precursors, showed asynchronous expression of *rag-2* and *Ig* (immunoglobulin gene expression was analyzed using a probe from the Cμ locus). The *rag-2* transcript encodes a protein that is involved in immunoglobulin locus rearrangement, and is expressed in B-lymphocyte development before *Ig*; the asynchrony observed here is likely to reflect differences in maturity of the two cultures. Supporting this suggestion is the correlation between the expression of *rag-2* and *c-myb* in the same samples (Fig. 5). As expected [6,7], the highest levels of *lck* expression were detected in T lymphocytes.

Overall, the results were compatible with our distinctions between precursors and terminally maturing cells. They also affirmed that our lineage assignments were correct and that mRNA sequences were faithfully preserved through the amplification procedure.

Mapping of gene expression patterns over the sample set

A large and growing number of genes are known whose products are involved in differentiation, growth, self-renewal or signalled cell death in hemopoietic cells. Although information is available on the stage- and lineage-specificity of RNA expression for many of these, the data have generally been derived from cell populations, frequently from immortalized cell lines, and little information is available on lineage- or stage-specificity in the normally differentiating hierarchy. As our panel of cDNA samples spanned a detailed spectrum of differentiation stages, we anticipated that hybridization analysis would extend our knowledge of expression patterns and possible roles played by the genes tested. Moreover, if our results matched specificities reported previously, the data

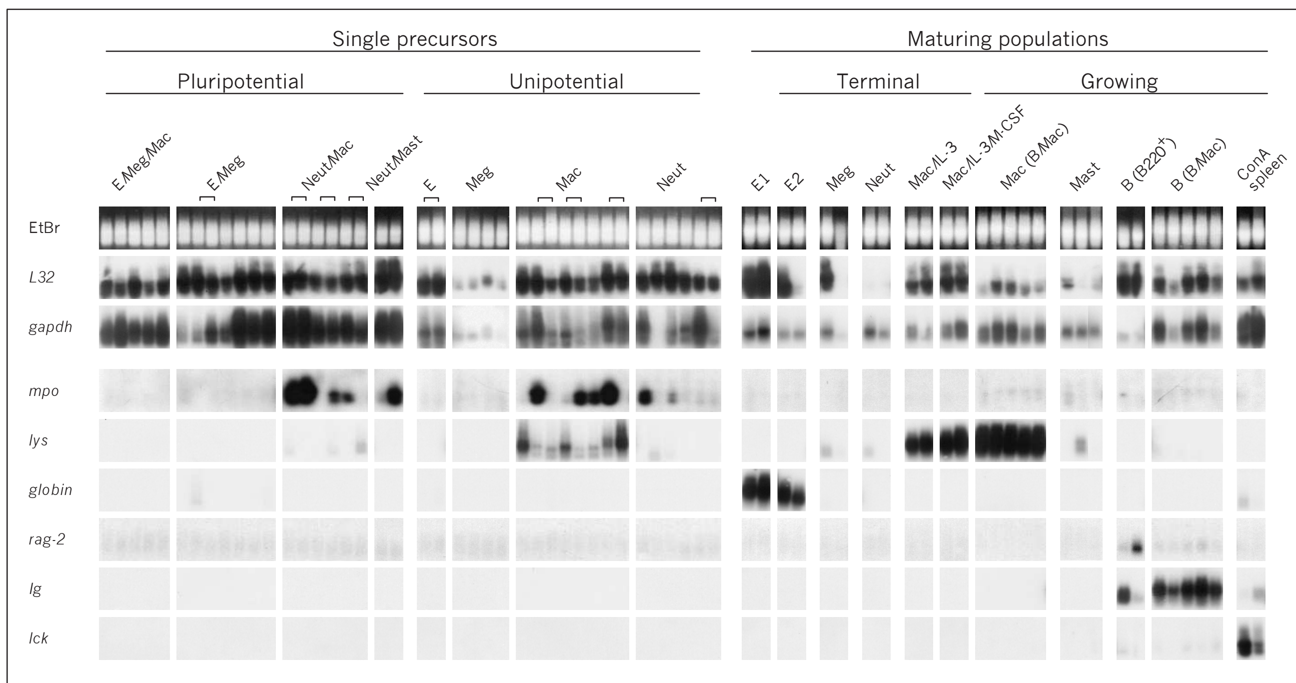

Fig. 4. Hybridization of probes for constitutive 'housekeeping' and lineage-specific genes to cDNA samples from diverse stages in the hemopoietic cell hierarchy. cDNA samples from the indicated cell types and stages were electrophoresed and Southern blotted. Bars join samples derived from sibling cells drawn from the same colony starts. The top row shows the set of samples stained with ethidium bromide before blotting, and indicates that roughly equivalent amounts of amplified cDNA were loaded in each lane. The rest of the figure shows hybridization of the indicated radiolabelled probes to each cDNA sample. For details of cells and culture conditions, see Materials and methods. Abbreviation as in Fig. 3, plus: M-CSF, macrophage colony-stimulating factor; ConA spleen, T lymphocytes grown from adult spleen cells stimulated with the T-cell mitogen concanavalin A.

would add to our confidence in interpreting novel results obtained with the system.

Hybridization results obtained with a broad panel of genes are shown in Figure 5. The first seven rows show patterns for a selection of genes that encode products involved in the control of gene expression, growth and differentiation. The *c-myb* gene encodes a transcription factor that is known to be expressed in immature hemopoietic cells [8]; its transcripts were detected broadly in pluripotential precursors, and in committed precursors transcripts were detected mainly in granulocytic and monocytic cells. Among maturing cells, expression was confined principally to growing B lymphocytes.

The RB-1 protein is a link in a chain that regulates entry of cells into the S phase of the cycle (9]. The expression of *RB-1* is upregulated in murine erythroleukemia cells after the induction of maturation [10], and blood cell formation is defective in *RB-1*$^{-/-}$ mouse embryos [11,12] . Observations in chimeric *RB-1*$^{-/-}$/*RB-1*$^{+/+}$ mice have suggested that the essential function provided by RB-1 may not be intrinsic to hemopoietic precursors [11,12]. Expression of *RB-1* predominated in pluripotential and committed precursors of the erythroid and megakaryocytic lineages (Fig. 5). This result provides direct evidence for the expression of *RB-1* RNA in normal hemopoietic precursors, and specifically in precursors with erythrocytic and megakaryocytic potential.

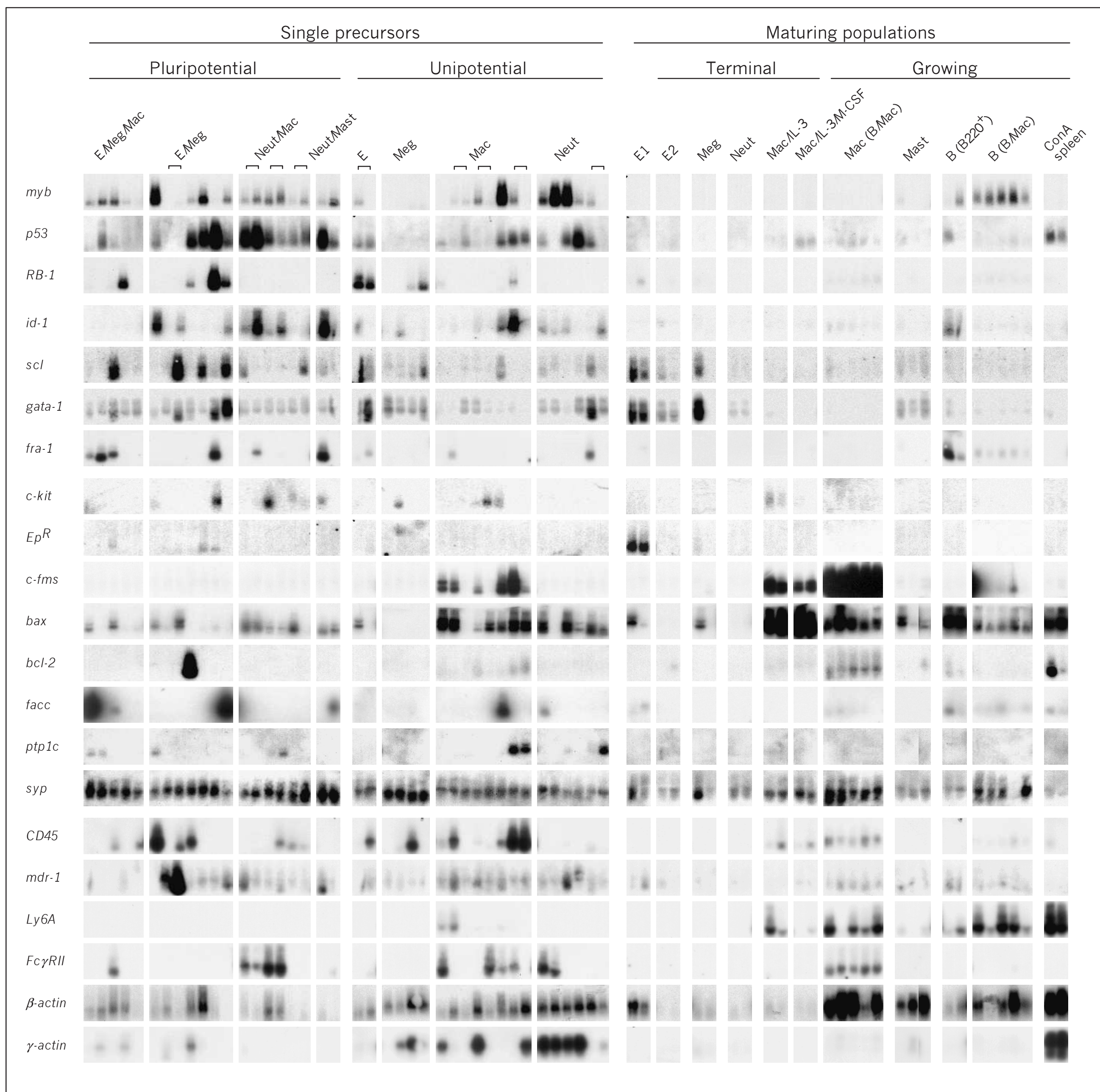

Fig. 5. Hybridization of probes for regulatory and structural genes to the cDNA sample set. Abbreviations as in Figs 3,4.

The observation suggests that RB-1 may indeed play a direct role in early erythroid differentiation, and would be compatible with a direct basis for the defective blood cell production observed in RB-1$^{-/-}$ embryos.

The p53 protein plays a role in signalling events that lead to cell death after DNA damage or cytokine withdrawal [13,14]. Its presence has been demonstrated directly in enriched marrow precursors [15,16], but not in mature myeloid cells. Our observations add further detail, showing p53 expression throughout pluripotential and committed myeloid precursor stages in neutrophil, monocyte and erythroid lineages (Fig. 5). Expression persisted in growing T lymphocytes, and at lower levels in B lymphocytes and macrophages, but was not detected in terminally maturing megakaryocytes, neutrophils or erythroid cells that lacked further growth potential. This expression pattern would be consistent with a role for p53 in the signalled death of growing hemopoietic cells on cytokine withdrawal.

GATA-1 is a zinc-finger transcription factor [17], and SCL [18] and Id-1 [19] are of the helix-loop-helix (HLH) family of transcription factors; all are expressed in immature hemopoietic cells. Id-1 lacks a DNA-binding domain but retains the capacity to hetero-oligomerize with, and thereby inactivate, other HLH family members. Transcripts of *scl* and *gata-1* were detected broadly in pluripotential and committed precursors, but persisted only into maturing cells in the erythroid, megakaryocytic and mast cell lineages (Fig. 5). Pluripotential cells also expressed *id-1*, but we observed a lower frequency of expression in committed progenitors. In maturing cells, *id-1* transcripts were detected in B lymphocytes and, at a very low level, in growing macrophages, but not in committed and maturing erythroid, megakaryocytic and mast cells. Fra-1 is a Fos-family AP-1 transcription factor [20], expression of which in hemopoietic cells has not been reported previously. We detected general *fra-1* expression in early myeloid precursors, but persistence only into the B-lymphocyte lineage (Fig. 5), a pattern shared with *id-1* and *c-myb*.

Figure 5 also shows hybridization patterns of probes for transcripts for three cytokine receptors. Expression of *c-fms* is known to be restricted to cells of the monocytic lineage [21]. We detected *c-fms* transcripts in precursors committed exclusively to monocytic differentiation, but not in earlier pluripotential precursors; expression persisted into maturing monocytic cells, as expected. The absence of detected *c-fms* transcripts in immediate bipotential precursors of the lineage or earlier cells was unexpected. The observation suggests that the expression of *c-fms* may be upregulated only after commitment to the monocyte lineage, and that the receptor it encodes may therefore be excluded from a role in earlier decision processes.

The erythropoietin receptor (EpoR) is known to be expressed in cells near the CFUE/proerythroblast stage of the red-cell lineage and, as expected, we observed strongest hybridizations to samples from this stage. Weak hybridization signals were also detected in 3 out of 12 pluripotential erythroid precursors, although early precursors are not known to be directly responsive to erythropoietin. The receptor encoded by *c-kit* plays an essential role in growth and maintenance of primitive hemopoietic precursors and also in early steps in the erythroid lineage [22]. Hybridization to our probe was detected in bipotential and committed precursors in various lineages, but was weak to undetectable in mast cells and other terminally maturing populations. The sensitivity of this probe was reduced as a result of the high proportion of A and T residues in the 3' untranslated region of the target cDNA.

Most of the other genes whose hybridization results are shown in Figure 5 are known to be expressed in hemopoietic cells, but little information is available concerning their stage- and lineage-specificity. The genes *bcl-2* and *bax* encode components of a system that controls signalled cell death in T and B lymphocytes and in myeloid leukemias [23,24]. Bcl-2 inhibits signalled death by heterodimerizing with Bax [23]. There is evidence that the p53 protein upregulates the expression of bax and represses that of *bcl-2* [25]. We detected *bcl-2* transcripts in monocytes and their committed precursors, and in growing lymphocytes, but not in other lineages or in most pluripotential precursors. In contrast, *bax* transcripts were nearly universally expressed. Our findings are compatible with the suggested relationships between the expression of *p53* and *bax/bcl-2*, and suggest that the involvement of Bcl-2 in regulation of myeloid precursors may be confined to the monocytic lineage.

The phosphatase encoded by *syp* is known to be widely expressed in hemopoietic lineages [26], and our blots confirmed its expression in all of our samples including pluripotential and unipotential precursors (Fig. 5). The murine probe to *facc* - the gene defective in human Fanconi anemia (group C) [27] - hybridized strongly but at low frequency among pluripotential precursors, and weakly to samples from growing B and T lymphocytes and monocytes. This pattern - intense expression in a minority of precursor samples and lower-level expression in growing, terminally differentiating populations of monocytes and lymphocytes - resembles that of *bcl-2*. The protein tyrosine phosphatase encoded by *ptp1c* has been mapped recently to the *motheaten* locus in the mouse [28]; the defective allele leads to a syndrome involving excessive numbers and constitutive activation of tissue macrophages. Like *facc*, *ptp1c* transcripts are difficult to detect in RNA from whole bone marrow. Our blots detected *ptp1c* expression in unipotential progenitors of monocytes and neutrophils, and in earlier pluripotential hemopoietic precursors. *CD45* also encodes a widely expressed hemopoietic protein phosphatase that is anchored in the surface membrane [29]. Transcripts were detected in precursor cells at each stage, but in maturing cells persisted only in the monocytic and lymphoid lineages.

We also probed for transcripts encoding cell-surface or structural proteins that are often used for phenotypic characterization of hemopoietic cells. The transport protein encoded by *mdr-1* confers multiple drug resistance when overexpressed [30]. Transcripts were expressed in precursor samples throughout the hierarchy, as well as in growing, terminally maturing cells in all lineages (Fig. 5). *Ly6A* (*Sca-1*) encodes a surface antigen that is known to be expressed on lymphocytes and monocytes [31], and in certain mouse strains it also marks primitive multipotential stem cells that have long-term *in vivo* reconstitution capacity [1,32]. Our samples were derived from CBA/J mice whose stem cells are not marked. Among precursor cells, only those committed uniquely to monocyte differentiation contained detectable transcripts, as did maturing monocytes and lymphocytes, as expected.

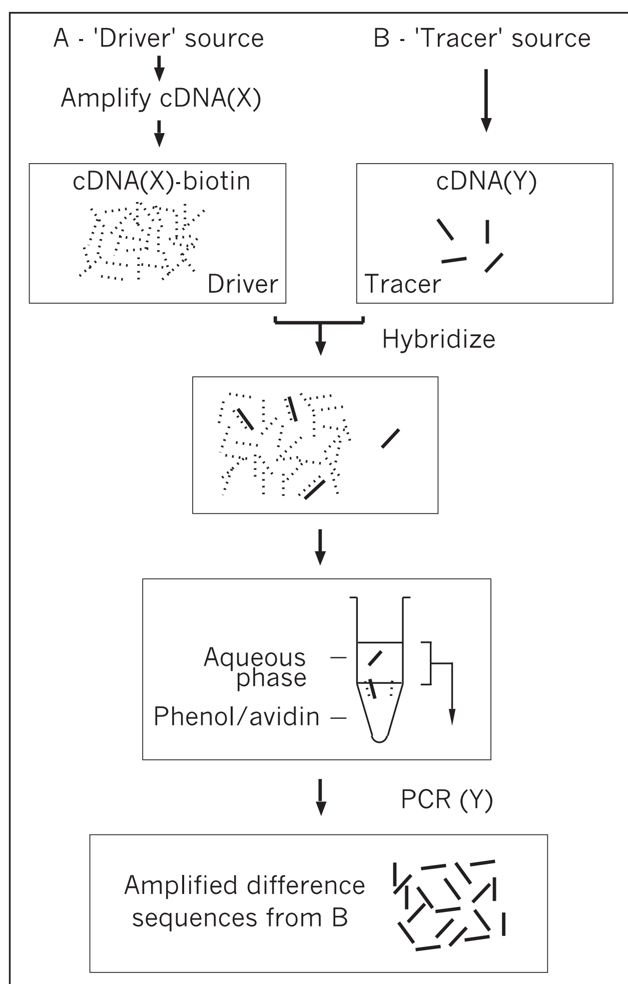

Fig. 6. Isolation of differentially expressed cDNAs by subtractive hybridization. The diagram illustrates the specific isolation of sequences present in the 'tracer' cDNA sample but not in the 'driver' sample A. Sample A is amplified using a primer containing a unique sequence, X, and is then biotinylated. Sample B is amplified using a primer that contains, a unique sequence, Y. After hybridization of B with an excess of A, streptavidin is added and the mixture is extracted with phenol. The avidin localizes to the phenol phase, taking with it the biotinylated cDNA from A together with sequences from B that have hybridized to A. Unhybridized cDNA from B remains free in the aqueous phase, and is specifically amplified using a primer containing the sequence Y.

The *FcγRIIβ1* probe detects transcripts encoding a class of low -affinity cell surface receptors which recognize the Fc portion of IgG. The probe also detects *FcγRIII* transcripts, which encode a second low-affinity receptor for IgG and IgE [33,34]. *FcγRII/III* transcripts were detected in monocytes, as expected [33], but not in lymphocytes. Transcripts were also detected in early pluripotential and committed precursors of granulocytic and monocytic lineages, but not in erythroid or megakaryocytic precursors. Although *FcγRII/III* transcripts have been detected in an enriched murine marrow precursor population [35], this is the first description of their differential expression in precursors of differing commitment states. If accompanied by protein expression, this would have practical implications for experiments in which antibodies are used to mark precursor cells, and it also suggests a potential strategy for the separation of neutrophil and macrophage progenitors from erythroid and megakaryocyte precursors.

The cytoskeletal β- and γ-actins are the major components of microfilaments. Although the relative levels of *β-* and *γ-actin* mRNA [36,37] and protein *[38]* are known to differ in different cell types and tissues, little is known of their expression patterns in hemopoiesis. The *β-actin* gene was generally expressed in growing cells throughout the hierarchy (Fig. 5). In contrast, *γ-actin* expression was detected mainly in precursor cells, particularly in committed megakaryocytic, monocytic and granulocytic precursors; among maturing cells, expression persisted only in Concanavalin-A-stimulated T cells. These results add support to the notion that the two actin isoforms may have distinct functions [39]. Together, the results add significant detail to our knowledge of expression of these genes in normal precursor populations.

Identification of genes expressed at specific stages in the hierarchy by subtractive hybridization

In order to isolate sequences that are expressed at specific stages in the hierarchy, we developed a protocol for subtractive hybridization that builds on advantages inherent in PCR-derived samples (Fig. 6). Essential features are the availability of 'driver' and 'tracer' cDNA in unlimited quantities, the specific amplification of unhybridized tracer sequences using a primer that distinguishes them from driver sequences, and the specific generation of unlimited quantities of subtraction product from even molecular amounts of unhybridized tracer.

To test the utility of the approach, we conducted a pilot experiment in which a cDNA sample from a bipotential neutrophil/macrophage precursor (driver) was subtracted from an erythroid/megakaryocyte precursor sample (tracer), and a second experiment in the reverse orientation. Difference sequences were subjected to a total of five rounds of sequential hybridization and reamplification. The final result from each orientation was cloned into a plasmid library, and 48 cDNA clones were picked at random from each. Dot blots were prepared from the sampled cDNA clones and hybridized to probes prepared

from total driver or tracer cDNA; 10-30% of clones hybridized to tracer but not driver. The remainder of the cDNA clones reflected the incomplete removal of shared sequences or contamination of the subtracted samples with inappropriately amplified driver sequences. Cross-hybridization experiments showed that the selected sets were almost entirely non-redundant.

To obtain a broader picture of specificity of expression of the selected clones within the hierarchy, they were individually labelled and hybridized to the master sample set. A sampling of the results is shown in Figure 7. Included are examples hybridizing to tracer but not driver, hybridizing to driver but not tracer, hybridizing to both, and hybridizing uniquely to the single cDNA sample from which the cDNA clone arose. Although numerous cDNA clones were identified whose expression differed between the individual cells that originated the driver and tracer cDNAs, only one example (*gb32*) was identified that was expressed in one precursor cell class but not in any other members of the precursor class exemplified by the subtraction partner. The clones *m13* and *gb32* show patterns of specificity that make them attractive candidates for further investigation. An unexpected observation was that many of the probes derived from the two precursor cells hybridized strongly to hemopoietic precursors but not to maturing cells in various lineages, even though many of the latter examples were still actively growing when sampled.

These results show that it is feasible to perform subtractive hybridization on cDNA samples derived from single cells, and that the procedure can yield difference sequences in substantial numbers. Transcripts that are uniquely expressed in rare progenitor cells may be too dilute in conventionally purified populations or cell lines to be captured with standard approaches. Our approach avoids dilution by non-desired cell types, and maximizes the opportunity of capturing precursor-specific transcripts. Moreover, once a difference tag is identified, it can be rapidly screened for consistency and specificity of expression by hybridization to the set of lineage- and stage-specific cDNA samples.

Discussion

This study introduces a novel system for high-resolution mapping of gene expression in a complex cellular hierarchy. The system is based on three principal elements. First, a polyA-PCR technique provides for unbiased amplification of all polyadenylated RNA transcripts present in a cellular sample, preserves abundance relationships, and performs reliably with samples as small as a

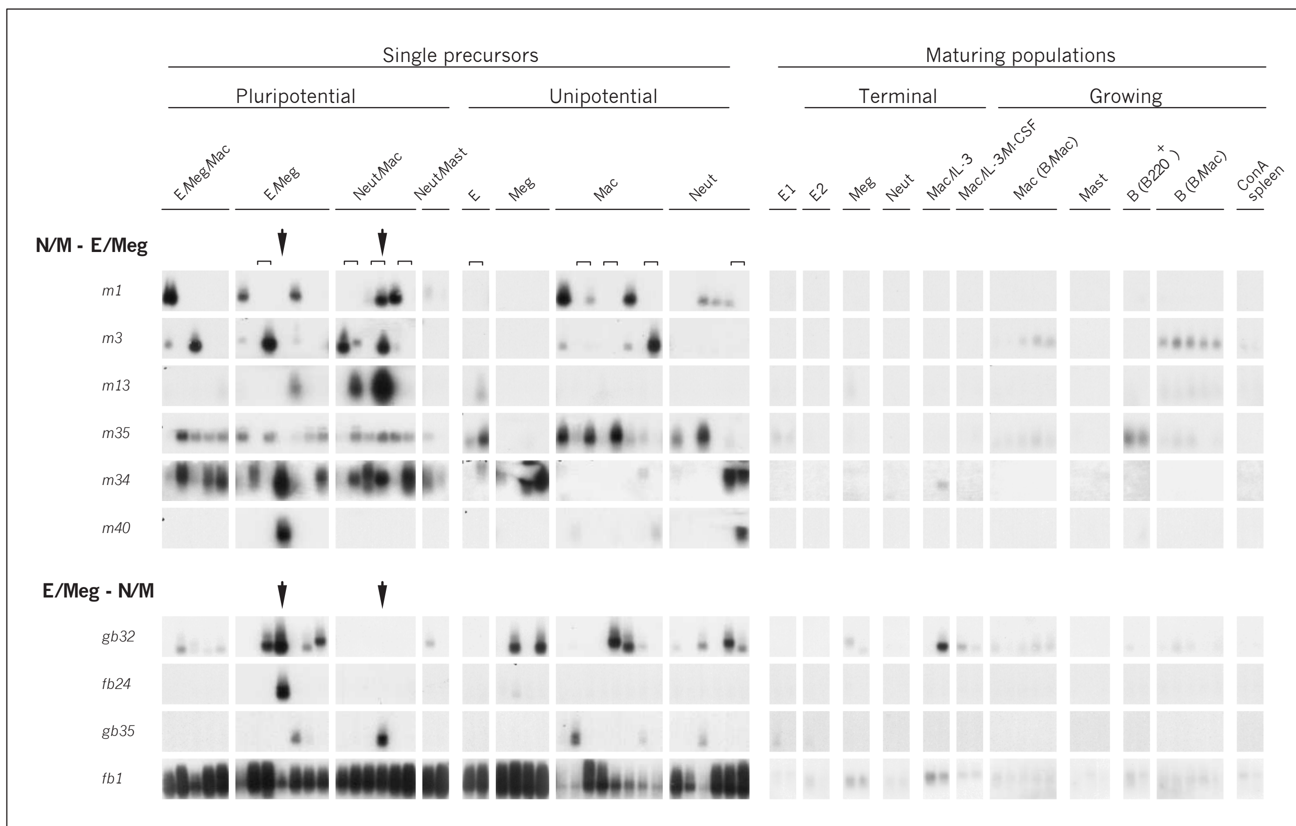

Fig. 7. Expression patterns of probes obtained from subtractive hybridization of cDNA from two individual cells. The cDNA preparations from a single E/Meg precursor and from a single Neut/Mac precursor, indicated with arrows, were subtractively hybridized in both possible orientations. The difference cDNAs were cloned into libraries and individual clones were randomly isolated, radiolabelled and hybridized to Southern blots as in Figs 4-6. Abbreviations as in Figs 3,4.

single cell. The resulting samples are permanently renewable by reamplification. Second, the technique is applied to rare precursor cells of diverse developmental potential by sampling them as single cells, thereby assuring homogeneity of the source of amplified cDNA. Third, sampling of single cells is coupled with direct observation of the biological potential of sibling cells, which thus serve as reporters of the probable biological potential of the processed cells.

We used the cDNA set as a matrix for mapping the stage- and lineage-specificity of expression of a large selection of genes. In returning answers expected on the basis of earlier work in conventional systems, the results confirmed the coherence and fidelity of the amplification procedure, and the accuracy of our phenotypic assignments based on sibling outcomes. A selection of illustrative results is shown schematically in Figure 8. We detected general expression of *p53* in precursors; expression was downregulated in mature progeny that lack further growth potential. *RB-1* illustrates an unexpectedly specific pattern that was confined principally to erythroid and megakaryocytic cells and their uncommitted precursors. The expression of *FcγRIIβ1* also presented an unanticipated pattern of specificity among pluripotential and committed precursors.

We also tested the ability of cDNA-based subtractive hybridization to select transcripts expressed in one precursor cell and not in another. Difference transcripts were readily identified, and further information on their specificity was gained by probing for their expression in the master sample collection. The hybridization pattern of one such clone, *gb32,* is also shown in Figure 8. Sequence analysis of these clones suggests that many will turn out to represent previously unreported genes (data not shown).

A striking observation in this study was the marked variability of hybridization of many of our probes to cDNA samples derived from cells that ostensibly shared the same developmental potential. Several possible sources of the variation can be envisioned. Variability might be generated artefactually during reverse transcription and amplification in vitro. A number of observations suggest that this is not the case. First, the same samples hybridized probes for constitutively expressed genes, such as *L32*, *gapdh, syp* and others, with relative uniformity. Moreover, control experiments were performed in which lysates were prepared from 100 pooled cells and then distributed in aliquots representing as little as 1/200 of the total, prior to reverse transcription and general amplification. Variability in sample-to-sample hybridization should be reproduced in such experiments if it arises in the process of amplification from small amounts of template, but should not be otherwise observed because each aliquot should be identical. Variation in probe-hybridization to these samples was relatively minor compared to the variation observed in samples prepared from single hemopoietic precursors, even where the amount of template used in the pooled controls was less than a single cell-equivalent.

We have reported previously that this procedure preserves relative abundance relationships, even when initiated with amounts of template typical of single cells, and that sensitivity of detection extends at least to transcripts represented at an initial abundance of 0.25 % [2]. Given the relative constancy of hybridization of the same samples to probes for constitutively expressed genes such as *L32* and *gapdh,* the repeatedly observed preservation

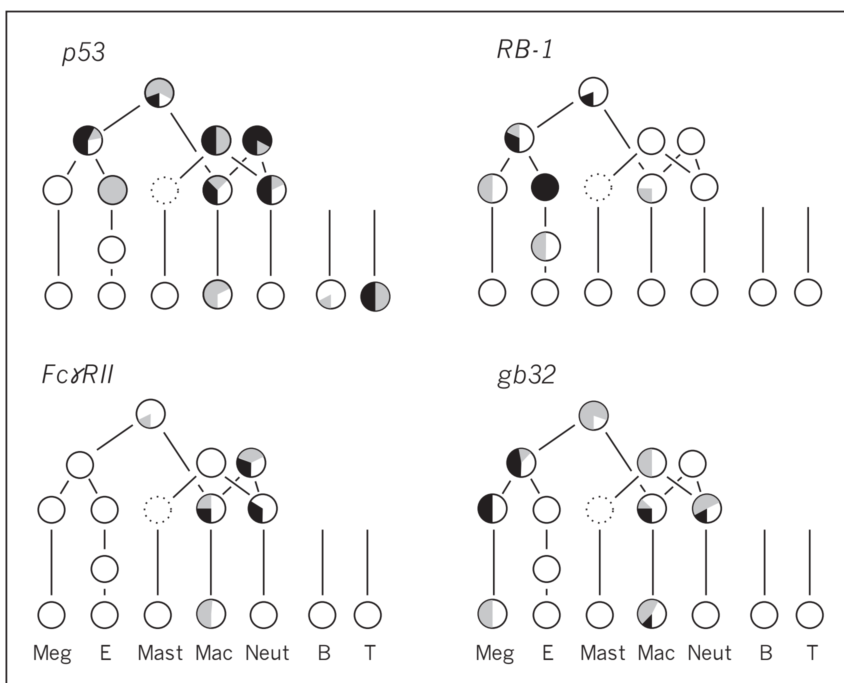

Fig. 8. Examples of differing expression patterns within the hemopoietic cell hierarchy. Expression patterns shown in Figs 5, 6 and 8 are positioned schematically on the sample hierarchy illustrated in Fig. 3. The solid black pattern represents strong hybridization, grey weaker hybridization, and white undetected hybridization. The proportions of the sample sets labelled at each position are indicated by the sector sizes. For example, the p53 probe hybridized strongly to four of seven cDNA samples from bipotential erythroid/megakaryocyte precursors, and more weakly to one of them. Abbreviations as in Fig. 3.

of abundance relationships (especially in samples drawn from cell populations) and the apparent sensitivity limits, we conclude that threshold effects are unlikely to account for extremes of intense hybridization alternating with undetectable hybridization – seen, for example, with the probes for *bc1-2* and *facc.*

Another possible source of variation to be considered is position in the cell cycle, an issue which will require systematic experimental analysis. However, we observed marked variation even in cDNA samples derived from siblings that were positioned within minutes of each other in the cell cycle. Moreover, although we knew that some of the cells were sampled in late G2 and others in early G1, corresponding correlations with expression were not observed. Finally, we consider that marked fluctuation in transcript abundance may be an inherent aspect of the transcription process itself. Transcription complexes assemble by random collisions of multiple proteins with promoter/enhancer sites, their transcriptional activity can pause, and formed complexes can dissociate. Because only two such complexes can exist in a diploid cell for a given gene, the timing of transcriptional activity is bound to be stochastic [40,41]. If the intervals between active transcription are significant relative to transcript half-lives, large temporal variations in transcript concentrations would result which would be observable in individual cells but not in populations. Although this is our favoured explanation, proof will require exclusion of the alternatives.

Whatever its sources, sample-to-sample variation in hybridization has practical implications for experimental design and interpretation. Assessment of the expression of a gene at a particular developmental stage requires the analysis of adequate numbers of individual cDNA samples. Intensity of expression may be better inferred from the proportion of positive hybridizations within a set of samples than from the intensity of individual hybridizations. High sample-to-sample variations may obscure correlations in expression of different genes if looked for only at the single-cell level; significant correlations should however become apparent at the level of sample sets. Similarly, if subtraction is to be used to identify differences in gene expression between two developmental stages, it should be performed with pools of individual cell samples to avoid the ‘quantum noise’ that would be generated by subtraction of single-cell samples. Numerous clones yielded by subtraction of cDNA from single cells represented true expressional differences between the individual subtraction partners, but not true differences between the developmental stages they represented (see Fig. 7).

In the course of this study, we frequently encountered extreme examples of intermittency in which individually cloned transcripts hybridized strongly to cDNA from their cell of origin but not to our other cDNA samples (for example, *fb24*; see Fig. 7). It is possible that such instances reflect low-probability events that have no important role in cell function. Together they would contribute to a significant background of ‘inappropriate’ transcripts that are readily detected in individual cells but not in cell populations.

The polyA-PCR procedure has unique advantages in a wide variety of settings. Applied to populations, polyA-PCR requires only a few cells and yields hybridization results quantitatively similar to those obtainable with bulk mRNA, with comparable limits of sensitivity [2]. Applied to single cells, it can be used to investigate transcriptional frequencies, simultaneous expression patterns of any desired number of genes and, most importantly, allows the preparation of cDNA from absolutely homogeneous cell sources even when those cells are drawn from mixed or asynchronous populations. Once generally amplified cDNA is made, specific PCR can be applied to detect species that are too rare to be detectable on Southern blots of the primary cDNA samples (FB., unpublished observations). Amplified cDNA also has unique advantages as a starting point for subtraction. As rare unhybridized sequences can subsequently be reamplified, subtractive approaches have a high degree of sensitivity, and exhaustive subtraction with repetitive rounds of subtraction and reamplification becomes feasible without concern for quantity of subtracted material. As an alternative to subtraction, differential display can also be applied to amplified cDNA pairs for rapid identification, particularly of higher-abundance differences [42].

The experiments described here have also identified directions for further refinement of the procedures. The polyA-PCR protocol amplifies only a few hundred bases of 3' terminal sequence, preserving relative abundance relationships but sacrificing direct access to coding regions. Recently described strategies relieve the length constraints on the PCR procedure and are likely to be applicable to polyA-PCR [43]. The subtractive protocol can also be made more efficient than the double-stranded procedure described here. Preliminary experiments indicate that nearly complete subtraction of shared sequences is possible with only one round of hybridization to single-stranded driver cDNA.

Conclusions

This study was designed to determine patterns of gene expression within a multilineage precursor hierarchy, using a novel approach to resolve precursors at individual developmental stages. The results established that a procedure for universal amplification of polyadenylated mRNA can be applied productively to single hemopoietic precursor cells and can be used to analyze much of the transcriptional repertoire in each cell examined. The experiments also established that sibling cells in colony starts can report reliably on the developmental potential of a single cell taken for PCR. By combining global cDNA analysis of single precursor cells with the identification of their potential by sibling analysis, it was possible to map expression, at the RNA level, of a large panel of

genes known to be involved in hemopoietic cell growth, differentiation or function. The results supply the first view of expression of these genes at a level of resolution that distinguishes precursor cells in differing commitment and maturational states, and thus provide insights into specificity of expression that were not available previously. Subtractive hybridization of cDNA from one precursor cell from that of another representing a different commitment state can provide a rich harvest of differentially expressed transcripts, and it appears that many will prove to be novel. By hybridizing the difference transcripts to the master set of precursor cDNA samples, those with interesting and consistent patterns of specificity over the entire lineage tree can be rapidly identified.

Materials and methods

Cells and cultures

Marrow cells were flushed from femurs of 8-12-week-old CBA/J mice. Where indicated, cells were enriched for primitive, more slowly cycling cells by incubation for 20 h with 100 μg ml^{-1} thymidine in IMDM with 5 % FBS, 1 ng ml^{-1} human IL-1P and 15 U ml^{-1} IL-3, as described [4]. Liver cells were obtained from day 12 C57BL/6J fetuses and subsequently enriched by panning for $B220^{+}$ cells, or for $B220^{-}$/Mac-1^{-}, $AA4.1^{+}Ly6A^{+}$ cells containing bipotential precursors of B cells and macrophages, as described [44].

Colony starts were initiated in ‘complete' methyl cellulose cultures containing IMDM, 4 % FBS, serum fraction V, lipids, transferrin, insulin, 15 % conditioned medium from human 5637 cells, IL-1, IL-3, murine c-Kit ligand and erythropoietin, as described [1,4]. Where indicated, cultures contained only G-CSF, or only IL-3, without the other hemopoietic cytokines.

Terminally maturing erythroid cells, megakaryocytes, neutrophils and macrophages were sampled from single-lineage colonies grown from marrow-cell precursors in methyl cellulose containing IL-1, IL-3 and erythropoietin. Where indicated, some macrophage colonies were grown in the additional presence of L929 cell-conditioned medium as a source of M-CSF. Maturation state and lineage composition were assessed by examination of May-Grunwald-Giemsa-stained spreads from the same colonies. Maturing B cells were sampled from bulk 7-day cultures of $B220^{+}$ fetal liver cells growing in response to IL-7, as described [44]. Maturing B cells and macrophages were also taken from clonal liquid cultures initiated from single bipotential precursors purified from day 12 fetal liver [44]. B-cell subcultures were grown with IL-7 and fibroblast feeder cells and sampled 12-15 days after initiation from fetal liver. Macrophage subcultures were grown with M-CSF, IL-3 and c-Kit ligand and sampled at 14-20 days, as described [44]. T cells were taken from 4-day cultures of adult spleen cells growing in response to 2 μg ml^{-1} concanavalin A. Mast cells were sampled from bulk cultures of adult marrow cells passaged for 25-46 days in IMDM containing 15 U ml^{-1} IL-3, 2 μg ml^{-1} concanavalin A, 1 % FBS, 0.05 % bovine serum fraction V, 5 μg ml^{-1} transferrin and 5 μg ml^{-1} insulin. Homogeneity was assessed by staining with Alcian Blue [45].

Sibling precursor cells from colony starts were plated in gridded plates in complete methyl cellulose. Secondary colonies were visualized daily and sampled for morphological assessment when growth ceased. Peripheral cells from large colonies that were still growing at b-7 days were sampled for morphology and specific staining at that time, and at 2-3-day intervals until growth stopped. Cell morphology was evaluated after staining with May-Grunwald-Giemsa. Megakaryocytic cells and precursors were identified by histochemical staining for acetylcholinesterase [46], and mast cells were identified by Alcian Blue staining. Some colonies that were composed predominantly of terminally maturing neutrophils at 6-8 days were overgrown in the succeeding week by slowly-dividing, highly refractile cells that were tentatively classified as mast cells or eosinophils, without confirmation by specific staining.

General amplification of cDNA

The procedure is shown schematically in Fig. 1 and described in detail elsewhere [2,47]. Briefly, cells were lysed in a small volume of buffer containing NP40 detergent and RNAse inhibitors; cDNA with an average length of 400 bases was generated on the liberated mRNA templates with reverse transcriptase and an oligo$(dT)_{24}$ primer. After addition of a homopolymeric 3' (dA) tail with terminal deoxynucleotide transferase, the resulting cDNA mixture was amplified by PCR using a 3'$(dT)_{24}$-containing 60-base primer [47]. Additional quantities of amplified cDNA were generated as required by reamplification using the same primer [47]. Relative hybridization patterns of various probes did not change with repeated reamplification.

Southern blotting and probing

Amplified cDNA, normally 0.1 μg per lane (0.5 μg for *mpo*, *lck*, *id-1*, *scl*, *gata-1*, *bcl-2*, *facc*, *ptp1c*, *CD45*, *mdr1*, *Ly6A*, *FcγRIIβ1*, was electrophoresed in 1.5 % agarose in Tris-borate buffer, denatured in NaOH/NaCl and transferred to $HybondN^{+}$ nylon membranes (Amersham) as described [47]. Probes labelled with ^{32}P were prepared by random priming (Oligolabelling Kit, Pharmacia) from templates which included the extreme 3' untranslated sequences of the genes of interest. Hybridization and washing were carried out as described [47]. Digitized images of labelled blots were obtained from PhosphorImager screens (Molecular Dynamics) for *gata-1*, *id-1* and *scl*, or by digital densitometric scanning of autoradiograms (Computing Densitometer, Molecular Dynamics) for all other probes.

Probes

Unless otherwise noted, probes were derived from the corresponding murine cDNA's and contained terminal 3' untranslated sequence located no more than 50 bp from the polyadenylation sites; each was validated by size determination and hybridization to appropriate positive and negative controls. The sequence references are indicated together with the donors of probes whose gifts are gratefully acknowledged: ribosomal *L32*, murine genomic 1.6 kb *SacI* fragment encompassing the final exon ([48]; C. Paige); *gapdh*, rat 1.3 kb cDNA ([49]; P. Dubrevil); *mpo*, human cDNA, 2 kb *EcoRI* fragment, pMP503 ([50]; ATCC); lysozyme M (Iys), 500 bp 3' *EcoRI* fragment ([2,51]; M. Cross); (*α-globin*, 1.6 kb 3' *HindIII-Aat2* fragment [52]; *rag-2*, 2.0 kb ([53]; D. Schatz); *Ig* (Cμ), 300 bp 3' *PstI* fragment ([54]; G. Wu); *lck*: 30 bp synthetic oligonucleotide 200 bp upstream of the polyA tail [55,5b]; *c-myb*, 600 bp 3' *BglII* fragment ([57]; ATCC); *p53*, 720 bp 3' fragment (D. Monroe and S. Benchimol, unpublished); *RB1*, 950 bp 3' *ScaI-HindIII* fragment (P, Hamel and R.A. Phillips); *id-1*, 927 bp full-length fragment including 3' untranslated region ([19]; H. Weintraub); *scl*, 438 bp fragment, 46 bp upstream of the polyadenylation signal, prepared by PCR (T.H.) [58];

gata-1, 443 bp terminal 3' fragment, prepared by PCR (T. H.) [59]; *fra1*, 444 bp 3' *Xba1* fragment ([20]; E. Wagner); *c-kit*, 315 bp 3' fragment prepared by PCR and confirmed by sequencing (G.B.); *EpR,* 1.75 kb fragment ([60]; G. Wong*); cfms,* 450 bp 3' ApII-EcoRI fragment ([21]; T. Pawson); *bax*, 943 bp terminal EcoRI fragment ([23]; Z. Oltvai and S J. Korsmeyer); *bcl-2*, 841 bp fragment ending 150 bp upstream from the polyA tail ([61]; Z. Altvai and S J. Korsmeyer); *facc*, 708 bp 3' *NheI-EcoRI* fragment [27]; *ptp1c*, 1.9 kb fragment, ending 100 bp upstream from the polyA tail [28]; *syp*, 2.0 kb fragment, ending 100 bp upstream from the polyA tail [62]; *CD45*, 295 bp 3' *BamHI-C1aI* fragment ([29]; *Ly5-68*, ATCC; *mdr1*, 4 kb *HindIII-EcoRI* fragment ([30], P. Gros); *Ly6A.2*, 6.3 kb *EcoRI* fragment ([31], A. Bothwell); *FcγRIIβ1*, 1.3 kb *PstI* fragment ([63]; P. M. Hogarth); *β-actin*, human cDNA, 404 bp *EcoRI* fragment ending 200 bp upstream from the polyA tail ([64]; L. Kedes); and *γ-actin*, human cDNA, 322 bp *EcoRI* fragment ending 200 bp upstream from the polyA tail ([65]; L. Kedes).

Subtractive hybridization

The following procedures were used to enrich for cDNA species present in sample A but not in sample B. The general procedure is adapted from [66]. Tracer cDNA was prepared from sample A in a conversion reamplification using ‘KV(dT)’ primer — (3'-5' (dT)24 CTTAAGTAGAGGGATATCACT-CAGCATAATCAATGG) — for 30 cycles (94 °C 1 min, 42 or 50 °C 2 min, 72 °C 2 min). Driver cDNA was prepared from sample B using ‘NotI(dT)’ primer — 3'-5' (dT) 17 CGCC GGCG) for 30 cycles (as above). The converted driver and tracer cDNAs were purified from primers and enzymes (Qiagen Q100 columns, from Qiagen, or Wizard PCR Prep columns, from Promega, respectively) and final DNA concentration was determined by OD_{260}.

Driver cDNA was photobiotinylated following the general procedures described in [66]. cDNA (20 µg) in HE buffer (10 mM HEPES (Sigma), 1 mM EDTA, pH 8.0) was boiled for 2 min and then kept immersed in ice water. Biotin Photoactivatable (40 µg; Calbiochem) in water was added and the mixture was directly illuminated (optimum wavelength 350 nm) using either a GE RSM 110V 275W sun lamp at a distance of 10 cm or an unfiltered 12V 75W halogen projector lamp. After 15 or 30 min, respectively, an additional 40 µg photobiotin was added and irradiation was continued for 5 or 10 min. An equal volume of 200 mM TrisHCl, pH 9.0 was added to stop the reaction. Unbound biotin was repeatedly extracted in equal volumes of TE-saturated 2-butanol until the organic phase was uncoloured. After chloroform extraction, the derivatized cDNA was precipitated overnight with LiCl/ethanol, centrifuged, washed in 70 % ethanol and pelleted again to yield a distinctly orange-red pellet.

For hybridization, 4000 ng biotinylated driver and 200 ng tracer were suspended in 50 µl EPPS buffer 10mM with 5 µg tRNA (Sigma). The mixture was brought to 100 °C for 2 min, chilled and precipitated in ethanol/acetate overnight. The pellet was dissolved in 4 µl 5/4 hybridization buffer (EPPS/EDTA/ SDS); 1 µl 5 M NaCl was added, followed by an overlay of mineral oil, and the mixture was brought to 100 °C for 5 min. Hybridization was carried out at 68 °C for 48 h in a final buffer composition of 10 mM EPPS (pH 8.25), 1 mM EDTA, 1 % SDS and 1 M NaCl. Extraction buffer (95 µl; 50 mM EPPS, pH 8.5, 500 mM NaCl, 2 mM EDTA) was then added.

For isolation of unhybridized tracer species, 20 µl of the mixture was withdrawn to a fresh tube and diluted to 100 µl with extraction buffer. Streptavidin (Sigma) (4 µg from 4 mg ml^{-1} in 50 mM EPPS, pH 8.25, stored at 20 °C) was added. After incubation for 2 min at room temperature, the mixture was extracted with 100 µl TE-saturated phenol/chloroform, vortexed and centrifuged. From the aqueous supernatant, an aliquot of 80 µl was withdrawn, carefully avoiding the interface where biotinylated driver and hybridized tracer were concentrated. Extraction buffer (25 µl) was added to the phenol/chloroform, vortexed and centrifuged, and an additional 20 µl aqueous supernatant harvested. The 100 µ1 aqueous material was subjected to an additional round of streptavidin addition and phenol/chloroform extraction as above, and then a final round of extraction with phenol/chloroform followed by chloroform alone. From a 5 µl aliquot, tracer species were specifically reamplified in a 100 µl PCR reaction primed with ‘KV’ primer — 3'-5' GATATCACTCAGCATAAT-CAATGG, lacking the 3' oligo(dT) sequence but identical in 5' terminal sequence to the Kv(dT) primer — for 30 cycles (94 °C 1 min, 50 °C 2 min, 72 °C 2 min). The resulting amplified difference cDNA was subjected to four additional rounds of subtraction against driver.

For library cloning, amplified difference cDNA was cleaved within the KV(dT) primer sequence with *EcoRI*, purified to eliminate the liberated primer fragment, and ligated into *EcoRI*-digested, dephosphorylated pBlueScript II KS+ plasmid (Stratagene) following procedures as detailed [47].

Acknowledgements: The first two authors contributed equally to this work. We thank M. Barbara and D. Hyam for expert technical assistance, Genetics Institute, Glaxo Research Institute and Kirin Brewery for generous gifts of hemopoietic cytokines, F.C.H. Franklin and H.C.C. Foot for advice on design of the subtraction protocol, the donors listed elsewhere of probes used in this study, F. Kiefer for hybridization data for *fra-1*, and F. Kiefer and S. Benchimol for critical reading of the manuscript. The work was supported by operating grants to N.N.I., T.H., R.H,, M.B., N.M. and K,S. from the National Cancer Institute of Canada and the Medical Research Council of Canada. G. Brady was a recipient of a Leukemia Society of America Fellowship and G. Boehmelt a Fellowship from the Human Frontiers of Science Program Organization.

References

1. Trevisan M, Iscove NN: **Phenotypic analysis of murine long-term hemopoietic reconstituting cells quantitated competitively in vivo and comparison with more advanced colony-forming progeny.** *J Exp Med* 1995, **181:**93–103.
2, Brady G, Barbara M, Iscove NN: **Representative in vitro cDNA amplification from individual hemopoietic cells and colonies.** *Meth Mol Cell Bio1* 1990, **2:**17-25.
3. Suda J, Suda T, Ogawa M: **Analysis of differentiation of mouse hemopoietic stem cells in culture by sequential replating of paired progenitors.** *Blood* 1984, **64:**393-399.
4. Iscove NN, Yan XQ: **Precursors (preCFCmulti) of multilineage hemopoietic colony-forming cells quantitated in vitro: uniqueness of IL-1 requirement, partial separation from pluripotential colony forming cells, and correlation with long term reconstituting cells in vivo.** *J 1mmunol* 1990, **145:**190-195.
5. Sagoh T, Yamada M: **Transcriptional regulation of myeloperoxidase gene expression in myeloid leukemia HL60 cells during differentiation into granulacytes and macrophages.** *Arch Biochem Biophys* 1988, **262:**599-604.
6. Marth JD, Lewis DB, Wilson CB, Gearn ME, Krebs EG, Perlmutter RM: **Regulation of p56lck during T-cell activation: functional implications for the src-like protein tyrosine kinases.** *EMBO J* 1987, **6;**2727-2734.
7. Bolen JB, Rowley RB, Spana C, Tsygankov AY: **The Src family of tyrosine protein kinases in hemopoietic signal transduction.** *FASEB* 1992, **6:**3403-3409.
8. Mucenski M, McLain K, Kier AB, Swerdlaw 5H, Schreiner CM, Miller TA, *et al.*: **A functional c-myb gene is required for normal murine fetal hepatic hematopoiesis.** *Cell* 1991, **65:**677-689.

9. Wang JYJ, Knudsen ES, Welch PJ: **The retinoblastoma suppressor protein.** *Adv Cancer Res 1994,* **64:**25-85.
10. Coppola JA, Lewis BA, Cole MD: **Increased retinoblastoma gene expression is associated with late stages of differentiation in many different cell types.** *Oncogene 1990,* **5:**1731-1733.
11. Williams BO, Schmitt EM, Remington L, Bronson RT, Albert DM, Weinberg RA, Jacks T: **Extensive contribution of Rb-deficient cells to adult chimeric mice with limited histopathological consequences.** *EMBO J* 1994,**13:**4251-4259.
12. Maandag EC, van der Valk M, Vlaar M, Feltkamp C, O'Brien J, van Roon M, *et al.*: **Developmental rescue of an embryonic-lethal mutation in the retinoblastoma gene in chimeric mice.** *EMBO J 1994,* **13:**4260-4268.
13. Gottlieb E, Haffner R, von Ruden T, Wagner EF, Oren M: **Down-regulation of wild-type p53 activity interferes with apoptosis of IL-3-dependent hematapoietic cells following IL-3 withdrawal.** *EMBO J* 1994, **13**:1368-1374.
14. Lotem J, Sachs L: **Hematopoietic cells from mice deficient in wildtype *p53* are more resistant to induction of apoptosis by some agents.** *Blood 1993,* **82**:1092-1096.
15. Rivas CI, Wisniewski D, Strife A, Perez A, Lambek C, Bruno S, *et al.*: **Constitutive expression of p53 protein in enriched normal human marrow blast cell populations.** *Blood* 1992, **79:**1982-1986.
16. Bi S, Lanza F, Goldman JM: **The involvement of 'tumor suppressor' p53 in normal and chronic myelogenous leukemia hemopoiesis.** *Cancer Res* 1994, **54:**582-586.
17. Pevny L, Lin CS, D'Agati V, Simon MC, Orkin SH, Costantini F: **Development of hematopoietic cells lacking transcription factor GATA1.** *Development* 1995,**121:**163-172.
18. Shrivdasani RA, Mayer EL, Orkin SH: **Absence of blood formation in mice lacking the T-cell leukaemia oncoprotein tal1/SCL.** *Nature 1995,* **373**:432-434.
19. Benezra R, Davis RL, Lockshon D, Turner DL, Weintraub H: **The protein Id: a negative regulator of helix-loop-helix binding proteins.** *Cell* 1990, **61:**49-59.
20. Cohen DR, Curran T: ***Fra-1*: a serum-inducible, cellular immediate-early gene that encodes a fos-related antigen.** *Mol Cell Biol*1988, **8:**2063-2069.
21. Rothwell VM, Rohrschneider LR: **Murine *c-fms* cDNA: cloning, sequence analysis and retroviral expression.** *Oncogene Res* 1987,**1:**311-324.
22. Chabot B, Stephensan DA, Chapman VM, Besmer P, Bernstein A: **The proto-oncogene *c-kit* encoding a transmembrane tyrosine kinase receptor maps to the mouse *W* locus.** *Nature* 1988, **335:**88--89.
23. Oltvai ZN, Milliman CL, Korsmeyer SJ: **Bcl-2 heterodimerizes in vivo with a conserved homolog, Bax, that accelerates programed cell death.** *Cell* 1993, **74:**609-619.
24. Yang E, Zha J, Jockel J, Boise LH, Thompson CB, Korsmeyer SJ: **Bad, a heterodimeric partner for Bcl-xL and Bcl-2, displaces Bax and promotes cell death.** *Cell* 1995, **80:**285-291.
25. Miyashita T, Krajewski S, Krajewska M, Wang HG, Lin HK, Liebermann DA, *et al.*: **Tumor suppressor p53 is a regulator of *bcl-2* and *bax* gene expression *in vitro* and *in vivo*.** *Oncogene* 1994, **9:**1799-1805.
26. Freeman RM, Plutzky J, Neel BJ: **Identification of a human src homology 2-containing protein-tyrosine-phosphatase: a putative homolog of Drosophila corkscrew.** *Proc Natl Acad Sci USA* 1992, **89:**11239-11243.
27. Wevrick R, Clarke CA, Buchwald M: **Cloning and analysis of the murine Fanconi anemia group C cDNA.** *Hum Mol Genet* 1993, **2:** 655-662.
28. Kozlowski M, Mlinaric-Rascan I, Feng GS, Shen R, Pawson T, Siminovitch KA: **Expression and catalytic activity of the tyrosine phosphatase PTP1C is severely impaired in *motheaten* and *viable motheaten* mice.** *J Exp Med* 1993,**178:**2157-216 3.
29. Shen FW, Saga Y, Litman G, Freeman G, Tung JS, Cantor H, Boyse EA: **Cloning of Ly5 cDNA.** *Proc Natl Acad Sci USA* 1985, **82:**7360-7363.
30. Gros P, Croop J, Hausman D: **Mammalian multidrug resistance gene: complete cDNA sequence indicates strong homology to bacterial transport proteins.** *Cell* 1986, **47:**371-380.
31. Reiser H, Coligan J, Palmer E, Benacerraf B, Rock KL: **Cloning and expression of a cDNA for the T-cell-activating protein TAP.** *Proc Natl Acad Sci USA* 1988, **85:**2255-2259.
32. Spangrude GJ, Brooks DM: **Mouse strain variability in the expression of the hematopoietic stem cell antigen Ly6A/E by bone marrow cells,** *Blood* 1993, **82:**3327-3332.
33. Ravetch J, Kinet JP: **Fc Receptors.** *Annu Rev Imm* 1991, **9:**457-492.
34. Takizawa F, Adamczewski M, Kinet JP: **Identification of low affinity receptor for immunoglobulin E on mouse mast cells and macrophages as FcγRII and FcγRIII.** *J Exp Med* 1992,**176:**469-475.
35. Lantz CS, Huff TF: **Murine KIT+ lineage bone marrow progenitors express FcγRII but do not express FcεRI until mast cell granule formation.** *J Immunol* 1995,**154:**355-362.
36. Erba HP, Eddy R, Shows T, Kedes L, Gunning P: **Structure, chromosome location, and expression of the human gamma-actin gene: differential evolution, location, and expression of the cytoskeletal beta and gamma-actin genes.** *Mol Cell Biol* 1988, **8:**1775-1789.
37. Tokunaga K, Takeda K, Kamiyama K, Kageyama H, Takenaga K, Sakiyama S: **Isolation of cDNA clones for mouse cytoskeletal gamma-actin and differential expression of cytoskeletal actin mRNAs in mouse cells.** *Mol Cell 8iol*1988, **8:**3929-3933.
38. Otey CA, Kalnoski MH, Bulinski JC: **Identification and quantification of actin isoforms in vertebrate cells and tissues.** *J Cell Biochem* 1987, **34:**113-124.
39. Hill MA, Gunning P: **Beta and gamma actin mRNAs are differentially located within myoblasts.** *J Cell Biol*1993,**122:**825-832.
40. Ross IL, Browne CM, Hume DA: **Transcription of individual genes in eukaryotic cells occurs randomly and infrequently.** *Immuno! Cell Biol* 1994, **72:**177-185.
41. Michaelson J: **Cellular selection in the genesis of multicellular organization.** *Lab Invest* 1993, **69:**136-151.
42. Liang P, Pardee AB: **Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction.** *Science* 1992, **257:**967-971.
43. Barnes WM: **PCR amplification of up to 35 kb DNA with high fidelity and high yield from lambda bacteriophage templates.** *Proc Nat1 Acad Sci USA* 1994, **91:**2216-2220.
44. Cumano A, Paige CJ, Iscove NN, Brady G: **Bipotential precursors of B cells and macrophages identified in murine fetal liver at day 12 of gestation.** *Nature* 1992, **356:**612-615,
45. Kitamura Y: **Heterogeneity of mast cells and phenotypic change between subpopulations.** *Annu Rev Immunol* 1989, **7:**59-76.
46. Williams N, Jackson H, Sheridan AP, Murphy MJ Jr, Elste A, Moore MA: **Regulation of megakaryopoiesis in long-term murine bone marrow cultures.** *Blood* 1978, **51:**245-255.
47. Brady G, Iscove NN: **Construction of cDNA libraries from single cells.** *Meth Enzymol* 1993, **225:**611-623.
48. Dudov KP, Perry RP: **The gene family encoding the mouse ribosomal protein L32 contains a uniquely expressed intron-containing gene and an unmutated processed gene.** *Cell* 1984, **37:**457-468.
49. Tso JY, Sun XH, Kao TH, Reece KS, Wu R: **Isolation and characterization of rat and human glyceraldehyde-3-phosphate dehydrogenase cDNAs: genomic complexity and molecular evolution of the gene.** *Nucleic Acids Res* 1985,**13:**2485-2502.
50. Chang KS, Trujillo JM, Cook RG, Stass SA: **Human myeloperoxidase gene: molecular cloning and expression in ieukemic cells.** *Blood* 1986, **68:**1411-1414,
51. Cross M, Mangelsdorf I, Wedel A, Renkawitz R: **Mouse lysozyme M gene: isolation, characterization, and expression studies.** *Proc Natl Acad Sci USA* 1988, **85:**623--626.
52. Leung S, Miyamato NG: **Point mutational analysis of the human c-fos serum response factor binding site.** *Nucl Acids Res* 1989,**17:**1171-1195.
53. Oettinger MA, Schatz DG, Gorka C, Baltimore D: ***Rag-1* and *Rag-2*, adjacent genes that synergistically activate V(D)J recombination.** *Science* 1990, **248:**1517-1523
54. Bothwell AL, Paskind M, Reth M, ImanishiKari T, Rajewsky K, Baltimore D: **Heavy chain variable region contribution to the NPb family of antibodies: somatic mutation evident in a gamma 2a variable region.** *Cell* 1981, **24:**625-637.
55. Marth JD, Peet R, Krebs EG, Perlmutter RM: **A lymphocyte-specific protein-tyrosine kinase gene is rearranged and overexpressed in the murine T cell lymphoma LSTRA.** *Cell* 1985, **43:**393-404,
56. Voronova AF, Sefton BM: **Expression of a new tyrosine protein kinase is stimulated by retrovirus promoter insertion.** *Nature* 1986, **319:**682-685.
57. Castle S, Sheiness D: **Structural organization of the mouse *protomyb* gene.** *Biochem Biophys Res Commun* 1985, **132:**688-695.
58. Begley CG, Visvader J, Green AR, Aplan PD, Metcalf D, Kirsch IR, Gough NM: **Molecular cloning and chromosomal localization of the murine homolog of the human helix-loop-helix gene *SCL*.** *Proc Natl Acad Sci USA* 1991, **88:**869-873.
59. Tsai SF, Martin DIK, Zon LI, D'Andrea AD; Wong GG, Orkin SH: **Cloning of cDNA for the major DNA-binding protein of the erythroid lineage through expression in mammalian cells.** *Nature* 1989, **339:**446-451.
60. D'Andrea AD, Lodish HF, Wong GG: **Expression cloning of the murine erythropoietin receptor.** *Cell* 1989, **572:**77-85.
61. Negrini M, Sillini E, Kozak C, Tsujimoto Y, Croce CM: **Molecular analysis of *mbcl2*: structure and expression of the murine gene homologous to the human gene involved in follicular lymphoma.** *Cell* 1987, **49:**455-463.
62. Feng GS, Hui CC, Pawson T: **SH2-containing phosphotyrosine phosphatase as a target of protein-tyrosine kinases.** *Science* 1993, **259:** 1607-1611.
63. Hogarth PM, Hibbs ML, Bonadonna L, Scott BM, Witort E, Pietersz GA, McKenzie IF: **The mouse Fc receptor for IgG (Ly17); molecular cloning and specificity.** *Immunogenetics* 1987, **26:**16-18.
64. Ponte P, Ng SY, Engel J, Gunning P, Kedes L: **Evolutionary conservation in the untranslated regions of actin mRNAs: DNA sequence of a human beta-actin cDNA.** *Nucleic Acids Res* 1984,**12:**1687-1696.
65. Erba HP, Gunning P, Kedes L: **Nucleotide sequence of the human γ cytoskeletal actin mRNA: anomolous evolution of vertebrate nonmuscle genes.** *Nucleic Acids Res* 1986,**14:**5275-5294.
66. Barr FG, Emanuel BS: **Application of a subtraction hybridization technique involving photoactivatable biotin and organic extraction to solution hybridization analysis of genomic DNA.** *Anal Biochem* 1990, **186:**369-373.

Received: **12 May 1995**; revised: **2 June 1995.**
Accepted: **7 June 1995**.