RNA analysis of B cell lines arrested at defined stages of differentiation allows for an approximation of gene expression patterns during B cell development

Panagiotis Tsapogas,* Thomas Breslin,† Sven Bilke,† Anna Lagergren,* Robert Månsson,* David Liberg,* Carsten Peterson,† and Mikael Sigvardsson*

Laboratory for Cellular Differentiation, Department for Stemcell Biology, and †Department of Theoretical Physics, Lund University, Sweden

Abstract: The development of a mature B lymphocyte from a bone marrow stem cell is a highly ordered process involving stages with defined features and gene expression patterns. To obtain a deeper understanding of the molecular genetics of this process, we have performed RNA expression analysis of a set of mouse B lineage cell lines representing defined stages of B cell development using Affymetrix™ microarrays. The cells were grouped based on their previously defined phenotypic features, and a gene expression pattern for each group of cell lines was established. The data indicated that the cell lines representing a defined stage generally presented a high similarity in overall expression profiles. Numerous genes could be identified as expressed with a restricted pattern using dCHIP-based, quantitative comparisons or presence/absence-based, probabilistic state analysis. These experiments provide a model for gene expression during B cell development, and the correctly identified expression patterns of a number of control genes suggest that a series of cell lines can be useful tools in the elucidation of the molecular genetics of a complex differentiation process. *J. Leukoc. Biol.* 74: 102–110; 2003.

Key Words: *gene expression · immunoglobulin · progenitor B cells*

INTRODUCTION

B lymphocyte development is a highly ordered process proceeding from the progenitor cells in the bone marrow (BM) to the immunoglobulin (Ig)-secreting plasma cell in the spleen, gut, or BM [1, 2]. The early steps of this developmental pathway can be divided into distinct stages based on the recombination status of the Ig genes and on the expression pattern of surface markers and the presence of intracellular proteins [1–6]. The early progenitor B (pro-B) cells in the mouse have their Ig genes in a germ-line configuration and express the surface molecules B220 and AA4.1, the signal-transducing molecule Igβ (B29), and the α-subunit of the interleukin (IL)-7 receptor [5, 6]. Subsequent differentiation results in the expression of the recombination-activating genes *Rag-1* and *Rag-2* and initiation of Ig recombination events [7]. This generates a functional Ig heavy-chain (IgH) gene that is transcribed, translated, and displayed on the cell surface in complex with the surrogate light-chain components λ5 and VpreB, as well as the signal-transduction molecules Igα (*mb-1*) and Igβ (*B29*) [8]. Subsequent differentiation allows for rearrangements of the Ig light-chain (IgL) genes that replace the surrogate light-chain genes on the surface of the B cell [8]. This immature cell is then subjected to negative selection to delete self-reactive cells before it leaves the BM to enter peripheral lymphoid organs, where it becomes a mature B cell [9]. If the cells are activated by interaction with antigens and obtain T cell help, they mature into terminally differentiated plasma cells secreting large amounts of antibodies [10–12]. The extensively studied biology of B cell development, in combination with the defined stages of differentiation, makes it a useful system for investigations of complex molecular events that might provide clues to general features of cellular development.

To obtain a deeper understanding of the molecular processes involved in B cell development and to create a map over stage-restricted gene expression, we wanted to establish a model system that allowed for a reasonable approximation of the gene expression profile during B lymphoid differentiation. Features such as varying proliferation status, heterogeneous populations, and difficulties to obtain sufficient amounts of material limit the use of primary sorted cells. The existence of numerous B cell lines arrested at defined differentiation stages would then pose a possibility to overcome some of these problems. They also provide a highly reproducible source of material that allows for the performance of large-scale gene expression analysis without the use of intermediate amplification steps. The use of cell lines does, however, introduce a risk of obtaining cell line-specific features as a result of the transformation process. To reduce the risk of analyzing cell line-specific features, we used several representative cell lines for each of four major stages in B cell development: pro-B, pre-B,

Correspondence: Mikael Sigvardsson, Laboratory for Cellular Differentiation, Department for Stemcell Biology, BMC B12, Lund, 22184, Sweden. E-mail: Mikael.Sigvardsson@stemcell.lu.se

Received January 9, 2003; revised March 6, 2003; accepted March 24, 2003; doi: 10.1189/jlb.0103008.

B, and plasma cells, and investigated the gene expression pattern in these cell lines by Affymetrix™ microarrays containing ~12,000 gene tags. This allowed for the correct classification of a large number of control genes using dCHIP-based, relative expression level analysis [13] or presence/absence (P/A)-based, probabilistic state analysis (S. Bilke et al., submitted). We also identified a large number of additional genes that now can be considered as candidates to display stage-restricted expression patterns during B cell development.

MATERIALS AND METHODS

Tissue-culture conditions

All cells were grown at 37°C and 5% CO_2 in RPMI supplemented with 7.5% fetal calf serum, 10 mM HEPES, 2 mM pyruvate, 50 μM 2-mercaptoethanol, and 50 μg/ml gentamycin (all purchased from Life Technologies AB, Täby, Sweden). The pro-B cell lines were grown in RPMI as above, supplemented with 10% of IL-3 containing WEHI3-conditioned media. The Ba/F3 subclones were kind gifts from Drs. Rudolf Grosschedl (Gene Center, Munich), Ramiro Gisler (Department for Stemcell Biology, Lund University), and Johan Forssel (Department of Cell and Molecular Biology, Umeå University), and the Ly9D cells were a gift from Dr. Meinrad Busslinger (IMP, Wienna). The pre-B cell line 18–81 was a gift from Dr. Inge-Lill Mårtensson (The Babraham Institute, Cambridge), and all the other cell lines were gifts from Dr. Thomas Leanderson (Department of Immunology, Lund University).

Gene expression analysis

RNA was prepared using Trizol (Gibco, Grand Island, NY), and 7.5 μg of total RNA was annealed to a T7-oligo T primer by denaturation at 70°C for 10 min followed by 10 min of incubation of the samples on ice. First-strand synthesis was performed for 2 h at 42°C using 20 U Superscript reverse transcriptase (RT; Gibco) in buffers and nucleotide mixes according to the manufacturer's instructions. This was followed by a second-strand synthesis for 2 h at 16°C, using RNaseH, *Escherichia coli* DNA polymerase I, and *E. coli* DNA ligase (all from Gibco), according to the manufacturer's instructions. The obtained, double-stranded cDNA was then blunted by the addition of 20 U T4 DNA polymerase and incubated for 5 min at 16°C. The material was then purified by phenol:cloroform:isoamyl alcohol extraction followed by precipitation with NH_4Ac and ethanol. The cDNA was then used in an in vitro transcription reaction for 6 h at 37°C using a T7 IVT kit and biotin-labeled ribonucleotides. The obtained cRNA was purified from unincorporated nucleotides on an RNAeasy column (Qiagen, Valencia, CA). The eluted cRNA was then fragmented by incubation of the products for 2 h in fragmentation buffer (40 mM Tris-acetate, pH 8.1, 100 mM KOAc, 150 mM MgOAc). The final, fragmented cRNA (20 μg) was hybridized to Affymetrix™ chip U74Av2 (Affymetrix, Santa Clara, CA) in 200 μl hybridization buffer [100 mM 2-(N-morpholino)-ethanesulfonic buffer, pH 6.6, 1 M NaCl, 20 mM EDTA, 0.01% Tween 20], supplemented with Herring sperm DNA (100 μg/ml) and acetylated bovine serum albumin (500 μg/ml) in an Affymetrix Gene Chip Hybridization Oven 320. The chip was then developed by the addition of fluorescein isothiocyanate (FITC)-streptavidin followed by washing using an Affymetrix Gene Chip Fluidics Station 400. Scanning was performed using a Hewlett Packard Gene Array scanner.

Data analysis

Probabilistic estimation of gene expression pattern was performed using the Breslin/Bilke method (S. Bilke et al., submitted). Hierarchical tree clusters were generated using the dCHIP program (Li and Wong [13], <http://biosun1.harvard.edu/complab/dchip/>). The initial analysis (see Fig. 2) was performed using the perfect match (PM)-only model, and genes were filtered according to 0.50 <standard deviation/mean between <10.00, and P call in the array used ≥20%. The pro-B, pre-B, B, and plasma cell lines (see Fig. 4 and Supplemental Figs. 1–4 available at http://www.jleukbio.org/) were treated as replicates using the same model as above but a P call above 10%. Classification of genes with an apparent restricted expression pattern in the P/A analysis into functional groups was performed manually using the National Center for Biotechnology Information (NCBI; National Institutes of Health, Bethesda, MD) database (see supplementary tables). Nondefined genes in the data set were resubmitted in a Blast search into Genebank allowing for the identification of most of these entries.

RT and polymerase chain reactions (PCRs)

RNA was prepared from cells using Trizol (Life Technology), and cDNA was generated by annealing 1 μg total RNA to 0.5 μg random hexamers in 10 μl diethylpyrocarbonate-treated water. RT reactions were performed with 200 U Superscript RT (Life Technologies) in the manufacturer's buffer supplemented with 0.5 mM dNTP, 10 mM dithiothreitol, and 20 U RNase inhibitor (Boehringer Mannheim, Bromma, Sweden) in a total volume of 20 μl at 37°C for 1 h. One-twentieth of the RT reactions were used in the PCR assays. PCR reactions were performed with 1 U Taq-polymerase (Life Technologies) in the manufacturer's buffer, supplemented with 0.2 mM dNTP in a total volume of 25 μl. Primers were added to a final concentration of 1 mM.

For all PCRs, the program was the same except for the number of cycles (Y) and the annealing temperature (XX). The common parts of the program were (Y cycles) 95°C for 2 min, 95°C for 45 s, XX°C for 45 s, 72°C for 1 min, and 72°C for 2 min. Annealing temperatures–cycles were: actin, 57ºC, 25 cycles; hypoxanthine guanine phosphoribosyl transferase (HPRT), 52ºC, 25 cycles; λ5, 60ºC, 30 cycles; Pax-5, 60ºC, 30 cycles; J-chain, 60ºC, 28 cycles; Bach1, 55°C, 30 cycles; RhoB, 56°C, 30 cycles; Yes, 55°C, 30 cycles; Sel1, 55°C, 25 cycles; Arhgef, 55°C, 27 cycles; protein kinase C (PKC)-β, 55°C, 28 cycles; and Pftaire-1, 55°C, 30 cycles.

Oligonucleotides used for RT-PCR were: actin, sense 5′GTTTGAGACCTTCAACACC, antisense 5′GTGGCCATCTCCTGCTCGAAGTC; B29, sense 5′GGTGAGCCGGTACCAGCAATG, antisense 5′AGTTCCGTGCCACAGCTGTCG; λ5, sense 5′TGTGAAGTTCTCCTCCTGCTC, antisense 5′ACCACCAAAGTACCTGGGTAG; Pax-5, sense 5′CTACAGGCTCCGTGACGCAG, antisense 5′GTCTCGGCCTGTGAAATAGG; Bach-1, sense 5′ACTCTCAGTTCCGTCAACTGC, antisense 5′TTCCTCTTGCGACAGCGTTGC; Arhgef3 (EST-1), sense 5′AAACATCCGTCCACTCTCTTCC, antisense 5′TACTGTACACATGGGTCATGTGC; Pftaire-1, sense 5′TGCTCTAGCATACATTGAACC, antisense 5′CTCCCCACTTAAAGAACTCC; PKC-β-II, sense 5′ATCCACCAGTCCTAACACC, antisense 5′AAGCAAGCATTTTCTCTCC; RhoB, sense 5′CTGATCGTGTTCAGTAAAGACGAATTCC, antisense 5′TTGTTGGCCACCAGGATGATGG; Yes-associated protein, sense 5′GCAGTTACAGATGGAGAAGGAG, antisense 5′TTGCATCTCCTTCCAGTGTGC; HPRT, sense 5′GCTGGTGAAAAGGACCTCT, antisense 5′CACAGGACTAGAACACCTGC; J-chain, sense 5′GTAGGTGGTACCTATACAATAACA, antisense 5′AGGGTAGCAAGAATCGGGGGTCAA.

Isolation and purification of BM progenitors and mature peripheral B cells

BM cells were sorted on a FACSVantage Cell Sorter (Becton Dickinson, San Jose, CA), equipped with a 488-nm argon ion (Coherent Enterprise II, Santa Clara, CA) and a 633-nm He-Ne (Model 127, Spectra-Physics, Mountain View, CA) laser. Antibodies used were B220 antigen-presenting cell, CD43 phycoerythrin (PE), IgM biotin (Streptavidin TRI), CD19 FITC, and CD138 (Syndecan–1) PE (all from PharMingen, San Diego, CA). The purity of all sorted cell populations is reproducible over 95%. To obtain activated and mature B cells, magnetic cell sorter-purified $B220^+$ spleen cells were incubated in 50 ng/ml lipopolysaccharide (LPS; Sigma Chemical Co., St. Louis, MO) at 37°C for 72 h.

RESULTS

B cell lines arrested at defined stages of development generally display gene expression profiles typical for the differentiation stage

To identify cell lines for our RNA analysis, we searched the literature and selected representatives based on previously published results. As representatives of the pro-B cell stage, we used Ly9D cells and three different subclones of the cell

line Ba/F3 [14, 15]. These are IL-3-dependent lines without Ig rearrangements but with expression of sterile Ig transcripts [14–16]. Ba/F3 cells also express low levels of the B lineage-restricted *mb-1* and *B29* genes [16] and have been differentiated into B cells in vivo [14]. The nitrous–urea-induced cell line 70Z/3 as well as the Abelson virus-transformed lines 230–238, 40EI, and 18–81 represented the pre-B cells. As representatives for the B cell stage, we used the cell lines WEHI231, M12, K46, and A20. The mouse myeloma cell lines J558, S194, MPCII, and SP2.0 represented the Ig-secreting plasma cell stage. To obtain some information concerning gene expression pattern in the cell lines as compared with primary sorted B lineage cells, we extracted RNA from $B220^+$, $CD43^+$, IgM^- (pro-B) cells, $B220^+$, $CD43^-$, IgM^-, $CD19^+$ (pre-B) BM cells, and $B220^+$ spleen cells (B cells). Plasma cells were obtained by sorting Syndecan-1^+, $B220^-$ BM cells or by 72 h LPS stimulation of $B220^+$ splenocytes. The RNA expression patterns in the different cell populations were then analyzed by RT-PCR experiments (**Fig. 1A**) using amplification of actin and HPRT message as a control for the quality of the cDNA. This suggested that the *B29* gene was expressed at all stages of development [17], and the surrogate light-chain gene λ*5* [18] was expressed mainly in pre-B cells. Some λ5 expression was also detected in the Syndecan-1^+ $B220^-$ BM cells representing primary plasma cells. This probably reflects contamination with pre-B cells rather than true expression of this gene at this late developmental stage, as no λ5 expression could be detected in the LPS-stimulated splenocytes. The expression of the transcription factor Pax-5 [19] was high in the pre-B and B cells, and it appeared to be lower in the plasma-cell populations. In contrast, the Ig-associated J-chain was expressed at a higher level in the plasma cells than in the other populations [20]. Analysis of the expression pattern of the same genes in a selection of cell lines (Fig. 1B) indicated that all these expressed the B lineage-restricted *B29* (Igβ) gene, and only the pre-B cell lines expressed λ*5* message. *Pax-5* was not expressed in the pro-B cell lines or in the plasma cell lines, and the pre-B and B cell lines expressed this transcription factor. Message encoding the J-chain was present in one of the pre-B cell lines (230–238) and in cell lines representing the mature or the plasma cell stage. This suggests that the cell lines display expression of B cell markers in patterns comparable with those observed in primary sorted cells.

To expand the analysis of the gene expression patterns in the different B lineage cell lines, we analyzed RNA from these on Affymetrix™ gene chip microarrays containing ~12,000 sequence tags. The data were then analyzed using the dCHIP program [13], allowing for the identification of differentially expressed genes in the cell line samples (**Fig. 2**). This analysis indicated that although there were differences in expression levels, the cell lines previously defined as belonging to a specific stage of development generally displayed a similar cluster of differentially expressed genes. The exception was the B cell line WEHI231, which appeared to group with the pre-B cells rather than with the B cells, possibly reflecting that this cell represents an immature BM-derived B cell rather than a

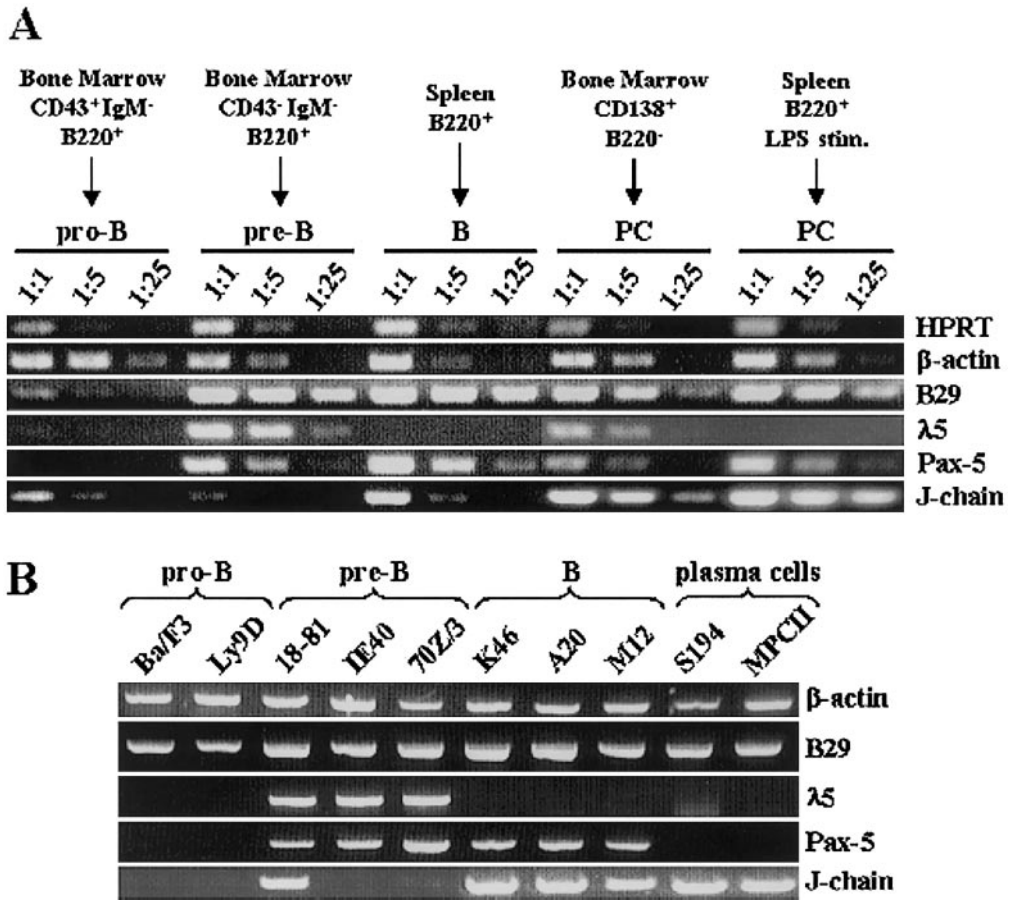
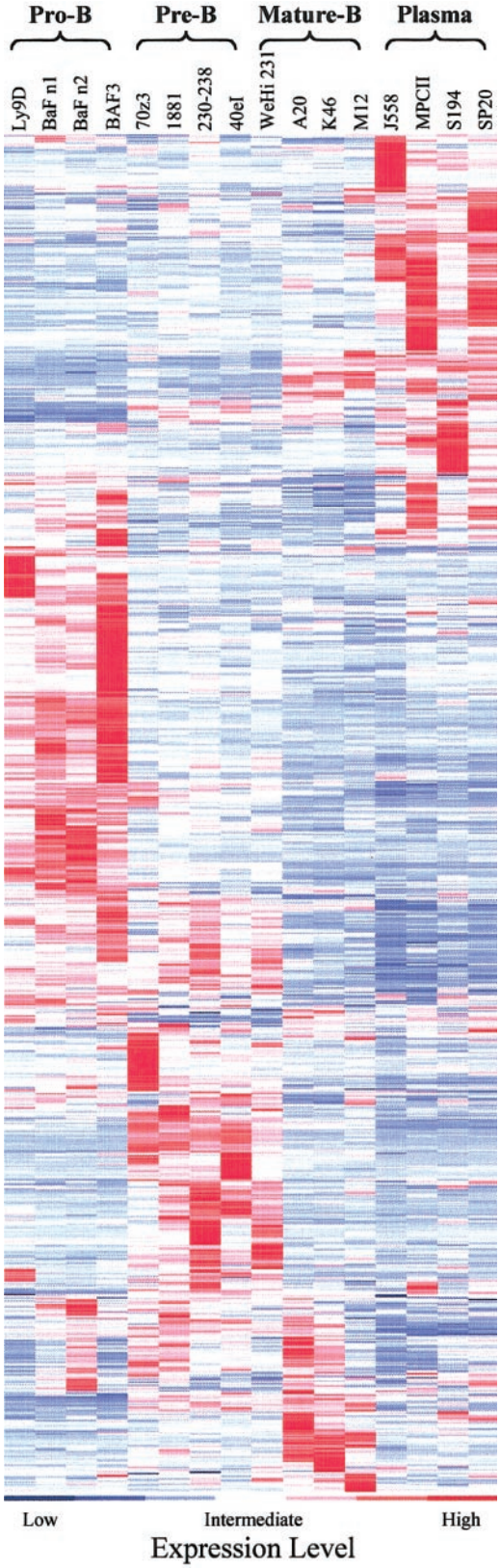


Fig. 1. Cell lines arrested in development express stage-specific genes in patterns resembling primary-sorted B lineage cells. (A) Ethidium bromide-stained agarose gels with PCR products obtained by RT-PCR analysis of primary-sorted cells as indicated. The cDNA was diluted in three steps to allow for a degree of quantification of the expressed transcripts. (B) A gel with a RT-PCR experiment displaying the expression of the same genes in a panel of cell lines as indicated.

mature peripheral cell [21, 22]. The homogeneity in expression levels within the cell line groups representing the different stages was also analyzed by the extraction of normalized expression levels for a set of genes linked to B cell development [1, 3–6] (**Fig. 3**). The expression of the transcription factor Id-1 was reduced upon progression into the pre-B cell stage, and the levels of the signal transducer B29 was increased and maintained in all the cell line groups. mRNA encoding λ5, V-preB1, VpreB-3, and the IL-7 receptor α subunit were all transiently up-regulated in the pre-B cell lines. Mb-1, EBF, and CD19 expression was high in the pre-B cells, but the mRNA could also be detected in the B cell lines. CD24a [human serum albumin (HSA)] was expressed in the pre-B and mature B cells as was PKC-β. The germinal center-specific transcription factor BCL-6 [23, 24] was expressed specifically in the B cell lines, and the Ig-associated J-chain was expressed in the B and the plasma cells. Only the latter cells expressed the plasma-cell transcription factor BLIMP [25]. Thus, although the standard deviations in some cases were substantial, these data indicate that the cell lines representing a specific developmental stage present rather homogenous gene expression patterns and support the idea that stage-specific genes can be identified using a series of B cell lines. To reduce the impact of cell line-specific features, we then treated all cell lines belonging to a certain differentiation stage as replicates and performed another dCHIP analysis. This resulted in the identification of a large number of genes, including several genes with previously defined expression patterns, as expressed in a stage-specific manner (**Fig. 4** and **Attachment Figs. 1–4**).



Affymetrix™ P/A analysis allows for the characterization of stage-restricted gene expression

The design of the Affymetrix™ microarrays, with one set of matching and one set of mismatching oligonucleotides for each gene, allows for a comparison of the obtained signals from the two probe sets. The data are then evaluated by Affymetrix™ array analysis software, allowing for the classification of all the studied genes as present (P) or absent (A) in each of the samples. This transforms the data set into binary values creating novel possibilities for mathematical analysis of the obtained data. This does, however, delete all information about relative expression levels; so to investigate if binary values could be used for the identification of stage-specific genes, we constructed a Hamming distance matrix based on P/A analysis. This generates a measure of similarity between any two samples based on the number of genes for which the Affymetrix™ P/A calls differ (**Fig. 5**). The pro-B and plasma cell groups were the most homogenous, and the B cell group displayed a poorer similarity. M12, K46, and A20 appeared similar, while the WEHI231 cells rather resembled the pre-B cells. This indicates that P/A analysis allows for the correct stage classi-

Fig. 2. dCHIP analysis of gene expression data suggests that B cell lines belonging to a certain differentiation stage generally display similar expression patterns of stage-specific genes. The figure displays a dCHIP analysis of the RNA expression patterns in the different cell lines selected for our investigations after filtering and hierarchical clustering. The name of the cell line and the earlier classification are displayed above the data panel. The criteria selected for the definition of differentially expressed genes were a 20% present count and a maximum standard deviation of 0.5 (0.50<standard deviation/mean between <10.00). Expression scales ranging from –3 (blue) to +3 (red) are indicated below the data display.

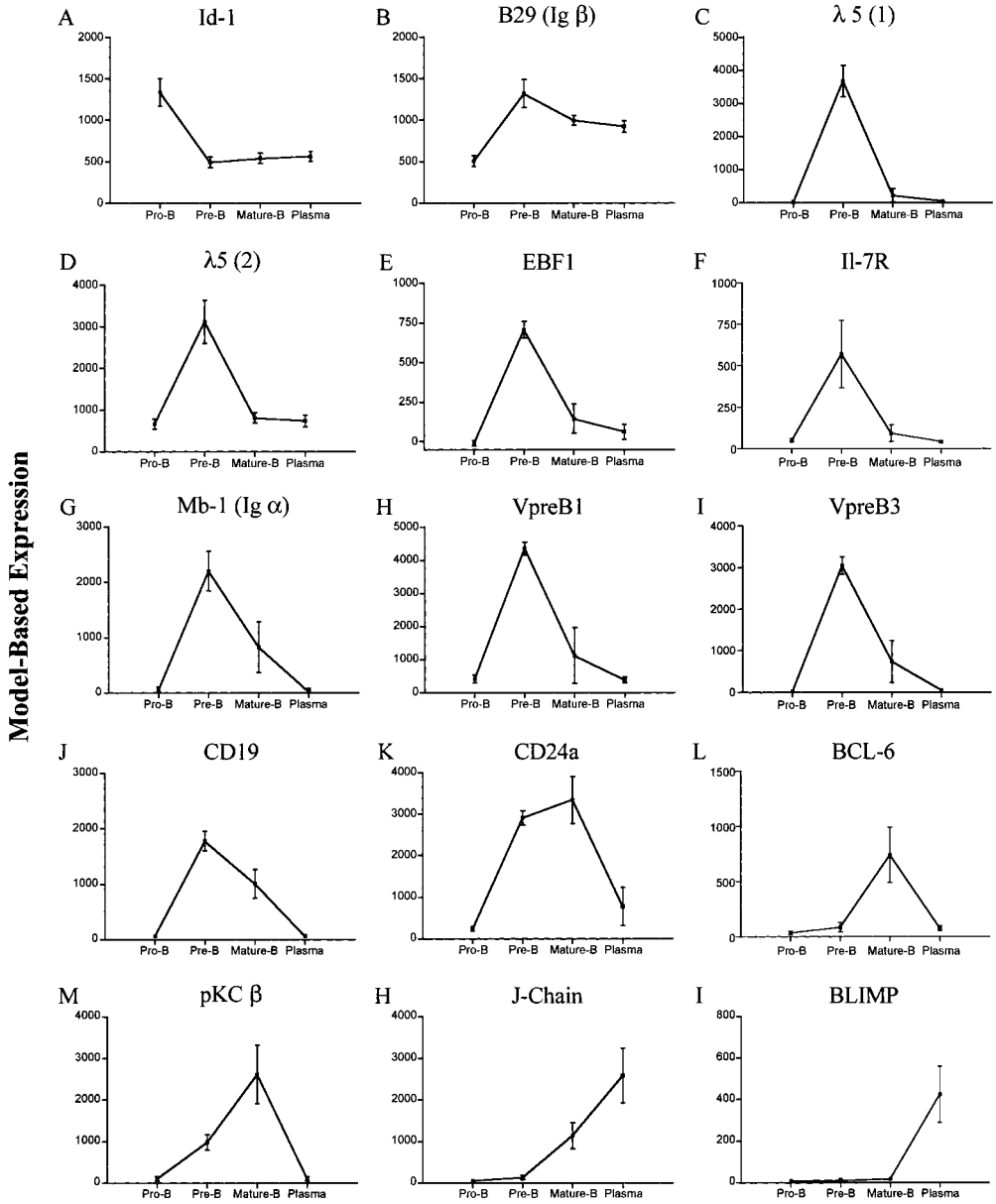


Fig. 3. Investigations of expression pattern of individual control genes indicate that RNA expression data from a group of cell lines can be used for the delineation of stage-specific genes. The figure displays diagrams over normalized expression levels of genes associated with B cell development as indicated on top of each display. The λ5 gene is represented by two probe sets. The standard deviations have been calculated by treatment of the cell lines belonging to a defined stage as replicates. Id-1, Inhibitor of DNA binding-1; EBF1, early B cell factor; BCL-6, B cell lymphoma-6; BLIMP, B lymphocyte-induced maturation protein-1.

fication of the cell lines and that this analysis method could be used to obtain information about stage-specific gene expression.

To extract the collected information from the P/A analysis, we applied the probabilistic estimation method (S. Bilke et al., submitted). Based on the expression of a given gene in the four samples representing each stage, this method yields the conditional probability as to whether the gene should be regarded as absent (A) mixed (M), i.e., expressed in some but not all the cell lines within the group, or present (P) at that specific stage. The analysis scheme further enables us to compute the probability of each possible expression profile over the four developmental stages, resulting in the fact that the most and second-to-most likely expression profile can be extracted. To investigate the feasibility of this analysis model, we extracted expression information from a number of genes with established expression patterns in B cell development (**Table 1**) and compared the result with that obtained with dCHIP analysis. Out of 37 genes, we found five that were not correctly classified as stage-specific from the P/A analysis and 10 that we did not detect as stage-restricted in the dCHIP analysis. The full P/A analysis is shown in **Supplementary Tables 1–7**, available online at http://www.jleukbio.org/. Thus, P/A-based probabilistic state analysis allowed for correct classification of several control genes, some of which could not be defined from the dCHIP analysis and vice versa, suggesting the two methods of data analysis to be complementary.

Predicted stage-restricted gene expression patterns can be verified by RT-PCR analysis

To validate the result of our data analysis, we randomly selected a number of genes suggested to be stage-specifically expressed and investigated their expression by RT-PCR analysis of a set of cell lines (**Fig. 6**). The transcription factor BACH-1, suggested from the P/A analysis to be specifically repressed at the B cell stage (supplementary tables), was

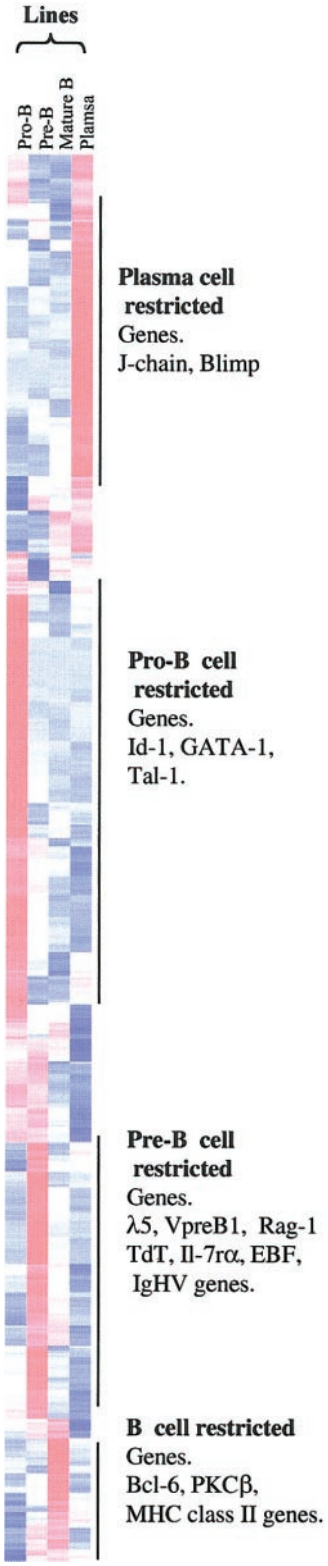


Fig. 4. Treatment of the data from different cell lines representing the same differentiation stage as replicates allows for the identification of stage-specific genes. The figure displays a dCHIP-generated cluster analysis of differential gene expression in the groups of cell lines after filtering and hierarchical clustering. The identification of control genes within the groups is indicated to the right of the color scheme. The full analysis with gene names of differentially expressed genes can be found as Attachment Figures 1–4. Expression scales ranging from –3 (blue) to +3 (red) are indicated below the data display. The figure only displays genes classified as present on more than one chip and to 0.50 < standard deviation/mean between <10.00. TdT, Terminal deoxynucleotidyl transferase; IgHV, Ig heavy-chain variable region; MCH, major histocompatibility complex.

expressed in pro-, pre-, and plasma cells but not in the B cell lines. The same data analysis indicated that the signal-transduction molecule Arhgef-3 was expressed specifically in pre-B cells, and although low levels of message could be detected in cell lines not belonging to the pre-B cell group, the expression appeared to be largely stage-restricted. A similar observation was made for the progenitor cell-restricted protein kinase PFTAIRE, which appeared to be expressed at high levels in the pro-B cells and only at a low level in one of the B cell lines. The P/A and the dCHIP analysis suggested another protein kinase, PKC-β, to be expressed mainly in the pre-B and B cells. mRNA encoding this protein could be detected in the pre-B and B cell lines and also in the Ly9D pro-B cells but not in BaF/3 or plasma cell lines. The dCHIP analysis suggested plasma cell-restricted expression of the signal-transduction molecules RhoB and Yes-associated protein, a finding confirmed by the RT-PCR analysis of the cell lines. This indicated that although specific discrepancies can be found, the overall picture of gene expression patterns using P/A or dCHIP analysis was well supported by the RT-PCR analysis.

DISCUSSION

A set of B cell lines can be used to define stage-specific genes

Here, we report a large-scale expression analysis of genes expressed in cell lines arrested at specific stages of B cell development. The calculated expression patterns of a large number of control genes, previously defined as expressed in the predicted pattern based on earlier experiments using cell line data and analysis of primary cells, indicate that analyzing B

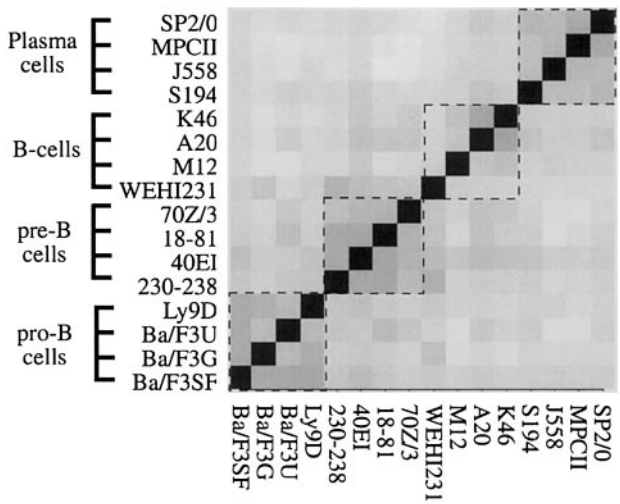


Fig. 5. P/A analysis allows for stage determination of B cell lines. The figure shows a distance matrix based on P/A analysis of the cell lines used for the generation of the data set. Ba/F3 1, 2, and 3 are differentially obtained subclones of the pro-B cell line Ba/F3, and Ly9D is an independently generated pro-B cell clone. 230–238, 40E1, and 18–81 are pre-B cell lines generated by Abelson virus transformation. 70/Z3 is a nitros–urea-induced pre-B cell line. WEHI231, A20, K46, and M12 are all defined as B cell lines, and S194, J558, SP2.0, and MPCII represent plasma cells. The scale ranges from highest similarity (black) to lowest (white).

TABLE 1. dCHIP and Probabilistic State Analysis Can Be Used As Complementary Methods for Gene Expression Analysis

Expression pattern	Probability	Secondary expression pattern	Probability 2	Gene name/reference	dCHIP classification
PAAA	0.934	MAAA	0.019	GATA-1 [26]	Pro-B
PAAA	0.934	MAAA	0.019	GATA-2 [26]	—
PAAA	0.934	MAAA	0.019	Id-1 [5]	Pro-B
APAA	0.790	APMA	0.157	Lef-1 [27]	—
APAA	0.934	AMAA	0.021	Sox-4 [28]	—
APAA	0.712	AMAA	0.243	Rag-1 [4]	Pre-B
APAA	0.790	APMA	0.157	λ5 [4]	Pre-B
APAA	0.790	APMA	0.157	VpreB [4]	Pre-B
AAPA	0.709	AAMA	0.246	SpiB [29]	—
APPA	0.924	APMA	0.024	CD19 [5, 30]	Pre-B-B
APPP	0.910	APPM	0.029	BOB-1 [31–33]	—
APPP	0.910	APPM	0.029	BLNK [34]	—
AMAA	0.955	AMAM	0.016	TdT [4]	Pre-B
APMM	0.922	APAM	0.024	EBF [35]	Pre-B
AAPM	0.702	AAMM	0.244	CD20 [36]	—
—	—	—	—	HSA [3]	Pre-B-B
APMA	0.922	APPA	0.027	Mb-1 [4]	Pre-B
—	—	—	—	Bcl-6 [23, 24]	B
—	—	—	—	Blimp-1 [34]	Plasma
—	—	—	—	J-chain [20]	Plasma
APMA	0.864	APAA	0.082	Pax-5 [19]	—
AMAA	0.931	AAAA	0.025	Rag-2 [4]	—
AAAM	0.539	AAAP	0.418	Syndecan [37]	—
—	—	—	—	Il-7rα [6]	Pre-B
APAA	0.934	AMAA	0.021	Clone BPS3.23 germline Ig variable region heavy chain precursor gene	Pre-B
APAA	0.934	AMAA	0.021	Immunoglobulin H-chain V-region pseudogene	Pre-B
APAA	0.934	AMAA	0.021	Clone BPS3.19 immunoglobulin heavy chain variable region precursor gene	Pre-B
APAA	0.934	AMAA	0.021	Germline immunoglobulin V(H)II gene H17	Pre-B
APAA	0.934	AMAA	0.021	Germline immunoglobulin V(H)II gene H8	Pre-B
APAA	0.934	AMAA	0.021	Clone BPS5.16 immunoglobulin heavy chain variable region precursor gene	Pre-B
APAA	0.934	AMAA	0.021	Ig B cell antigen receptor gene	Pre-B
APAA	0.790	APMA	0.157	Immunoglobulin heavy chain V DSP2.7-JH2 region (Igh) gene	Pre-B
APAA	0.790	APMA	0.157	Clone BHS2.19 immunoglobulin heavy chain variable region precursor gene	Pre-B
APAA	0.790	APMA	0.157	Immunoglobulin heavy and light chain variable region mRNA	Pre-B
APAA	0.790	APMA	0.157	Recombinant antineuraminidase single chain Ig VH and VL domains mRNA	Pre-B
APAA	0.790	APMA	0.157	Clone N1.1.b immunoglobulin heavy chain VDJ region gene	Pre-B
APAA	0.790	APMA	0.157	Immunoglobulin heavy chain gene, CDR3 region	Pre-B
APAA	0.712	AMAA	0.243	Germline immunoglobulin V(H)II gene H18	Pre-B
APAA	0.712	AMAA	0.243	Clone BPS3.26 immunoglobulin heavy chain variable region precursor, gene	Pre-B

Table 1 shows the calculated expression pattern of a set of control genes using dCHIP or P/A analysis. Genes were considered as present (P), absent (A), or mixed (M; S. Bilke et al., submitted) within the samples from each specific cell line group. The first classification goes for pro-B, the second to pre-B, the third to B, and the last for plasma-cell expression. For every gene, the second-highest probability to have another expression pattern is indicated as Probability 2. The gene name as well as reference to expression patterns are indicated in a separate column. The obtained information was compared with that resulting from a dCHIP analysis of the same data set using the same criteria as in Figure 4 (minimum present count 10% and 0.50<standard deviation/mean between <10.00).

cell development by the use of microarrays and cell lines arrested in defined stages results in reasonable approximation of gene expression patterns in B cell development. The analysis also suggests that cell lines defined as belonging to a specific developmental stage display a rather homogenous gene expression pattern. The group displaying the largest differences was the B cell group, possibly because these cell lines can arise at different anatomical sites and at different substages of differentiation. WEHI231 cells displayed an RNA expression profile closer related to the pre-B cells than to the B cells (Fig. 2). As a second, independently generated sample from this cell line gave the same result, we believe this might reflect that WEHI231 cells represent an immature BM-derived B cell. This idea is supported by reports suggesting that WEHI231 cells display defined features of immature B cells [21, 22]. The potential to obtain a degree of dynamic informa-

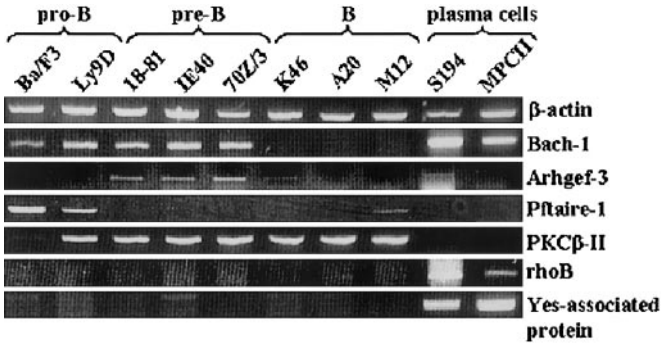


Fig. 6. Calculated gene expression patterns can be verified by RT-PCR analysis. The figure displays agarose gels, with the PCR products obtained using primers amplifying genes predicted to display restricted expression patterns (Attachment Figs. 1–4 and supplementary tables). The identities of the amplified mRNAs are indicated to the right, and the cell line used to generate cDNA is indicated on top of the panel. The PCR product has been visualized by ethidium bromide staining.

tion using stage-specifically arrested cell lines is also supported by analysis of the genes defined as expressed at more than one stage in the probabilistic state analysis. This is because the groups representing continuous expression patterns contained, on average, 17 genes (supplementary tables), and those representing discontinuous patterns contained, on average, six genes, a finding that indicates the existence of a flow of genetic information from the progenitor cell to the plasma cell. Our limited comparison of gene expression patterns in primary cells and cell lines indicates that the general feature of gene regulation is reasonably conserved in the transformed cells. This also suggests that the use of cell-surface markers on primary cells allows for a reasonable degree of enrichment of specific cell stages. This is also supported from microarray analysis of primary BM pre-B and B cells, where several of the same genes could be defined as stage-restricted [38, 39]. One major difference is lack of identification of cell-cycle genes that constitute a prominent group of genes when primary cells are used in expression analysis [38, 39]. This is probably explained by the fact that although only some of the primary cell stages are in cycle, all the cell lines are constitutively cycling, reducing the complexity of this part of the analysis.

Expression analysis of pre-B cell lines suggests simultaneous stage-restricted expression of several nonrearranged IgH genes

Another aspect of B cell development that does not become apparent when using sorted primary cell populations for gene expression analysis is reflected in the detection of RNA encoded by several V-region, heavy-chain (VH) genes, including pseudo-genes, specifically in the pre-B cell lines (Table 1). As the cell lines are of clonal origin, this could indicate that one and the same pre-B cell has the ability to express several VH genes simultaneously. These transcripts were in most cases not detected at the later stages of differentiation, and the mature cells, to a larger extent, expressed IgVL (V-region, light chain) genes. This is likely to reflect an ongoing rearrangement process of the heavy-chain gene in pre-B cell lines [40], with sterile expression of VH genes making them accessible for the recombination machinery [7, 41]. The expression of these V genes appears to be silenced at the later stages of development, possibly to ensure that no additional rearrangements of the heavy-chain genes occur during the assembly of the light-chain genes [8]. It may also be a mechanism contributing to allelic exclusion of the heavy chain to ensure that each single B cell only expresses one type of surface-bound Ig to avoid cross-reactive immune responses [8]. Thus, there may be a biological necessity in this rather complicated expression pattern, possibly demanding differential regulation of IgH promoters during B cell development. This type of information would not be extracted from the use of primary, sorted cells, as a broad expression of Ig genes could be explained by the heterogeneity of the sorted cell populations.

The Affymetrix™ P/A analysis is useful for the identification of stage-specific genes

Although the general picture of gene expression patterns was the same, independently of which method we used for the analysis of our data, the results from the (dCHIP) analysis differed to some extent from that obtained by P/A-based, probabilistic state analysis. As the dCHIP analysis takes into regard the relative transcription levels, there will be one group of genes that is expressed at all developmental stages but at different relative levels, which will be detected using dCHIP but not P/A analysis. These genes will be classified as present in all the groups and therefore not be detected as stage-specifically regulated in a P/A analysis. A bit more surprising was the detection of genes by the P/A-based method that we could not get classified in the dCHIP analysis. This is probably a result of that fact that rather small changes in relative expression values could change the classification from absent to present. Such an alteration might be classified as insignificant in the dCHIP analysis. This means that the P/A method gives a higher sensitivity of the data analysis, but at the same time, it will also increase the probability of detecting nonregulated genes. However, it appears that we in some cases detect different control genes using different analysis methods, indicating that the two approaches are complementary to each other.

We have not performed any extended analysis of the data we obtained as a result of the large amounts of information and the validity of the expression profile, as any individual gene needs further investigation. The analysis does, however, provide information that can be used to create a preliminary map of gene expression patterns that can be used to formulate working hypotheses for complex molecular events in B cell development.

ACKNOWLEDGMENTS

The Swedish Research Council, The Knut and Alice Wallenberg Foundation through the Swegene consortium, and The Swedish Cancer, Barncancer, Magnus Bergwall, Åke Wibergs, Österlunds, and The Crafoord foundations funded this work. P. T. and T. B. contributed equally to this work. We personally

thank the Swegen micro array facility in Lund as well as Drs. S. Cardell and Ramiro Gisler for critical reading of the manuscript.

REFERENCES

1. Ghia, P., ten Boekel, E., Rolink, A. G., Melchers, F. (1998) B-cell development: a comparison between mouse and man. *Immunol. Today* **19,** 480–485.
2. Jelinek, D. F. (2000) Regulation of B lymphocyte differentiation. *Ann. Allergy Asthma Immunol.* **84,** 375–385.
3. Hardy, R. R., Carmack, C. E., Shinton, S. A., Kemp, J. D., Hayakawa, K. (1991) Resolution and characterization of pro-B and pre-pro-B cell stages in normal mouse bone marrow. *J. Exp. Med.* **173,** 1213–1225.
4. Li, Y. S., Hayakawa, K., Hardy, R. R. (1993) The regulated expression of B lineage associated genes during B cell differentiation in bone marrow and fetal liver. *J. Exp. Med.* **178,** 951–960.
5. Li, Y. S., Wasserman, R., Hayakawa, K., Hardy, R. R. (1996) Identification of the earliest B lineage stage in mouse bone marrow. *Immunity* **5,** 527–535.
6. Tudor, K. S., Payne, K. J., Yamashita, Y., Kincade, P. W. (2000) Functional assessment of precursors from murine bone marrow suggests a sequence of early B lineage differentiation events. *Immunity* **12,** 335–345.
7. Grawunder, U., West, R. B., Lieber, M. R. (1998) Antigen receptor gene rearrangement. *Curr. Opin. Immunol.* **10,** 172–180.
8. Melchers, F., ten Boekel, E., Yamagami, T., Andersson, J., Rolink, A. (1999) The roles of preB and B cell receptors in the stepwise allelic exclusion of mouse IgH and L chain gene loci. *Semin. Immunol.* **11,** 307–317.
9. Nussenzweig, M. C. (1998) Immune receptor editing: revise and select. *Cell* **95,** 875–878.
10. Liu, Y. J., Arpin, C. (1997) Germinal center development. *Immunol. Rev.* **156,** 111–126.
11. Liu, Y. J., Banchereau, J. (1997) Regulation of B-cell commitment to plasma cells or to memory B cells. *Semin. Immunol.* **9,** 235–240.
12. MacLennan, I. C. M., Gulbranson-Judge, A., Toellner, K-M., Casamayor-Palleja, M., Chan, E., Sze, D. M-Y., Luther, S. A., Orbea, H. A. (1997) The changing preference of T and B cell partners as T-dependent antibody responses develop. *Immunol. Rev.* **156,** 53–66.
13. Li, C., Wong, W. (2001) Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proc. Natl. Acad. Sci. USA* **98,** 31–36.
14. Palacios, R., Steinmetz, M. (1985) Il-3-dependent mouse clones that express B-220 surface antigen, contain Ig genes in germ-line configuration, and generate B lymphocytes in vivo. *Cell* **41,** 727–734.
15. Palacios, R., Karasuyama, H., Rolink, A. (1987) Ly1+ PRO-B lymphocyte clones. Phenotype, growth requirements and differentiation in vitro and in vivo. *EMBO J.* **6,** 3687–3693.
16. Sigvardsson, M., O'Riordan, M., Grosschedl, R. (1997) EBF and E47 collaborate to induce expression of the endogenous immunoglobulin surrogate light chain genes. *Immunity* **7,** 25–36.
17. Hermanson, G. G., Eisenberg, D., Kincade, P. W., Wall, R. (1988) B29: a member of the immunoglobulin gene superfamily exclusively expressed on beta-lineage cells. *Proc. Natl. Acad. Sci. USA* **85,** 6890–6894.
18. Sakaguchi, N., Melchers, F. (1986) Lambda 5, a new light-chain-related locus selectively expressed in pre-B lymphocytes. *Nature* **324,** 579–582.
19. Adams, B., Dorfler, P., Aguzzi, A., Kozmik, Z., Urbanek, P., Maurer-Fogy, I., Busslinger, M. (1992) Pax-5 encodes the transcription factor BSAP and is expressed in B lymphocytes, the developing CNS, and adult testis. *Genes Dev.* **6,** 1589–1607.
20. Lamson, G., Koshland, M. E. (1984) Changes in J chain and mu chain RNA expression as a function of B cell differentiation. *J. Exp. Med.* **160,** 877–892.
21. Monroe, J. G., Seyfert, V. L., Owen, C. S., Sykes, N. (1989) Isolation and characterization of a B lymphocyte mutant with altered signal transduction through its antigen receptor. *J. Exp. Med.* **169,** 1059–1070.
22. Yellen, A. J., Glenn, W., Sukhatme, V. P., Cao, X. M., Monroe, J. G. (1991) Signaling through surface IgM in tolerance-susceptible immature murine B lymphocytes. Developmentally regulated differences in transmembrane signaling in splenic B cells from adult and neonatal mice. *J. Immunol.* **146,** 1446–1454.
23. Cattoretti, G., Chang, C. C., Cechova, K., Zhang, J., Ye, B. H., Falini, B., Louie, D. C., Offit, K., Chaganti, R. S., Dalla-Favera, R. (1995) BCL-6 protein is expressed in germinal-center B cells. *Blood* **86,** 45–53.
24. Falini, B., Bigerna, B., Pasqualucci, L., Fizzotti, M., Martelli, M. F., Pileri, S., Pinto, A., Carbone, A., Venturi, S., Pacini, R., Cattoretti, G., Pescarmona, E., Lo Coco, F., Pelicci, P. G., Anagnastopoulos, I., Dalla-Favera, R., Flenghi, L. (1996) Distinctive expression pattern of the BCL-6 protein in nodular lymphocyte predominance Hodgkin's disease. *Blood* **87,** 465–471.
25. Turner, C. A., Mack, D. H., Davis, M. M. (1994) Blimp-1, a novel zinc finger-containing protein that can drive the maturation of B lymphocytes into immunoglobulin-secreting cells. *Cell* **77,** 297–306.
26. Weiss, M. J., Orkin, S. H. (1995) GATA transcription factors: key regulators of hematopoiesis. *Exp. Hematol.* **23,** 99–107.
27. Travis, A., Amsterdam, A., Belanger, C., Grosschedl, R. (1991) LEF-1, a gene encoding a lymphoid-specific protein with an HMG domain, regulates T-cell receptor α enhancer function. *Genes Dev.* **5,** 880–894.
28. van de Wetering, M., Oosterwegel, M., van Norren, K., Clevers, H. (1993) Sox-4, an Sry-like HMG box protein, is a transcriptional activator in lymphocytes. *EMBO J.* **12,** 3847–3854.
29. Su, G. H., Ip, H. S., Cobb, B. S., Lu, M-M., Chen, H-M., Simon, M. C. (1996) The Ets protein Spi-B is expressed exclusively in B cells and T cells during development. *J. Exp. Med.* **184,** 203–214.
30. Kozmik, Z., Wang, S., Dorfler, P., Adams, B., Busslinger, M. (1992) The promoter of the CD19 gene is a target for the B-cell-specific transcription factor BSAP. *Mol. Cell. Biol.* **12,** 2662–2672.
31. Luo, Y., Roeder, R. G. (1995) Cloning, functional characterization, and mechanism of action of the B-cell-specific transcriptional coactivator OCA-B. *Mol. Cell. Biol.* **15,** 4115–4124.
32. Strubin, M., Newell, J. W., Matthias, P. (1995) OBF-1, a novel B cell-specific coactivator that stimulates immunoglobulin promoter activity through association with octamer-binding proteins. *Cell* **80,** 497–506.
33. Gstaiger, M., Knoepfel, L., Georgiev, O., Schaffner, W., Hovens, C. M. (1995) A B-cell coactivator of octamer-binding transcription factors. *Nature* **373,** 360–362.
34. Fu, C., Turck, C. W., Kurosaki, T., Chan, A. C. (1998) BLNK: a central linker protein in B cell activation. *Immunity* **9,** 93–103.
35. Hagman, J., Belanger, C., Travis, A., Turck, C. W., Grosschedl, R. (1993) Cloning and functional characterization of early B-cell factor, a regulator of lymphocyte-specific gene expression. *Genes Dev.* **7,** 760–773.
36. Tedder, T. F., Klejman, G., Disteche, C. M., Adler, D. A., Schlossman, S. F., Saito, H. (1988) Cloning of a complementary DNA encoding a new mouse B lymphocyte differentiation antigen, homologous to the human B1 (CD20) antigen, and localization of the gene to chromosome 19. *J. Immunol.* **141,** 4388–4394.
37. Sanderson, R., Lalor, P., Bernfield, M. (1989) B lymphocytes express and lose syndecan at specific stages of differentiation. *Cell Regul.* **1,** 27–35.
38. Hoffman, R., Seidl, T., Neeb, M., Rolink, A., Melchers, F. (2002) Changes in gene expression profiles in developing B cells of murine bone marrow. *Genome Res.* **12,** 98–111.
39. Hoffman, R., Bruno, L., Siedl, T., Rolink, A., Melchers, F. (2003) Rules for gene usage inferred from a comparison of large-scale gene expression profiles of T and B lymphocyte development. *J. Immunol.* **170,** 1339–1353.
40. Schlissel, M. S., Corcoran, L. M., Baltimore, D. (1991) Virus-transformed pre-B cells show ordered activation but not inactivation of immunoglobulin gene rearrangement and transcription. *J. Exp. Med.* **173,** 711–720.
41. Yancopoulos, G. D., Alt, F. W. (1985) Developmentally controlled and tissue-specific expression of unrearranged VH gene segments. *Cell* **40,** 271–281.