Characterization of Early Stages of Human B Cell Development by Gene Expression Profiling


*J Immunol* 2007; 179:3662-3671; doi: 10.4049/jimmunol.179.6.3662

http://www.jimmunol.org/content/179/6/3662

Supplementary Material

http://www.jimmunol.org/content/suppl/2008/03/12/179.6.3662.DC1

Why *The JI*?

- **Rapid Reviews!** 30 days* from submission to initial decision
- **No Triage!** Every submission reviewed by practicing scientists
- **Speedy Publication!** 4 weeks from acceptance to publication

*average

References

This article cites 59 articles, 29 of which you can access for free at:

http://www.jimmunol.org/content/179/6/3662.full#ref-list-1

Subscription

Information about subscribing to *The Journal of Immunology* is online at:

http://jimmunol.org/subscription

Permissions

Submit copyright permission requests at:

http://www.aai.org/About/Publications/JI/copyright.html

Email Alerts

Receive free email-alerts when new articles cite this article. Sign up at:

http://jimmunol.org/alerts
Characterization of Early Stages of Human B Cell Development by Gene Expression Profiling

Marit E. Hystad,* June H. Myklebust,*‡ Trond H. Bø,† Einar A. Sivertsen,* Edith Rian,* Lise Forfang,* Else Munthe,‡ Andreas Rosenwald,§ Michael Chiorazzi,* Inge Jonassen,*# Louis M. Staudt,¶ and Erlend B. Smeland*||

We have characterized several stages of normal human B cell development in adult bone marrow by gene expression profiling of hematopoietic stem cells, early B (E-B), pro-B, pre-B, and immature B cells, using RNA amplification and Lymphochip cDNA microarrays (n = 6). Hierarchical clustering of 758 differentially expressed genes clearly separated the five populations. We used gene sets to investigate the functional assignment of the differentially expressed genes. Genes involved in VDJ recombination as well as B lineage-associated transcription factors (TCF3 (E2A), EBF, BCL11A, and PAX5) were turned on in E-B cells, before acquisition of CD19. Several transcription factors with unknown roles in B lymphoid cells demonstrated interesting expression patterns, including ZCCHC7 and ZHX2. Compared with hematopoietic stem cells and pro-B cells, E-B cells had increased expression of 18 genes, and these included IGJ, ILIRAP, BCL2, and CD62L. In addition, E-B cells expressed T/NK lineage and myeloid-associated genes including CD2, NOTCH1, CD99, PECAM1, TNFSF13B, and MPO. Expression of key genes was confirmed at the protein level by FACS analysis. Several of these Ags were heterogeneously expressed, providing a basis for further subdivision of E-B cells. Altogether, these results provide new information regarding expression of genes in early stages of human B cell development. The Journal of Immunology, 2007, 179: 3662–3671.

Early human B cell development occurs in the bone marrow (BM) as hematopoietic stem cells (HSC) develop via various B lineage-restricted precursors into immature B (IM-B) cells, which then leave the BM and enter the periphery. The different stages can be identified by the expression of CD Ags and the rearrangement status of Ig H and L chains, and the current consensus is that B lineage-committed cells pass through a CD34+CD10+CD19− common lymphoid progenitors (CLP) early B (E-B) stage before they mature via CD34+CD19+CD10− pro-B, CD34+CD19+ large pro-B I and II, and CD34−CD19− small pre-B II into CD34−CD19−sIgM− IM-B cells (1, 2). Galy et al. (3) were the first to identify human CLP by demonstrating that CD34LinCD10+ cells had B, T, NK, and dendritic cell potential. Later studies have indicated that this population is biased toward B lineage development and termed the population E-B cells (4, 5). Rearrangements of the VDJ H chain locus are characteristic of pro-B cells and require the expression of the lymphocyte-specific recombination enzymes RAG1, RAG2, and TdT (6). The expression of pre-BCR, composed of IgH chains and surrogate L chains (VpreB and lμ1.1), is a hallmark for the pre-B cell population. Signaling through pre-BCR promotes L chain (VJL) rearrangement and allelic exclusion at the Ig H chain locus. Once VJL rearrangements are successful, L chains are expressed and combine with H chains as well as Igα/Igβ to form a functional BCR expressed on IM-B cells.

Early B cell development is controlled by a hierarchical regulatory network of transcription factors including PU.1, E2A, EBF, and Pax5 (7). E2A and EBF are required for the initiation of E-B cell development by regulating the expression of B lineage-specific genes, including PAX5, which is essential for B cell commitment maintenance (8, 9). Furthermore, development of lymphoid cells is supported by cytokines and interactions with stromal cells, as demonstrated by in vitro studies of human B cell development (10).

Genome-wide gene expression profiling has been performed in murine precursor B cells (11, 12). Recently, van Zelm et al. (13) investigated the Ig gene rearrangement steps in correlation to the transcription factor expression of developing B cells from the BM of healthy children. However, whereas the later stages of human B lineage development are well-characterized, the transitions from HSC into B lineage-committed cells have been explored in less detail. The aim of this study was to characterize the earliest steps of human B cell development in adult BM by gene expression profiling. Total RNA was extracted from HSC, E-B, pro-B, pre-B, and IM-B cell populations from human BM, amplified, and hybridized to Lymphochip cDNA microarrays. Altogether, this study provides new insight into the molecular processes that take place in early human B cell differentiation.

Materials and Methods

Cell separation and FACS analysis

BM aspirates from healthy adult donors were obtained after informed consent and approval by the Ethics Committee of the Norwegian Radium...
Hospital. CD34<sup>+</sup> or CD10<sup>+</sup> cells were isolated from BM by the use of immunomagnetic beads (CD34 Dynabeads, or sheep anti-mouse IgG-coated Dynabeads) labeled with anti-CD10 mAb; Invitrogen Life Technologies) as previously described (14, 15). CD38<sup>+</sup> CD43<sup>+</sup> HSC and CD38<sup>+</sup> CD43<sup>+</sup> HSC<sup>+</sup> CD19<sup>+</sup> E-B cells were obtained by FACS sorting on a FACSDiVa cell sorter (BD Biosciences) after labeling of CD34<sup>+</sup> BM cells with anti-CD38 FITC, anti-CD34 allophycocyanin, anti-CD10 PE/Cy7, and anti-CD19 PE/Cy5. CD34<sup>+</sup> CD10<sup>+</sup> CD19<sup>+</sup> IgM<sup>+</sup> pro-B, CD34<sup>+</sup> CD10<sup>+</sup> CD19<sup>+</sup> IgM<sup>-</sup> pre-B and CD34<sup>+</sup> CD10<sup>+</sup> CD19<sup>+</sup> IgM<sup>-</sup> IM-B cells were obtained by FACS sorting after labeling of CD10<sup>+</sup> BM cells with anti-CD10 PE/Cy7, anti-CD19 PE/Cy5, anti-IgM FITC, and anti-CD34 allophycocyanin (details in Table I). All cell populations were collected in 5-ml tubes, washed once in PBS, and subjected to RNA isolation. Detection of intracellular proteins was performed after labeling CD34<sup>+</sup> BM cells with Abs used to identify the various subpopulations, followed by fixation and permeabilization and subsequent staining with Abs (anti-Bcl-2, anti-myeloperoxidase (MPO), anti-J chain, or anti-TdT; Table I). The labeled cells were analyzed in LSR II (BD Biosciences) using FlowJo software (www.flowjo.com).

RNA extraction, amplification, labeling, and hybridization

Each cell population was isolated from six donors and analyzed independently. Total RNA was extracted from 1 × 10<sup>6</sup>–8 × 10<sup>6</sup> cells from the CD34<sup>+</sup> BM microprep kit (Stratagene), and subjected to two cycles of in vitro transcription according to a slightly modified protocol from Baugh et al. (16), previously described by Nygaard et al. (17). The quality of the rRNA was determined by the RNA Microprep kit (dual (A)) and compared to the significant probe sets, for probe sets annotated with GO terms representing genes associated with the variable lineages B, T, NK, dendritic cell, and myeloid, or genes associated with apoptosis, adhesion, regulation of transcription, and signaling. In the cell lineage marker gene sets, we also included selected genes (supplementary table I) (18).

Gene pairs for discrimination of HSC, E-B, and pro-B stages, and E-B gene expression profile

Seventy-two genes were selected because they characterized differences between the three first stages (one stage vs rest or one stage vs next stage), had at least a 2-fold change, and were coding for membrane proteins (selected by GO terms) or because they were of particular interest. All pairs of these 72 genes were examined to determine whether the expression levels of some of these gene pairs discriminated between the three stages. To test whether sample groups could be separated given the expression of two genes, we used the scoring function:

\[ \sum_{i,j} \frac{(x_i - \bar{x}_j)^2}{\sigma_i \sigma_j} \]

where \(x_i\) and \(x_j\) are the expression levels of gene A and gene B, respectively, in sample i and sample j, respectively; \(\bar{x}_i\) and \(\bar{x}_j\) are the average expression levels of gene A and gene B, respectively, in all samples; and \(\sigma_i\) and \(\sigma_j\) are the standard deviations of gene A and gene B, respectively, in all samples. The online version of this article contains supplemental material.

Table I. List of Abs and fixation/permeabilization methods used in FACS analysis and in cell-sorting experiments

<table>
<thead>
<tr>
<th>Ab</th>
<th>Company (Prod. No.)</th>
<th>Used at Concentration (µg/ml)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IgM FITC</td>
<td>DakoCytomation (F0203)</td>
<td>1/40 dilution</td>
</tr>
<tr>
<td>CD10 PE-Cy7</td>
<td>BD Biosciences (341102)</td>
<td>0.50</td>
</tr>
<tr>
<td>CD10 aliphycocyanin</td>
<td>BD Biosciences (332777)</td>
<td>0.60</td>
</tr>
<tr>
<td>CD34 aliphycocyanin</td>
<td>BD Biosciences (345818)</td>
<td>2.50</td>
</tr>
<tr>
<td>CD34 PE</td>
<td>BD Biosciences (345802)</td>
<td>2.50</td>
</tr>
<tr>
<td>CD38 FITC</td>
<td>Beckman Coulter (A07778)</td>
<td>0.30</td>
</tr>
<tr>
<td>CD38 PC5</td>
<td>Beckman Coulter (A07780)</td>
<td>0.80</td>
</tr>
<tr>
<td>CD38 PE</td>
<td>DakoCytomation (R7144)</td>
<td>2.50</td>
</tr>
<tr>
<td>CD19 PC5</td>
<td>Beckman Coulter(A07771)</td>
<td>1.25</td>
</tr>
<tr>
<td>CXCR4 PE</td>
<td>BD Biosciences (559794)</td>
<td>1/10 dilution</td>
</tr>
<tr>
<td>CD62L</td>
<td>BD Biosciences (341012)</td>
<td>1.00</td>
</tr>
<tr>
<td>CD2 FITC</td>
<td>DakoCytomation (F9767)</td>
<td>2.50</td>
</tr>
<tr>
<td>CD127 PE</td>
<td>Immunotech (PN IM1980)</td>
<td>1/5 dilution</td>
</tr>
</tbody>
</table>

* The cells were fixed in 4% paraformaldehyde for 10 minutes at room temperature, washed once in PBS, and then permeabilized in 0.1% Triton-X 100 for 30 min at room temperature, washed once in PBS, and incubated with 50% methanol for 10 min at room temperature, washed twice in PBS, and then stained with the intracellular Ab. The anti-J chain Ab was a gift from P. O. Brandtzæg, Oslo, Norway.

* The cells were fixed and permeabilized by incubating with 100% methanol for 10 min at room temperature, washed once in PBS, and then stained with anti-J chain Ab. Mouse anti-rabbit IgG PE was used as secondary layer.

* The Intrastain kit from DakoCytomation was used as described by the manufacturer.

Statistics analysis

Genes with >30% missing values were removed from the data set. The remaining missing values were estimated and imputed with LSimpute, adaptive (19). Subsequently, each gene expression profile was mean centered. These steps were performed using J-Express software (Molmine)
FIGURE 1. Differential gene expression during human B cell development. A, Immunomagnetic selection and subsequent FACS were used to isolate the five populations from adult human BM. Shown are the FACS dot plots with sorting gates to obtain CD34⁺CD38⁻ hematopoietic stem cells (HSC), CD34⁺CD38⁺CD10⁺CD19⁻ E-B cells (E-B), CD34⁺CD10⁻CD19⁻/IgM⁻ progenitor B cells (pro-B), CD34⁺CD10⁺CD19⁺/IgM⁻ precursor B cells (pre-B), and CD34⁻CD10⁺CD19⁺/IgM⁺ IM-B cells. B, Expression of the genes encoding the selection markers confirmed the validity of the approach. C, Hierarchical clustering of the 758 genes that differed significantly between two subsequent stages of differentiation (≥2-fold change, FDR 10%). Total RNA was extracted from the purified cell populations, amplified, and hybridized to Lymphochip cDNA microarrays. Independent experiments from six different donors were performed for each cell population. Color changes within a row indicate expression levels relative to the average of the sample population.

where

\[ \xi = \frac{1}{m} \sum_{i,j} \sqrt{d(x_{ij}, \bar{x}_j)^2} \]

\( i \) is group number, \( j \) is group element number, \( x_{ij} \) and \( \bar{x}_j \) are the two-dimensional coordinates (expression level of the two genes) of each samples and group centroid, respectively, and \( d(i) \) denotes the euclidean distance.

**Results**

**Differential gene expression in HSC, E-B, pro-B, pre-B, and IM-B cell populations**

The objective of this study was to perform gene expression profiling of the earliest developmental stages of human B cell differentiation. RNA was extracted from 1 × 10⁴–8 × 10⁴ cells of each of the five populations from normal adult human BM: CD34⁺CD38⁻ HSC cells; CD34⁺CD38⁺CD10⁺CD19⁻ E-B cells; CD34⁺CD10⁺CD19⁻/IgM⁻ pro-B cells; CD34⁺CD10⁺CD19⁺/IgM⁻ pre-B cells, and CD34⁻CD10⁺CD19⁺/IgM⁺ IM-B cells, followed by linear RNA amplification and hybridization to Lymphochip cDNA microarrays (Fig. 1A). The mRNA expression of the genes encoding the selection markers corresponded closely to the expression of the respective proteins, confirming the validity of the approach (Fig. 1B). Furthermore, expression of 758 genes differed significantly between two subsequent stages of differentiation (≥2-fold change, FDR 10%). Hierarchical clustering of these genes revealed a pattern that clearly separated the five consecutive cell populations (Fig. 1C). Indeed, each population displayed a distinct gene expression profile that was remarkably consistent among the individual donors (n = 6). Furthermore, there was a stepwise change in the gene expression pattern from HSC throughout the other developmental stages. A total of 191, 189, 234, and 363 genes changed in the HSC to E-B, E-B to pro-B, pro-B to pre-B, and pre-B to IM-B cell transitions, respectively (total gene list in supplementary table IB).

**Lineage-associated markers**

The expression of genes known to participate in B cell differentiation and activation (B cell-associated markers), illustrated the shift in gene expression during B lymphopoiesis (Fig. 2, B cell markers). As expected, several B lineage-associated genes, including CD19, CD24, CD79B, and VPREB3, showed highest expression levels from the pro-B stage and through the subsequent stages (cluster IV). Interestingly, ADA, DNTT (TdT), and the Ig recombination genes RAG1 and RAG2 showed increased expression levels already in the CD34⁺CD38⁺CD10⁺CD19⁻ E-B population (4.5-, 26.1-, 7.3-, and 5.6-fold increase, respectively, compared with the expression levels in HSC), together with HMGB1, HMGB2, and EZH2, suggesting that the rearrangement machinery is activated in CD19-negative cells (cluster II). Additionally, TCF3 (E2A), EBF, and PAX5, which are of crucial importance for B cell development and function, were expressed at higher levels in E-B vs HSC (1.7-, 3.5-, and 8.0-fold increase, respectively), and were further increased in the E-B to pro-B transition (2.9-, 1.9-, and 4.4-fold, respectively; Figs. 2 (cluster IV) and 3B, and supplementary table IB). As expected, genes known to be involved in immune responses and Ag binding, including HLAs, TAP2, IGHM, CD83, CD86, BLR1, and BTG1, showed the highest expression levels in the pre-B and IM-B populations (cluster III).

The E-B population also expressed several genes encoding T or NK lineage-associated markers, including CD2, SELL (CD62L), CD99, CD244, and NOTCH1 (Fig. 2, T/NK cell markers). Moreover, NOTCH1, which is crucial for T lineage commitmenet and development, was expressed in HSC, E-B and pro-B cells, but was down-regulated in later stages. In contrast, GATA3 which also is important for T lineage commitmenet was down-regulated in the HSC to E-B transition. Interestingly, E-B cells had higher expression levels of myeloid-associated markers MPO, PECAM1, and TNFSF13B, than HSC and pro-B cells (Fig. 2, myeloid cell markers).
Gene expression patterns related to function

To further investigate the functional assignment of the differentially expressed genes, we used the proliferation signature (18), or created gene sets based on information from GO (details in Materials and Methods and supplementary table IA). The proliferation signature included different cell cycle control genes, cell cycle checkpoint genes, and genes involved in DNA synthesis and replication. Interestingly, this signature divided the populations largely in two groups: the HSC and IM population showed low expression of virtually all genes in this signature (Fig. 3A, proliferation signature). In contrast, the pro-B and pre-B populations comprised high expression levels of nearly all the proliferation-associated genes, in agreement with the clonal expansion and rapid proliferation which follow successful rearrangement of the H and L chains. Interestingly, this signature divided the populations largely in two groups: the HSC and IM population showed low expression of virtually all genes in this signature (Fig. 3A, proliferation signature). In contrast, the pro-B and pre-B populations comprised high expression levels of nearly all the proliferation-associated genes, in agreement with the clonal expansion and rapid proliferation which follow successful rearrangement of the H and L chains. Interestingly, this signature suggested that the E-B population was not as quiescent as HSC cells. Thirty proliferation-associated genes were increased ≥2-fold in the HSC to E-B transition, including several genes involved in S phase, such as CKEK1, TOP2A, UHRF1, PCNA, BZW1, and ILF3, and genes involved in the G2/M phase transition (KIF11, KIF14, and BUB1) (Fig. 3B). These results corresponded well with the findings that 13 ± 2% of E-B cells were in S/G2/M phase of cell cycle, compared with 0.4 ± 0.2% and 30 ± 2% of HSC and pro-B cells, respectively (n = 4, data not shown). The high fraction of quiescent cells in the HSC population is in accordance with the finding of others (22, 23), and a progressive reduction in the percentage of GO cells during differentiation has also been detected in mice (24).

The apoptosis gene set included 38 differentially expressed genes which could be separated in three main clusters (Fig. 3A, apoptosis gene set). Of note, genes with highest mRNA expression levels in HSC and particularly in E-B, included the antiapoptotic gene BCL2 in addition to TNFSF13B (encoding the cytokine BAFF), but also the proapoptotic gene DAPK1 (cluster II). Several antiapoptotic genes such as BIRC5 (survivine), MYBL2 (B-myb), and FAIM, as well as suppressor of cytokine signaling 2 showed highest expression levels in pro-B and pre-B cells (cluster III).

Finally, genes with highest expression levels in IM-B cells and partially also in pre-B cells, included several genes involved in induction of apoptosis like BTG1, HRK, TRAF5, STK17B, TNFRSF21, and BIK (cluster I). Taken together, the populations with highest similarity regarding expression of the genes in the apoptosis gene set turned out to be the HSC and E-B populations.

The adhesion gene set generated four main clusters which clearly illustrated the shift in expression of cell-cell and cell-matrix molecules as the cells mature (Fig. 3A, adhesion gene set). Genes with high expression in HSC and E-B and low expression in the next three populations included ITGB2, CD44, and SELL (CD62L) (cluster I), while CD34, CDH2 (N-cadherin), CD99, and ITGA6
FIGURE 3. Gene expression changes according to function. The 758 genes that differed significantly between two subsequent stages of differentiation (≥2-fold change, FDR 10%) were merged to visualize their expression during the E-B cell development and shown according to their functional assignment to the different gene sets (apoptosis, adhesion, regulation of transcription or signaling). The proliferation signature was developed previously (18). The other gene sets were obtained based on information from GO (details in supplementary table IA). A. Analyses of changes in gene expression pattern between the five consecutive cell populations. B, Changes in gene expression pattern in the HSC to E-B transition and the E-B to pro-B transition.
showed high expression in the first three populations and were down-regulated in the pre-B and IM-B cell populations (cluster II). CD44 and CDH2 have recently been implicated to play a role in interaction with the stem cell niche (25, 26). Interestingly, CTNNA1 (α-catenin), ICAM3, and SELP (P-selectin) are the only genes in the gene set that were significantly down-regulated in the HSC to E-B transition, whereas all others were abundantly expressed also in E-B cells (Fig. 3B). As the cells have rearranged the Ig H (pre-B) and L chains (IM-B), and the stroma dependency is declining (27), a new set of adhesion genes were up-regulated, including CD72, CD9, and VCAM1 (cluster IV).

Totally, 111 genes associated with transcription regulation were differentially expressed in the five populations (Fig. 3A, transcription gene set). Transcription factors with high mRNA expression in HSC and partly also in E-B, which were down-regulated in the next three consecutive populations, included homeobox genes known to play a crucial role in normal development like HOXB6, HOXB2, and MEIS1, and genes important for the self-renewal potential such as STAT5A (cluster I). Genes with high expression levels in the first three populations (HSC, E-B, and pro-B cells), and then down-regulated in the pre-B and IM-B populations included TIEG, XBP1, ERG, and TCF7L2 (cluster III). As earlier described, the essential transcription factors EBF, TCF3, PAX5, and LEF1 were up-regulated in HSC to E-B and in E-B to pro-B transitions and continued to be high during the next differentiation steps (cluster VI). Interestingly, several other transcription factors showed similar expression patterns including MEF2A, ZHX2, ZCCHC7, DMTF1, NR3C1, FOXL1A, BACH2, and RB1. Cluster V consisted of genes with high expression levels in 1) E-B and pro-B or 2) pre-B and pre-B. The genes up-regulated in E-B and pro-B cells included BCL11A, which is essential for normal lymphopoiesis (28) (cluster Va). Among the genes up-regulated in pro-B and pre-B cells were transcription factors known to be involved in regulation of proliferation or differentiation including MYBL2, FOXL1, GFI1 and TCF19 (Cluster Vb). Finally, several of the transcription factors with high expression levels in the pre-B and IM-B cells included the transcription factors IRF4, ICSBP1 (IRF8), NFATC4 and VDR, (cluster IV). Interestingly, several of the transcription factors that were expressed during the earliest stages of B cell development were also found to be highly expressed in germinal center (GC) cells (GC B cells or CD195 peripheral blood B cells, data not shown), including XBP1, BACH2, HMGB2, EZH2, MCM2, MCM7, TCF19, FOXL1, NR3C1, RB1, UHRF1, E2F1, and SMARCA4.

Finally, 127 genes coupled to intracellular signaling generated six clusters (Fig. 3A, signaling gene set). Genes with low expression in HSC, E-B, and pro-B, but that were highly up-regulated in pre-B and IM-B cells, included the chemokine and cytokine receptors CXCR4 and IL4R, and several kinases including FGR, TRAF5, and LYN (cluster II). Interestingly, the TNF family members, TNFSR13B (BlyS/BAFF) and TNFSF4 (OX40L), were down-regulated in pre-B and IM-B cells (cluster IV). Additionally, adapter proteins and kinases were represented in this cluster, including LCP2 (SLP76), FYN, MAP4K3 and MAPKAPK3. Thus, the mRNA levels of LYN and FYN, which are important in (pre-) BCR signaling, were inversely expressed in the five populations. Finally, genes included in cluster VI were up-regulated in pro-B and pre-B cells, and partly in E-B cells. In addition to genes involved in cell cycle regulation, this cluster also included BLNK (SLP65), an important downstream mediator of (pre-) BCR signaling.

**E-B gene expression profile**

To characterize the earliest steps in B cell development in further detail, we next compared genes differentially expressed in E-B compared with the neighbor populations HSC and pro-B cells (2-fold change, 30% FDR), giving a list of 22 genes which constituted the E-B gene expression profile (Table II). Among these genes, 18 showed higher and 4 showed lower expression levels in E-B compared with the HSC and pro-B populations. Interestingly, IGF1 (3 chain) was the gene showing the highest differential expression in the E-B population with a 9.4-fold increase compared with the two neighboring populations. Additionally, the other up-regulated genes included the antiapoptotic gene BCL2, ILIRAP (essential signal
transducing component of the functional IL-1R), CYBB (a subunit of cytochrome b), the MHC class II regulatory factor RFX1, the inhibitory smad MADH7, S100A8 (a calcium-binding S100 protein involved in the regulation of cell cycle progression and differentiation) and MPO (a heme protein synthesized during myeloid differentiation).

Characterization of HSC, E-B, and pro-B cells by gene pairs

A gene pair class separation test was performed to identify combinations of markers that could better discriminate the E-B population. The ability of the gene pairs to discriminate between the stages was ranked by a score, and a score above 10 gave excellent separation. Shown are dot plots for three selected gene pairs, where each dot represents the median mRNA from individual donors: green, HSC; red, E-B; and blue, pro-B.

Protein expression of selected gene pairs, including CD38 and CD44, were also analyzed at the single-cell level by FACS analysis. As the protein expression of several of these gene pairs was relatively heterogeneous, particularly in pro-B cells, we obtained a less
criminators for the three populations with score 12.1, 10.9, and 26.8, respectively (Fig. 4 and Table III).

Confirmation of the differential expression of selected genes by FACS analysis

Among the genes identified in the E-B gene expression profile, we selected a set of genes to be examined at the protein level based on feasibility in FACS analyses. In addition, we included receptors known to be involved during B lymphopoiesis, like CD127 (IL-7Ra) and CXCR4. Thus, CD34 results were stained with Abs to identify the HSC, E-B, and pro-B populations, in addition to Abs to the selected proteins. Overall, there were good correlations between mRNA and protein levels in the three populations (Fig. 5). E-B cells demonstrated a strong and relatively homogenous expression of TdT protein and most of these cells also expressed IL-1RacP (IL1RAP), Bcl-2, CD127, and CXCR4. Moreover, we detected a weak expression of J chain and CD2 in E-B, while HSC and pro B cells did not express detectable levels. The expression level of CD62L (SELL) was relatively homogeneous and strong in E-B cells, in contrast to a weaker and more heterogeneous expression in HSC and pro-B cells, respectively. Despite higher MPO mRNA levels in E-B cells than in HSC and pro-B cells, we did not observe detectable levels of MPO protein in these cells (Fig. 5). A gene pair class separation test was performed to identify combinations of markers that could better discriminate the E-B population, in addition to Abs to the selected proteins.

Table III. Gene pairs that separate HSC, E-B, and pro-B cell populations

<table>
<thead>
<tr>
<th>Rank</th>
<th>Score</th>
<th>Gene 1</th>
<th>Gene 2</th>
<th>Rank</th>
<th>Score</th>
<th>Gene 1</th>
<th>Gene 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>26.8</td>
<td>IL1RAP</td>
<td>CD44</td>
<td>25</td>
<td>12.7</td>
<td>DNTT</td>
<td>IGJ</td>
</tr>
<tr>
<td>2</td>
<td>20.7</td>
<td>RASGRP2</td>
<td>CD44</td>
<td>28</td>
<td>12.1</td>
<td>SELL</td>
<td>CD72</td>
</tr>
<tr>
<td>3</td>
<td>19.7</td>
<td>MEIS1</td>
<td>CD44</td>
<td>41</td>
<td>11.2</td>
<td>CD72</td>
<td>CD72</td>
</tr>
<tr>
<td>4</td>
<td>18.4</td>
<td>CD86</td>
<td>CD44</td>
<td>44</td>
<td>11.0</td>
<td>CD79B</td>
<td>CD79B</td>
</tr>
<tr>
<td>5</td>
<td>17.5</td>
<td>CD79B</td>
<td>CD44</td>
<td>29</td>
<td>12.1</td>
<td>SELL</td>
<td>MEIS1</td>
</tr>
<tr>
<td>6</td>
<td>17.5</td>
<td>CTGF</td>
<td>CD44</td>
<td>38</td>
<td>11.5</td>
<td>MEIS1</td>
<td>ADRA2A</td>
</tr>
<tr>
<td>7</td>
<td>16.9</td>
<td>MERTK</td>
<td>CD44</td>
<td>43</td>
<td>11.1</td>
<td>PAX5</td>
<td>CD44</td>
</tr>
<tr>
<td>8</td>
<td>16.9</td>
<td>DNTT</td>
<td>CD44</td>
<td>64</td>
<td>10.1</td>
<td>CTGF</td>
<td>CD44</td>
</tr>
<tr>
<td>9</td>
<td>16.8</td>
<td>CD38</td>
<td>CD44</td>
<td>30</td>
<td>12.1</td>
<td>MEIS1</td>
<td>CD79B</td>
</tr>
<tr>
<td>10</td>
<td>16.7</td>
<td>MKI67</td>
<td>CD44</td>
<td>37</td>
<td>11.5</td>
<td>CD72</td>
<td>CD72</td>
</tr>
<tr>
<td>11</td>
<td>16.1</td>
<td>PAX5</td>
<td>CD44</td>
<td>42</td>
<td>11.2</td>
<td>CTGF</td>
<td>CD44</td>
</tr>
<tr>
<td>12</td>
<td>15.6</td>
<td>DDR1</td>
<td>CD44</td>
<td>47</td>
<td>11.0</td>
<td>CYBB</td>
<td>CD44</td>
</tr>
<tr>
<td>13</td>
<td>15.5</td>
<td>EFNB2</td>
<td>CD44</td>
<td>36</td>
<td>11.5</td>
<td>PAX5</td>
<td>CD72</td>
</tr>
<tr>
<td>14</td>
<td>14.3</td>
<td>CD81</td>
<td>CD44</td>
<td>61</td>
<td>10.2</td>
<td>EFNB2</td>
<td>CD44</td>
</tr>
<tr>
<td>15</td>
<td>14.1</td>
<td>PTPRE</td>
<td>CD44</td>
<td>18</td>
<td>13.3</td>
<td>CTGF</td>
<td>CYBB</td>
</tr>
<tr>
<td>16</td>
<td>13.4</td>
<td>SELL</td>
<td>CD44</td>
<td>26</td>
<td>12.4</td>
<td>BCL2</td>
<td>CD44</td>
</tr>
<tr>
<td>17</td>
<td>13.1</td>
<td>CXCR4</td>
<td>CD44</td>
<td>33</td>
<td>11.8</td>
<td>ADRA2A</td>
<td>CD44</td>
</tr>
<tr>
<td>18</td>
<td>12.9</td>
<td>ADRA2A</td>
<td>CD44</td>
<td>46</td>
<td>11.0</td>
<td>CD38</td>
<td>CD44</td>
</tr>
<tr>
<td>19</td>
<td>12.8</td>
<td>CD50</td>
<td>CD44</td>
<td>54</td>
<td>10.8</td>
<td>LAIR1</td>
<td>CD44</td>
</tr>
<tr>
<td>20</td>
<td>12.7</td>
<td>CD72</td>
<td>CD44</td>
<td>55</td>
<td>10.8</td>
<td>IGJ</td>
<td>CD44</td>
</tr>
<tr>
<td>21</td>
<td>12.4</td>
<td>DLG5</td>
<td>CD44</td>
<td>56</td>
<td>10.7</td>
<td>DLG5</td>
<td>CD44</td>
</tr>
<tr>
<td>22</td>
<td>11.9</td>
<td>ADAM28</td>
<td>CD44</td>
<td>57</td>
<td>10.7</td>
<td>CD45</td>
<td>CD45</td>
</tr>
<tr>
<td>23</td>
<td>11.0</td>
<td>CD24</td>
<td>CD44</td>
<td>65</td>
<td>10.1</td>
<td>FLNB</td>
<td>CD45</td>
</tr>
<tr>
<td>24</td>
<td>10.8</td>
<td>CD10</td>
<td>CD44</td>
<td>66</td>
<td>10.1</td>
<td>DDR1</td>
<td>CD45</td>
</tr>
<tr>
<td>25</td>
<td>10.8</td>
<td>IL1RAP</td>
<td>CD44</td>
<td>68</td>
<td>10.0</td>
<td>CD50</td>
<td>CD45</td>
</tr>
<tr>
<td>26</td>
<td>10.1</td>
<td>CD45</td>
<td>CD44</td>
<td>31</td>
<td>11.9</td>
<td>ADRA2A</td>
<td>CD79B</td>
</tr>
<tr>
<td>27</td>
<td>14.2</td>
<td>RASGRP2</td>
<td>CD44</td>
<td>58</td>
<td>10.4</td>
<td>CYBB</td>
<td>CD44</td>
</tr>
<tr>
<td>28</td>
<td>13.0</td>
<td>CD44</td>
<td>CD44</td>
<td>59</td>
<td>10.4</td>
<td>CYBB</td>
<td>CD44</td>
</tr>
<tr>
<td>29</td>
<td>11.6</td>
<td>EFNB2</td>
<td>CD44</td>
<td>60</td>
<td>10.3</td>
<td>CD45</td>
<td>CD45</td>
</tr>
<tr>
<td>30</td>
<td>11.2</td>
<td>CTGF</td>
<td>CD44</td>
<td>62</td>
<td>10.2</td>
<td>DDR1</td>
<td>CD45</td>
</tr>
<tr>
<td>31</td>
<td>12.8</td>
<td>CD44</td>
<td>CD44</td>
<td>67</td>
<td>10.0</td>
<td>DLG5</td>
<td>CD45</td>
</tr>
<tr>
<td>32</td>
<td>11.7</td>
<td>CTGF</td>
<td>CD44</td>
<td>67</td>
<td>10.0</td>
<td>DLG5</td>
<td>CD45</td>
</tr>
<tr>
<td>33</td>
<td>11.3</td>
<td>ADRA2A</td>
<td>CD44</td>
<td>67</td>
<td>10.0</td>
<td>DLG5</td>
<td>CD45</td>
</tr>
<tr>
<td>34</td>
<td>10.9</td>
<td>CD44</td>
<td>CD44</td>
<td>67</td>
<td>10.0</td>
<td>DLG5</td>
<td>CD45</td>
</tr>
<tr>
<td>35</td>
<td>10.9</td>
<td>CD38</td>
<td>CD44</td>
<td>67</td>
<td>10.0</td>
<td>DLG5</td>
<td>CD45</td>
</tr>
<tr>
<td>36</td>
<td>10.8</td>
<td>DLG5</td>
<td>CD44</td>
<td>67</td>
<td>10.0</td>
<td>DLG5</td>
<td>CD45</td>
</tr>
</tbody>
</table>
that this population encompasses several subpopulations (1). In humans, most investigators have identified CLP as CD34+/CD10−/CD19− (3, 5, 29). Of interest, CD34+/CD10+/CD19− cells expressing CD127 (IL-7Ra), were found to be biased toward the B lineage (30). We found that the majority of CD34+/CD10+/CD19− E-B cells expressed CD127 (77 ± 1%), while only a small fraction coexpressed CD7 (3 ± 1%, data not shown). This is in line with other data, which also indicated that CD34+/CD7+ cells are biased toward T lineage development (31). Interestingly, we found that the T/NK cell-associated gene CD2 was specifically expressed in E-B cells both at the mRNA and the protein level. CD2 expression has previously been detected on normal and neoplastic B progenitors (32), but the striking expression of CD2 on CD34+/CD38−/CD10−/CD19− cells has not been previously recognized. In addition, other T/NK-associated genes were also expressed in E-B cells, including IL1RAP, SELL, CD244, CD99, and NOTCH1. Furthermore, E-B cells expressed mRNA for the myeloid-associated genes MPO, PECAM1, and TNFSF13B, although MPO protein expression could not be detected. Taken together, these data suggest that the E-B population might contain cells with multilineage potential. Of note, Rumfelt et al. (33) recently demonstrated that murine CD19+ B cell precursor populations revealed considerable lineage plasticity and could be divided into three subpopulations, including multilineage progenitor cells, CLP cells which possessed lymphoid and myeloid potential (MLP), and pro-pre-B cells which possessed B/T potential. Investigations to further subdivide human CD34+ cells on the basis of CD2, CD10, CD22, CD62L, CD244, CXCR4, and CD127 expression, in combination with functional studies to explore the lineage plasticity, might be of value for better characterization of different E-B lineage populations in humans.

How early the initiation of VDJ rearrangement in B lineage cells starts is still debated and has been a focus in recent studies. We found that the E-B population had turned on transcription of several genes encoding proteins involved in gene rearrangement, including RAG1, RAG2, DNTT (TdT), and ADA, indicating that the recombine machinery is initiated before expression of CD19. This is in line with other studies, which also found high expression of TdT in human CD34+CD19−/CD10−/lin− cells (4), and RAG1 and RAG2 in CD34+Lin−/CD10−/CD7− cells (5). Moreover, it has been shown that CD34+/CD10+/CD127−/CD19− cells have initiated rearrangement of DJH (34). The initiation of rearrangement of DJH in CD34+/CD22+/CD19− cells was shown by van Zelm et al. (13), although they found that RAG1 or RAG2 expression levels were hardly up-regulated in this population, suggesting that rearrangement can be initiated even with low levels of RAG1/RAG2. It should be noted that van Zelm et al. (13) defined pro-B cells as CD34+Lin−/CD22+. Analysis of CD22 expression revealed that among CD34+CD19− cells, 10.6% were CD22+ CD10−, 0.5% were CD10+CD22−, and 2.4% were CD22− CD10+ (data not shown), indicating that our E-B subpopulation (composed of CD22+ CD10− and CD10+CD22−) was only partly overlapping with the CD34+Lin−/CD22+ population. In line with our findings, RAG2 and TdT expression was already detectable in murine MLP cells, while CLP cells were found to express TdT, RAG1, and RAG2 at significant levels (33). We observed a homogenous and strong TdT protein staining of E-B cells, whereas HSC had no detectable levels. This could indicate that there exist intermediate precursors between CD34+/CD38+ cells and the CD34−/CD38−/CD10−/CD19− E-B population with low or intermediate levels of TdT and/or RAG genes, which were not included in our study. Indeed, recent data suggest that CD79A and CD127 expression...
of J chain in E-B cells, but not in HSC or in pro-B cells (Table 2). Some genes known to be activated by PAX5 were found to be increased in the E-B population including BLNK (8.4-fold) and CD79A (1.9-fold), whereas another target gene, CD19, was turned on in the pro-B cells. IGJ and XBP1, genes known to be repressed by PAX5, were down-regulated in pro-B and pre-B cells, respectively. The differential effects of PAX5 on various target genes is in line with previous data (37) and is likely to due to variable expression and/or activity of transcriptional coregulators interacting with PAX5 at the different stages in B cell development. Of note, we also identified a number of genes encoding transcription factors with expression pattern similar to the described essential factors. EWSR1, ZCCHC7, MEF2A, and NR3C1 showed exactly the same expression pattern as EBF, whereas the expression of ARID5B, ZHX2, DMTF1, FOXXO1A, BACH2, ZEP36L1, and RB1 was similar to the expression of E2A and PAX5. Many of these genes have a well-characterized role in cell cycle regulation, whereas little is known regarding other genes, in particular the zinc finger transcription factors ZHX2 and ZCCHC7. Thus, these two transcription factors represent interesting candidates to be further explored regarding a putative role in E-B cell development.

Of note, several of the transcription factors showed a biphasic expression pattern during B cell development. We observed that IRF4 and ICSBP1 (IRF8) showed expression peaks in the pre-B and IM-B cells, respectively. In line with this, double knockout mice (IRF4−/−/IRF8−/−) were shown to have a complete block at the pre-B cell stage, demonstrating a crucial role for these factors in the transition from pre-B to IM-B cells (38). Both genes are down-regulated at later stages until re-expressed in a subpopulation of centrocytes and in plasma cells (39), where they control class-switch recombination and plasma cell differentiation, respectively (40, 41). BCL11A also demonstrated a biphasic expression pattern with up-regulation in the E-B and pro-B populations. BCL11A has previously been demonstrated to be essential for lineage commitment of both B and T cells in mice (28), and is expressed also in GC cells, where it functions as a transcriptional repressor (42, 43). Interestingly, we identified several other transcription factors with biphasic expression patterns. Most of these factors are known to be important for the cell cycle, RNA transcription, or cell proliferation, but also included factors involved in V(D)J recombination such as HMGB2 and EZH2 (44, 45). In addition, we observed high expression levels in early stages of B cell development of other transcription factors with known essential roles in GC B cells or plasma cells, including BACH2 and XBP1 (46, 47). However, there is no evidence for a functional role for these factors in E-B cell progenitors. Taken together, the biphasic expression of several transcription factors during B cell differentiation suggest the use of similar transcriptional networks in E-B cell stages and in GC and plasma cells.

One of the surprising findings in our study was that IGJ gene (J chain) expression was significantly up-regulated in E-B cells compared with HSC and pro-B cells. The mRNA expression in E-B cells was, however, much lower than in IM-B cells. In line with this, we could detect weak, but distinct, protein expression of J chain in E-B cells, but not in HSC or in pro-B cells (Table II and Fig. 5). The J chain determines the correct assembly of IgA dimers and IgM pentamers (48), and accordingly shows a strong increase during the last stages of B cell activation (49). In line with our data, J-chain mRNA expression has previously been detected in B cell lines representing progenitors and also in primary cells including E-B progenitors (CD34+/CD19+), pro-B cells, and in early fetal thymocytes (50–52). The transcriptional regulation of J-chain expression has been well-studied in mice (53, 54), and myocyte enhancer factor-2 (MEF2) is one of the known positive regulators of IGJ gene expression. Of note, we found that MEF2A was highly up-regulated in E-B cells, suggesting that the increased MEF2A levels could contribute to J-chain expression in E-B cells. Pax5, a negative regulator of IGJ, was slightly up-regulated in E-B cells, and showed a further increase in the E-B/pro-B transition. It has been hypothesized that high levels of Pax5 prevent IGJ expression during B cell development until Pax5 is down-regulated at the plasma cell stage (53, 54). Thus, the transient IGJ expression observed at the E-B stage could be due to high expression of MEF2A, preceding high levels of Pax5 in pro-B cells. The transient expression of J chain in early lymphoid progenitors, before Ig expression, is intriguing, but so far there is no evidence for a functional role for J chain in early progenitors.

Identification and characterization of the gene expression in normal precursor B cell subsets are valuable for the study of malignant B lineage cells, and can give a more precise reference for the cellular origin of precursor B cell acute lymphoblastic leukemia (ALL). ALL is a heterogeneous disease with a number of genetically distinct leukemia subtypes, and can arise from any stage of development, including primitive cells with multilineage potential. Gene expression profiling (55–57), and single nucleotide polymorphism array analysis have provided new insights into the underlying biology of ALL subtypes (58). Interestingly, Mullighan et al. (59) recently used single nucleotide polymorphism arrays to detect genetic alterations in leukemic cells from 242 ALL patients. Pax5 turned out to be the most frequently mutated gene, giving reduced levels of Pax5 protein or generated hypomorphic alleles. Furthermore, deletions were also detected in TCF3, LEF1, IKZF1, and IKZF3, suggesting that direct disruption of pathways controlling B cell development contributes to B-ALL pathogenesis. Hence, comparison of gene expression profiles of normal and neoplastic B cell progenitors may identify genes involved in the oncogenic transformation and gene products of potential prognostic and therapeutic relevance.

Acknowledgment

We thank Prof. P. O. Brandtzaeg for kindly providing us with the anti-J-chain Ab.

Disclosures

The authors have no financial conflict of interest.

References


