Ig Gene Rearrangement Steps Are Initiated in Early Human Precursor B Cell Subsets and Correlate with Specific Transcription Factor Expression


*J Immunol* 2005; 175:5912-5922; doi: 10.4049/jimmunol.175.9.5912
http://www.jimmunol.org/content/175/9/5912

**Supplementary Material**
http://www.jimmunol.org/content/suppl/2005/10/19/175.9.5912.DC1

**References**
This article cites 51 articles, 21 of which you can access for free at:
http://www.jimmunol.org/content/175/9/5912.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
Ig Gene Rearrangement Steps Are Initiated in Early Human Precursor B Cell Subsets and Correlate with Specific Transcription Factor Expression

Menno C. van Zelm,*† Mirjam van der Burg,* Dick de Ridder,*‡ Barbara H. Barendregt,*† Edwin F. E. de Haas,* Marcel J. T. Reinders,‡ Arjan C. Lankester,§ Tom Révész,¶ Frank J. T. Staal,* and Jacques J. M. van Dongen*‡

The role of specific transcription factors in the initiation and regulation of Ig gene rearrangements has been studied extensively in mice models, but data on normal human precursor B cell differentiation are limited. We purified five human precursor B cell subsets, and assessed and quantified their IGH, IGK, and IGL gene rearrangement patterns and gene expression profiles. Pro-B cells already massively initiate D_{H1}-J_{H} rearrangements, which are completed with V_{H}-D_{JH} rearrangements in pre-B-I cells. Large lymphoblastic leukemia and B cell differentiation blocks in primary Ab deficiencies. PAX5, and IRF4, were specifically up-regulated at stages undergoing Ig gene rearrangements. Based on the combined Ig gene rearrangement status and gene expression profiles of consecutive precursor B cell subsets, we identified 16 candidate genes involved in initiation and/or regulation of Ig gene rearrangements. These analyses provide new insights into early human precursor B cell differentiation steps and represent an excellent template for studies on oncogenic transformation in precursor B acute lymphoblastic leukemia and B cell differentiation blocks in primary Ab deficiencies. The Journal of Immunology, 2005, 175: 5912–5922.

Precursor B cells develop from hemopoietic stem cells and differentiate through a number of stages in the bone marrow (BM) before they migrate to the periphery as naive mature B lymphocytes. The ultimate purpose of B cell differentiation is to produce the broad repertoire of B cell Ag receptors, which are composed of two identical Ig heavy chains and two identical Ig light chains (reviewed in Ref. 1). Early in differentiation, V(D)J recombination is initiated in the Ig H chain (IGH) locus with D_{H1} to J_{H} rearrangements in pro-B cells. The lymphocyte-specific RAG1 and RAG2 proteins introduce a single-stranded nick between the recombination signal sequence (RSS) and the flanking D or J gene segment, which results in the generation of a dsDNA break (2, 3). The DNA-bending high mobility group proteins HMGB1 and HMGB2 stimulate the RAG proteins in DNA binding and the generation of the dsDNA breaks (4), which are subsequently repaired via nonhomologous end joining (NHEJ) (5). In the next stage (pre-B-I), a V_{H} segment is rearranged to the DJ_{H} element, and if an in-frame VDJ_{H} exon is formed, a pre-BCR is expressed, which is composed of IgH chains and surrogate light chains (VpreB and λ14.1). The expression of this receptor initiates several cycles of proliferation (large cycling pre-B-II cell stage). After this proliferation phase, Ig L chain rearrangements are initiated in the small pre-B-II cells. If a functional Ig L chain (IgL or Igα) is expressed that is able to assemble with the Ig H chain, the cell becomes a surface membrane Ig 

Copyright © 2005 by The American Association of Immunologists, Inc.

0022-1767/05/$02.00
partner of IRF4 and IRF8 in large pre-B cells, where they inhibit proliferation and initiate Ig L chain rearrangements (19, 20). In addition, histone-remodeling factors, EZH2 and BRG1, and DNA demethylases are likely to be involved in the induction of Ig gene rearrangements (21–23).

The majority of our understanding of the initiation and regulation of Ig gene rearrangements in precursor B cell differentiation comes from studies in mouse models, and genome-wide gene expression profiling has been performed exclusively in murine precursor B cells (24, 25). However, these studies did not show how V(D)J recombination is strictly regulated per locus in each specific stage of differentiation.

Therefore, we aimed at purifying cells in the five main stages of human precursor B cell differentiation, thus creating the opportunity to study the initiation of IGH and IGK/IGL rearrangements independently from each other and independently from the selection processes. The precursor B cell subsets were purified with membrane markers only, to allow both DNA extraction for analysis of the Ig gene rearrangements using real-time quantitative PCR (RQ-PCR) and GeneScan assays, and RNA extraction for genome-wide gene expression profiling using Affymetrix GeneChip arrays. This approach enabled us for the first time to study the networks of factors that might be involved in the initiation and regulation of V(D)J recombination processes in the different Ig loci in human precursor B cell differentiation.

Materials and Methods

Purification of CD34+lin- and (precursor) B cell populations

Precursor B cells were obtained from freshly isolated BM samples of healthy children (3–16 years of age), who served as donors for BM transplantation in a sibling. The BM samples were taken for quality control of the graft. Remaining BM (generally 0.5–2.0 ml) was used for the precursor B cell studies. CD34+lin- cells were obtained from umbilical cord blood (UCB), because it was not possible to obtain enough CD34+lin- cells from BM samples.

For isolation of CD34+lin- cells and pro-B cells, MACS CD34 beads (Miltenyi Biotec) were used to positively select CD34+ progenitor cells. CD34+lin- cells were further purified by sorting on a FACSdiVa cell sorter (BD Biosciences) after labeling with CD34-APC (8G12), and CD33-PE (Leu-4), CD16-PE (B73.1), CD19-PE (4G7), CD56-PE (M431; all from BD Biosciences) and CD13-RED (MY7), CD33-RED (906; both from Beckman Coulter) for exclusion. Pro-B cells were further purified after labeling with CD34-FITC, CD19-PE, CD123-PE (95F), and CD22-APC (S-HCL-1; all from BD Biosciences). Four CD19+ precursor B cell subsets were MACS sorted using CD19 beads (Miltenyi Biotec) and staining with CD34-FITC, CD20-PE (Leu-16), CD10-APC (H110a; both from BD Biosciences), and anti-IgM-Cy5 (Jackson Immunoresearch Laboratories). Additional markers for defining and characterization of the precursor B cell fractions were: Ki67-FITC, anti-IgM-Cy5 (both from DakoCytomation); anti-IgA-PE (Southern Biotechnology Associates); and CD79b-FITC (SN8; BD Biosciences). All fractions were collected in 2-ml tubes containing filtered heptane-inactivated FCS and were kept at 4°C until reanalysis on a FACSCalibur (BD Biosciences) and subsequent isolation of DNA and/or RNA.

Mature Igκ+ and Igλ- B lymphocytes were sorted after mechanical disruption of tonsillar tissue obtained from children. Ficoll density centrifugation and direct staining with anti-Igκ-FITC, anti-Igλ-PE and CD19-APC (SJ25C1; BD Biosciences).

When possible, DNA and RNA were isolated from subsets of three donors. DNA was isolated twice for CD34+lin- , pro-B, mature Igκ+, and mature Igλ- cells. RNA was isolated twice from CD34+lin- and large pre-B cells and once from pro-B cells. All fractions were obtained with a purity of >95%.

All cell samples were obtained according to the informed consent guidelines of the Medical Ethics Committees of the University Medical Center Utrecht (Utrecht, The Netherlands), the Leiden University Medical Center (Leiden, The Netherlands), and the Erasmus MC, University Medical Center Rotterdam (Rotterdam, The Netherlands).

RQ-PCR and GeneScan analysis of Ig gene rearrangements

A multiplex TaqMan-based RQ-PCR was used to quantify IGH, IGK, and IGL gene rearrangements. Family-specific V and D segment forward primers, J segment-specific reverse primers, and J segment-specific probes were newly designed or adapted for such a way that >95% of all rearrangements could be detected (26–31). The RQ-PCR mix consisted of 25 µl contained TaqMan Universal MasterMix (Applied Biosystems), 900 nM concentrations of each primer (300 nM in case of multiplex mixtures), 100 nM concentrations of each FAM-TAMRA-labeled probe, 50 ng of DNA, and 0.4 ng of BSA, and was run on the ABIPRISM 7700 sequence detection system (Applied Biosystems) (32, 33).

Primers RQ-PCR were used for quantification of the DNA input (34). For quantification of the Ig gene rearrangement, a standard curve was made on a 10-fold dilution series of DNA from one or more cell lines harboring monoallelic Ig gene rearrangements, diluted in a background of germline (unrearranged) DNA (total input, 50 ng). Each assay reproducibly reached a sensitivity of at least 1% rearrangement in germline background and was performed in duplicate on each DNA sample. The rearrangements were quantified relative to the monially rearranged control cell lines, which were set at 100%. Consequently, it is possible to measure a level of 200% rearrangements in a cell population with all cells having biallelic rearrangements.

To analyze the in-frame selection of Ig gene rearrangements in precursor B cell subsets, complete IgH, IgK, and IgL rearrangements were amplified and subjected to GeneScan analysis with modified V segment forward primers positioned in such a way that in-frame rearrangements result in triplet spacing of the GeneScan peaks (primer sequences available on request) (30).

RNA isolation and gene expression profiling

RNA isolation and gene expression profiling were essentially described as performed before (35) and according to MIAME guidelines (36) using Affymetrix HG-U133 set GeneChip arrays and corresponding hybridization, washing, and scanning equipment (Affymetrix).

Due to the limited amount of material, 100 ng of total RNA were subjected to two cycles of in vitro transcription according to the Affymetrix small sample target labeling assay, version 2. The scaling factor, noise, and presence cells were comparable between all samples. GeneChip array data were quantile normalized (37), and background was removed using robust multichip analysis (38). Array groups corresponding to the development stages were compared based on the perfect match probe intensity levels only (38), by performing a per probe set two-way ANOVA (with factors probe and stage). This resulted in average expression levels for each probe set in each stage as well as p values for the significance of the difference between the stages. The p values were adjusted for multiple testing using Šidák stepdown adjustment (39) and all differences with adjusted p values <0.05 were considered significant, resulting in 5365 significant probe sets. All raw GeneChip array data are freely available at http://franklin.et.tudelft.nl/.

Selection of genes likely to be involved in the initiation and regulation of Ig gene rearrangements

The expression values of the 5365 significant probe sets were normalized per-probe set to zero mean and unit SD (z score). Hierarchical clustering (complete linkage) based on Pearson correlation was then performed on the significant probe sets using the Genlab software toolbox (www.genlab.tudelft.nl), running in the Matlab programming environment. The Gene Ontology (GO) Consortium website (http://www.godatabase.org) was used to select, within the significant probe sets, for probe sets annotated with GO terms representing genes involved in regulation of transcription, DNA recombination, and DNA repair: GO:3676, GO:3677, GO:3684, GO:3700, GO:3712, GO:3713, GO:3714, GO:5667, GO:5668, GO:6281, GO:6282, GO:6302, GO:6303, GO:6319, GO:6530, GO:6531, GO:6535, GO:6536, GO:6637, GO:16481, GO:16563, GO:16564, GO:45739, GO:45941. Genes not involved in regulation of transcription, DNA recombination, and DNA repair, such as RNA-editing enzymes and histones, were manually removed from the list.

Confirmation of gene expression patterns by RQ-PCR

To confirm the gene expression patterns of the RAG genes and 16 newly identified genes in the isolated populations, RNA was reverse transcribed into cDNA, and RQ-PCR was performed with newly designed primers and FAM-TAMRA-labeled probes (primer sequences available on request). The expression levels were corrected by comparison with control genes Abelson (ABL) and β-glucuronidase (GUS) as described previously (40). The expression values of the genes were normalized per-gene to zero mean and unit SD (z score).
Results
Isolation of stem cell-like and (precursor) B cell populations
The objective of this study was to correlate Ig gene rearrangements with gene expression profiles in human B cell differentiation. Therefore, only membrane markers were used for isolation of the relevant cell populations to reliably obtain both DNA and RNA from these cells (Fig. 1). Stem cell-like CD34+lin− (defined as CD3+CD13−CD19−CD33+CD56+) cells from UCB and pro-B cells (CD22+CD34+CD19+) could be isolated directly using previously established markers. However, for the remaining subpopulations, alternative markers CD10, CD20, and CD34 were required instead of cytoplasmic VpreB or cyIg. CyIg pre-B-I cells were isolated based on the phenotype CD19−CD34−CD10−CD20+. In addition, large cycling pre-B-II cells (CD19−CD34−CD10−CD20dimIgM−), small pre-B-II cells (CD19−CD34−CD10−CD20−), and immature B cells (CD19+CD34−CD10+CD20high) were isolated. Mature B lymphocytes (CD19+CD20+) were obtained from tonsillar tissue and separated in either SmIg+ or SmIg− populations based on membrane Ig L chain expression (Fig. 1).

Incomplete D_{H}–J_{H} gene rearrangements are initiated in pro-B cells, directly followed by complete V_{H}–D_{H}–J_{H} rearrangements in pre-B-I cells
The human IGK locus contains 76 rearrangeable V,27D, and 5J, gene segments. In addition to the V_{H}–J_{H} rearrangements there are two

FIGURE 2. Incomplete D_{H}–J_{H} gene rearrangements are initiated in pro-B cells and are directly followed by complete V_{H}–D_{H}–J_{H} rearrangements in pre-B-I cells. A, Organization of the human IGK locus; B, D_{H}–J_{H} rearrangements were quantified in all isolated subpopulations in two separate multiplex TaqMan RQ-PCR assays with primers and probes that detect most members of the six main D_{H} gene segment families and the six J_{H} gene segments; C, quantification of V_{H}–D_{H}–J_{H} rearrangements in two separate multiplex TaqMan RQ-PCR assays using primers that recognize most members of the seven V_{H} gene segment families and the six J_{H} gene segments. Data are obtained from duplicate experiments in two to three donor samples and are quantified (mean ± SD) relative to (a mixture of) monoallelically rearranged control cell lines.

IGK and IGL rearrangements are initiated in small pre-B-II cells, but V_{κ}–J_{κ} rearrangements are rare in mature Igα− cells
The human IGL locus contains 76 rearrangeable V_{κ}, 5 N, and 5 J_{κ} gene segments. In addition to the V_{κ}–J_{κ} rearrangements there are two
that detect most members of seven V\textsubscript{\textit{\textalpha{}}} types in a single multiplex TaqMan RQ-PCR assay with primers and probes. The vast majority of V\textsubscript{\textalpha{}} gene segment families and the five J\textsubscript{\textalpha{}} clusters (Fig. 4\textit{C}), of which J\textsubscript{1}–3 are used in 99\% of the V\textsubscript{\textalpha{}}–J\textsubscript{\textalpha{}} rearrangements (42). The V\textsubscript{\textalpha{}}–J\textalpha{1} rearrangements in precursor B cell subsets showed the same pattern as V\textsubscript{\textalpha{}}–Jk\textsubscript{\textalpha{}}1, but the levels of V\textsubscript{\textalpha{}}–J\textalpha{1} rearrangements appeared to be lower and were about one-half of the V\textsubscript{\textalpha{}}–J\textalpha{1} levels in mature Ig\textalpha{+} B cells (Fig. 4\textit{B}). The absence of V\textsubscript{\textalpha{}}–J\textalpha{1} rearrangements in mature Ig\textalpha{+} B cells (<5\%) indicated that most IGL rearrangements were initiated only if no functional Ig\textalpha{+} chain could be formed.

Complete V–(D)J rearrangements are followed by in-frame selection

After complete V–(D)J recombination, only B cells with in-frame rearrangements, expressing functional proteins, are selected (positive selection). Due to exonuclease activity and random nontemplated (N)-nucleotide addition by TdT, the size of the junctional region is variable, and only one-third of the rearrangements is in-frame. Using GeneScan analysis, the V\textsubscript{\textalpha{1}}–DJ\textsubscript{\textalpha{1}}, V\textsubscript{\textalpha{κ}}–Jk\textsubscript{\textalpha{1}}, and V\textalpha{λ}–J\textalpha{1} types of IGK-deleting rearrangements involving the \textkappa-deleting element (Kde) that make the IGK allele nonfunctional (41). V\textsubscript{\textalpha{κ}}–Kde rearrangements can delete an initial V\textsubscript{\textalpha{κ}}–Jk and the Ck\textsubscript{\textalpha{}} exon. Kde can also rearrange to the intronRSS, which is located between Js\textsubscript{\textalpha{}} and Ck\textsubscript{\textalpha{}}. The intronRSS–Kde rearrangement deletes Ck\textsubscript{\textalpha{}} but keeps the initial V\textsubscript{\textalpha{κ}}–Jk rearrangement present on the locus (Fig. 3\textit{A}).

In CD34\textsuperscript{+}lin\textsuperscript{−}, pro-B, pre-B-I, and large pre-B-II cells, V\textsubscript{\textalpha{κ}}–Jk rearrangements were barely detectable (Fig. 3\textit{B}). An enormous increase was found in small pre-B-II and immature B cells, which represented ~75\% of the frequency of V\textsubscript{\textalpha{κ}}–Jk rearrangements observed in mature Ig\textalpha{+} B cells. The majority of the V\textsubscript{\textalpha{κ}}–Kde and the intronRSS–Kde rearrangements were induced in pre-B-II small cells as well. In contrast to V\textsubscript{\textalpha{κ}}–Jk rearrangements, the Kde rearrangements occurred at low frequency in mature Ig\textalpha{+} B cells. The fact that they occurred at maximal levels in mature Ig\textalpha{+} cells (Fig. 3, C and D) indicates that most IGK alleles are deleted in mature Ig\textalpha{+} B cells.

The human IGL locus consists of 56 rearrangeable \textI\textkappa gene segments and 7 J\textkappa-CA clusters (Fig. 4\textit{A}), of which J\textkappa1–3 are used in 99\% of the V\textI–J\textkappa rearrangements (42). The V\textI–J\textkappa rearrangements in precursor B cell subsets showed the same pattern as V\textsubscript{\textalpha{κ}}–Jk\textsubscript{\textalpha{1}}, but the levels of V\textI–J\textkappa rearrangements appeared to be lower and were about one-half of the V\textI–J\textkappa levels in mature Ig\textkappa+ B cells (Fig. 4\textit{B}). The absence of V\textI–J\textkappa rearrangements in mature Ig\textkappa+ B cells (<5\%) indicated that most IGL rearrangements were initiated only if no functional Ig\textkappa+ chain could be formed.

Complete V–(D)J rearrangements are followed by in-frame selection

After complete V–(D)J recombination, only B cells with in-frame rearrangements, expressing functional proteins, are selected (positive selection). Due to exonuclease activity and random nontemplated (N)-nucleotide addition by TdT, the size of the junctional region is variable, and only one-third of the rearrangements is in-frame. Using GeneScan analysis, the V\textsubscript{\textI1}–DJ\textsubscript{\textI1}, V\textsubscript{\textkappa}–Jk\textsubscript{\textalpha{1}}, and V\textalpha{λ}–J\textalpha{1}
rearrangements were evaluated for the size of the junctional regions in the (precursor) B cell subsets in which these rearrangements were found previously with RQ-PCR. Complete VH–DJH rearrangements were present in pre-B-I cells, but the size distribution of the junctional regions showed a merely random pattern. In large pre-B-II cells and all subsequent stages, the GeneScan pattern was symmetrical unimodal with peaks at every third nucleotide (triplet peaks), representing in-frame rearrangements (Fig. 5A). These results show that in-frame selection of VH–DJH rearrangements apparently was completed in large pre-B-II cells.

The GeneScan patterns of the IGK/IGL gene rearrangements showed a smaller size of the junctional regions distribution than in case of IGH, due to the absence of Dγ1 gene segments and lower levels of N-nucleotide insertion (Fig. 5A). In agreement with the RQ-PCR results, pre-B-I and large pre-B-II cells hardly contained Vκ–Jκ and Vλ–Jλ rearrangements. Interestingly, the average size of the Vκ–Jκ junctional regions, but not the Vλ–Jλ junctional regions, was larger in the pre-B-I and large pre-B-II subsets compared with subsequent stages. The size distribution of the junctional regions in small pre-B-II cells was unimodal without triplet peaks, but in immature B cells some triplet peaks could be identified for both Vκ–Jκ and Vλ–Jλ, indicating that immature B cells have undergone in-frame selection. Mature Igκ+ B cells showed more clear triplet peaks for Vκ–Jκ rearrangements, whereas mature Igλ+ B cells showed triplet peaks for Vλ–Jλ rearrangements, but not for Vκ–Jκ rearrangements. The triplet peaks of the Vκ–Jκ and the Vλ–Jλ rearrangements in immature B cells were less easily identifiable compared with mature Igκ+ B cells and Igλ+ B cells, respectively, because the immature B cell subset consisted of a mixture of Igκ+ and Igλ+ cells.

Gene expression profiling of precursor B cell populations
Gene expression profiles of CD34-lin- cells from UCB and the five main precursor B cell subsets from BM were determined on several biological repeats using the Affymetrix HG-U133 set GeneChip arrays that contain 45,000 probe sets. The six subsets were distinguished by the significant differential expression of 5365 probe sets that were at least once differentially expressed between 2 subsequent stages of differentiation (Fig. 6A and Supplemental Table I). Hierarchical clustering was performed to group the expression patterns in 18 different clusters (Fig. 6, A and B).

The expression patterns were determined of a number of genes encoding proteins that have been described to function in human and/or murine precursor B cell differentiation (Fig. 6B and Supplemental Table II). The gene expression patterns of the markers that were used to isolate the six subsets (CD34, CD19, CD10, CD10, CD10, CD10, CD10).
CD22, CD20, and KI67) corresponded well with their protein expression (Fig. 6B and Supplemental Table II). Stem cell markers (FLT3, KIT), myeloid genes (MPO), and TCR germline transcripts were down-regulated from stem cells to early stages of B cell differentiation. Inversely, B cell commitment transcription factors (E2A, EBF, E2-2, PU.1, and PAX5) were found to be up-regulated early in B cell differentiation. These transcription factors are involved in opening of Ig loci and positively regulate transcription of (pre-)-BCR complex and signaling molecules. Polycomb protein EZH2 and chromatin remodeling protein BRG1, which are implicated in complete IGH gene rearrangements, were highly expressed in pre-B-I cells. IGHM, IGKC, and IGLC transcripts were already up-regulated early in B cell differentiation, even before Ig gene rearrangements were found. These early transcripts were most likely germline transcripts. Transcripts of (pre-) BCR complex members LI4.1, VPREB1, CD79A, and CD79B were significantly up-regulated in large pre-B-II cells, and the same was seen for many (pre-) BCR signaling molecules, such as BTK, SYK, and LYN. HMG-box proteins LEF1 and SOX4 were highly expressed in both pre-B-I and pre-B-II small, whereas IRF4 and IRF8 were up-regulated in large pre-B-II and immature B cells, respectively. A number of genes were recognized by multiple probe sets, which showed a slightly different expression pattern and were therefore in some cases assigned to more than one cluster.

The expression patterns of RAG1 and RAG2 showed peaks in pre-B-I and small pre-B-II cells, the populations in which complete IGH and IGK/IGL gene rearrangements are formed, respectively. HMGB1 and HMGB2 expression was high throughout B cell development with a peak in pre-B-I and large pre-B-II cells. TdT expression was high only in pro-B and pre-B-I cells and not in small pre-B-II cells that undergo IGH/IGL gene rearrangements. Genes encoding proteins involved in recognition of dsDNA breaks induced by cleavage of the RAG proteins (KU70, KU80, DNA-PKcs) were found to be specifically up-regulated in one or more stages of B cell differentiation undergoing Ig gene rearrangements. DNA ligase IV (LIG4) was the only member of the NHEJ dsDNA break repair machinery that was differentially expressed in parallel to RAG1 and RAG2 with peaks of expression in pre-B-I and small pre-B-II cells.

We have thus shown that the expression of well-defined B cell differentiation and V(D)J recombination-associated genes correlated well with the rearrangement status of the precursor B cell subsets. This provides a firm basis to identify molecules that are likely to be involved in the initiation and regulation of the Ig gene rearrangements.

Initiation and regulation of Ig gene rearrangements

A restricted set of GO terms of the Gene Ontology Consortium was used to select for categories of genes required for V(D)J recombination: 1) DNA-binding transcription factors; 2) enzymes involved in DNA methylation; 3) histone and chromatin remodeling factors; 4) recombination-related enzymes; 5) DNA repair molecules. This yielded a set of 476 genes that were at least once differentially expressed between two subsequent stages in differentiation (Fig. 7 and Supplemental Table III). From CD34+lin to pro-B cells, about the same number of genes (~35 genes) were up-regulated as were down-regulated. This was clearly not the case for the transition from pro-B to pre-B-I where complete IGH rearrangements are initiated; almost 4 times as many genes were up-regulated (40 genes) than down-regulated (143 genes). In contrast, twice as many selected genes were down-regulated rather than up-regulated in the transitions from pre-B-I to large pre-B-II (155 vs 85 genes) and from small pre-B-II to immature B (71 vs 39

**FIGURE 6.** Gene expression profiling of human CD34+lin and precursor B cell subsets. Heat map with the normalized data of significantly differentially expressed probe sets. The expression patterns were grouped into 18 clusters. Expression pattern of all probe sets per cluster are shown as mean with SD. For each cluster, the total number of genes and the names of several genes reported to be involved in precursor B cell differentiation or commitment to other hematopoietic lineages are shown. A few genes can be found in several clusters, when multiple probe sets representing the same gene showed different expression patterns; within each cluster a gene is displayed only once. The following genes were studied as well but were not found to be differentially expressed: ARTEMIS, HEL, ID1, ID4, NOTCH1, PLCG2, STAT5A, TCRA/D, XRCC4, IKAROS, HELIOS, AIOLOS.
DH–JH, V H–DJH, and VL–JL rearrangements take place, respectively, peaks in pro-B, pre-B-I, and/or small pre-B-II cells during which compared with subsets undergoing V(D)J recombination.

Interestingly, this indicates a trend that many transcription-regulated genes are down-regulated in subsets undergoing selection. It was possible to identify 16 novel genes that are likely to be involved in initiation and regulation of Ig gene rearrangements within the large set of differentially expressed probe sets.

Discussion

In this study, for the first time cells representing the five main stages of precursor B cell differentiation were isolated from human BM and were completely analyzed at the DNA level for IGH, IGK, and IGL gene rearrangements and at the RNA level for genomewide gene expression profiles. Incomplete D J H rearrangements were initiated in pro-B cells, whereas V H–DJH rearrangements were initiated in pre-B-I cells. Large cycling pre-B cells were selected for in frame V H–DJH rearrangements and small pre-B-II cells initiated IGK/IGL gene rearrangements. Finally, immature B cells contained in frame IGK/IGL gene rearrangements indicating that they had undergone positive selection for the presence of productive rearrangements that result in membrane expression of a complete BCR (summarized in Fig. 8).

Incomplete IGH rearrangements were already abundantly present in pro-B cells, whereas RAG1 and RAG2 expression levels were hardly up-regulated (Fig. 6B and Fig. 8B). These data suggest that D J H rearrangements can be initiated with much lower levels of RAG1/RAG2 than complete V H–DJH and IGK/IGL rearrangements. In contrast to human pro-B cells, murine CD19− pro-B cells have germline IGH alleles (43), most likely because different phenotypic markers have been used for their isolation. Complete V H to DJH rearrangements started in CD19+ pre-B-I cells. This is in line with up-regulated PAX5 expression, which is critical for the
initiation of complete IGH gene rearrangements and for the expression of CD19 (44).

Rearrangements of the three types of IGK/IGL gene rearrangements, Vκ–Jκ, Kde, and Vα–Jα, were all initiated massively in small pre-B-II cells. Apparently, the current markers are not sufficient to separate precursor B cells according to these three consecutive IGK/IGL rearrangement steps. Low levels of IGK/IGL gene rearrangements were present in pre-B-I and large pre-B-II cells. The average size of the Vκ/H9260–Jκ/H9260 junctional regions was larger in pre-B-I and large pre-B-II cells than in pre-B-II small cells (Fig. 5B), indicating more N-nucleotide addition. This was in line with the higher TdT expression in pre-B-I cells than in large pre-B-II cells.

Table I. Transcription-related genes in clusters that correlate with the initiation of Ig gene rearrangements

<table>
<thead>
<tr>
<th>Peak in Expression</th>
<th>IGH</th>
<th>IGH</th>
<th>IGH</th>
<th>IGH</th>
<th>IGH</th>
<th>IGK/IGL</th>
<th>IGH</th>
<th>IGK/IGL</th>
<th>IGH</th>
<th>IGK/IGL</th>
<th>IGH</th>
<th>IGK/IGL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category/Cluster</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. DNA-binding TFa</td>
<td>10</td>
<td>11</td>
<td>17</td>
<td>18</td>
<td>12</td>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E2-2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HMG1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IRF1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC148203</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MYB2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UHRF1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WHSC1L</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ZCCHC7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ZNF367</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. DNA methylation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HEMK</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAP30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMARCA4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Chromatin and</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>histone remodeling</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BRG1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H2AFY</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PHF1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Recombination</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>POLS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TdT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RAG1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. DNA repair</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>XRN2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FancI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KU70</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSH2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NBS1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRKDC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TREX1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KUB3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>POLI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ANKR1D7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Btg2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Csnk1E</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NcoA6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rad21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ubev2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NEIL1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CSNK1E</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gadd45g</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tdp1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Xrn2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TF, Transcription factor.

Genes set in bold are 16 likely candidate genes involved in initiation and/or regulation of Ig gene rearrangements.
cells. We suggest that the low levels of Vκ-Jκ rearrangements found before the pre-B-II small stage indicate early opening of the IGK locus before down-regulation of TdT. Interestingly, this phenomenon was not clearly found for IGL gene rearrangements.

The amount of Kde and Vα-Jα rearrangements in mature Igκ+ cells was low compared with mature Igλ+ cells, whereas Vκ-Jκ was high in both. This confirms the hierarchical order that has been found in previous studies (31, 45); Kde and Vα-Jα rearrangements are induced only after Vκ-Jκ rearrangements.

Correlating the Ig gene rearrangement status to the genome-wide gene expression profiles of all five precursor B cell subsets enabled us to study the networks of genes involved in the initiation of different types of Ig gene rearrangements. The expression of RAG1 and RAG2 showed peaks in pre-B-I and small pre-B-II cells, which correlated well with the initiation of complete IGH and IGK/IGL rearrangements. Interestingly, LIG4 was the only member of the NHEJ dsDNA break repair machinery that was also up-regulated in these stages. In addition to RAG1 and RAG2, apparently also higher levels of LIG4 are required for V(D)J recombination.

The use of mouse models has contributed enormously to the understanding of lymphocyte differentiation. However, it is still unclear which factors determine the hierarchical order of Ig gene rearrangements. In this study, a selection was made within the differentially expressed genes that regulate transcription to identify the networks of factors involved in the process of V(D)J recombination. The E2A, EBF, and E2-2 transcription factors are known to be involved in the initiation of V(D)J recombination on the IGH locus (8–10). Transcripts encoding these proteins were directly up-regulated in the pro-B cell stage that contains abundant Dij−Jij rearrangements. Furthermore, pro-B cells showed up-regulated expression of genes encoding EZH2 and SMARCA4, which modify the chromatin structure, thereby making the IGH locus accessible for V(D)J recombination (21, 22).

PAX5 is required for Vκ-Dij rearrangements (11) and was found to be strongly up-regulated in pre-B-I. Furthermore, RAG1, RAG2, HMGB1, HMGB2, LIG4, and TdT transcripts were strongly up-regulated in pre-B-I cells. This could indicate that higher levels of the recombinase machinery are needed for initiation of Vκ to DJκ gene rearrangements. In mice, IL-7R signaling induces Vκ-Dij rearrangements via STAT5 activation (46). STAT5B gene expression was indeed up-regulated in human pre-B-I cells, but IL7RA was not expressed before the stage of large pre-B-II cells. Apparently, STAT5 might be involved in initiation of Vκ-Dij rearrangements in humans as well. However, STAT5 is probably not activated by IL-7Rα but by an alternative signaling pathway.

In addition to the well-described genes, we identified a number of genes the expression pattern of which correlated with the initiation of complete IGH gene rearrangements. COPEB was significantly up-regulated in pre-B-I cells. In thymocytes, COPEB was found to bind to the Dβ1 promoter, but in contrast to KLF5 it did not enhance the promoter activity in murine pro-T cell lines (47). The up-regulation of COPEB in pre-B-I cells suggests that it plays a role in IGH gene rearrangements by binding to the enhancer in the same way KLF5 can do this for TRCB, thereby opening the IGH locus. Furthermore, the expression of three Ets family protein

FIGURE 8. RQ-PCR analysis confirms the correlation of 16 newly identified transcription factors with the initiation of Ig gene rearrangements in precursor B cell differentiation. A, Hypothetical scheme of precursor B cell differentiation, summarizing the quantitative Ig gene rearrangement data. The horizontal bars represent the rearrangement of the Ig loci. The arrows show the timing of in frame selection of Vγ–DJγ and Vκ–Jκ/VA–Ja after the pre-B-I and pre-B-II small cells stage, respectively. B, Heat map with the normalized RQ-PCR data of RAG1 and RAG2, as well as the 16 candidate genes involved in initiation and/or regulation of Ig gene rearrangements. The expression patterns fully confirm the patterns observed with DNA microarray analysis with peaks in expression in one or more stages that underlie Ig gene rearrangements. Slight differences were seen for ELF1 and OCAB when compared with the cluster patterns, because the cluster patterns were the averages of multiple genes expression patterns.
transcripts, ETS2, ELK3, and ERG, was up-regulated specifically in pre-B-I cells. Of these, ERG has been found to bind to a promoter sequence in IGH and was able to activate a reporter construct synergistically with E2A splice variant E12 (48), making it a likely cofactor of E box proteins for IGH locus opening in pre-B-I cells.

The expression of RAG transcripts was found to be down-regulated in large pre-B-II cells and subsequently up-regulated in small pre-B-II cells that initiated IGH/IGL gene rearrangements. This up-regulation of the RAG transcripts was accompanied by a up-regulation of IRF4, which has been described to be involved in initiation of Ig L chain rearrangements (19, 20). In addition, E2A, EBF, E2-2, and PAX5 were highly expressed in small pre-B-II cells. EBF and E2A splice variant E47 have been shown to induce Vκ-Jκ rearrangements in a nonlymphoid cell line (10), and PAX5 is required for IGK germline transcription (12). COPEB was expressed again after down-regulation in large pre-B-II cells, suggesting that it might be involved in both IGH and IGL/IGK gene rearrangements. Ets family factors ETS1 and SPIB were specifically up-regulated in small pre-B-II cells. In mice, loss of SPIB did not affect B cell differentiation in BM (49), but pro-B cell lines derived from PU.1/SPIB double knockout mice showed decreased IGL germline transcription (50). Rearrangements of the IGL locus are strictly regulated and usually occur only after the production of a functional IGK allele has failed. SPIB might act to specifically target the recombination machinery to the IGL locus. HIVEP3 showed specific up-regulation in the small pre-B-II stage. Its protein product is capable of binding the E box transcriptional enhancer motif and to the canonical RSS heptamer and nonamer (51).

HIVEP3 might function to make the RSS elements accessible for recombination, thereby creating the opportunity for many different gene segments to be recombinated. Although it was found to be dispensable for precursor B cell differentiation (52), as yet studies have addressed a function in gene segment selection in IGH/IGL gene rearrangements.

With the currently available markers, it was not possible to separate the initiation of Vκ-Jκ, Kde, and VA-Jα rearrangements, because they seem to coincide in a single stage of precursor B cell differentiation. However, it can be anticipated that different networks of genes regulate these three consecutive types of rearrangements, thereby creating the hierarchical order found in mature B lymphocytes (31, 45).

In summary, we determined the quantitative and qualitative status of all major Ig gene rearrangements in human precursor B cell subsets. On the basis of the Ig gene rearrangement patterns, we divided human precursor B cell differentiation into five functional stages. Careful analysis of genome-wide expression profiles of these 5 subsets enabled us to identify 16 novel candidate genes for initiation and regulation of IGH and IGL/IGK gene rearrangements, thereby providing insight in the mechanism of precursor B cell differentiation. This detailed analysis represents an excellent template for studies on oncogenic transformation in precursor B cell acute lymphoblastic leukemia and B cell differentiation blocks in primary Ab deficiencies.

Acknowledgments

We thank W. M. Comans-Bitter for assistance with preparing the figures.

Disclosures

The authors have no financial conflict of interest.

References


Ig GENE REARRANGEMENTS IN HUMAN PRECURSOR B CELLS


