B-cell Development and Primary Antibody Deficiencies

B-celontwikkeling en primaire antistofdeficiënties

ISBN-13: 978-90-73436-76-3

"The quest of the B cell"
Developing B cells have a long way to go before they are ready to encounter disease-causing pathogens. In order to make it a successful journey, multiple factors are required, without which the guest will fail. The ultimate dream of the B cell is to fight a pathogen and become an antigen-experienced B cell that can generate a fast and strong immune response, which contributes to antigen-specific immunity.

No part of this thesis may be reproduced or transmitted in any form by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system, without permission in writing from the publisher (M.C. van Zelm, Department of Immunology, Erasmus MC, P.O. Box 2040, 3000 CA Rotterdam, the Netherlands).

B-cell Development and Primary Antibody Deficiencies

B-celontwikkeling en primaire antistofdeficiënties

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Erasmus Universiteit Rotterdam
op gezag van de rector magnificus
Prof.dr. S.W.J. Lamberts
en volgens besluit van het College voor Promoties.
De openbare verdediging zal plaatsvinden op
woensdag 20 juni 2007 om 13.45 uur

door

Menno Cornelis van Zelm

geboren te Utrecht

PROMOTIECOMMISSIE

Promotoren:	Prof.dr. J.J.M. van Dongen Prof.dr. R. de Groot
Overige leden:	Prof.dr. H. Hooijkaas Prof.dr. C. Murre Prof.dr. C.J.M. van Noesel
Copromotor:	Dr. M. van der Burg

The studies described in this thesis were performed at the Department of Immunology, Erasmus MC, Rotterdam, the Netherlands and partly at the division of Biological Sciences, University of California San Diego, La Jolla, CA, USA.

The studies were financially supported by 'Sophia Kinderziekenhuis Fonds' (SKF; grant 349), and partly by Ter Meulen Fund - Royal Netherlands Academy of Arts and Sciences (KNAW), Trustfonds - Erasmus University Rotterdam and the Dutch Society for Immunology (NVvI).

The printing of the thesis was supported by Erasmus University Rotterdam, J.E. Jurriaanse Stichting, Stichting Kind & Afweer, Baxter B.V., Dr. Ir. Van de Laar Stichting, Beckman Coulter Nederland B.V. and BD Biosciences.

Illustrations : W. Marieke Comans-Bitter
Printing : Ridderprint Offsetdrukkerij B.V., Ridderkerk
Cover : Frans Tutert
Lay-out : Wendy Netten

Voor mijn ouders

B-cell Development and Primary Antibody Deficiencies

B-celontwikkeling en primaire antistofdeficiënties

CONTENTS

Chapter I	**General Introduction**	**9**
Chapter II	**Ig Gene Rearrangement Steps are Initiated in Early Human Precursor-B-cell Subsets and Correlate with Specific Transcription Factor Expression** *J Immunol 2005;175(9):5912-22*	**41**
Chapter III	**3D-Architecture of the Immunoglobulin Heavy Chain Locus** *Submitted*	**67**
Chapter IV	**Gross Deletions involving *IGHM*, *BTK* or *Artemis*: A model for Genomic Lesions Mediated by Transposable Elements** *Submitted*	**93**
Chapter V	**An Antibody-Deficiency Syndrome Due to Mutations in the *CD19* Gene** *N Engl J Med 2006;354(18):1901-12*	**115**
Chapter VI	**Replication History of B lymphocytes Reveals Homeostatic Proliferation and Extensive Antigen-induced B-cell Expansion** *J Exp Med 2007;204(3):645-55*	**133**
Chapter VII	**General Discussion**	**153**

Chapter VIII Appendices **167**
1 PCR primers and TaqMan probes for quantification of Ig gene rearrangements 168
2 PCR primers for GeneScan analysis of complete *IGH*, *IGK* and *IGL* rearrangements 170
3 PCR primers and TaqMan probes for quantification of expression levels of 16 candidate genes for induction of Ig gene rearrangements 171
4 Full width at half the maximum (FWHM) values of the point spread function for the indicated colors 172
5. Chromatic shifts 172
6 Resolution equivalents before and after application of chromatic corrections 172
7 *IGHM* gross deletion primers for deletion-scanning and breakpoint PCR 173
8 *BTK* gross deletion primers for deletion-scanning and breakpoint PCR 177
9 *Artemis* gross deletion primers for deletion-scanning and breakpoint PCR 178
10 PCR primers for amplification and sequencing of *CD19* coding regions from genomic DNA 179
11 PCR primers and TaqMan probe for quantification of *CD19* transcripts 179
12 PCR primers and TaqMan probes for KREC and IgκREHMA assays 180

Abbreviations **181**

Summary **183**

Samenvatting **186**

Dankwoord **190**

Curriculum Vitae **192**

Publications **194**

I

General Introduction

The immune system of higher vertebrates, such as humans, is a complex interplay of cells and protein products that protects the body from invading micro-organisms. The first line of defense is formed by the skin and mucous membranes. The second line consists of proteins (e.g. complement) and immune cells (e.g. macrophages, granulocytes, natural killer (NK) cells) that are part of the innate immune system. T and B lymphocytes are part of the specific adaptive immune system: the third line of defense. These lymphocytes adapt their response during an infection to improve recognition of the pathogen and this improved response is then retained after the pathogen has been eliminated in the form of an immunological memory. Classically, the cells of the innate immune system were thought to react a-specifically, whereas the adaptive immune cells carried out a specific immune response. However, more recently innate immune cells were found to carry evolutionarily conserved receptors, named Toll-like receptors (TLRs). These receptors play a crucial role in early host defense against invading pathogens and are specific enough to discriminate self and pathogens.

B lymphocytes contribute to the immune system by production of antigen-specific antibodies (immunoglobulins; Ig). The antibody response to a pathogen is the classical example to demonstrate the enormous capacity and specificity of the adaptive immune response (Figure 1). Similar to most immune cells, B lymphocytes are generated throughout life from hematopoietic stem cells in bone marrow (Figure 2). Every newly generated B-lymphocyte carries a specific Ig receptor, which has been generated during differentiation in bone marrow

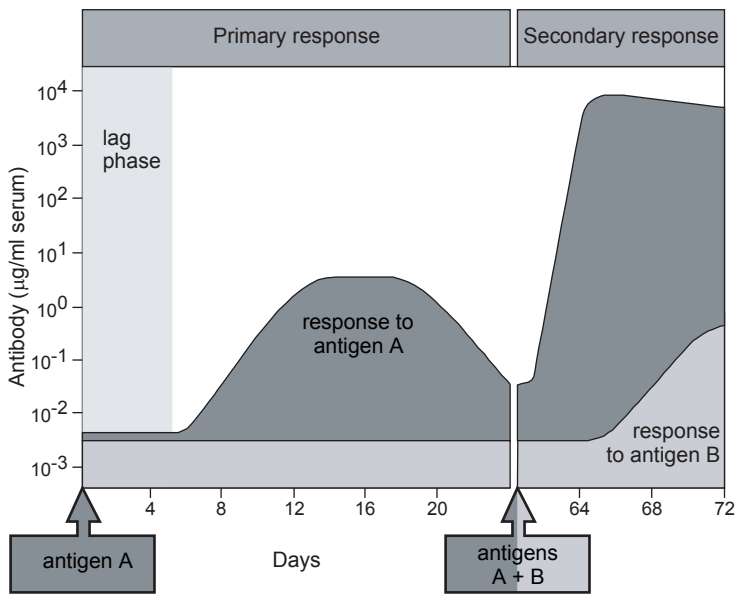


Figure 1. The course of a typical immune response.
Upon first encounter with antigen A, an immune response is initiated and antibodies against antigen A are produced after a lag phase of several days. The antibody levels reach a plateau and then decrease. When the animal is later challenged with antigen A, a very rapid and intense response occurs. The response to antigen B at the same time of the booster with antigen A, resembles the initial and primary response to antigen A. This illustrates the antigen-specific build up of memory. Figure adapted from Janeway *et al.* Immunobiology, 2001, Garland Science.

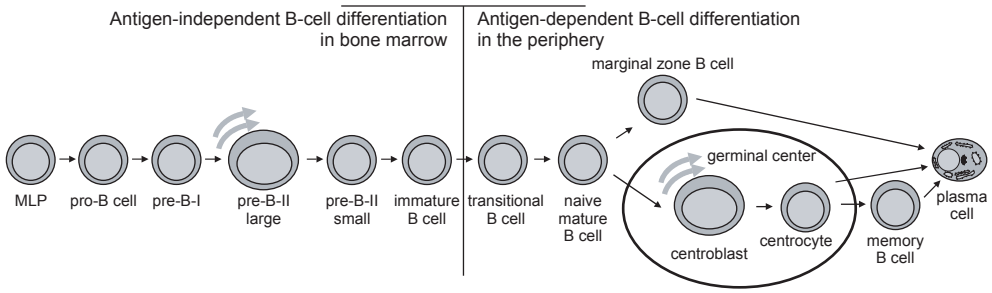


Figure 2. Scheme of B-cell differentiation.
B cells develop from hematopoietic stem cells through various precursor stages in bone marrow. In the periphery, naive mature B lymphocytes can respond in absence of T-cell help to antigen in the marginal zone, or they undergo a more powerful T-cell dependent response in the germinal center reaction. Eventually, Ig-secreting plasma cells and long-lived memory B cells are formed.

by somatic rearrangement of the gene complexes encoding the Ig heavy chain (IgH) and Ig light chain (Igκ or Igλ) molecules. When it recognizes an antigen with its specific Ig receptor, the B-lymphocyte undergoes clonal expansion to generate daughter cells that differentiate into either memory B cells or plasma cells, which produce and excrete soluble Ig molecules. As shown in Figure 1, this primary response takes a couple of days and solely targets one specific pathogen. Another property of the specific immune system is the build-up of memory. When the same pathogen is encountered for the second time, the immune system will respond quicker and more powerful, both with respect to the number and the affinity of the antibodies. However, when a different pathogen is encountered that has not been 'seen' before, a primary response will be initiated.

Generation of a large repertoire of B lymphocytes and the response of one of these B lymphocytes to antigen is a multi-step process. Multiple proteins are critical for proper B-cell differentiation in bone marrow and the B-cell antigen response.The lack of one of these proteins leads to a problem with the humoral immunity of the individual, which is seen in patients with primary antibody deficiency diseases. These children or young adults carry rare inherited disorders and are subject to multiple, recurrent (mainly) bacterial infections.

This thesis focuses on the process of B-cell differentiation and the defects that lead to primary antibody deficiency diseases.

1. HUMAN B-CELL DIFFERENTIATION

B-cell commitment

Hematopoietic stem cells (HSC) in bone marrow are long-lived and self-renewing. They generate multilineage progenitor cells (MLP) that are capable of developing into erythroid, myeloid and lymphoid cells (Figure 3). While they are capable of developing into each of these cell types, every single MLP will only develop into one specific cell type.

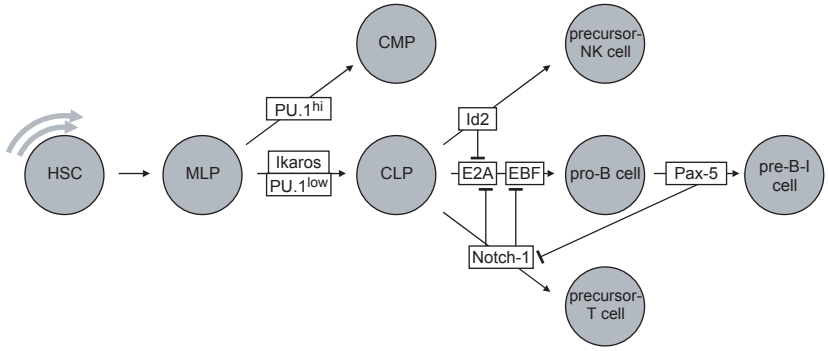


Figure 3. Schematic overview of differentiation and commitment from hematopoietic stem cells (HSC) to the B-cell lineage.
Multilineage progenitor cells (MLPs) derive from HSC and subsequently depend on Ikaros and low PU.1 expression to develop into common lymphoid progenitors (CLPs). The transcription factors E2A and EBF support further differentiation into pro-B cells, whereas their expression is inhibited in precursor-NK and precursor-T cells by Id2 and Notch1, respectively. Pax5 ensures commitment to the B-cell lineage. Figure adapted from Busslinger.[5]

This progressive restriction in developmental potential is thought to occur in intermediate steps.[1] Due to the low numbers of early progenitor cells and the fact that they are located in bone marrow, limited data is available from B-cell commitment in man. In contrast, many studies have been performed in mice. Therefore, mainly data obtained from mouse models will be discussed in this paragraph.

The MLPs generate populations restricted to either lymphoid (common lymphoid progenitor; CLP) or erythroid/myeloid (common myeloid progenitor; CMP) cell lineages (Figure 3).[2, 3] CMPs can develop into erythrocytes and myeloid cells, e.g. monocytes and granulocytes. CLPs give rise to natural killer (NK) cells, T cells and B cells. Commitment to the B-cell lineage takes place in several consecutive steps. B-lineage specification occurs already at the MLP/CLP stage.[4] Although B-cell fate is already induced early, the progenitor cell keeps a high degree of lineage plasticity. The myeloid potential is not lost until the pro-B-cell stage, and B/T lineage plasticity persists until the pre-B-I cell stage.[4] From this stage onwards the cell is fully committed to the B-cell lineage (Figure 4).

Multiple proteins, particularly transcription factors, are involved in controlling cell fate decisions in HSC, MLP, CMP and CLP (reviewed in refs 5, 6). Ikaros and PU.1 act in parallel pathways to control the development of CLP from MLP.[7, 8] This is done in part by regulating the expression of essential signaling receptors (Flt3, c-Kit, and IL-7Rα).[8, 9]

The generation of pro-B cells depends on transcription factors encoded by the *E2A* and *EBF* genes. The *E2A* gene encodes the two splice variants E12 and E47, which belong to the basic helix-loop-helix (bHLH) family of transcription factors.[10, 11] Together with E2-2 and HEB, these proteins belong to class 1 bHLH proteins, also called E-proteins. E12 and E47 are widely expressed and bind specific DNA sequences, E-boxes, as homo- or heterodimers with other E-proteins. In developing B cells, mainly E47 homodimers are detected. It is currently

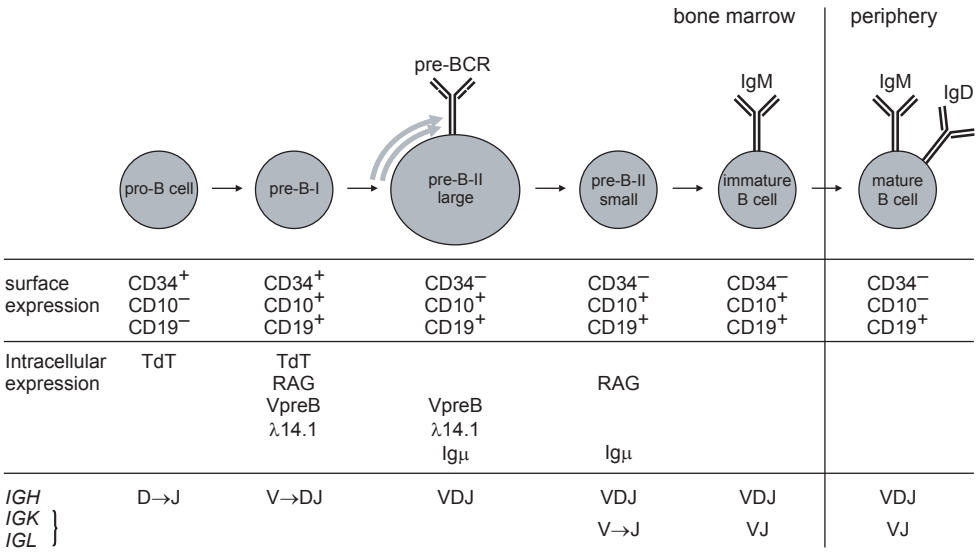


Figure 4. Scheme of human precursor-B-cell differentiation.
Precursor-B cells differentiate through five functionally distinct stages in bone marrow during which they rearrange their Ig genes before they migrate to the periphery with a functional BCR. The distinct stages can be identified by membrane and cytoplasmic expression patterns of several proteins.

unclear what factors regulate E47 expression and homodimerization in pro-B cells; the latter is probably controlled by hypophosphorylation of N-terminal serine residues.[12]

Within the hematopoietic system, EBF is exclusively expressed in the B-cell lineage and binds as homodimer to conserved palindromic DNA sequences.[13-15] E2A proteins and PU.1 are important regulators of EBF expression.[16, 17] E47 and EBF coordinately regulate the expression of multiple genes required for proper B-cell differentiation, including components of the pre-BCR and the transcription factor Pax-5.[18] EBF and Pax-5 initiate a B-cell specific transcription program and suppress the expression of genes required for development of other hematopoietic lineages (e.g. Notch1, Flt3), thereby committing the pro-B cell to the B-cell lineage.[19-22] Analogously, regulating factors of other lymphoid lineages, such as Notch1 in early T-cell progenitors and Id2 in precursor-NK cells, can suppress the genes that regulate early B-cell differentiation (Figure 3).

IGH gene rearrangements

The primary objective of every precursor-B cell in bone marrow is to generate a unique B-cell antigen receptor (BCR). In doing so, every mature B-lymphocyte in peripheral lymphoid organs contributes to the enormous antigen-recognition repertoire, which is estimated to be over 10^{12}. In order to generate a specific antigen receptor, every precursor-B cell initiates V(D)J recombination on the Ig receptor encoding loci (*IGH*, *IGK* and *IGL*). In contrast to other genes, these loci lack a functional first coding exon. Instead they carry multiple variable

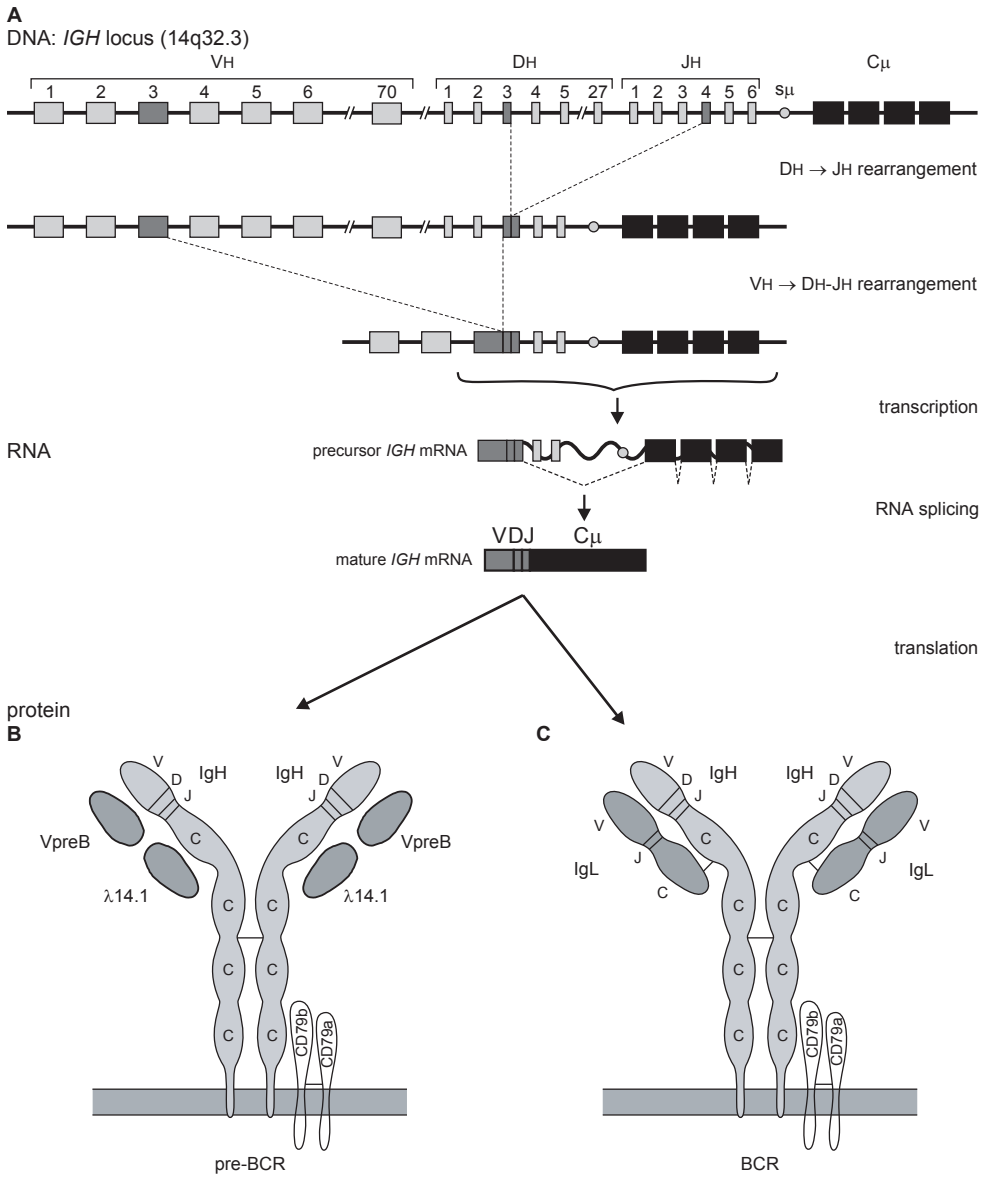


Figure 5. Schematic representation of *IGH* gene rearrangement.
A. First, a DH gene segment is joined to a JH gene segment. Subsequently, a VH gene segment is coupled to the DJH joint. The VDJ exon is transcribed and spliced to the Cμ exons. **B, C.** A functional in-frame Igμ protein will be transported to the plasma membrane with anchoring molecules CD79a and CD79b, and surrogate light chain molecules λ14.1 and VpreB (pre-BCR in large pre-B-II cells), or a functional Ig light chain protein (BCR in immature and mature B cells).

(V), diversity (D) and joining (J) gene segments (Figure 5). A functional first exon is formed by genomic coupling of one V, one D and one J gene segment.[23] Because *IGK* and *IGL* lack D gene segments, direct V–J gene rearrangement is initiated on these loci.

The first Ig gene rearrangements that occur during precursor-B-cell differentiation are DH–JH rearrangements.[24, 25] These rearrangements are initiated on both *IGH* alleles.[26, 27] In mouse models, DH–JH rearrangements are already detectable in some CLP, but the majority is initiated in pro-B cells.[4] Although CLP have the potential to differentiate into multiple cell lineages, DH–JH rearrangements are fairly specific for B lymphocytes: only low levels of DH–JH have been found in T cells, NK cells and myeloid cells. Furthermore, no VH–DJH, Ig light chain or TCR gene rearrangements are induced in CLPs and pro-B cells indicating that additional factors specifically target the common V(D)J recombinase machinery to DH and JH gene segments. The transcription factors E2A and EBF appear to have an important role in making the *IGH* loci accessible for DH–JH recombination. This is mediated in part by the intron enhancer (iEμ) downstream of the JH gene segments, which consists of several conserved motifs, among which are five E-boxes.

Complete VH–DJH rearrangements are initiated in pre-B-I cells (Figure 4). Initially, one allele is poised to rearrange and the other allele only rearranges when rearrangement of the first allele was not successful. Because V and J gene segments have one reading frame and stop codons can be introduced in the coding joint, less than one-third of the rearrangements can generate functional proteins. VH–DJH rearrangements are highly B-cell specific and are not present in other cell lineages. Murine studies have indicated prominent roles for the Pax5 transcription factor in making the VH gene segment region accessible to the RAG proteins. Furthermore, signaling via the IL-7R and downstream transcription factor Stat5 was found to be crucial for initiation of VH–DJH rearrangements in mouse models.

Prior to V(D)J recombination, two gene segments have to be spatially co-localized. Since the Ig loci can be over 1 Mb in size, gene segments poised to rearrange can be located in genomically distant locations. Recent studies in mouse models aimed to visualize the spatial co-localization of gene segments in the *IGH* locus in cells poised to rearrange.[28, 29] Using 3D fluorescence *in situ* hybridization techniques, the *IGH* locus was shown to undergo large-scale compaction and looping of distant regions. These processes are independent of RAG-2 and are probably mediated by transcription factor activity.[29] The nature of the compaction and looping as well as the DNA motifs mediating these processes are topics for further research.

Molecular mechanisms of V(D)J recombination

V(D)J recombination is a highly specific process that is only initiated in precursor-B and precursor-T cells in order to generate antigen receptors. During Ig gene rearrangements, double stranded DNA breaks are induced. These breaks are unwanted in other genes, thus requiring very specific targeting of the Ig loci.

The double stranded DNA breaks are induced by a heterodimer composed of the recombinase activating gene protein products RAG-1 and RAG-2 (Figure 6). This RAG-1/RAG-2 dimer specifically recognizes a DNA sequence, the recombination signal sequence (RSS), composed of conserved heptamer and nonamer sequences. These conserved sequences are separated by either a 12-bp or a 23-bp DNA sequence (spacer). All Ig gene segments are flanked by an RSS, and rearrangements commonly occur between gene segments of which

one is flanked by a 12-bp RSS and the other by a 23-bp RSS.[30] This is known as the 12-23 rule. The RAG-dimers juxtapose the two gene segments poised to rearrange and nick the DNA of both at the border of the gene segment and flanking RSS.[31-33] The 3´OH at each nick attacks the antiparallel strand to create a DNA hairpin at the gene segment. Blunt ends remain on the RSS sequences and are directly joined by non-homologous end joining (NHEJ) to create a circular excision product. The hairpins on the coding end are processed prior to ligation by NHEJ components.[34] The proteins Ku70 and Ku80 first bind the coding ends and recruit the

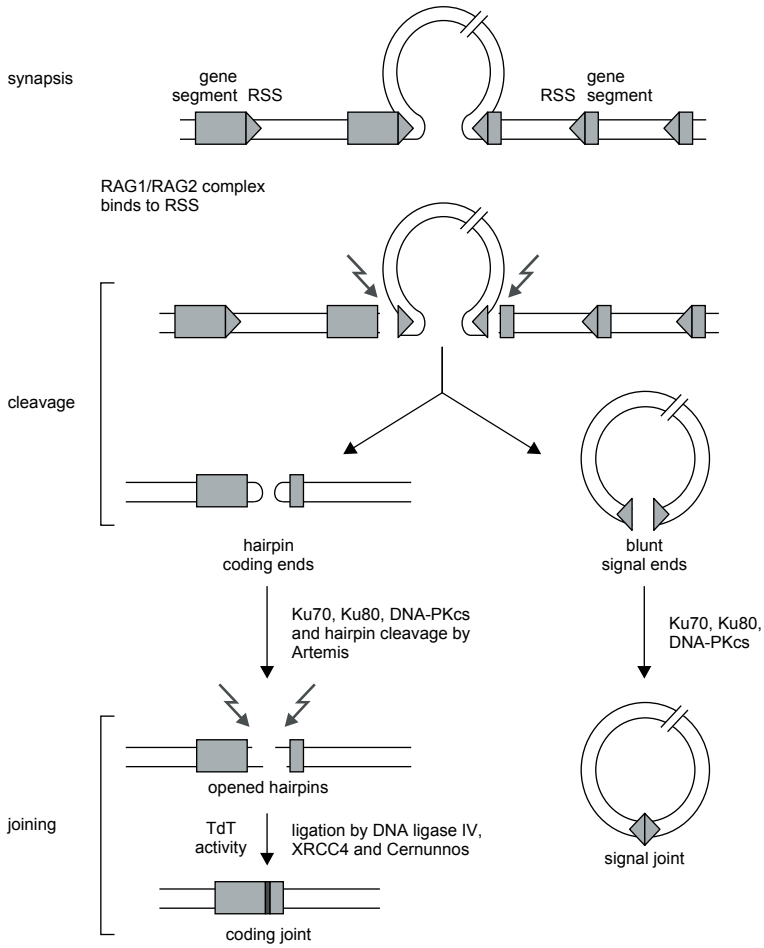


Figure 6. Schematic overview of V(D)J recombination.
The RAG proteins cleave two gene segments at the border of the RSSs. This results in two hairpinned coding end and two blunt signal ends. The blunt signal ends are directly joined by a complex of Ku70, Ku80 and DNA-PKcs. These proteins also bind the two coding ends before Artemis cleaves the hairpins. Subsequently, the coding ends are processed by endonuclease and TdT activity. Finally, a complex of DNA ligase IV, XRCC4 and Cernunnos is required to ligate the ends into a coding joint.

catalytic subunit of DNA-dependent protein kinase (DNA-PKcs) and Artemis.[35, 36] The protein Artemis nicks the hairpin preferentially at the tip or 1-2 bp 5' of the tip.[37, 38] When the nick is introduced 5'of the tip, this results in a 3' overhang. Complementary nucleotides can be inserted from the 5' end to generate a blunt end, resulting in the generation of palindromic (P-)nucleotides. Furthermore, exonuclease proteins can remove nucleotides, whereas the terminal deoxynucleotidyl transferase (TdT) protein can add random (N-)nucleotides.[39, 40] Finally, the Ku heterodimer recruits a protein complex, including XRCC4, DNA ligase IV and Cernunnos (XLF), that ligates the coding ends.[41-43] The processing of the coding ends before ligation results in a unique coding joint sequence.

Pre-B-cell receptor signaling

Once a VDJH exon is formed by Ig gene rearrangements, RNA transcripts are produced and the exon is spliced to the Igμ encoding exons (Figure 5). If the resulting Igμ heavy chain protein is functional and is able to pair with the surrogate light chain proteins λ14.1 and VpreB, this pre-BCR complex is expressed on the plasma membrane of the cell with anchoring molecules CD79a and CD79b. Expression of a functional pre-BCR is an important checkpoint in precursor-B-cell differentiation.[44, 45] The receptor induces multiple processes via downstream signaling pathways (Figure 7). First, further rearrangement of the *IGH* alleles is

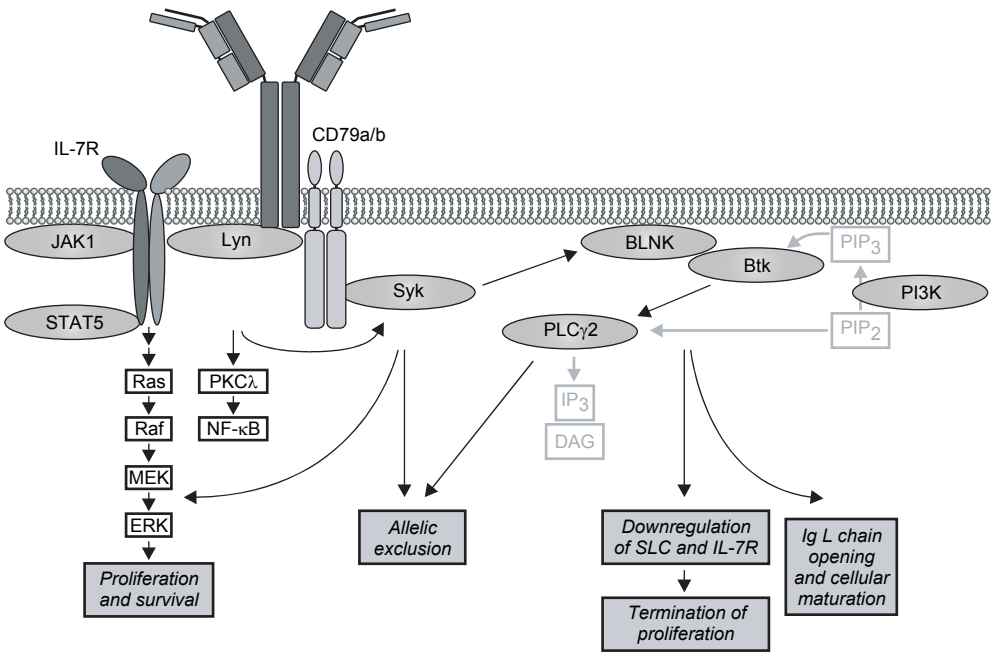


Figure 7. Schematic overview of factors and downstream processes regulated by pre-BCR signaling.
The pre-BCR and the IL-7R signal via Lyn, Syk, Stat5 and the Ras-Raf-MEK-ERK pathway for proliferation. Signaling of the pre-BCR via Syk is also required for allelic exclusion. Furthermore, the pre-BCR signals via BLNK and Btk to limit proliferation and induce Ig light chain rearrangements. Figure adapted from Hendriks and Middendorp.[44]

inhibited by direct downregulation of the RAG proteins.[46] In addition, the *IGH* alleles are made inaccessible for V(D)J recombination. This process, which is called allelic exclusion, inhibits recombination on the second allele ensuring that only one functionally rearranged *IGH* allele is produced. Second, the cell will undergo several cycles of proliferation. Third, transcription of the genes encoding components of the surrogate light chain is inhibited.[47] As a result, the pre-BCR complex is quickly downregulated and proliferation will be limited. Finally, the cell will upregulate RAG protein levels and initiates Ig light chain gene rearrangements. This complete sequence of events is regulated by a complex network of downstream signaling events initiated upon cross-linking of pre-BCR complexes. It is still controversial whether cross-linking is induced by a ligand, or directly by interaction between the surrogate light chains of pre-BCR complexes.[48-50] Upon cross-linking, the intracellular protein-tyrosine kinase Lyn phosphorylates immunoreceptor tyrosine-based activation motifs (ITAMs) in CD79a and CD79b. As illustrated in Figure 7, upon ITAM phosporylation multiple signal transduction molecules are recruited to the pre-BCR complex.

The use of mouse models has contributed greatly in dissecting the pathways through which the pre-BCR signals for the above-mentioned four processes, which have been difficult to address due to the (partial) redundancy of some proteins. Lyn recruitment to the pre-BCR complex activates the Ras-Raf-MEK-ERK pathway to signal for proliferation.[51] This pathway is thought to intersect with IL7R signaling, which also activates this pathway via Lyn.[51, 52] Allelic exclusion is established through signaling via Syk and signaling via PLCγ.[53, 54] BLNK and Btk have been shown to activate PLCγ, however, both proliferation and allelic exclusion were found to be established in absence of both BLNK and of Btk.[55-57] Therefore, PLCγ could be activated through other mechanisms, e.g. via Ca^{2+} signaling, or homologs of BLNK and/or Btk are partially redundant for PLCγ activation. Termination of proliferation and initiation of Ig light chain rearrangements are dependent on signaling via BLNK and Btk. However, the mechanisms through which the surrogate light chain expression is inhibited and the *IGK* locus is made accessible are largely unknown.

Although human and mouse precursor-B-cell differentiation are very similar, observations show that several aspects of pre-BCR signaling may be different in humans. First, in contrast to mouse models, IL7R signaling is dispensable for human precursor-B-cell differentiation.[58, 59] Furthermore, Btk has been shown to be important for pre-BCR induced proliferation of human pre-B-II cells.[60]

BCR selection in bone marrow

Signaling from the pre-BCR induces Ig light chain rearrangements. Similar to VH–DJH rearrangements, the initiation of V–J rearrangements on the *IGK* and *IGL* light chain alleles appears to be highly regulated.[61] First, one *IGK* allele is poised to rearrange. When this renders a non-functional rearrangement, the other *IGK* allele will be rearranged. *IGL* gene rearrangements will be induced when both *IGK* alleles are non-functional. These four Ig light chain alleles provice the cell for multiple options to produce a functional Ig light chain (Figure 8).

In addition, because only single step rearrangements are required to produce a functional VJ exon, multiple attempts are possible on a single allele. After a rearrangement renders a non-functional allele, an upstream V and a downstream J gene segment can rearrange, thereby

removing the pre-existing V–J rearrangement from the genome.

Once an Ig light chain protein is generated that can pair with the Ig heavy chain to form a functional BCR, the cell will be positively selected into the next differentiation stage (immature B cell). Immature B cells that carry an autoreactive BCR go into apoptosis (negative selection), induce additional Ig gene rearrangements, or become anergic. Before the B cell can create another Ig light chain protein, the allele that encodes the autoreactive protein has to be removed from the genome. In addition to subsequent Vκ–Jκ rearrangements, the *IGK* locus can initiate rearrangements between Vκ gene segments and the kappa-deleting element (Kde), which is located downstream of the Cκ exons (Figure 8).[62] These rearrangements remove the rearrangement and the Cκ exons from the genome. The Kde element can also rearrange with an RSS sequence located upstream of the Cκ exons and downstream of the Jκ gene segments. This rearrangement removes the Cκ exons from the genome. Both types of *IGK*-deleting rearrangements are commonly found in mature B lymphocytes of healthy individuals.

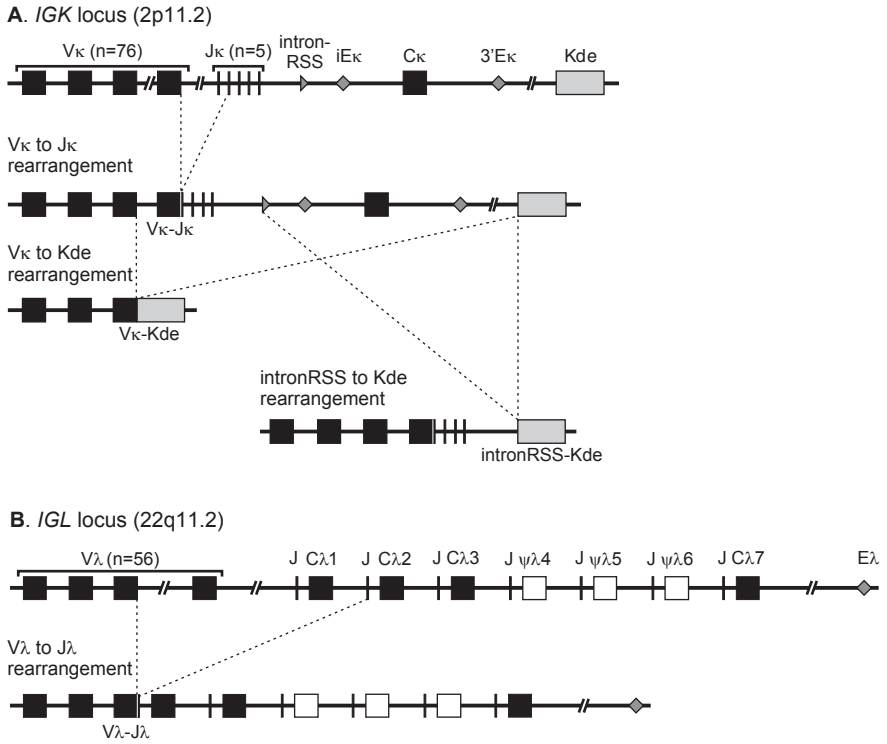


Figure 8. Schematic overview of Ig light chain gene rearrangements.
IGK contains 76 functional V and five functional J gene segments that can be rearranged to form a functional VJκ exon. In addition, Vκ gene segments and an intron-RSS can rearrange with the kappa-deleting element (Kde). These kappa-deleting rearrangements remove the enhancer and the Cκ region, thereby making the allele non-functional. The *IGL* locus contains 56 functional V and three functional J-C clusters that are often used Vλ-Jλ in rearrangements.

Naive B-lymphocyte homeostasis

Once a functional BCR is formed that does not display high affinity for autoantigens, the immature B cell will further maturate and migrate via peripheral blood to peripheral lymphoid organs as a transitional B lymphocyte. In the periphery, the transitional B-lymphocyte will develop into a naive mature B lymphocyte. It is currently unclear which processes and factors are involved in this development. Transitional and naive mature B lymphocytes are functionally different: recent reports demonstrate that transitional B lymphocytes are immature in their migration capacity and response to antigen.[63, 64] Naive mature B lymphocytes are thought to be short-lived unless they are activated upon antigen encounter.

The absolute number of B lymphocytes is tightly regulated. In principle, four processes can contribute to this homeostasis: 1, output from bone marrow; 2, survival; 3, loss of naive B lymphocytes via further maturation after antigen encounter;[65] and 4, antigen-independent (homeostatic) proliferation. Homeostatic proliferation without differentiation of naive mature B lymphocytes independent of antigen has been reported in mouse models in response to a B-cell deficit.[66-68] However, it remains unclear to what extent proliferation contributes to naive B-lymphocyte numbers in healthy individuals.

T-cell dependent and T-cell independent B-cell responses to antigen

Naive mature B lymphocytes undergo an immune response upon recognition of antigen with their specific BCR. The response of a B cell to antigen depends greatly on the strength of the interaction and the presence of co-stimulatory signals. Upon antigen binding, the BCR complex translocates to glycolipid and cholesterol-rich membrane microdomains (‘lipid rafts’). The B-cell co-receptor complex, consisting of CD19, CD21, CD81 and CD225 signals in conjunction with the BCR,[69] thereby reducing the threshold for antigen-dependent stimulation.[70, 71] Since CD21 is a receptor for complement fragment C3d, C3d-loaded antigen can be bound by CD21 and the BCR, thus resulting in co-localization of the two complexes (Figure 9).

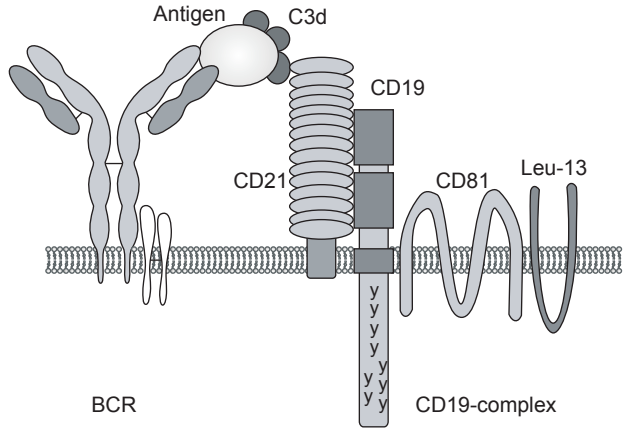


Figure 9. Schematic representation of antigen binding by the BCR and the CD19-complex.
The CD19-complex, consisting of CD19, CD21, CD81 and CD225, signals together with the BCR. C3d-bound antigen can be recognized by CD21, which is a complement receptor, thus bringing the BCR- and the CD19-complex together.

Most B-cell responses are mediated via cognate $CD4^+$ T-cell help. These T cells are activated by exposure to dendritic cells (DC) presenting antigen-derived peptides, and subsequently express B-cell activating membrane-bound and soluble factors. Important membrane-bound receptors are CD40L and ICOS (inducible co-stimulor), which interact with respectively CD40 and ICOSL on the B cell. Upon T-B interaction, the activated B cells undergo extensive proliferation and somatic hypermutation (SHM) of their Ig genes in a highly organized structure in lymph nodes and other secondary lymphoid organs: the germinal center. In this anatomical site, an apical light and a basal dark zone can be identified. The dark zone mainly consists of proliferating B cells (centroblasts) in a network of follicular dendritic cells (FDC). These cells from stromal origin express molecules that are important for B cell proliferation and survival. In this respect, the FDCs are important for affinity maturation by selectively providing survival signals to B cells that bind antigen with high affinity. To this purpose, FDCs present antigen bound to C3d via CD21 on their plasma membrane. After proliferation and SHM, the centroblasts become resting centrocytes and undergo Ig class switch recombination (CSR) in the light zone of the germinal center. Subsequently, the germinal center-derived B lymphocytes differentiate into antibody-producing plasma cells or long-lived memory B cells.

The splenic marginal zone contains B lymphocytes that show a fast response to blood borne pathogens. The BCRs of these marginal zone B cells often recognize repetitive components. Consequently, multiple BCRs can bind to one antigen and are capped together thereby generating a strong signal. The strong signal resulting from this capping enables these cells to generate an immune response independently of T-cell help. Interestingly, these cells have been found to carry SHM on their Ig loci. Because they did not interact with T cells, this molecular process was potentially initiated via a different mechanism.[72]

Molecular processes of somatic hypermutation and Ig class switch recombination

During the proliferation phase of a B-cell antigen response in the germinal center, the DNA encoding the variable regions of the Ig heavy and Ig light chain proteins is the target of SHM. These mutations can be induced throughout the variable regions, although mutations are preferentially found in the DGYW sequence motif (D denotes G, A or T; Y denotes C or T; W denotes A or T).[73] Modifications in the amino acid sequence of Ig molecules are unwanted in the domains that are important for structural organization of the molecule, i.e. framework (FR-)regions. In addition, mutations that negatively influence antigen binding in the antigen-recognition domains, the complementary-determining regions (CDR), are unfavorable. Consequently, the B cells that acquire these mutations go into apoptosis due to a lack of BCR-mediated proliferative signals. In contrast, mutations in CDRs that result in higher affinity interaction with antigen lead to a selective advantage for the cell to survive and proliferate. Consequently, SHM and selection greatly contribute to antigen-dependent affinity maturation.

In addition to V(D)J recombination and SHM, a third molecular mechanism is involved in the antibody response: CSR. The *IGH* locus contains multiple constant regions downstream of Cμ. In precursor- and naive mature B lymphocytes, these regions are not used because the rearranged VDJ_H exon is spliced to the Cμ exons. In the germinal center, the B cell is capable of rearranging the Ig switch region upstream of Cμ with one of the switch regions upstream

of the other constant regions. This process results in deletion of the intervening DNA and places a different constant region directly downstream of the JH region. CSR does not alter the antigen specificity or membrane expression of the receptor on the B cell. The constant region is important for mediating immune responses by secreted Ig after binding to antigen. Because the different isotypes are recognized by specific receptors on immune cells, they exhibit different effector functions.

Key player in the molecular processes of SHM and CSR in mature B lymphocytes is the activation-induced cytidine deaminase (AID). Although AID shares homology with RNA-editing enzymes, it is currently believed that AID deaminates cytidine (C) residues to uracil (U) in single stranded genomic DNA instead of RNA transcripts. AID probably acts during the DNA replication phase, when single strand formation of genomic DNA is temporarily induced. The enzymatic function of AID ultimately results in G•U mismatches that are repaired via different mechanisms for SHM and CSR.

The uracil residues induced by AID can either be replicated in a proliferating cell, resulting in a transition mutation (C to T, and G to A), or be removed by uracil-N-glycosylase (UNG) creating an abasic site (Figure 10).[74] This site can attract either members of the base excision repair (BER), or the mismatch repair (MMR) machinery. The BER machinery specifically targets the abasic site and error-prone DNA polymerases add a random nucleotide generating

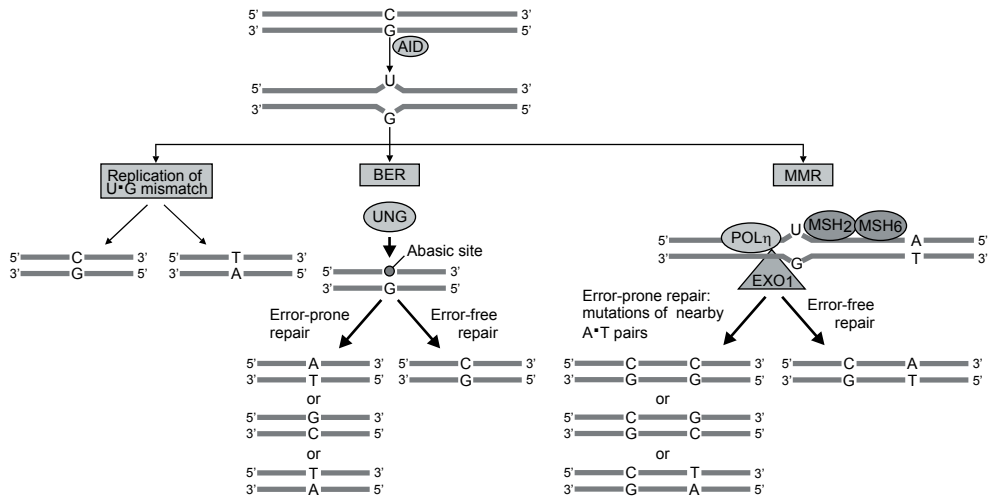


Figure 10. The error-prone repair of AID-induced lesions.
AID initiates SHM by the deamination of cytidine (C) nucleotides on single-stranded DNA. The resulting uracil (U)•guanosine (G) mismatch can be repaired by one of three pathways. (1) For a mutation to be fixed at the site of deamination, error-free DNA repair must be 'perturbed' to become error-prone. If the mismatch is carried unrepaired to replication, DNA polymerases will insert an adenosine (A) opposite the U nucleotide, ultimately creating C to T, and G to A transition mutations. (2) In base-excision repair (BER), replication over an abasic site created by uracil-DNA glycosylase (UNG) will give rise to both transition and transversion mutations. (3) A U•G mismatch also recruits the mismatch repair (MMR) machinery, which is thought to create mutations at A•T base pairs near the initiating U•G lesion. EXO1, exonuclease 1; MSH, homologue of Escherichia coli MutS; POLη, polymerase (DNA directed)-η. Figure adapted from Odegard and Schatz.[74]

transition and transversion mutations. Consequently, this process only generates mutations at G•C bases, which reflect direct targets of AID. *In vivo*, G•C bases constitute only half of the mutations generated, whereas the other half are A•T bases. These mutations are indirect targets and induced by patch repair of the MMR machinery. The G•U mismatches are recognized by a heterodimer of MutS homology proteins MSH2 and MSH6. This heterodimer recruits additional mismatch repair proteins, among which are MLH1, PMS2 and exonuclease 1 that excise the DNA strand containing the G•U mismatch. This strand is newly synthesized by error-prone translesional polymerase eta (polη). The error-prone nature of this polymerase

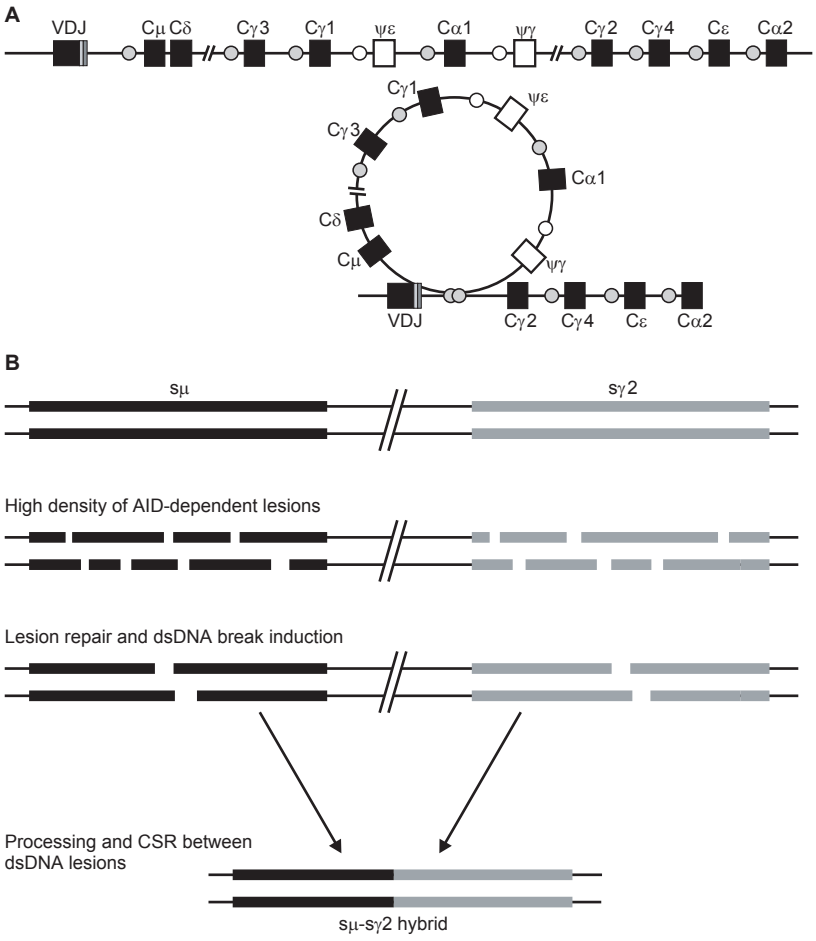


Figure 11. Molecular processes of CSR.
The Ig switch regions (s) contain multiple hot-spots for AID targeting. Consequently, multiple single-stranded DNA lesions can occur in close proximity. These can result in double-stranded DNA breaks and when two Ig switch regions are brought into close proximity, they can be recombined. This results in a hybrid switch region and excision of the intervening DNA as a circular product. **A.** Co-localization of Ig switch versions, Sμ and Sγ2. **B.** High density AID-induced lesions can result in double started DNA breaks and recombination between the Ig switch regions.

can result in the generation of mutations in A•T (and G•C) bases that were not original targets of AID.

The Ig switch regions (s) upstream of the constant gene regions consist of tandem repeats of 20-80 bp sequences, which contain numerous AID hotspot motifs (WRC/GYW). Similar to SHM, the U induced by AID can be removed by UNG and the strand excised by mismatch repair proteins. Because of the high density of hotspot motifs, two abasic sites can be induced within a small range on opposite strands (Figure 11). When both breaks result in excision of DNA strands, this will lead to a double stranded (ds)DNA break instead of just a single stranded break. Consequently, NHEJ proteins are recruited similar to V(D)J recombination. The induction of dsDNA breaks in Sμ and one of the other switch regions will lead to rearrangement between the switch regions and excision of the intervening DNA.

Control of B-lymphocyte homeostasis and antigen responses by TNF family ligands BAFF and APRIL

B-cell activating factor (BAFF) and a proliferation-inducing ligand (APRIL) are two tumor necrosis factors (TNF) that are implicated in several aspects of B-cell survival and Ig isotype switching and production.[75, 76] BAFF and APRIL can both bind the BCMA (B-cell maturation antigen) and TACI (transmembrane activator and CAML interactor) receptors,[77-81] whereas BAFF specifically binds to a third receptor: BAFF-R,[82, 83] and APRIL can interact with proteoglycans (Figure 12).[84, 85] Similar to other TNF receptors, BCMA, TACI and BAFF-R form homo-trimeric complexes on the plasma membrane. Both ligands are expressed on the plasma membrane and excreted as trimeric molecules by multiple cell types from both hematopoietic (myeloid cells) and non-hematopoietic origin (reviewed in ref. 86). In addition, 20 BAFF trimers can form a 60-mer virus-like structure.[87, 88] Although BAFF and APRIL

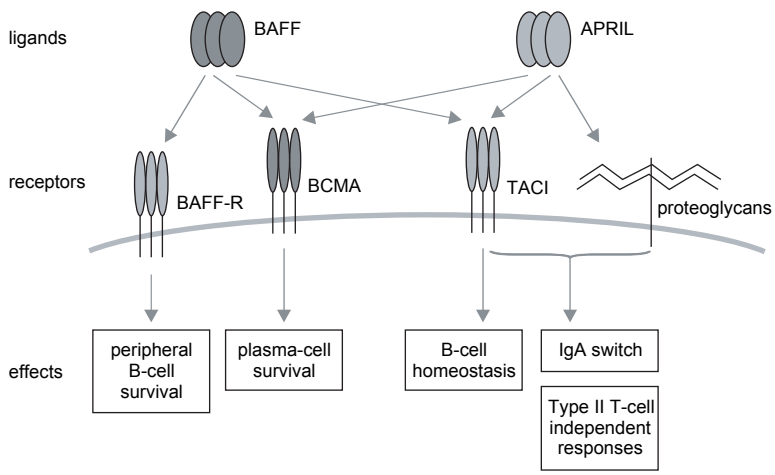


Figure 12. Ligand specificity of BAFF and APRIL and function assignment of their receptors.
BAFF and APRIL can both bind TACI and BCMA, whereas BAFF specifically recognizes BAFF-R, and APRIL can specifically interact with proteoglycans. The receptors are differentially expressed during B-cell differentiation and signal for unique processes, as indicated. Figure adapted from Schneider.[92]

show clear overlap in receptor binding and they are widely expressed, they exhibit unique functions. These unique features mainly arise from the different affinities with which they bind to their receptors and the restricted expression pattern of those receptors (reviewed in refs. 86, 89).

In addition to constant and ligand-independent tonic signaling via the BCR,[90, 91] naive mature B lymphocytes critically depend on BAFF – BAFF-R interaction for survival.[75, 92]

In mice, genetic ablation of BAFF or BAFF-R results in a dramatic reduction in nearly peripheral B-cell subsets.[93, 94]

TACI appears to have a dual role in B-cell activation. TACI$^{-/-}$ mice have enlarged spleens and lymph nodes, suggesting that TACI functions to regulate B-cell homeostasis.[95-97] On the other hand, TACI$^{-/-}$ mice have reduced serum IgM and IgA levels, and they mount a deficient response to type II T-cell independent antigens.[95] In contrast to BAFF, APRIL can bind TACI and proteoglycans simultaneously. The opposing roles of TACI in B-cell activation could be the result of these two different interactions: TACI acts as a negative regulator of BAFF-induced B-cell responses and a positive regulator of APRIL induced IgA production and humoral responses to T-cell independent antigens.[98] BCMA was only found expressed on the surface of a proportion of germinal center B cells.[99] The interaction of BCMA with APRIL is stronger than with BAFF. BCMA$^{-/-}$ mice do not show any B-cell defects,[100] but BCMA appears to be involved in antigen presentation and plasma cell survival.[101, 102]

2. PRIMARY ANTIBODY DEFICIENCIES

Proper functioning of the complex network of cells and components that form the immune system is required for proper host defence. This is illustrated by patients that suffer from primary immunodeficiencies. These patients have a rare inherited disorder that affects the immune system. As a result, they present with multiple recurrent infections during childhood or as young adult, depending on the defect. In the past five decades, over 120 inherited primary immunodeficiencies have been discovered, which have recently been grouped into eight categories:[103]

1. Combined T- and B-cell immunodeficiencies
2. Predominantly antibody deficiencies
3. Other well-defined immunodeficiency syndromes
4. Diseases of immune dysregulation
5. Congenital defects of phagocyte number, function, or both
6. Defects in innate immunity
7. Autoinflammatory disorders
8. Complement deficiencies

B-cell defects can be found in two categories: 1, 2 and 3. Category 1 contains the B-cell deficiencies that are combined with a T-cell deficiency. Due to the T-cell deficiency, these patients can suffer from severe and often opportunistic infections caused by a wide range of pathogens already in the first months of life. Most deficiencies in this category were

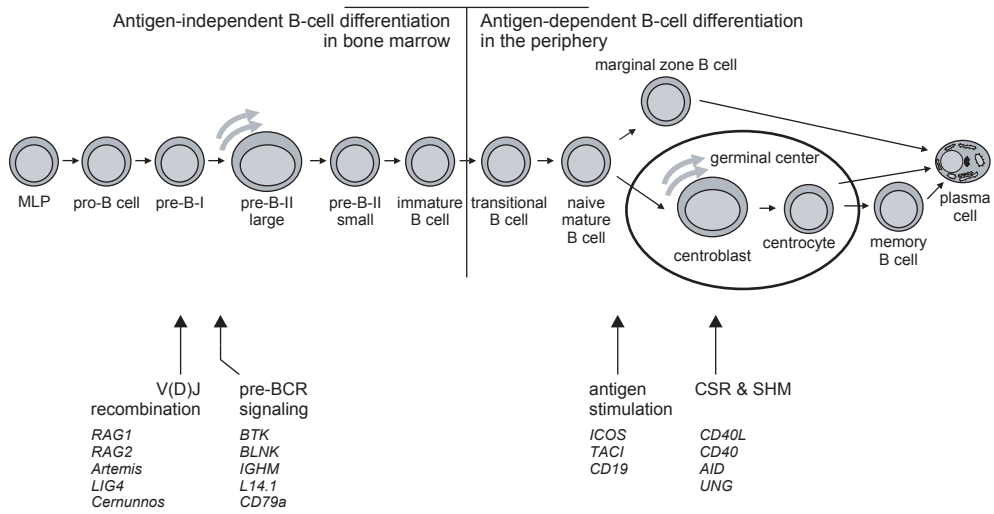


Figure 13. B-cell differentiation blocks in primary immunodeficiency diseases.
Defects in genes encoding V(D)J recombination or pre-BCR signaling components result in a block in precursor-B-cell differentiation prior to the pre-B-II cell stage. ICOS, TACI and CD19 (**Chapter V**) deficiencies lead to a defect in B-cell maturation and CD40L, CD40, AID and UNG defects result in impaired CSR and SHM in the germinal center reaction.

previously denoted as severe combined immunodeficiency (SCID). Category 2 concerns B-cell specific deficiencies. These patients are unable to generate a proper antibody response. As a result, they commonly present with multiple recurrent bacterial infections, e.g. of the gut and upper respiratory tract. Category 3 contains syndromes with a defect in the immune system and additional clinical features that are not related to the immunodeficiency. The different types of B-cell defects and affected genes are discussed below and indicated in Figure 13.

V(D)J recombination defects

Defects in the genes required for V(D)J recombination lead to a combined T- and B-cell deficiency and are grouped into category 1 and 3. In these patients, B- and T-cell differentiation are blocked in early stages in bone marrow and thymus due to the inability to produce a functional BCR or TCR. Consequently, hardly any B or T lymphocytes are found in peripheral blood and serum Igs are not detectable. The first defects in these $T^{-}B^{-}$SCID patients were identified in the *RAG* genes.[104] These patients show a full block in precursor-B-cell differentiation in bone marrow at the pre-B-I cell stage and lack Ig gene rearrangements.[105] Since *RAG* gene expression is restricted to precursor-B and precursor-T cells, these patients have no other clinical features. Hypomorphic mutations in the *RAG* genes that allow restricted residual activity lead to Omenn syndrome. These patients do have peripheral T cells, but in general these are of oligoclonal origin with restricted TCR heterogeneity. In addition to a profound immunodeficiency, these patients commonly have severe atopic dermatitis, histiocytosis X, and symptoms

similar to graft-versus-host disease (GVHD). For diagnosis of Omenn syndrome, maternal T-cell engraftment or transfusion-associated GVHD should be excluded.

More recently, defects in T^-B^-SCID patients have been identified in three genes encoding proteins involved in dsDNA break repair: Artemis, DNA ligase IV and Cernunnos.[37, 106, 107] These proteins are expressed in all cell types, where they function in recognition and repair of dsDNA breaks. As a result, patients with a defect in one of these genes are radiation sensitive, which is illustrated by the fact that fibroblasts cultured from the skin of these patients are highly sensitive to ionizing radiation.[37, 106, 107] Furthermore, patients with DNA ligase IV and Cernunnos deficiencies can present with microcephaly and facial dystrophy.[107, 108]

Pre-BCR signaling defects

Several defects have been described in genes that form the pre-BCR or that are required for downstream signaling from this receptor. These genes are crucial for proper precursor-B-cell differentiation, and defects result in a profound decrease or absence of B lymphocytes in the periphery and profound decreased serum Ig levels. Previously, this group of patients was classified as agammaglobulinemia. This disease was first described by Bruton in 1952 and seemed to affect only boys, thus suggesting X-linked inheritance.[109] The affected gene was not identified until 1993.[110, 111] It appeared to be a protein-tyrosine kinase that signals downstream of the pre-BCR, and it was later named BTK. More recent observations indicate that about 85 % of agammaglobulinemia patients are boys that carry a genetic defect in the *BTK* gene on chromosome *X*.[112] In some of the residual cases, genetic defects have been found in *IGHM*,[113] *L14.1*,[114] *CD79A*,[115] and *BLNK*.[116] These defects are all autosomal recessive and are rarely found. Defects in all five genes result in a block in precursor-B-cell differentiation before the pre-B-II cell stage. In IGHM-deficient and in CD79A-deficient patients, this is a full block, whereas in *L14.1*, *BTK* and *BLNK* deficiencies low numbers of pre-B-II, immature and mature B lymphocytes can be found.

Class switch recombination and/or somatic hypermutation defects

Deficiencies of four genes have been identified that result in a defect in CSR and/or SHM. Previously, these deficiencies were named hyper-IgM syndromes. These patients have normal numbers of B lymphocytes and normal to high levels of serum IgM, but due to defects in CSR, they lack or have strongly reduced serum IgG and IgA.

Originally, the hyper-IgM syndrome was described in boys and the first genetic defects were identified in a gene located on the X-chromosome encoding CD154 (CD40L).[117] More recently, an autosomal recessive variant with similar clinical features was shown to be caused by mutations in the gene encoding CD40.[118] The CD40 and CD40L are required for, but not restricted to, T-cell dependent B-cell responses, and defects in these genes also affect T-cell and innate immunity. Therefore, in addition to the antibody deficiency, these patients can suffer from neutropenia and opportunistic infections. Consequently, CD40 and CD40L deficiencies have recently been grouped in the combined T- and B-cell immunodeficiencies.[101] CD40- and CD40L-deficient patients lack class-switched memory B cells in peripheral blood, whereas they do have $CD27^+IgM^+IgD^+$ natural effector cells, which are thought to be

recirculating splenic marginal zone B lymphocytes.[119] The Ig genes of these natural effector B cells contain somatic hypermutations.[72] This indicates that CD40-CD40L interaction is required for the germinal center reaction, but not for induction of AID in T-cell independent responses in the marginal zone.[71, 116]

Defects in *AID* and *UNG* have recently been identified in patients with severely reduced serum IgG and IgA.[120, 121] These defects result in a B-cell specific defect. Null mutations in *AID* completely abolish both SHM and CSR, whereas hypomorphic mutations have been found that specifically restricts one of the two molecular processes. UNG-deficient B cells fail to generated dsDNA breaks in Ig switch regions and cannot undergo CSR. SHM can take place, however, mutations at C and G residues are skewed towards transition mutations.[121]

Common variable immunodeficiencies

Although many gene defects have been shown to underlie primary antibody deficiencies, no molecular defect has been found in large groups of immunodeficient patients. This concerns patients with a phenotype that resembles T^-B^-SCID, agammaglobulinemia or a hyper-IgM syndrome, but also patients with a more diffuse clinical picture. The latter group is currently diagnosed as common variable immunodeficiency (CVID) and includes patients with both childhood and early adult onset of predominantly recurrent bacterial infections. These observations suggest that other genes might be affected than those previously identified to cause antibody deficiency when disrupted. Similar to the identified genes, these as yet unidentified genes might be involved in the V(D)J recombination, pre-BCR signaling, or CSR/SHM processes. However, genes involved in other processes that are critical for B-cell differentiation and antigen maturation might be mutated. Recently, the first monogenetic defects have been identified in CVID patients. A single founder mutation in the *ICOS* gene was found in nine patients, corresponding to <1 % of CVID patients.[122, 123] ICOS is upregulated in T cells and is required for normal germinal center formation. The ICOS deficiency therefore results in reduced Ig switched memory B cells, plasma cells and serum IgA and IgG. The majority of the ICOS-deficient patients presented in adulthood, suggesting that partly compensatory mechanisms exist to maintain immunity during childhood. The second gene that was found to be associated with CVID is *TACI*.[124, 125] Whereas patients with homozygous mutations have been found, the majority of identified patients carries a single affected allele and family pedigree analysis indicates autosomal dominant inheritance.[126] Approximately 8 % of CVID patients carries a mutant *TACI* allele.[124, 125] In <10 % of antibody deficient patients, a genetic defect has been found that underlies the disease. Patients diagnosed as having CVID present with highly variable susceptibility to and severity of infections. In addition, an increased incidence of autoimmune diseases and malignancies is found in cohorts of CVID patients. The clinical variation restricts optimal individual patient treatment and although patients are currently treated with IgG substitution therapy and (prophylactic) antibodies,[127] their quality of life is significantly worse than patients with other chronic illnesses .[128]

3. GENOMIC MUTATIONS IN HUMAN DISEASE

Genomic mutations underlie human inherited diseases, such as primary immunodeficiencies. Characterization of the gene lesions resulted in the identification of multiple different types of mutation: single base pair substitutions, deletions, insertions, duplications, inversions, and repeat expansions. Functional studies to address the mechanism underlying these mutations are difficult to perform and data on this topic is scarce. However, molecular data on disease-causing mutations is widely available thanks to inclusion in databases, either genome-wide (Online Mendelian Inheritance in Man; OMIM,[129] and Human Gene Mutation Database; HGMD),[130] or gene-specific (e.g. immunodeficiency mutation databases; IDbases).[131] The collected data on disease-causing mutations allows sequence analysis for common motifs in order to identify the processes mediating the mutations.

Within the total group of disease-causing mutations, a crude division is made between small lesions (<20 bp) and gross lesions. Small lesions include single-base-pair substitutions, micro-deletions (<20 bp), micro-insertions (<20 bp) and micro-indels (<20 bp). Indels are lesions, in which one or more nucleotides were deleted and the same or a different number of new nucleotides were inserted. Recent analysis of the mutations recorded in HGMD showed that single basepair substitutions are most frequently observed: 68 % of all mutations (Figure 14). This is clearly more than micro-deletions (16 %), micro-insertions (7 %) and micro-indels (1.5 %).

In 2000, Antonarakis and colleagues performed a detailed analysis of small lesions recorded in HGMD.[132] They found a significant excess of transition (purine - purine or pyrimidine - pyrimidine exchange) over transversion (purine - pyrimidine exchange) mutations as compared

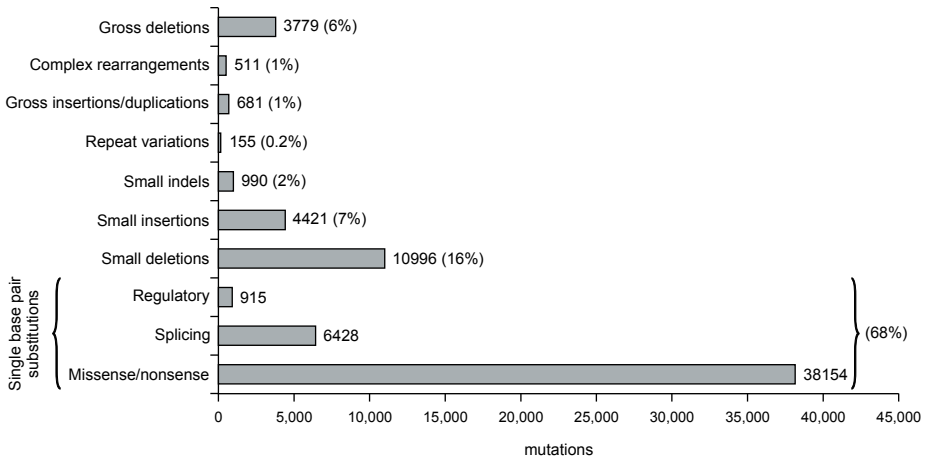


Figure 14. Spectrum of the different types of human gene mutations recorded in the human genome mutation database (HGMD) as of 3 January 2007 containing 67030 mutations.
Micro-lesions are the most commonly found gene disruptions underlying human disease, with the vast majority being single base pair substitution. Although not common, a substantial number of gross lesions has been identified. The majority of these are gross deletions.

to random expectation. This was mainly, but not solely, affected by the hypermutability of the CG dinucleotide to TG or CA. Nucleotide substitutions are thought to occur in a non-randomly throughout the genome.[133, 134] Especially CG dinucleotides have an increased susceptibility to be mutated, and has been observed in a wide range of human genes.[135-138] This is thought to be related to the methylation of the cytosine in CG dinucleotides.[139] Deamination of cytosine results in uracil, whereas deamination of 5-methylcytosine results in a thymidine residue. Uracil is recognized and excised by uracil DNA glycosylase. However, thymidine is not specifically excised, as it is one of the four natural bases in DNA. Thus, methylated CG dinucleotides have increased mutation susceptibility into TG or CA.

Micro-deletions and -insertions are associated with short repeats that are usually between 2 and 8 bp.[140] These elements can be direct repeats, palindromes (inverted repeats) and symmetric elements. Most proposed mechanisms that underlie microdeletions are based on the slipped-mispairing hypothesis,[141] which states that during replication one DNA strand can be misaligned via a repeat sequence with the complementary strand.[142, 143] The resulting loop structure can be excised thereby generating the small deletion. This type of deletion is induced by mutagens.[144]

Gross lesions (>20 bp) include gross deletions (6 % of total deleterious mutations), gross insertions (1 %), repeat variations (0.2 %), or other complex rearrangements (1 %) such as inversions. The exact molecular mechanisms by which gross lesions arise are not completely understood, but based on limited sequence data of gross deletion junctions in disease-causing alleles, three mechanisms resulting in gross deletions have been proposed: 1) Mispairing of homologous sequences and unequal crossing over; 2) Non-homologous deletions; 3) Mispairing between short repeat elements and crossing-over.[145] Non-homologous gross deletion breakpoints that fall in the latter two categories are frequently located in or near interspersed repeat elements.[146]

Whereas coding sequences comprise only 5 % of the human genome, repeat sequences account for at least 50 %.[147] The majority of these sequences are transposon-derived repeats (interspersed elements). In mammals, four classes of transposon-derived repeats can be identified, which make up about 45 % of the total sequence (Figure 15). DNA transposons

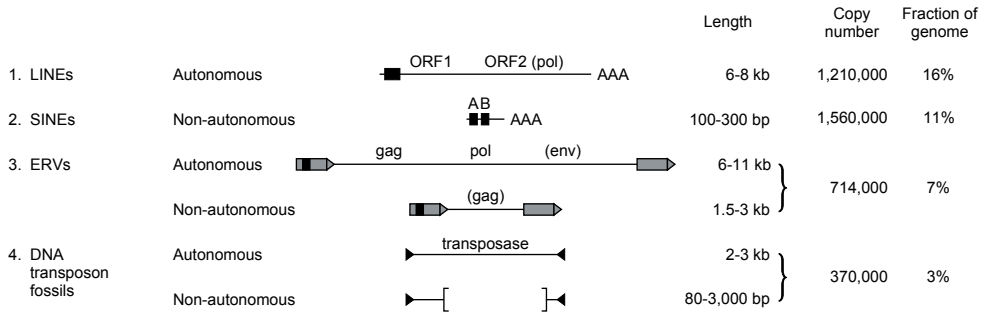


Figure 15. Classes of transposon-derived elements in the human genome.
Four main classes of transposable elements can be identified based on their functional characteristics. Three of these, LINEs, SINEs and ERVs, transpose with RNA intermediates, whereas the fourth, DNA transposons, does not. Figure modified from Lander *et al.*;[147] the copy numbers and fraction of genome for each class were determined on the human genome sequence in December 2006 using the CENSOR software tool of the Repbase database in normal mode.[148]

transpose directly as DNA, whereas LTR retrotransposons, long and short interspersed elements (LINE and SINE) transpose via RNA intermediates. LINEs are autonomous elements; they harbor an internal polymerase II promoter to generate transcripts. These transcripts encode two proteins that assemble with the transcript and display reverse transcriptase activity enabling integration of RNA transcripts in genomic DNA. Of the three LINE families (LINE1-3), only LINE1 elements are still active.

SINEs harbor an internal polymerase III promoter, but do not contain an open reading frame (ORF). Consequently, these elements cannot transpose autonomously. They are thought to use the LINE1 machinery for transposition. Of the three SINE families in the human genome, Alu elements are active, whereas MIR and MIR3 are not.

Full-length LTR retrotransposons are autonomous. They contain long terminal 5' and 3' direct repeats that contain transcriptional regulatory elements. Furthermore, LTR retrotransposons contain the *GAG* and *POL* genes that mediate the integration of RNA intermediates in genomic DNA. In mammals, the only active LTR retrotransposons are endogenous retroviruses (ERV), which fall into three classes (I-III).

DNA transposons have terminal inverted repeats that flank the transposase-encoding gene. The transposase recognizes the inverted repeats and is able to cut the DNA transposon from the genome and paste it in a different position without forming an RNA intermediate.

4. AIMS OF THE THESIS

B lymphocytes and the immunoglobulins they produce are critical components of the immune system. This is illustrated by patients with primary antibody deficiency. These patients have a defect in the generation of B lymphocytes or in the production of immunoglobulins by their B lymphocytes. Genetic disruption of a single gene is often the cause of the disease. Although in the past 15 years defects in multiple genes have been identified in patients with primary antibody deficiency, in many cases a molecular diagnosis has not been made. In order to identify the disrupted gene, a better understanding of B-cell differentiation is required.

Primary antibody deficient patients in whom the underlying genetic defect has been identified provide the unique opportunity to study the role of a single protein in the immune system. Studying this defect will result in better understanding of the human immune system. Subsequently, this knowledge can be applied to develop novel approaches to increase the efficiency and accuracy of diagnosis and chronic treatment of patients with primary immunodeficiency.

This thesis describes studies aimed at understanding important checkpoints and key players in B-cell differentiation and defects underlying primary antibody deficient patients. **Chapter II** describes the initiation of *IGH* and *IGK/IGL* rearrangements in human precursor-B-cell differentiation and the networks of factors that might be involved. In **Chapter III**, mouse models are used to visualize the *IGH* locus conformation in early lymphoid and B-progenitor cells. Large-scale compaction of *IGH* is required for V(D)J recombination, but the underlying mechanisms are largely unknown. The majority of human disease-causing gene disruptions are small lesions. This is not the case for several genes disrupted in primary immunodeficiencies and leading to a block in precursor-B-cell differentiation. **Chapter IV**

describes gross deletions in primary immunodeficiency genes and potential mechanisms that mediate these deletions. In **Chapter V**, a novel antibody deficiency syndrome is described in four patients due to mutations in *CD19*. Immunological studies were performed to identify the B-cell defects. The contribution of proliferation to naive B-lymphocyte homeostasis and antigen-induced B-cell expansion are described in **Chapter VI**. The **General Discussion** deals with the implications of the studies described in this thesis and with directions for future research.

REFERENCES

1. Kondo M, Wagers AJ, Manz MG, Prohaska SS, Scherer DC, Beilhack GF, Shizuru JA, Weissman IL. Biology of hematopoietic stem cells and progenitors: implications for clinical application. *Annu Rev Immunol* 2003;**21**: 759-806.
2. Kondo M, Weissman IL, Akashi K. Identification of clonogenic common lymphoid progenitors in mouse bone marrow. *Cell* 1997;**91**:661-72.
3. Akashi K, Traver D, Miyamoto T, Weissman IL. A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. *Nature* 2000;**404**:193-7.
4. Rumfelt LL, Zhou Y, Rowley BM, Shinton SA, Hardy RR. Lineage specification and plasticity in CD19- early B cell precursors. *J Exp Med* 2006;**203**:675-87.
5. Busslinger M. Transcriptional control of early B cell development. *Annu Rev Immunol* 2004;**22**:55-79.
6. Roessler S, Grosschedl R. Role of transcription factors in commitment and differentiation of early B lymphoid cells. *Semin Immunol* 2006;**18**:12-9.
7. DeKoter RP, Singh H. Regulation of B lymphocyte and macrophage development by graded expression of PU.1. *Science* 2000;**288**:1439-41.
8. Nichogiannopoulou A, Trevisan M, Neben S, Friedrich C, Georgopoulos K. Defects in hemopoietic stem cell activity in Ikaros mutant mice. *J Exp Med* 1999;**190**:1201-14.
9. DeKoter RP, Lee HJ, Singh H. PU.1 regulates expression of the interleukin-7 receptor in lymphoid progenitors. *Immunity* 2002;**16**:297-309.
10. Murre C, McCaw PS, Baltimore D. A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. *Cell* 1989;**56**:777-83.
11. Kadesch T. Helix-loop-helix proteins in the regulation of immunoglobulin gene transcription. *Immunol Today* 1992;**13**:31-6.
12. Sloan SR, Shen CP, McCarrick-Walmsley R, Kadesch T. Phosphorylation of E47 as a potential determinant of B-cell-specific activity. *Mol Cell Biol* 1996;**16**:6900-8.
13. Hagman J, Travis A, Grosschedl R. A novel lineage-specific nuclear factor regulates mb-1 gene transcription at the early stages of B cell differentiation. *Embo J* 1991;**10**:3409-17.
14. Hagman J, Belanger C, Travis A, Turck CW, Grosschedl R. Cloning and functional characterization of early B-cell factor, a regulator of lymphocyte-specific gene expression. *Genes Dev* 1993;**7**:760-73.
15. Travis A, Hagman J, Hwang L, Grosschedl R. Purification of early-B-cell factor and characterization of its DNA-binding specificity. *Mol Cell Biol* 1993;**13**:3392-400.
16. Kee BL, Murre C. Induction of early B cell factor (EBF) and multiple B lineage genes by the basic helix-loop-helix transcription factor E12. *J Exp Med* 1998;**188**:699-713.
17. Medina KL, Pongubala JM, Reddy KL, Lancki DW, Dekoter R, Kieslinger M, Grosschedl R, Singh H. Assembling a gene regulatory network for specification of the B cell fate. *Dev Cell* 2004;**7**:607-17.
18. O'Riordan M, Grosschedl R. Coordinate regulation of B cell differentiation by the transcription factors EBF and E2A. *Immunity* 1999;**11**:21-31.
19. Nutt SL, Heavey B, Rolink AG, Busslinger M. Commitment to the B-lymphoid lineage depends on the transcription factor Pax5. *Nature* 1999;**401**:556-62.
20. Nutt SL, Morrison AM, Dorfler P, Rolink A, Busslinger M. Identification of BSAP (Pax-5) target genes in early B-cell development by loss- and gain-of-function experiments. *Embo J* 1998;**17**:2319-33.

21. Schebesta M, Pfeffer PL, Busslinger M. Control of pre-BCR signaling by Pax5-dependent activation of the BLNK gene. *Immunity* 2002;**17**:473-85.
22. Ikawa T, Kawamoto H, Wright LY, Murre C. Long-term cultured E2A-deficient hematopoietic progenitor cells are pluripotent. *Immunity* 2004;**20**:349-60.
23. Tonegawa S. Somatic generation of antibody diversity. *Nature* 1983;**302**:575-81.
24. Ghia P, ten Boekel E, Sanz E, de la Hera A, Rolink A, Melchers F. Ordering of human bone marrow B lymphocyte precursors by single-cell polymerase chain reaction analyses of the rearrangement status of the immunoglobulin H and L chain gene loci. *J Exp Med* 1996;**184**:2217-29.
25. Ghia P, ten Boekel E, Rolink AG, Melchers F. B-cell development: a comparison between mouse and man. *Immunol Today* 1998;**19**:480-5.
26. Ehlich A, Schaal S, Gu H, Kitamura D, Muller W, Rajewsky K. Immunoglobulin heavy and light chain genes rearrange independently at early stages of B cell development. *Cell* 1993;**72**:695-704.
27. Alt FW, Yancopoulos GD, Blackwell TK, Wood C, Thomas E, Boss M, Coffman R, Rosenberg N, Tonegawa S, Baltimore D. Ordered rearrangement of immunoglobulin heavy chain variable region segments. *Embo J* 1984;**3**:1209-19.
28. Roldan E, Fuxa M, Chong W, Martinez D, Novatchkova M, Busslinger M, Skok JA. Locus 'decontraction' and centromeric recruitment contribute to allelic exclusion of the immunoglobulin heavy-chain gene. *Nat Immunol* 2005;**6**:31-41.
29. Sayegh C, Jhunjhunwala S, Riblet R, Murre C. Visualization of looping involving the immunoglobulin heavy-chain locus in developing B cells. *Genes Dev* 2005;**19**:322-7.
30. Hiom K, Gellert M. Assembly of a 12/23 paired signal complex: a critical control point in V(D)J recombination. *Mol Cell* 1998;**1**:1011-9.
31. van Gent DC, Hiom K, Paull TT, Gellert M. Stimulation of V(D)J cleavage by high mobility group proteins. *Embo J* 1997;**16**:2665-70.
32. McBlane JF, van Gent DC, Ramsden DA, Romeo C, Cuomo CA, Gellert M, Oettinger MA. Cleavage at a V(D)J recombination signal requires only RAG1 and RAG2 proteins and occurs in two steps. *Cell* 1995;**83**:387-95.
33. Hiom K, Gellert M. A stable RAG1-RAG2-DNA complex that is active in V(D)J cleavage. *Cell* 1997;**88**: 65-72.
34. Grawunder U, Harfst E. How to make ends meet in V(D)J recombination. *Curr Opin Immunol* 2001;**13**: 186-94.
35. Gottlieb TM, Jackson SP. The DNA-dependent protein kinase: requirement for DNA ends and association with Ku antigen. *Cell* 1993;**72**:131-42.
36. Nussenzweig A, Chen C, da Costa Soares V, Sanchez M, Sokol K, Nussenzweig MC, Li GC. Requirement for Ku80 in growth and immunoglobulin V(D)J recombination. *Nature* 1996;**382**:551-5.
37. Moshous D, Callebaut I, de Chasseval R, Corneo B, Cavazzana-Calvo M, Le Deist F, Tezcan I, Sanal O, Bertrand Y, Philippe N, Fischer A, de Villartay JP. Artemis, a novel DNA double-strand break repair/V(D)J recombination protein, is mutated in human severe combined immune deficiency. *Cell* 2001;**105**:177-86.
38. Schlissel MS. Structure of nonhairpin coding-end DNA breaks in cells undergoing V(D)J recombination. *Mol Cell Biol* 1998;**18**:2029-37.
39. Benedict CL, Gilfillan S, Thai TH, Kearney JF. Terminal deoxynucleotidyl transferase and repertoire development. *Immunol Rev* 2000;**175**:150-7.
40. Benedict CL, Gilfillan S, Kearney JF. The long isoform of terminal deoxynucleotidyl transferase enters the nucleus and, rather than catalyzing nontemplated nucleotide addition, modulates the catalytic activity of the short isoform. *J Exp Med* 2001;**193**:89-99.
41. Critchlow SE, Bowater RP, Jackson SP. Mammalian DNA double-strand break repair protein XRCC4 interacts with DNA ligase IV. *Curr Biol* 1997;**7**:588-98.
42. Grawunder U, Wilm M, Wu X, Kulesza P, Wilson TE, Mann M, Lieber MR. Activity of DNA ligase IV stimulated by complex formation with XRCC4 protein in mammalian cells. *Nature* 1997;**388**:492-5.
43. Ahnesorg P, Smith P, Jackson SP. XLF interacts with the XRCC4-DNA ligase IV complex to promote DNA nonhomologous end-joining. *Cell* 2006;**124**:301-13.
44. Hendriks RW, Middendorp S. The pre-BCR checkpoint as a cell-autonomous proliferation switch. *Trends Immunol* 2004;**25**:249-56.

45. Melchers F. The pre-B-cell receptor: selector of fitting immunoglobulin heavy chains for the B-cell repertoire. *Nat Rev Immunol* 2005;**5**:578-84.
46. Grawunder U, Leu TM, Schatz DG, Werner A, Rolink AG, Melchers F, Winkler TH. Down-regulation of RAG1 and RAG2 gene expression in preB cells after functional immunoglobulin heavy chain rearrangement. *Immunity* 1995;**3**:601-8.
47. Parker MJ, Licence S, Erlandsson L, Galler GR, Chakalova L, Osborne CS, Morgan G, Fraser P, Jumaa H, Winkler TH, Skok J, Martensson IL. The pre-B-cell receptor induces silencing of VpreB and lambda5 transcription. *Embo J* 2005;**24**:3895-905.
48. Gauthier L, Rossi B, Roux F, Termine E, Schiff C. Galectin-1 is a stromal cell ligand of the pre-B cell receptor (BCR) implicated in synapse formation between pre-B and stromal cells and in pre-BCR triggering. *Proc Natl Acad Sci U S A* 2002;**99**:13014-9.
49. Bradl H, Wittmann J, Milius D, Vettermann C, Jack HM. Interaction of murine precursor B cell receptor with stroma cells is controlled by the unique tail of lambda 5 and stroma cell-associated heparan sulfate. *J Immunol* 2003;**171**:2338-48.
50. Ohnishi K, Melchers F. The nonimmunoglobulin portion of lambda5 mediates cell-autonomous pre-B cell receptor signaling. *Nat Immunol* 2003;**4**:849-56.
51. Fleming HE, Paige CJ. Pre-B cell receptor signaling mediates selective response to IL-7 at the pro-B to pre-B cell transition via an ERK/MAP kinase-dependent pathway. *Immunity* 2001;**15**:521-31.
52. Seckinger P, Fougereau M. Activation of src family kinases in human pre-B cells by IL-7. *J Immunol* 1994;**153**:97-109.
53. Schweighoffer E, Vanes L, Mathiot A, Nakamura T, Tybulewicz VL. Unexpected requirement for ZAP-70 in pre-B cell development and allelic exclusion. *Immunity* 2003;**18**:523-33.
54. Wen R, Chen Y, Schuman J, Fu G, Yang S, Zhang W, Newman DK, Wang D. An important role of phospholipase Cgamma1 in pre-B-cell development and allelic exclusion. *Embo J* 2004;**23**:4007-17.
55. Xu S, Lee KG, Huo J, Kurosaki T, Lam KP. Combined deficiencies in Bruton's Tyrosine Kinase and Phospholipase C{gamma}2 arrest B cell development at a pre-BCR-positive stage. *Blood* 2007;**109**:3377-84.
56. Hayashi K, Yamamoto M, Nojima T, Goitsuka R, Kitamura D. Distinct signaling requirements for Dmu selection, IgH allelic exclusion, pre-B cell transition, and tumor suppression in B cell progenitors. *Immunity* 2003;**18**:825-36.
57. Xu S, Wong SC, Lam KP. Cutting edge: B cell linker protein is dispensable for the allelic exclusion of immunoglobulin heavy chain locus but required for the persistence of CD5+ B cells. *J Immunol* 2000;**165**: 4153-7.
58. Noguchi M, Yi H, Rosenblatt HM, Filipovich AH, Adelstein S, Modi WS, McBride OW, Leonard WJ. Interleukin-2 receptor gamma chain mutation results in X-linked severe combined immunodeficiency in humans. *Cell* 1993;**73**:147-57.
59. Puel A, Ziegler SF, Buckley RH, Leonard WJ. Defective IL7R expression in T(-)B(+)NK(+) severe combined immunodeficiency. *Nat Genet* 1998;**20**:394-7.
60. Nomura K, Kanegane H, Karasuyama H, Tsukada S, Agematsu K, Murakami G, Sakazume S, Sako M, Tanaka R, Kuniya Y, Komeno T, Ishihara S, Hayashi K, Kishimoto T, Miyawaki T. Genetic defect in human X-linked agammaglobulinemia impedes a maturational evolution of pro-B cells into a later stage of pre-B cells in the B-cell differentiation pathway. *Blood* 2000;**96**:610-7.
61. Korsmeyer SJ, Hieter PA, Sharrow SO, Goldman CK, Leder P, Waldmann TA. Normal human B cells display ordered light chain gene rearrangements and deletions. *J Exp Med* 1982;**156**:975-85.
62. Siminovitch KA, Bakhshi A, Goldman P, Korsmeyer SJ. A uniform deleting element mediates the loss of kappa genes in human B cells. *Nature* 1985;**316**:260-2.
63. Sims GP, Ettinger R, Shirota Y, Yarboro CH, Illei GG, Lipsky PE. Identification and characterization of circulating human transitional B cells. *Blood* 2005;**105**:4390-8.
64. Cuss AK, Avery DT, Cannons JL, Yu LJ, Nichols KE, Shaw PJ, Tangye SG. Expansion of functionally immature transitional B cells is associated with human-immunodeficient states characterized by impaired humoral immunity. *J Immunol* 2006;**176**:1506-16.
65. Agenes F, Rosado MM, Freitas AA. Peripheral B cell survival. *Cell Mol Life Sci* 2000;**57**:1220-8.
66. Agenes F, Freitas AA. Transfer of small resting B cells into immunodeficient hosts results in the selection of a self-renewing activated B cell population. *J Exp Med* 1999;**189**:319-30.

67. Cabatingan MS, Schmidt MR, Sen R, Woodland RT. Naive B lymphocytes undergo homeostatic proliferation in response to B cell deficit. *J Immunol* 2002;**169**:6795-805.
68. Woodland RT, Schmidt MR. Homeostatic proliferation of B cells. *Semin Immunol* 2005;**17**:209-17.
69. Cherukuri A, Cheng PC, Sohn HW, Pierce SK. The CD19/CD21 complex functions to prolong B cell antigen receptor signaling from lipid rafts. *Immunity* 2001;**14**:169-79.
70. Carter RH, Fearon DT. CD19: lowering the threshold for antigen receptor stimulation of B lymphocytes. *Science* 1992;**256**:105-7.
71. van Noesel CJ, Lankester AC, van Lier RA. Dual antigen recognition by B cells. *Immunol Today* 1993;14: 8-11.
72. Weller S, Faili A, Garcia C, Braun MC, Le Deist FF, de Saint Basile GG, Hermine O, Fischer A, Reynaud CA, Weill JC. CD40-CD40L independent Ig gene hypermutation suggests a second B cell diversification pathway in humans. *Proc Natl Acad Sci U S A* 2001;**98**:1166-70.
73. Rogozin IB, Diaz M. Cutting edge: DGYW/WRCH is a better predictor of mutability at G:C bases in Ig hypermutation than the widely accepted RGYW/WRCY motif and probably reflects a two-step activation-induced cytidine deaminase-triggered process. *J Immunol* 2004;**172**:3382-4.
74. Odegard VH, Schatz DG. Targeting of somatic hypermutation. *Nat Rev Immunol* 2006;**6**:573-83.
75. Mackay F, Schneider P, Rennert P, Browning J. BAFF AND APRIL: a tutorial on B cell survival. *Annu Rev Immunol* 2003;21:231-64.
76. Dillon SR, Gross JA, Ansell SM, Novak AJ. An APRIL to remember: novel TNF ligands as therapeutic targets. *Nat Rev Drug Discov* 2006;**5**:235-46.
77. Gross JA, Johnston J, Mudri S, Enselman R, Dillon SR, Madden K, Xu W, Parrish-Novak J, Foster D, Lofton-Day C, Moore M, Littau A, Grossman A, Haugen H, Foley K, Blumberg H, Harrison K, Kindsvogel W, Clegg CH. TACI and BCMA are receptors for a TNF homologue implicated in B-cell autoimmune disease. *Nature* 2000;**404**:995-9.
78. Marsters SA, Yan M, Pitti RM, Haas PE, Dixit VM, Ashkenazi A. Interaction of the TNF homologues BLyS and APRIL with the TNF receptor homologues BCMA and TACI. *Curr Biol* 2000;**10**:785-8.
79. Thompson JS, Schneider P, Kalled SL, Wang L, Lefevre EA, Cachero TG, MacKay F, Bixler SA, Zafari M, Liu ZY, Woodcock SA, Qian F, Batten M, Madry C, Richard Y, Benjamin CD, Browning JL, Tsapis A, Tschopp J, Ambrose C. BAFF binds to the tumor necrosis factor receptor-like molecule B cell maturation antigen and is important for maintaining the peripheral B cell population. *J Exp Med* 2000;**192**:129-35.
80. Wu Y, Bressette D, Carrell JA, Kaufman T, Feng P, Taylor K, Gan Y, Cho YH, Garcia AD, Gollatz E, Dimke D, LaFleur D, Migone TS, Nardelli B, Wei P, Ruben SM, Ullrich SJ, Olsen HS, Kanakaraj P, Moore PA, Baker KP. Tumor necrosis factor (TNF) receptor superfamily member TACI is a high affinity receptor for TNF family members APRIL and BLyS. *J Biol Chem* 2000;**275**:35478-85.
81. Yu G, Boone T, Delaney J, Hawkins N, Kelley M, Ramakrishnan M, McCabe S, Qiu WR, Kornuc M, Xia XZ, Guo J, Stolina M, Boyle WJ, Sarosi I, Hsu H, Senaldi G, Theill LE. APRIL and TALL-I and receptors BCMA and TACI: system for regulating humoral immunity. *Nat Immunol* 2000;**1**:252-6.
82. Thompson JS, Bixler SA, Qian F, Vora K, Scott ML, Cachero TG, Hession C, Schneider P, Sizing ID, Mullen C, Strauch K, Zafari M, Benjamin CD, Tschopp J, Browning JL, Ambrose C. BAFF-R, a newly identified TNF receptor that specifically interacts with BAFF. *Science* 2001;**293**:2108-11.
83. Yan M, Brady JR, Chan B, Lee WP, Hsu B, Harless S, Cancro M, Grewal IS, Dixit VM. Identification of a novel receptor for B lymphocyte stimulator that is mutated in a mouse strain with severe B cell deficiency. *Curr Biol* 2001;**11**:1547-52.
84. Hendriks J, Planelles L, de Jong-Odding J, Hardenberg G, Pals ST, Hahne M, Spaargaren M, Medema JP. Heparan sulfate proteoglycan binding promotes APRIL-induced tumor cell proliferation. *Cell Death Differ* 2005;**12**:637-48.
85. Ingold K, Zumsteg A, Tardivel A, Huard B, Steiner QG, Cachero TG, Qiang F, Gorelik L, Kalled SL, Acha-Orbea H, Rennert PD, Tschopp J, Schneider P. Identification of proteoglycans as the APRIL-specific binding partners. *J Exp Med* 2005;**201**:1375-83.
86. Kalled SL. Impact of the BAFF/BR3 axis on B cell survival, germinal center maintenance and antibody production. *Semin Immunol* 2006;**18**:290-6.
87. Liu Y, Xu L, Opalka N, Kappler J, Shu HB, Zhang G. Crystal structure of sTALL-1 reveals a virus-like assembly of TNF family ligands. *Cell* 2002;**108**:383-94.

88. Cachero TG, Schwartz IM, Qian F, Day ES, Bossen C, Ingold K, Tardivel A, Krushinskie D, Eldredge J, Silvian L, Lugovskoy A, Farrington GK, Strauch K, Schneider P, Whitty A. Formation of virus-like clusters is an intrinsic property of the tumor necrosis factor family member BAFF (B cell activating factor). *Biochemistry* 2006;**45**:2006-13.
89. Bossen C, Schneider P. BAFF, APRIL and their receptors: Structure, function and signaling. *Semin Immunol* 2006;**18**:263-75.
90. Lam KP, Kuhn R, Rajewsky K. *In vivo* ablation of surface immunoglobulin on mature B cells by inducible gene targeting results in rapid cell death. *Cell* 1997;**90**:1073-83.
91. Kraus M, Alimzhanov MB, Rajewsky N, Rajewsky K. Survival of resting mature B lymphocytes depends on BCR signaling via the Igalpha/beta heterodimer. *Cell* 2004;**117**:787-800.
92. Schneider P. The role of APRIL and BAFF in lymphocyte activation. *Curr Opin Immunol* 2005;**17**:282-9.
93. Gross JA, Dillon SR, Mudri S, Johnston J, Littau A, Roque R, Rixon M, Schou O, Foley KP, Haugen H, McMillen S, Waggie K, Schreckhise RW, Shoemaker K, Vu T, Moore M, Grossman A, Clegg CH. TACI-Ig neutralizes molecules critical for B cell development and autoimmune disease. impaired B cell maturation in mice lacking BLyS. *Immunity* 2001;**15**:289-302.
94. Schiemann B, Gommerman JL, Vora K, Cachero TG, Shulga-Morskaya S, Dobles M, Frew E, Scott ML. An essential role for BAFF in the normal development of B cells through a BCMA-independent pathway. *Science* 2001;**293**:2111-4.
95. von Bulow GU, van Deursen JM, Bram RJ. Regulation of the T-independent humoral response by TACI. *Immunity* 2001;**14**:573-82.
96. Yan M, Wang H, Chan B, Roose-Girma M, Erickson S, Baker T, Tumas D, Grewal IS, Dixit VM. Activation and accumulation of B cells in TACI-deficient mice. *Nat Immunol* 2001;**2**:638-43.
97. Seshasayee D, Valdez P, Yan M, Dixit VM, Tumas D, Grewal IS. Loss of TACI causes fatal lymphoproliferation and autoimmunity, establishing TACI as an inhibitory BLyS receptor. *Immunity* 2003;**18**:279-88.
98. Sakurai D, Hase H, Kanno Y, Kojima H, Okumura K, Kobata T. TACI regulates IgA production by APRIL in collaboration with HSPG. *Blood* 2006.
99. Zhang X, Park CS, Yoon SO, Li L, Hsu YM, Ambrose C, Choi YS. BAFF supports human B cell differentiation in the lymphoid follicles through distinct receptors. *Int Immunol* 2005;**17**:779-88.
100. Xu S, Lam KP. B-cell maturation protein, which binds the tumor necrosis factor family members BAFF and APRIL, is dispensable for humoral immune responses. *Mol Cell Biol* 2001;**21**:4067-74.
101. O'Connor BP, Raman VS, Erickson LD, Cook WJ, Weaver LK, Ahonen C, Lin LL, Mantchev GT, Bram RJ, Noelle RJ. BCMA is essential for the survival of long-lived bone marrow plasma cells. *J Exp Med* 2004;**199**: 91-8.
102. Yang M, Hase H, Legarda-Addison D, Varughese L, Seed B, Ting AT. B cell maturation antigen, the receptor for a proliferation-inducing ligand and B cell-activating factor of the TNF family, induces antigen presentation in B cells. *J Immunol* 2005;**175**:2814-24.
103. Notarangelo L, Casanova JL, Conley ME, Chapel H, Fischer A, Puck J, Roifman C, Seger R, Geha RS. Primary immunodeficiency diseases: an update from the International Union of Immunological Societies Primary Immunodeficiency Diseases Classification Committee Meeting in Budapest, 2005. *J Allergy Clin Immunol* 2006;**117**:883-96.
104. Schwarz K, Gauss GH, Ludwig L, Pannicke U, Li Z, Lindner D, Friedrich W, Seger RA, Hansen-Hagge TE, Desiderio S, Lieber MR, Bartram CR. RAG mutations in human B cell-negative SCID. *Science* 1996;**274**: 97-9.
105. Noordzij JG, de Bruin-Versteeg S, Verkaik NS, Vossen JM, de Groot R, Bernatowska E, Langerak AW, van Gent DC, van Dongen JJM. The immunophenotypic and immunogenotypic B-cell differentiation arrest in bone marrow of RAG-deficient SCID patients corresponds to residual recombination activities of mutated RAG proteins. *Blood* 2002;**100**:2145-52.
106. van der Burg M, van Veelen LR, Verkaik NS, Wiegant WW, Hartwig NG, Barendregt BH, Brugmans L, Raams A, Jaspers NG, Zdzienicka MZ, van Dongen JJM, van Gent DC. A new type of radiosensitive T-B-NK+ severe combined immunodeficiency caused by a LIG4 mutation. *J Clin Invest* 2006;**116**:137-45.
107. Buck D, Malivert L, de Chasseval R, Barraud A, Fondaneche MC, Sanal O, Plebani A, Stephan JL, Hufnagel M, le Deist F, Fischer A, Durandy A, de Villartay JP, Revy P. Cernunnos, a novel nonhomologous end-joining factor, is mutated in human immunodeficiency with microcephaly. *Cell* 2006;**124**:287-99.

108. Ben-Omran TI, Cerosaletti K, Concannon P, Weitzman S, Nezarati MM. A patient with mutations in DNA Ligase IV: clinical features and overlap with Nijmegen breakage syndrome. *Am J Med Genet A* 2005;**137**: 283-7.
109. Bruton OC. Agammaglobulinemia. *Pediatrics* 1952;**9**:722-8.
110. Vetrie D, Vorechovsky I, Sideras P, Holland J, Davies A, Flinter F, Hammarstrom L, Kinnon C, Levinsky R, Bobrow M, et al. The gene involved in X-linked agammaglobulinaemia is a member of the src family of protein-tyrosine kinases. *Nature* 1993;**361**:226-33.
111. Tsukada S, Saffran DC, Rawlings DJ, Parolini O, Allen RC, Klisak I, Sparkes RS, Kubagawa H, Mohandas T, Quan S, et al. Deficient expression of a B cell cytoplasmic tyrosine kinase in human X-linked agammaglobulinemia. *Cell* 1993;**72**:279-90.
112. Conley ME, Mathias D, Treadaway J, Minegishi Y, Rohrer J. Mutations in btk in patients with presumed X-linked agammaglobulinemia. *Am J Hum Genet* 1998;**62**:1034-43.
113. Yel L, Minegishi Y, Coustan-Smith E, Buckley RH, Trubel H, Pachman LM, Kitchingman GR, Campana D, Rohrer J, Conley ME. Mutations in the mu heavy-chain gene in patients with agammaglobulinemia. *N Engl J Med* 1996;**335**:1486-93.
114. Minegishi Y, Coustan-Smith E, Wang YH, Cooper MD, Campana D, Conley ME. Mutations in the human lambda5/14.1 gene result in B cell deficiency and agammaglobulinemia. *J Exp Med* 1998;**187**:71-7.
115. Minegishi Y, Coustan-Smith E, Rapalus L, Ersoy F, Campana D, Conley ME. Mutations in Igalpha (CD79a) result in a complete block in B-cell development. *J Clin Invest* 1999;**104**:1115-21.
116. Minegishi Y, Rohrer J, Coustan-Smith E, Lederman HM, Pappu R, Campana D, Chan AC, Conley ME. An essential role for BLNK in human B cell development. *Science* 1999;**286**:1954-7.
117. Allen RC, Armitage RJ, Conley ME, Rosenblatt H, Jenkins NA, Copeland NG, Bedell MA, Edelhoff S, Disteche CM, Simoneaux DK, et al. CD40 ligand gene defects responsible for X-linked hyper-IgM syndrome. *Science* 1993;**259**:990-3.
118. Ferrari S, Giliani S, Insalaco A, Al-Ghonaium A, Soresina AR, Loubser M, Avanzini MA, Marconi M, Badolato R, Ugazio AG, Levy Y, Catalan N, Durandy A, Tbakhi A, Notarangelo LD, Plebani A. Mutations of CD40 gene cause an autosomal recessive form of immunodeficiency with hyper IgM. *Proc Natl Acad Sci U S A* 2001;**98**:12614-9.
119. Weller S, Braun MC, Tan BK, Rosenwald A, Cordier C, Conley ME, Plebani A, Kumararatne DS, Bonnet D, Tournilhac O, Tchernia G, Steiniger B, Staudt LM, Casanova JL, Reynaud CA, Weill JC. Human blood IgM "memory" B cells are circulating splenic marginal zone B cells harboring a prediversified immunoglobulin repertoire. *Blood* 2004;**104**:3647-54.
120. Revy P, Muto T, Levy Y, Geissmann F, Plebani A, Sanal O, Catalan N, Forveille M, Dufourcq-Labelouse R, Gennery A, Tezcan I, Ersoy F, Kayserili H, Ugazio AG, Brousse N, Muramatsu M, Notarangelo LD, Kinoshita K, Honjo T, Fischer A, Durandy A. Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2). *Cell* 2000;**102**:565-75.
121. Imai K, Slupphaug G, Lee WI, Revy P, Nonoyama S, Catalan N, Yel L, Forveille M, Kavli B, Krokan HE, Ochs HD, Fischer A, Durandy A. Human uracil-DNA glycosylase deficiency associated with profoundly impaired immunoglobulin class-switch recombination. *Nat Immunol* 2003;**4**:1023-8.
122. Grimbacher B, Hutloff A, Schlesier M, Glocker E, Warnatz K, Drager R, Eibel H, Fischer B, Schaffer AA, Mages HW, Kroczek RA, Peter HH. Homozygous loss of ICOS is associated with adult-onset common variable immunodeficiency. *Nat Immunol* 2003;**4**:261-8.
123. Salzer U, Maul-Pavicic A, Cunningham-Rundles C, Urschel S, Belohradsky BH, Litzman J, Holm A, Franco JL, Plebani A, Hammarstrom L, Skrabl A, Schwinger W, Grimbacher B. ICOS deficiency in patients with common variable immunodeficiency. *Clin Immunol* 2004;**113**:234-40.
124. Castigli E, Wilson SA, Garibyan L, Rachid R, Bonilla F, Schneider L, Geha RS. TACI is mutant in common variable immunodeficiency and IgA deficiency. *Nat Genet* 2005;**37**:829-34.
125. Salzer U, Chapel HM, Webster AD, Pan-Hammarstrom Q, Schmitt-Graeff A, Schlesier M, Peter HH, Rockstroh JK, Schneider P, Schaffer AA, Hammarstrom L, Grimbacher B. Mutations in TNFRSF13B encoding TACI are associated with common variable immunodeficiency in humans. *Nat Genet* 2005;**37**:820-8.
126. Castigli E, Geha RS. Molecular basis of common variable immunodeficiency. *J Allergy Clin Immunol* 2006;**117**:740-6; quiz 7.

127. Di Renzo M, Pasqui AL, Auteri A. Common variable immunodeficiency: a review. *Clin Exp Med* 2004;**3**: 211-7.
128. Tcheurekdjian H, Palermo T, Hostoffer R. Quality of life in common variable immunodeficiency requiring intravenous immunoglobulin therapy. *Ann Allergy Asthma Immunol* 2004;**93**:160-5.
129. Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res* 2005;**33**:D514-7.
130. Stenson PD, Ball EV, Mort M, Phillips AD, Shiel JA, Thomas NS, Abeysinghe S, Krawczak M, Cooper DN. Human Gene Mutation Database (HGMD): 2003 update. *Hum Mutat* 2003;**21**:577-81.
131. Piirila H, Valiaho J, Vihinen M. Immunodeficiency mutation databases (IDbases). *Hum Mutat* 2006;**27**: 1200-8.
132. Antonarakis SE, Krawczak M, Cooper DN. Disease-causing mutations in the human genome. *Eur J Pediatr* 2000;**159 Suppl 3**:S173-8.
133. Maki H. Origins of spontaneous mutations: specificity and directionality of base-substitution, frameshift, and sequence-substitution mutageneses. *Annu Rev Genet* 2002;**36**:279-303.
134. Rogozin IB, Pavlov YI. Theoretical analysis of mutation hotspots and their DNA sequence context specificity. *Mutat Res* 2003;**544**:65-85.
135. Youssoufian H, Kazazian HH, Jr., Phillips DG, Aronis S, Tsiftis G, Brown VA, Antonarakis SE. Recurrent mutations in haemophilia A give evidence for CpG mutation hotspots. *Nature* 1986;**324**:380-2.
136. Youssoufian H, Antonarakis SE, Bell W, Griffin AM, Kazazian HH, Jr. Nonsense and missense mutations in hemophilia A: estimate of the relative mutation rate at CG dinucleotides. *Am J Hum Genet* 1988;**42**:718-25.
137. Rideout WM, 3rd, Coetzee GA, Olumi AF, Jones PA. 5-Methylcytosine as an endogenous mutagen in the human LDL receptor and p53 genes. *Science* 1990;**249**:1288-90.
138. Green PM, Montandon AJ, Bentley DR, Ljung R, Nilsson IM, Giannelli F. The incidence and distribution of CpG----TpG transitions in the coagulation factor IX gene. A fresh look at CpG mutational hotspots. *Nucleic Acids Res* 1990;**18**:3227-31.
139. Cooper DN, Youssoufian H. The CpG dinucleotide and human genetic disease. *Hum Genet* 1988;**78**:151-5.
140. Krawczak M, Cooper DN. Gene deletions causing human genetic disease: mechanisms of mutagenesis and the role of the local DNA sequence environment. *Hum Genet* 1991;**86**:425-41.
141. Streisinger G, Okada Y, Emrich J, Newton J, Tsugita A, Terzaghi E, Inouye M. Frameshift mutations and the genetic code. *Cold Spring Harb Symp Quant Biol* 1966;**31**:77-84.
142. Kunkel TA, Soni A. Mutagenesis by transient misalignment. *J Biol Chem* 1988;**263**:14784-9.
143. Kunkel TA. Misalignment-mediated DNA synthesis errors. *Biochemistry* 1990;**29**:8003-11.
144. Kimura H, Iyehara-Ogawa H, Kato T. Slippage--misalignment: to what extent does it contribute to mammalian cell mutagenesis? *Mutagenesis* 1994;**9**:395-400.
145. Abeysinghe SS, Chuzhanova N, Krawczak M, Ball EV, Cooper DN. Translocation and gross deletion breakpoints in human inherited disease and cancer I: Nucleotide composition and recombination-associated motifs. *Hum Mutat* 2003;**22**:229-44.
146. McNaughton JC, Cockburn DJ, Hughes G, Jones WA, Laing NG, Ray PN, Stockwell PA, Petersen GB. Is gene deletion in eukaryotes sequence-dependent? A study of nine deletion junctions and nineteen other deletion breakpoints in intron 7 of the human dystrophin gene. *Gene* 1998;**222**:41-51.
147. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M,

Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blocker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ. Initial sequencing and analysis of the human genome. *Nature* 2001;**409**:860-921.

148. Kohany O, Gentles AJ, Hankus L, Jurka J. Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. *BMC Bioinformatics* 2006;**7**:474.

II

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[*]*

[*]Erasmus MC, Department of Immunology, Rotterdam, the Netherlands; [†]Erasmus MC, Department of Pediatrics, Rotterdam, the Netherlands; [‡]Information & Communication Theory Group, Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, the Netherlands; [§]Department of Pediatrics, Leiden University Medical Center, Leiden, the Netherlands; [¶]Hematology-Oncology Unit, Wilhelmina Children's Hospital, University Medical Center, Utrecht, the Netherlands.

J Immunol 2005;175(9):5912-22

Erratum: J Immunol 2006;176(12):7787

Copyright 2005 and 2006 The American Association of Immunologists, Inc.

ABSTRACT

The role of specific transcription factors in the initiation and regulation of Ig gene rearrangements has been studied extensively in mouse 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_H-J_H rearrangements, which are completed with V_H-DJ_H rearrangements in pre-B-I cells. Large cycling pre-B-II cells are selected for in-frame *IGH* gene rearrangements. The first *IGK/IGL* gene rearrangements were initiated in pre-B-I cells, but their frequency increased enormously in small pre-B-II cells, and in-frame selection was found in immature B cells. Transcripts of the *RAG1* and *RAG2* genes, and earlier defined transcription factors, such as *E2A*, *EBF*, *E2-2*, *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-cell acute lymphoblastic leukemia and B-cell differentiation blocks in primary antibody deficiencies.

KEY WORDS

B-cell differentiation, V(D)J recombination, expression profiling, *IGH*, *IGK*, *IGL*.

The online version of this article contains supplemental material.

INTRODUCTION

Precursor-B cells develop from hematopoietic 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 antigen receptors (BCR), 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 heavy chain (*IGH*) locus with D_H 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 non-homologous 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 light chain rearrangements are initiated in the small pre-B-II cells. If a functional Ig light chain (Igκ or Igλ) is expressed that is able to assemble with the IgH chain, the cell becomes a surface membrane (Sm)Ig^+ immature B cell. If this cell is non-autoreactive, it migrates to the periphery as a naive mature B cell.

The differential expression of marker molecules is used to define the main stages of B-cell differentiation. Five main human precursor-B-cell differentiation stages can be recognized using the following markers: 1) $CD22^+CD19^-$ pro-B cells; 2) $CD19^+$ cytoplasmic (Cy)$Ig\mu^-$ pre-B-I cells; 3) $CyIg\mu^+VpreB^+$ pre-B-II large cells; 4) $CyIg\mu^+VpreB^-$ pre-B-II small cells; 5) $SmIgM^+$ immature B cells. Naive mature B lymphocytes in the periphery are $SmIgM^+SmIgD^+$.[6, 7]

Although there is a common V(D)J recombinase machinery, Ig gene rearrangements are separated in time and restricted to one locus. Transcription factors PAX5, early B-cell factor (EBF), and the E-box proteins E2A, HEB and E2-2 appear to act in a hierarchical order for commitment to B-cell fate and the initiation of Ig gene rearrangements by opening Ig loci,[8-12] and by inducing expression of recombination-related proteins (RAG1, RAG2, TdT) and (pre-)BCR proteins (VpreB, λ14.1, CD79A).[13-16] The Ets transcription factor PU.1 is required for generation of lymphoid as well as myeloid cells,[17, 18] and is suggested to be the partner of IRF4 and IRF8 in large pre-B cells, where they inhibit proliferation and initiate Ig light 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 (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 (age 3-16), 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, Cologne, Germany) were used to positively select $CD34^+$ progenitor cells. $CD34^+lin^-$ cells were further purified by sorting on a FACSDiVa cell sorter (BD Biosciences, Santa Clara, CA) after labeling with CD34-APC (8G12), and CD3-PE (Leu-4), CD16-PE (B73.1), CD19-PE (4G7), CD56-PE (M431; all from BD Biosciences), CD13-RDI (MY7), CD33-RDI (906; both from Beckman Coulter, Fullerton, CA) for exclusion. Pro-B cells were further purified after labeling with CD34-FITC, CD19-PE, CD123-PE (9F5), 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 (HI10a; both from BD Biosciences), and anti-IgM-Cy5 (Jackson Immunoresearch Laboratories, Inc, West Grove, PA). Additional markers for defining and characterization of the precursor-B-cell fractions were: Ki67-FITC, anti-Igκ-FITC (both from DakoCytomation, Glostrup, Denmark), anti-Igλ-PE (Southern Biotechnology Associates, Inc, Birmingham, AI) and CD79b-PE (SN8; BD Biosciences). All fractions were collected in 2 ml tubes containing filtered heat-inactivated FCS and were kept at 4°C until reanalysis on a FACSCalibur (BD Biosciences) and subsequent isolation of DNA and/or RNA.

Mature $Ig\kappa^+$ and $Ig\lambda^+$ 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 $Ig\kappa^+$ and mature $Ig\lambda^+$ B 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, the Leiden University Medical Center and the Erasmus MC, University Medical Center Rotterdam.

Real-time quantitative PCR (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 in such a way that >95 % of all rearrangements could be detected (Appendix 1).[26-31] The RQ-PCR mixture of 25 µl contained TaqMan Universal MasterMix (Applied Biosystems, Foster City, CA), 900 nM of each primer (300 nM in case of multiplex mixtures), 100 nM of each FAM-TAMRA labeled probe, 50 ng of DNA, 0.4 ng BSA, and was run on the ABIPRISM 7700 sequence detection system (Applied Biosystems).[32, 33]

An albumin RQ-PCR was 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 (un-rearranged) 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 monoallelically 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 (Appendix 2).[30]

RNA isolation and gene expression profiling

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

Due to the limited amount of material, 100 ng of total RNA was 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 calls were comparable between all samples. GeneChip array data was 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 analysis of variance (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 step-down adjustment,[39] and all differences with adjusted *p*-values < 0.05 were considered significant, resulting in 5,365 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 5,365 significant probe sets were normalized per-probe set to zero mean and unit standard deviation (*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:6266, GO:6281, GO:6282, GO:6302, GO:6303, GO:6310, GO:6350, GO:6351, GO:6355, GO:6366, GO:6367, 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 (Appendix 3). The expression levels were corrected by comparison with control genes Abelson (*ABL*) and beta-glucuronidase (*GUSB*) as described previously.[40] The expression values of the genes were normalized per-gene to zero mean and unit standard deviation (*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 in order to reliably obtain both DNA and RNA from these cells (Figure 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 cyIgμ. CyIgμ$^-$ pre-B-I cells were isolated based on the phenotype: $CD19^+CD34^+CD10^+CD20^-$. In addition, large cycling pre-B-II cells ($CD19^+CD34^-CD10^+CD20^{dim}IgM^-$), small pre-B-II cells ($CD19^+CD34^-CD10^+CD20^-$), and immature B cells ($CD19^+CD34^-CD10^+CD20^{hi}$) were isolated. Mature B lymphocytes ($CD19^+CD20^+$) were obtained from tonsillar tissue and separated in either SmIgκ$^+$ or SmIgλ$^+$ populations based on membrane Ig light chain expression (Figure 1).

Incomplete D_H–J_H gene rearrangements are initiated in pro-B cells, directly followed by complete V_H–DJ_H rearrangements in pre-B-I cells

The human *IGH* locus contains 66 rearrangeable V, 27 D, and 6 J gene segments (Figure 2A) that can be involved in rearrangements in B cells to form a VDJ_H exon.

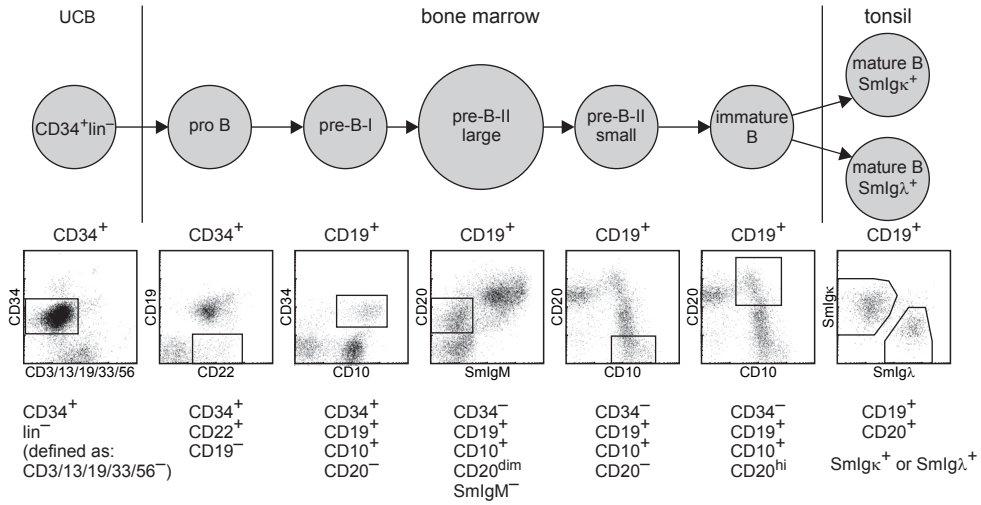


Figure 1. Isolation of CD34$^+$lin$^-$ and (precursor-)B-cell subsets using membrane markers only.
The subsets studied here, their origin, and the markers that were used to isolate them are indicated.

Two multiplex RQ-PCR assays were designed that recognize nearly all these gene segments for both the incomplete (DH–JH) and complete (VH–DJH) rearrangements. No DH–JH rearrangements were found in DNA of CD34$^+$lin$^-$ cells, but they were abundantly present in pro-B, pre-B-I and large pre-B-II cells (Figure 2B). In all subsequent stages the percentage of DH–JH rearrangements was lower, due to initiation of VH–DJH rearrangements, which results in loss of the intron sequence upstream of the D segments in which the forward DH primers are located. VH–DJH rearrangements were not detectable until the pre-B-I stage (Figure 2C). In large pre-B-II cells the frequency of VH–DJH rearrangements reached its maximum and remained constant in all subsequent stages.

These results show that already in the first stage of precursor-B-cell differentiation a large number of incomplete *IGH* rearrangements are formed. This is directly followed by complete VH–DJH rearrangements in pre-B-I cells, which reach maximal levels in large pre-B-II cells.

IGK and *IGL* rearrangements are initiated in small pre-B-II cells, but Vλ–Jλ rearrangements are rare in mature Igκ$^+$ B cells

The human *IGK* locus contains 76 rearrangeable Vκ and 5 Jκ gene segments. In addition to the Vκ–Jκ rearrangements there are two types of *IGK* deleting rearrangements involving the kappa-deleting element (Kde) that make the *IGK* allele non-functional.[41] Vκ–Kde rearrangements can delete an initial Vκ–Jκ and the Cκ exon. Kde can also rearrange to the intronRSS, which is located in between Jκ5 and Cκ. The intronRSS–Kde rearrangement deletes Cκ, but keeps the initial Vκ–Jκ rearrangement present on the locus (Figure 3A).

In CD34$^+$lin$^-$, pro-B, pre-B-I and large pre-B-II cells, Vκ–Jκ rearrangements were barely detectable (Figure 3B). An enormous increase was found in small pre-B-II and immature B cells, which represented approximately 75 % of the frequency of Vκ–Jκ rearrangements

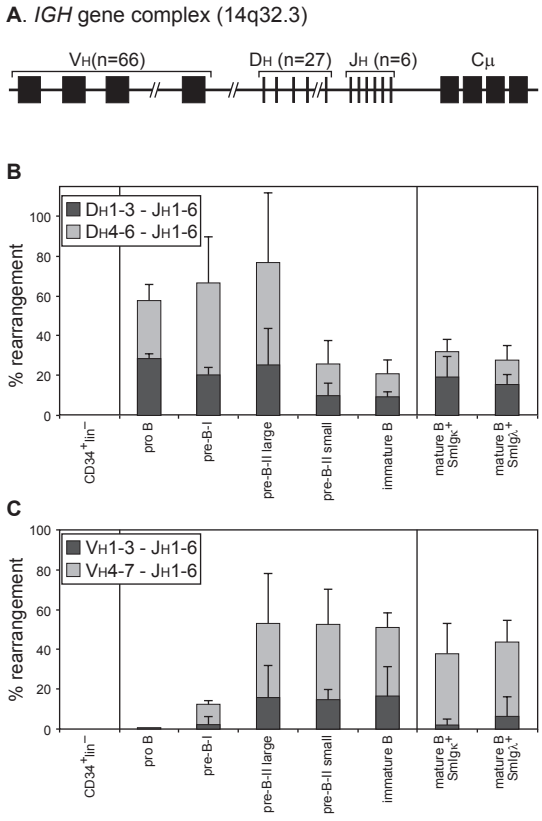


Figure 2. Incomplete DH–JH gene rearrangements are initiated in pro-B cells and are directly followed by complete VH–DJH rearrangements in pre-B-I cells. **A.** Organization of the human *IGH* locus; **B.** DH–JH rearrangements were quantified in all isolated subpopulations in two separate multiplex TaqMan RQ-PCR assays with primers and probes that detect most members of the six main DH gene segment families and the six JH gene segments; **C.** Quantification of VH–DJH rearrangements in two separate multiplex TaqMan RQ-PCR assays using primers that recognize most members of the seven VH gene segment families and the six JH 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.

observed in mature Igκ$^+$ B cells. The majority of the Vκ–Kde and the intronRSS–Kde rearrangements were induced in pre-B-II small cells as well. In contrast to Vκ–Jκ rearrangements, the Kde rearrangements occurred at low frequency in mature Igκ$^+$ B cells. The fact that they occurred at maximal levels in mature Igλ$^+$ B cells (Figure 3, C and D) indicates that most *IGK* alleles are deleted in mature Igλ$^+$ B cells.

The human *IGL* locus consists of 56 rearrangeable V gene segments and 7 Jλ-Cλ clusters (Figure 4A), of which Jλ1-3 are used in 99 % of the Vλ–Jλ rearrangements.[42] The Vλ–Jλ rearrangements in precursor-B-cell subsets showed the same pattern as Vκ–Jκ, but the levels of Vλ–Jλ rearrangements appeared to be lower and were about half of the Vλ–Jλ levels in mature Igλ$^+$ B cells (Figure 4B). The absence of Vλ–Jλ rearrangements in mature Igκ$^+$ B cells (<5 %) indicated that most *IGL* rearrangements were initiated only if no functional Igκ 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 non-templated (N)-nucleotide addition by TdT, the size of the junctional

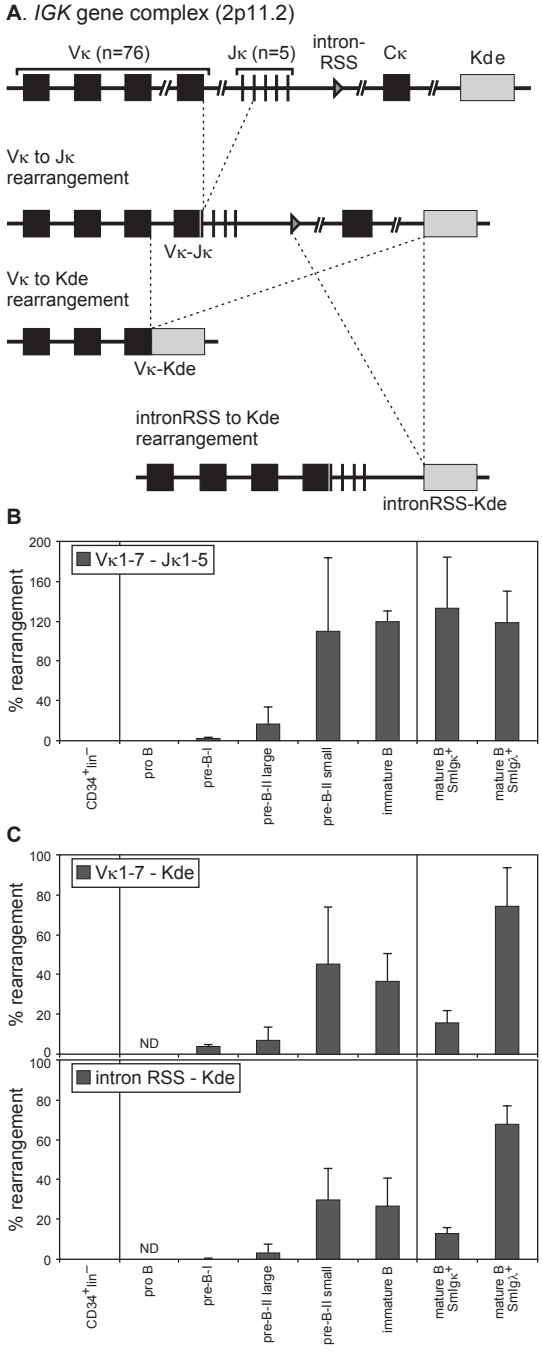


Figure 3. The vast majority of Vκ–Jκ and Kde rearrangements are initiated in small pre-B-II cells.

A. Organization of the human *IGK* locus and the Vκ–Jκ and the Kde rearrangements (Vκ–Kde and intronRSS–Kde); **B.** Vκ–Jκ rearrangements were quantified in all isolated subpopulations in a single multiplex TaqMan RQ-PCR assay with primers and probes that detect most members of seven Vκ gene segment families and the five Jκ gene segments; **C.** Vκ–Kde and intronRSS–Kde rearrangements were quantified in the isolated subpopulations in two separate TaqMan RQ-PCR assays with primers and probes that detect most members of seven Vκ gene segment families, intronRSS and Kde. The Kde assays were not performed on pro-B cells due to limited availability of material. 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. The y-axis of panel B is extended to 200 %, because the amount of Vκ–Jκ rearrangements exceeds the 100 % level in several stages, indicating the presence of biallelic rearrangements.

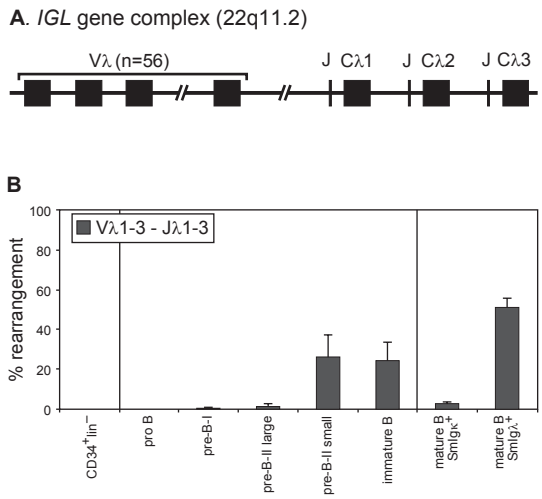


Figure 4. The vast majority of Vλ–Jλ rearrangements are initiated in small pre-B-II cells, but seem to occur only after rearrangement of the *IGK* alleles. **A.** Organization of the human *IGL* locus; **B.** Vλ–Jλ rearrangements were quantified for all isolated subpopulations in a single multiplex TaqMan RQ-PCR assay with primers and probes that detect most members of the three most frequently used Vλ gene segment families and the three dominant Jλ 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.

region is variable and only one-third of the rearrangements is in-frame. Using GeneScan analysis, the VH–DJH, Vκ–Jκ, and Vλ–Jλ 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 symmetric unimodal with peaks at every third nucleotide (triplet peaks), representing in-frame rearrangements (Figure 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 absence of D gene segments and lower levels of N-nucleotide insertion (Figure 5, B and C). 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 to 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 to mature Igκ$^+$ B cells and Igλ$^+$ B cells, respectively, because the immature B-cell subset consisted of a mixture of Igκ$^+$ and Igλ$^+$ B cells.

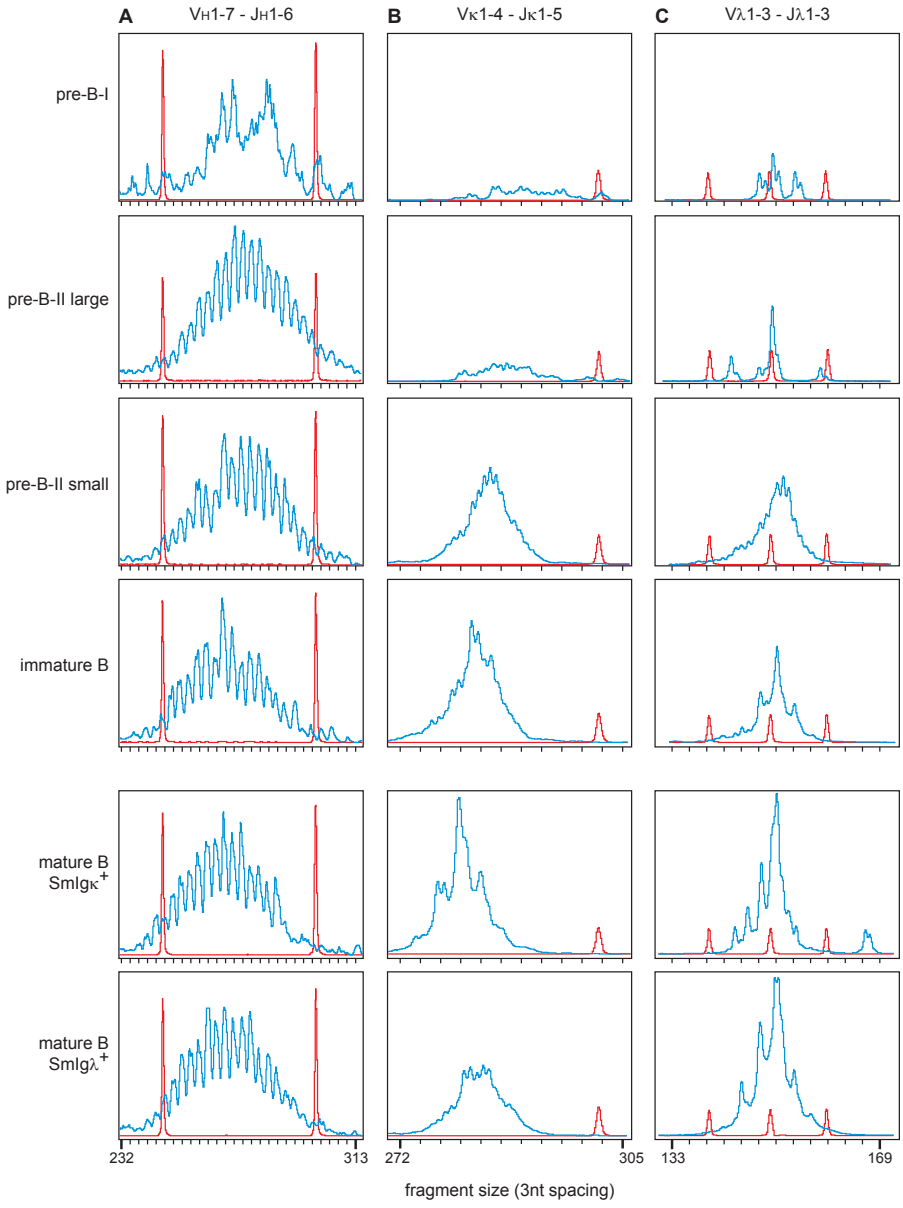
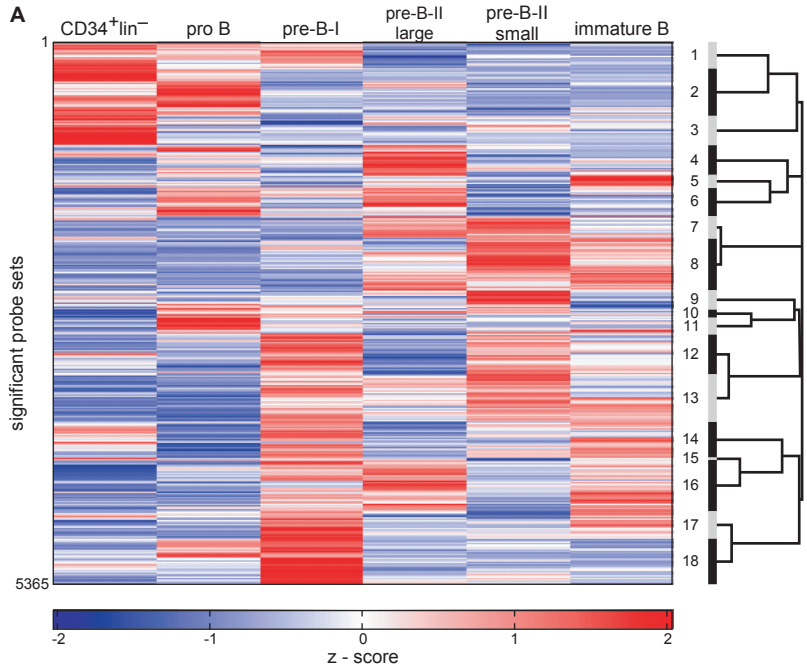


Figure 5. In-frame selection of *IGH* and *IGK/IGL* gene rearrangements occurs in the stage directly after initiation of the respective complete rearrangements.
Analysis of in-frame selection on VH–DJH, Vκ–Jκ, Vλ–Jλ rearrangements using GeneScan analysis, as assessed via triplet peaks, of which the position is indicated at the bottom of each panel. In addition to the size distribution of the junctional regions, the internal size standard peaks are shown. The triplet peaks of Vκ–Jκ and Vλ–Jλ rearrangements were most clear in mature Igκ$^+$ and Igλ$^+$ B cells, respectively. In immature B cells, the triplet peaks were less visible because this was a mixture of Igκ$^+$ and Igλ$^+$ B cells. Data are representative of two or three independent experiments.



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 5,365 probe sets that were at least once differentially expressed between two subsequent stages of differentiation (Figure 6A and Supplemental Table I, online). Hierarchical clustering was performed to group the expression patterns in 18 different clusters (Figure 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 (Figure 6B and Supplemental Table II, online). The gene expression patterns of the markers that were used to isolate the six subsets (*CD34*, *CD19*, *CD10*, *CD22*, *CD20*, and *KI67*) corresponded well with their protein expression (Figure 6B and Supplemental Table II, online). Stem cell markers (*FLT3*, *KIT*), myeloid genes (*MPO*) and T-cell receptor 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

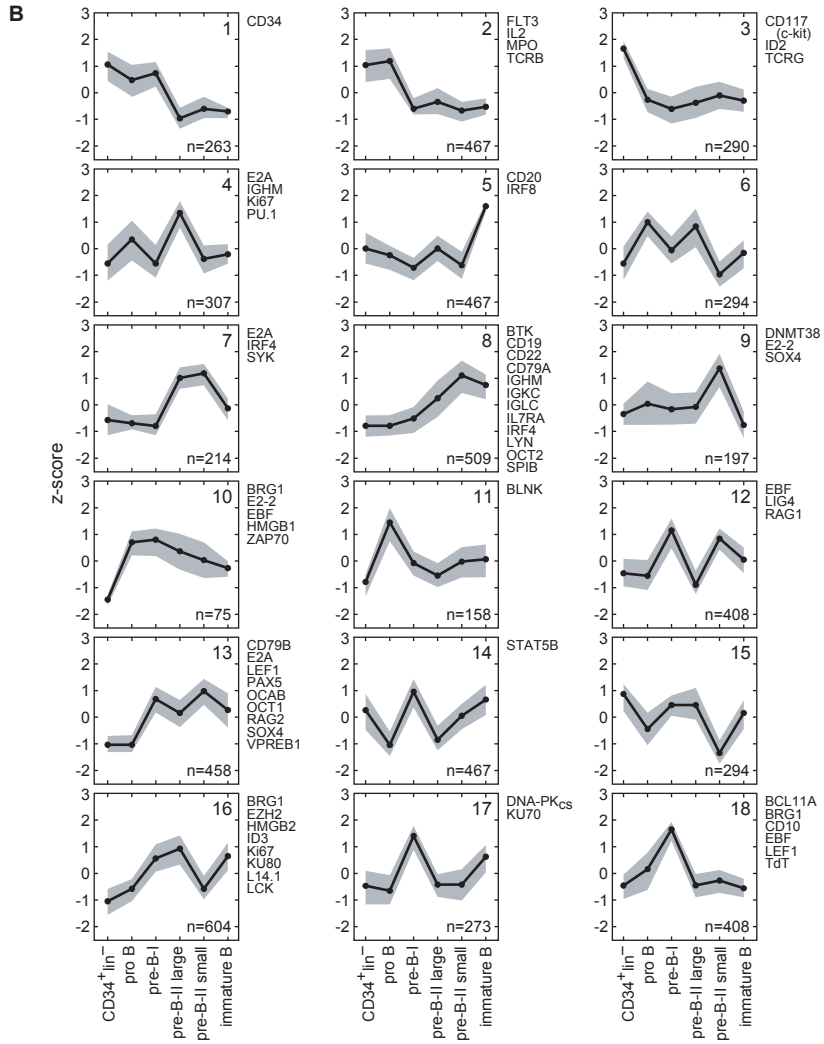


Figure 6. Gene expression profiling of human CD34$^+$lin$^-$- and precursor-B-cell subsets.
A. (facing page) Heat map with the normalized data of significantly differentially expressed probe sets. The expression patterns were grouped into 18 clusters; **B.** Expression pattern of all probe sets per cluster 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*, *HEB*, *ID1*, *ID4*, *NOTCH1*, *PLCG2*, *STAT5A*, *TCRA/D*, *XRCC4*, *IKAROS*, *HELIOS*, *AIOLOS*.

Ig gene rearrangements were found. These early transcripts were most likely germline transcripts. Transcripts of (pre-)BCR complex members *L14.1*, *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 *IGK/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 (Figure 7 and Supplemental Table III, online). 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 four times as many genes were up-regulated (143 genes) than down-regulated (40 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 genes). Interestingly, this indicates a trend that many transcription-related genes are down-regulated in subsets undergoing selection compared with subsets undergoing V(D)J recombination.

The prototypes of nine of the eighteen defined clusters (Figure 6B) showed peaks in pro-B, pre-B-I and/or small pre-B-II cells during which D_H-J_H, V_H-DJ_H and V_L-J_L rearrangements take place, respectively. Therefore, these clusters were more likely to contain genes that are involved in the initiation of Ig gene rearrangements than the other nine clusters (Table I). These nine clusters contained 243 genes, of which most belong to families of

transcription factors based on sequence and/or functional homology. Most genes shown in Table I contained zinc fingers and were related to the Krüppel-like factor (KLF) family of DNA-binding transcriptional regulators. A large number of zinc finger proteins were annotated solely based on sequence homology and due to the lack of additional data from literature they will not be further considered here.

Within the remaining set of genes, a selection of 16 genes was made based on an earlier suggested role of their protein products or proteins related to them in lymphoid development (shown in bold in Table I). The expression patterns of these 16 genes, and *RAG1* and *RAG2* were confirmed with RQ-PCR (Figure 8). *KLF2* was highly up-regulated in pre-B-I cells, in which *IGH* gene rearrangements were initiated, whereas *KLF4*, *COPEB* and *KLF12* were up-regulated in both pre-B-I and small pre-B-II, where *IGK/IGL* gene rearrangements were initiated. Six Ets-family members were found in the nine clusters of Table I. *ERG* was specifically up-regulated in pre-B-I cells. Furthermore, *ELK3* and *ETS2* expression levels showed peaks in both pre-B-I and small pre-B-II cells, and *ELF1*, *ETS1* and *SPIB* were not up-regulated before the pre-B-II large stage. Four POU-domain proteins were found to be up-regulated in precursor-B cells undergoing complete V to (D)J gene rearrangements. *OCT1* was highly expressed in both pre-B-I and small pre-B-II cells, whereas *OCAB*, *OCT2* and *POU4F1* were up-regulated specifically in small pre-B-II cells. Not only *EZH2* and *BRG1*, but also transcripts of the chromatin-remodeling factor *SMARCA5* were differentially expressed with peaks in pre-B-I and small pre-B-II cells. The κB DNA binding and recognition component *HIVEP3* is a member of the ZAS family and was found to be up-regulated in small pre-B-II cells.

In summary, a selection was made within the 5,365 significant probe sets for genes that encode transcriptional regulators. By relating their gene expression profiles to the rearrangement status of the precursor-B-cell subsets, 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 genome-wide gene expression profiles. Incomplete D_H–J_H rearrangements were initiated in pro-B cells, whereas V_H–DJ_H rearrangements were initiated in pre-B-I cells. Large cycling pre-B cells were selected for in-frame V_H–DJ_H 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 Figure 8).

Incomplete *IGH* rearrangements were already abundantly present in pro-B cells, while *RAG1* and *RAG2* expression levels were hardly up-regulated (Figure 6B and Figure 8B). These data suggest that D_H–J_H rearrangements can be initiated with much lower levels of

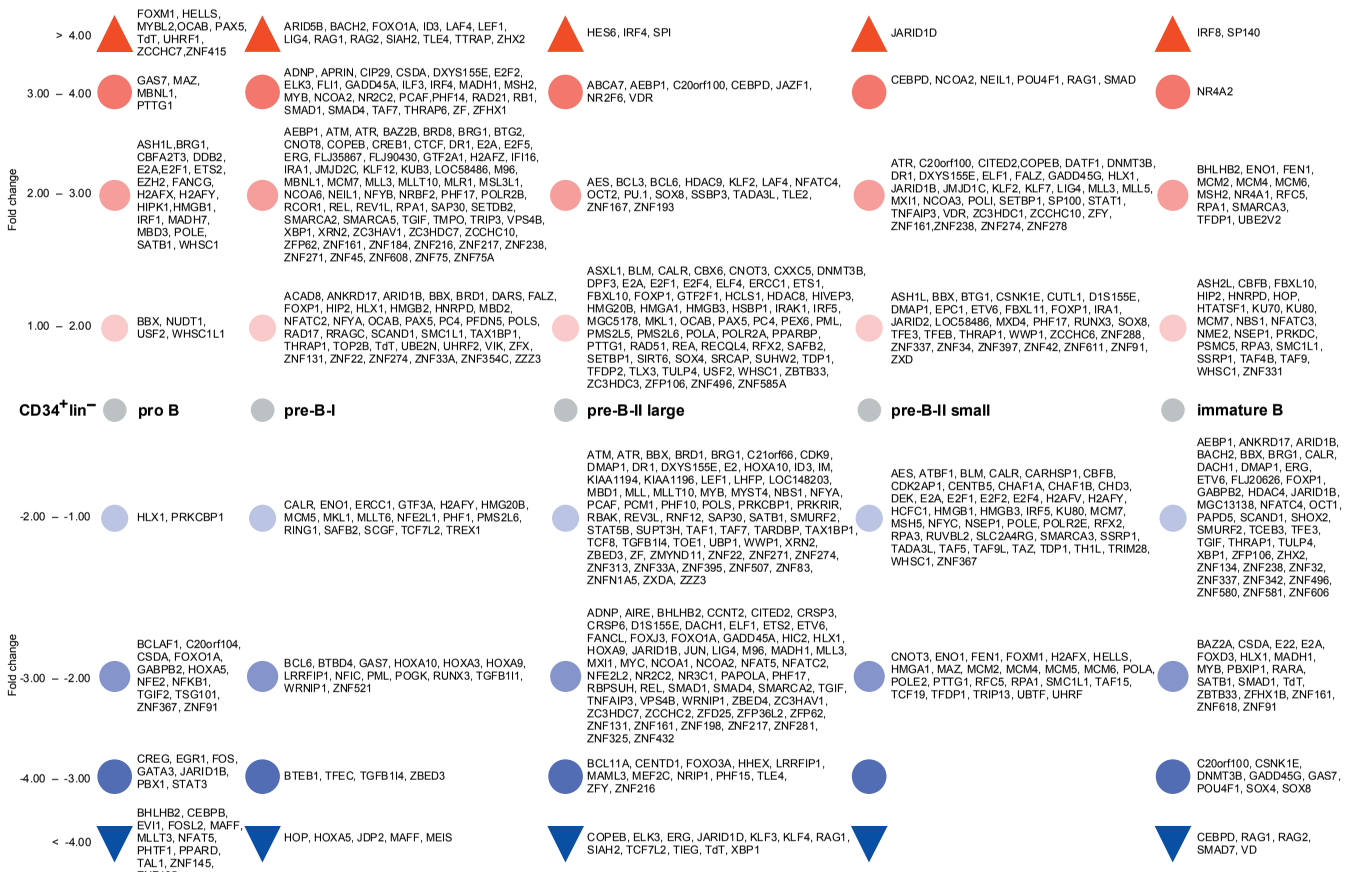


Figure 7. Differential expression of transcription factors and recombination-related genes.
Graphs show differential expression of all genes encoding transcriptional regulators between subsequent stages of differentiation. Within each category, the genes are ordered alphabetically. In the cases in which more than one probe set representing the same gene showed a significant difference between two subsequent precursor-B-cell subsets, only the data from the probe set that showed the largest fold difference is shown.

RAG1/RAG2 than complete VH–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 VH 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 Vλ–Jλ, 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–Jκ junctional regions was larger in pre-B-I and large pre-B-II cells than in pre-B-II small cells (Figure 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. 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 Vλ–Jλ rearrangements in mature Igκ$^+$ B cells was low compared with mature Igλ$^+$ B cells, whereas Vκ–Jκ was high in both. This confirms the hierarchical order that has been found in previous studies;[31, 45] Kde and Vλ–Jλ rearrangements are only induced 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 DH–JH rearrangements. Furthermore, pro-B cells showed up-regulated expression of genes encoding EZH2 and SMARCA4, which modify the chromatin structure and thereby make the *IGH* locus accessible for V(D)J recombination.[21, 22]

PAX5 is required for VH–DJH rearrangements,[11] and it 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 VH to DJH gene rearrangements. In mice, IL-7R signaling induces VH–DJH rearrangements via STAT5 activation.[46] *STAT5B* gene expression was indeed up-regulated in human pre-B-I cells, but *IL7RA* was not expressed

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

Peak in expression	IGH				IGH and IGK/IGL			IGK/IGL	
Category/Cluster	10	11	17	18	12	13	7	8	9
1. DNA binding TF [a]	*E2-2*	*ASH1L*	*ARID5B*	*AIRE*	*BAZ2B*	*ASXL1*	*AEBP1*	*ABCA7*	*BAZ2A*
	HMGB1	*CBFA2T3*	*C21orf66*	*BCL11A*	*BBX*	*BACH2*	*E2A*	*ACAD8*	*C20orf100*
	IRF1	*HIPK1*	*CIP29*	***COPEB*** [b]	*BRD1*	*BBX*	*FOXP1*	*BTG1*	*CALR*
	LOC148203	*LRRFIP1*	*FBXL10*	*DACH1*	*CENTD1*	*BRD8*	*HES6*	*CXXC5*	*CEBPD*
	MYBL2	*MBNL1*	*HIP2*	***ELK3***	*CITED2*	*BTG2*	*IRF4*	*DATF1*	*CUTL1*
	UHRF1	*MLLT6*	*HTATSF1*	***ERG***	*CRSP3*	*CNOT8*	***KLF2***	*DPF3*	*E2-2*
	WHSC1L1	*ZNF288*	*ILF3*	***ETS2***	*DR1*	*CREB1*	*NFATC4*	*DR1*	*FLJ20626*
	ZCCHC7	*ZNF42*	*ING5*	*JUN*	*DXYS155E*	*CSDA*	*NR2F6*	*E2F5*	*FOXD3*
	ZNF367		*KIAA1194*	*LEF1*	***ELF1***	*CTCF*	*RARA*	***ETS1***	*GAS7*
			MLLT10	*M96*	*FBXL11*	*DXYS155E*	*SHOX2*	*FALZ*	*PBXIP1*
			MLR1	*MEF2C*	*FOXJ3*	*E2A*	*SIRT6*	*FOXP1*	*SATB1*
			PCM1	*MLL*	*FOXO1A*	*FLI1*	*TADA3L*	*HCLS1*	*SCAND1*
			PSMC5	*PHF15*	*HLX1*	*FLJ35867*	*TLE2*	***HIVEP3***	*SMAD7*
			TARDBP	*RBAK*	***KLF4***	*FOXO1A*	*ZBTB33*	*IRAK1*	*SOX4*
			ZC3HAV1	*SIAH2*	*LOC58486*	*GABPB2*	*ZFP106*	*IRF4*	*SOX8*
			ZF	*TAF9L*	*MADH1*	*GTF2A1*	*ZHX2*	*JARID2*	*VDR*
			ZNF608	*TIEG*	*MLL5*	*IRA1*	*ZNF167*	*JAZF1*	*ZFHX1B*
			ZZZ3	*TOE1*	*MXI1*	*JMJD2C*	*ZNF193*	*JMJD1C*	*ZNF134*
				XBP1	*NFYA*	***KLF12***	*ZNF342*	***KLF2***	*ZNF32*
				ZFP62	*NR2C2*	*LEF1*	*ZNF496*	*LAF4*	*ZNF337*
				ZNF216	*NR3C1*	*MYB*		*NCOA2*	*ZNF581*
				ZNF22	*PHF17*	***OCAB***		*NCOA3*	*ZNF606*
				ZNF33A	*SMAD1*	***OCT1***		***OCT2***	*ZNF618*
				ZNF507	*SMURF2*	*NFATC3*		***POU4F1***	*ZNF91*
					TAF1	*PAX5*		*SP100*	
					TAF7	*PC4*		***SPIB***	
					TAX1BP1	*RB1*		*STAT1*	
					TGIF	*RRAGC*		*TFDP2*	
					TLE4	*SCAND1*		*TFEB*	
					UBP1	*SOX4*		*TULP4*	
					WWP1	*THRAP1*		*ZC3HDC1*	
					ZBED4	*THRAP6*		*ZCCHC10*	

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

Peak in expression	IGH				IGH and IGK/IGL			IGK/IGL	
Category/Cluster	10	11	17	18	12	13	7	8	9
1. DNA binding TF [a]					*ZFD25*	*UHRF2*		*ZNF34*	
					ZNF161	*ZCCHC6*		*ZNF611*	
					ZNF198	*ZNF238*			
					ZNF217	*ZNF75A*			
					ZNF274				
					ZNF278				
					ZNF281				
					ZNF325				
					ZNF397				
					ZNF432				
					ZNFN1A5				
					ZXDA				
2. DNA methylation				*HEMK*	*DMAP1*	*MBD2*			*DNMT3B*
3. Chromatin and	*BRG1*	*H2AFY*		*SAP30*	*MLL3*	*ARID1B*		*HDAC9*	*HDAC4*
Histone		*PHF1*		*SMARCA4*		*EPC1*		*MSL3L1*	
Remodeling						*NCOA6*		*VIK*	
						PCAF			
						SETDB2			
						SMARCA5			
4. Recombination			*POLS*	*TdT*	*RAG1*	*PAPD5*			
						RAG2			
5. DNA repair		*XRN2*	*ATM*		*KUB3*	*ANKRD17*		*NEIL1*	*CSNK1E*
			FANCL		*POLI*	*ATM*			*GADD45G*
			KU70			*ATR*			*TDP1*
			MSH2			*BTG2*			*XRN2*
			NBS1			*CSNK1E*			
			PRKDC			*NCOA6*			
			TREX1			*RAD21*			
						UBE2V2			

[a] TF, transcription factor.

[b] 16 likely candidate genes involved in initiation and/or regulation of Ig gene rearrangements are shown in bold font

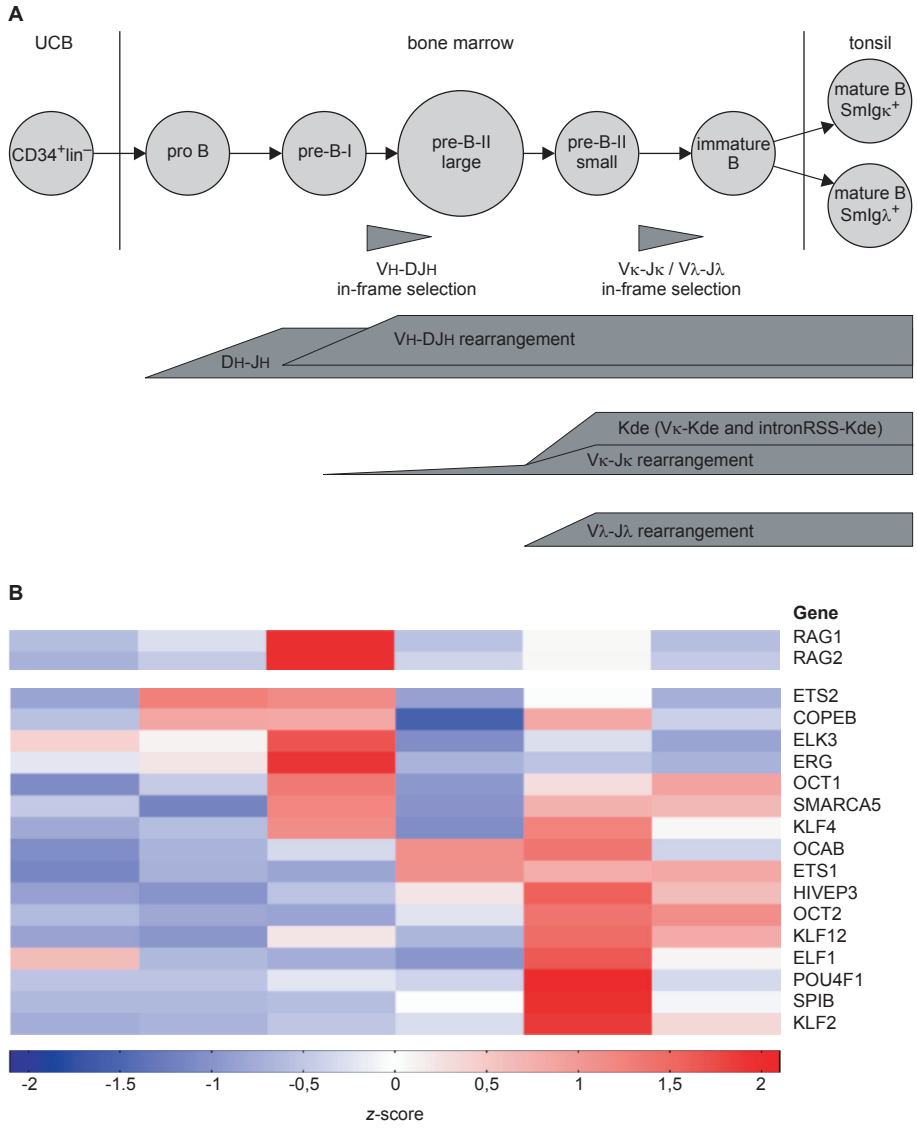


Figure 8. RQ-PCR analysis confirms the correlation of 16 newly identified transcription factors with the initiation of Ig gene rearrangements in precursor-B-cell differentiation.
A. Hypothetical scheme of precursor-B-cell differentiation, summarizing the quantitative Ig gene rearrangement data. The horizontal bars represent the rearrangement of the Ig loci. The arrows show the timing of in-frame selection of VH–DJH and Vκ–Jκ/Vλ–Jλ after the pre-B-I and pre-B-II small cell 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 undergo Ig gene rearrangements. Slight differences were seen for *ELF1* and *OCAB* when compared to the cluster patterns, because the cluster patterns were the averages of multiple genes expression patterns.

before the stage of large pre-B-II cells. Apparently, STAT5 might be involved in initiation of VH–DJH rearrangements in man as well. However, STAT5 is probably not activated by IL7Rα, but by an alternative signaling pathway.

In addition to the well-described genes, we identified a number of genes of which the expression pattern 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 *TCRB*, thereby opening the *IGH* locus. Furthermore, the expression of three Ets-family protein 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 were found to initiate *IGK/IGL* gene rearrangements. This up-regulation of the *RAG* transcripts was accompanied by an up-regulation of *IRF4*, which has been described to be involved in initiation of Ig light 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 non-lymphoid 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 it might be involved in both *IGH* and *IGK/IGL* 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 κB 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 recombined. Although it was found to be dispensable for precursor-B-cell differentiation,[52] as yet, no studies have addressed a function in gene segment selection in *IGK/IGL* gene rearrangements.

With the currently available markers, it was not possible to separate the initiation of Vκ–Jκ, Kde and Vλ–Jλ rearrangements, since 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. Based on 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 five subsets enabled us to

identify 16 novel candidate genes for initiation and regulation of *IGH* and *IGK/IGL* 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 antibody deficiencies.

ACKNOWLEDGEMENTS

The authors acknowledge Mrs. W.M. Comans-Bitter for assistance with preparing the figures.

DISCLOSURES

This work was supported by grant 349 from the foundation "Sophia Kinderziekenhuis Fonds" (MCvZ), and Veni grant 916.56.107 from ZonMw (MvdB).

The authors have no financial conflict of interest.

REFERENCES

1. Ghia P, ten Boekel E, Rolink AG, Melchers F. B-cell development: a comparison between mouse and man. *Immunol Today* 1998;**19**:480-5.
2. Oettinger MA, Schatz DG, Gorka C, Baltimore D. RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. *Science* 1990;**248**:1517-23.
3. McBlane JF, van Gent DC, Ramsden DA, Romeo C, Cuomo CA, Gellert M, Oettinger MA. Cleavage at a V(D)J recombination signal requires only RAG1 and RAG2 proteins and occurs in two steps. *Cell* 1995;**83**:387-95.
4. van Gent DC, Hiom K, Paull TT, Gellert M. Stimulation of V(D)J cleavage by high mobility group proteins. *Embo J* 1997;**16**:2665-70.
5. Grawunder U, Harfst E. How to make ends meet in V(D)J recombination. *Curr Opin Immunol* 2001;**13**: 186-94.
6. Ghia P, ten Boekel E, Sanz E, de la Hera A, Rolink A, Melchers F. Ordering of human bone marrow B lymphocyte precursors by single-cell polymerase chain reaction analyses of the rearrangement status of the immunoglobulin H and L chain gene loci. *J Exp Med* 1996;**184**:2217-29.
7. Noordzij JG, de Bruin-Versteeg S, Comans-Bitter WM, Hartwig NG, Hendriks RW, de Groot R, van Dongen JJM. Composition of precursor B-cell compartment in bone marrow from patients with X-linked agammaglobulinemia compared with healthy children. *Pediatr Res* 2002;**51**:159-68.
8. Schlissel M, Voronova A, Baltimore D. Helix-loop-helix transcription factor E47 activates germ-line immunoglobulin heavy-chain gene transcription and rearrangement in a pre-T-cell line. *Genes Dev* 1991;**5**: 1367-76.
9. Choi JK, Shen CP, Radomska HS, Eckhardt LA, Kadesch T. E47 activates the Ig-heavy chain and TdT loci in non-B cells. *Embo J* 1996;**15**:5014-21.
10. Romanow WJ, Langerak AW, Goebel P, Wolvers-Tettero IL, van Dongen JJM, Feeney AJ, Murre C. E2A and EBF act in synergy with the V(D)J recombinase to generate a diverse immunoglobulin repertoire in nonlymphoid cells. *Mol Cell* 2000;**5**:343-53.
11. Fuxa M, Skok J, Souabni A, Salvagiotto G, Roldan E, Busslinger M. Pax5 induces V-to-DJ rearrangements and locus contraction of the immunoglobulin heavy-chain gene. *Genes Dev* 2004;**18**:411-22.
12. Sato H, Saito-Ohara F, Inazawa J, Kudo A. Pax-5 is essential for kappa sterile transcription during Ig kappa chain gene rearrangement. *J Immunol* 2004;**172**:4858-65.

13. Hsu LY, Lauring J, Liang HE, Greenbaum S, Cado D, Zhuang Y, Schlissel MS. A conserved transcriptional enhancer regulates RAG gene expression in developing B cells. *Immunity* 2003;**19**:105-17.
14. Sigvardsson M, O'Riordan M, Grosschedl R. EBF and E47 collaborate to induce expression of the endogenous immunoglobulin surrogate light chain genes. *Immunity* 1997;**7**:25-36.
15. Akerblad P, Rosberg M, Leanderson T, Sigvardsson M. The B29 (immunoglobulin beta-chain) gene is a genetic target for early B-cell factor. *Mol Cell Biol* 1999;**19**:392-401.
16. Sigvardsson M, Clark DR, Fitzsimmons D, Doyle M, Akerblad P, Breslin T, Bilke S, Li R, Yeamans C, Zhang G, Hagman J. Early B-cell factor, E2A, and Pax-5 cooperate to activate the early B cell-specific mb-1 promoter. *Mol Cell Biol* 2002;**22**:8539-51.
17. Scott EW, Simon MC, Anastasi J, Singh H. Requirement of transcription factor PU.1 in the development of multiple hematopoietic lineages. *Science* 1994;**265**:1573-7.
18. McKercher SR, Torbett BE, Anderson KL, Henkel GW, Vestal DJ, Baribault H, Klemsz M, Feeney AJ, Wu GE, Paige CJ, Maki RA. Targeted disruption of the PU.1 gene results in multiple hematopoietic abnormalities. *Embo J* 1996;**15**:5647-58.
19. Pongubala JM, Nagulapalli S, Klemsz MJ, McKercher SR, Maki RA, Atchison ML. PU.1 recruits a second nuclear factor to a site important for immunoglobulin kappa 3' enhancer activity. *Mol Cell Biol* 1992;**12**: 368-78.
20. Lu R, Medina KL, Lancki DW, Singh H. IRF-4,8 orchestrate the pre-B-to-B transition in lymphocyte development. *Genes Dev* 2003;**17**:1703-8.
21. Su IH, Basavaraj A, Krutchinsky AN, Hobert O, Ullrich A, Chait BT, Tarakhovsky A. Ezh2 controls B cell development through histone H3 methylation and Igh rearrangement. *Nat Immunol* 2003;**4**:124-31.
22. Morshead KB, Ciccone DN, Taverna SD, Allis CD, Oettinger MA. Antigen receptor loci poised for V(D)J rearrangement are broadly associated with BRG1 and flanked by peaks of histone H3 dimethylated at lysine 4. *Proc Natl Acad Sci U S A* 2003;**100**:11577-82.
23. Busslinger M. Transcriptional control of early B cell development. *Annu Rev Immunol* 2004;**22**:55-79.
24. Hoffmann R, Seidl T, Neeb M, Rolink A, Melchers F. Changes in gene expression profiles in developing B cells of murine bone marrow. *Genome Res* 2002;**12**:98-111.
25. Hoffmann R, Melchers F. A genomic view of lymphocyte development. *Curr Opin Immunol* 2003;**15**: 239-45.
26. Beishuizen A, de Bruijn MA, Pongers-Willemse MJ, Verhoeven MA, van Wering ER, Hahlen K, Breit TM, de Bruin-Versteeg S, Hooijkaas H, van Dongen JJM. Heterogeneity in junctional regions of immunoglobulin kappa deleting element rearrangements in B cell leukemias: a new molecular target for detection of minimal residual disease. *Leukemia* 1997;**11**:2200-7.
27. Pongers-Willemse MJ, Seriu T, Stolz F, d'Aniello E, Gameiro P, Pisa P, Gonzalez M, Bartram CR, Panzer-Grumayer ER, Biondi A, San Miguel JF, van Dongen JJM. Primers and protocols for standardized detection of minimal residual disease in acute lymphoblastic leukemia using immunoglobulin and T cell receptor gene rearrangements and TAL1 deletions as PCR targets: report of the BIOMED-1 CONCERTED ACTION: investigation of minimal residual disease in acute leukemia. *Leukemia* 1999;**13**:110-8.
28. Verhagen OJ, Willemse MJ, Breunis WB, Wijkhuijs AJ, Jacobs DC, Joosten SA, van Wering ER, van Dongen JJM, van der Schoot CE. Application of germline IGH probes in real-time quantitative PCR for the detection of minimal residual disease in acute lymphoblastic leukemia. *Leukemia* 2000;**14**:1426-35.
29. van der Velden VH, Willemse MJ, van der Schoot CE, Hahlen K, van Wering ER, van Dongen JJM. Immunoglobulin kappa deleting element rearrangements in precursor-B acute lymphoblastic leukemia are stable targets for detection of minimal residual disease by real-time quantitative PCR. *Leukemia* 2002;**16**: 928-36.
30. van Dongen JJM, Langerak AW, Bruggemann M, Evans PA, Hummel M, Lavender FL, Delabesse E, Davi F, Schuuring E, Garcia-Sanz R, Van Krieken JH, Droese J, Gonzalez D, Bastard C, White HE, Spaargaren M, Gonzalez M, Parreira A, Smith JL, Morgan GJ, Kneba M, Macintyre EA. Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: Report of the BIOMED-2 Concerted Action BMH4-CT98-3936. *Leukemia* 2003;**17**: 2257-317.

31. van der Burg M, Tumkaya T, Boerma M, de Bruin-Versteeg S, Langerak AW, van Dongen JJM. Ordered recombination of immunoglobulin light chain genes occurs at the IGK locus but seems less strict at the IGL locus. *Blood* 2001;**97**:1001-8.
32. Moppett J, van der Velden VH, Wijkhuijs AJ, Hancock J, van Dongen JJM, Goulden N. Inhibition affecting RQ-PCR-based assessment of minimal residual disease in acute lymphoblastic leukemia: reversal by addition of bovine serum albumin. *Leukemia* 2003;**17**:268-70.
33. van der Velden VH, Hochhaus A, Cazzaniga G, Szczepanski T, Gabert J, van Dongen JJM. Detection of minimal residual disease in hematologic malignancies by real-time quantitative PCR: principles, approaches, and laboratory aspects. *Leukemia* 2003;**17**:1013-34.
34. Pongers-Willemse MJ, Verhagen OJ, Tibbe GJ, Wijkhuijs AJ, de Haas V, Roovers E, van der Schoot CE, van Dongen JJM. Real-time quantitative PCR for the detection of minimal residual disease in acute lymphoblastic leukemia using junctional region specific TaqMan probes. *Leukemia* 1998;**12**:2006-14.
35. Staal FJ, van der Burg M, Wessels LF, Barendregt BH, Baert MR, van den Burg CM, van Huffel C, Langerak AW, van der Velden VH, Reinders MJ, van Dongen JJM. DNA microarrays for comparison of gene expression profiles between diagnosis and relapse in precursor-B acute lymphoblastic leukemia: choice of technique and purification influence the identification of potential diagnostic markers. *Leukemia* 2003;**17**: 1324-32.
36. Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, Aach J, Ansorge W, Ball CA, Causton HC, Gaasterland T, Glenisson P, Holstege FC, Kim IF, Markowitz V, Matese JC, Parkinson H, Robinson A, Sarkans U, Schulze-Kremer S, Stewart J, Taylor R, Vilo J, Vingron M. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat Genet* 2001;**29**:365-71.
37. Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 2003;**19**:185-93.
38. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 2003;**4**: 249-64.
39. Ge U, Dudoit S, Speed TP. Resampling-based multiple testing for microarray analysis. *Dept of Statistics, University of California, Berkeley, CA, USA* 2003.
40. Beillard E, Pallisgaard N, van der Velden VH, Bi W, Dee R, van der Schoot E, Delabesse E, Macintyre E, Gottardi E, Saglio G, Watzinger F, Lion T, van Dongen JJM, Hokland P, Gabert J. Evaluation of candidate control genes for diagnosis and residual disease detection in leukemic patients using 'real-time' quantitative reverse-transcriptase polymerase chain reaction (RQ-PCR) - a Europe against cancer program. *Leukemia* 2003;**17**: 2474-86.
41. Siminovitch KA, Bakhshi A, Goldman P, Korsmeyer SJ. A uniform deleting element mediates the loss of kappa genes in human B cells. *Nature* 1985;**316**:260-2.
42. Tumkaya T, van der Burg M, Garcia Sanz R, Gonzalez Diaz M, Langerak AW, San Miguel JF, van Dongen JJM. Immunoglobulin lambda isotype gene rearrangements in B cell malignancies. *Leukemia* 2001;**15**: 121-7.
43. Lu LS, Tung J, Baumgarth N, Herman O, Gleimer M, Herzenberg LA. Identification of a germ-line pro-B cell subset that distinguishes the fetal/neonatal from the adult B cell development pathway. *Proc Natl Acad Sci U S A* 2002;**99**:3007-12.
44. Nutt SL, Urbanek P, Rolink A, Busslinger M. Essential functions of Pax5 (BSAP) in pro-B cell development: difference between fetal and adult B lymphopoiesis and reduced V-to-DJ recombination at the IgH locus. *Genes Dev* 1997;**11**:476-91.
45. Engel H, Rolink A, Weiss S. B cells are programmed to activate kappa and lambda for rearrangement at consecutive developmental stages. *Eur J Immunol* 1999;**29**:2167-76.
46. Goetz CA, Harmon IR, O'Neil JJ, Burchill MA, Farrar MA. STAT5 activation underlies IL7 receptor-dependent B cell development. *J Immunol* 2004;**172**:4770-8.
47. Yang XO, Doty RT, Hicks JS, Willerford DM. Regulation of T-cell receptor D beta 1 promoter by KLF5 through reiterated GC-rich motifs. *Blood* 2003;**101**:4492-9.
48. Rivera RR, Stuiver MH, Steenbergen R, Murre C. Ets proteins: new factors that regulate immunoglobulin heavy-chain gene expression. *Mol Cell Biol* 1993;**13**:7163-9.

49. Su GH, Chen HM, Muthusamy N, Garrett-Sinha LA, Baunoch D, Tenen DG, Simon MC. Defective B cell receptor-mediated responses in mice lacking the Ets protein, Spi-B. *Embo J* 1997;**16**:7118-29.
50. Schweitzer BL, DeKoter RP. Analysis of gene expression and Ig transcription in PU.1/Spi-B-deficient progenitor B cell lines. *J Immunol* 2004;**172**:144-54.
51. Allen CE, Mak CH, Wu LC. The kappa B transcriptional enhancer motif and signal sequences of V(D)J recombination are targets for the zinc finger protein HIVEP3/KRC: a site selection amplification binding study. *BMC Immunol* 2002;**3**:10.
52. Allen CE, Muthusamy N, Weisbrode SE, Hong JW, Wu LC. Developmental anomalies and neoplasia in animals and cells deficient in the large zinc finger protein KRC. *Genes Chromosomes Cancer* 2002;**35**: 287-98.

III

3D-Architecture of the Immunoglobulin Heavy Chain Locus

Suchit Jhunjhunwala,[1] Menno C. van Zelm,[1,2] Mandy Peak,[1] Roy Riblet,[3] Jacques J.M. van Dongen,[2] Frank G. Grosveld,[4] Tobias A. Knoch,[4,5] and Cornelis Murre[1]

[1]Division of Biological Sciences, 0377, University of California, San Diego, La Jolla, CA 92093, USA; [2]Erasmus MC, Department of Immunology, Dr. Molewaterplein 50, 3015 GE Rotterdam, the Netherlands; [3]Torrey Pines Institute for Molecular Studies, San Diego, CA 92121, USA; [4]Erasmus MC, Departments of Biophysical Genomics, Cell Biology and Genetics, Dr. Molewaterplein 50, 3015 GE Rotterdam, the Netherlands; [5]Heidelberg University, Kirchhoff Institute for Physics, Department of Biophysical Genomics, Im Neuenheimfer Feld 280, 69120 Heidelberg, Germany

Submitted

ABSTRACT

The immunoglobulin heavy chain (*IGH*) locus is organized in distinct regions that contain multiple variable (V_H), diversity (D_H), joining (J_H) and constant (C_H) coding elements. Spatial distance distributions were determined between 12 genomic markers that span the entire *IGH* locus. In pro-B cells, the *IGH* locus topology brought V_H segments into relative close physical proximity is on the D_H-J_H elements. Although separated by large genomic distances, the pro-B proximal and distal V_H regions were positioned at similar spatial distances from the D_HJ_H cluster. Statistical analysis showed that the pro-B *IGH* locus configuration allowed V_H regions to interact with the D_HJ_H elements with relatively high probability. Comparison of the distance distribution to simulated data described for different configurations of chromatin structure, indicated that the experimental data are most consistent with an *IGH* locus organized in rosette-like structure(s) containing bundles of loops, connected by spacers. These data indicate that the *IGH* locus displays a distinct topology, to allow long-range genomic interactions to occur with relatively high frequency.

KEY WORDS

Immunoglobulin heavy chain locus, immunoglobulin gene rearrangement, chromatin structure, spatial nuclear architecture, long-range interaction.

INTRODUCTION

It is well established that higher order chromatin organization plays a pivotal role in genome function.[1-3] For more than a century, the organization of chromosomes and its functional implications in eukaryotes has been extensively studied using light microscopy.[4, 5] Electron micrographs of chromosome spreads have suggested the presence of loops, with sizes of ~90 kb, that interact with the nuclear matrix and aggregate during mitosis into rosettes containing ~18 loops, resulting in ~100 rosettes per average chromosome.[6, 7] Similar rosette-like structures have been detected in interphase cells.[8] Electron tomography has suggested another structure, a "chromonema" fiber with a diameter of 60-130 nm and agglomerations along the chromatin fiber.[9]

Initial studies have suggested that the chromatin fiber behaves like a random walk that spans the entire nucleus.[10] However, the confinement to territories, chromosome arms and bands, clearly indicated the presence of spatial constraints. More recent observations demonstrated that the spatial distance depends on the genomic distance according to a power law with exponents of 0.5 below and 0.32 above a genomic separation of 4 Mb.[11-14] The constraints and the scaling behavior suggested a two level folding of the chromatin fiber, resulting in the Random-Walk/Giant-Loop (RW/GL) model.[12] In the RW/GL model, the 30 nm fiber forms 2 to 5 Mb loops that are attached to a polymer backbone. The backbone and the chromatin fiber within the loops have been proposed to follow random walk dynamics. However, distance measurements between genetic markers with genomic separations of less than 4 Mb are incompatible with the RW/GL model, whereas they are consistent with another model of chromatin structure, named the Multi-Loop-Subcompartment (MLS) model.[3, 11] The MLS model suggests that the 30 nm fiber is folded into bundles of loops, connected by spacers of variable sizes.

Recently, computer models have been developed to evaluate and test experimental results, designs and hypotheses about the three-dimensional genome organization.[3, 11, 15] Beyond supporting the chromatin organization into chromosome territory, arm and band domains, these simulations may reveal how the local, global and dynamic characteristics of cell nuclei are inter-connected.[3, 15]

How genes are regulated over long distances has been a topic of intensive study. In prokaryotes, transcriptional enhancers act through looping or tracking of the intervening DNA.[16, 17] In eukaryotic cells, chromatin compaction and looping, the presence of DNase I hypersensitive sites at regulatory elements, including transcriptional enhancers, insulators and locus control regions, influence gene expression over large genomic distances.[18] The globin locus control region was shown to act long-range by associating with actively transcribing β-globin genes.[19, 20] Other loci also have been shown to bring distant enhancer elements into proximity of promoter regions by looping, including the Th2 cytokine locus and the interferon gamma gene.[21, 22] How these genes and regulatory elements are organized in 3D-space remains to be determined.

The murine immunoglobulin heavy chain locus (*IGH*), is organized into distinct regions that contain the variable (V_H), diversity (D_H), joining (J_H), and constant (C_H) coding elements. Eight constant regions encode the various isotypes: Cμ, Cδ, Cγ1, Cγ2a, Cγ2b, Cγ3, Cα and Cε. Twelve D_H and four J_H segments are positioned immediately upstream of the Cμ constant

region. There are at least 134 VH gene segments that belong to fifteen partially interspersed families.[23] The VH, DH and JH coding elements are flanked by recombination signal sequences (RSSs) that act as recognition elements for the V(D)J-recombinase: RAG1 and RAG2.[24, 25]

The *IGH* and Ig light chain loci undergo ordered rearrangement during B-lineage development.[26] In pro-B cells, DH to JH rearrangement precedes VH to DHJH joining. Once an in-frame VDJH joint has been generated, a pre-B-cell receptor is formed, suppressing RAG1/2 activity, and preventing continued *IGH* gene rearrangement.[27, 28] The temporal and lineage specificity of V(D)J recombination is controlled at distinct levels, including DNA methylation, chromatin remodeling, histone acetylation, germline transcription and transcription elongation.[24] Fluorescence *in situ* hybridization (FISH) analysis has demonstrated that the *IGH* alleles are located in close proximity to the nuclear membrane in T-lineage cells.[29] In B-lineage cells, the *IGH* locus is located away from the nuclear membrane and the VH regions undergo contraction and looping.[29-32] However, it is unclear how a large region spanning 3 Mb allows efficient joining of VH to DH gene segments. It seems plausible that within such a large genomic locus an underlying order facilitates this interaction.[33]

To explore the 3D-architecture of the *IGH* locus, we combined 3D structure-preserving FISH with multiple probes that span the entire *IGH* locus and computer modeling to determine and analyze spatial distances between small genetic markers as a function of their genomic separation. Collectively, the data are most compatible with a topology in which the *IGH* chromatin polymer is organized in rosette-like structures, containing multiple loops, which allow long-range genomic interactions to occur with relatively high frequency.

RESULTS

High Resolution Spatial Distance Measurements between Genomic Markers in B-lineage Cells

We previously used 100-150 kb BAC probes to simultaneously visualize three distinct regions of the *IGH* locus.[32] However, the resolution obtained using this approach was limited, due to the large sizes of the probes. Here we use relatively small DNA probes (10 kb) to visualize the *IGH* locus. To label 10 kb DNA segments, a direct probe labeling procedure was employed (1-4 labeled nucleotides per 100 bp; see Materials and Methods). E2A-deficient pre-pro-B cells and RAG2-deficient pro-B cells were fixed, permeabilized and hybridized with one 235 kb BAC and two 10 kb probes to visualize the *IGH* locus (Figure 1A). The fluorescent signals emitted by the 10 kb probes were clearly detectable, but the intensity varied between the different probes (Figure 1A). The *IGH* locus in E2A-deficient pre-pro-B cells is located closer to the nuclear membrane as compared to RAG-deficient pro-B cells, confirming previous observations (Figure 1B and C).[29]

To determine the smallest measurable distance, we calculated the full width at half maximum (FWHM) of the point spread function of the optical system used in the experiments described below (400-1,800 nm range; Appendices 4 and 5). Although the resolution obtained was rather limited, measurements of considerably smaller distances were achieved using different absorption/emission spectra and correction for chromatic shifts. The average distance measured for the different combinations of the fluorochromes ranged between 35 and 47 nm,

with standard deviations of 18 and 24 nm, respectively (Appendix 6). Thus, this approach allowed us to measure spatial distances with a 3D-resolution better than 100 nm.

3D-Architecture of the *IGH* Locus

To gain insight into the topology of the *IGH* locus, we measured the spatial distances between an anchor, BAC probe RP23-201H14, and eleven 10 kb probes that span the entire locus (Figure 2A). Both pre-pro-B and pro-B cells were examined. We noted that the average distances and standard deviations were substantially higher than the smallest measurable distance (Figure 2B; Appendix 6). Consequently, this approach allowed us to accurately determine spatial distances in interphase nuclei. The distance distributions are not well described by a Gaussian distribution as expected for a random walk unaffected by tethering and confinement, but rather can be fitted as Rayleigh distributions (Figure 2B and S.J., unpublished observations). The average spatial distances separating the anchor RP23-201H14 and the markers located within the CH region cluster (h2 and h3) increased with increasing genomic separation (Figure 2B and C). The pattern leveled off beyond the intronic enhancer in pro-B cells indicating confinement within the pre-pro-B-cell nucleus (Figure 2C). Local inversions in the pattern were observed in the region containing the *IGH* intronic enhancer and the region separating the proximal and distal VH regions (Figure 2C). When compared to the spatial organization in pre-pro-B cells, the *IGH* locus configuration in pro-B cells showed substantial spatial contraction beyond the *IGH* intronic enhancer (Figure 2C; Table 1). The differences in compaction values between pre-pro-B and pro-B cells, were particularly striking in the distal VH cluster (1284 versus 2059 between RP23-201H14 and h11; Table 1). Taken together, the spatial organization and compaction values indicate that the 3D-architecture of the *IGH* locus differs between pre-pro-B and pro-B cells.

Long-range Genomic Separation but Similar Spatial Distances

The data described above showed a high degree of compaction between the distal VH and CH regions (Figure 2C). Because during B-cell differentiation VH gene segments rearrange to DHJH gene segments, we measured the spatial distances separating the VH regions from the DHJH cluster (Figure 3A). The average spatial distances between the DHJH elements and the proximal VH regions were similar for pre-pro-B and pro-B cells (Figure 3B-C and Table 1). In contrast, the spatial distances separating the distal VH and DHJH regions were substantially larger in pre-pro-B cells as compared to pro-B cells (Figure 3C). The degree of compaction was particularly high beyond the proximal VH region in pro-B cells (Table 1). Notably, in pro-B cells the spatial distances separating the proximal and distal VH regions from the DHJH cluster were similar (Figure 3C). On the other hand, the spatial distances separating the DHJH elements from the CH regions were larger in pro-B cells as compared to pre-pro-B cells (Figure 3C). These data suggest that upon commitment to the B-cell lineage, the *IGH* locus is remodeled to allow the proximal and distal VH regions equal access to the DHJH elements.

Long-range Interaction Probability and VH to DHJH Joining

The V(D)J recombination process requires that the VH and DHJH segments are mobile, thus allowing them to interact with each other with detectable frequency. What are the probabilities of such encounters in pre-pro-B versus pro-B cells? The frequency by which VH regions are

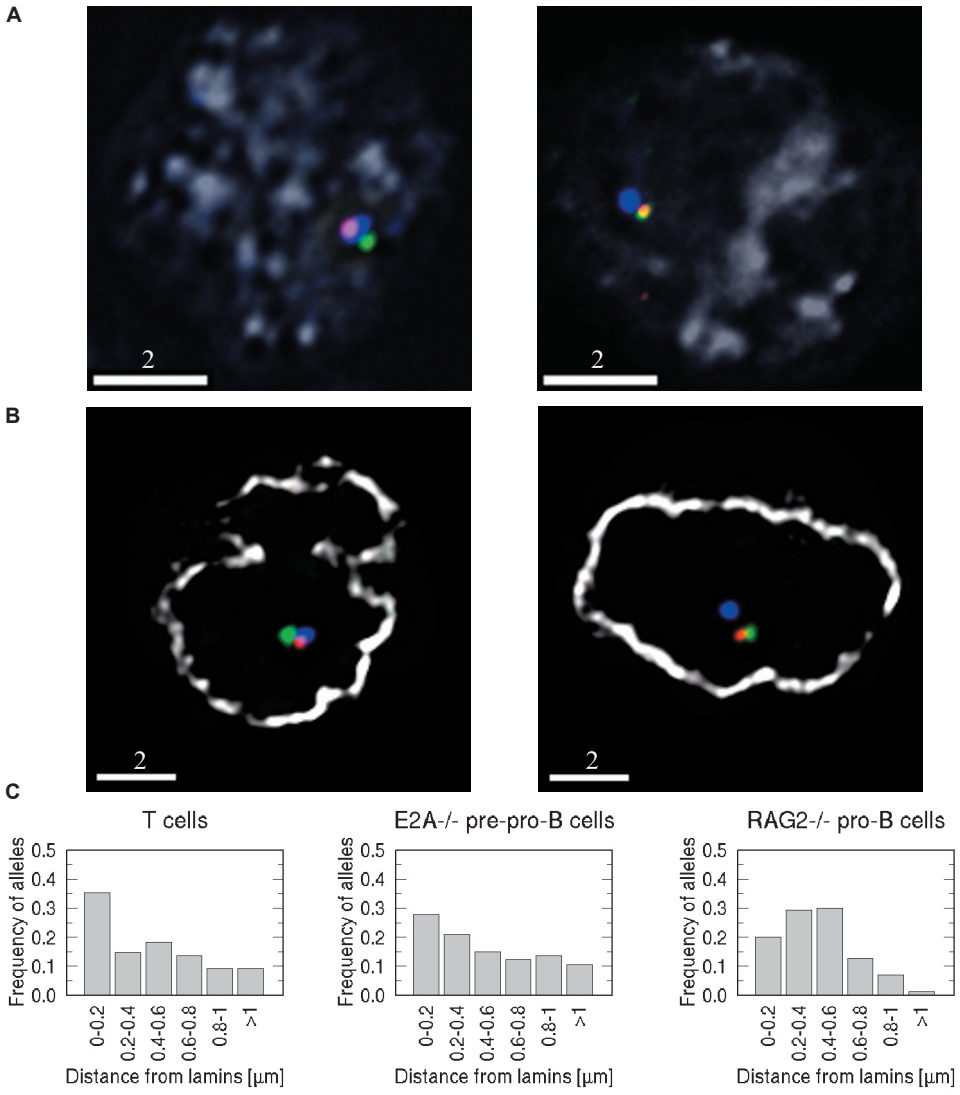


Figure 1. 3D structure-preserving fluorescence *in situ* hybridization and spectral precision distance fluorescence microscopy.

A. 3D FISH in nuclei derived from RAG2-deficient pro-B cells using 10 kbp probes. Two 10 kb probes (shown in red and green) were labeled with aminoallyl-dUTP using nick-translation followed by incubation with succinimidyl-ester conjugates of Alexa fluorochromes. The BAC probe RP23-201H14 (shown in blue) was labeled as described previously by Sayegh et al.[32] Digitally magnified images of *IGH* stainings are shown. Nuclei are visualized by DAPI staining. White bars represent 2µM **B.** Localization of *IGH* locus in E2A-deficient pre-pro B and RAG2-deficient pro-B cells. The distances separating each of the three BAC probes CT7-526A21(green), CT7-34H6 (red), and RP23-24I12 (blue) from the nuclear membrane were determined. The nuclear lamina were stained with anti-lamin antibodies and are indicated in gray. **C.** Distances separating the *IGH* constant region (BAC-CT7-526A21) from the nuclear membrane in T cells, E2A-deficient pre-pro-B cells and RAG2-deficient pro-B cells.

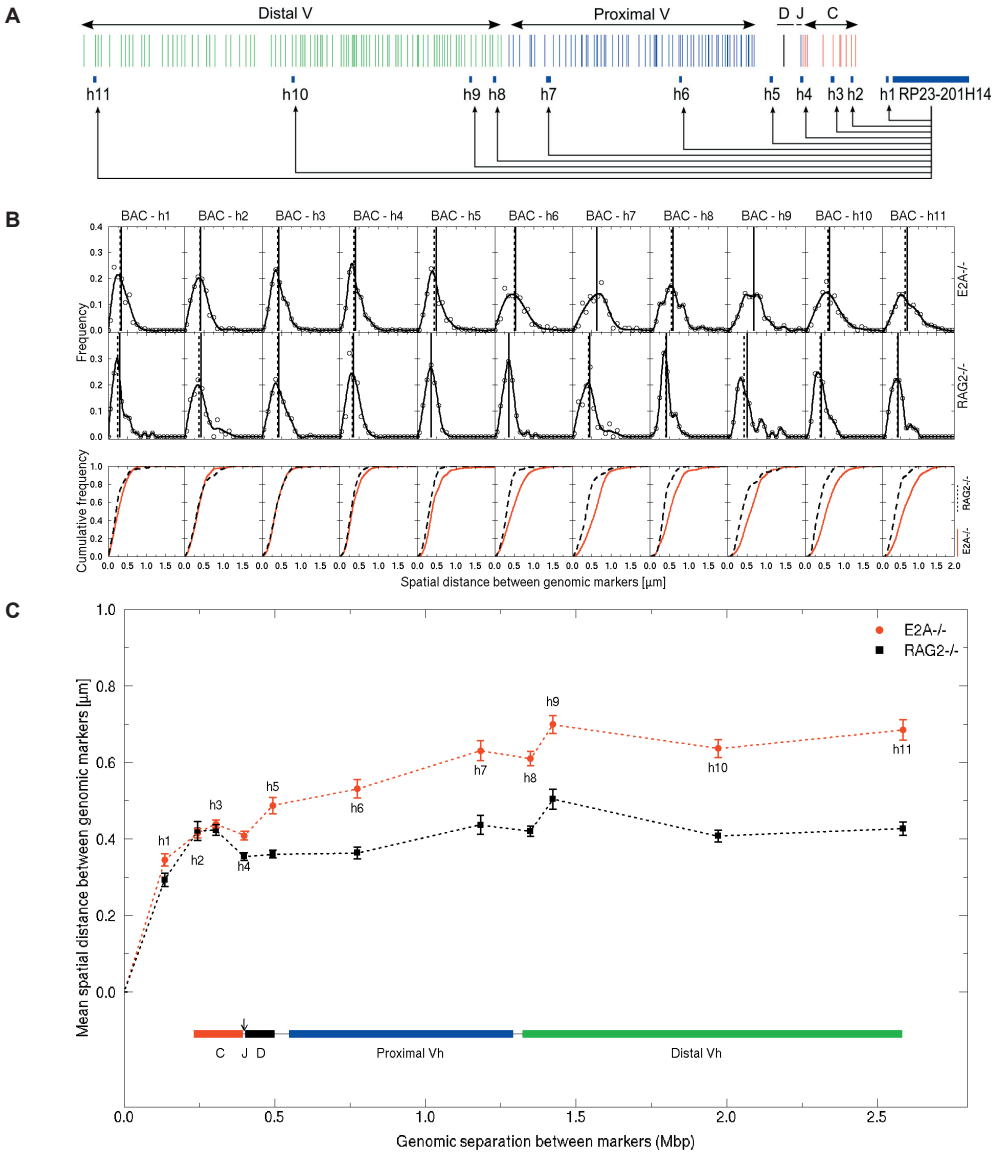


Figure 2. Spatial distances within the *IGH* locus as a function of genomic separation in pre-pro-B and pro-B cells.

A. Strategy to generate a physical map of the 2.77 Mbp *IGH* locus using anchor probe RP23-201H14. The 3D-distances separating the probes from the anchor were measured. **B.** Frequency plots showing the distribution of physical distances for each 10 kbp probe as measured from the anchor (RP23-201H14) in E2A-deficient pre-pro-B cells and RAG2-deficient pro-B cells. Distances were determined between the centers of mass for each of the two probes. X-axis shows spatial distances in µm. Y-axis indicates frequency monitored for each of the spatial distances. Distances were corrected for chromatic shifts. **C.** Average spatial distances were plotted as a function of genomic separation. Mean distances are indicated ± standard error. The arrow indicates the position of the intronic enhancer. The dotted lines only indicate connectivity, but the exact path between two data points is not known.

Table 1. Genomic and spatial distances and compaction values between genetic markers spanning the *IGH* locus

Marker pair	Genomic separation (in Mbp)	E2A-deficient pre-pro-B cells				RAG2-deficient pro-B cells			
		Detection	Nuclei [n]	<x> ± SD ± SE (in nm)	compaction	Detection	Nuclei [n]	<x> ± SD ± SE (in nm)	compaction
BAC - h1	0.14	A647 / A488	131	331.6 ± 181.4 ± 15.8	138	A647/ A488	139	293.2 ± 199.2 ± 16.9	156
BAC - h2	0.24	A647 / A594	262	412.4 ± 203.6 ± 12.6	201	A647 / A594	91	421.1 ± 236.5 ± 24.8	197
BAC - h3	0.3	A647 / A488	391	429.4 ± 196.9 ± 10.0	241	A647 / A488	185	423.4 ± 194.0 ± 14.3	244
BAC - h4	0.4	A647 / A594	300	408.1 ± 206.5 ± 11.9	332	A647 / A594	193	354.1 ± 147.1 ± 10.6	382
BAC - h5	0.49	A647 / A488	185	483.9 ± 286.1 ± 21.0	346	A647 / A488	192	360.2 ± 139.3 ± 10.1	465
BAC - h6	0.77	A647 / A594	135	528.1 ± 278.8 ± 24.0	498	A647 / A594	110	363.4 ± 158.9 ± 15.1	724
BAC - h7	1.18	A647 / A594	114	630.0 ± 278.2 ± 26.1	638	A647 / A594	82	436.8 ± 227.1 ± 25.1	920
BAC - h8	1.35	A647 / A488	227	593.8 ± 278.6 ± 18.5	772	A647 / A488	127	420.3 ± 154.8 ± 13.7	1091
BAC - h9	1.42	A647 / A594	154	670.4 ± 284.9 ± 23.0	722	A647 / A594	125	504.0 ± 293.5 ± 26.2	960
BAC - h10	1.97	A647 / A488	143	624.5 ± 275.0 ± 23.0	1073	A647 / A488	119	407.7 ± 164.9 ± 15.1	1644
BAC - h11	2.59	A647 / A594	143	684.4 ± 317.4 ± 26.5	1284	A647 / A594	115	426.8 ± 185.5 ± 17.3	2059
h4 - h1	0.09	A594 / A488	94	350.1 ± 155.4 ± 16.0	256	A488 / A594	58	440.0 ± 251.4 ± 33.0	204
h4 - h2	0.16	A488 / A594	100	246.0 ± 150.8 ± 15.1	214	A488 / A594	59	357.8 ± 256.6 ± 33.4	147
h4 - h3	0.26	A594 / A488	188	258.8 ± 161.9 ± 11.8	124	A488 / A594	67	337.4 ± 241.5 ± 29.5	95
h4 - h5	0.1	A594 / A488	97	252.3 ± 155.9 ± 15.8	128	A594 / A488	38	255.5 ± 140.7 ± 22.8	126
h4 - h6	0.38	A488 / A594	47	436.0 ± 260.3 ± 38.0	293	A488 / A594	63	465.6 ± 310.7 ± 39.1	274
h4 - h7	0.78	A488 / A594	73	479.4 ± 233.0 ± 27.3	556	A488 / A594	55	407.5 ± 287.5 ± 38.8	654
h4 - h8	0.95	A594 / A488	59	483.5 ± 227.5 ± 29.6	668	A594 / A488	68	440.0 ± 309.6 ± 37.5	734
h4 - h9	1.03	A488 / A594	30	774.3 ± 362.4 ± 66.2	450	A488 / A594	42	524.3 ± 336.8 ± 52.0	665
h4 - h10	1.57	A488 / A594	70	607.2 ± 301.6 ± 36.1	881	A488 / A594	55	325.7 ± 145.7 ± 19.6	1642
h4 - h11	2.19	A488 / A594	41	641.9 ± 346.7 ± 54.1	1158	A488 / A594	74	353.5 ± 194.1 ± 22.6	2103

n refers to number of nuclei analyzed; <x> indicates average spatial distance; SD indicates standard deviation; SE refers to standard error;

The average DNA compaction based on the linear contour length of 340 nm for 1.0 kbp was calculated from: $\frac{\text{genomic lenght [kb]}}{\text{spatial distance [nm]}} \cdot 340\,\text{[nm/kb]}$

positioned in close proximity to DHJH elements, can be derived from the cumulative distance distributions (Figure 3B). The distance distributions for the proximal VH regions were similar for pre-pro-B cells and pro-B cells (Figure 3B). In contrast, the cumulative distance distributions for distal VH regions (h4-h10; h4-h11), differed substantially between pre-pro-B and pro-B cells (Figure 3B). Specifically, the probability that a distal VH region was located within relative close proximity (sub 100-200 nm range) to the DHJH elements was substantially lower in pre-pro-B cells as compared to pro-B cells (Figure 3B).

To study the probabilities that VH regions are located in relative close proximity (100-200 nm) to the DHJH cluster, the cumulative distance distributions separating the VH from the DHJH elements were plotted (Figure 4). In pre-pro-B cells, the probabilities of VH regions to interact with DHJH elements correlated well with increasing genomic separation (Figure 4A). In contrast, in pro-B cells the probability distribution was clustered at low spatial distances (sub 350 nm) for the majority of VH regions (Figure 4B). Collectively, these data show that in pro-B cells, the probabilities of VH regions to interact with the DHJH elements are similar, regardless of their genomic separation.

IGH Locus Configuration and the Random Walk Model

Previous observations indicated that the chromatin fiber in budding yeast (*Saccharomyces cerevisiae*) can be described as a worm-like chain.[34] Our analysis of the average spatial distances, the locus dynamics and interaction probabilities suggest that the *IGH* chromatin fiber configuration is incompatible with that of a free worm-like chain. To determine whether there is any condition under which the *IGH* 3D-architecture can be explained by random walk behavior, the experimental spatial distances were compared to the Kratky-Porod description of a worm-like chain.[35] In this approach, only the flexibility of the chromatin fiber, the persistence length L, and the density of the chromatin fiber c determine, for a specific genomic separation, the average theoretical spatial distance (see Material and Methods for details). No reasonable values could be found for either the pre-pro-B or the pro-B cells, using different values for both parameters (Figure 5A and B). A least square optimization analysis also showed values with $L = 0.85$ nm and $c = 5.9$ bp/nm, respectively, that are not compatible with the physical properties of the chromatin fiber (S.J., unpublished observations). These data indicate that unlike budding yeast chromatin, the spectrum of *IGH* conformations cannot be described in terms of a worm-like chain.

Chromatin Fiber Statistics in *Saccharomyces cerevisiae* and B-lineage Cells

In budding yeast, long-range genomic interactions have not been reported. In contrast, vertebrate organisms have the ability to modulate gene expression from regulatory elements that are separated by large genomic distances.[18] Why is it that DNA segments, separated by large genomic distances in budding yeast, are not able to encounter each other, whereas in multicellular organisms enhancers and locus control regions physically interact with promoter regions with detectable frequency?[19, 20] Our observations raise the possibility that it is the distinct chromatin topology in mammalian cells that allows genomic elements to interact long-range with relatively high probability. To address this possibility, we directly compared the theoretical probabilities by which DNA elements interact in budding yeast, modeled as a 3D random walk, to our experimental values obtained from B-lineage cells.

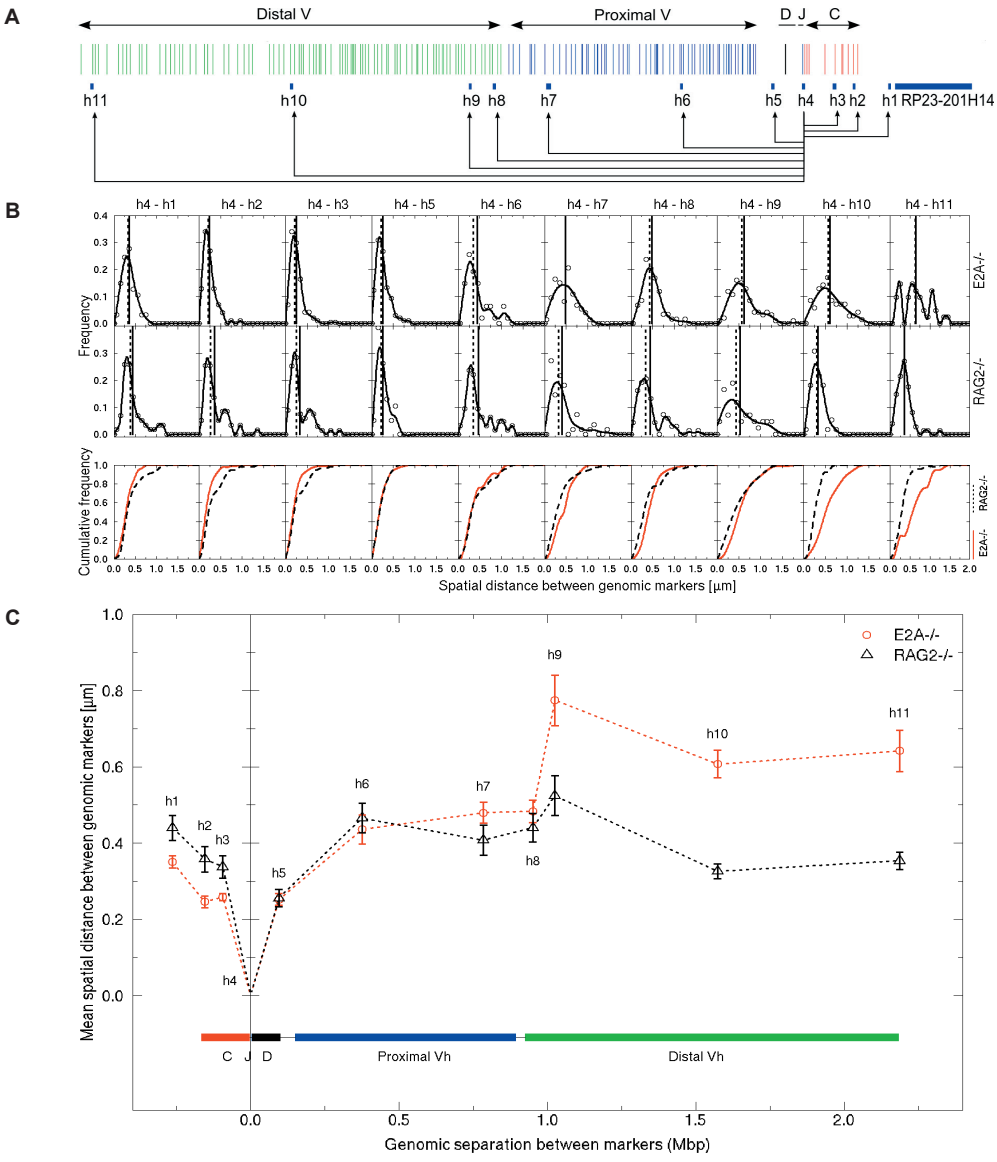


Figure 3. Altered *IGH* locus Tolopogy in Pro-B cells Brings Distal VH Regions and DHJH and Enhancer Elements in Close Proximity.

A. Strategy to generate a physical map of the *IGH* locus using a region containing the DHJH elements and the intronic enhancer as an anchor. **B.** Frequency plots showing the distribution of spatial distances for each genomic marker from the DHJH cluster in pre-pro-B cells and pro-B cells. X-axis shows spatial distances in µm. Y-axis indicates frequency monitored for each of the spatial distances. Distances were corrected for chromatic shifts. **C.** Average spatial distances were plotted as a function of genomic separation for each of the probes. Mean distances are indicated ± standard error. The dotted lines only indicate connectivity, but the exact path between two data points is not known.

Cumulative frequency distributions for yeast chromatin were determined for different persistence lengths (see Materials and Methods for details). For the shortest genomic separation (h4-h5) and using 50 nm as a persistence length value, the theoretical distance distribution for a random walk approaches that of pre-pro-B and pro-B cells (Figure 5C-F). However, for larger genomic separations, the cumulative frequency distributions, using the same range of values for persistence lengths, do not compare well with those obtained from the experimental data (Figure 5C-F). Thus, it is the *IGH* locus topology in pro-B cells that allows genomic and regulatory elements to encounter one another long-range with relatively high frequency.

Computer Modeling of Chromatin Structure and Comparison of Simulated to Experimental Data

To determine the 3D-architecture of the *IGH* locus in more detail, computer simulations were performed using Monte Carlo and Brownian Dynamics methods (Figure 6A and B).[3] The RW/GL and MLS models were used as starting configurations. The 30 nm fiber was modeled as an elastic polymer chain. The persistence length L_p and the density c of the chromatin fiber were based on a previous consensus comparison and integration of experimental data.[3] We note that variations in persistence length and chromatin density do not have a large influence on this analysis. Different loop and spacer sizes were used at ~5,000 bp resolution, which is >19 times smaller than the smallest genomic separation investigated in this study. The different architecture and morphology of chromatin structure that these models

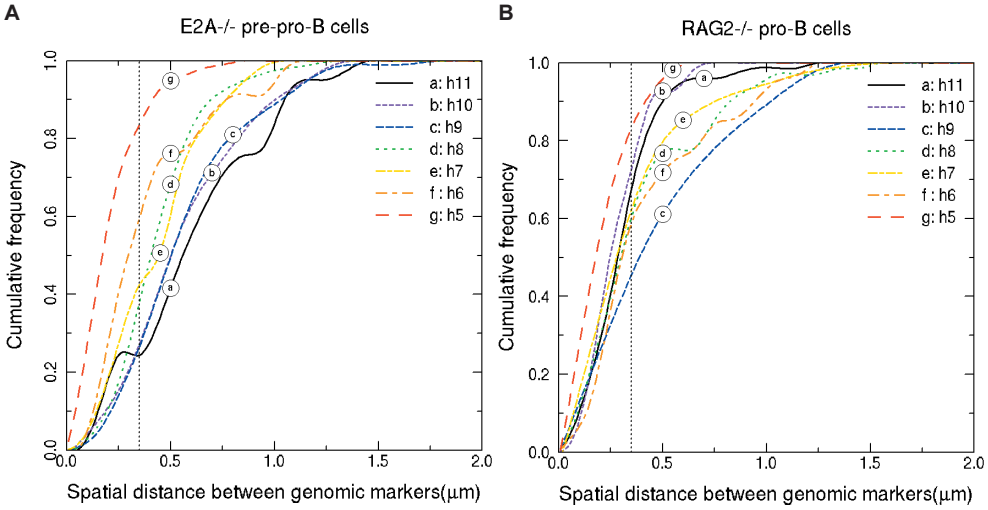


Figure 4. Probabilities for Proximal and Distal VH Regions to Interact with DHJH Elements are Similar in Pro-B Cells Regardless of Genomic Separation.

A, B. Probabilities of proximal and distal VH regions to interact with DHJH elements in pre-pro-B cells and pro-B cells. Cumulative frequencies were obtained by accumulating the frequency values corresponding to the spatial distances in intervals of 100 nm using the DHJH elements (probe h4) as an anchor. The dotted line indicating the spatial distance below which the cumulative frequencies cluster in pro-B cells.

show provide a qualitative idea about the outcome concerning the average spatial distances, the locus dynamics and interaction probabilities (Figure 5A).

There are now two challenges to be met by comparing the experimental with the simulated data:

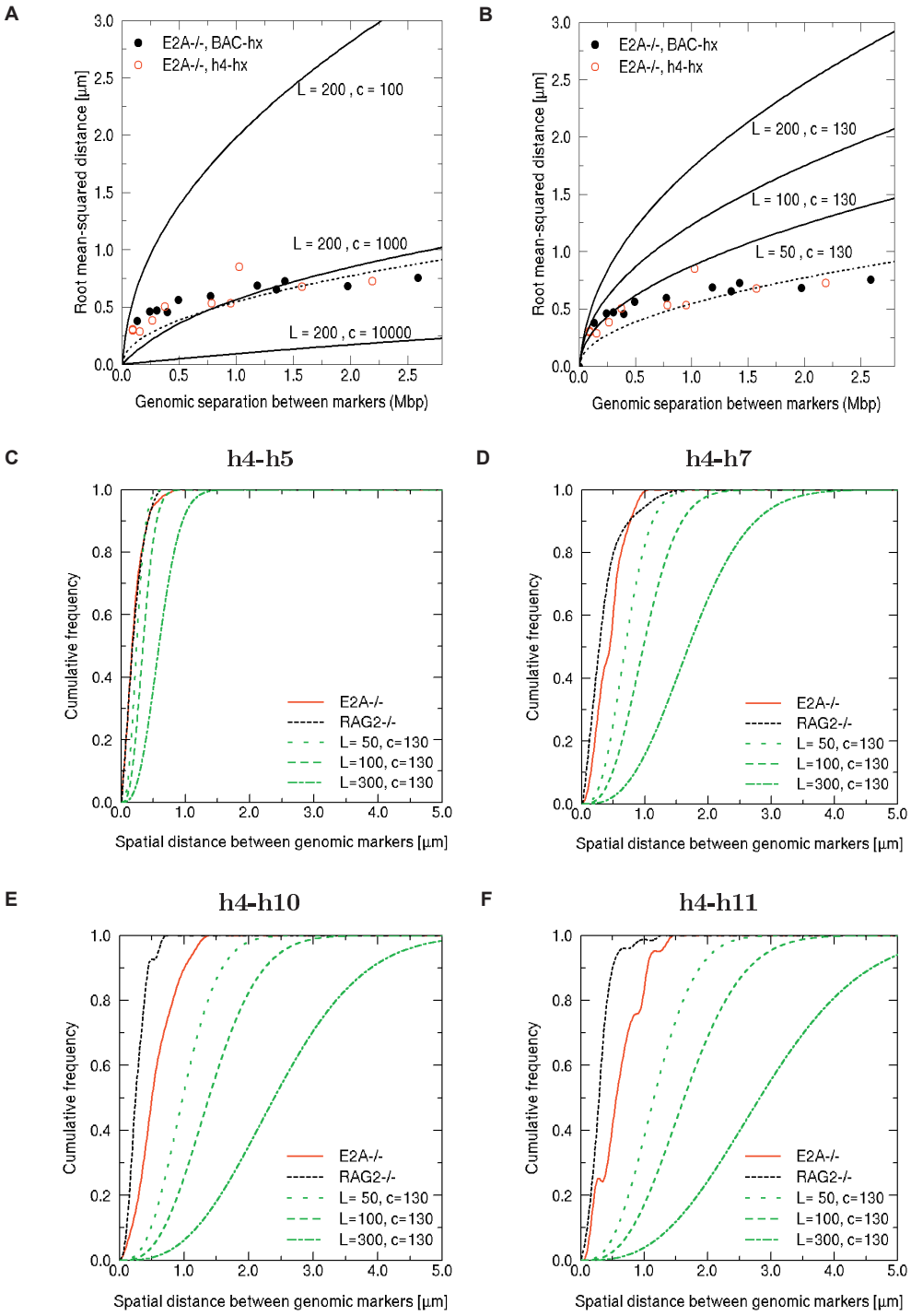
(1) Spatial distances depend strongly on the probe placement and whether the 3D-architecture is fixed or variable relative to the DNA sequence (i.e. whether loops with a defined rigid/ ”covalent” loop base exist or not);

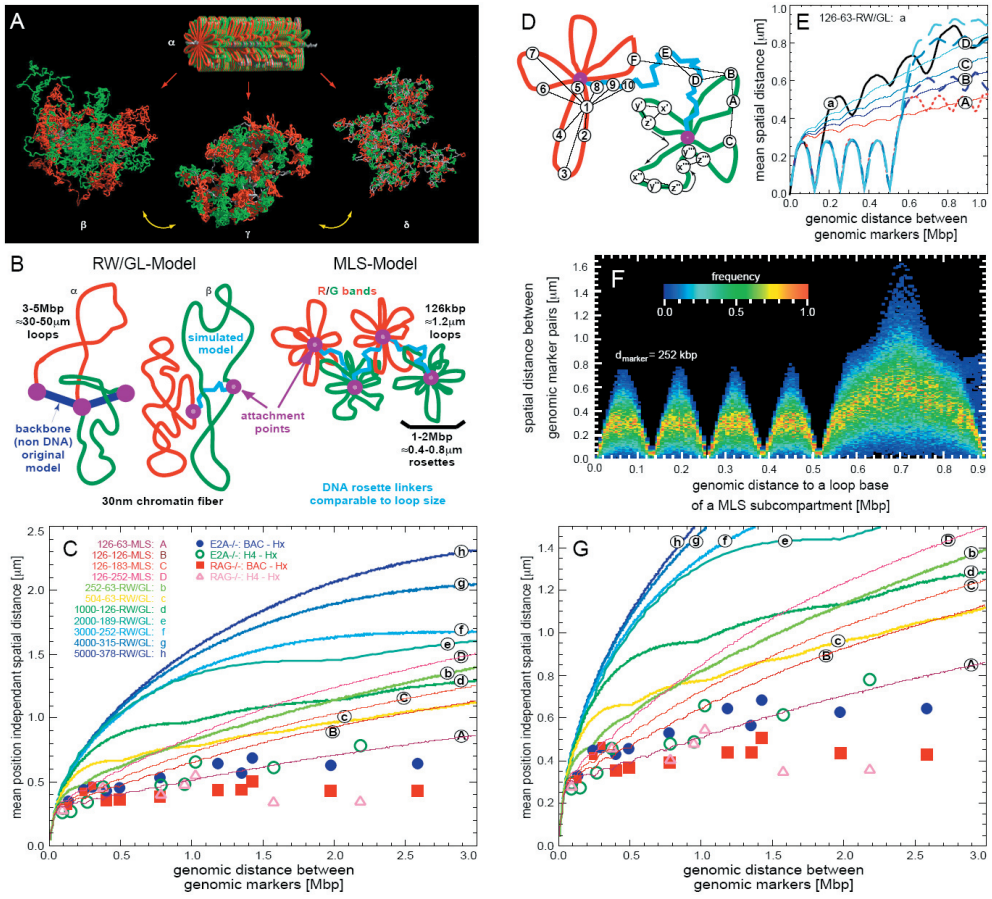
(2) *A priori* neither is the coverage of the *IGH* locus with genomic markers high enough, nor do the results of the quantitative measurements indicate the location of the experimental probes or the variability of the underlying chromatin architecture.

Therefore, the best approach is to assume both and to place the virtual markers in a position-independent and position-dependent manner on simulated chromatin architecture. Then the average is taken over several states of the same architecture. For position-dependent placement, the virtual anchor point was placed at the base of a loop (Figure 6D). For the position-independent placing of the virtual markers, the anchor point was placed randomly within the chain of segments. This sort of unconstrained measurement also reflects experimental imperfections. Vice versa these two types of analysis also reflect the variability of genome architecture: Changes in the architecture will be reflected by changing the position of the marker with respect to the architecture! This is obvious e.g. for virtual average spatial distances (Figure 6C-E) or for virtual distance distributions, i.e. interaction probabilities (Figure 6F).

When first considering position-independent achieved virtual average spatial distances, it is immediately apparent that already above 250 kb the experimental data do not agree with RW/GL models, even with loops as small as 252 kb (Figure 6C and G). These data are consistent with the analysis described above that excluded random walk behavior. Instead, the experimental data are most consistent with that described for an MLS model for both pre-pro-B and pro-B cells. Estimated from the genomic separations below 500 kb, the loop sizes range from around 80 to 150 kb (for more detailed analysis of loop size variation see Knoch[3]). The virtual spatial distance behavior also supports the observation that there might be possible agreements with a wormlike chain for very small genomic separations. Thus, within a loop, the chromatin could behave like a random walk. However, the smallest genomic separation we studied was 94 kb. Therefore, we could not perform a reasonable analytical analysis with low influence of the chromatin folding to determine the persistence length and the internal chromatin fiber density. Spacer sizes of 60 to 70 kb show the best

Figure 5. (facing page) The *IGH* Chromatin Topology cannot be Described as a Worm-like Chain.
A. Using the Kratky-Porod equation,[35] spatial distances were plotted as a function of genomic separation for different chromatin densities, but with a constant persistence length. Spatial distances were plotted as a function of genomic separation for a worm-like chain with a persistence length of 200 nm and a range of chromatin densities (100-10,000 bp/nm). **B.** Using the Kratky-Porod equation, spatial distances were plotted as a function of genomic separation with a range in persistence length of 50-200 nm and a constant chromatin density of 130 bp/nm. **C-F.** Probabilities of VH regions to interact with DHJH elements in pre-pro-B and pro-B cells compared to theoretical probabilities of a 3D random walk. Cumulative frequencies were obtained by accumulating the frequency values corresponding to the spatial distances in intervals of 100 nm using the DHJH elements (probe h4) as an anchor. Cumulative frequency distributions for the random walk were determined for different persistence lengths (see Materials and Methods).


A
E2A-/-, BAC-hx
E2A-/-, h4-hx
L = 200 , c = 100
L = 200 , c = 1000
L = 200 , c = 10000
Root mean-squared distance [µm]
Genomic separation between markers (Mbp)
B
E2A-/-, BAC-hx
E2A-/-, h4-hx
L = 200 , c = 130
L = 100 , c = 130
L = 50 , c = 130
Root mean-squared distance [µm]
Genomic separation between markers (Mbp)
C
h4-h5
D
h4-h7
E
h4-h10
F
h4-h11
Cumulative frequency
E2A-/-
RAG2-/-
L= 50, c=130
L=100, c=130
L=300, c=130
Spatial distance between genomic markers [µm]



agreements for genomic separations up to ~1.0 Mb. For larger separations, the virtual spatial distances would correspond to much smaller spacers for both pre-pro-B and pro-B cells and independent of the experimental anchor point. It now becomes obvious that the plateau behavior in pro-B cells, for genomic separations above 0.5 to 1.0 Mb is not arbitrary but distinct: Thus, the entire *IGH* locus in pro-B cells might span one single loop cluster. However, it is conceivable that two or more loop clusters are present, connected by relatively small sized spacers. The virtual markers placed in a position-dependent way in fixed chromatin architecture clarify this (Figure 6E). Notably, this domain contains the distal VH regions and shows distinct conformations in pre-pro-B and pro-B cells.

The computer modeling also confirms that the local inversions and peaks as well as the functional differences mentioned above are relevant and can be indeed interpreted as rearrangement of the locus architecture. These interpretations are even more striking when considering the *IGH* locus dynamics: The interaction probability between VH and DHJH gene segments is reflected by changing the position-dependent placed makers as function

Figure 6. (facing page) Comparison and Evaluation of Spatial Distances Between Genomic Markers in the Murine *IGH* Locus by Computer Simulations.
A. Simulation of decondensation of a metaphase-like chromosome (α) consisting of stacked rosettes linked by a chromatin loop into thermodynamically equilibrated interphase configurations. γ depicts a Multi-Loop-Subcompartment (MLS) organization with 126 kb rosettes and 126 kb spacers. β depicts a Random-Walk/Giant-Loop (RW/GL) organization with 5 Mb loops and 378 kb spacers, whereas the RW/GL organization in δ consists of 126 kb loops and 63 kb spacers. The fiber diameter was set at 30 nm and adjacent loops or rosettes are shown in red and green. **B.** Diagrams of hypothetical chromatin organization according to the RW/GL model and the MLS model. The RW/GL model is depicted with large loops are attached to a non-DNA backbone (α), and with large loops simulated with a chromatin spacer in-between (β). The MLS model shows ~100 kb loops that form rosettes, which are connected by a chromatin spacer. **C.** Spatial distances of the experimental data and position dependent distance simulations of the RW/GL and MLS models with various parameters plotted as a function of genomic separation. The experimental data is plotted as independent data points, whereas the average of the distance distribution for the various models are depicted as solid lines. Nomenclature for the different models: nomenclature: (loop size [kb])-(linker size [kb])-(modelname). **D.** Schematic representation of two rosettes and a spacer according to the MLS model. For position-independent spatial distances, markers with a given genetic separation are placed randomly with respect to the chromatin topology reflecting also a dynamic chromatin organization, although with certain loop and spacer properties. In contrast, for position-dependent spatial distances, distances are measured from one marker to other markers reflecting a static chromatin organization (1-10). Thus, strong dependencies occur for position-dependent spatial distances if an ensemble of genomic markers is shifted through given chromatin topology (x',x",x'"). **E.** The comparison between simulated position-dependent and position-independent spatial distances reveal in the first case the detailed loop structure and transition into a spacer and the next loop or rosette (see model legend in **C**). **F.** The distributions of position-dependent spatial distances of a genetic marker pair with 252 kb genetic distance shifted through a rosette and into a spacer of an 126-126-MLS model reveals the dependence on the marker position. **G.** Spatial distances of the experimental data and position-dependent simulations of the RW/GL and MLS models with various parameters plotted as a function of genomic separation (see model legend in **C**).

of their location, thus completely supporting the interpretation from above (Figure 6F). Consequently, analysis of the 3D chromatin architecture in the *IGH* locus by comparison to structural computer models reveals that an organization into at least one or two rosette-like clusters of small loops connected by spacers (Figure 7). Taken together, the experimental data are most consistent with a MLS configuration in which the *IGH* locus is organized in 1-2 Mb subcompartments, which contain multiple loops, separated by spacers.

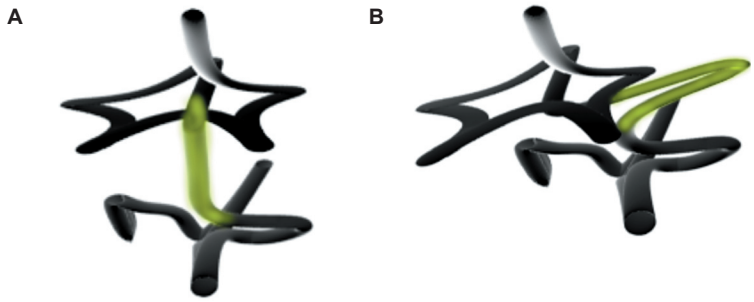


Figure 7. Hypothetical diagram of *IGH* rosettes with a sizeable spacer (in green).
We hypothesize that pre-pro-B cells have a stretched spacer (**A**), whereas the spacer in-between the *IGH* rosettes is contracted in pro-B cells (**B**), thereby repositioning the distal V_H gene segments within the same spatial distance from the D_HJ_H cluster as the proximal V_H gene segments.

DISCUSSION

The DNA sequences of a wide variety of genomes have been characterized in substantial detail. How the genome is tangled in 3D-space, however, is largely unknown. Here we used the *IGH* locus as a model system to examine the spectrum of structural conformations in interphase nuclei. The *IGH* locus is of interest and suitable for these studies for the following reasons: (1) The *IGH* locus is large, spanning over 3 Mb of genomic sequence, making it amenable to physical scrutiny; (2) The *IGH* locus contains multiple well-characterized coding and non-coding elements; (3) The *IGH* regulatory elements are well characterized; (4) The *IGH* locus undergoes genomic rearrangement between distant elements in a temporal and lineage-specific fashion. Here we demonstrate that the *IGH* locus, while highly dynamic, is structured and how the *IGH* locus configuration is altered to affects its function.

High-resolution 3D Structure-preserving Fluorescence Microscopy and *IGH* Locus Topology

In the studies described here we use spectral precision distance fluorescence microscopy to determine spatial distances between genetic markers. Measuring the point spread function and chromatic shifts, we demonstrate that fluorescence microcopy using two fluorochrome-labeled probes, unlike single labeled probes, allows a resolution (less than 50 nm) that is well below the average spatial distances and standard deviations recorded in this study. Consequently, spectral precision distance fluorescence microscopy allows accurate spatial measurements between probes that are labeled with different fluorochromes to determine the *IGH* locus topology.

We used a structure-preserving fixation and permeabilization protocol to prepare nuclei for spectral precision distance fluorescence microscopy. However, it remains to be determined to what degree our experimental design affects *IGH* locus spatial organization. Although the cells were fixed with paraformaldehyde in an isotonic environment, we cannot exclude the possibility that shifting of the *IGH* fiber may have occurred during the fixation procedure. Furthermore, we cannot exclude the possibility that DNA denaturation leads to distortions in *IGH* topology. In *Saccharomyces cerevisiae*, both 3D structure-preserving fluorescence microscopy and live-imaging of fluorescently labeled proteins targeted to the chromatin fiber have been used to determine spatial distances.[34] Direct comparison of these two approaches, has indicated that structure-preserving fixation and denaturation conditions do not substantially alter spatial distances.[34] Based on these results, it seems unlikely that the fixation and denaturation conditions used in this analysis substantially interfere with the spatial organization of the *IGH* fiber. Nevertheless, we note that ultimately it will be important to target fluorescently labeled proteins to the *IGH* chromatin fiber, to examine the *IGH* locus conformation and dynamics in live B cells.

IGH Locus Topology

Previous studies have established that the *IGH* locus shows long-range complexity in both B-and T-lineage cells.[29-32] However, it has remained unclear whether the *IGH* locus configuration can be described as a random walk or is specifically structured to facilitate long-range interactions. As a first approach to answer this question, we performed distance

measurements between multiple regions within the *IGH* locus, and compared the obtained experimental data to different models of polymer structure. The data are not at all compatible with an *IGH* topology described as a 3D random walk/worm-like chain. Rather, comparison of the experimental observations with that of simulated data suggests that the *IGH* locus is structured into subcompartments that contain bundles of loops (multi-loop subcompartments). While the overall topology of the *IGH* locus is most consistent with that of the MLS model, there are substantial differences between the *IGH* configurations in pre-pro-B and pro-B cells. The best agreement for the *IGH* configuration in pre-pro-B cells is obtained when we assume 126 kb size loops that are organized as rosettes (1-2 Mb) separated by spacers of 65-126 kb. The *IGH* topology in pro-B cells differs from pre-pro-B cells beyond a region separating the proximal and distal VH regions (Figure 6). How is the *IGH* locus organized in pro-B cells? It is plausible that in pro-B cells, the proximal and distal VH regions are localized in a single rosette. However, we favor a configuration in which multiple rosettes, separated by spacers, comprise the *IGH* locus in both pre-pro-B and pro-B cells (Figure 7). We note that shortening the spacer is an attractive mechanism to decrease the spatial distance separating compartments that contain DHJH elements and distal VH regions (Figures 3C and 7). A candidate for modulating the spacer size is Pax5, which has been demonstrated to promote *IGH* locus compaction in pro-B cells.[30]

Rosette formation of the *IGH* Fiber

The organization of the *IGH* locus in clusters of loops raises the question how these loops are formed. Previous studies have suggested that the *IGH* fiber is attached to the nuclear matrix, a component that is resistant to salt extraction and DNase I digestion.[36] In principle, the loops may attach to the matrix associating regions (MARs) that have been identified flanking the *IGH* intronic enhancer and flanking the majority of VH gene segments.[37, 38] The MAR sequences present within the *IGH* locus are not critical for VH to DHJH gene rearrangement.[39] Still, it is conceivable that deleting a single MAR does not substantially interfere with overall *IGH* topology, and it will be of interest to examine the *IGH* locus topology in pro-B cells deficient for MAR-binding proteins such as Cux/CDP, Bright and SATB2.

Does the comparison of experimental and simulated data imply that the *IGH* fiber is organized into rosettes consisting of 126 kb loops? We consider this highly unlikely. Rather, we suggest that loop size within the chromatin territories varies substantially and might be regulated, at least in part, by factors that interact with specific DNA sequences localized at the "bases" of loops. These factors could also determine the spacing between the loops within the subcompartments. The MLS model suggests that a bundle of loops forms a complete rosette. However, we emphasize that the folding of loops into higher order structures will not necessarily lead to the formation of complete rosettes. Rather, it is conceivable that rosettes only partially form and that the degree of rosette formation will depend on the presence of DNA sequences to which "loop-base" interacting factors bind. We also suggest that loops are not stable configurations, but fold and unfold continuously, regulated by transient interactions that involve "loop-base" interacting proteins.

Long-range Genomic Separation and Enhancer Function

Previous studies have established that enhancer elements act long-range in terms of genomic distance and independent of orientation in vertebrate organisms.[18] The *IGH* intronic

enhancer plays a critical role in regulating VH to DHJH gene rearrangements. Whereas distal and proximal VH to DHJH gene rearrangements are severely perturbed in mice that carry deletions of the intronic enhancer, DH to JH gene rearrangements are only modestly impaired.[40] Although still to be proven, we presuppose that the *IGH* intronic enhancer modulates accessibility of the VH gene segments to the recombination machinery. This implies that the enhancer acts over large genomic distances (up to 2 Mb). How does the *IGH* intronic enhancer act over such large genomic distances to modulate VH to DHJH gene rearrangement? We propose that the *IGH* locus is organized into bundles of loops, resulting in 1 Mb subcompartments. This topology allows regulatory elements, e.g. promoters and enhancers, separated by large genomic distances, to interact at close spatial distances.

Unlike mammalian cells, long-range genomic interactions in budding yeast have not been described. Why do enhancer elements have the ability to activate transcription from DNA elements that are separated by large genomic distances in the majority of eukaryotic organisms but not in budding yeast? We show here that if the *IGH* configuration is organized as a simple random walk, as observed for chromosomal DNA in budding yeast, the probability by which the majority of the proximal and all of the distal VH region promoters find the intronic enhancer would be overwhelmingly close to zero. Thus, our data imply that the statistical properties of the yeast chromatin fiber are incompatible with long-range genomic encounters (>250 kb). Furthermore, they suggest that the physical properties and structural organization of the *IGH* locus polymer in pro-B cells are the underlying mechanisms that allow DNA elements to act long range in genomic terms.

VH to DHJH Recombination and Long-range Interactions

The VH families have been categorized into proximal and distal classes, based on their genomic separation from the DHJH cluster. In general, no correlation exists between VH region usage and genomic location.[41-43] We found that the cumulative distance distributions indicate that distal and proximal VH regions have similar probabilities in finding the DHJH cluster in pro-B cells. Thus, these data explain previous observations indicating that VH regions scattered throughout the *IGH* locus rearrange with frequencies that seem independent of genomic location. We did observe a relatively high probability for VH regions located in immediate proximity of the DH region to encounter DHJH elements (Figures 3B and 4). These VH gene segments are more frequently found in VH to DHJH rearrangements.[44, 45] It is conceivable that these VH regions are located within the same loop or possibly spacer as the DHJH elements.

Statistical analysis indicates that the probabilities of distal VH regions to encounter the DHJH cluster in pre-pro-B cells are low when compared to pro-B cells. Thus, upon commitment to the B-cell lineage, the spectrum of *IGH* conformations is altered, to allow VH regions to associate with DHJH elements with higher frequency. What is the underlying mechanism that allows distal VH regions, separated by 1-2 Mb of genomic distance from the DHJH cluster, to rearrange with similar frequencies as VH regions localized in closer proximity? We speculate that the DH, JH and CH elements and the *IGH* intronic enhancer are packed into one subcompartment. The proximal and distal VH regions are organized in two distinct chromatin territories, each comprised of bundles of loops and spanning large genomic distances. In pro-B cells, both territories are positioned at equal distances from the DHJH elements, and consequently have the ability to interact with equal probabilities with DHJH elements (Figure 7).

Conclusion

The data described in this study provide insight into the structural organization of the *IGH* locus. It is unlikely that the topology described here is restricted to that of the *IGH* locus, and therefore it will be of interest to examine whether other genetic loci also show this configuration. Our observations also provide physical insight into how DNA elements that are separated by large genomic distances, encounter each other. It demonstrates that the chromatin topology determines the probability of these encounters and it will be of interest to examine how chromatin topology has evolved in multicellular organisms to allow high-density packing and promote long-range genomic interactions.

EXPERIMENTAL PROCEDURES

Mice and Cell Culture

All mice used were maintained in a C57BL/6 background. $Rag2^{-/-}$ pro-B cells and $E2A^{-/-}$ hematopoietic progenitors were isolated and grown as described previously.[32, 46] Two cell types were used in these studies: (1) An E2A-deficient hematopoietic progenitor cell line was examined, because these cells are arrested at the pre-pro-B cell stage and have not yet committed to the B-cell lineage;[46] (2) RAG2-deficient pro-B cells were chosen, because they are committed to the B-cell lineage but unlike wild-type pro-B cells they carry the *IGH* locus in germline configuration.[47] Since the genomic organization of the *IGH* locus differs substantially between C57Bl/6 and 129 mice strains, genomic DNA was isolated from E2A-deficient pre-pro-B and RAG2-deficient pro-B cells to determine their origins. PCR analysis confirmed that both cell types were derived from C57Bl/6 mice (R.R., data not shown). Thus, the genomic distances calculated in the experiments described in this study were derived from the C57Bl/6 genome.

Design, Cloning and Labeling of 10 kb Probes

A percentage identity plot of the *IGH* locus was generated using the PipMaker tool.[48] Eleven unique 10 kb DNA segments, i.e. regions that were not duplicated within the *IGH* locus and showing minimal genomic repeats, were identified on the generated percentage identity plot. Long-range PCR (Eppendorf) using newly designed primers was used to amplify the selected 10 kb stretches. The obtained PCR products were cloned into the pGEMTeasy vector system (Promega), and the identities of individual clones were confirmed by restriction mapping and DNA sequencing. One μg of DNA was labeled with aminoallyl-dUTP (ARES labeling kits, Invitrogen) by nick-translation (Roche) in a 20 μl reaction for 3 hours and 45 minutes. The nick-translated products were size selected on a 0.6 % agarose gel and 10-500 bp fragments were purified using a Qiagen gel extraction kits. Aminoallyl-modified DNA was fluorescently labeled with succinimidyl ester derivatives of Alexa fluorochromes (Alexa-488, -594 or -647).

High-resolution 3D Structure-preserving FISH

Approximately 40 μl of a 1 x 10^6 cells/ml suspension of cells was directly attached to coverslips. The cells were fixed in 4 % paraformaldehyde for 10 min at room temperature

(RT) and subsequently incubated with 0.1M Tris-Cl pH7.2 for 10 min at RT and washed with PBS. The cells were permeabilized for 10 min at RT with PBS plus 0.1 % triton X-100 plus 0.1 % saponin (WBI) and then incubated for 20 min in 20 % glycerol in PBS solution. Subsequently, the coverslips were immersed in liquid nitrogen and thawed for three consecutive times. The cells were washed once with PBS and incubated at 37°C in PBS plus 5 % bovine serum albumin and 0.1% Triton X-100 (BBI). The cells were subsequently incubated for 5 min in the presence of 0.1M Tris-Cl pH7.2 and then washed in phosphate buffered saline. The coverslips were then treated for 30 minutes with a 0.1 N HCl solution at room temperature, washed once in PBS, and incubated with 0.1 μg/ml DNase-free RNase in BB1 for 1 hour at 37 °C. The cells were washed once with PBS, incubated in PBS supplemented with 0.5 % triton X-100, 0.5 % saponin solution for 30 minutes at room temperature, and washed once with PBS. Nuclear DNA was denatured by incubating coverslips for 2 minutes and 30 seconds at 73°C in 2x SSC, 70 % formamide solution, followed by an incubation of 1 minute in 2x SSC plus 50 % formamide. Excess liquid was removed and 10 μl of hybridization cocktail was added to each of the coverslips. The coverslips were mounted, sealed with rubber-cement, and incubated overnight at 37°C. The hybridization solution contained 400 ng of each labeled probe, 4 μg of mouse C_0t DNA, 1 μg of sheared salmon-sperm DNA dissolved in 50 % formamide, 2x SSC and 10 % dextran sulfate. The probes were denatured at 70°C for 5 minutes and chilled on ice prior to incubation with the cells on the coverslips. The following day, the coverslips were removed and washed once with 2x SSC and 50 % formamide for 15 minutes, and 3 times with 2x SSC for 5 minutes at 37 °C. The coverslips were subsequently incubated in 2x SSC, 3 % BSA, 0.1 % Tween-20 (BB2) at room temperature for 30 minutes. The coverslips were then washed twice for 10 minutes in 2x SSC plus 0.1 % Tween-20 and once in PBS at room temperature. Excess PBS was removed and the coverslips were mounted on slides.

Image Acquisitions, Distance Calculations, Calibration and Statistics

Images were captured with a DeltaVision deconvolution microscope system (Applied Precision, Inc.) located at the UCSD cancer center microscope facility. Using a 100x (NA 1.4) lens, images of approximately 40 serial optical sections spaced by 0.2 μm were acquired from experiments involving BAC DNA probes only, and 80 serial optical sections spaced by 0.1 μm were acquired from experiments involving 10 kb DNA probes. The data sets were deconvolved and optical sections merged to produce 3D images using SoftWorx software (Applied Precision, Inc) on a Silicon Graphics Octane workstation. The 3D-coordinates of the centre of mass of each probe were obtained to calculate the distances separating each probe according to the equation

$$\sqrt{(X_a - X_b)^2 + (Y_a - Y_b)^2 + (Z_a - Z_b)^2}$$

where X,Y, and Z are the coordinates of object a or b.

To determine the smallest measurable distance by spectral precision distance fluorescence microscopy, the "resolution equivalent" (RE) was determined. Both the point spread function as well as chromatic shifts were determined. The PSF with its full width at half the maximum (FWHM) of the microscope was measured with fluorescent tetraspeck beads (Invitrogen, 0.2 μm) at 0.1 μm intervals.

To exclude positional shifts of signals representing probes labeled with different fluorochromes, the distance measurements were corrected for chromatic shifts. To measure chromatic shifts, 0.2 μm tetraspeck beads were used. Each bead was uniformly stained with a mixture of fluorescent dyes that emit in the blue, green, orange and dark red wavelengths. The coordinates of the center of the beads were determined for each of the colors that were used in the FISH analysis. For any two colors, the average separation between the coordinates (when observed under the filters for those colors) was used as the chromatic shift. The chromatic shifts were determined separately for the x, y and z coordinates (Appendix 4).

The average DNA compaction based on the linear contour length of 340 nm for 1.0 kb was calculated as follows:

$$\frac{\text{genomic lenght [kb]}}{\text{spatial distance [nm]}} \cdot 340\,[\text{nm/kb}]$$

Kratky-Porod Description of a Semi-flexible Polymer

The Kratky-Porod description of a semi-flexible polymer was used to compare the experimental data to that of the worm-like chain:[35]

$$\left\langle R^2 \right\rangle = 2 \cdot L^2 \cdot \left(\frac{d}{cL} - 1 + e^{-\frac{d}{cL}} \right)$$

where R indicates physical distance between genetic markers (nm), d refers to genomic separation between genetic markers (bp), c indicates DNA mass density (bp/nm), and L refers to the persistence length (nm).

Probability Distribution for a 3D Random Walk

The frequency distribution of the physical distances between the probes was determined in intervals of 100 nm, and the cumulative frequencies were plotted against the physical distances. The cumulative frequency indicates the fraction of the population of probes that lies within a certain physical distance from each other. Cumulative frequencies were also determined for the random walk model for a polymer chain. For this purpose, DNA was assumed to have a mass density of 130 bp/nm.[34] Consider two genomic probes separated by 'd' base pairs. If one probe is considered as a reference, the probability that the displacement of the second probe with respect to the first probe is '**R**' is given by:

$$P(R,a) = \left(\frac{3c}{2\pi da} \right)^{\frac{3}{2}} e^{\frac{3cR^2}{2da}}$$

where 'a' is the Kuhn length of DNA (nm) and c is the DNA mass density (bp/nm).[49] To obtain the probability that the second probe lies within a physical distance R nm from the first probe, one needs to evaluate the probability that the second probe lies anywhere within a sphere of radius R nm with the first probe at its center. This is given by:

$$\int_0^R \left(\frac{3c}{2\pi da} \right)^{\frac{3}{2}} e^{\frac{3cR^2}{2da}} 4\pi R^2 dR$$

This is the fraction of the population of probes that are separated by a distance of up to R nm. This gives the cumulative distribution function for the random walk model for polymers. The cumulative frequencies were determined for different persistence lengths and compared with the experimental data.

Computer Simulations

Using a two-step combined Monte-Carlo and Brownian Dynamics method, the Random-Walk/Giant-Loop (RW/GL) model[12] and the Multi-Loop Subcompartment (MLS) model[11] were simulated for human interphase chromosome 15.[3, 15, 50]

Assuming a flexible polymer chain, the chromatin fiber was split into ~3,300 segments of 300 nm (~31 kb). To each segment a harmonic stretching potential and between two segments a harmonic bending potential were assigned. To avoid self-crossing of the polymer chain, a short ranged excluded volume potential was introduced, whose potential barrier could be changed to facilitate chain dis-entanglement to speed up simulations. The simulation of single chromosomes necessitated placement into an embedding potential simulating the surrounding nuclear chromosomes. In the RW/GL model loops of 5.0, 4.0, 3.0, 2.0, 1.0, 0.504 and 0.252 Mb were used and connected by a chromatin spacer whose length was adjusted such that the global territory behavior yielded comparable results as in the MLS model. In the MLS model the loop size was 126 kb with spacer sizes of 63, 126, 189 and 252 kb. The number of loops in the rosettes varied, since the DNA content of the rosettes was assumed to be that of the metaphase ideogram banding pattern[51] divided by three to account for the transition into interphase ideogram bands.[52] The starting configuration of a chromosome had the approximate form and size of a metaphase-chromosome whose de-condensation resembles the natural process. Typically ~400,000 Monte Carlo steps were needed to generate enough statistically independent configurations at thermodynamic equilibrium. For comparison with experimental spatial distance measurements between genetic markers as function of their genetic separation, 100 to 150 statistically independent Monte Carlo configurations were taken as starting points for relaxation at higher spatial resolution by Brownian Dynamics methods using a decreased segment length of 50 nm (~5.2 kb), corresponding to 20,000 segments for chromosome 15. 2,000 Brownian Dynamics steps were performed until equilibration was reached again. The simulated spatial distances were determined position dependently, i.e. the marker pairs were placed in respect to the topological folding of the simulated chromosome and position independently, i.e. the pairs of markers were placed randomly and therefore regardless of any folding topology on the chromosome (for the underlying assumption see Results). All pairs of positions for one genomic separation were taken, the resulting spatial distance determined and averaged over one and the 100 to 150 statistically independent chromosome configurations.

ACKNOWLEDGMENTS

C.M. thanks Doug Smith, David Meijer and Rob Philips for stimulating discussions. Suchit Jhunjhunwala was supported by the NIH. S.J., M.C.v.Z. and M.P. thank the Imaging Core Facility at UCSD for assistance and encouragement. Menno van Zelm was supported by a fellowships from the Ter Meulen Fund – Royal Netherlands Academy of Arts and Sciences,

and the Dutch Society for Immunology. M.C.v.Z. and J.J.M.v.D. were supported by grant 349 from Sophia Kinderziekenhuis Fonds. Mandy Peak was supported by a Haemoglobin/ Blood Training Grant. The authors gratefully thank the High-Performance Computing Center Stuttgart (HLRS; grant HumNuc), the Supercomputing Center Karlsruhe (SCC; grant ChromDyn), and the Computing Facility of the German Cancer Research Center (DKFZ) for access to their CRAY T3E and IBM SP2s and the Bundesministerium für Bildung und Forschung (BMBF) under grant # 01 KW 9602/2 (German Human Genome Project) for financial support and the German Cancer Research Center (DKFZ) for a dissertation grant to Tobias Knoch for the start-up of the initial simulations part. The Erasmus Medical Center is thanked for providing T.A.K. with the capacity to build up the Biophysical Genomics Group and their general infrastructure. This work was supported by grants from the NIH to R.R and C.M. The authors have no financial interest related to this work.

REFERENCES

1. Cremer T, Cremer C. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Genet* 2001;**2**:292-301.
2. Dundr M, Misteli T. Functional architecture in the cell nucleus. *Biochem J* 2001;**356**:297-310.
3. Knoch TA. Approaching the three-dimensional organization of the human genome: structural-, scaling- and dynamic-properties in the simulation of interphase chromosomes and cell nuclei, long-range correlations in complete genomes, *in vivo* quantification of the chromatin distribution, construct conversions in simultaneous co-transfections. PhD thesis: Ruperto-Carola University, Heidelberg, Germany, and TAK Press; 2002.
4. Bover T. Die Blastomerenkerne von Ascaris meglocephala und die Theorie der Chromosomenindividualitaet. *Archiv für Zellforschung* 1909;**3**:181-268.
5. Rabl C. Über Zellteilung. *Morphologisches Jahrbuch* 1885;**10**:214-330.
6. Paulson JR, Laemmli UK. The structure of histone-depleted metaphase chromosomes. *Cell* 1977;**12**:817-28.
7. Pienta KJ, Coffey DS. A structural analysis of the role of the nuclear matrix and DNA loops in the organization of the nucleus and chromosome. *J Cell Sci Suppl* 1984;**1**:123-35.
8. Okada TA, Comings DE. Higher order structure of chromosomes. *Chromosoma* 1979;**72**:1-14.
9. Belmont AS, Bruce K. Visualization of G1 chromosomes: a folded, twisted, supercoiled chromonema model of interphase chromatid structure. *J Cell Biol* 1994;**127**:287-302.
10. Trask BJ, Massa H, Kenwrick S, Gitschier J. Mapping of human chromosome Xq28 by two-color fluorescence in situ hybridization of DNA sequences to interphase cell nuclei. *Am J Hum Genet* 1991;**48**:1-15.
11. Munkel C, Langowski J. Chromosome structure described by a polymer model. *Phys Rev sect E* 1998;**57**: 5888-96.
12. Sachs RK, van den Engh G, Trask B, Yokota H, Hearst JE. A random-walk/giant-loop model for interphase chromosomes. *Proc Natl Acad Sci U S A* 1995;**92**:2710-4.
13. Trask BJ, Allen S, Massa H, Fertitta A, Sachs R, van den Engh G, Wu M. Studies of metaphase and interphase chromosomes using fluorescence in situ hybridization. *Cold Spring Harb Symp Quant Biol* 1993;**58**:767-75.
14. Warrington JA, Bengtsson U. High-resolution physical mapping of human 5q31-q33 using three methods: radiation hybrid mapping, interphase fluorescence in situ hybridization, and pulsed-field gel electrophoresis. *Genomics* 1994;**24**:395-8.
15. Knoch TA, Münkel C, Langowski J. Three-dimensional organization of chromosome territories in the human interphase nucleus. In: Krause E, Jäger W, eds. High Performance Computing in Science and Engineering 1999: Springer Berlin-Heidelberg-New York; 2000:229-38.
16. Herendeen DR, Kassavetis GA, Geiduschek EP. A transcriptional enhancer whose function imposes a requirement that proteins track along DNA. *Science* 1992;**256**:1298-303.
17. Rombel I, North A, Hwang I, Wyman C, Kustu S. The bacterial enhancer-binding protein NtrC as a molecular machine. *Cold Spring Harb Symp Quant Biol* 1998;**63**:157-66.

18. Banerji J, Rusconi S, Schaffner W. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell* 1981;**27**:299-308.
19. Carter D, Chakalova L, Osborne CS, Dai YF, Fraser P. Long-range chromatin regulatory interactions *in vivo*. *Nat Genet* 2002;**32**:623-6.
20. Tolhuis B, Palstra RJ, Splinter E, Grosveld F, de Laat W. Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol Cell* 2002;**10**:1453-65.
21. Eivazova ER, Aune TM. Dynamic alterations in the conformation of the Ifng gene region during T helper cell differentiation. *Proc Natl Acad Sci U S A* 2004;**101**:251-6.
22. Spilianakis CG, Flavell RA. Long-range intrachromosomal interactions in the T helper type 2 cytokine locus. *Nat Immunol* 2004;**5**:1017-27.
23. Chevillard C, Ozaki J, Herring CD, Riblet R. A three-megabase yeast artificial chromosome contig spanning the C57BL mouse Igh locus. *J Immunol* 2002;**168**:5659-66.
24. Jung D, Giallourakis C, Mostoslavsky R, Alt FW. Mechanism and control of V(D)J recombination at the immunoglobulin heavy chain locus. *Annu Rev Immunol* 2006;**24**:541-70.
25. Schatz DG, Spanopoulou E. Biochemistry of V(D)J recombination. *Curr Top Microbiol Immunol* 2005;**290**: 49-85.
26. Alt FW, Yancopoulos GD, Blackwell TK, Wood C, Thomas E, Boss M, Coffman R, Rosenberg N, Tonegawa S, Baltimore D. Ordered rearrangement of immunoglobulin heavy chain variable region segments. *Embo J* 1984;**3**:1209-19.
27. Manz J, Denis K, Witte O, Brinster R, Storb U. Feedback inhibition of immunoglobulin gene rearrangement by membrane mu, but not by secreted mu heavy chains. *J Exp Med* 1988;**168**:1363-81.
28. Nussenzweig MC, Shaw AC, Sinn E, Danner DB, Holmes KL, Morse HC, 3rd, Leder P. Allelic exclusion in transgenic mice that express the membrane form of immunoglobulin mu. *Science* 1987;**236**:816-9.
29. Kosak ST, Skok JA, Medina KL, Riblet R, Le Beau MM, Fisher AG, Singh H. Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development. *Science* 2002;**296**:158-62.
30. Fuxa M, Skok J, Souabni A, Salvagiotto G, Roldan E, Busslinger M. Pax5 induces V-to-DJ rearrangements and locus contraction of the immunoglobulin heavy-chain gene. *Genes Dev* 2004;**18**:411-22.
31. Roldan E, Fuxa M, Chong W, Martinez D, Novatchkova M, Busslinger M, Skok JA. Locus 'decontraction' and centromeric recruitment contribute to allelic exclusion of the immunoglobulin heavy-chain gene. *Nat Immunol* 2005;**6**:31-41.
32. Sayegh C, Jhunjhunwala S, Riblet R, Murre C. Visualization of looping involving the immunoglobulin heavy-chain locus in developing B cells. *Genes Dev* 2005;**19**:322-7.
33. Hesslein DG, Schatz DG. Factors and forces controlling V(D)J recombination. *Adv Immunol* 2001;**78**:169-232.
34. Bystricky K, Heun P, Gehlen L, Langowski J, Gasser SM. Long-range compaction and flexibility of interphase chromatin in budding yeast analyzed by high-resolution imaging techniques. *Proc Natl Acad Sci U S A* 2004;**101**:16495-500.
35. Kratky O, Porod G. Röntgenuntersuchung gelöster Fadenmoleküle. *Rec Trav Chim Pays-Bas* 1949;**68**:1106-33.
36. Cockerill PN, Garrard WT. Chromosomal loop anchorage of the kappa immunoglobulin gene occurs next to the enhancer in a region containing topoisomerase II sites. *Cell* 1986;**44**:273-82.
37. Goebel P, Montalbano A, Ayers N, Kompfner E, Dickinson L, Webb CF, Feeney AJ. High frequency of matrix attachment regions and cut-like protein x/CCAAT-displacement protein and B cell regulator of IgH transcription binding sites flanking Ig V region genes. *J Immunol* 2002;**169**:2477-87.
38. Webb CF, Das C, Eneff KL, Tucker PW. Identification of a matrix-associated region 5' of an immunoglobulin heavy chain variable region gene. *Mol Cell Biol* 1991;**11**:5206-11.
39. Sakai E, Bottaro A, Davidson L, Sleckman BP, Alt FW. Recombination and transcription of the endogenous Ig heavy chain locus is effected by the Ig heavy chain intronic enhancer core region in the absence of the matrix attachment regions. *Proc Natl Acad Sci U S A* 1999;**96**:1526-31.
40. Perlot T, Alt FW, Bassing CH, Suh H, Pinaud E. Elucidation of IgH intronic enhancer functions via germ-line deletion. *Proc Natl Acad Sci U S A* 2005;**102**:14362-7.
41. Brodeur PH, Riblet R. The immunoglobulin heavy chain variable region (Igh-V) locus in the mouse. I. One hundred Igh-V genes comprise seven families of homologous genes. *Eur J Immunol* 1984;**14**:922-30.
42. Gu H, Tarlinton D, Muller W, Rajewsky K, Forster I. Most peripheral B cells in mice are ligand selected. *J Exp Med* 1991;**173**:1357-71.

43. Love VA, Lugo G, Merz D, Feeney AJ. Individual V(H) promoters vary in strength, but the frequency of rearrangement of those V(H) genes does not correlate with promoter strength nor enhancer-independence. *Mol Immunol* 2000;**37**:29-39.
44. Malynn BA, Yancopoulos GD, Barth JE, Bona CA, Alt FW. Biased expression of JH-proximal VH genes occurs in the newly generated repertoire of neonatal and adult mice. *J Exp Med* 1990;**171**:843-59.
45. Yancopoulos GD, Desiderio SV, Paskind M, Kearney JF, Baltimore D, Alt FW. Preferential utilization of the most JH-proximal VH gene segments in pre-B-cell lines. *Nature* 1984;**311**:727-33.
46. Ikawa T, Kawamoto H, Wright LY, Murre C. Long-term cultured E2A-deficient hematopoietic progenitor cells are pluripotent. *Immunity* 2004;**20**:349-60.
47. Shinkai Y, Rathbun G, Lam KP, Oltz EM, Stewart V, Mendelsohn M, Charron J, Datta M, Young F, Stall AM, et al. RAG-2-deficient mice lack mature lymphocytes owing to inability to initiate V(D)J rearrangement. *Cell* 1992;**68**:855-67.
48. Schwartz S, Zhang Z, Frazer KA, Smit A, Riemer C, Bouck J, Gibbs R, Hardison R, Miller W. PipMaker--a web server for aligning two genomic DNA sequences. *Genome Res* 2000;**10**:577-86.
49. Grosberg AY, Khokhlov AR. Giant Molecules: Academic Press; 1997.
50. Knoch TA. Towards a hollistic understanding of the human genome by determination and integration of its sequential and three-dimensional organization. In: Krause E, Jäger W, eds. High Performance Computing in Science and Engineering 1999: Springer Berlin-Heidelberg-New York; 2003:421-40.
51. Franke UC, Scambler PJ, Loffler C, Lons P, Hanefeld F, Zoll B, Hansmann I. Interstitial deletion of 22q11 in DiGeorge syndrome detected by high resolution and molecular analysis. *Clin Genet* 1994;**46**:187-92.
52. Yunis JJ. Mid-prophase human chromosomes. The attainment of 2000 bands. *Hum Genet* 1981;**56**:293-8.

IV

Gross Deletions involving *IGHM*, *BTK* or *Artemis*: A model for Genomic Lesions Mediated by Transposable Elements

Menno C. van Zelm, Corinne Geertsema, Nicole Nieuwenhuis, Dick de Ridder, Mary Ellen Conley, Claudine Schiff, Ilhan Tezcan, Eva Bernatowska, Nico G. Hartwig, Elisabeth A.M. Sanders, Jiri Litzman, Irina Kondratenko, Jacques J.M. van Dongen, and Mirjam van der Burg

From the Erasmus MC, Dept. Immunology (M.C.v.Z.; C.G.; N.N.; J.J.M.v.D.; M.v.d.B.), and Dept. Pediatrics, Rotterdam, the Netherlands (M.C.v.Z.; N.G.H.); the Information & Communication Theory Group, Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, the Netherlands (D.R.); Dept. Pediatrics, University of Tennessee College of Medicine and St. Jude Children's Research Hospital, Memphis, TN, USA (M.E.C.); Centre d'Immunologie de Marseille-Luminy, Marseille, France (C.S.); the Div. Pediatric Immunology, University of Hacettepe, Ankara, Turkey (I.T.); Dept. Immunology, The Children's Memorial Health Institute, Warsaw, Poland (E.B.); Dept. Pediatrics, Wilhelmina Children's Hospital, UMCU, Utrecht, the Netherlands (E.A.M.S.); Dept. Clinical Immunology and Allergology, St Anne's University Hospital, Masaryk University, Brno, Czech Republic (J.L.); and Dept. Clinical Immunology, Russian Clinical Children Hospital, Moscow, Russia (I.K.).

Submitted

ABSTRACT

Most genetic disruptions underlying human disease are micro-lesions, whereas gross lesions are rare with gross deletions being most frequently found (6 %). Similar observations have been made in primary immunodeficiency genes, such as *BTK*, but for unknown reasons the *IGHM* and *DCLRE1C* (*Artemis*) gene defects frequently represent gross deletions (58 % and 46 %, respectively). We characterized the gross deletion breakpoints in IGHM-, BTK- and Artemis-deficient patients. The *IGHM* deletion breakpoints did not show involvement of recombination signal sequences or immunoglobulin switch regions. Instead, similar to *BTK* and *Artemis* deletions, most *IGHM* breakpoints were located in or near sequences derived from transposable elements (TE). A role for these elements in mediating gross deletions in other genes disrupted in human disease has been suggested, but has never been specified. We found that *IGHM*, *BTK* and *Artemis* gross deletion breakpoints were frequently located in TEs belonging to types that were specifically overrepresented in the involved gene as compared to the average in the human genome. This concerned both long (LINE1) and short (*Alu*, MIR) interspersed elements, as well as LTR retrotransposons (ERV). Furthermore, a high total TE content (>40 %) was associated with an increased frequency of gross deletions. Both findings were further investigated and confirmed in a total set of 20 genes disrupted in human disease. Thus, for the first time, we provide evidence that a high TE content, irrespective of the type of element, results in the increased incidence of gross deletions as gene disruption underlying human disease.

KEY WORDS

Gross deletion, transposable element, *IGH*, *BTK*, *Artemis*.

INTRODUCTION

The majority of gene disruptions underlying human inherited diseases are micro-lesions, i.e. single basepair substitutions (68 %) and small (<20 nt) deletions or insertions (25 %).[1] Gross lesions are less common, with gross deletions being most frequently found (6 %).[1] For gross deletions to occur, three conditions have to be met: (a) two double-stranded DNA breaks in distinct genomic locations that are (b) physically located in close proximity in order to (c) undergo incorrect repair. Based on limited sequence data of the junctions of gross deletion breakpoints in disease-causing alleles, three mechanisms have been described: 1) Mispairing of homologous sequences and unequal crossing over; 2) Non-homologous deletions; 3) Mispairing between two *Alu* short interspersed elements (SINE) and crossing-over [2]. Recent analysis of several gross deletion breakpoints indicates a role for single-strand annealing of repeated fragments in gross lesions.[3] Although this mechanism could explain incorrect repair of complex DNA structures, it remains unclear how two genomically distinct sites become located in close proximity.

The human genome contains four classes of interspersed repeats that are derived from transposable elements (TE) and comprise about 45 % of the total sequence.[4] Besides the above-mentioned SINEs, these are long interspersed elements (LINE), LTR retrotransposons and DNA transposons. Non-homologous gross deletion breakpoints are frequently located in or near these TEs,[5, 6] but the role of these repeats in mediating gross deletions is currently unknown.

Although inherited disorders of the immune system are rare, multiple gene defects have been identified during the last 15 years, mainly concerning micro-lesions.[7] However, gross deletions have also been found, for example involving the *DCLRE1C* (*Artemis*), *BTK* and immunoglobulin Cμ heavy chain (*IGHM*) genes.[8-13] The gross deletion frequency of *BTK* disruptions is 6 %, whereas about half of the reported disease-causing alleles of *IGHM* and *Artemis* are disrupted by gross deletions.[1, 8, 13] A mechanism underlying this striking difference has never been reported.

The immunoglobulin heavy chain (IgH) is part of the antibody molecule and is produced by B cells after V(D)J recombination of the *IGH* locus. The *IGH* locus contains multiple V_H, D_H and J_H gene segments and constant (C_H) regions. The V_H, D_H and J_H gene segments are flanked by recombination signal sequences (RSS) that are recognized by the recombinase machinery,[14] which is specifically expressed during precursor-B-cell differentiation to generate a functional VDJ exon.[15] The VDJ exon is initially spliced to the Cμ exons. During an immune response, genomic Ig class switch recombination can take place to replace the Cμ exons for one of the other C_H regions. This process is mediated by the Ig switch regions.[16] Similar to V(D)J recombination, Ig class switch recombination involves the genomic deletion of a large intervening DNA sequence.

Gross deletions involving *IGHM* can be up to 300 kb and can include upstream D_H and V_H gene segments and downstream constant gene regions.[11-13] We hypothesized that the complexity of the *IGH* locus contributes to the increased incidence of *IGHM*-disrupting gross deletions. Possible mediators are the RSS and Ig switch regions, which are target sequences for B-cell specific recombination events. Furthermore, germline duplications and deletions of Ig subclasses have been reported.[17-20] The C_H regions contain high degrees of sequence

homology and the altered haplotypes are thought to be the result of mispairing between highly homologous regions and subsequent unequal crossing-over during meiosis.[17-22]

In this study, we aimed to identify mechanisms that underlie the high frequency of gross deletions disrupting *IGHM* and *Artemis* as compared to *BTK*. Breakpoint analysis suggested involvement of TEs rather than recombination motifs. We hypothesized that the TE content of a gene is related to the gross deletions frequency, and this hypothesis was tested on 20 genes disrupted in human disease.

MATERIALS AND METHODS

Patient DNA isolation and PCR amplification

Genomic DNA was isolated from post-Ficoll PB granulocytes with the GenElute Mammalian Genomic DNA Miniprep Kit (Sigma-Aldrich, St. Louis, MO). The sizes of the gross deletions in IGHM, BTK and Artemis deficient patients were determined with scanning PCRs for gene segments, exons and subsequently for intronic regions. Seven unrelated IGHM-deficient patients with a homozygous gross deletion were studied. Of these patients, three were previously published as having a gross deletion (F2, F7 and SIOC),[11-13] and the other four were newly identified (BN, ID392, ID393 and ID394). Furthermore, three unrelated patients with a hemizygous deletion of *BTK* (ID113, ID434, ID440), and five unrelated patients with a homozygous deletion of *Artemis* (ID020, ID024, ID124, ID176, ID389) were studied. PCR reactions were essentially performed as described before with primers were designed to specifically amplify 200-600 bp of DNA (primer sequences shown in Appendices 7-9).[23]

Long-range (LR-)PCR, ligation-mediated (LM-)PCR and DNA sequencing for analysis of the breakpoint junctions

LR-PCR was performed as described before using the appropriate forward and reverse primers that were used to map the gross deletion boundaries.[24]

LM-PCR was essentially performed as described before with newly designed primers. Aliquots of 1 mg high molecular weight DNA were digested with blunt-end restriction enzymes (DraI, HincII, PvuII and StuI) and 50 mM of an adaptor (Clontech, Palo Alto, CA, USA) was ligated to both ends of the restriction fragments.[25] The ligation products were subjected to two rounds of PCR with nested adaptor-specific primers AP1 and AP2 (Clontech) and sets of primers designed upstream of the 5' end of the breakpoint. Atypical bands that appeared from patient's DNA, but not from control DNA were excised from the gel, purified using the QIAquick Gel Extraction Kit (Qiagen, Valencia, CA**)** and sequenced on an ABI Prism 7700 sequence detection system (Applied Biosystems, Foster City, CA).

Human disease-causing mutation data

Mutation data of genes disrupted in human disease were extracted from the February 2007 release of the human gene mutation database (HGMD).[1] To calculate the gross deletion frequency, we included the newly described gross deletion alleles from IGHM- and Artemis-deficient patients from this study, and newly identified *BLNK* mutations (unpublished results).

Sequence analysis of genes and breakpoint regions

Sequences of the *IGH* locus (NCBI: NG_001019.5), and the *BTK*, *Artemis* (*DCLRE1C*), *BLNK*, *APC*, *ATM*, *BRCA1*, *BRCA2*, *CFTR*, *DMD*, *Factor VIII*, *FANCA*, *HPRT1*, *LDLR*, *MECP2*, *MLH1*, *MSH2*, *NF1*, *RB1* and *VWF* genes including 10 kb upstream and downstream sequences, extracted from Ensembl v42 – Dec2006,[26] were annotated with TE-derived interspersed repeats by the CENSOR software tool of the Repbase database.[27] To obtain representative TE content frequencies, only genes spanning more than 50 kb were included.

The identified breakpoint regions were aligned with the gene sequences extracted from public databases. The Genewindow website was used to identify whether mismatches with standard sequences were previously described polymorphisms.[28] Complexity analysis of ± 25 bp flanking the breakpoint regions was performed to examine the potential contribution of local sequence structure to the mechanism of gross deletion in the *IGH* locus and the *BTK* and *Artemis* genes.[3, 29] In addition, the same region was scanned for the presence of 24 sequence motifs known to be associated with site-specific recombination, mutation, cleavage and gene rearrangement.[2] The sequences ± 1,000 bp flanking the breakpoint regions were annotated with TEs by the CENSOR software tool of the Repbase database.[27]

Statistical analysis

Differences in gross deletion frequencies were analyzed by using the non-parametric Mann-Whitney test (one-tailed; $P > 0.05$) in the GraphPad Prism program (GraphPad Software, San Diego, CA, USA).

RESULTS

Identification of five unique gross deletion breakpoint junctions in seven homozygous *IGHM*-deficient patients

DNA material was collected from four previously described and three newly identified unrelated patients with a full block in precursor-B-cell differentiation in bone marrow resulting from a homozygous deletion of *IGHM*. For each of these seven patients, the extent of the deletions was studied with newly designed and tested PCR amplicons recognizing unique VH and DH gene segments and *IGH* constant gene regions (Figure 1A). Subsequently, additioned PCR amplicons were designed to refine the position of the breakpoints within 2-3 kb to enable cloning of the breakpoint fusion with either LM-PCR or LR-PCR. This approach resulted in the characterization of five unique breakpoint junctions within the *IGH* loci of the seven IGHM-deficient patients (Figure 1A).

Patient BN was found to carry a homozygous *IGHM* deletion of 732 kb including VH3-37 until *IGHGP* (Figure 1A). The junction contained nine bp of unique sequence that could not be assigned to the 5' or 3' breakpoint region sequences (Figure 1B). The two mismatches with the 5' breakpoint region corresponded to previously described single nucleotide polymorphisms (SNP).

In addition to the *IGHM* deletion, the homozygous *IGH* locus of patient BN had a 33 kb deletion involving the VH3-39, VH7-40 and VH3-41 gene segments. About 1.3 kb sequence in-between the two gross deletions including the VH3-38 gene segment was still present

on the locus. The junction of the 33 kb deletion contains 2 bp microhomology and shows high degrees of sequence homology between the regions downstream of the breakpoints (Figure 1B). None of the other six *IGHM* patients carried this deletion. To test whether the 33 kb VH deletion is a polymorphism that is not related to the *IGHM* disruption, 66 controls were screened for this deletion. We found that 15 of 66 controls carried this deletion, but not the *IGHM* deletion; 7 were heterozygous and 8 homozygous. The breakpoint flanking sequences (200 bp) of the 23 control alleles were identical to patient BN including the six nucleotides that differed from the standard *IGH* sequence (Figure 1B). From this, we concluded that the 33 kb deletion of VH3-39, VH7-40 and VH3-41 is apparently a common polymorphism as it was present in 17 % of the *IGH* alleles in controls.

Identical breakpoint junctions were found in two unrelated patients: ID393 and ID394. Both patients lack 578 kb of genomic DNA starting from the VH3-19 gene segment and ending 3' of the *IGHG4* constant region. The two breakpoints were fused with a three-nucleotide microhomology region (Figure 1B). PCR amplicons flanking the breakpoint, VH3-20, *IGHE* and *IGHA2* were screened for 21 previously identified SNPs. The amplicons showed identical sequences for both patients: 11 sites differed from the standard sequence, 5 of which are known SNPs. Thus, these patients carry the same allele, which they inherited from a common ancestor.

The two unrelated patients ID392 and F7 also carried deleted alleles with identical breakpoint junctions. The deletions span 126 kb of genomic DNA starting at the KIAA0125 element in-between the VH6-1 and DH1-1 gene segments and ending 3' of *IGHD* (Figure 1B). One nucleotide microhomology was found at the breakpoint fusion (Figure 1B). Interestingly, 5' of the breakpoint junction of both patients, a small deletion of 19 bp with 6 bp microhomology in the fusion was found. This small deletion was not seen in 82 controls, indicating that this is not a common polymorphism. Taking into account the close proximity to the gross deletion breakpoint, we suggest that both the small and the gross deletion occurred in the same event. SNP analysis of the breakpoint fusion region, VH3-20, VHIV-20.1 and *IGHE* PCR amplicons showed no polymorphic differences between the alleles of patients ID392 and F7, indicating that both patients inherited the allele from a common ancestor.

The gross deletion in patient F2 spanned 71 kb from DH1-1 to *IGHM*. Interestingly, the junction contained a 51 nt insertion, which aligned almost perfectly with the intronic region in-between DH4-23 and DH5-24, 37 kb downstream of the 5' breakpoint (Figure 1A). Consequently, two deletions were found; the two junctions contained 1 and 3 nucleotides of microhomology, respectively (Figure 1B). The 5' deletion that does not involve *IGHM* was not found in 66 controls. Therefore, it does not appear to be a common polymorphism.

Patient SIOC had a gross deletion of 146 kb extending from DH6-13 to *IGHG3* (Figure 1A). The junction contains 4 nucleotides of microhomology (Figure 1B).

We identified 5 unique breakpoint junctions in 7 unrelated homozygous IGHM-deficient patients. Two previously identified IGHM-deficient patients, who were heterozygous for a gross deletion, did not carry one of the identified breakpoint junctions and are likely to carry unique deletions.[13] All 5 identified breakpoint junctions demonstrated the characteristics of non-homologous recombination with four out of five showing microhomology and one with an insertion of six nucleotides. Furthermore, both the 5' and the 3' breakpoints were scattered throughout the *IGH* locus and did not co-localize with RSS or Ig switch regions, thus

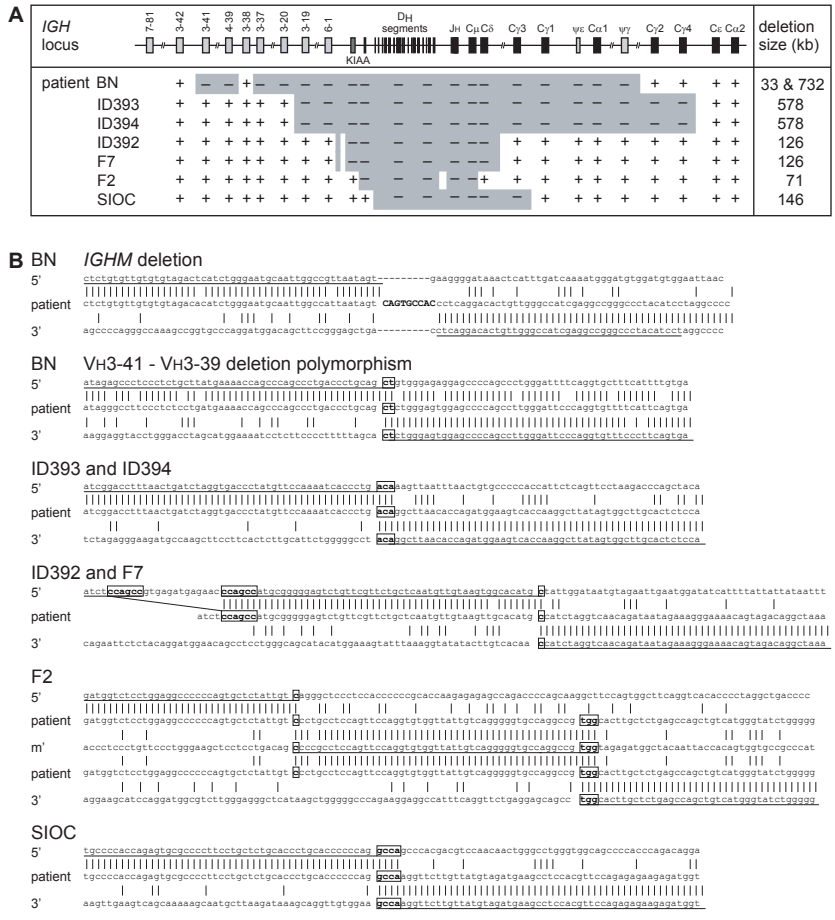


Figure 1. Gross deletion breakpoints of seven unrelated homozygous IGHM-deficient patients.
A. *IGHM* gross deletions ranging in size from 71 to 732 kb were identified using a PCR approach. An additional 33 kb deletion 1.6 kb upstream of the *IGHM* deletion disrupts three VH gene segments in patient BN. **B.** Sequences of the five unique *IGHM*-deleting breakpoint junctions and one novel VH deletion polymorphism. The *IGHM*-deletion breakpoint junction of patient BN shows a 9-nucleotide insertion (caps). The two mismatches with the 5' breakpoint region are previously described SNPs. The 33 kb VH-deletion polymorphism in patient BN shows a dinucleotide microhomology (boxed). The *IGHM*-deletion in patients ID393 and ID394 shows a trinucleotide microhomology. The *IGHM*-deletion in patients ID392 and F7 shows one nucleotide microhomology. The small deletion and the mismatch with the 5' breakpoint region are not previously described polymorphisms. The *IGHM* deletion in patient F2 shows a 55-nucleotide insertion that correspond to the intronic region in-between DH4-23 and DH5-24. The 5' junctions showed one nucleotide microhomology, whereas the 3' junction showed three nucleotides of microhomology. The mismatch within the 55 nt insertion is not a previously described SNP. Complexity analysis of ± 25 bp flanking the breakpoints showed that the deletions in patients ID393, ID394 and SIOC are potentially mediated by a hairpin loop (inverted repeats). The deletion in patients ID392 and F7 could have been mediated by a knot structure. Furthermore, the deletion in patient F2 is potentially mediated by combination of a knot structure for the 5' joint and a Möbius loop for the 3' joint.

indicating that no preferential well-established recombination sites were targeted.
No common sequence motifs were found in the breakpoint regions, but complexity analysis of the breakpoint regions showed the likely involvement of repeated fragments (Figure 1B).

IGH breakpoints are associate with TEs

The five identified *IGH* breakpoints were located in intronic sequences without any apparent homology between the 5' and 3' breakpoint flanking regions. In order to study the involvement of TEs, 2 kb DNA sequences flanking the breakpoint regions were screened against a reference collection of repeats. As shown in Figure 2, we found that 2 of the 10 *IGHM*-deleting breakpoints were located in or near a LINE 1 and 6 were located in or near endogenous retroviral (ERV) sequences, which belong to the class of LTR retrotransposons. The remaining 2 breakpoints were located in unique sequences. Consequently, in almost all gross deletions involving *IGHM*, the breakpoints of fusion partners are located in or near repetitive elements of the same class. Since TEs are present throughout the genome and they have been implicated in genomic rearrangements,[5, 6] the localization of gross deletions in or near TEs could be a general phenomenon.

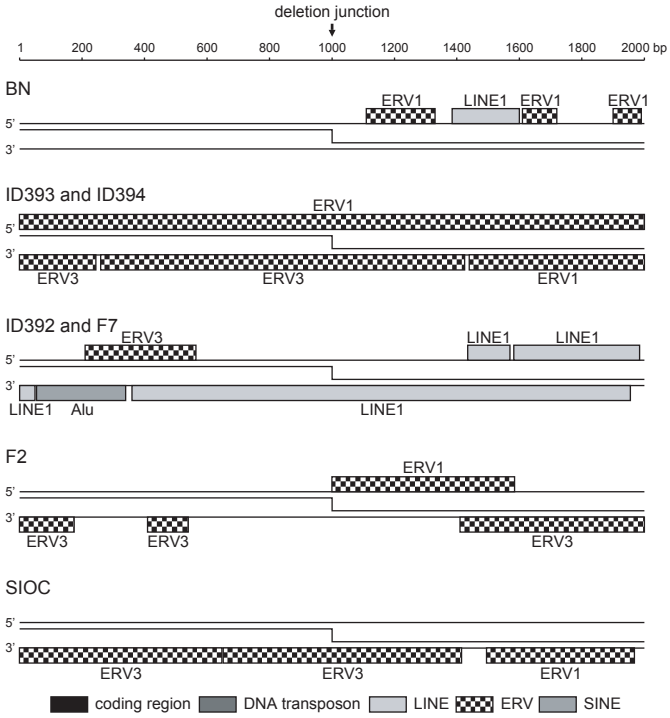


Figure 2. *IGHM*-deletion breakpoints are located in or near repetitive elements.
± 1,000 bp flanking the 10 unique breakpoints involved in *IGHM* deletions were analyzed for the presence of transposon-derived repetitive elements. No V_H, D_H, J_H gene segments or constant gene exons were present in these regions.

Similar to *IGHM*, gross deletion breakpoints in *BTK* and *Artemis* frequently occur in TEs

In addition to *IGH*, gross deletions disrupting *BTK* or *Artemis* have been found in primary immunodeficiencies (PID). Although only 5 of 37 identified breakpoints have been analyzed at sequence level, similar to *IGHM* gross deletions, the majority of the *BTK* breakpoints was located in repetitive elements.[30, 31]

In addition to the published sequences, we identified three additional gross deletions in BTK-deficient patients. The eight deletions are spread throughout the gene and range in size from 2.6 to 38.2 kb deleting one or more exons (Figure 3A). The results of the breakpoint sequence analyses of our three patients were in line with previous observations (Figure 3B and C). In 6 of the 8 patients, 5' and 3' breakpoints were located in *Alu* repeats and showed high degrees of sequence homology (Figure 3B and refs. 30, 31). Interestingly, 2 *Alu* motifs were found to be involved in two independent recombination events. In both cases, the breaks occurred at distinct sites within the *Alu* motifs, although the 3' breakpoint of patient P4 is located closely to the 3' breakpoint of patient ID434 (Figure 3B).

The gross deletion breakpoints in patients 0703 and ID440 were not located in *Alu* elements and did not show sequence homology within 50 bp. Patient 0703 had 1 nucleotide of microhomology at the junction,[30] whereas 2 nucleotides were inserted in the junction of patient ID440. The 5' breakpoint of patient 0703 was located near a MIRb (SINE2) sequence, whereas the 3' breakpoint was located in a LINE1 element and near an *Alu* element (Figure 3C). In patient ID440, the breakpoints were also located closely to SINE elements; the 5' breakpoint downstream of a MIRb and the 3' breakpoint downstream of an *Alu* repeat.

In general, the gross deletions in *BTK* were much smaller than those characterized in the *IGH* locus. This was especially noted for the *Alu – Alu* recombination events. Detailed TE annotation of breakpoint flanking regions (Figure 3C) showed that all eight gross deletion breakpoints in *BTK* were located in or near TEs.

In contrast to *BTK*, *Artemis* gross deletion breakpoints have not been sequenced to date. We identified two unique breakpoint regions in five homozygous Artemis-deficient patients (Figure 4A). Patient ID124 carried a deletion of ~82 kb including exons 1, 2 and 3 of the *Artemis* gene. A homozygous deletion of ~82 kb including exons 1, 2 and 3 of *Artemis* was found in patients ID020, ID024 and ID176. Patient ID389 had an 11.1 kb genomic deletion including *Artemis* exons 10, 11 and 12. The 5' and 3' breakpoints were located in *Alu* sequences and the joint showed high degrees of sequence homology (Figure 4B and C).

High total TE content associated with increased incidence of gene-disrupting gross deletions

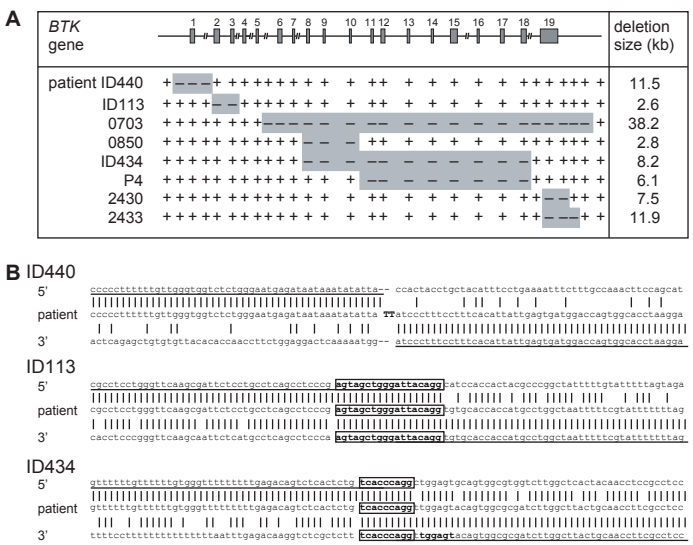
Although the incidence of gross deletions in *IGHM*, *BTK* and *Artemis* differs greatly (58 %, 6 % and 46 % of disease-causing mutations, respectively), the majority of gross deletions in all three genes show involvement of TEs. To address this difference, we determined the total TE content in these genes and compared this TE content to that of the total human genome (Figure 5A). We found that the TE content of *IGH* (41 %) and *Artemis* (45 %) was higher than the average of the human genome (37 %). In contrast, only 29 % of the *BTK* gene consists of TE-derived repeats. Similar analysis was performed for *BLNK*, which is also a PID gene of >50 kb in size. Only three disease-causing mutations in *BLNK* have been

identified, which were all micro-lesions. Similar to *BTK*, a reduced fraction of the *BLNK* gene consisted of TEs (Figure 5A; 30 %). These observations indicate that a high total TE content of a gene is associated with an increased gross deletion frequency found in gene disruptions underlying human disease.

Of the four types of TEs that are most frequently found in the human genome (LINE1, ERV, *Alu*, MIR), the ERV and LINE 1 content is over-represented in *IGH* compared to the average in the human genome (Figure 5B). In contrast, *BTK* and *Artemis* have very few of these TE elements, whereas twice as much of their genomic sequence is derived from *Alu* motifs as compared to the average in the human genome. None of the four dominant types of TEs was overrepresented in *BLNK*. Whereas the total TE content is correlated to the gross deletion frequency, gross deletions involve mainly TEs belonging to a class that is overrepresented in a gene, irrespective of the total TE content.

TEs and gross deletions in human disease genes

We demonstrated an association between the TE content and the gross deletion frequency of four genes implicated in primary immunodeficiencies. This is a small set of genes and with the exception of *BTK*, only a limited number of disease-causing mutations have been identified in these genes: 12 *IGHM* mutations; 13 *Artemis* mutations; 3 *BLNK* mutations (Table 1). Therefore, we decided to test our model in a larger set of extensively studied human disease genes. A literature search yielded a set of 16 additional genes. The gross deletion frequency of these genes in human disease was extracted from the HGMD,[1] and the TE content was determined with CENSOR.[27] Similar to what was observed for the initial four genes, the genes with a higher TE content tended to show an increased frequency of gross deletions (Table 1). A strong increase in the gross deletion frequency was noted in genes that showed >40 % TE content (Table 1). The median gross deletion frequency was 6 % for genes with ≤40 % repetitive elements, compared with 14 % for genes with >40 % repetitive elements



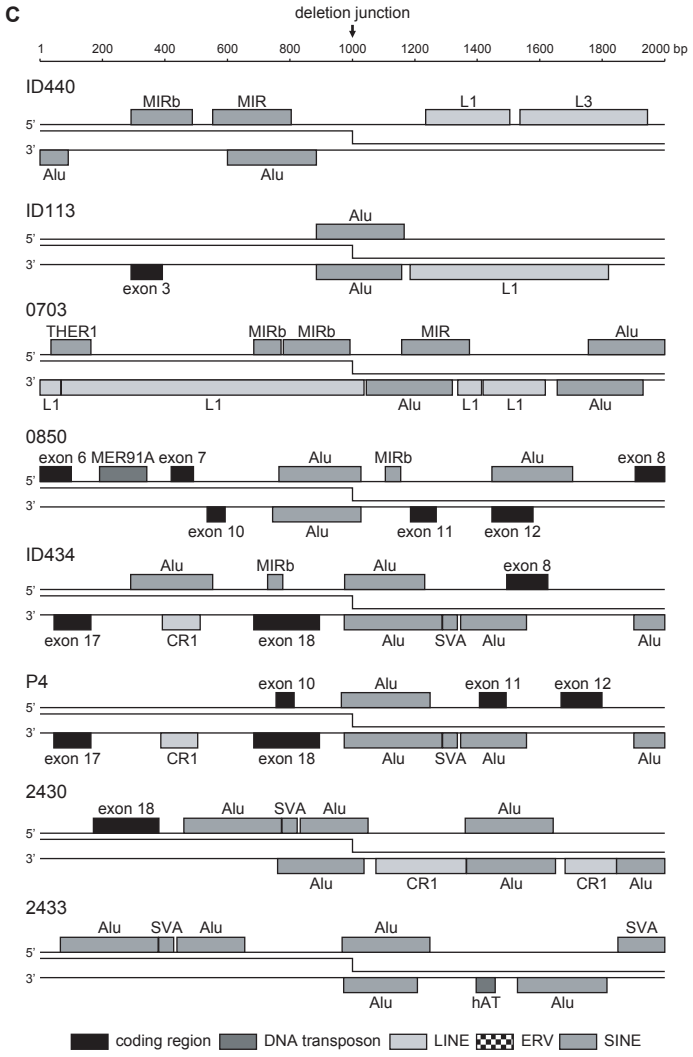


Figure 3. Gross deletions in *BTK* are mainly associated with *Alu* – *Alu* mispairing and crossing over.
A. (facing page) Mapping of the *BTK* gross deletions in three unrelated patients using a PCR approach. The gross deletions in patients 0703, 0850, 2430, 2433 [30] and P4 [31] have been described before. **B.** (facing page) Sequences of the three newly identified breakpoint junctions compared to control sequences. The breakpoint junction of patient ID440 shows a TT insertion, which could be palindromic nucleotides. Microhomology regions at the junctions of patients ID113 and ID434 are boxed. Complexity analysis of ± 25 bp flanking the breakpoints did not result in a likely mediator for patient ID440, whereas the breakpoint in patients ID113, 0850, ID434, P4, 2430 and 2433 appear to be mediated by homologous recombination and in patient 0703 by a knot structure (inversion of inverted repeats). The 3' breakpoint of patient P4 was located close to the 3' breakpoint of patient ID434 and the 6 nucleotides of microhomology are indicated by bold font in the 3' breakpoint region of ID434. **C.** ± 1,000 bp flanking the 16 unique *BTK* gross deletion breakpoints were analyzed for transposon-derived repetitive elements. In addition, *BTK* exons are indicated. The 3' breakpoints of patients ID434 and P4 were located in the same *Alu* element, and the 3' breakpoint of patient 0850 and the 5' breakpoint of patient P4 were located in the same *Alu* element.

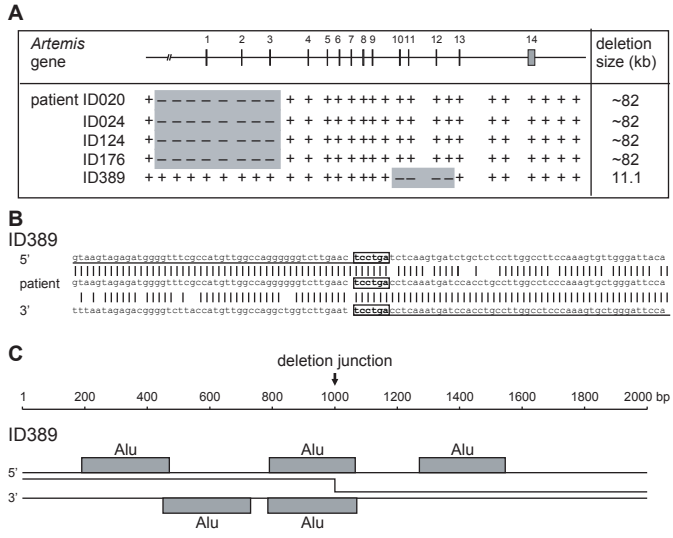


Figure 4. Gross deletions in the *Artemis* gene suggest *Alu* – *Alu* mispairing and crossing over.
A. Mapping of the homozygous *Artemis* gross deletions in five unrelated patients using a PCR approach. **B.** The deletion in patient ID389 is potentially mediated by homologous recombination. **C.** The deletion breakpoints of patient ID389 are located in *Alu* elements.

(Figure 6A). This difference was found to be highly significant between both groups of genes ($p = 0.006$).

Gross deletions are currently defined as > 20 bp, which is a crude cut-off value. TEs in the human genome range in size from about 300 bp to several kb. Consequently, this would imply that only gross deletions spanning more than a few hundred bp can show involvement of two independent TEs in a gross deletion. To test whether this holds true, additional analyses were performed with cut-off values for gross deletions of 200 and 400 bp. These analyses mainly showed a remarkable reduction in gross deletion frequency of *MECP2* from 34 % to 13 % with a cut-off of 400 bp. No difference was found for the median values and the significance for the 200 bp cut-off, whereas the medians were 5 % and 13 % for genes with ≤40 % and >40 % sequence derived from TEs, respectively, with $p = 0.005$ (Figure 6B). In conclusion, genes with a TE content of >40 % have a significantly increased gross deletion frequency in mutations underlying human disease. The significance increased slightly when the cut-off for gross deletions is increased to 400 bp.

In addition to the total TE content, the content of the four most abundant TE subclasses (ERV, LINE1, *Alu* and MIR) was determined for the 20 genes to study whether the overrepresented elements were specifically involved in gross deletions (Table 1).

Although the breakpoint fusion sequence of only a small fraction of the reported gross deletions has been determined, mainly the overrepresented TE subclasses appeared to be involved in gross deletion breakpoints (Table 1).

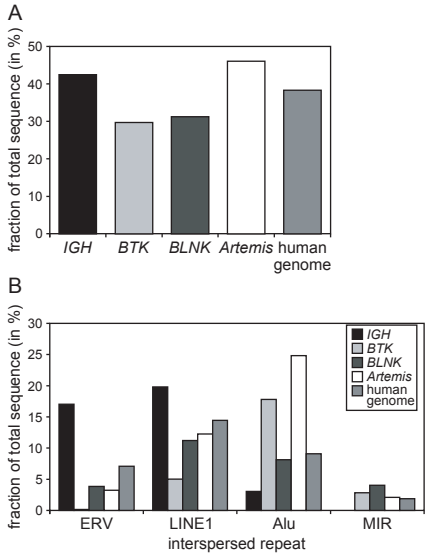


Figure 5. Highly variable TE content in four genes.
A. The *IGH* and *Artemis* loci contain a high TE content compared to the average of the human genome, whereas the *BTK* and *BLNK* genes have decreased TE content. **B.** The LINE, ERV, *Alu* and MIR elements are the most common TEs found in the human genome. *IGH* contains an increased ERV and LINE content, whereas *BTK* and *Artemis* mainly contain *Alu* elements. *BLNK* does not display increased presence of any of the four types of elements as compared to the average of the human genome.

DISCUSSION

We characterized the gross deletion breakpoints in IGHM-, BTK-, and Artemis-deficient patients to explain the high frequency of gross deletions in *IGHM* and *Artemis* compared to *BTK*. *IGHM* gross deletion breakpoints were not located in or near RSS or Ig switch regions, but similar to gross deletions involving *BTK* or *Artemis*, they were associated with TE-derived repeats. A model based on differential TE involvement and gross deletion frequencies of these genes was tested on a total set of 20 genes disrupted in human disease. We found that gross deletion breakpoints in each gene were frequently located in TE types that were specifically overrepresented in that gene as compared to the average in the human genome. Furthermore, we found that genes with a high TE content (>40 %) showed an increased frequency of gross deletions when disrupted in human disease. We conclude that, similar to suggestions based on limited sequence data from other genes disrupted in human disease, repetitive elements are likely mediators of gross deletions in *IGHM*, *BTK* and *Artemis*.

In this study, five unique breakpoint regions were characterized in seven unrelated *IGHM*-deficient patients. The breakpoints were scattered throughout the locus in introns within the VH, DH and CH regions, thus indicating that there was no preferential targeting of a specific region. Interestingly, no co-localization with RSS or Ig switch regions was found.

Table 1. Correlation gross deletions incidence and TE content.

gene	Size (kb)	gross deletion frequency [a] # gross deletions / # total mutations	motif fraction [b] % total sequence	ERV [c] % total sequence	LINE1 [c] % total sequence	Alu [c] % total sequence	MIR [c] % total sequence	breakpoint mediators	references
IGHM [d]	1280	**58** % 7 / 12	**41** %	16.1	**19.1**	2.8	0.2	ERV, LINE1	
BTK	56	7 % 37 / 523	29 %	0.1	4.6	18.0	**2.9**	*Alu*, MIR	30, 31
BLNK [e]	100	0 % 0 / 3	30 %	3.8	10.9	7.9	4.2	–	
Artemis [d]	66	**46** % 6 / 13	**45** %	3.1	11.3	23.3	**4.0**	*Alu*	
CFTR	209	3 % 35 / 1208	27 %	5.2	9.8	4.1	3.0	?	
DMD	2112	**35** % 255 / 735	29 %	5.3	11.8	5.3	1.9	*Alu*, LINE1	5
MECP2	96	**34 %** 108 / 319	31 %	1.7	8.8	19.4	1.0	*Alu*	37, 38
APC	128	5 % 34 / 696	32 %	3.5	7.9	13.3	**2.2**	*Alu*	39, 40
VWF	196	6 % 12 / 206	33 %	3.3	12.1	10.9	**2.7**	*Alu*	41
MLH1	77	**11** % 46 / 415	36 %	2.4	8.2	16.5	2.3	*Alu*, LINE1	42
NF1	299	6 % 45 / 718	37 %	4.6	13.3	13.8	1.5	*Alu*	43
BRCA2	104	1 % 7 / 504	39 %	4.2	8.3	13.8	2.0	*Alu*	44
BRCA1	101	6 % 45 / 750	40 %	2.1	1.7	31.5	**2.9**	*Alu*	45
ATM	166	8 % 35 / 421	40 %	1.3	20.3	**11.7**	1.3	*Alu*, LINE1	46
FANCA	99	**28** % 66 / 234	**42** %	4.3	3.7	30.4	**2.6**	*Alu*	47, 48
HPRT1	60	**12 %** 28 / 232	43 %	3.8	6.4	21.9	**4.2**	*Alu*, LINE1	49
MSH2	100	**16** % 56 / 346	**45** %	4.1	5.9	26.2	**5.7**	*Alu*	42, 50
RB1	198	**8 %** 35 / 418	47 %	4.1	26.4	6.9	1.7	?	
LDLR	64	**11** % 94 / 866	**51** %	2.0	1.9	41.0	**4.5**	*Alu*	51-55
Factor VIII	207	**11** % 95 / 885	**58** %	10.7	**32.3**	7.0	1.0	*Alu*, LINE1	56-60
Total	3.2 x 10^6 [f]	6 % 2962 / 53200	37 % [f]	6.8 [f]	14.6 [f]	9.3 [f]	1.8 [f]	–	

[a] HGMD February 2007
[b] TE content was determined with CENSOR
[c] The four most abundant TE subclasses; ERVs are LTR retrotransposons, LINE1s are the most common LINEs, and Alu and MIR are the most common SINEs
[d] Statistics include patients that are described here
[e] Statistics include unpublished patients (M.v.d.B.)
[f] Data obtained on the total human genome sequence

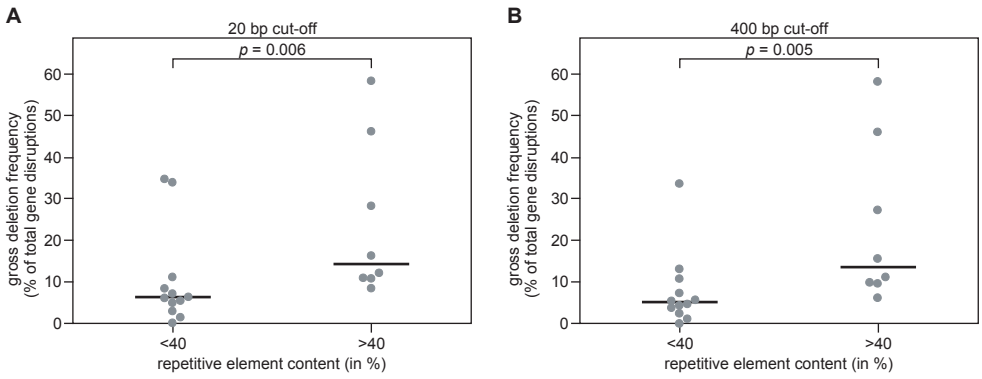


Figure 6. Genes disrupted in human inherited diseases with a transposon-derived repetitive element content of > 40 % are significantly more frequently disrupted by gross deletions than genes with ≤40 % TE content. The frequency of gross deletions was determined from 20 genes disrupted in human inherited disease. A surprising increase in gross deletion frequency was found in genes with >40 % TE content. **A.** The gross deletion frequency of 20 genes is shown for genes with ≤40 % and >40 % TE content. Every dot represents an individual gene. The gross deletion frequency was significantly increased in genes with >40 % TE content as determined with a non-parametric test. **B.** The significance increases slightly when the cut-off set for gross deletions is increased from >20 bp to >400 bp.

Whereas such sequences are potent recombination sequences in specific stages during B-cell development, they do not appear to be recombination-prone in germline cells. It is therefore highly unlikely that the B-cell specific recombination characteristics of the *IGH* locus underlie the high frequency of gross deletions involving *IGHM*.

The five *IGHM* deletions did not show high sequence homology between the 5' and the 3' breakpoint regions. Therefore, it is unlikely that the gross deletions were the result of mispairing of homologous sequences and unequal crossing-over as has been observed in recurrent Ig subclass deletions.[22] The five breakpoint junctions did show the characteristics of NHEJ with microhomology in four joints and an insertion of six nucleotides in the fifth. No common sequence motifs were found in the breakpoint regions, whereas complexity analysis revealed that short repetitive sequences are likely mediators of the unequal repair.

In addition to the disease-causing *IGHM* deletion in patient BN, a 33 kb deletion of three VH gene segments was found. Because 17 % of the alleles in healthy controls carried the exact same 33 kb deletion, but not the *IGHM* deletion, we conclude that this is a common polymorphism. This polymorphism likely arose by homologous unequal recombination prior to and independently from the *IGHM* deletion. Since the polymorphism is located quite distant from the 5' *IGHM* deletion breakpoint (1.6 kb), it is unlikely that this polymorphism made the *IGH* allele more susceptible to the deleterious deletion.

The gross deletion junction of patients ID392 and F7 and of patient F2 also revealed that two deletions had disrupted the *IGH* locus (Figure 1). In both cases, the 3' deletion disrupts *IGHM* and therefore underlies the disease. The 5' micro-deletion in patients ID392 and F7 and the 5' gross deletion in patient F2 were located in close proximity to the 5' gross deletions (54 and 51 bp, respectively) and not found in controls. Although the 5' deletion restricts

the DH repertoire by removing 23 of the 27 DH gene segments, the remaining four DH gene segments should be sufficient for B-cell development and immune function.[32] Because deleterious *IGHM* mutations are very rare, it is unlikely that the double deletions are the result of two independent hits. Furthermore, similar observations were made in a *TCF2* gross deletion in a patient with Maturity-onset diabetes of the young type 5.[33] A 55-bp insertion corresponding to part of the deleted DNA sequence was inserted in the inverse orientation. These observations led the authors to conclude that the deletion was the result of a single event. Considering the rarity of gross deletions and the similarities with the *TCF2* deletion, we propose that the deletions in patients ID392 and F7 and in patient F2 arose in a single event during which two regions were looped out and incorrectly repaired, thus resulting in the complicated junction we observed (Figure 1).

The majority of *IGHM*-deleting gross deletion breakpoints were located in or near TE-derived repeats. Interestingly, the involved elements (LINEs and LTR retrotransposons) were overrepresented in the *IGH* locus as compared to the average in the human genome. On the other hand, the gross deletions identified in BTK- and Artemis-deficient patients mainly involved SINEs. In contrast to *IGH*, these SINEs are specifically overrepresented in *BTK* and *Artemis*. Furthermore, the high frequency of gross deletions in *IGH* and *Artemis* as compared to *BTK* and *BLNK* was associated with a high total TE content. These observations suggest a role for TEs in mediating gross deletions. Involvement of TEs has been suggested before.[5, 6] However, by analysis of 20 genes disrupted in human disease, we are the first to show that the total TE content is associated with a significantly increased frequency of gross deletions as cause of gene disruption.

The role of TEs in mediating gross deletions is difficult to examine. In general, the breakpoints analyzed in this study involved at least two double-stranded DNA breaks in distant genomic locations (2.6 – 732 kb) that were placed in physical proximity and incorrectly repaired. Complexity analysis showed that small repeat fragments are likely mediators of the incorrect repair. This mechanism is quite similar to what is seen in micro-lesions, which are thought to result from slipped mispairing of single-stranded DNA in the replication fork. Since the stretches of single-stranded DNA in the replication fork do not extend beyond 1,000-2,000 bp, additional factors are required to mediate co-localization of two distant genomic regions and double-stranded DNA break induction.

Studies in *Drosophila* have demonstrated the accumulation of transposon-derived repetitive elements in heterochromatin and this appears to be widespread among higher eukaryotes (reviewed by Dimitri *et al.*[34]). Since genes packed in heterochromatin are commonly repressed, the accumulation of TEs in heterochromatin could result from the preferential insertion of these elements in DNA sequences packed in heterochromatin, or from a mechanism that recruits DNA with high TE content to heterochromatin. Although the latter mechanism was observed recently in *Arabidopsis*,[35] this does not exclude the other mechanism. Heterochromatic regions are compact clusters of genomic DNA. It seems therefore likely that genomic regions with high TE content are tightly packed in heterochromatin, and this is the mechanism by which TEs mediate co-localization of distant genomic regions.

Micro-insertions and micro-deletions are assumed to result from slipped mispairing, but this mechanism has been proposed to underlie some of the identified deletions that were >20 bp.[36] This is not remarkable as it is suggested to result from mispairing of ssDNA in

the replication fork and the length of ssDNA is 1,000-2,000 bp. However, TEs are a few hundred or a few thousand bp in size. Therefore, the involvement of TEs in the generation of deletions smaller than a few hundred bp is unlikely. We suggest defining gross deletions to be larger than 400 or even 1,000 bp because this better reflects the mechanism underlying these lesions. Ultimately, additional sequence data and analysis of deletion breakpoints should allow for a more accurate discrimination of micro- and gross deletions based on the underlying mechanism.

In this study, we identified and characterized *IGHM*, *BTK* and *Artemis* deletions in patients with primary immunodeficiency. Analysis of these deletion junctions formed the basis of our model, which was confirmed by analysis of 20 genes disrupted in human disease. This model states that: (a) Gross deletions are mediated by TE subclasses that are specifically overrepresented in the involved gene as compared to the average in the human genome; and (b) a high total TE content results in an increased frequency of gene disruption by gross deletions. Whereas more sequence data are needed to draw firm conclusions on the underlying mechanisms, this study shows for the first time the impact of TEs on the gross deletion frequency as gene disruption in human disease.

ACKNOWLEDGEMENTS

The authors acknowledge Mrs. W.M. Comans-Bitter for assistance with preparing the figures. This work was supported by a grant from the foundation “Sophia Kinderziekenhuis Fonds” (SKF; grant 349 to M.C. van Zelm, Dr. R. de Groot, and Dr. J.J.M. van Dongen).

DISCLOSURES

The authors state that they have no competing financial interests.

REFERENCES

1. Stenson PD, Ball EV, Mort M, Phillips AD, Shiel JA, Thomas NS, Abeysinghe S, Krawczak M, Cooper DN. Human Gene Mutation Database (HGMD): 2003 update. *Hum Mutat* 2003;**21**:577-81.
2. Abeysinghe SS, Chuzhanova N, Krawczak M, Ball EV, Cooper DN. Translocation and gross deletion breakpoints in human inherited disease and cancer I: Nucleotide composition and recombination-associated motifs. *Hum Mutat* 2003;**22**:229-44.
3. Chuzhanova N, Abeysinghe SS, Krawczak M, Cooper DN. Translocation and gross deletion breakpoints in human inherited disease and cancer II: Potential involvement of repetitive sequence elements in secondary structure formation between DNA ends. *Hum Mutat* 2003;**22**:245-51.
4. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL,

Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blocker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ. Initial sequencing and analysis of the human genome. *Nature* 2001;**409**:860-921.

5. McNaughton JC, Cockburn DJ, Hughes G, Jones WA, Laing NG, Ray PN, Stockwell PA, Petersen GB. Is gene deletion in eukaryotes sequence-dependent? A study of nine deletion junctions and nineteen other deletion breakpoints in intron 7 of the human dystrophin gene. *Gene* 1998;**222**:41-51.
6. Hedges DJ, Deininger PL. Inviting instability: Transposable elements, double-strand breaks, and the maintenance of genome integrity. *Mutat Res* 2007;**616**:46-59.
7. Notarangelo L, Casanova JL, Conley ME, Chapel H, Fischer A, Puck J, Roifman C, Seger R, Geha RS. Primary immunodeficiency diseases: an update from the International Union of Immunological Societies Primary Immunodeficiency Diseases Classification Committee Meeting in Budapest, 2005. *J Allergy Clin Immunol* 2006;**117**:883-96.
8. Moshous D, Callebaut I, de Chasseval R, Corneo B, Cavazzana-Calvo M, Le Deist F, Tezcan I, Sanal O, Bertrand Y, Philippe N, Fischer A, de Villartay JP. Artemis, a novel DNA double-strand break repair/V(D)J recombination protein, is mutated in human severe combined immune deficiency. *Cell* 2001;**105**:177-86.
9. Noordzij JG, Verkaik NS, van der Burg M, van Veelen LR, de Bruin-Versteeg S, Wiegant W, Vossen JM, Weemaes CM, de Groot R, Zdzienicka MZ, van Gent DC, van Dongen JJM. Radiosensitive SCID patients with Artemis gene mutations show a complete B-cell differentiation arrest at the pre-B-cell receptor checkpoint in bone marrow. *Blood* 2003;**101**:1446-52.
10. Conley ME, Mathias D, Treadaway J, Minegishi Y, Rohrer J. Mutations in btk in patients with presumed X-linked agammaglobulinemia. *Am J Hum Genet* 1998;**62**:1034-43.
11. Yel L, Minegishi Y, Coustan-Smith E, Buckley RH, Trubel H, Pachman LM, Kitchingman GR, Campana D, Rohrer J, Conley ME. Mutations in the mu heavy-chain gene in patients with agammaglobulinemia. *N Engl J Med* 1996;**335**:1486-93.
12. Milili M, Antunes H, Blanco-Betancourt C, Nogueiras A, Santos E, Vasconcelos J, Castro e Melo J, Schiff C. A new case of autosomal recessive agammaglobulinaemia with impaired pre-B cell differentiation due to a large deletion of the IGH locus. *Eur J Pediatr* 2002;**161**:479-84.
13. Lopez Granados E, Porpiglia AS, Hogan MB, Matamoros N, Krasovec S, Pignata C, Smith CI, Hammarstrom L, Bjorkander J, Belohradsky BH, Casariego GF, Garcia Rodriguez MC, Conley ME. Clinical and molecular analysis of patients with defects in micro heavy chain gene. *J Clin Invest* 2002;**110**:1029-35.
14. Schatz DG. V(D)J recombination. *Immunol Rev* 2004;**200**:5-11.
15. Tonegawa S. Somatic generation of antibody diversity. *Nature* 1983;**302**:575-81.
16. Odegard VH, Schatz DG. Targeting of somatic hypermutation. *Nat Rev Immunol* 2006;**6**:573-83.
17. Bech-Hansen NT, Cox DW. Duplication of the human immunoglobulin heavy chain gamma 2 gene. *Am J Hum Genet* 1986;**38**:67-74.
18. Bottaro A, Cariota U, DeMarchi M, Carbonara AO. Pulsed-field electrophoresis screening for immunoglobulin heavy-chain constant-region (IGHC) multigene deletions and duplications. *Am J Hum Genet* 1991;**48**:745-56.

19. Brusco A, Cariota U, Bottaro A, Boccazzi C, Plebani A, Ugazio AG, Galanello R, van Leeuwen AM, DeLange GG, Depelchin S, et al. Structural and immunologic analysis of gene triplications in the Ig heavy chain constant region locus. *J Immunol* 1994;**152**:129-35.
20. Brusco A, Cariota U, Bottaro A, Boccazzi C, Plebani A, Ugazio AG, Galanello R, Guerra MG, Carbonara AO. Variability of the immunoglobulin heavy chain constant region locus: a population study. *Hum Genet* 1995;**95**:319-26.
21. Rabbani H, Pan Q, Kondo N, Smith CI, Hammarstrom L. Duplications and deletions of the human IGHC locus: evolutionary implications. *Immunogenetics* 1996;**45**:136-41.
22. Brusco A, Saviozzi S, Cinque F, Bottaro A, DeMarchi M. A recurrent breakpoint in the most common deletion of the Ig heavy chain locus (del A1-GP-G2-G4-E). *J Immunol* 1999;**163**:4392-8.
23. van Dongen JJM, Langerak AW, Bruggemann M, Evans PA, Hummel M, Lavender FL, Delabesse E, Davi F, Schuuring E, Garcia-Sanz R, Van Krieken JH, Droese J, Gonzalez D, Bastard C, White HE, Spaargaren M, Gonzalez M, Parreira A, Smith JL, Morgan GJ, Kneba M, Macintyre EA. Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: Report of the BIOMED-2 Concerted Action BMH4-CT98-3936. *Leukemia* 2003;**17**: 2257-317.
24. Noordzij JG, de Bruin-Versteeg S, Comans-Bitter WM, Hartwig NG, Hendriks RW, de Groot R, van Dongen JJM. Composition of precursor B-cell compartment in bone marrow from patients with X-linked agammaglobulinemia compared with healthy children. *Pediatr Res* 2002;**51**:159-68.
25. Przybylski GK, Dik WA, Wanzeck J, Grabarczyk P, Majunke S, Martin-Subero JI, Siebert R, Dolken G, Ludwig WD, Verhaaf B, van Dongen JJM, Schmidt CA, Langerak AW. Disruption of the BCL11B gene through inv(14)(q11.2q32.31) results in the expression of BCL11B-TRDC fusion transcripts and is associated with the absence of wild-type BCL11B transcripts in T-ALL. *Leukemia* 2005;**19**:201-8.
26. Hubbard TJ, Aken BL, Beal K, Ballester B, Caccamo M, Chen Y, Clarke L, Coates G, Cunningham F, Cutts T, Down T, Dyer SC, Fitzgerald S, Fernandez-Banet J, Graf S, Haider S, Hammond M, Herrero J, Holland R, Howe K, Johnson N, Kahari A, Keefe D, Kokocinski F, Kulesha E, Lawson D, Longden I, Melsopp C, Megy K, Meidl P, Ouverdin B, Parker A, Prlic A, Rice S, Rios D, Schuster M, Sealy I, Severin J, Slater G, Smedley D, Spudich G, Trevanion S, Vilella A, Vogel J, White S, Wood M, Cox T, Curwen V, Durbin R, Fernandez-Suarez XM, Flicek P, Kasprzyk A, Proctor G, Searle S, Smith J, Ureta-Vidal A, Birney E. Ensembl 2007. *Nucleic Acids Res* 2007;**35**: D610-7.
27. Kohany O, Gentles AJ, Hankus L, Jurka J. Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. *BMC Bioinformatics* 2006;**7**:474.
28. Staats B, Qi L, Beerman M, Sicotte H, Burdett LA, Packer B, Chanock SJ, Yeager M. Genewindow: an interactive tool for visualization of genomic variation. *Nat Genet* 2005;**37**:109-10.
29. Gusev VD, Nemytikova LA, Chuzhanova NA. On the complexity measures of genetic sequences. *Bioinformatics* 1999;**15**:994-9.
30. Rohrer J, Minegishi Y, Richter D, Eguiguren J, Conley ME. Unusual mutations in Btk: an insertion, a duplication, an inversion, and four large deletions. *Clin Immunol* 1999;**90**:28-37.
31. Jo EK, Wang Y, Kanegane H, Futatani T, Song CH, Park JK, Kim JS, Kim DS, Ahn KM, Lee SI, Park HJ, Hahn YS, Lee JH, Miyawaki T. Identification of mutations in the Bruton's tyrosine kinase gene, including a novel genomic rearrangements resulting in large deletion, in Korean X-linked agammaglobulinemia patients. *J Hum Genet* 2003;**48**:322-6.
32. Schelonka RL, Ivanov, II, Jung DH, Ippolito GC, Nitschke L, Zhuang Y, Gartland GL, Pelkonen J, Alt FW, Rajewsky K, Schroeder HW, Jr. A single DH gene segment creates its own unique CDR-H3 repertoire and is sufficient for B cell development and immune function. *J Immunol* 2005;**175**:6624-32.
33. Bellanne-Chantelot C, Clauin S, Chauveau D, Collin P, Daumont M, Douillard C, Dubois-Laforgue D, Dusselier L, Gautier JF, Jadoul M, Laloi-Michelin M, Jacquesson L, Larger E, Louis J, Nicolino M, Subra JF, Wilhem JM, Young J, Velho G, Timsit J. Large genomic rearrangements in the hepatocyte nuclear factor-1beta (TCF2) gene are the most frequent cause of maturity-onset diabetes of the young type 5. *Diabetes* 2005;**54**:3126-32.
34. Dimitri P, Corradini N, Rossi F, Mei E, Zhimulev IF, Verni F. Transposable elements as artisans of the heterochromatic genome in Drosophila melanogaster. *Cytogenet Genome Res* 2005;**110**:165-72.

35. Lippman Z, Gendrel AV, Black M, Vaughn MW, Dedhia N, McCombie WR, Lavine K, Mittal V, May B, Kasschau KD, Carrington JC, Doerge RW, Colot V, Martienssen R. Role of transposable elements in heterochromatin and epigenetic control. *Nature* 2004;**430**:471-6.
36. Chen JM, Chuzhanova N, Stenson PD, Ferec C, Cooper DN. Meta-analysis of gross insertions causing human genetic disease: novel mutational mechanisms and the role of replication slippage. *Hum Mutat* 2005;**25**:207-21.
37. Laccone F, Junemann I, Whatley S, Morgan R, Butler R, Huppke P, Ravine D. Large deletions of the MECP2 gene detected by gene dosage analysis in patients with Rett syndrome. *Hum Mutat* 2004;**23**:234-44.
38. Schollen E, Smeets E, Deflem E, Fryns JP, Matthijs G. Gross rearrangements in the MECP2 gene in three patients with Rett syndrome: implications for routine diagnosis of Rett syndrome. *Hum Mutat* 2003;**22**:116-20.
39. Su LK, Steinbach G, Sawyer JC, Hindi M, Ward PA, Lynch PM. Genomic rearrangements of the APC tumor-suppressor gene in familial adenomatous polyposis. *Hum Genet* 2000;**106**:101-7.
40. Cao X, Eu KW, Seow-Choen F, Zhao Y, Cheah PY. Topoisomerase-I- and Alu-mediated genomic deletions of the APC gene in familial adenomatous polyposis. *Hum Genet* 2001;**108**:436-42.
41. Xie F, Wang X, Cooper DN, Chuzhanova N, Fang Y, Cai X, Wang Z, Wang H. A novel Alu-mediated 61-kb deletion of the von Willebrand factor (VWF) gene whose breakpoints co-locate with putative matrix attachment regions. *Blood Cells Mol Dis* 2006;**36**:385-91.
42. van der Klift H, Wijnen J, Wagner A, Verkuilen P, Tops C, Otway R, Kohonen-Corish M, Vasen H, Oliani C, Barana D, Moller P, Delozier-Blanchet C, Hutter P, Foulkes W, Lynch H, Burn J, Moslein G, Fodde R. Molecular characterization of the spectrum of genomic deletions in the mismatch repair genes MSH2, MLH1, MSH6, and PMS2 responsible for hereditary nonpolyposis colorectal cancer (HNPCC). *Genes Chromosomes Cancer* 2005;**44**:123-38.
43. Gervasini C, Venturin M, Orzan F, Friso A, Clementi M, Tenconi R, Larizza L, Riva P. Uncommon Alu-mediated NF1 microdeletion with a breakpoint inside the NF1 gene. *Genomics* 2005;**85**:273-9.
44. Casilli F, Tournier I, Sinilnikova OM, Coulet F, Soubrier F, Houdayer C, Hardouin A, Berthet P, Sobol H, Bourdon V, Muller D, Fricker JP, Capoulade-Metay C, Chompret A, Nogues C, Mazoyer S, Chappuis P, Maillet P, Philippe C, Lortholary A, Gesta P, Bezieau S, Toulas C, Gladieff L, Maugard CM, Provencher DM, Dugast C, Delvincourt C, Nguyen TD, Faivre L, Bonadona V, Frebourg T, Lidereau R, Stoppa-Lyonnet D, Tosi M. The contribution of germline rearrangements to the spectrum of BRCA2 mutations. *J Med Genet* 2006;**43**:e49.
45. Armaou S, Konstantopoulou I, Anagnostopoulos T, Razis E, Boukovinas I, Xenidis N, Fountzilas G, Yannoukakos D. Novel genomic rearrangements in the BRCA1 gene detected in greek breast/ovarian cancer patients. *Eur J Cancer* 2007;**43**:443-53.
46. Cavalieri S, Funaro A, Porcedda P, Turinetto V, Migone N, Gatti RA, Brusco A. ATM mutations in Italian families with ataxia telangiectasia include two distinct large genomic deletions. *Hum Mutat* 2006;**27**:1061.
47. Levran O, Doggett NA, Auerbach AD. Identification of Alu-mediated deletions in the Fanconi anemia gene FAA. *Hum Mutat* 1998;**12**:145-52.
48. Centra M, Memeo E, d'Apolito M, Savino M, Ianzano L, Notarangelo A, Liu J, Doggett NA, Zelante L, Savoia A. Fine exon-intron structure of the Fanconi anemia group A (FAA) gene and characterization of two genomic deletions. *Genomics* 1998;**51**:463-7.
49. Williams M, Rainville IR, Nicklas JA. Use of inverse PCR to amplify and sequence breakpoints of HPRT deletion and translocation mutations. *Environ Mol Mutagen* 2002;**39**:22-32.
50. Li L, McVety S, Younan R, Liang P, Du Sart D, Gordon PH, Hutter P, Hogervorst FB, Chong G, Foulkes WD. Distinct patterns of germ-line deletions in MLH1 and MSH2: the implication of Alu repetitive element in the genetic etiology of Lynch syndrome (HNPCC). *Hum Mutat* 2006;**27**:388.
51. Hobbs HH, Brown MS, Goldstein JL. Molecular genetics of the LDL receptor gene in familial hypercholesterolemia. *Hum Mutat* 1992;**1**:445-66.
52. Kim SH, Bae JH, Chae JJ, Kim UK, Choe SJ, Namkoong Y, Kim HS, Park YB, Lee CC. Long-distance PCR-based screening for large rearrangements of the LDL receptor gene in Korean patients with familial hypercholesterolemia. *Clin Chem* 1999;**45**:1424-30.
53. Chae JJ, Park YB, Kim SH, Hong SS, Song GJ, Han KH, Namkoong Y, Kim HS, Lee CC. Two partial deletion mutations involving the same Alu sequence within intron 8 of the LDL receptor gene in Korean patients with familial hypercholesterolemia. *Hum Genet* 1997;**99**:155-63.

54. Peeters AV, Van Gaal LF, du Plessis L, Lombardi MP, Havekes LM, Kotze MJ. Mutational and genetic origin of LDL receptor gene mutations detected in both Belgian and Dutch familial hypercholesterolemics. *Hum Genet* 1997;**100**:266-70.
55. Simard LR, Viel J, Lambert M, Paradis G, Levy E, Delvin EE, Mitchell GA. The Delta>15 Kb deletion French Canadian founder mutation in familial hypercholesterolemia: rapid polymerase chain reaction-based diagnostic assay and prevalence in Quebec. *Clin Genet* 2004;**65**:202-8.
56. Woods-Samuels P, Kazazian HH, Jr., Antonarakis SE. Nonhomologous recombination in the human genome: deletions in the human factor VIII gene. *Genomics* 1991;**10**:94-101.
57. Vidal F, Farssac E, Tusell J, Puig L, Gallardo D. First molecular characterization of an unequal homologous alu-mediated recombination event responsible for hemophilia. *Thromb Haemost* 2002;**88**:12-6.
58. van de Water N, Williams R, Ockelford P, Browett P. A 20.7 kb deletion within the factor VIII gene associated with LINE-1 element insertion. *Thromb Haemost* 1998;**79**:938-42.
59. Nakaya SM, Hsu TC, Geraghty SJ, Manco-Johnson MJ, Thompson AR. Severe hemophilia A due to a 1.3 kb factor VIII gene deletion including exon 24: homologous recombination between 41 bp within an Alu repeat sequence in introns 23 and 24. *J Thromb Haemost* 2004;**2**:1941-5.
60. Rossetti LC, Goodeve A, Larripa IB, De Brasi CD. Homeologous recombination between AluSx-sequences as a cause of hemophilia. *Hum Mutat* 2004;**24**:440.

V

An Antibody-Deficiency Syndrome Due to Mutations in the *CD19* Gene

Menno C. van Zelm, M.Sc., Ismail Reisli, M.D., Mirjam van der Burg, Ph.D., Diana Castaño, M.Sc., Carel J.M. van Noesel, M.D., Ph.D., Maarten J.D. van Tol, Ph.D., Cristina Woellner, M.Sc., Bodo Grimbacher, M.D., Pablo J. Patiño, M.D., Ph.D., Jacques J.M. van Dongen, M.D., Ph.D., and José L. Franco, M.D., Ph.D.

From the Department of Immunology, Erasmus MC (M.C.Z., M.B., J.J.M.D.), and the Department of Pediatrics, Erasmus MC-Sophia (M.C.Z.), Rotterdam, the Netherlands; the Department of Pediatric Immunology and Allergy, Meram Medical Faculty, Selçuk University, Konya, Turkey (I.R.); the Group of Primary Immunodeficiencies, University of Antioquia, Medellín, Colombia (D.C., P.J.P., J.L.F.); the Department of Pathology, Academic Medical Center, Amsterdam (C.J.M.N.); the Department of Pediatrics, Leiden University Medical Center, Leiden, the Netherlands (M.J.D.T.); and the Division of Rheumatology and Clinical Immunology, University Hospital Freiburg, Freiburg, Germany (C.W., B.G.)

N Engl J Med 2006;354(18):1901-12

Copyright 2006 Massachusetts Medical Society

ABSTRACT

Background The CD19 protein forms a complex with CD21, CD81, and CD225 in the membrane of mature B cells. Together with the B-cell antigen receptor, this complex signals the B cell to decrease its threshold for activation by the antigen.

Methods We evaluated four patients from two unrelated families who had increased susceptibility to infection, hypogammaglobulinemia, and normal numbers of mature B cells in blood. We found a mutation in the CD19 gene in all four patients. The CD19 gene in the patients and their first-degree relatives was sequenced, and flowcytometric immunophenotyping of B cells, immunohistochemical staining of lymphoid tissues, and DNA and messenger RNA analysis were performed. B-cell responses on the triggering of the B-cell receptor were investigated by *in vitro* stimulation; the antibody response after vaccination with rabies vaccine was also studied.

Results All four patients had homozygous mutations in the CD19 gene. Levels of CD19 were undetectable in one patient and substantially decreased in the other three. Levels of CD21 were decreased, whereas levels of CD81 and CD225 were normal, in all four patients. The composition of the precursor B-cell compartment in bone marrow and the total numbers of B cells in blood were normal. However, the numbers of $CD27^+$ memory B cells and $CD5^+$ B cells were decreased. Secondary follicles in lymphoid tissues were small to normal in size and had a normal cellular composition. The few B cells that showed molecular signs of switching from one immunoglobulin class to another contained V_H-Cα and V_H-Cγ transcripts with somatic mutations. The response of the patients' B cells to *in vitro* stimulation through the B-cell receptor was impaired, and in all four patients, the antibody response to rabies vaccination was poor.

Conclusions Mutation of the CD19 gene causes a type of hypogammaglobulinemia in which the response of mature B cells to antigenic stimulation is defective.

The online version of this article contains supplemental material.

INTRODUCTION

Primary antibody deficiencies, which are associated mainly with susceptibility to bacterial infections, can be caused by mutations in genes involved in B-cell differentiation.[1] Such genetic defects block the differentiation of immature B cells in bone marrow, thereby causing both a deficiency of mature B cells in the peripheral circulation and agammaglobulinemia.[2-8]

In disorders involving defects in the maturation process of antigen-responsive B cells, by contrast, there are mature B cells in blood and a deficiency of some, but not all, immunoglobulin classes (dysgammaglobulinemia) or hypogammaglobulinemia. These disorders fall into two categories. The hyper-IgM syndromes are those in which a genetic defect impairs immunoglobulin-class switching (e.g., IgM to IgG) or impairs somatic mutation of immunoglobulin genes (a molecular sign of a response to an antigen), or both.[9-12]

The second category is common variable immunodeficiency, a syndrome involving hypogammaglobulinemia, recurrent bacterial infections, and a lack of antibody production in response to vaccination. Genetic defects have been found in some patients with common variable immunodeficiency,[13, 14] but no genetic abnormality is evident in more than 80 percent of such patients.

During B-cell differentiation in the bone marrow, the surface molecule CD19 appears early and remains on the B cell until it differentiates into a plasma cell. Four proteins on the surface of mature B cells – CD19, CD21, CD81, and CD225 – form the CD19-complex, which signals in conjunction with the B-cell antigen receptor, thereby decreasing the threshold for receptor-dependent signaling.[15]

We describe four patients, from two unrelated families, who had hypogammaglobulinemia and defective CD19. These patients had normal levels of B cells in blood but had undetectable or very low levels of surface CD19, owing to a homozygous mutation in the *CD19* gene. The antibody deficiency in these patients resulted from a subnormal response to antigen by mature B cells.

CASE REPORTS

Patient 1, a girl of Turkish descent and the second child of consanguineous parents (second cousins), was referred to the Selçuk University Hospital in Konya, Turkey, at ten years of age with a seven-month history of intermittent hematuria. She also had a history of recurrent bronchiolitis and bronchopneumonia starting at one year of age and meningitis starting at eight years of age. On physical examination, her heart rate was 118 beats per minute and auscultation of the lung revealed rhonchi and wheezing. The differential blood count was normal, but urinalysis showed microscopic hematuria without proteinuria. The complement levels were normal, tests for antinuclear antibodies were negative, the erythrocyte sedimentation rate was 41 mm per hour, and the level of C-reactive protein was 93 mg per deciliter. The dermal response to purified protein derivative after two vaccinations with bacille Calmette-Guérin was 12 mm. Renal ultrasonography showed no abnormalities; a renal biopsy revealed proliferation of mesangial cells, interstitial edema, and an apparently normal basement membrane of the glomerulus. She was given a diagnosis

Table 1. Baseline characteristics and laboratory findings in four patients with CD19 deficiency.

Characteristic	Patient 1	Patient 2	Patient 3	Patient 4
Sex	female	male	female	female
Age at onset - yr	1	7	6	5
Age at diagnosis - yr	10	35	33	49
Lymphocytes - cells/mm^3				
Total lymphocytes	4480	2182	2508	2059
CD3$^+$ T cells	3270	1520	1855	1384
CD4$^+$ T cells	1792	713	1070	620
CD8$^+$ T cells	1478	720	692	696
CD20$^+$ B cells	806	286	521	268
CD16$^+$ CD56$^+$ Natural Killer cells	313	277	348	288
Serum immunoglobulin level - mg/dl (95 % CI for age-matched controls)*				
IgG level	**325** (842-1943)	**204** (804-2212)	**198** (814-2047)	**256** (814-2047)
IgA level	292 (62-390)	**18** (99-489)	**7** (99-489)	**19** (81-538)
IgM level	**25** (54-392)	**47** (58-324)	**30** (68-379)	63 (42-600)
Blood group	A, Rh$^+$	0, Rh$^+$	0, Rh$^+$	0, Rh$^+$
Isohemagglutinin levels	Not detectable	Low	Low	Low

* Serum Ig levels at diagnosis are shown. Subnormal values are given in bold type; CI denotes confidence interval. Age-matched controls have been described previously.[27]

of postinfectious glomerulonephritis and was also found to have hypogammaglobulinemia (Table 1).

Patients 2, 3, and 4 were siblings, children of unrelated parents of Colombian descent, and were examined at the Primary Immunodeficiencies Clinic of the University of Antioquia in Medellín, Colombia, at 35, 33, and 49 years of age, respectively. All three had had otitis media, sinusitis, and pharyngitis during childhood. In addition, Patient 2 (the index patient) had had four bouts of pneumonia between the ages of 18 and 35 years; more recently, he had received diagnoses of bacterial conjunctivitis and chronic gastritis (*Heliobacter pylori* infection). He was given a diagnosis of hypogammaglobulinemia due to common variable immunodeficiency (Table 1). His two sisters also had hypogammaglobulinemia (Table 1) and histories of recurrent upper respiratory tract infections. Patient 3 first had pneumonia at 20 years of age, herpes zoster at 30 years of age, and more recently, recurrent bacterial conjunctivitis with dacryocystitis and diarrhea. Patient 4 had had pneumonia at eight years of age, recurrent skin abscesses and more bouts of pneumonia during childhood, and chronic diarrhea, bronchitis, and recurrent bacterial conjunctivitis during adulthood. These three patients had had surgery for chronic sinusitis.

Intravenous immune globulin therapy was initiated in all four patients soon after the diagnosis of hypogammaglobulinemia. All patients are now clinically well, with no signs of

lymphoproliferation, cancer, or autoimmunity. The other siblings of both families have normal serum immunoglobulin levels and no symptoms of immunodeficiency. However, one sibling of Patients 2, 3, and 4 was given a diagnosis of systemic lupus erythematosus at 13 years of age, according to the criteria of the American Rheumatism Association.[16]

METHODS

Diagnostic workups of blood, biopsy of bone marrow, cervical lymph node, and tonsil, and rabies vaccination were carried out in patients and family members after they provided oral informed consent and according to the guidelines of the local medical ethics committees in Turkey, Colombia, and the Netherlands. Controls provided written consent.

Amplification and sequence analysis of the *CD19* gene

The *CD19* gene on chromosome 16 (16p11.2) spans 7.5 kb of genomic DNA and codes for a protein of 556 amino acids, including the signal peptide.[17] DNA was isolated from granulocytes that had been isolated with the use of Ficoll gradient centrifugation. All 15 exons were amplified by the polymerase chain reaction (PCR; the primer sequences are listed in Appendix 10), and were sequenced on an ABI Prism 377 fluorescent sequencer (Applied Biosystems).

Total complementary DNA (cDNA) was prepared from messenger RNA (mRNA) that had been isolated from the patients' mononuclear cells, as described previously.[18] Fragments of *CD19* cDNA were amplified and sequenced to verify the mutations found in the genomic DNA.

CD19 transcripts

Levels of the *CD19* transcript in blood B cells were quantified from reverse-transcribed mRNA from patients' blood mononuclear cells with the use of TaqMan real-time quantitative PCR with newly designed primers and a FAM–TAMRA-labeled fluorogenic probe (sequences are given in Appendix 11). The transcript levels were normalized against those of a control *ABL* gene and corrected for the numbers of B cells as determined by means of flowcytometric analysis.[19]

Analysis of somatic hypermutation

Hypermutation of immunoglobulin genes was studied in the VH3-Cα, VH4-Cα, VH3-Cγ, and VH4-Cγ fragments, which were amplified from the cDNA and cloned into a pGEM-T easy vector (Promega). Twenty-one clones of Patient 1, 44 clones of Patient 2, 39 clones of Patient 3, and 39 clones of Patient 4 were sequenced. The International ImMunoGeneTics information system (http://imgt.cines.fr/) was used to identify the V, D, and J segments and somatic mutations.[20] The ratio of replacement and silent mutations was determined for the framework regions and the complementarity-determining regions, and the distribution of replacement mutations in the framework regions was analyzed according to the binomial distribution model of Chang and Casali.[21]

Flow cytometry

Flowcytometric immunophenotyping of blood samples from four patients, nine relatives (including eight carriers), and eight controls was performed to evaluate markers and subgroups of B cells. We also analyzed bone marrow samples from Patient 1 and from five controls, as previously described.[22] Antibodies against CD21 (LB21), CD81 (JS-81), and CD225 (Leu13) were used to further analyze the CD19-complex.[15, 23]

Western blotting

B cells in blood were sorted from mononuclear cells with the use of the magnetic affinity cell-sorting B-cell purification kit (Miltenyi Biotec). Lysates of B cells ($2x10^6$) were separated by sodium dodecyl sulfate–polyacrylamide-gel electrophoresis and immunoblotted with the use of a polyclonal rabbit antiserum against the intracellular portion of human CD19.[24]

Histology and immunohistochemistry

The biopsy specimen of a cervical lymph node from Patient 1 and the tonsil-biopsy specimen from Patient 2 were fixed in formalin, embedded in paraffin, and stained with hematoxylin and eosin. Immunohistochemistry was performed according to standard procedures involving monoclonal antibodies against CD20 (L26), CD21 (1F8), CD79a (JCB117), BCL-2 (124), BCL-6 (PG-B6p), and Ki67 (MIB-1), all obtained from DakoCytomation.

B-cell activation

Blood mononuclear cells were incubated with 6 μg of indo-1 per milliliter of cell suspension (molecular probes) to evaluate calcium fluxes after stimulation. Levels of free intracellular calcium were measured in $CD20^+$ B cells with the use of flow cytometry in a FACSVantage station (BD Biosciences) before and after stimulation with *Staphylococcus aureus* Cowan I (Calbiochem) in a 1:500 dilution or with 20 μg of goat antihuman IgM $F(ab')_2$ per milliliter (Jackson Immunoresearch Laboratories). Subsequently, 2 μg of ionomycin per milliliter (molecular probes) was added to control for intracellular loading of indo-1.

B-cell proliferation

Enriched blood B cells ($1x10^5$ per well) were cultured in 96-well flat-bottomed plates for 48 or 72 hours at 37°C and stimulated with *S. aureus* Cowan I in a 1:10,000 dilution or with 20 μg of anti-IgM $F(ab')_2$ per milliliter. Proliferation of the B cells was assessed by incorporating 0.5 μCi of [^{3}H]thymidine in each well.

***In vivo* vaccination with rabies vaccine**

Patients, relatives, and controls were vaccinated twice with the human diploid-cell rabies vaccine (Institut Pasteur Merieux, Merck Sharp & Dohme) according to standard protocols. Antibody levels and avidity were measured as previously described.[25]

RESULTS

Identification of CD19 defects

Flowcytometric immunophenotyping with antibodies against the cell-surface markers CD3 (on T cells), CD16 or CD56 or both (on natural killer cells), and CD19 or CD20 (on B cells) results in the identification of more than 95 percent of lymphocytes. All four patients had normal absolute numbers of $CD20^+$ B cells. In contrast, membrane expression of CD19 on B cells was not detectable in Patient 1 and was barely detectable in Patients 2, 3, and 4 (Figure 1A).

Sequencing of the *CD19* gene in Patient 1 revealed an insertion of an adenine in exon 6 of both alleles of *CD19*. This insertion causes a frame shift, and thus there is a premature stop codon at the intracellular amino acid 328 (Figure 1B). Patients 2, 3, and 4 were homozygous for a dinucleotide deletion of guanine and adenine in exon 11, resulting in a premature stop codon at the intracellular amino acid 465 (Figure 1B). Living parents and several members of both families carried one mutated allele (Figure 1C and 1D) but had no clinical features of immunodeficiency.

Normal levels of *CD19* transcripts were found in B cells from all patients (data not shown). Western blotting did not detect the CD19 protein in B cells from Patient 1, because the CD19 in this patient lacked most of the cytoplasmic domains. However, a reduction in intracellular levels of CD19 was observed in B cells from Patient 4 (data not shown), confirming the flowcytometric results.

Membrane expression of the CD19-complex

The level of expression of CD19 in the carriers of the mutations was lower than that in controls (Figure 2A). In addition, CD21 expression on the surface of patients' B cells was also reduced (Figure 2B). CD81 and CD225 molecules were expressed at normal levels by the B cells of all patients and carriers (data not shown).

B-cell differentiation

The precursor-B-cell compartment in the bone marrow of Patient 1 showed a normal relative distribution of all differentiation stages (data not shown).[22] $Ig\kappa^+$ B cells and $Ig\lambda^+$ B cells were present in normal ratios in blood samples from all patients. Moreover, B cells from all four patients had normal expression of other membrane-bound molecules, including IgM, IgD, CD20, CD22, CD38, CD40, β2-microglobulin, and HLA-DR. However, all four patients had decreased levels of $CD5^+$ B cells (8 to 12 percent) and $CD27^+$ memory B cells (1 to 6 percent) in blood, as compared with levels in age-matched controls and carriers (19 to 41 percent for $CD5^+$ and 17 to 28 percent for $CD27^+$) (Figure 2C and 2D). Memory B cells that switched immunoglobulin class were $CD27^+IgD^-$, lacking IgM and IgD, and were present in decreased numbers in blood samples from all patients. Sequencing of the V_H-Cα and V_H-Cγ transcripts showed somatic hypermutations in a pattern suggesting the potential to produce an antigen-selected B-cell-receptor repertoire, which is a molecular sign of a response to antigenic stimulation (data not shown).

The numbers of $CD27^+$ memory B cells in carriers in the family of Patient 1 were similar to those in controls, whereas carriers in the family of Patients 2, 3, and 4 had fewer memory

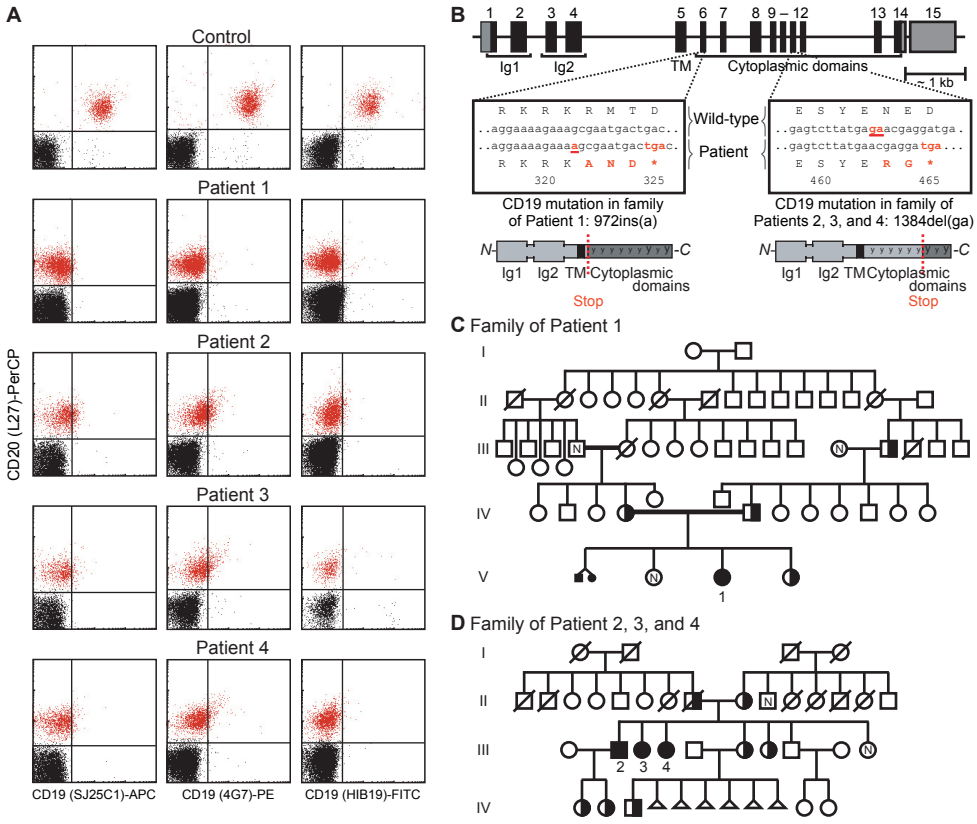


Figure 1. Mutations in the *CD19* gene.
Panel **A** shows the results of flowcytometric analysis of the membrane expression of CD19 on B cells in blood from the four patients and a control. B cells are shown in red; black represents non-B cells; both the x-axis and y-axis show a log (base-10) scale. Patient 1 showed a complete lack of CD19 on $CD20^{+}$ blood B cells, and the three other patients had variable traces. Three CD19 monoclonal antibodies (SJ25C1, 4G7, and HIB19) and one CD20 monoclonal antibody (L27) were used. PerCP denotes perdinin chlorophyll protein, APC allophycocyanin, PE phycoerythrin, and FITC fluorescein isothiocyanate. Control B cells stained positive for both CD19 and CD20. Panel **B** shows the location of mutations in the *CD19* gene. Exons 1 through 14 code for CD19, which consists of two immunoglobulin-like domains (Ig1 and Ig2), a transmembrane domain (TM), and many cytoplasmic domains containing nine conserved tyrosines. Patient 1 had a homozygous insertion of an adenine in exon 6, resulting in three altered amino acids and a premature stop codon at position 328 of the protein. Patients 2, 3, and 4 had a homozygous dinucleotide deletion in exon 11, resulting in two altered amino acids and a premature stop codon at position 465. The premature stop codons lead to the deletion of part of the cytoplasmic domains, including the two tyrosine residues that are critical for signal transduction through CD19 (indicated with a large y in the diagram of the protein).[24] Panels **C** and **D** show the pedigrees of Patient 1 and Patients 2, 3, and 4, respectively. Half-solid symbols denote known carriers of the mutation, solid symbols family members who are homozygous for the mutation, symbols with a slash deceased members, squares male family members, circles female family members, and N family members who are known not to be carriers. The parents of Patient 1 were consanguineous. No consanguinity could be confirmed in the family of Patients 2, 3, and 4.

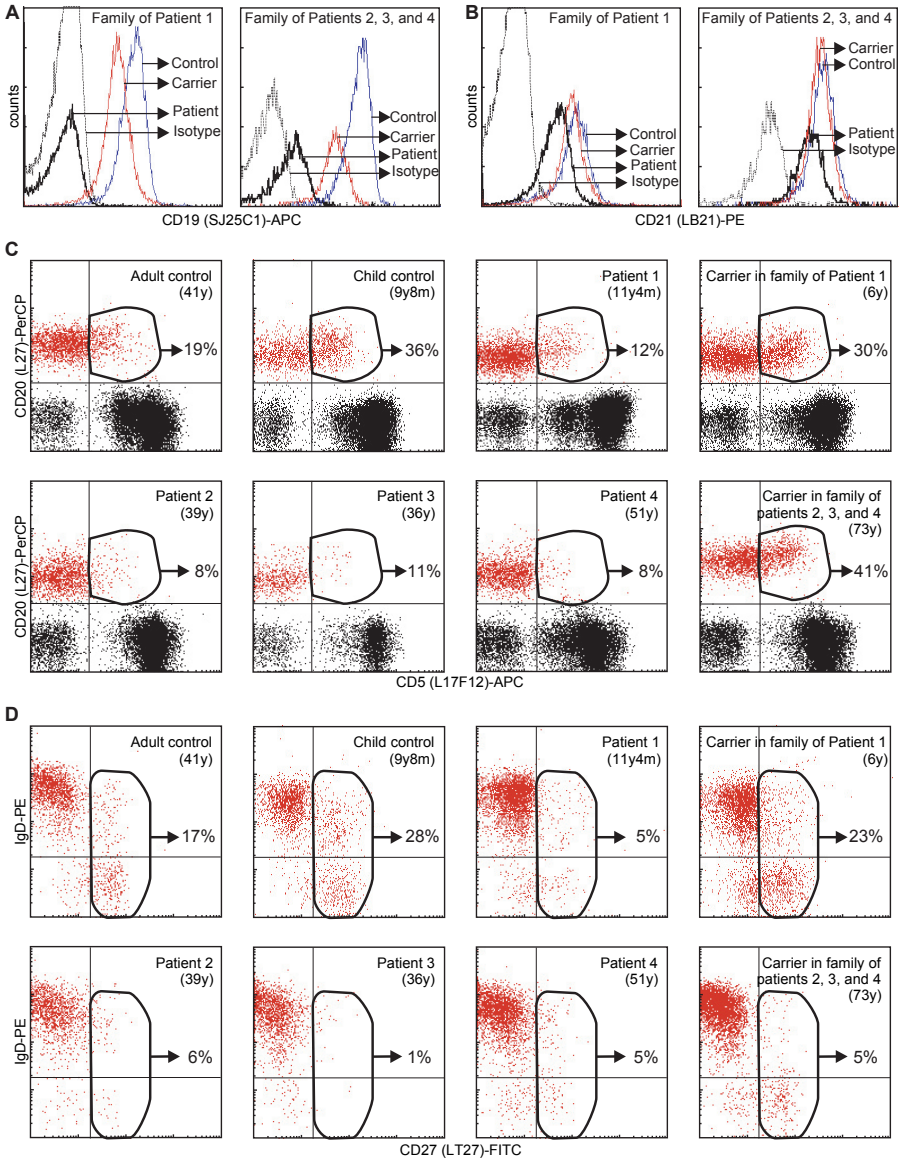


Figure 2. Immunophenotype of mature B cells in patients with a CD19 mutation.
All four patients had a reduced level of expression of CD19 or no expression on B cells in blood, as compared with levels in the controls and the mouse IgG1 isotype control (Panel **A**). In addition, the level of expression of CD19 on B cells in blood from carriers was lower than that in the controls. The level of expression of CD21 on B cells in blood from the patients was lower than that in carriers and controls (Panel **B**). The level of expression of CD5 on $CD20^+$ B cells (shown in red; black represents non-B cells) was decreased in all patients as compared with carriers and controls, whereas it was normal on $CD20^-$ cells (Panel **C**; log scale). The level of expression of $CD27^+IgD^+$ and $CD27^+IgD^-$ memory B cells was lower in the patients than in age-matched controls (Panel **D**; log scale). PerCP denotes perdinin chlorophyll protein, APC allophycocyanin, PE phycoerythrin, and FITC fluorescein isothiocyanate.

B cells than controls (Figure 2D). In general, however, in contrast to the patients, the carriers in the family of Patients 2, 3, and 4 had mainly memory B cells that had switched immunoglobulin class and normal serum immunoglobulin levels (data not shown).

Germinal centers in secondary lymphoid organs

Histologic analysis of a sample of cervical lymph node from Patient 1 showed multiple B-cell follicles of small-to-moderate size with proliferation centers on immunohistochemical staining for CD79a, BCL-2, BCL-6, CD21, and Ki67 (Figure 3A). Follicles of normal size and architecture were also found in the tonsil specimen from Patient 2 (Figure 3B).

Calcium mobilization and proliferation on stimulation of the B-cell receptor

B-cell receptor-mediated activation causes an initial flux of calcium from the rough endoplasmic reticulum into the cytoplasm, followed by a sustained influx of extracellular calcium.[26] As compared with control B cells, the B cells from all four patients with CD19 deficiency exhibited defective calcium fluxes after stimulation with an anti-IgM antibody, which cross-links IgM molecules on the B-cell surface. No calcium influx or efflux was observed in the B cells from Patient 1, whereas the initial calcium influx was delayed in B cells from Patients 2, 3, and 4. Furthermore, calcium efflux in all four patients was delayed and remained low for the duration of the experiment (Figure 4A; data for Patients 2 and 3 not shown). B cells from heterozygous relatives had normal calcium fluxes (data not shown). Normal calcium fluxes were found in Patients 2, 3, and 4 in response to *S. aureus* Cowan I, which activates B cells by means of innate immune receptors and the B-cell receptor (data not shown).

B cells from Patient 1 proliferated poorly after stimulation with *S. aureus* Cowan I or an anti-IgM antibody, as compared with B cells from the control (Figure 4B). B cells from the parents of Patient 1, who are carriers of the mutation, showed intermediate proliferation in response to both *S. aureus* Cowan I and anti-IgM antibody (Figure 4B). B cells from carriers in the family of Patients 2, 3, and 4 proliferated normally after stimulation with *S. aureus* Cowan I or anti-IgM antibody (not shown).

Antibody response to rabies vaccine

To test the primary and secondary antibody responses *in vivo*, all patients were vaccinated twice with rabies vaccine (at week 0 and week 13). The primary response of production of IgG anti-rabies antibody in Patient 1 was low but detectable, whereas Patients 2, 3, and 4 did produce IgG anti-rabies antibodies after the first vaccination. The secondary IgG response, after week 13, in all patients was below the 95 percent confidence interval of the normal response (Figure 4C).[25] Furthermore, no increase in avidity was observed in Patient 2, 3, or 4, and Patient 1 had only a marginal increase in avidity during the secondary response (Figure 4D). The second antibody response and the increase in avidity maturation were normal in the two carriers evaluated in the family of Patients 2, 3, and 4 (Figure 4C and 4D).

DISCUSSION

We described four patients (from two different families) with homozygous mutations in *CD19* and showed that their profound hypogammaglobulinemia resulted from a poor response to antigen by mature B cells. Both CD19 mutations were frameshift mutations, causing premature stop codons that result in the deletion of the total cytoplasmic domain (in the family of Patient 1) or a major part of the cytoplasmic domain (in family B). Both truncations result in the lack of the three C-terminal tyrosine residues, two of which have been shown to be critical for signal transduction through CD19 (Figure 1B).[24]

These mutations are not likely to be polymorphisms; CD19 is the most broadly used B-cell marker for flowcytometric immunophenotyping of lymphocytes and has not been found to be lacking in large-scale studies of healthy children.[27, 28] Moreover, the B cells of heterozygous carriers of the mutation expressed CD19, albeit at slightly lower levels than noncarriers. Futhermore, CD19 - together with CD21, CD81, and CD225 - forms a signaling complex that decreases the threshold for the activation of B cells as mediated by the B-cell receptor.[29]

B cells from CD81-knockout mice have decreased CD19 expression.[30, 31] The mature B cells from our patients, however, had a normal level of expression of CD81 and CD225 but a reduced level of expression of CD21, suggesting a dependency of CD21 on CD19. Normal levels of *CD19* transcripts in the B cells of these patients indicate that the reduction in the level of the protein does not result from nonsense-mediated decay of mRNA, but rather from the instability of the truncated CD19 proteins. This interpretation is consistent with the observation that truncated (but otherwise stable) CD19 molecules can associate with CD81 and CD225 on the membranes of B cells in which 95 percent of the cytoplasmic domains of CD19 are deleted.[32]

The CD19 deficiency did not disturb the differentiation of precursor-B cells in the bone marrow. Still, CD19-deficient B cells may undergo differentiation less efficiently than $CD19^+$ B cells, as has been observed in CD19-knockout mice.[33, 34] Nevertheless, our patients had normal numbers of mature B cells in blood.

The numbers of $CD5^+$ B cells in our patients were decreased, as they are in CD19-knockout mice.[35, 36] CD5 is a negative regulator of B-cell receptor signaling, and in mice, this marker defines the subgroup of autoreactive B-1a cells.[37, 38] It is likely that, in our patients, the inhibitory effect of CD5 on B-cell-receptor signaling is superfluous in the absence of CD19.

The numbers of $CD27^+IgD^+$ and $CD27^+IgD^-$ memory B cells in blood from all four patients were decreased. Together with the finding of decreased serum immunoglobulin levels, this finding suggests that terminal differentiation of B cells into memory B cells and plasma cells is affected. In two patients, secondary lymphoid tissues contained follicles of small-to-moderate size with germinal centers that appeared normal. CD19-knockout mice lack germinal centers, but centers appear when the animals are challenged with live virus.[39] IgG antibodies and circulating memory B cells are defective in these mice,[39] as they are in our four patients.

IgG and IgA transcripts with antigen-driven hypermutation were found in our patients at variable levels, indicating that CD19-deficient B cells are in principle able to undergo antigen-dependent differentiation, albeit with reduced efficiency. In mice, defects in CD19 signaling due to mutations in critical intracellular tyrosines result in the absence of high-affinity antibodies owing to low B-cell proliferation and abnormal immunoglobulin-class switching.[24, 40, 41]

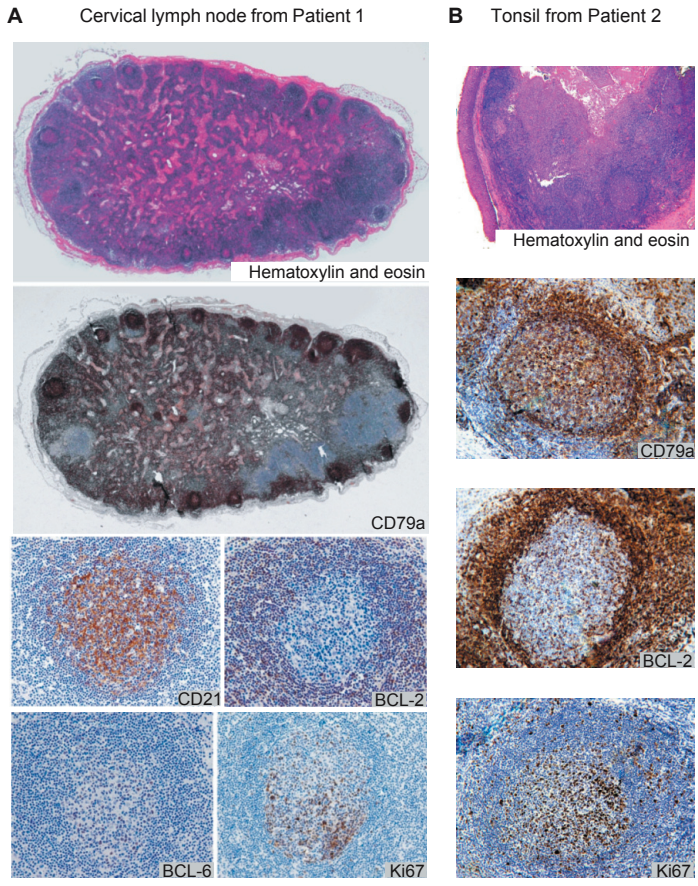


Figure 3. Normal formation of the germinal center.
Panel **A** shows cross sections of a cervical lymph node from Patient 1 with secondary lymphoid follicles of small-to-moderate size. The whole lymph node was stained with hematoxylin and eosin (top section) and CD79a (second section). The dense light-blue areas of CD79a-negative cells consisted mainly of T cells, as observed on staining with CD3 (not shown). The follicles containing B cells are shown (bottom four sections) after staining with monoclonal antibodies against CD21, BCL-2, BCL-6, and Ki67. The germinal centers consist of BCL-2$^-$BCL-6$^+$ B cells, proliferating in a network of CD21$^+$ follicular dendritic cells. The proliferation marker Ki67 distinguishes dark and light areas, with the number of cycling cells greater in the germinal center. Panel **B** shows a sample of a tonsil from Patient 2, stained with hematoxylin and eosin (top section) or with monoclonal antibodies against CD79a (second section), BCL-2 (third section), or Ki67 (bottom section). The germinal centers consist of CD79a$^+$BCL-2$^-$ B cells, with Ki67-positive cycling cells.

The inefficient response of CD19-deficient B cells in these patients probably results from the loss of dual antigen recognition through the B-cell receptor and the CD19-complex.[15, 42] The B cells from all four patients had abnormal calcium fluxes on stimulation, but the differences between Patient 1 and the other three patients may be due to the presence of different mutations. Patient 1 lacked all the intracellular domains responsible for signal transduction[32]; in Patients 2, 3, and 4, the CD19 molecule lacked only 92 amino acids at the intracellular C-terminal end, which contains two critical tyrosines that are required for most functions of CD19.[24] Most of the intracellular domains were present, potentially permitting some B-cell signaling.[43] Nevertheless, all four patients with CD19 deficiency had substantially decreased antibody responses (levels and avidity) after vaccination with rabies vaccine. We conclude that the CD19 signals that normally enable B-cell responses to antigenic stimuli are lacking in our patients.

This study shows that the disruption of CD19 signaling results in a primary antibody deficiency, mainly characterized by a poor antigen-specific response. On the basis of the crucial role of CD19 in signaling by the B-cell receptor on antigen recognition, it is likely that defects in other members of the CD19-complex (CD21, CD81, and CD225) also lead to antibody deficiency in humans.

ACKNOWLEDGEMENTS

We are indebted to Mrs. B.H. Barendregt, Mrs. S. de Bruin-Versteeg, and Mr. E.F.E. de Haas for their support with flowcytometric immunophenotyping and functional assays; to S. Evans, Rockwell Park Institute, for the antihuman CD225; to R. Carter, University of Alabama at Birmingham, for the polyclonal rabbit anti-CD19 antiserum; to Mrs. C.M. Jol-van der Zijde for performing the serologic antirabies response experiments; to Ms. C. Rubeles and Mrs. B. Vieco, Dr. C. Montoya, Dr. M. Schlesier, Dr. U. Salzer, and Mrs. D. Thiel for their technical assistance; to Dr. H. Arbag and Dr. M. Cihat Avunduk for their support in the diagnostic procedures; and to Mrs. W.M. Comans-Bitter for assistance with the figures.

DISCLOSURES

Supported by grants from Sophia Kinderziekenhuis Fonds (349), Revolving Fund 2000, the Vereniging Trustfonds Erasmus Universiteit Rotterdam, the Colombian Institute for Scientific and Technological Development (Colciencias; 1115-04-11924 and 1115-05-16874), and the Deutsche Forschungsgemeinschaft (GR1617/3 and SFB620). No potential conflict of interest relevant to this article was reported.

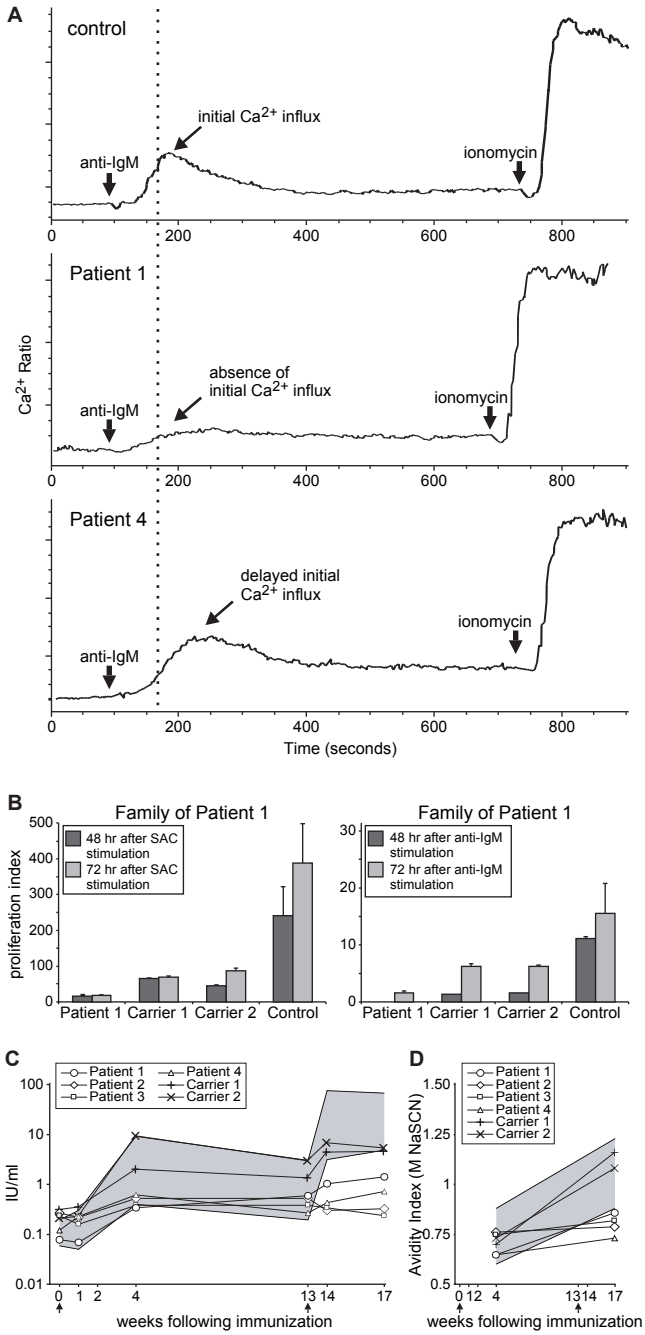

A
control
initial Ca2+ influx
anti-IgM
ionomycin
0
200
400
600
800
Patient 1
absence of
initial Ca2+ influx
anti-IgM
ionomycin
Ca2+ Ratio
Patient 4
delayed initial
Ca2+ influx
anti-IgM
ionomycin
Time (seconds)
B
Family of Patient 1
48 hr after SAC stimulation
72 hr after SAC stimulation
proliferation index
Patient 1
Carrier 1
Carrier 2
Control
Family of Patient 1
48 hr after anti-IgM stimulation
72 hr after anti-IgM stimulation
C
Patient 1
Patient 2
Patient 3
Patient 4
Carrier 1
Carrier 2
IU/ml
weeks following immunization
D
Avidity Index (M NaSCN)
weeks following immunization

Figure 4 (facing page). *In vitro* responses of B cells and antibody responses to rabies vaccine.
The stimulation of blood mononuclear cells with anti-IgM antibody did not result in an initial calcium influx in $CD20^+$ B cells from Patient 1 and resulted in a delayed initial calcium influx in cells from Patient 4, as compared with the control cells (Panel **A**; the dashed line indicates the peak of the initial calcium influx in the control). In Patient 4, the subsequent calcium flux was sustained and relatively high. No defect in calcium flux was seen in carriers of the CD19 mutation (not shown). Ionomycin was added as a control for intracellular loading of indo-1. The proliferation of B cells from Patient 1 was less than that in two carriers and a control from her family after stimulation with *Staphylococcus aureus* Cowan I (SAC) or anti-IgM antibody (Panel **B**). The blood B cells were harvested and cultured for 48 or 72 hours in medium alone or medium with either SAC or anti-IgM antibody. The proliferation index is the ratio of the counts per minute of stimulated and unstimulated B cells. Each experiment was performed in triplicate, and T bars represent standard errors. The levels of anti-rabies IgG antibodies were subnormal in the four patients after an initial vaccination at week 0 (arrow) and a booster vaccination at week 13 (arrowhead) (Panel **C**). The relative avidity index of the antibody was also low in these patients (Panel **D**). These patients were compared with 18 controls vaccinated according to the same methods;[25] the gray area represents the 95 percent confidence interval of the anti-rabies IgG level and avidity index. Two carriers from the family of Patients 2, 3, and 4 were also tested and had a normal quantitative and qualitative response. NaSCN denotes sodium thiocyanate.

REFERENCES

1. Ballow M. Primary immunodeficiency disorders: antibody deficiency. *J Allergy Clin Immunol* 2002;**109**: 581-91.
2. Vetrie D, Vorechovsky I, Sideras P, Holland J, Davies A, Flinter F, Hammarstrom L, Kinnon C, Levinsky R, Bobrow M, et al. The gene involved in X-linked agammaglobulinaemia is a member of the src family of protein-tyrosine kinases. *Nature* 1993;**361**:226-33.
3. Tsukada S, Saffran DC, Rawlings DJ, Parolini O, Allen RC, Klisak I, Sparkes RS, Kubagawa H, Mohandas T, Quan S, et al. Deficient expression of a B cell cytoplasmic tyrosine kinase in human X-linked agammaglobulinemia. *Cell* 1993;**72**:279-90.
4. Yel L, Minegishi Y, Coustan-Smith E, Buckley RH, Trubel H, Pachman LM, Kitchingman GR, Campana D, Rohrer J, Conley ME. Mutations in the mu heavy-chain gene in patients with agammaglobulinemia. *N Engl J Med* 1996;**335**:1486-93.
5. Minegishi Y, Coustan-Smith E, Wang YH, Cooper MD, Campana D, Conley ME. Mutations in the human lambda5/14.1 gene result in B cell deficiency and agammaglobulinemia. *J Exp Med* 1998;**187**:71-7.
6. Minegishi Y, Hendershot LM, Conley ME. Novel mechanisms control the folding and assembly of lambda5/14.1 and VpreB to produce an intact surrogate light chain. *Proc Natl Acad Sci U S A* 1999;**96**:3041-6.
7. Minegishi Y, Rohrer J, Coustan-Smith E, Lederman HM, Pappu R, Campana D, Chan AC, Conley ME. An essential role for BLNK in human B cell development. *Science* 1999;**286**:1954-7.
8. Conley ME, Rohrer J, Rapalus L, Boylin EC, Minegishi Y. Defects in early B-cell development: comparing the consequences of abnormalities in pre-BCR signaling in the human and the mouse. *Immunol Rev* 2000;**178**: 75-90.
9. Allen RC, Armitage RJ, Conley ME, Rosenblatt H, Jenkins NA, Copeland NG, Bedell MA, Edelhoff S, Disteche CM, Simoneaux DK, et al. CD40 ligand gene defects responsible for X-linked hyper-IgM syndrome. *Science* 1993;**259**:990-3.
10. Revy P, Muto T, Levy Y, Geissmann F, Plebani A, Sanal O, Catalan N, Forveille M, Dufourcq-Labelouse R, Gennery A, Tezcan I, Ersoy F, Kayserili H, Ugazio AG, Brousse N, Muramatsu M, Notarangelo LD, Kinoshita K, Honjo T, Fischer A, Durandy A. Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2). *Cell* 2000;**102**:565-75.
11. Ferrari S, Giliani S, Insalaco A, Al-Ghonaium A, Soresina AR, Loubser M, Avanzini MA, Marconi M, Badolato R, Ugazio AG, Levy Y, Catalan N, Durandy A, Tbakhi A, Notarangelo LD, Plebani A. Mutations of CD40 gene cause an autosomal recessive form of immunodeficiency with hyper IgM. *Proc Natl Acad Sci U S A* 2001;**98**:12614-9.

12. Imai K, Slupphaug G, Lee WI, Revy P, Nonoyama S, Catalan N, Yel L, Forveille M, Kavli B, Krokan HE, Ochs HD, Fischer A, Durandy A. Human uracil-DNA glycosylase deficiency associated with profoundly impaired immunoglobulin class-switch recombination. *Nat Immunol* 2003;**4**:1023-8.
13. Grimbacher B, Hutloff A, Schlesier M, Glocker E, Warnatz K, Drager R, Eibel H, Fischer B, Schaffer AA, Mages HW, Kroczek RA, Peter HH. Homozygous loss of ICOS is associated with adult-onset common variable immunodeficiency. *Nat Immunol* 2003;**4**:261-8.
14. Salzer U, Chapel HM, Webster AD, Pan-Hammarstrom Q, Schmitt-Graeff A, Schlesier M, Peter HH, Rockstroh JK, Schneider P, Schaffer AA, Hammarstrom L, Grimbacher B. Mutations in TNFRSF13B encoding TACI are associated with common variable immunodeficiency in humans. *Nat Genet* 2005;**37**:820-8.
15. Carter RH, Fearon DT. CD19: lowering the threshold for antigen receptor stimulation of B lymphocytes. *Science* 1992;**256**:105-7.
16. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, Healey LA, Kaplan SR, Liang MH, Luthra HS, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum* 1988;**31**:315-24.
17. Tedder TF, Isaacs CM. Isolation of cDNAs encoding the CD19 antigen of human and mouse B lymphocytes. A new member of the immunoglobulin superfamily. *J Immunol* 1989;**143**:712-7.
18. van Dongen JJM, Macintyre EA, Gabert JA, Delabesse E, Rossi V, Saglio G, Gottardi E, Rambaldi A, Dotti G, Griesinger F, Parreira A, Gameiro P, Diaz MG, Malec M, Langerak AW, San Miguel JF, Biondi A. Standardized RT-PCR analysis of fusion gene transcripts from chromosome aberrations in acute leukemia for detection of minimal residual disease. Report of the BIOMED-1 Concerted Action: investigation of minimal residual disease in acute leukemia. *Leukemia* 1999;**13**:1901-28.
19. Beillard E, Pallisgaard N, van der Velden VH, Bi W, Dee R, van der Schoot E, Delabesse E, Macintyre E, Gottardi E, Saglio G, Watzinger F, Lion T, van Dongen JJM, Hokland P, Gabert J. Evaluation of candidate control genes for diagnosis and residual disease detection in leukemic patients using 'real-time' quantitative reverse-transcriptase polymerase chain reaction (RQ-PCR) - a Europe against cancer program. *Leukemia* 2003;**17**: 2474-86.
20. Lefranc MP. IMGT databases, web resources and tools for immunoglobulin and T cell receptor sequence analysis, http://imgt.cines.fr. *Leukemia* 2003;**17**:260-6.
21. Chang B, Casali P. The CDR1 sequences of a major proportion of human germline Ig VH genes are inherently susceptible to amino acid replacement. *Immunol Today* 1994;**15**:367-73.
22. Noordzij JG, de Bruin-Versteeg S, Verkaik NS, Vossen JM, de Groot R, Bernatowska E, Langerak AW, van Gent DC, van Dongen JJM. The immunophenotypic and immunogenotypic B-cell differentiation arrest in bone marrow of RAG-deficient SCID patients corresponds to residual recombination activities of mutated RAG proteins. *Blood* 2002;**100**:2145-52.
23. Evans SS, Lee DB, Han T, Tomasi TB, Evans RL. Monoclonal antibody to the interferon-inducible protein Leu-13 triggers aggregation and inhibits proliferation of leukemic B cells. *Blood* 1990;**76**:2583-93.
24. Wang Y, Brooks SR, Li X, Anzelon AN, Rickert RC, Carter RH. The physiologic role of CD19 cytoplasmic tyrosines. *Immunity* 2002;**17**:501-14.
25. Brinkman DM, Jol-van der Zijde CM, ten Dam MM, Vossen JM, Osterhaus AD, Kroon FP, van Tol MJ. Vaccination with rabies to study the humoral and cellular immune response to a T-cell dependent neoantigen in man. *J Clin Immunol* 2003;**23**:528-38.
26. Scharenberg AM, Kinet JP. PtdIns-3,4,5-P3: a regulatory nexus between tyrosine kinases and sustained calcium signals. *Cell* 1998;**94**:5-8.
27. Comans-Bitter WM, de Groot R, van den Beemd R, Neijens HJ, Hop WC, Groeneveld K, Hooijkaas H, van Dongen JJM. Immunophenotyping of blood lymphocytes in childhood. Reference values for lymphocyte subpopulations. *J Pediatr* 1997;**130**:388-93.
28. Ikinciogullari A, Kendirli T, Dogu F, Egin Y, Reisli I, Cin S, Babacan E. Peripheral blood lymphocyte subsets in healthy Turkish children. *Turk J Pediatr* 2004;**46**:125-30.
29. Bradbury LE, Kansas GS, Levy S, Evans RL, Tedder TF. The CD19/CD21 signal transducing complex of human B lymphocytes includes the target of antiproliferative antibody-1 and Leu-13 molecules. *J Immunol* 1992;**149**:2841-50.
30. Maecker HT, Levy S. Normal lymphocyte development but delayed humoral immune response in CD81-null mice. *J Exp Med* 1997;**185**:1505-10.

31. Shoham T, Rajapaksa R, Boucheix C, Rubinstein E, Poe JC, Tedder TF, Levy S. The tetraspanin CD81 regulates the expression of CD19 during B cell development in a postendoplasmic reticulum compartment. *J Immunol* 2003;**171**:4062-72.
32. Bradbury LE, Goldmacher VS, Tedder TF. The CD19 signal transduction complex of B lymphocytes. Deletion of the CD19 cytoplasmic domain alters signal transduction but not complex formation with TAPA-1 and Leu 13. *J Immunol* 1993;**151**:2915-27.
33. Otero DC, Rickert RC. CD19 Function in Early and Late B Cell Development. II. CD19 Facilitates the Pro-B/Pre-B Transition. *J Immunol* 2003;**171**:5921-30.
34. Diamant E, Keren Z, Melamed D. CD19 regulates positive selection and maturation in B lymphopoiesis: Lack of CD19 imposes developmental arrest of immature B cells and consequential stimulation of receptor editing. *Blood* 2005.
35. Rickert RC, Rajewsky K, Roes J. Impairment of T-cell-dependent B-cell responses and B-1 cell development in CD19-deficient mice. *Nature* 1995;**376**:352-5.
36. Engel P, Zhou LJ, Ord DC, Sato S, Koller B, Tedder TF. Abnormal B lymphocyte development, activation, and differentiation in mice that lack or overexpress the CD19 signal transduction molecule. *Immunity* 1995;**3**: 39-50.
37. Gary-Gouy H, Bruhns P, Schmitt C, Dalloul A, Daeron M, Bismuth G. The pseudo-immunoreceptor tyrosine-based activation motif of CD5 mediates its inhibitory action on B-cell receptor signaling. *J Biol Chem* 2000;**275**:548-56.
38. Berland R, Wortis HH. Origins and functions of B-1 cells with notes on the role of CD5. *Annu Rev Immunol* 2002;**20**:253-300.
39. Fehr T, Rickert RC, Odermatt B, Roes J, Rajewsky K, Hengartner H, Zinkernagel RM. Antiviral protection and germinal center formation, but impaired B cell memory in the absence of CD19. *J Exp Med* 1998;**188**:145-55.
40. Wang Y, Carter RH. CD19 regulates B cell maturation, proliferation, and positive selection in the FDC zone of murine splenic germinal centers. *Immunity* 2005;**22**:749-61.
41. Barrington RA, Zhang M, Zhong X, Jonsson H, Holodick N, Cherukuri A, Pierce SK, Rothstein TL, Carroll MC. CD21/CD19 coreceptor signaling promotes B cell survival during primary immune responses. *J Immunol* 2005;**175**:2859-67.
42. van Noesel CJ, Lankester AC, van Lier RA. Dual antigen recognition by B cells. *Immunol Today* 1993;**14**:8-11.
43. Brooks SR, Kirkham PM, Freeberg L, Carter RH. Binding of cytoplasmic proteins to the CD19 intracellular domain is high affinity, competitive, and multimeric. *J Immunol* 2004;**172**:7556-64.

VI

Replication History of B lymphocytes Reveals Homeostatic Proliferation and Extensive Antigen-induced B-cell Expansion

Menno C. van Zelm,[1,2] Tomasz Szczepański,[1,3] Mirjam van der Burg,[1] and Jacques J.M. van Dongen[1]

[1]Erasmus MC, Department of Immunology and [2]Department of Pediatrics, 3015 GE Rotterdam, the Netherlands; [3]Department of Pediatric Hematology and Oncology, Silesian Medical Academy, 41-800 Zabrze, Poland

J Exp Med 2007;204(3):645-55

Copyright 2007 Rockefeller University Press

ABSTRACT

The contribution of proliferation to B-lymphocyte homeostasis and antigen responses is largely unknown. We quantified the replication history of mouse and human B-lymphocyte subsets by calculating the ratio between genomic coding joints and signal joints on kappa-deleting recombination excision circles (KREC) of the *IGK*-deleting rearrangement. This approach was validated with *in vitro* proliferation studies. We demonstrate that naive mature B lymphocytes, but not transitional B lymphocytes, undergo *in vivo* homeostatic proliferation in the absence of somatic mutations in the periphery. T-cell dependent B-cell proliferation was substantially higher and showed higher frequencies of somatic hypermutation than T-cell independent responses, fitting with the robustness and high affinity of T-cell dependent antibody responses. More extensive proliferation and somatic hypermutation in antigen-experienced B lymphocytes from human adults compared to children indicated consecutive responses upon additional antigen exposures. Our combined observations unravel the contribution of proliferation to both B-lymphocyte homeostasis and antigen-induced B-cell expansion. We propose an important role for both processes in humoral immunity. These new insights will support the understanding of peripheral B-cell regeneration after hematopoietic stem cell transplantation or B cell-directed antibody therapy, and the identification of defects in homeostatic or antigen-induced B-cell proliferation in patients with common variable immunodeficiency or another antibody deficiency.

KEY WORDS

B-cell development, V(D)J recombination, B-cell memory, Homeostatic Proliferation, KREC.

INTRODUCTION

B lymphocytes appear in peripheral blood and secondary lymphoid organs after they have differentiated from precursor B cells in bone marrow[1] and have created a unique B-cell antigen receptor by V(D)J rearrangement of their Ig heavy and Ig light chain loci.[2] In the periphery, multiple B-cell subsets can be identified, which are either naive or have been exposed to antigen.

The mouse spleen has been extensively studied for B-cell maturation (for review see reference 3). Although several models exist,[4, 5] the general consensus is that recent bone marrow emigrants are $AA4.1^+$ and defined as transitional B cells. Subsequently, these cells can give rise to either naive follicular B cells, or marginal zone (MZ) B cells, which are named based on their anatomical location. MZ B cells show a rapid and predominantly T-cell independent response to pathogens, whereas follicular B cells show a T-cell dependent response in a germinal center reaction, where they undergo somatic hypermutation (SHM) and class-switch recombination (CSR) to generate an optimized high affinity antibody response.

In man, the recent bone marrow emigrants, i.e., the transitional B lymphocytes, are $CD38^+CD24^+IgD^+$ and represent a relatively small subset compared with the $CD38^-IgM^+IgD^+CD27^-$ naive mature B lymphocytes in blood and peripheral lymphoid organs.[6, 7] In peripheral blood, a substantial fraction of the naive B-lymphocyte population expresses the CD5 antigen and is prone to produce polyreactive antibodies.[8] Splenic MZ B lymphocytes in man are $IgM^+IgD^+CD27^+$. They show a rapid IgM response to blood-borne T-cell independent antigens,[9, 10] and their Ig loci carry somatic mutations.[11] Counterparts of splenic MZ B lymphocytes are found in lymph nodes, tonsils, and Peyer's patches, and they circulate in blood.[12-15] Because these cells are not anatomically located in the splenic MZ, it was suggested to refer to them as natural effector B lymphocytes.[16] T-cell dependent responsive B lymphocytes proliferate in a germinal center reaction as centroblasts ($CD38^+CD77^+IgD^-$) (for review see reference 17). These centroblasts undergo SHM before further differentiation into centrocytes ($CD38^+CD77^-IgD^-$), which are subsequently selected based on the highest affinity for their cognate antigen, undergo CSR, and give rise to either $CD27^+$ memory B lymphocytes or $CD38^{hi}$ antibody-producing plasma cells.[18] B-cell proliferation upon antigen encounter in the germinal center is a key step in the humoral immune response, which has been extensively studied,[19, 20] but the extent of antigen-induced proliferation of B lymphocytes remains largely unknown.

Recent reports suggest that in addition to antigen-induced proliferation, naive B lymphocytes can proliferate without further differentiation.[21-23] This homeostatic proliferation of naive B lymphocytes has been described in mouse models in response to a B-cell deficit.[22] Theoretically, the absolute number of naive B lymphocytes is regulated by four processes: (a) output from bone marrow; (b) survival; (c) loss of naive B lymphocytes via further maturation after antigen encounter;[24] and (d) antigen-independent (homeostatic) proliferation. However, similar to the extent of antigen-induced B-cell proliferation, it remains unclear to what extent proliferation contributes to naive B-lymphocyte numbers; to answer both questions, a robust assay is needed to assess the replication history of B cells.

Here we introduce the use of kappa-deleting recombination excision circles (KRECs)

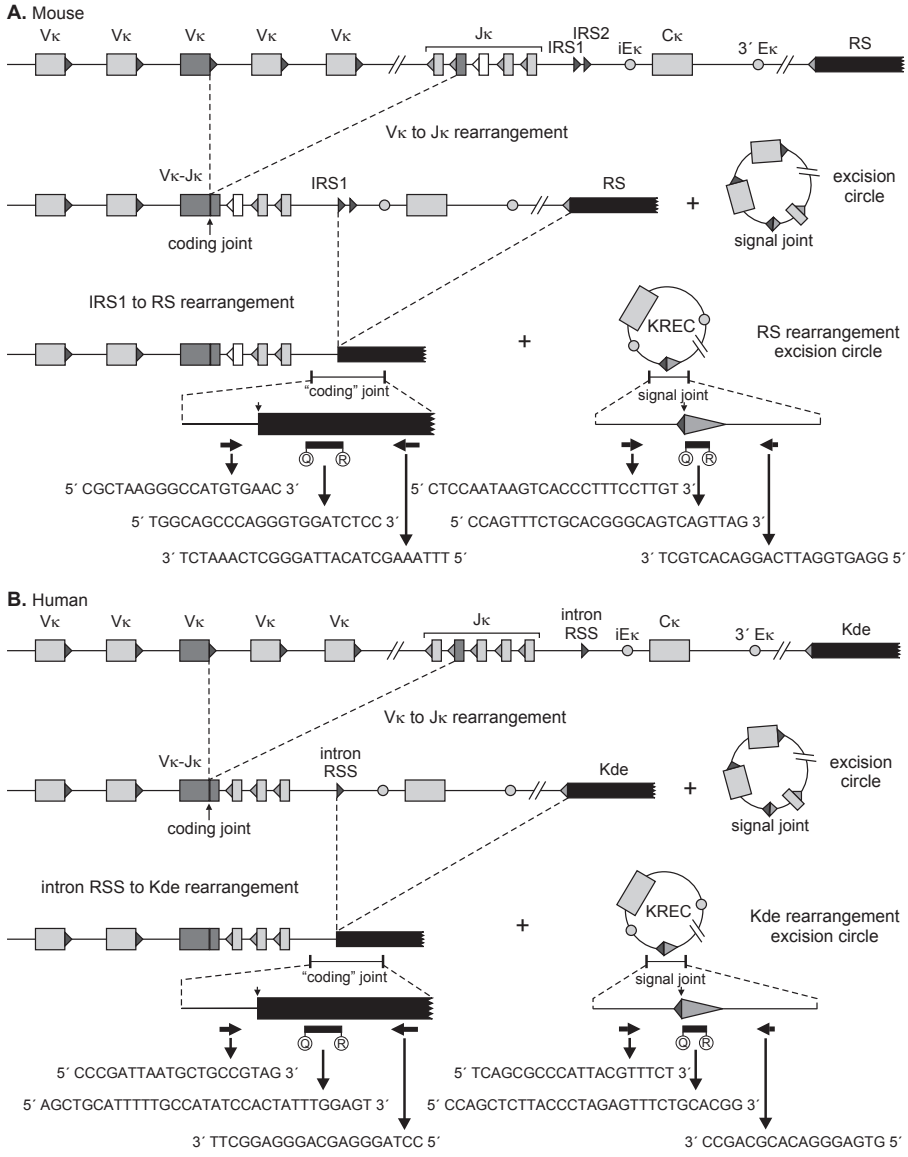


Figure 1. Detection of coding joints and signal joints of kappa-deleting rearrangements in mice (A) and man (B). V(D)J recombination on the *IGK* locus results in a Vκ-Jκ coding joint. Subsequent rearrangement between the IRS1 (mouse) or intronRSS (human) and the RS (mouse) and Kde (human) elements can make the *IGK* allele nonfunctional by deleting the Cκ exons and the enhancers. Consequently, the coding joint precludes any further rearrangements in the *IGK* locus and therefore remains present in the genome. The KREC with the corresponding signal joint is a stable double-stranded, circular DNA structure. The coding joint in the genome and the signal joint on the episomal excision circle can be quantifi ed via RQ-PCR using the indicated primers and TaqMan probes. The murine *IGK* locus contains two intronRSS sequences, IRS1 and IRS2, of which IRS1 has the most conserved RSS sequence and is >10-fold more frequently found in rearrangements to RS (reference 62 and unpublished data).

to determine the replication history of both human and mouse mature B-cell subsets. Rearrangement of the Jκ-Cκ intron recombination signal sequence (IRS1 in mouse and intronRSS in man) to the kappa-deleting element (RS in mouse and Kde in man) renders the *IGK* locus nonfunctional, precludes any further rearrangements in the *IGK* locus, and therefore the coding joint of this rearrangement remains stably present in the genome (Figure 1, A and B).[25, 26] It is a common rearrangement because it is present in 30 % of Igκ$^+$ and in almost all Igλ$^+$ mature human B lymphocytes and does not influence antigen specificity.[27, 28] The chromosomal DNA located in between the two rearranging elements is excised from the genome and circularized into stable episomal structures, which are diluted twofold by each cell division (Figure 2A).[29-32] Previously, we demonstrated the use of T-cell receptor excision circles (TRECs) for studying rearrangement and proliferation processes.[30, 31] Subsequently, TRECs of δREC–ψJα rearrangements have been used as markers for recent thymic emigrants and T-cell expansion in patients treated for HIV infection.[33, 35]

In this study, we introduce the use of KRECs to accurately assess the replication history of isolated mouse and human B-cell subsets. Using this approach we demonstrate homeostatic proliferation of mature B lymphocytes, but not of transitional B lymphocytes, in healthy individuals. Furthermore, mouse MZ and human natural effector B lymphocytes, which mainly respond to T-cell independent antigens, undergo additional antigen-induced proliferation. The T-cell dependent antigen response in the germinal center was found to be much stronger with regard to both proliferation and SHM.

RESULTS

The KREC assay is a robust technique to determine the replication history of B cells

The ratio between the kappa-deleting rearrangement and the corresponding excision circle can be used as measure for the *in vivo* replication history of an isolated B-cell subset, assuming that the excision circle is a stable DNA structure, which is diluted twofold in every cell division (Figure 2A). Consequently, real-time quantitative (RQ)-PCR assays were designed for detection of the coding joints and the corresponding signal joints of both the mouse IRS1–RS (Figure 1A) and the human intronRSS–Kde (Figure 1B) rearrangements. The two mouse RQ-PCR assays were tested for equal efficiency on serial dilutions of constructs containing either the coding joint or the signal joint. To minimize the chance of contamination in potential future clinical studies, a special cell line control was created for the human KREC studies. The Igκ$^+$ B-cell line U698-M was selected because it has two Vκ-Jκ–rearranged *IGK* alleles, one of which is out-of-frame and contains an intronRSS–Kde rearrangement (Figure 3, A and B). To obtain an equal ratio between the intronRSS–Kde coding and signal joints, an intronRSS–Kde signal joint construct was inserted into the genome of the U698-M cell line using retroviral transduction. Individual clones were sorted, and Southern blot analysis revealed a single insertion per genome for three clones: DB01, DB11, and DE09 (Figure 3B). Each of these stable B-cell lines contained both the coding joint and the corresponding signal joint of the intronRSS–Kde rearrangement and were therefore used as positive control for assessment of the efficiency of both human RQ-PCR assays.

The stability and twofold dilution of excision circles were investigated by *in vitro*

proliferation experiments. Resting B lymphocytes were isolated from wild-type mouse spleen, labeled with CFSE, and cultured with IL-4, LPS, and anti-IgM for 2-3 d. DNA was isolated from B cells before culture and from sorted B cells that had undergone 1, 2, 3, or 4 cell divisions to quantify the kappa-deleting rearrangement and the excision circle. The combined data from three individual experiments show that resting B lymphocytes contain a coding joint to signal joint ratio of 1:1 ($\Delta C_T = 0$). Furthermore, the ΔC_T ($C_{T_{coding\ joint}} - C_{T_{signal\ joint}}$) of the RQ-PCRs corresponded well with the number of cell divisions these cells underwent in culture as observed with CFSE staining (Figure 2B). Therefore, the KREC assay can be used as a robust technique to determine the replication history of B-cell subsets *in vitro*.

In vivo replication history of mouse B-cell subsets

To determine the *in vivo* replication history using the KREC assay in a well-defined system, mouse B-cell subsets ($B220^+$) from spleen and Peyer's patches were isolated (Figure 4A).

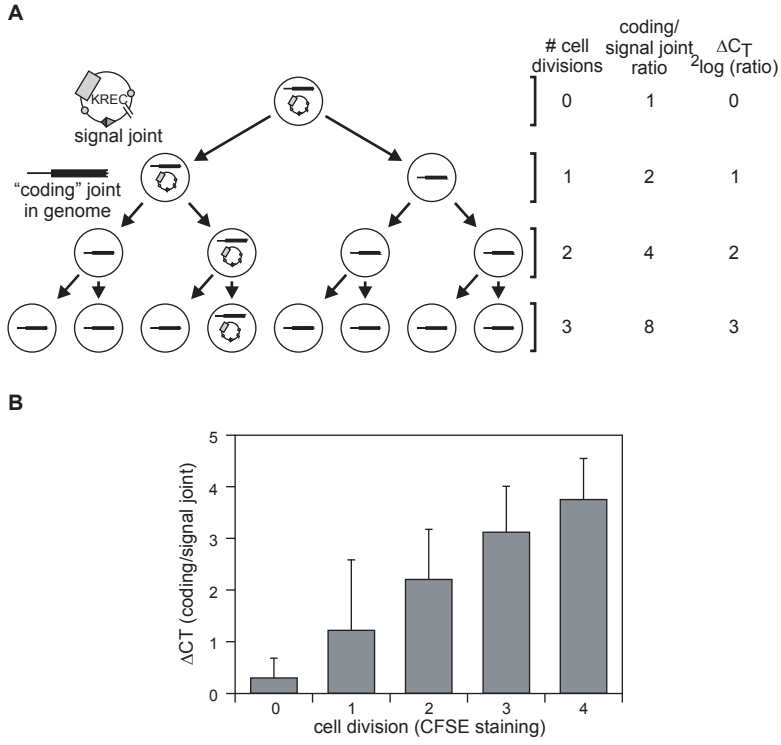


Figure 2. Quantification of the replication history of B cells using KRECs.
A. When a B lymphocyte with an intronRSS–Kde rearrangement divides, both daughter cells inherit the intronRSS–Kde coding joint in the genome. However, the signal joint, which is on the episomal KREC, will be inherited by only one of the two daughter cells. Crucially, the ΔC_T of the PCRs detecting the coding joint and the signal joint exactly represent the number of cell divisions a B lymphocyte has undergone because both processes have an exponential increase with base number 2. **B.** The ΔC_T between the coding joint and the signal joint is shown for mouse splenic B cells that were cultured *in vitro* with polyclonal stimulation and were sorted based on the CFSE staining intensity as measure for 0, 1, 2, 3, and 4 cell divisions.

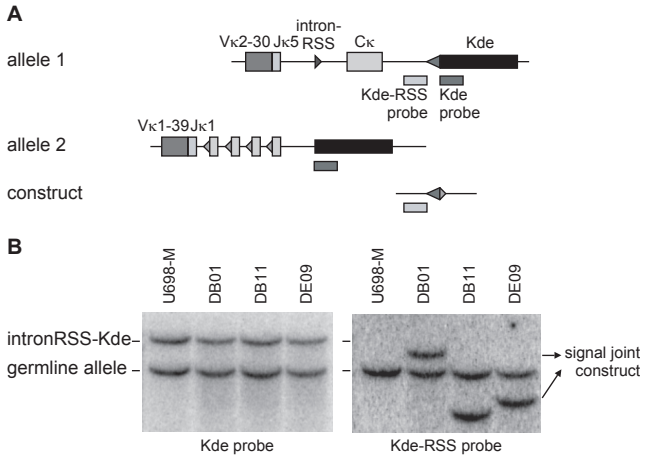


Figure 3. Generation of a human control B-cell line with one intronRSS–Kde coding joint and one artificially introduced signal joint per genome.
A. Schematic representation of the Southern blot probes that recognize the *IGK* alleles of the U698-M cell line and the KREC signal joint construct that was introduced into the cell line. The Kde probe recognizes both *IGK* alleles, whereas the Kde-RSS probe recognizes allele 1, which does not harbor an intronRSS–Kde rearrangement, and the KREC construct. **B.** Southern blot of the original U698-M cell line and three single clones that contain a unique insertion of the signal joint construct.

From spleen, T1 ($AA4.1^+CD23^-$) and T2 ($AA4.1^+CD23^+$) transitional B cells and naive follicular ($AA4.1^-CD21^{low}CD23^{hi}$) and MZ ($AA4.1^-CD21^{hi}CD23^{low}$) B lymphocytes were purified, whereas IgA^+ class-switched B lymphocytes were isolated from Peyer's patches. Total bone marrow samples were used to determine the replication history of precursor B cells, in which the Igκ-deleting rearrangement is initiated (Figure 4A).

Precursor-B cells in bone marrow and transitional B cells in spleen did not show replication (Figure 4B). In contrast, naive follicular B cells showed limited proliferation (1 cell division). MZ B cells and class-switched IgA^+ B cells showed a more extensive replication history of 4 and 5.5 cell divisions, respectively. As expected, the observed replication histories increased with more progressive differentiation stages. Consequently, the KREC assay is a powerful tool to study not only the *in vitro* replication history but also the *in vivo* replication history of sorted B-cell subsets.

The extent of antigen-induced B-cell proliferation in man

The majority of the intronRSS–Kde rearrangements in man is initiated in small pre-B-II cells.[27] Therefore, B-cell subsets starting with the small pre-B-II stage were isolated from bone marrow, tonsil, and peripheral blood using previously established markers (Figure 5A and Figure 6A) to quantify intronRSS–Kde coding joints in the genome and the intronRSS–Kde signal joints on the KRECs.[6, 15, 18, 27, 36] The coding joints were easily detected; 35–50 % of the alleles in all peripheral B-cell subsets carried this rearrangement (not depicted). In small pre-B-II and immature B cells the ΔC_T was 0.4 (Figure 5B). This indicates that

precursor-B cells in human bone marrow hardly proliferate after rearrangement of intronRSS to Kde.

Transitional B lymphocytes in tonsil did not show replication, whereas naive mature B lymphocytes in tonsil showed a replication history of 2.5 cell divisions (Figure 5B), indicating a substantial proliferation in this subset. Natural effector B lymphocytes (counterparts of splenic MZ B lymphocytes) underwent more proliferation: on average 4.1 cell divisions. Centroblasts and centrocytes in the germinal center, memory B cells, and plasma cells all showed a replication history of 8 cell divisions. As expected by their nature,

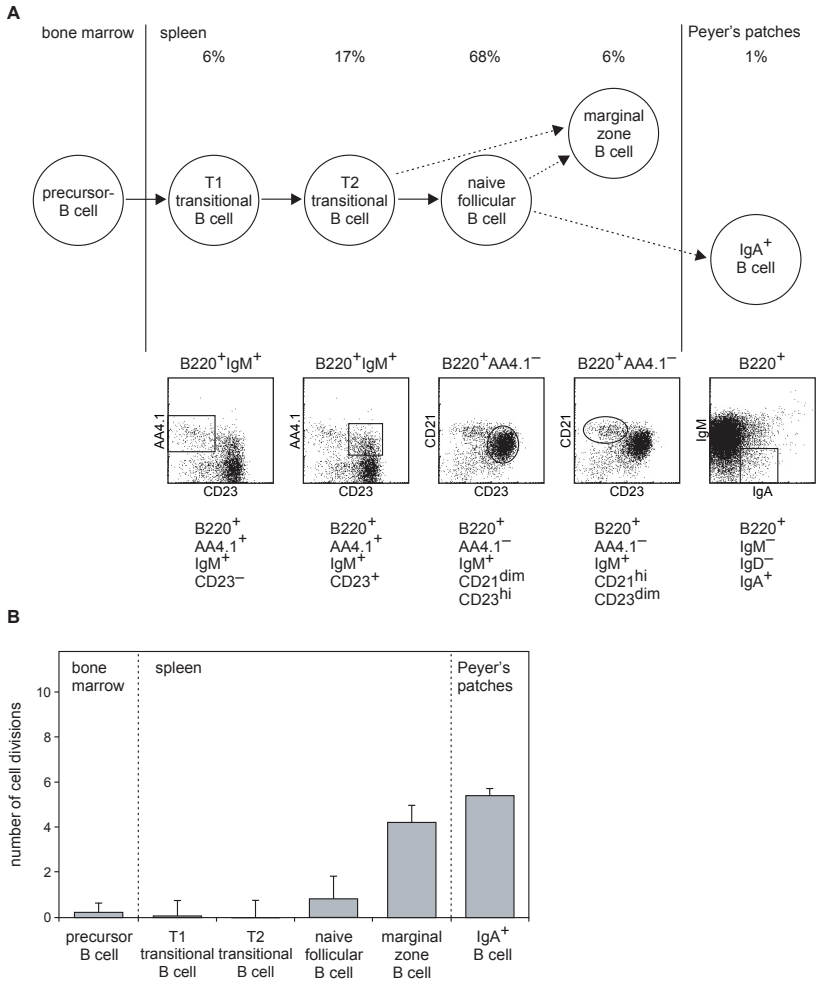


Figure 4. Replication history of mouse B-cell subsets.
A. Scheme of the differentiation stages of mature B-cell development in mice. B-cell subsets were isolated based on the indicated marker expression from spleen and Peyer's patches. **B.** The replication history of each B-cell subset. The ΔC_T between the coding joint and the signal joint PCRs represents the number of cell divisions, which are given for each subset. The data are the means of three independent sorts and are presented as mean ± SD.

the centroblasts in human tonsil form the major proliferative subset. It can thus be concluded that naive mature B lymphocytes have already proliferated but massive proliferation is found in subsequent germinal center reactions where each cell generates 45 ($2^{[8–2.5]}$) or more daughter cells. Several mature B-lymphocyte subsets circulate in blood and can be identified as counterparts of subsets found in secondary lymphoid organs (Figure 6A). The four main human B-lymphocyte subsets were isolated from blood of healthy adults and analyzed for their replication history. Naive $CD5^-$ B lymphocytes in adult peripheral blood were found to have a replication history of 1.9 (Figure 6B), which is almost similar to the total bulk of naive mature B lymphocytes in tonsils from children. Furthermore, natural effector B lymphocytes and memory B lymphocytes in adult blood were found to have a replication history of 5.6 and 10.9 cell divisions, respectively. For both populations, these

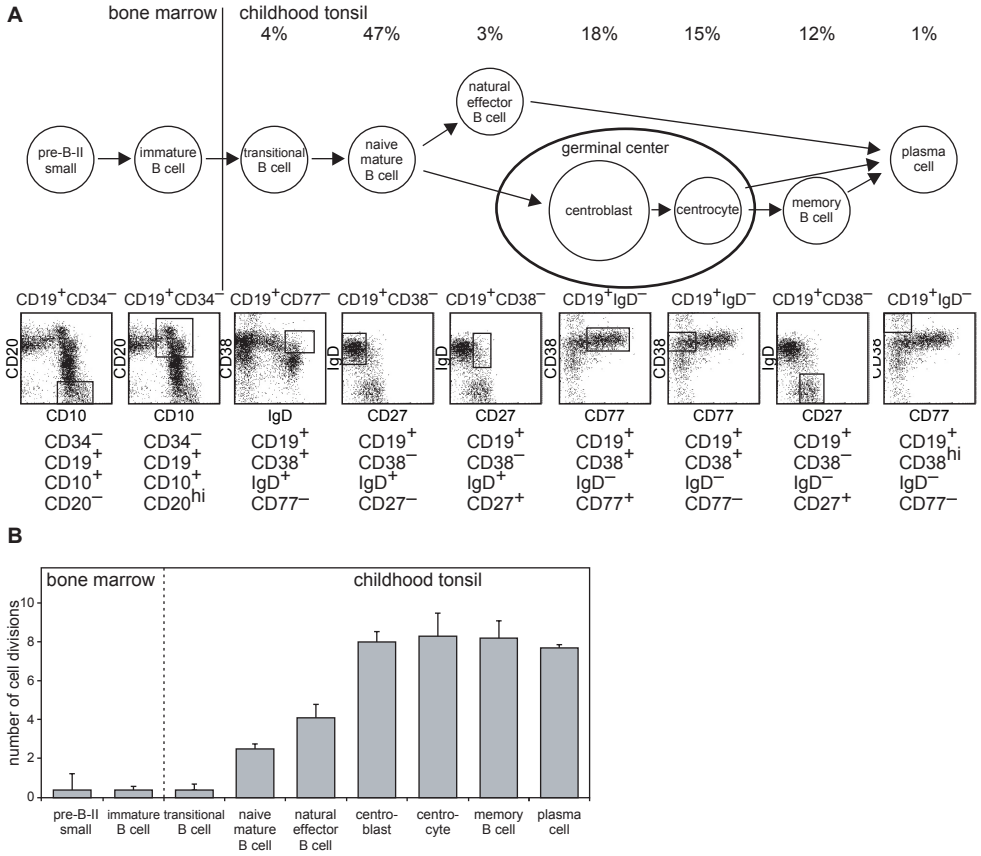


Figure 5. Replication history of human B-lymphocyte subsets from childhood tonsil.
A. Scheme of the differentiation stages of mature B-cell development. The B-cell subsets that contain the intronRSS–Kde rearrangement were isolated from bone marrow and tonsil samples of young children using the indicated markers. **B.** The replication history of each B-cell subset. The ΔC_T between the coding joint and the signal joint PCRs, which represents the number of cell divisions is given for each subset. The data are the means of independently sorted subsets from three children and are presented as mean ± SD.

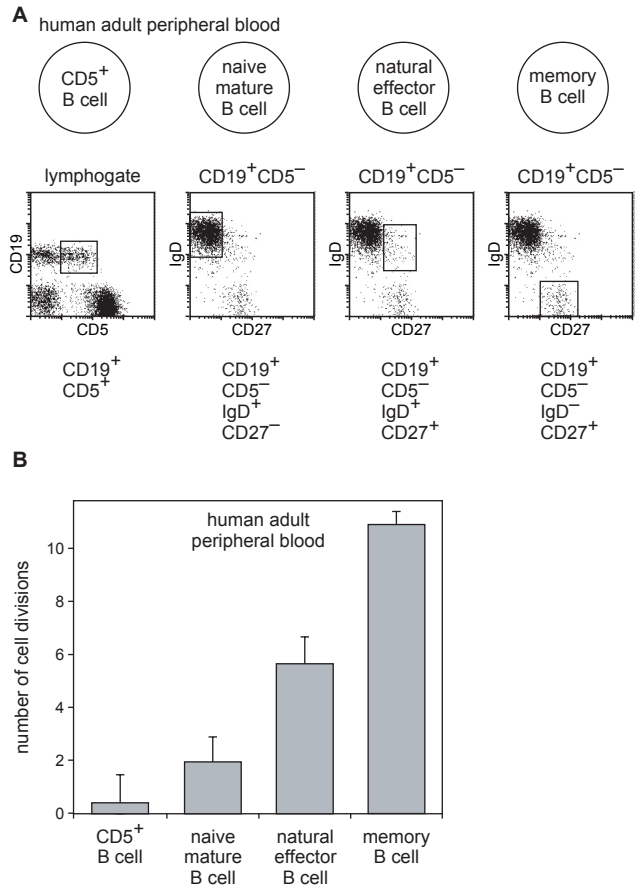


Figure 6. Replication history of the four main B-lymphocyte subsets in adult peripheral blood.
A. Schematic overview of the four main blood B-lymphocyte subsets and the markers that were used to isolate them from adult peripheral blood. **B.** The replication history of these subsets was determined as the ΔC_T of the coding joint and the signal joint PCR of subsets isolated from four or five donor samples (represented as mean ± SD).

numbers are higher than what was found in childhood tonsil. This indicates that most memory B lymphocytes in adults have been exposed to their cognate antigen multiple times and have accordingly undergone additional proliferation in consecutive germinal center reactions.

The fourth major population isolated from blood was the $CD5^+$ naive B-lymphocyte subset. This subset had a replication history of 0.4, similar to small pre-B-II cells and immature B cells in bone marrow and transitional B lymphocytes in tonsil. Thus, contrary to $CD5^-$ naive mature B lymphocytes, $CD5^+$ naive B lymphocytes do not proliferate in the peripheral B-cell compartment.

Naive B lymphocytes undergo homeostatic proliferation in healthy individuals

Using the KREC assay, we found that the mature B-lymphocyte subsets, except for the CD5$^+$ naive B lymphocytes, had undergone proliferation in the periphery. Antigen-induced proliferation is usually accompanied by the induction of SHM. To provide a quantitative measurement for the presence of SHM, we determined the presence of mutations in a hot-spot motif in rearranged *IGK* (Vκ–Jκ) alleles involving the Vκ3-20 gene segment using a modified version of the Igκ restriction enzyme-based hot-spot mutation assay (IgκREHMA; Figure 7, A–C).[37]

Neither transitional and naive mature B lymphocytes from tonsil, nor CD5$^+$ and CD5$^-$ naive B lymphocytes from blood carried mutated Vκ3-20 to Jκ rearrangements (<1.5 %; Figure 7, D and E). In contrast, germinal center and memory B-lymphocyte subsets contained

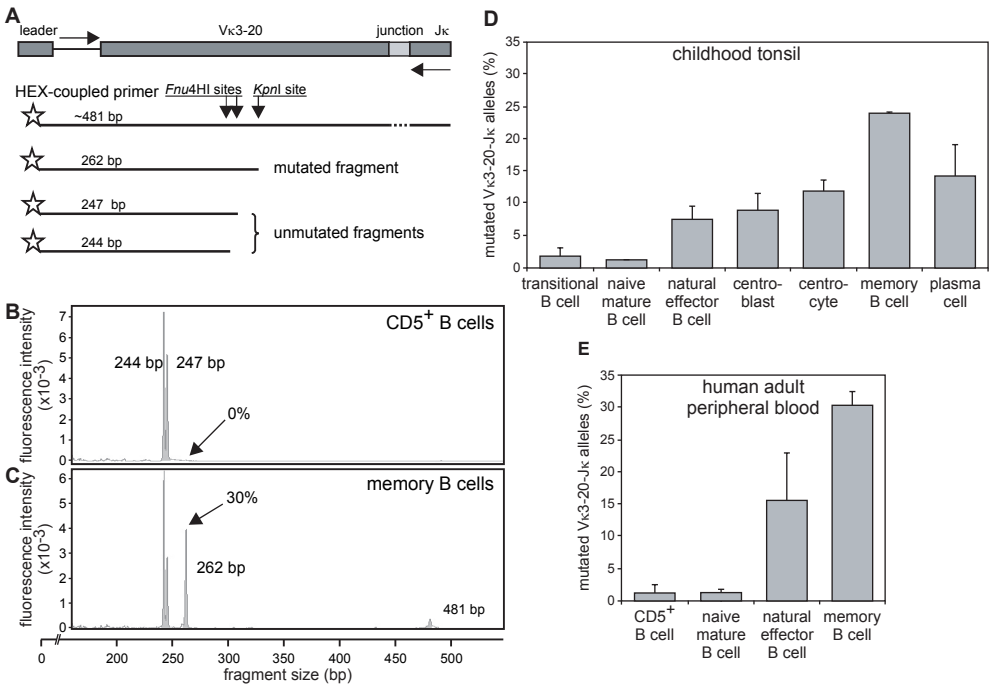


Figure 7. Quantification of SHM in human B-cell subsets.
The IgκREHMA assay was modified from Andersen et al. for use on genomice DNA.[37] **A.** Schematic representation of a Vκ3-20 to Jκ rearrangement on genomic DNA. The Vκ3-20 gene segment contains two Fnu4HI restriction sites in a SHM hot-spot in the CDR1 region. In addition, a KpnI site is located downstream of the Fnu4HI sites in the FR2 region. Vκ3-20–Jκ rearrangements were PCR amplified with one HEX-labeled Vκ3-20-intron forward and two Jκ reverse primers that recognize all five Jκ gene segments.[61] Digestion of these PCR products with both Fnu4HI and KpnI resulted in unmutated fragments of 244 and 247 bp and mutated fragments of 262 bp. **B and C.** Spectratyping of Vκ3-20–Jκ rearrangements of CD5$^+$ B lymphocytes and memory B lymphocytes after double digest with Fnu4HI and KpnI. CD5$^+$ B lymphocytes did not have a replication history and showed no SHM, whereas memory B lymphocytes in adult blood had undergone 11 cell cycles in the periphery and showed 30 % mutated Vκ3-20–Jκ rearrangements. **D and E.** The percentage of mutated Vκ3-20–Jκ rearrangements was measured in the same DNA samples that were previously used for determining the replication history (data represent means ± SD).

substantial numbers of mutated Vκ3-20 alleles. Of the rearranged Vκ3-20 alleles, 7.5 % was mutated in natural effector B lymphocytes from childhood tonsil, and 16 % in natural effector B lymphocytes from adult blood. Centroblasts and centrocytes from childhood tonsil showed 9 % and 12 % mutated rearranged Vκ3-20 alleles, respectively. Memory B lymphocytes from childhood tonsil had 24 % mutated alleles, and memory B lymphocytes from adult blood had 30 % mutated alleles. Finally, plasma cells in childhood tonsil showed 14 % mutated alleles.

It can thus be concluded that naive mature B lymphocytes in blood and tonsil undergo limited proliferation in the absence of SHM, indicative of antigen-independent homeostatic proliferation. In contrast, the proliferation of centroblasts in tonsil is associated with SHM. The frequency of mutated alleles was higher in centrocytes, memory B cells, and plasma cells, probably as the result of selection for antigen binding. Finally, the frequency of mutated alleles is higher in adult natural effector B lymphocytes and memory B lymphocytes than in their counterparts isolated from tonsils of children. This correlated with the higher proliferation rate, fitting with additional antigen-induced proliferation and SHM in adults.

DISCUSSION

In this study, we demonstrate the contribution of proliferation to B-cell homeostasis and the extent of antigen-induced B-cell expansion in mouse and man. Assessment of the replication history and the SHM status allowed us to discriminate homeostatic proliferation from antigen-induced proliferation. The homeostatic proliferation in naive B lymphocytes did not differ between human childhood and adult naive B lymphocytes, whereas the extensive proliferation and SHM of natural effector and memory B cells in adults was greater than in children, indicating additional antigen encounters in adults.

In our novel RQ-PCR-based KREC assay, we use the coding joint versus signal joint ratio of the intronRSS–Kde rearrangement to obtain highly accurate quantification of the replication history of mature B lymphocytes. This approach differs from studies in which the kinetics of the peripheral B-cell pool was assessed with BrdU or 6,6-$^{2}H_{2}$-glucose.[38, 39] The latter studies have provided insight into the recent turnover of B-cell subsets but could not quantify the contribution of proliferation to B-cell homeostasis. Because of time-lapse differences (measurement 3-4 d after deuterium glucose infusion) it was not possible to determine whether proliferation occurred in the investigated B-cell differentiation stage or in an earlier (preceding) differentiation stage.[39] The KREC assay allows us to assess the total number of cell divisions that B cells, in a well-defined subset isolated from lymphoid tissue, have undergone since their emigration from bone marrow. Consequently, differences in replication history between B-cell subsets can be straightforwardly assessed by measuring the KREC content of the B-cell subsets under study.

The applicability of KRECs depends on their stability, which has been suggested before but never formally proven.[29] In theory, the excision circles can be subject to transcriptional activity and recombinase activity in precursor-B cells because of the presence of *IGK* enhancers. Furthermore, it is unclear whether excision circles remain stably present in the nucleus of one of the two daughter cells after cell division. We show here that there is very little or no breakdown of KRECs in precursor-B cells and that the KREC content is diluted twofold with every cell division.

The use of KRECs to measure the replication history of B-cell subsets is analogous to the use of δREC–ψJα rearrangements and their corresponding TRECs to study T-cell dynamics. Although the use of TRECs provided insight into the effects of novel antiviral therapies on the thymic output in patients with HIV infection,[33, 34] TRECs could not be used to quantify proliferation because most δREC–ψJα coding joints are deleted from the genome by subsequent Vα–Jα rearrangements.[30, 31] Therefore, TREC data as measured in T lymphocytes of healthy and diseased individuals are difficult to interpret.[31, 35] Our KREC assay does not have these limitations. In contrast, the intronRSS–Kde rearrangement has several advantages for studying the replication history of B lymphocytes: (a) it is a frequently occurring *IGK* gene rearrangement in both normal and malignant B cells;[27, 28] (b) it is one of the last Ig gene rearrangements in bone marrow-derived B cells before obtaining a functional Ig molecule,[27] ensuring that the corresponding KRECs are abundantly present in naive B lymphocytes; (c) it is a single-step rearrangement, which allows easy design of RQ-PCR primers and probes for accurate detection of the coding joints and signal joints;[40] and (d) it is an end-stage rearrangement, precluding further rearrangements,[41] which implies that the coding joint is retained in the genome while the corresponding signal joint on the KREC is diluted twofold by each cell division. These characteristics make the intronRSS–Kde rearrangement a solid RQ-PCR target for accurate quantification of mature B-lymphocyte proliferation.

The intronRSS–Kde rearrangement is initiated in precursor-B cells in bone marrow and is frequently found to inactivate functionally rearranged *IGK* loci, suggesting actively induced receptor editing in bone marrow.[42] Furthermore, receptor revision has been speculated to occur in peripheral B lymphocytes, based on low levels of *RAG* gene transcription and the presence of double-stranded DNA breaks in Jκ gene segments.[43-48] If intronRSS–Kde rearrangements would be induced in the process of receptor revision in peripheral B lymphocytes, the observed number of cell divisions measured with KRECs should be an underestimate. However, because the *RAG* mRNA and double-stranded DNA break levels are clearly lower in peripheral B-cell subsets compared with bone marrow and *RAG* expression is inhibited by BCR signaling,[49, 50] the influence of receptor revision on measuring proliferation using KRECs will be limited.

The contribution of proliferation to B-lymphocyte homeostasis in healthy individuals has never been shown. Using the KREC assay we were able to address this issue. We found no replication history for transitional B lymphocytes in mouse spleen and childhood tonsil. In addition, no replication was seen in the $CD5^+$ B-cell subset in human peripheral blood. Recently, it has been shown that nearly all human transitional B cells express CD5, whereas only a fraction of the naive mature B lymphocytes express CD5.[7] Thus, the $CD5^+$ B-lymphocyte subset we isolated from peripheral blood mainly consists of transitional B lymphocytes. Consequently, the limited proliferation seen in $CD5^+$ B lymphocytes is likely to be caused by the high abundance of excision circles from transitional B lymphocytes. Therefore, it remains unclear whether the absence of proliferation in the $CD5^+$ B-cell subset reflects a difference in maturation stage or is a direct effect of CD5 expression.

Naive follicular B lymphocytes in mice and naive mature B lymphocytes obtained from human blood and tonsil showed proliferation in the absence of SHM. By definition, this is homeostatic proliferation, which occurred in the periphery, because no replication history was found for precursor-B cells and transitional B lymphocytes. Homeostatic proliferation has been observed before in naive T and NK cells.[51-54] It has been suggested to occur in B lymphocytes

also,[23] but it was only observed in response to a B-cell deficit.[21, 22] Our results show for the first time that homeostatic proliferation of naive mature B lymphocytes occurs in healthy individuals and contributes greatly to the homeostasis of absolute B-lymphocyte numbers by two- to fourfold expansion of the naive B-lymphocyte pool. The factors driving this homeostatic proliferation need to be identified in future studies.

Splenic MZ B lymphocytes in mice showed more replications than naive follicular B lymphocytes. In absence of B-cell influx from bone marrow, MZ B lymphocytes persist much longer than naive follicular B lymphocytes.[55] Our results support earlier suggestions that this persistence partly results from an increased proliferation capacity. It remains unclear, however, whether this proliferation is antigen induced, especially because these cells carry nonmutated Ig genes.[10] In contrast, human MZ B lymphocytes have a memory phenotype ($CD27^+$) and carry mutated Ig genes generated mainly in response to T-cell independent antigens.[9, 10, 15] We found that the replication history of their counterparts in childhood tonsil, natural effector B lymphocytes, was comparable to that of mouse MZ B lymphocytes. Studies in children younger than 2 yr suggest that natural effector B lymphocytes undergo SHM during ontogeny.[15] This observation supports the idea that similar to mouse human MZ B lymphocytes are generated independent of antigen. However, in this study we observed increased proliferation and SHM in natural effector B lymphocytes from adult blood compared with childhood tonsil, consistent with antigen-induced maturation. Consequently, more insight into the induction of SHM is needed to solve the question whether MZ B lymphocytes and natural eff ector B lymphocytes in man are generated dependently or independently of antigen.

The T-cell dependent response of B lymphocytes in a germinal center reaction is thought to be more powerful than the T-cell independent response. Our analyses indeed consistently showed that the germinal center response was greater with regard to proliferation and SHM. Interestingly, the replication history of germinal center-derived B cells differed greatly among mice, children, and adults. A plausible explanation is the exposure to antigen. Mice that are 5-12 wk old and kept under specific pathogen-free conditions are exposed to a limited load and spectrum of antigens and likely have less extensive B-cell maturation. Furthermore, the increased proliferation of memory B lymphocytes from adults compared with children correlated with increased frequencies of mutated *IGK* alleles. This could reflect additional maturation resulting from secondary and tertiary immune responses to a specific antigen, or additional activation over time induced by innate stimuli acting on innate antigen receptors, such as Toll-like receptors.[56]

In this study, we introduced the robust KREC assay to quantify the replication history of mouse and human B-lymphocyte subsets *in vivo*. We observed similar trends in both species. In combination with the IgκREHMA assay to determine SHM, this resulted in new and quantitative insights into homeostatic proliferation of naive mature B lymphocytes and the extent of the humoral immune response in healthy individuals. Importantly, the techniques and insights demonstrated in this study provide a firm basis for clinical evaluation of the regeneration of the peripheral B-cell compartment after hematopoietic stem cell transplantation or B cell-directed antibody therapy. Furthermore, these new insights will contribute to the understanding of defects in homeostatic or antigen-induced B-cell proliferation in patients suffering from common variable immunodeficiency or another antibody deficiency.

MATERIALS AND METHODS

Mice, B-cell culture, and B-cell subset isolation

The animal studies were approved by the Institutional Animal Care and Use Committees of the Erasmus University Rotterdam and the University of California, San Diego.

Resting B cells were purified from total splenocytes of young adult C57BL/6 mice (5-12 wk) using the MACS B-cell isolation kit (Miltenyi Biotec), labeled for 10 min at 37°C with 7.5 μM CFSE, and cultured for 2-3 d in RPMI medium containing 10 % fetal calf serum, 200 U/ml penicillin, 200 μg/ml streptomycin, 4 mM l-glutamine, and 50 μM β-mercaptoethanol supplemented with IL-4, anti-IgM, and LPS at 1 ng/ml.

B-cell subsets were purified from spleen and Peyer's patches using a FACSAria cell sorter (BD Biosciences). The following monoclonal antibodies were used: CD21-FITC (7G6), CD23-PE (B3B4), B220-PerCP (RA3-6B2), IgM-biotin (II/41), IgA-biotin (C10-1; all from BD Biosciences), AA4.1-APC, streptavidin PE-cy7, IgM-APC (II/41; all from eBiosciences), and IgD-PE (11–26; Southern Biotechnology Associates, Inc.).

Isolation of (precursor-)B-cell subsets from human bone marrow, peripheral blood, and tonsil

All bone marrow, peripheral blood, and tonsil samples were obtained with informed consent and according to the guidelines of the Medical Ethics Committee of Erasmus MC.

Precursor B cells were obtained from freshly isolated bone marrow samples of three healthy children (age 3-16) as described previously.[27]

Four B-cell subsets were purified from blood samples of five healthy adults using a FACSDiVa cell sorter (BD Biosciences) after staining of post-Ficoll mononuclear cells that were MACS sorted using CD19 beads (Miltenyi Biotec) with CD27-FITC (LT27; Serotec), IgD-PE, CD19-PE-Cy7 (SJ25C1), and CD5-APC (L17F12; all from BD Biosciences).

Seven tonsillar B-lymphocyte subsets were sorted from tonsils of three children on a FACSDiVa cells sorter after direct staining on freeze-thawed material.[17] Additional monoclonal antibodies used were CD77-FITC (5B5) and CD38-APC (HB7; both from BD Biosciences).

All fractions were obtained with a purity of >95 %.

Generation of control constructs and control cell line for the KREC assay

Mouse kappa-deleting coding and signal joints were PCR amplified from total bone marrow and cloned into the pGEMT-easy vector (Promega). Efficiency was tested on dilution series of the construct. A pGEMT-easy RQ-PCR was performed to correct for input.

A human KREC signal joint was PCR amplified in two parts from bone marrow mononuclear cell DNA to introduce an EcoRI restriction site 60 bp upstream of the Kde RSS heptamer sequence, and subsequently cloned into the retroviral LZRS-IRES-eGFP vector (Figure 3A). The LZRS-KREC construct was transfected into the Phoenix amphotropic packaging cell lines using Fugene-6 (Roche Molecular Biochemicals). Stable hightiter producer clones were selected with puromycin (1 μg/ml). The U698-M pre–B-cell line was cultured for several days in RPMI 1640 medium containing 10 % FCS and antibiotics before transduction using Retronectin coated Petridishes (Takara) and recombinant retrovirus containing supernatant for 2 d, with daily replenishing of retroviral supernatant. GFP-positive cells were single cell

sorted using a FACSDiVa cell sorter. Individual clones were selected for dim GFP expression, suggesting a single genomic integration.

Southern blot analysis of individual U698-M KREC cell line clones

DNA isolation and Southern blot analysis was performed as previously described.[57] The configuration of the *IGK* locus was determined using ^{32}P-labeled probes: the *IGK*-Kde probe is specific for the Kde region,[58] and the Kde-RSS probe is specific for the region upstream of Kde, which is deleted from the genome after an intronRSS–Kde rearrangement and are present on the KREC construct (Figure 3B).

RQ-PCR and calculation of the replication history of B-cell populations

Primers and probes were designed to specifically amplify the IRS1–RS and intronRSS–Kde rearrangements (coding joint) and the corresponding signal joint using TaqMan-based RQ-PCR from DNA isolated from cell lines and primary material (Figure 1, A and B). The RQ-PCR mixture of 25 μl contained TaqMan Universal MasterMix (Applied Biosystems), 900 nM of each primer, 100 nM of each FAM-TAMRA-labeled probe, 50 ng of DNA, and 0.4 ng BSA, and was run on the ABIPRISM 7700 sequence detection system (Applied Biosystems).[59, 60] The primer-probe sets were tested for comparable efficiencies using dilution series of plasmid constructs (for mouse reagents), and DNA was isolated from the U698-M clones DB01, DB11, and DE09, which contain one intronRSS–Kde rearrangement and one KREC construct per genome (for human reagents).

In all experiments, the cycle threshold (C_T) was set at 0.03, and the C_T values of the coding joint and the signal joint were compared for each sample. Because both PCR amplification and cell division are exponential multiplication processes with base 2, the ΔC_T ($C_{T_{intronRSS–Kde}} - C_{T_{KREC}}$) from a given cell fraction represents the mean number of cell divisions (Figure 2 A).

SHM analysis using a Vκ3-20–specific restriction enzyme hot-spot mutation assay (IgκREHMA) on genomic DNA

To investigate the occurrence of SHM in specific B-cell subsets, we modified the IgκREHMA assay developed by Andersen et al.[37] for use on genomic DNA (Figure 7A; Appendix 12).

In short, a PCR reaction was performed using a HEX-coupled Vκ3-20 intron forward primer and two FAM-coupled Jκ reverse primers recognizing all five Jκ gene segments.[61] The PCR products (500 bp) were digested with KpnI and Fnu4HI and run on a Capillary Sequencer ABI 3100 (Applied Biosystems). The Fnu4HI restriction enzyme recognizes two adjacent sites in the unmutated gene product in the SHM hot-spot region of CDR1. Unmutated gene products can therefore be visualized as 244 or 247-bp HEX-coupled fragments. KpnI cuts the gene products in FR2 downstream of the Fnu4HI sites, resulting in a 262-bp HEX-coupled mutated fragment. The unmutated B-cell line CLL-1 was used as a positive control for complete digestion with Fnu4HI. The digests hardly contained undigested gene products of 481 bp, indicating complete digestion by KpnI.

ACKNOWLEDGEMENTS

We thank Dr. C. Murre, Dr. R.W. Hendriks, and Mr. A. van Oudenaren for support with murine studies, Dr. L.J. Hoeve (Erasmus MC-Sophia, Rotterdam, the Netherlands) for providing childhood tonsil material, Mr. E.F.E. de Haas for high-speed cell sorting, Ms. E.J. van Gastel-Mol for Southern blot analysis, Mrs. W.M. Comans-Bitter for assistance in preparing the figures, and Drs. R. Benner and J.D. Laman for critically reading the manuscript.

This work was supported by grants from the foundation "Sophia Kinderziekenhuis Fonds" (grant 349 to M.C. van Zelm, R. de Groot, and J.J.M. van Dongen) and the Dutch Organization for Scientific Research (NWO/ZonMw veni grant 916.56.107 to M. van der Burg).

The authors have no conflicting financial interests.

REFERENCES

1. Ghia P, ten Boekel E, Rolink AG, Melchers F. B-cell development: a comparison between mouse and man. *Immunol Today* 1998;**19**:480-5.
2. Tonegawa S. Somatic generation of antibody diversity. *Nature* 1983;**302**:575-81.
3. Thomas MD, Srivastava B, Allman D. Regulation of peripheral B cell maturation. *Cell Immunol* 2006;**239**: 92-102.
4. Loder F, Mutschler B, Ray RJ, Paige CJ, Sideras P, Torres R, Lamers MC, Carsetti R. B cell development in the spleen takes place in discrete steps and is determined by the quality of B cell receptor-derived signals. *J Exp Med* 1999;**190**:75-89.
5. Allman D, Lindsley RC, DeMuth W, Rudd K, Shinton SA, Hardy RR. Resolution of three nonproliferative immature splenic B cell subsets reveals multiple selection points during peripheral B cell maturation. *J Immunol* 2001;**167**:6834-40.
6. Sims GP, Ettinger R, Shirota Y, Yarboro CH, Illei GG, Lipsky PE. Identification and characterization of circulating human transitional B cells. *Blood* 2005;**105**:4390-8.
7. Cuss AK, Avery DT, Cannons JL, Yu LJ, Nichols KE, Shaw PJ, Tangye SG. Expansion of functionally immature transitional B cells is associated with human-immunodeficient states characterized by impaired humoral immunity. *J Immunol* 2006;**176**:1506-16.
8. Casali P, Burastero SE, Nakamura M, Inghirami G, Notkins AL. Human lymphocytes making rheumatoid factor and antibody to ssDNA belong to Leu-1$^+$ B-cell subset. *Science* 1987;**236**:77-81.
9. Spencer J, Perry ME, Dunn-Walters DK. Human marginal-zone B cells. *Immunol Today* 1998;**19**:421-6.
10. Martin F, Kearney JF. Marginal-zone B cells. *Nat Rev Immunol* 2002;**2**:323-35.
11. Dunn-Walters DK, Isaacson PG, Spencer J. Analysis of mutations in immunoglobulin heavy chain variable region genes of microdissected marginal zone (MGZ) B cells suggests that the MGZ of human spleen is a reservoir of memory B cells. *J Exp Med* 1995;**182**:559-66.
12. van den Oord JJ, de Wolf-Peeters C, Desmet VJ. The marginal zone in the human reactive lymph node. *Am J Clin Pathol* 1986;**86**:475-9.
13. Dono M, Zupo S, Augliera A, Burgio VL, Massara R, Melagrana A, Costa M, Grossi CE, Chiorazzi N, Ferrarini M. Subepithelial B cells in the human palatine tonsil. II. Functional characterization. *Eur J Immunol* 1996;**26**: 2043-9.
14. Spencer J, Finn T, Pulford KA, Mason DY, Isaacson PG. The human gut contains a novel population of B lymphocytes which resemble marginal zone cells. *Clin Exp Immunol* 1985;**62**:607-12.
15. Weller S, Braun MC, Tan BK, Rosenwald A, Cordier C, Conley ME, Plebani A, Kumararatne DS, Bonnet D, Tournilhac O, Tchernia G, Steiniger B, Staudt LM, Casanova JL, Reynaud CA, Weill JC. Human blood IgM "memory" B cells are circulating splenic marginal zone B cells harboring a prediversified immunoglobulin repertoire. *Blood* 2004;**104**:3647-54.
16. Weill, J.C. 2006. B cell memory. Abstract book of XIIth Meeting of European Society for Immunodeficiency, Budapest, Hungary. 13.

17. MacLennan IC. Germinal centers. *Annu Rev Immunol* 1994;**12**:117-39.
18. Pascual V, Liu YJ, Magalski A, de Bouteiller O, Banchereau J, Capra JD. Analysis of somatic mutation in five B cell subsets of human tonsil. *J Exp Med* 1994;**180**:329-39.
19. Wolniak KL, Shinall SM, Waldschmidt TJ. The germinal center response. *Crit Rev Immunol* 2004;**24**:39-65.
20. MacLennan IC. Germinal centers still hold secrets. *Immunity* 2005;**22**:656-7.
21. Agenes F, Freitas AA. Transfer of small resting B cells into immunodeficient hosts results in the selection of a self-renewing activated B cell population. *J Exp Med* 1999;**189**:319-30.
22. Cabatingan MS, Schmidt MR, Sen R, Woodland RT. Naive B lymphocytes undergo homeostatic proliferation in response to B cell deficit. *J Immunol* 2002;**169**:6795-805.
23. Woodland RT, Schmidt MR. Homeostatic proliferation of B cells. *Semin Immunol* 2005;**17**:209-17.
24. Agenes F, Rosado MM, Freitas AA. Peripheral B cell survival. *Cell Mol Life Sci* 2000;**57**:1220-8.
25. Siminovitch KA, Bakhshi A, Goldman P, Korsmeyer SJ. A uniform deleting element mediates the loss of kappa genes in human B cells. *Nature* 1985;**316**:260-2.
26. Inlay M, Alt FW, Baltimore D, Xu Y. Essential roles of the kappa light chain intronic enhancer and 3' enhancer in kappa rearrangement and demethylation. *Nat Immunol* 2002;**3**:463-8.
27. van Zelm MC, van der Burg M, de Ridder D, Barendregt BH, de Haas EF, Reinders MJ, Lankester AC, Revesz T, Staal FJ, van Dongen JJM. Ig gene rearrangement steps are initiated in early human precursor B cell subsets and correlate with specific transcription factor expression. *J Immunol* 2005;**175**:5912-22.
28. van der Burg M, Tumkaya T, Boerma M, de Bruin-Versteeg S, Langerak AW, van Dongen JJM. Ordered recombination of immunoglobulin light chain genes occurs at the IGK locus but seems less strict at the IGL locus. *Blood* 2001;**97**:1001-8.
29. Livak F, Schatz DG. T-cell receptor alpha locus V(D)J recombination by-products are abundant in thymocytes and mature T cells. *Mol Cell Biol* 1996;**16**:609-18.
30. Verschuren MC, Wolvers-Tettero IL, Breit TM, Noordzij J, van Wering ER, van Dongen JJM. Preferential rearrangements of the T cell receptor-delta-deleting elements in human T cells. *J Immunol* 1997;**158**:1208-16.
31. Breit TM, Verschuren MC, Wolvers-Tettero IL, Van Gastel-Mol EJ, Hahlen K, van Dongen JJM. Human T cell leukemias with continuous V(D)J recombinase activity for TCR-delta gene deletion. *J Immunol* 1997; **159**:4341-9.
32. Kong FK, Chen CL, Six A, Hockett RD, Cooper MD. T cell receptor gene deletion circles identify recent thymic emigrants in the peripheral T cell pool. *Proc Natl Acad Sci U S A* 1999;**96**:1536-40.
33. Douek DC, McFarland RD, Keiser PH, Gage EA, Massey JM, Haynes BF, Polis MA, Haase AT, Feinberg MB, Sullivan JL, Jamieson BD, Zack JA, Picker LJ, Koup RA. Changes in thymic function with age and during the treatment of HIV infection. *Nature* 1998;**396**:690-5.
34. Hazenberg MD, Otto SA, Cohen Stuart JW, Verschuren MC, Borleffs JC, Boucher CA, Coutinho RA, Lange JM, Rinke de Wit TF, Tsegaye A, van Dongen JJM, Hamann D, de Boer RJ, Miedema F. Increased cell division but not thymic dysfunction rapidly affects the T-cell receptor excision circle content of the naive T cell population in HIV-1 infection. *Nat Med* 2000;**6**:1036-42.
35. Hazenberg MD, Verschuren MC, Hamann D, Miedema F, van Dongen JJM. T cell receptor excision circles as markers for recent thymic emigrants: basic aspects, technical approach, and guidelines for interpretation. *J Mol Med* 2001;**79**:631-40.
36. Klein U, Tu Y, Stolovitzky GA, Keller JL, Haddad J, Jr., Miljkovic V, Cattoretti G, Califano A, Dalla-Favera R. Transcriptional analysis of the B cell germinal center reaction. *Proc Natl Acad Sci U S A* 2003;**100**:2639-44.
37. Andersen P, Permin H, Andersen V, Schejbel L, Garred P, Svejgaard A, Barington T. Deficiency of somatic hypermutation of the antibody light chain is associated with increased frequency of severe respiratory tract infection in common variable immunodeficiency. *Blood* 2005;**105**:511-7.
38. Deenen GJ, Kroese FG. Kinetics of B cell subpopulations in peripheral lymphoid tissues: evidence for the presence of phenotypically distinct short-lived and long-lived B cell subsets. *Int Immunol* 1993;**5**:735-41.
39. Macallan DC, Wallace DL, Zhang Y, Ghattas H, Asquith B, de Lara C, Worth A, Panayiotakopoulos G, Griffin GE, Tough DF, Beverley PC. B-cell kinetics in humans: rapid turnover of peripheral blood memory cells. *Blood* 2005;**105**:3633-40.
40. van der Velden VH, Willemse MJ, van der Schoot CE, Hahlen K, van Wering ER, van Dongen JJM. Immunoglobulin kappa deleting element rearrangements in precursor-B acute lymphoblastic leukemia are stable targets for detection of minimal residual disease by real-time quantitative PCR. *Leukemia* 2002;**16**:928-36.

41. Langerak AW, Nadel B, De Torbal A, Wolvers-Tettero IL, van Gastel-Mol EJ, Verhaaf B, Jager U, van Dongen JJM. Unraveling the consecutive recombination events in the human IGK locus. *J Immunol* 2004;**173**:3878-88.
42. Retter MW, Nemazee D. Receptor editing occurs frequently during normal B cell development. *J Exp Med* 1998;**188**:1231-8.
43. Han S, Zheng B, Schatz DG, Spanopoulou E, Kelsoe G. Neoteny in lymphocytes: Rag1 and Rag2 expression in germinal center B cells. *Science* 1996;**274**:2094-7.
44. Hikida M, Mori M, Takai T, Tomochika K, Hamatani K, Ohmori H. Reexpression of RAG-1 and RAG-2 genes in activated mature mouse B cells. *Science* 1996;**274**:2092-4.
45. Papavasiliou F, Casellas R, Suh H, Qin XF, Besmer E, Pelanda R, Nemazee D, Rajewsky K, Nussenzweig MC. V(D)J recombination in mature B cells: a mechanism for altering antibody responses. *Science* 1997;**278**:298-301.
46. Han S, Dillon SR, Zheng B, Shimoda M, Schlissel MS, Kelsoe G. V(D)J recombinase activity in a subset of germinal center B lymphocytes. *Science* 1997;**278**:301-5.
47. Hikida M, Ohmori H. Rearrangement of lambda light chain genes in mature B cells *in vitro* and *in vivo*. Function of reexpressed recombination-activating gene (RAG) products. *J Exp Med* 1998;**187**:795-9.
48. Hillion S, Saraux A, Youinou P, Jamin C. Expression of RAGs in peripheral B cells outside germinal centers is associated with the expression of CD5. *J Immunol* 2005;**174**:5553-61.
49. Hertz M, Kouskoff V, Nakamura T, Nemazee D. V(D)J recombinase induction in splenic B lymphocytes is inhibited by antigen-receptor signalling. *Nature* 1998;**394**:292-5.
50. Meffre E, Papavasiliou F, Cohen P, de Bouteiller O, Bell D, Karasuyama H, Schiff C, Banchereau J, Liu YJ, Nussenzweig MC. Antigen receptor engagement turns off the V(D)J recombination machinery in human tonsil B cells. *J Exp Med* 1998;**188**:765-72.
51. Schluns KS, Kieper WC, Jameson SC, Lefrancois L. Interleukin-7 mediates the homeostasis of naive and memory CD8 T cells *in vivo*. *Nat Immunol* 2000;**1**:426-32.
52. Khaled AR, Durum SK. Lymphocide: cytokines and the control of lymphoid homeostasis. *Nat Rev Immunol* 2002;**2**:817-30.
53. Prlic M, Blazar BR, Farrar MA, Jameson SC. *In vivo* survival and homeostatic proliferation of natural killer cells. *J Exp Med* 2003;**197**:967-76.
54. Jamieson AM, Isnard P, Dorfman JR, Coles MC, Raulet DH. Turnover and proliferation of NK cells in steady state and lymphopenic conditions. *J Immunol* 2004;**172**:864-70.
55. Hao Z, Rajewsky K. Homeostasis of peripheral B cells in the absence of B cell influx from the bone marrow. *J Exp Med* 2001;**194**:1151-64.
56. Ruprecht CR, Lanzavecchia A. Toll-like receptor stimulation as a third signal required for activation of human naive B cells. *Eur J Immunol* 2006;**36**:810-6.
57. van Dongen JJM, Wolvers-Tettero IL. Analysis of immunoglobulin and T cell receptor genes. Part I: Basic and technical aspects. *Clin Chim Acta* 1991;**198**:1-91.
58. Beishuizen A, Verhoeven MA, Mol EJ, van Dongen JJM. Detection of immunoglobulin kappa light-chain gene rearrangement patterns by Southern blot analysis. *Leukemia* 1994;**8**:2228-36; discussion 37-39.
59. Moppett J, van der Velden VH, Wijkhuijs AJ, Hancock J, van Dongen JJM, Goulden N. Inhibition affecting RQ-PCR-based assessment of minimal residual disease in acute lymphoblastic leukemia: reversal by addition of bovine serum albumin. *Leukemia* 2003;**17**:268-70.
60. van der Velden VH, Hochhaus A, Cazzaniga G, Szczepanski T, Gabert J, van Dongen JJM. Detection of minimal residual disease in hematologic malignancies by real-time quantitative PCR: principles, approaches, and laboratory aspects. *Leukemia* 2003;**17**:1013-34.
61. van Dongen JJM, Langerak AW, Bruggemann M, Evans PA, Hummel M, Lavender FL, Delabesse E, Davi F, Schuuring E, Garcia-Sanz R, Van Krieken JH, Droese J, Gonzalez D, Bastard C, White HE, Spaargaren M, Gonzalez M, Parreira A, Smith JL, Morgan GJ, Kneba M, Macintyre EA. Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: Report of the BIOMED-2 Concerted Action BMH4-CT98-3936. *Leukemia* 2003;**17**: 2257-317.
62. Shimizu T, Iwasato T, Yamagishi H. Deletions of immunoglobulin C kappa region characterized by the circular excision products in mouse splenocytes. *J Exp Med* 1991;**173**:1065-72.

VII

General Discussion

The research described in this thesis was started with the ultimate goal to identify new genetic defects in patients with a primary antibody deficiency. Since antibodies are produced by plasma cells, which are terminally differentiated B cells, a primary antibody deficiency is likely to be the result of a defect in B-cell development. Therefore, the studies in this thesis aimed to characterize important checkpoints in B-cell development, and when applicable, compare normal with deficient B-cell development.

By studying the immunoglobulin (Ig) gene rearrangement status and the gene expression profile of isolated precursor-B-cell subsets, it was shown that the initiation and selection on complete V to (D)J gene rearrangements were strictly controlled during precursor-B-cell differentiation (**Chapter II**). Furthermore, the *IGH* locus was shown to be structurally organized in multi-loop subcompartments that underwent large-scale compaction when the locus was poised to rearrange (**Chapter III**). In murine pro-B cells, distal V_H gene segments and their promoter regions were repositioned in close proximity of the intronic enhancer region, providing a physical explanation for the mechanisms underlying long-range genomic interactions. Analysis of gross deletion breakpoints involving the *IGHM*, *BTK* and *Artemis* genes in primary immunodeficient patients showed association of transposon-derived repetitive elements, thereby shedding some light on the mechanism underlying these disease-causing mutations (**Chapter IV**). A novel antibody-deficiency syndrome was found to be caused by mutations in the *CD19* gene (**Chapter V**). By comparing of normal and CD19-deficient B cells, it was shown that CD19 plays a crucial role in lowering the signal threshold for B-cell antigen receptor-mediated B-cell activation. The extent of the T-cell independent and the T-cell dependent B-cell proliferative responses were determined using kappa-deleting recombination excision circles (KREC; **Chapter VI**). In addition, homeostatic proliferation of naive B lymphocytes was shown to contribute substantially to naive B-cell homeostasis.

Regulation of Ig gene rearrangements: Ig locus conformation

One of the big unresolved questions regarding B-cell differentiation concerns the highly controlled order in which Ig gene rearrangements are induced. A common machinery is responsible for V(D)J recombination of three Ig and four T-cell receptor (TCR) loci, but how is it possible that in almost every precursor-B cell the same order of rearrangements occurs: *IGH* > *IGK* > *IGL*, and the TCR genes are not rearranged?

The targeting of recombinase activity to a specific locus (or even clusters of gene segments) is controlled at multiple levels, including DNA methylation,[1] chromatin remodeling,[2] histone acetylation,[3, 4] germ-line transcription,[5] and transcription elongation.[6] Epigenetic changes to the DNA or chromatin are induced by enzymes that are specifically recruited to these regions. This local recruitment is likely mediated by transcription factors that recognize specific DNA sequences. Some factors determining the induction of Ig gene rearrangements in precursor-B cells and TCR gene rearrangements in precursor-T cells have been identified. Although two key players, E2A splice variants E12 and E47, are shared by the two lineages, additional factors have been identified to direct induction of either Ig or TCR gene rearrangements. The EBF and Pax-5 transcription factors specifically regulate Ig gene rearrangements in precursor-B cells together with E47 homodimers, whereas HEB-E47 heterodimers are important for TCR gene rearrangements. Obviously these factors are not decisive for all steps, especially considering the fact that a hierarchical order is found during precursor-B-cell differentiation in which the

different Ig loci and even individual alleles are consecutively induced to rearrange.

In **Chapter II** we identified DNA-binding proteins that were specifically upregulated during stages in which *IGH* or *IGK/IGL* gene rearrangements were induced. This specific regulation led us to suggest that these are candidate genes involved in regulation of Ig gene rearrangements. Obviously, functional studies have to be performed to confirm a direct role in induction of Ig gene rearrangements. The Klf4 transcription factor, for example, was found to be specifically upregulated in pre-B-I and small pre-B-II cells. A recent study, however, has shown that Klf4 suppresses proliferation and induces cell death in transformed precursor-B cells.[7] It is well possible that upregulation in pre-B-I and small pre-B cells does not directly regulate Ig gene rearrangements, but is required to induce arrest in the G1 phase of the cell cycle to allow for the induction of Ig gene rearrangements. Consequently, similar types of overexpression studies, or studies in (conditional) knockout mouse models could help to address this issue. Another possibility is testing the effects of the identified candidate proteins on the induction of Ig gene rearrangements in non-lymphoid cells. Previously, (combinations of) the transcription factors E2A, HEB and EBF have been shown to induce different types of Ig and TCR gene rearrangements in an embryonic kidney cell line.[8, 9] Since the efficiencies of these proteins to induce rearrangements were fairly low, the additional effect of co-expression of the newly identified candidate genes could be measured.

The exact mechanism by which DNA-binding factors mediate Ig gene rearrangement remains mainly unclear. Most importantly, the DNA is made accessible to the RAG proteins and large-scale compaction is induced to allow the joining of two distantly located gene segments. It is commonly believed that promoter and enhancer regions play crucial roles in regulating accessibility of Ig loci to RAG proteins. This has been confirmed in several murine studies. The Pax5 and Stat5a transcription factors bind the promoters of multiple VH genes segments, and induction of VH–DJH gene rearrangements was impaired in Pax5$^{-/-}$ and Stat5a$^{-/-}$ pre-B-I cells.[10, 11] Furthermore, disruption of enhancer regions in the murine *IGH* and *IGK* loci was found to reduce activation of Ig gene rearrangements.[12-16] More specifically, the E2A binding sites in the intron enhancer of *IGK* are critical for the efficiency of Vκ–Jκ rearrangements.[17] In **Chapter III** we found that the murine *IGH* locus is organized in rosette-like structures that undergo compaction in pre-B-I cells. Since co-localization of distant regions within the *IGH* locus is required for VH–DJH rearrangement, the structural organization of *IGH* might be crucial for proper efficiency. The specific repositioning in pro-B cells of VH regions in close proximity of the intronic enhancer region, led us to speculate that factors binding the promoter and enhancer regions play crucial roles in structural organization of *IGH*. Consequently, it would be interesting to see whether disruption of the intron enhancer leads to impaired large-scale compaction of an Ig locus and thereby disturbs the V(D)J recombination process. In absence of Pax5, large-scale contraction of *IGH* and rearrangement of distal VH gene segments were disturbed.[18] Therefore, it would also be interesting to study the effects of absence of Pax5 and Stat5, or genetic disruption of their target sequences in VH promoter regions on the structural organization of *IGH*. Using the small 10-kb probes, local effects can be studied, thus addressing for example whether Pax5 is involved in repositioning of the entire VH region, or only of the distal VH region.

Alternatively, transcriptional regulators might not act long-range themselves, but are repositioned in close proximity of each other due to reorganization of *IGH* in pro-B cells.

In such a scenario, the reorganization of *IGH* is induced by other factors. Likely candidates are proteins that bind matrix associating regions (MAR) in *IGH*. The gene expression profiles of precursor-B-cell subsets in **Chapter II** showed specific upregulation of transcripts encoding ARID5B, a Bright-like MAR-binding protein, in human pre-B-I cells that initiate V_H–DJ_H gene rearrangements (data not shown). Thus, this would be an interesting protein to study with respect to the *IGH* locus conformation in developing B cells.

The peripheral B-cell compartment

The precursor-B-cell compartment has been extensively studied and most identified functional processes can now be linked to phenotypically well-defined subsets. Consequently, more attention is now given to the peripheral B-cell compartment with the aim to identify B-cell subsets with unique functional characteristics. The positioning of peripheral B-cell subsets in a scheme is more challenging, because there appears to be no linear differentiation pathway. However, recent studies have provided insight in the identification of several naive and antigen-experienced B-cell subsets (Figure 1).

Naive B-cell homeostasis

Recent bone marrow emigrants are classified as transitional B lymphocytes and can be phenotypically distinguished from naive mature B lymphocytes. In **Chapter VI**, we found that naive mature B lymphocytes undergo homeostatic proliferation, whereas no proliferation was seen in recent bone marrow emigrants, i.e. transitional B lymphocytes. This homeostatic proliferation contributes greatly to the size of the naive mature B-cell compartment. The development of recent bone marrow emigrants into naive mature B lymphocytes is an important checkpoint selection against self-reactive antibodies.[19] It remains unclear whether transitional B lymphocytes undergo selection before they can enter the naive mature B-cell compartment. Transitional B lymphocytes were found to die rapidly in culture, they are unresponsive to BAFF, and although they enter cell cycle upon BCR stimulation with IL-4 and

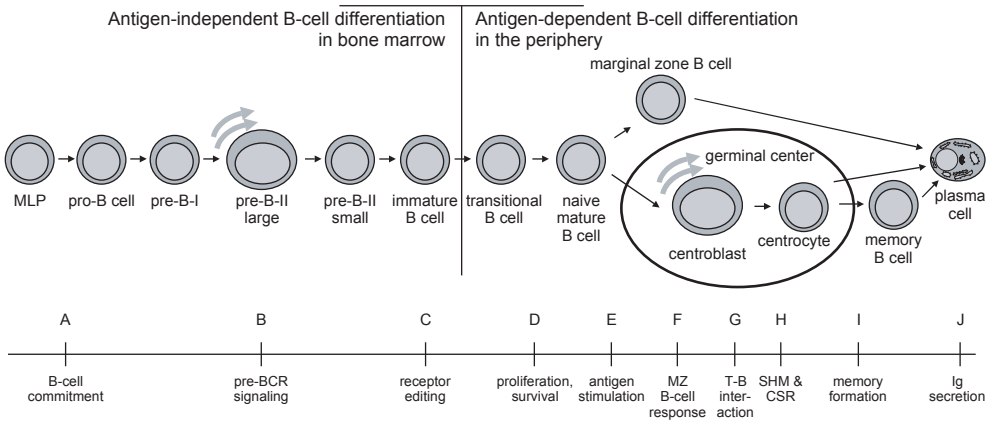


Figure 1. Important checkpoints in human B-cell differentiation.
Studies described in this thesis have contributed to the identification of 10 important checkpoints, which might be affected in patients with primary antibody deficiency.

CD40 ligation,[20] their proliferation, Ig secretion and migration *in vitro* were greatly reduced when compared to naive mature B lymphocytes.[21] Apparently, transitional B lymphocytes that survive proceed to become naive mature B lymphocytes, but considering the strong reduction of autoreactive clones, survival and/or selection is not dependent on strong BCR signals. On the other hand, the observed selection of transitional B lymphocytes into the naive mature B-cell compartment might be the result of selective proliferation of non-autoreactive cells. This would explain the reduction in autoreactivity and the observed homeostatic proliferation of naive mature B lymphocytes. The question then remains how specifically non-autoreactive transitional B lymphocytes are induced to proliferate.

Homeostatic proliferation of naive mature B lymphocytes has been observed before in mouse models in response to B-cell deficit and was independent from T-cell help.[22,23] The c-rel and p50 components of NF-κB were dispensable for this proliferation.[23] C-rel is critical for antigen and mitogen stimulation and p50 for the survival of peripheral B cells. Consequently, homeostatic proliferation occurs most likely independent from antigen or other means of BCR stimulation.

Homeostatic proliferation was found to critically depend on Btk, which does not only signal downstream of the BCR, but also other surface receptors mediating B-cell growth and proliferation (IL-5R, IL-10R, the Toll-like receptor RP-105) and migration (CD38). Consequently, there might be a role for cytokine stimulation of naive mature B lymphocytes to mediate homeostatic proliferation. In addition, the BAFF – BAFF-R interaction could be involved in induction of homeostatic proliferation. BAFF-R signals are crucial for B-cell survival, but initiate signaling and transcriptional programs that increase the metabolic activity.[24, 25] This metabolic fitness has been shown to result in a more rapid and robust response to BCR or TLR stimulation *in vitro*.[25] It is therefore likely that a combination of cytokine and BAFF-induced signaling initiates homeostatic proliferation.

During development, a B-cell undergoes consecutive stages of differentiation and proliferation. These processes are highly controlled and often E-proteins and their inhibitors (Id proteins) are involved. The E-protein dimers initiate differentiation, whereas Id proteins inhibit the E-protein function by forming E-Id heterodimers. The Id3 protein is expressed throughout B-cell differentiation except in plasma cells,[26] and B lymphocytes of $Id3^{-/-}$ mice show impaired proliferation upon stimulation with anti-IgM, LPS, anti-CD38, anti-CD40.[27] Considering the role of Id3 in peripheral B-cell proliferation, it can well be anticipated that this is one of the genes involved in the induction of homeostatic proliferation.

Theoretically, the absolute number of naive B lymphocytes is regulated by four processes: (a) output from bone marrow; (b) survival; (c) loss of naive B lymphocytes via further maturation after antigen encounter; and (d) antigen-independent homeostatic proliferation. Understanding the contribution of these four processes to naive B-lymphocyte homeostasis is of great importance for diagnosis and clinical evaluation of several patient categories. It would be interesting to determine the contribution of proliferation to regeneration of the B-cell compartment following anti-B-cell therapy (e.g. rituximab treatment) or stem cell transplantation in relation to the clinical status of the patient. Homeostatic proliferation can support the expansion of peripheral B-cell numbers, but the patient needs a B-cell compartment with a broad repertoire. Therefore, the balance between the output of B cells from bone marrow and peripheral expansion of these cells has to be tightly controlled.

Furthermore, the KREC assay can be a powerful tool to monitor the peripheral B-cell compartment in patients with a lymphoproliferative disorder. For instance, patients with XLP (X-linked lymphoproliferative disease) that carry a genetic defect in *SH2D1A* have an expansion of transitional B lymphocytes.[21, 28-30] The over-abundance of transitional B lymphocytes may contribute to the humoral deficiency that accompanies the disease. The question remains, however, whether this expansions results from extra output from bone marrow, a reduced loss of these cells, or increased proliferation.

The B-cell antigen response

The response of B lymphocytes upon antigen recognition was studied in **Chapter V** and **Chapter VI.** In **Chapter V**, we described four patients with a homozygous disruption in the *CD19* gene. *In vitro* Ca^{2+} flux and proliferation studies revealed that CD19-deficient B cells were not optimally activated upon BCR stimulation. As a result, the number of $CD27^+IgM^+IgD^+$ natural effector B cells and $CD27^+IgM^-IgD^-$ class-switched memory B cells, the Ig response to rabies vaccination and the total serum Ig levels were drastically reduced in CD19-deficient patients. This is most likely caused by a decreased number of naive B lymphocytes that are activated by antigen and/or centrocytes that are positively selected for by optimal antigen-binding capacities. These data confirm previous observations that CD19 acts in a complex with CD21, CD81 and CD225 to lower the threshold for signal transduction.[31, 32] The CD19 deficiency is the first genetic defect identified in man that specifically leads to a decreased response of B lymphocytes to antigen.

Recently, a $CD21^-$ mature B-cell subset identified in peripheral blood of rheumatoid arthritis patients was described to be anergic (E. Meffre, B cells in health and disease, Keystone symposia series, Banff, Alberta, Canada, 2007). Interestingly, mature B cells of CD19-deficient patients have decreased CD21 expression levels. Human B cells apparently become anergic when the CD19-complex, either by absence of CD19 or CD21, is disrupted.

The extent of B-cell proliferation upon antigen recognition was studied in **Chapter VI**. Two types of antigen responses can be identified. Splenic marginal zone B lymphocytes mainly undergo a fast, T-cell independent response (reviewed in ref. 33). Human marginal zone B lymphocytes carry a memory phenotype ($CD27^+$) and mutated Ig genes.[34, 35] Their counterparts, natural effector B lymphocytes, are found in lymph nodes, tonsils, Peyer's patches and blood.[36-39] Although these cells show the characteristics of antigen-experienced cells, it is currently under debate, whether these cells develop independently from antigen during ontogeny or in response to antigen.[39-41] Our data show increased proliferation and SHM in natural effector B lymphocytes from adult blood compared with childhood tonsil, consistent with antigen-induced maturation. On the other hand, we found that in young mice, kept under specific pathogen-free conditions, marginal zone B lymphocytes showed similar extensive proliferation as human natural effector B lymphocytes. Further analysis of SHM and proliferation in this subset could help to resolve this issue. Sequencing of the V_H-Cμ transcripts should reveal whether the pattern of SHM shows a molecular sign of a response to antigenic stimulation.[42] Furthermore, the KREC assay can be used to analyze whether marginal zone B lymphocytes undergo proliferation in animals that are kept completely in absence of pathogens.

B lymphocytes undergoing a T-cell dependent antigen response in the germinal center showed most proliferation (**Chapter VI**). Centroblasts, centrocytes, and germinal center-

derived class-switched memory B cells and plasma cells showed a replication history of 8 cycles in childhood tonsil. Interestingly, class-switched memory B cells in adult peripheral blood showed a clearly higher replication history of 11 cell cycles. Since this increase correlated with an increase in SHM, this is likely to be the result of additional maturation resulting from secondary and tertiary immune responses. A more detailed study on the replication history of B-cell subsets from healthy controls ranging in age from birth to 80 years would provide more insight into the build up of memory by additional proliferation in consecutive immune responses.

It is a long-standing concept that the adaptive immune system forms memory during a primary response to antigen. These cells will show a faster and more powerful response upon second contact with the same antigen, thus preventing disease and resulting in immunity. Although our results are in line with this concept and suggest additional proliferation, we were unable to demonstrate (a) the magnitude of the memory B-cell response and (b) the contribution of naive and memory B cells upon a second encounter with the same antigen (Figure 2). This multifactorial problem can probably not be solved with human studies. The use of mouse models, however, should facilitate this, because mice that are kept under specific pathogen-free conditions have only small germinal centers and limited numbers of class-switched memory B cells. Consequently, the immune response upon immunization with a specific hapten can be studied in a relatively 'clean' system. Furthermore, when using a conjugate of (4-hydroxy-3-nitrophenyl)acetyl (NP) with, for instance, keyhole limpet hemocyanin (KLH), a substantial population of antigen-specific B cells are formed.[43, 44] These cells can be identified flowcytometrically by their ability to bind 4-hydroxy-3-iodo-5-nitrophenylacetate (NIP) conjugated to PE. Thus, the replication history of these cells, in addition to other immunological parameters, can be determined after primary and secondary immunizations. The secondary response will involve both naive and memory B cells formed during the primary response. Additional experiments will have to be performed to determine the response of memory B cells. An option is to treat these mice with anti-BAFF prior to secondary immunization. All mature B-cell subsets, except for memory B cells, depend on BAFF and will go in apoptosis upon treatment with anti-BAFF (J.E. Crowley, B cells in health

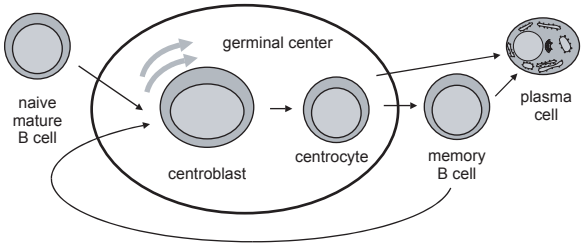


Figure 2. Model of a germinal center reaction during a secondary immune response.
Upon secondary encounter with an antigen, both naive and memory B cells can be activated and recruited in a germinal center reaction. The contribution and the extent of the primary and memory response upon secondary antigen encounter are currently unknown. Application of the KREC assay in murine immunization studies could provide more insight into the issue.

and disease, Keystone symposia series, Banff, Alberta, Canada, 2007). Thus, upon anti-BAFF treatment only memory B cells will be present for participation in the secondary response.

New approaches to identify B-cell defects in patients suffering from an antibody deficiency

In >90 % of patients that suffer from an antibody deficiency, no genetic defect has been found to underlie the disease. The majority of these patients are diagnosed as having common variable immunodeficiency (CVID) and although the presentation of the disease and the susceptibility to pathogens varies considerably, they are currently treated and studied as a single disease entity. Recently, we (**Chapter V**) and others (refs. 45-47) have identified new genetic defects in patients that were previously diagnosed as having CVID. These defects resulted in a block in peripheral B-cell development and thus showed that these molecules play crucial roles in important checkpoints in B-cell development. Previously identified genes disrupted in antibody deficiencies were found to be crucial for either pre-BCR signaling or Ig class switching in the germinal center.[48-57] We anticipate that in CVID patients, B-cell differentiation is blocked in these or other critical checkpoints.

Based on human studies in healthy controls, primary antibody deficient patients, and mouse models, we identified 10 critical checkpoints in B-cell development (step A-I in Figure 1). First, multipotent progenitor cells in bone marrow have to develop into pro-B cells. Multiple factors are required to induce specification and subsequently commitment to the B-cell lineage. The importance of this first checkpoint (step A) was demonstrated in multiple mouse models, such as the $E2A^{-/-}$ and the $EBF^{-/-}$ mice that show a complete absence of all precursor-B cells in bone marrow.[58, 59] Signaling via the pre-B cell receptor (pre-BCR) at the pre-B-II large stage is the second important checkpoint (step B) and precedes a proliferation phase.[60] A proportion of the immature B cells is autoreactive and is either deleted by apoptosis, becomes anergic, or generates an altered B-cell receptor via receptor editing (step C).[61] Immature B cells that are not autoreactive leave the BM and migrate to the periphery as transitional B lymphocytes and develop into naive mature B lymphocytes if they are able to survive and proliferate (step D). These naive mature B lymphocytes can be activated by antigen (step E) and differentiate into marginal zone or memory B cells depending on the type of antigen. Marginal zone B cells and their counterparts in peripheral blood and other lymphoid tissues clearly differ from class-switched memory B lymphocytes in their replication history and SHM load (**Chapter VI**). Furthermore, they show a fast and mainly T-cell independent response to antigen (step F; reviewed in ref. 33) A cognate B-T cell interaction is required for activation with a T-cell dependent (TD) antigen (step G) and the initiation of the affinity maturation process (step H), which aims for the generation of B cells with the highest affinity for the antigen in the germinal center response. The affinity maturation process is based on the generation of somatic hypermutations (SHM) and subsequent selection and class switch recombination to IgA, IgG or IgE. Finally, maturation into memory B cells (step I) or antibody-secreting plasma cells takes place (step J).[62] The techniques we used and the insights we obtained from the studies described in this thesis can be used to investigate possible B-cell defects in patients suffering from an antibody deficiency.

Biology of germline mutations: small and gross deletions

In **Chapter IV** we studied gross deletions in three primary immunodeficiency (PID) genes with the aim to identify underlying mechanisms involved in their induction. In contrast to micro-lesions, the molecular mechanisms mediating gross lesions are still poorly understood. Random mispairing of 2-8 bp repeats in the replication fork is a likely origin of micro-deletions and -insertions, because the involved repeats are in close proximity to each other. In contrast, gross lesions involve DNA sequences that can be hundreds of kb apart. Consequently, additional mechanisms have to be involved to bring these sequences close to each other.

In **Chapter III** we showed using distance measurements between multiple regions in the murine *IGH* locus that the locus has a very compact and organized structure in which subdomains can be identified. These data showed that genomically distant regions can be located in spatial proximity of each other in the nucleus. In **Chapter IV** we found that a high local content of transposon-derived elements was associated with increased incidence of gross deletions. There are several indications that high density of transposon-derived elements in a genomic region can alter the structural organization of that region. First, transposons greatly influence the GC/AT ratio of the genomic region in which they integrate. For instance, *Alu* repeats have a high GC content (~63 %). Second, transposon-derived elements alter the epigenetic regulatory structures. Transposon integration can trigger CpG methylation and histone deacetylation of the involved genomic region, thereby resulting in localization in heterochromatin.[63, 64] These combined observations indicate that transposon-derived elements are possible mediators of gross deletions by the co-localizing genomically distant regions.

Genetic diagnosis of primary antibody deficiencies: considerations and new approaches

In our diagnostic setting, we aim for the identification of a genetic defect with molecular methods in patients suspected to suffer from a primary immunodeficiency. A molecular diagnosis has several advantages over the clinical or immunological diagnosis.[65] First, the presence or absence of a mutation can be determined unambiguously. Second, the specific mutation may provide prognostic information. Third, molecular diagnostic testing allows the screening of family members for carriers or other affected individuals prior to the appearance of clinical symptoms. Fourth, molecular diagnosis is required for corrective treatment with gene therapy. Although gene therapy to treat primary immunodeficiencies has not been developed beyond clinical trials and has faced some setbacks,[66-70] it is still the most promising corrective treatment.

Besides the clear advantages, molecular testing also has several clear limitations. The techniques are relatively costly and laborious. Therefore, a candidate gene has to be selected before starting molecular analysis. Since the human genome contains >30,000 genes, the gene has to be carefully selected based on the clinical and immunological findings. Furthermore, it is not always clear whether the identified mutation is disease-causing. Single point mutations that alter one amino acid could be harmless polymorphisms. In such situations, the clinical and immunological findings or additional protein analysis might be required to support the genetic diagnosis. Finally, gross lesions can be very difficult to detect. The presence of a deletion is hard to confirm with a positive result (**Chapter IV**), but the detection of inversions is even more difficult. Because the genetic material is still present in the genome, PCR amplification and

sequencing will not result in abnormal findings. However, these lesions result in loss of normal transcripts. Therefore, these limitations should be taken into account and additional assays should be performed to reduce false positive or false negative results in molecular diagnosis.

Molecular diagnosis in CVID patients proves to be extremely difficult; until now in <10 % of patients a genetic defect has been identified. The difficulties result from (a) the heterogeneity of the disease and (b) the fact that in many cases not a monogenetic defect, but a complex genetic trait underlies the disease. As stated before, clinical and immunological characterization of the patient will help to guide molecular diagnostics. However, this could still generate multiple candidate genes. Consequently, a different approach using new techniques is needed for large-scale screening to identify a molecular basis of the disease in individual CVID patients.

In large families with multiple patients, linkage analysis can be performed to identify genomic regions with candidate genes. The standard method involves a whole-genome scan using 300-400 microsatellite markers spaced at ~10 Mb intervals.[71, 72] Although several monogenetic disorders were mapped with this approach, it is not optimal for the identification of disease susceptibility genes in complex traits.[73] More recently, DNA microarrays have been developed that detect >500,000 SNPs (Affymetrix). With a mean distance between markers of <8 kb, these SNP arrays can be much more powerful for linkage studies.[74] The exact strength of SNP chips in small families has to be determined, but in combination with the clinical and immunological characterization of the patient, this will further restrict the number of candidate genes that might be affected.

In addition to SNP detection, arrays have been developed to enable analysis of up to 300 kb of genomic sequence. These arrays enable sequence analysis of multiple candidate genes and can greatly facilitate molecular diagnosis. However, considering the possibility of technical errors due to the large-scale approach, identified mutations will have to be confirmed by standard sequencing methods. Furthermore, this approach mainly aims for the identification of small lesions and false negative results can be obtained when a gross lesion underlies the disease. Nevertheless, the use of these types of DNA microarrays will greatly support the identification of genetic defects in patients suffering from an antibody deficiency.

New insights in human B-cell differentiation resulting from the identification of novel monogenetic defects in patients with a primary antibody deficiency

As discussed above, our studies on normal and deficient B-cell differentiation can support future studies on and diagnostic testing of primary antibody deficiencies. The identification of an immunological defect and especially a molecular defect underlying the disease is a relevant ultimate aim of this type of research. However, we should realize that there is more: there is a lot to be learned from well-defined groups of patients. Especially when a monogenetic defect appears to cause the disease, the antibody-deficiency syndrome of a patient can be used as a model. The affected gene apparently has a critical role in maintaining immunity, and its absence results in a block in B-cell differentiation or in the antibody response. Therefore, studies on antibody-deficient patients can support our understanding of important checkpoints in normal B-cell differentiation as well as studies on normal B-cell differentiation do this to define the immunological defects in patients with an antibody deficiency.

REFERENCES

1. Mostoslavsky R, Singh N, Kirillov A, Pelanda R, Cedar H, Chess A, Bergman Y. Kappa chain monoallelic demethylation and the establishment of allelic exclusion. *Genes Dev* 1998;**12**:1801-11.
2. Maes J, O'Neill LP, Cavelier P, Turner BM, Rougeon F, Goodhardt M. Chromatin remodeling at the Ig loci prior to V(D)J recombination. *J Immunol* 2001;**167**:866-74.
3. McBlane F, Boyes J. Stimulation of V(D)J recombination by histone acetylation. *Curr Biol* 2000;**10**:483-6.
4. Chowdhury D, Sen R. Stepwise activation of the immunoglobulin mu heavy chain gene locus. *Embo J* 2001;**20**:6394-403.
5. Yancopoulos GD, Alt FW. Developmentally controlled and tissue-specific expression of unrearranged VH gene segments. *Cell* 1985;**40**:271-81.
6. Abarrategui I, Krangel MS. Regulation of T cell receptor-alpha gene recombination by transcription. *Nat Immunol* 2006;**7**:1109-15.
7. Kharas MG, Yusuf I, Scarfone VM, Yang VW, Segre JA, Huettner CS, Fruman DA. KLF4 suppresses transformation of pre-B cells by ABL oncogenes. *Blood* 2007;**109**:747-55.
8. Romanow WJ, Langerak AW, Goebel P, Wolvers-Tettero IL, van Dongen JJM, Feeney AJ, Murre C. E2A and EBF act in synergy with the V(D)J recombinase to generate a diverse immunoglobulin repertoire in nonlymphoid cells. *Mol Cell* 2000;**5**:343-53.
9. Langerak AW, Wolvers-Tettero IL, van Gastel-Mol EJ, Oud ME, van Dongen JJM. Basic helix-loop-helix proteins E2A and HEB induce immature T-cell receptor rearrangements in nonlymphoid cells. *Blood* 2001;**98**:2456-65.
10. Hesslein DG, Pflugh DL, Chowdhury D, Bothwell AL, Sen R, Schatz DG. Pax5 is required for recombination of transcribed, acetylated, 5' IgH V gene segments. *Genes Dev* 2003;**17**:37-42.
11. Corcoran AE, Riddell A, Krooshoop D, Venkitaraman AR. Impaired immunoglobulin gene rearrangement in mice lacking the IL-7 receptor. *Nature* 1998;**391**:904-7.
12. Chen J, Young F, Bottaro A, Stewart V, Smith RK, Alt FW. Mutations of the intronic IgH enhancer and its flanking sequences differentially affect accessibility of the JH locus. *Embo J* 1993;**12**:4635-45.
13. Serwe M, Sablitzky F. V(D)J recombination in B cells is impaired but not blocked by targeted deletion of the immunoglobulin heavy chain intron enhancer. *Embo J* 1993;**12**:2321-7.
14. Perlot T, Alt FW, Bassing CH, Suh H, Pinaud E. Elucidation of IgH intronic enhancer functions via germ-line deletion. *Proc Natl Acad Sci U S A* 2005;**102**:14362-7.
15. Xu Y, Davidson L, Alt FW, Baltimore D. Deletion of the Ig kappa light chain intronic enhancer/matrix attachment region impairs but does not abolish V kappa J kappa rearrangement. *Immunity* 1996;**4**:377-85.
16. Gorman JR, van der Stoep N, Monroe R, Cogne M, Davidson L, Alt FW. The Ig(kappa) enhancer influences the ratio of Ig(kappa) versus Ig(lambda) B lymphocytes. *Immunity* 1996;**5**:241-52.
17. Inlay MA, Tian H, Lin T, Xu Y. Important roles for E protein binding sites within the immunoglobulin kappa chain intronic enhancer in activating Vkappa Jkappa rearrangement. *J Exp Med* 2004;**200**:1205-11.
18. Fuxa M, Skok J, Souabni A, Salvagiotto G, Roldan E, Busslinger M. Pax5 induces V-to-DJ rearrangements and locus contraction of the immunoglobulin heavy-chain gene. *Genes Dev* 2004;**18**:411-22.
19. Wardemann H, Yurasov S, Schaefer A, Young JW, Meffre E, Nussenzweig MC. Predominant autoantibody production by early human B cell precursors. *Science* 2003;**301**:1374-7.
20. Sims GP, Ettinger R, Shirota Y, Yarboro CH, Illei GG, Lipsky PE. Identification and characterization of circulating human transitional B cells. *Blood* 2005;**105**:4390-8.
21. Cuss AK, Avery DT, Cannons JL, Yu LJ, Nichols KE, Shaw PJ, Tangye SG. Expansion of functionally immature transitional B cells is associated with human-immunodeficient states characterized by impaired humoral immunity. *J Immunol* 2006;**176**:1506-16.
22. Agenes F, Freitas AA. Transfer of small resting B cells into immunodeficient hosts results in the selection of a self-renewing activated B cell population. *J Exp Med* 1999;**189**:319-30.
23. Cabatingan MS, Schmidt MR, Sen R, Woodland RT. Naive B lymphocytes undergo homeostatic proliferation in response to B cell deficit. *J Immunol* 2002;**169**:6795-805.
24. Patke A, Mecklenbrauker I, Erdjument-Bromage H, Tempst P, Tarakhovsky A. BAFF controls B cell metabolic fitness through a PKC beta- and Akt-dependent mechanism. *J Exp Med* 2006;**203**:2551-62.
25. Woodland RT, Schmidt MR, Thompson CB. BLyS and B cell homeostasis. *Semin Immunol* 2006;**18**:318-26.

26. Meyer KB, Skogberg M, Margenfeld C, Ireland J, Pettersson S. Repression of the immunoglobulin heavy chain 3' enhancer by helix-loop-helix protein Id3 via a functionally important E47/E12 binding site: implications for developmental control of enhancer function. *Eur J Immunol* 1995;**25**:1770-7.
27. Pan L, Sato S, Frederick JP, Sun XH, Zhuang Y. Impaired immune responses and B-cell proliferation in mice lacking the Id3 gene. *Mol Cell Biol* 1999;**19**:5969-80.
28. Nichols KE, Harkin DP, Levitz S, Krainer M, Kolquist KA, Genovese C, Bernard A, Ferguson M, Zuo L, Snyder E, Buckler AJ, Wise C, Ashley J, Lovett M, Valentine MB, Look AT, Gerald W, Housman DE, Haber DA. Inactivating mutations in an SH2 domain-encoding gene in X-linked lymphoproliferative syndrome. *Proc Natl Acad Sci U S A* 1998;**95**:13765-70.
29. Coffey AJ, Brooksbank RA, Brandau O, Oohashi T, Howell GR, Bye JM, Cahn AP, Durham J, Heath P, Wray P, Pavitt R, Wilkinson J, Leversha M, Huckle E, Shaw-Smith CJ, Dunham A, Rhodes S, Schuster V, Porta G, Yin L, Serafini P, Sylla B, Zollo M, Franco B, Bolino A, Seri M, Lanyi A, Davis JR, Webster D, Harris A, Lenoir G, de St Basile G, Jones A, Behloradsky BH, Achatz H, Murken J, Fassler R, Sumegi J, Romeo G, Vaudin M, Ross MT, Meindl A, Bentley DR. Host response to EBV infection in X-linked lymphoproliferative disease results from mutations in an SH2-domain encoding gene. *Nat Genet* 1998;**20**:129-35.
30. Sayos J, Wu C, Morra M, Wang N, Zhang X, Allen D, van Schaik S, Notarangelo L, Geha R, Roncarolo MG, Oettgen H, De Vries JE, Aversa G, Terhorst C. The X-linked lymphoproliferative-disease gene product SAP regulates signals induced through the co-receptor SLAM. *Nature* 1998;**395**:462-9.
31. Carter RH, Fearon DT. CD19: lowering the threshold for antigen receptor stimulation of B lymphocytes. *Science* 1992;**256**:105-7.
32. van Noesel CJ, Lankester AC, van Lier RA. Dual antigen recognition by B cells. *Immunol Today* 1993;**14**: 8-11.
33. Lopes-Carvalho T, Kearney JF. Development and selection of marginal zone B cells. *Immunol Rev* 2004;**197**: 192-205.
34. Spencer J, Perry ME, Dunn-Walters DK. Human marginal-zone B cells. *Immunol Today* 1998;**19**:421-6.
35. Martin F, Kearney JF. Marginal-zone B cells. *Nat Rev Immunol* 2002;**2**:323-35.
36. van den Oord JJ, de Wolf-Peeters C, Desmet VJ. The marginal zone in the human reactive lymph node. *Am J Clin Pathol* 1986;**86**:475-9.
37. Dono M, Zupo S, Augliera A, Burgio VL, Massara R, Melagrana A, Costa M, Grossi CE, Chiorazzi N, Ferrarini M. Subepithelial B cells in the human palatine tonsil. II. Functional characterization. *Eur J Immunol* 1996;**26**:2043-9.
38. Spencer J, Finn T, Pulford KA, Mason DY, Isaacson PG. The human gut contains a novel population of B lymphocytes which resemble marginal zone cells. *Clin Exp Immunol* 1985;**62**:607-12.
39. Weller S, Braun MC, Tan BK, Rosenwald A, Cordier C, Conley ME, Plebani A, Kumararatne DS, Bonnet D, Tournilhac O, Tchernia G, Steiniger B, Staudt LM, Casanova JL, Reynaud CA, Weill JC. Human blood IgM "memory" B cells are circulating splenic marginal zone B cells harboring a prediversified immunoglobulin repertoire. *Blood* 2004;**104**:3647-54.
40. Weller S, Reynaud CA, Weill JC. Splenic marginal zone B cells in humans: where do they mutate their Ig receptor? *Eur J Immunol* 2005;**35**:2789-92.
41. Willenbrock K, Jungnickel B, Hansmann ML, Kuppers R. Human splenic marginal zone B cells lack expression of activation-induced cytidine deaminase. *Eur J Immunol* 2005;**35**:3002-7.
42. Chang B, Casali P. The CDR1 sequences of a major proportion of human germline Ig VH genes are inherently susceptible to amino acid replacement. *Immunol Today* 1994;**15**:367-73.
43. Jacob J, Kassir R, Kelsoe G. In situ studies of the primary immune response to (4-hydroxy-3-nitrophenyl)acetyl. I. The architecture and dynamics of responding cell populations. *J Exp Med* 1991;**173**:1165-75.
44. Wolniak KL, Noelle RJ, Waldschmidt TJ. Characterization of (4-hydroxy-3-nitrophenyl)acetyl (NP)-specific germinal center B cells and antigen-binding B220- cells after primary NP challenge in mice. *J Immunol* 2006;**177**:2072-9.
45. Grimbacher B, Hutloff A, Schlesier M, Glocker E, Warnatz K, Drager R, Eibel H, Fischer B, Schaffer AA, Mages HW, Kroczek RA, Peter HH. Homozygous loss of ICOS is associated with adult-onset common variable immunodeficiency. *Nat Immunol* 2003;**4**:261-8.
46. Castigli E, Wilson SA, Garibyan L, Rachid R, Bonilla F, Schneider L, Geha RS. TACI is mutant in common variable immunodeficiency and IgA deficiency. *Nat Genet* 2005;**37**:829-34.

47. Salzer U, Chapel HM, Webster AD, Pan-Hammarstrom Q, Schmitt-Graeff A, Schlesier M, Peter HH, Rockstroh JK, Schneider P, Schaffer AA, Hammarstrom L, Grimbacher B. Mutations in TNFRSF13B encoding TACI are associated with common variable immunodeficiency in humans. *Nat Genet* 2005;**37**:820-8.
48. Vetrie D, Vorechovsky I, Sideras P, Holland J, Davies A, Flinter F, Hammarstrom L, Kinnon C, Levinsky R, Bobrow M, et al. The gene involved in X-linked agammaglobulinaemia is a member of the src family of protein-tyrosine kinases. *Nature* 1993;**361**:226-33.
49. Tsukada S, Saffran DC, Rawlings DJ, Parolini O, Allen RC, Klisak I, Sparkes RS, Kubagawa H, Mohandas T, Quan S, et al. Deficient expression of a B cell cytoplasmic tyrosine kinase in human X-linked agammaglobulinemia. *Cell* 1993;**72**:279-90.
50. Yel L, Minegishi Y, Coustan-Smith E, Buckley RH, Trubel H, Pachman LM, Kitchingman GR, Campana D, Rohrer J, Conley ME. Mutations in the mu heavy-chain gene in patients with agammaglobulinemia. *N Engl J Med* 1996;**335**:1486-93.
51. Minegishi Y, Coustan-Smith E, Wang YH, Cooper MD, Campana D, Conley ME. Mutations in the human lambda5/14.1 gene result in B cell deficiency and agammaglobulinemia. *J Exp Med* 1998;**187**:71-7.
52. Minegishi Y, Coustan-Smith E, Rapalus L, Ersoy F, Campana D, Conley ME. Mutations in Igalpha (CD79a) result in a complete block in B-cell development. *J Clin Invest* 1999;**104**:1115-21.
53. Minegishi Y, Rohrer J, Coustan-Smith E, Lederman HM, Pappu R, Campana D, Chan AC, Conley ME. An essential role for BLNK in human B cell development. *Science* 1999;**286**:1954-7.
54. Allen RC, Armitage RJ, Conley ME, Rosenblatt H, Jenkins NA, Copeland NG, Bedell MA, Edelhoff S, Disteche CM, Simoneaux DK, et al. CD40 ligand gene defects responsible for X-linked hyper-IgM syndrome. *Science* 1993;**259**:990-3.
55. Ferrari S, Giliani S, Insalaco A, Al-Ghonaium A, Soresina AR, Loubser M, Avanzini MA, Marconi M, Badolato R, Ugazio AG, Levy Y, Catalan N, Durandy A, Tbakhi A, Notarangelo LD, Plebani A. Mutations of CD40 gene cause an autosomal recessive form of immunodeficiency with hyper IgM. *Proc Natl Acad Sci U S A* 2001;**98**:12614-9.
56. Revy P, Muto T, Levy Y, Geissmann F, Plebani A, Sanal O, Catalan N, Forveille M, Dufourcq-Labelouse R, Gennery A, Tezcan I, Ersoy F, Kayserili H, Ugazio AG, Brousse N, Muramatsu M, Notarangelo LD, Kinoshita K, Honjo T, Fischer A, Durandy A. Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2). *Cell* 2000;**102**:565-75.
57. Imai K, Slupphaug G, Lee WI, Revy P, Nonoyama S, Catalan N, Yel L, Forveille M, Kavli B, Krokan HE, Ochs HD, Fischer A, Durandy A. Human uracil-DNA glycosylase deficiency associated with profoundly impaired immunoglobulin class-switch recombination. *Nat Immunol* 2003;**4**:1023-8.
58. Bain G, Maandag EC, Izon DJ, Amsen D, Kruisbeek AM, Weintraub BC, Krop I, Schlissel MS, Feeney AJ, van Roon M. E2A proteins are required for proper B cell development and initiation of immunoglobulin gene rearrangements. *Cell* 1994;**79**:885-92.
59. Lin H, Grosschedl R. Failure of B-cell differentiation in mice lacking the transcription factor EBF. *Nature* 1995;**376**:263-7.
60. Hendriks RW, Middendorp S. The pre-BCR checkpoint as a cell-autonomous proliferation switch. *Trends Immunol* 2004;**25**:249-56.
61. Sandel PC, Monroe JG. Negative selection of immature B cells by receptor editing or deletion is determined by site of antigen encounter. *Immunity* 1999;**10**:289-99.
62. Shapiro-Shelef M, Calame K. Regulation of plasma-cell development. *Nat Rev Immunol* 2005;**5**:230-42.
63. Lavie L, Kitova M, Maldener E, Meese E, Mayer J. CpG methylation directly regulates transcriptional activity of the human endogenous retrovirus family HERV-K(HML-2). *J Virol* 2005;**79**:876-83.
64. Yu F, Zingler N, Schumann G, Stratling WH. Methyl-CpG-binding protein 2 represses LINE-1 expression and retrotransposition but not Alu transcription. *Nucleic Acids Res* 2001;**29**:4493-501.
65. Korf B. Molecular diagnosis (1). *N Engl J Med* 1995;**332**:1218-20.
66. Cavazzana-Calvo M, Hacein-Bey S, de Saint Basile G, Gross F, Yvon E, Nusbaum P, Selz F, Hue C, Certain S, Casanova JL, Bousso P, Deist FL, Fischer A. Gene therapy of human severe combined immunodeficiency (SCID)-X1 disease. *Science* 2000;**288**:669-72.
67. Hacein-Bey-Abina S, Le Deist F, Carlier F, Bouneaud C, Hue C, De Villartay JP, Thrasher AJ, Wulffraat N, Sorensen R, Dupuis-Girod S, Fischer A, Davies EG, Kuis W, Leiva L, Cavazzana-Calvo M. Sustained correction of X-linked severe combined immunodeficiency by ex vivo gene therapy. *N Engl J Med* 2002;**346**:1185-93.

68. Bordignon C, Notarangelo LD, Nobili N, Ferrari G, Casorati G, Panina P, Mazzolari E, Maggioni D, Rossi C, Servida P, Ugazio AG, Mavilio F. Gene therapy in peripheral blood lymphocytes and bone marrow for ADA-immunodeficient patients. *Science* 1995;**270**:470-5.
69. Aiuti A, Slavin S, Aker M, Ficara F, Deola S, Mortellaro A, Morecki S, Andolfi G, Tabucchi A, Carlucci F, Marinello E, Cattaneo F, Vai S, Servida P, Miniero R, Roncarolo MG, Bordignon C. Correction of ADA-SCID by stem cell gene therapy combined with nonmyeloablative conditioning. *Science* 2002;**296**:2410-3.
70. Hacein-Bey-Abina S, Von Kalle C, Schmidt M, McCormack MP, Wulffraat N, Leboulch P, Lim A, Osborne CS, Pawliuk R, Morillon E, Sorensen R, Forster A, Fraser P, Cohen JI, de Saint Basile G, Alexander I, Wintergerst U, Frebourg T, Auricchio A, Stoppa-Lyonnet D, Romana S, Radford-Weiss I, Gross F, Valensi F, Delabesse E, Macintyre E, Sigaux F, Soulier J, Leiva LE, Wissler M, Prinz C, Rabbitts TH, Le Deist F, Fischer A, Cavazzana-Calvo M. LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science* 2003;**302**:415-9.
71. Weissenbach J, Gyapay G, Dib C, Vignal A, Morissette J, Millasseau P, Vaysseix G, Lathrop M. A second-generation linkage map of the human genome. *Nature* 1992;**359**:794-801.
72. Sheffield VC, Weber JL, Buetow KH, Murray JC, Even DA, Wiles K, Gastier JM, Pulido JC, Yandava C, Sunden SL, et al. A collection of tri- and tetranucleotide repeat markers used to generate high quality, high resolution human genome-wide linkage maps. *Hum Mol Genet* 1995;**4**:1837-44.
73. Botstein D, Risch N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nat Genet* 2003;**33 Suppl**:228-37.
74. Papassotiropoulos A, Stephan DA, Huentelman MJ, Hoerndli FJ, Craig DW, Pearson JV, Huynh KD, Brunner F, Corneveaux J, Osborne D, Wollmer MA, Aerni A, Coluccia D, Hanggi J, Mondadori CR, Buchmann A, Reiman EM, Caselli RJ, Henke K, de Quervain DJ. Common Kibra alleles are associated with human memory performance. *Science* 2006;**314**:475-8.

VIII

Appendices

Appendix 1. PCR primers and TaqMan probes for quantification of Ig gene rearrangements.

Locus	Target	Forward primer(s)	Position	Primer sequence (5'-3')	Reverse primer(s)	Position	Primer sequence (5'-3')
IGH	DH1-3 - JH1-6						
		DH1 [a]	-211	GGCGGAATGTGTGCAGGC	JH1 [b]	+77	CGCTATCCCCAGACAGCAGA
		DH2 [a]	-190	GCACTGGGCTCAGAGTCCTCT	JH2 [b]	+99	GGTGCCTGGACAGAGAAGACT
		DH3	-75	CCTCAGGTCAGCCCTGGACA	JH3 [b]	+73	AGGCAGAAGGAAAGCCATCTTAC
					JH4 [b]	+84	CAGAGTTAAAGCAGGAGAGAGGTTGT
					JH5 [b]	+75	AGAGAGGGGGTGGTGAGGACT
					JH6 [b]	+98	GCAGAAAACAAAGGCCCTAGAGT
	DH4-6 - JH1-6						
		DH4 [a]	-134	AGATCCCCAGGACGCAGCA	JH1 [b]	+77	CGCTATCCCCAGACAGCAGA
		DH5 [a]	-185	CAGGGGGACACTGTGCATGT	JH2 [b]	+99	GGTGCCTGGACAGAGAAGACT
		DH6 [a]	-83	TGACCCCAGCAAGGGAAGG	JH3 [b]	+73	AGGCAGAAGGAAAGCCATCTTAC
					JH4 [b]	+84	CAGAGTTAAAGCAGGAGAGAGGTTGT
					JH5 [b]	+75	AGAGAGGGGGTGGTGAGGACT
					JH6 [b]	+98	GCAGAAAACAAAGGCCCTAGAGT
	VH1-3 - JH1-6						
		VH1-FR3	-51	GCTGAGCAGCCTGAGATCTGA	JH1 [b]	+77	CGCTATCCCCAGACAGCAGA
		VH2-FR3	-49	ACCAACATGGACCCTGTGGA	JH2 [b]	+99	GGTGCCTGGACAGAGAAGACT
		VH3-FR3	-53	CAAATGAACAGCCTGAGAGCC	JH3 [b]	+73	AGGCAGAAGGAAAGCCATCTTAC
					JH4 [b]	+84	CAGAGTTAAAGCAGGAGAGAGGTTGT
					JH5 [b]	+75	AGAGAGGGGGTGGTGAGGACT
					JH6 [b]	+98	GCAGAAAACAAAGGCCCTAGAGT
	VH4-7 - JH1-6						
		VH4-FR3	-47	AGCTCTGTGACCGCCGC	JH1 [b]	+77	CGCTATCCCCAGACAGCAGA
		VH5-FR3	-63	CGCCTACCTGCAGTGGAGC	JH2 [b]	+99	GGTGCCTGGACAGAGAAGACT
		VH6-FR3	-59	TCCCTGCAGCTGAACTCTGTG	JH3 [b]	+73	AGGCAGAAGGAAAGCCATCTTAC
		VH7-FR3	-66	CACGGCATATCTGCAGATCAG	JH4 [b]	+84	CAGAGTTAAAGCAGGAGAGAGGTTGT
					JH5 [b]	+75	AGAGAGGGGGTGGTGAGGACT
					JH6 [b]	+98	GCAGAAAACAAAGGCCCTAGAGT
IGK	Vκ1-7 - Jκ1-5						
		Vκ1/6 [c]	-108	GTTCAGCGGCAGTGGATCTG	Jκ1	+82	GATCACTTCATAGACACAGGGAACAG
		Vκ2 [c]	-83	ACAGATTTCACACTGAAAATCAG	Jκ2	+84	AATCCACACAAAGGAACACAAAGTA
		Vκ3 [c]	-86	GGGACAGACTTCACTCTCAC	Jκ3	+82	TCACAAAACTTGCACAAATGATACA
		Vκ4 [c]	-86	GGGACAGATTTCACTCTCAC	Jκ4	+95	AATTGACCTGAGCTAATATCTCAAACAC
		Vκ5 [e]	-72	CAGCGGGTATGGAACAGATTTTAC	Jκ5	+58	CCATCAGACCCAAATTTCAGAAGAA
		Vκ7	-82	CCGATTTCACCCTCACAATTAATC			
	Vκ1-7 - Kde						
		Vκ1/6 [c]	-108	GTTCAGCGGCAGTGGATCTG	Kde [c]	+91	TTCCTAGGGAGGTCAGACTC
		Vκ2 [c]	-83	ACAGATTTCACACTGAAAATCAG			
		Vκ3 [c]	-86	GGGACAGACTTCACTCTCAC			
		Vκ4 [c]	-86	GGGACAGATTTCACTCTCAC			
		Vκ5 [e]	-72	CAGCGGGTATGGAACAGATTTTAC			
		Vκ7	-82	CCGATTTCACCCTCACAATTAATC			
	IntronRSS - Kde						
		IntronRSS	-152	GGCACCGCGAGCTGTAGAC	Kde [c]	+91	TTCCTAGGGAGGTCAGACTC
IGL	Vλ1-3 - Jλ1-3						
		Vλ1/2	-108	CTCTGGCTCCAAGTCTGGC	Jλ1	+70	ACAGGCTGGGAAAGGTTGAG
		Vλ3	-118	GATCCCTGAGCGATTCTCTGG	Jλ2/3	+65	GGGAAGGAGAGGGGAGAAG

a. Van Dongen *et al.* (2003) *Leukemia* **17**:2257-2317
b. Verhagen *et al.* (2000) *Leukemia* **14**:1426-1435
c. Pongers-Willemse *et al.* (1999) *Leukemia* **13**:110-118

Appendix 1. PCR primers and TaqMan probes for quantification of Ig gene rearrangements (continued).

TaqMan probes	Position	Probe sequence (5'-3')	Control cell line(s)
JH1/2/4/5 [b]	+31/32/27/30	CCCTGGTCACCGTCTCCTCAGGTG	KCA + Ros-15
JH3 [b]	+20	CAAGGGACAATGGTCACCGTCTCTTCA	
JH6 [b]	+43	CACGGTCACCGTCTCCTCAGGTAAGAA	
JH1/2/4/5 [b]	+31/32/27/30	CCCTGGTCACCGTCTCCTCAGGTG	Cll-1
JH3 [b]	+20	CAAGGGACAATGGTCACCGTCTCTTCA	
JH6 [b]	+43	CACGGTCACCGTCTCCTCAGGTAAGAA	
JH1/2/4/5 [b]	+31/32/27/30	CCCTGGTCACCGTCTCCTCAGGTG	KCA + Nalm-6
JH3 [b]	+20	CAAGGGACAATGGTCACCGTCTCTTCA	
JH6 [b]	+43	CACGGTCACCGTCTCCTCAGGTAAGAA	
JH1/2/4/5 [b]	+31/32/27/30	CCCTGGTCACCGTCTCCTCAGGTG	U698
JH3 [b]	+20	CAAGGGACAATGGTCACCGTCTCTTCA	
JH6 [b]	+43	CACGGTCACCGTCTCCTCAGGTAAGAA	
Jκ1 MGB [d]	+26	CTCACGTTTGATTTCCA	U698 + Namalwa +
Jκ2/4 cons MGB	+26/27	CTTACGTTTGATCTCCA	KCA + Cll-1
Jκ3 MGB	+25	CTTACGTTTGATATCCAC	
Jκ5 MGB	+26	CTTACGTTTAATCTCCA	
Kde [f]	+43	AGCTGCATTTTTGCCATATCCACTATTTGGAGT	Ros-15 + 380 + Nalm-1 + Ros-5
Kde [f]	+43	AGCTGCATTTTTGCCATATCCACTATTTGGAGT	Nalm-1
Jλ1 MGB	+44	CACCGTCCTAGGTAAGT	KCA
Jλ2/3 cons MGB	+44	ACCGTCCTAGGTGAGT	

d. MGB = Minor Groove Binding
e. Beishuizen *et al.* (1997) *Leukemia* **11**:2200-2207
f. Van der Velden *et al.* (2002) *Leukemia* **16**:928-936

Appendix 2. PCR primers for GeneScan analysis of complete *IGH*, *IGK* and *IGL* rearrangements.

Locus	Target	Position	Type	Sequence (5' - 3')	Position
IGH	VH - DJH				
		VH1-FR2 [a]	Forward	CTGGGTGCGACAGGCCCCTGGACAA	-192
		VH2-FR2	Forward	GATCCGTCAGCCCCCAGGGAAGG	-188
		VH3-FR2 [a]	Forward	GGTCCGCCAGGCTCCAGGGAA	-189
		VH4-FR2	Forward	GATCCGCCAGCCCCCAGGGAAGG	-186
		VH5-FR2	Forward	GGTGCGCCAGATGCCCGGGAAAGG	-189
		VH6-FR2	Forward	GATCAGGCAGTCCCCATCGAGAG	-192
		VH7-FR2 [a]	Forward	TTGGGTGCGACAGGCCCCTGGACAA	-192
		JH FAM1	Reverse	CTTACCTGAGGAGACGGTGACC	+57
IGK	Vκ - Jκ				
		Vκ1	Forward	AGGAGACAGAGTCACCATCACT	-246
		Vκ2	Forward	CTCCATCTCCTGCAGGTCTAGTC	-246
		Vκ3	Forward	GGAAAGAGCCACCCTCTCCTG	-243
		Vκ4 [a]	Forward	CAACTGCAAGTCCAGCCAGAGTGTTTT	-243
		Jκ1-4 FAM [a]	Reverse	CTTACGTTTGATCTCCACCTTGGTCCC	+43
		Jκ5 FAM [a]	Reverse	CTTACGTTTAATCTCCAGTCGTGTCCC	+43
IGL	Vλ - Jλ				
		Vλ1/2	Forward	CTGACCGATTCTCTGGCTCCAAGTCT	-118
		Vλ3 [a]	Forward	GGCTCCCTGAGCGATTCTCTGG	-124
		Jλ1-3 FAM [a]	Reverse	CTAGGACGGTGAGCTTGGTCCC	+48

a. Van Dongen *et al.* (2003) *Leukemia* **17**:2257-2317

Appendix 3. PCR primers and TaqMan probes for quantification of expression levels of 16 candidate genes for induction of Ig gene rearrangements.

Target	Forward primer	Primer sequence (5'-3')	Reverse primer	Primer sequence (5'-3')	TaqMan probe	Probe sequence (5'-3')
RAG1 [a]	RAG1_F	TGAGTAATATCAACCAAATTGCAGACA	RAG1_R	GGATCTCACCCGGAACAGC	T-RAG1	CCCCAGATGAAATTCAGCACCCACATATTAA
RAG2 [a]	RAG2_F	GTTTAGCGGCAAAGATTCAGAGA	RAG2_R	GTCCATCAAAATTCATCAGTGAGAA	T-RAG2	CAAAAATCTACGTACCATCAGAAACTATGTCTCTGCA
COPEB (KLF6)	KLF6_F2	AGGCACTTCCGAAAGCACAC	KLF6_R_short	AGCCTGGAAGCCTTTTAGCCTA	T-KLF6	TCTGACCACCTGGCCCTGCACAT
ELF1	ELF1_F	GGAGCCTCATCACCAGAACAG	ELF1_R	GTAGCCTTGTCCTGGAGCAGTG	T-ELF1	CACCACGACCAGATTCCCCAGCC
ELK3	ELK3_F	CCCTCACCCCAGCCTTCTT	ELK3_R	GGAACTGGAACAGCGTGCTT	T-ELK3	CAGTTGCTCCGCTGAGTCCTGCCAG
ERG	ERG_F	GCCGACATCCTTCTCTCACATC	ERG_R	ATGGTTGAGCAGCTTTCGACTG	T-ERG	ACTCTCCACGGTTAATGCATGCTAGAAACACAG
ETS1	ETS1_F	GGGACCGTGCTGACCTCAA	ETS1_R	CCCAGCCATCTCCTGTCCA	Tr-ETS1	TCCACTGCCTGTGTAGCCAGCTAGGG
ETS2	ETS2_F	GGACTTCCCAGGCAGCAAC	ETS2_R	AGCAAGGACGACTGGCTGTT	T-ETS2	AAGACCACGACTCCCCTGAGAACGGT
HIVEP3	HIVEP3_F	GATGCTCCTGCCTCCCAGA	HIVEP3-R	CGAATTCCACACTCCTCACAAA	T-HIVEP3	TCGAAGGAGGGTACAAATCAAACGAAGAGTATGTA
KLF12	KLF12_F	GAAGGCTCACCGGAGGACA	KLF12_R	GCAGGGCCAAATGATCTGAC	T-KLF12	ACTGACGAGGCATTACCGCAAACATACG
KLF2	KLF2_F	CACCAAGAGTTCGCATCTGAAG	KLF2_R	ACAGATGGCACTGGAATGGC	T-KLF2	GACGGCTGCGGCTGGAAGTTTG
KLF4	KLF4_F	GAGTTCCCATCTCAAGGCACAC	KLF4_R	TGTTTACGGTAGTGCCTGGTCAG	T-KLF4	ACTGTGACTGGGACGGCTGTGGATG
OCAB	OCAB_F	TGCCTCTCCGCCACTCAT	OCAB_R	GCAGGGTGGAGGTGGGTAG	T-OCAB	CTCACCTATTTCCCGTGGCCTCAGC
OCT1	OCT1_F	GGGACCAGCAGCTCACCTATT	OCT1_R	GAAAGATTCGTCACAGCAGCAC	T-OCT1	CTCACAGTCAGCCCTGTCCTCCCTCTG
OCT2	OCT2_F	CCCTCCTGGAGAAGTGGCTC	OCT2_R	TTTGTCTCGATGCTGGTCCTC	T-OCT2	ACTCAAGCCTGCCCAGCCCCAAC
POU4F1	POU4F1_F	AAGTACCCGTCGCTGCACTC	POU4F1_R	GGCTTGAAAGGATGGCTCTTG	T-POU4F1	AGCAACCTCTTCGCCAGCCTGGAC
SMARCA5	SMARCA5_F	GCAACGAGCTCCAGGACATAG	SMARCA5_R	CAGCTGATGAAAAGGTGCTTTGTA	T-SMARCA5	CTCAGATTGAAAGGGGAGAGGCGAGAA
SPIB	SpiB_F	TGCAGGGAACCTCGAACTG	SpiB_R	CGCTGTCCGAGACCTCCAG	T-SpiB	CCTGTGCTATCAGAGGAGGAAGACTTACCGTT

a. Boeckx *et al.* (2002) *Leukemia* **16**:368-375

Appendix 4. Full width at half the maximum (FWHM) values of the point spread function for the indicated colors.

Color	dimension		
	X (nm)	Y (nm)	Z (nm)
Green (Alexa-488)	543.40	489.06	1800
Red (Alexa-594)	380.38	380.38	1400
Dark Red (Alexa-647)	434.72	434.72	1700

Appendix 5. Chromatic shifts.

	red vs. green (in nm)			red vs. dark red (in nm)			dark red vs. green (in nm)		
	X	Y	Z	X	Y	Z	X	Y	Z
average	31.7	-14.4	-133.0	-21.8	-11.6	26.1	53.5	-2.8	-159.1
standard deviation	16.5	17.2	35.3	22.9	25.0	36.6	30.2	31.5	24.9

Appendix 6. Resolution equivalents before and after application of chromatic corrections.

Color combination	average ± SD ± SE (in nm) before correction	average ± SD ± SE (in nm) after correction
Red (Alexa-594) - Green (Alexa-488)	139.7 ± 34.6 ± 2.3	34.9 ± 24.3 ± 1.6
Red (Alexa-594) - Dark Red (Alexa-647)	56.3 ± 24.6 ± 1.6	44.6 ± 22.1 ± 1.5
Dark Red (Alexa-647) - Green (Alexa-488)	173.2 ± 26.2 ± 1.7	46.8 ± 17.8 ± 1.2

SD: standard deviation
SE: standard error

Appendix 7. *IGHM* gross deletion primers for deletion-scanning and breakpoint PCR.

Target	Forward primer	Primer sequence (5'-3')	Reverse primer	Primer sequence (5'-3')	Size (bp)
deletion-scanning PCR					
VHIV-44.1	VHIV-44.1	CGCCTATGCACATTCAATTTC	VH4-44.1 R	GCAGGGCAGGTTTGTGTCTG	268
VHII-43.1	VHII-43.1_F	CTGGATTCCCCAACAGAAC	VHII-43.1_R	ATGTCACTCAGCCATACGAAGAG	185
VH3-42	VH3-42_F	TCTGGGAAAGGGCTGGAG	VH3-42_R	GTGTATGGGCCTACGTTA	215
VH3-42+3500	VH3-42+3500_F	AAGTGCGTTCCCTGAAGAGAG	VH3-42+3500_R	AGACGGCACAGTTGGTTTG	292
VH3-42+6700	VH3-42+6700_F	AGATTTGGGGTAGGACACAGA	VH3-42+6700_R	TATCGGTCAAAGGGCACAG	284
VH3-42+10000	VH3-42+10000_F	AGAGCATGTTGGCAGTTCCTT	VH3-42+10000_R	CGGGAGGCTGAGGCAGTGA	222
VH3-42+13200	VH3-42+13200_F	GATAACAAAACCTGGCAGAGACT	VH3-42+13200_R	ATGATTTTGGCTGTGGGTTCA	351
VH3-42+16500	VH3-42+16500_F	CTCATCCCACCTGACGAGAAA	VH3-42+16500_R	AAAGTGGCTTGATTGCAGCTTA	275
VH3-41	VH3-41_F	ACAGTGGATACGTGTGGCAGTT	VH3-41_R	TGGCCGTGTCCTTAGCTT	337
VH3-38	VH3-38_F	TGTGGCAGTTTCTGACCTTGG	VH3-38_R	CAATGACTTCCCCTCTGTGTA	360
VH3-37	VH3-37_F	GCCTGCGGATTCATCTTG	VH3-37_R	TACTGATTTCCCCTCACCA	236
VH7-27	VH7-27_F	CAGTCACCGGGCTGAGAG	VH7-27_R	GAAGATATGCCATGCTGACG	326
VHIV-26.2	VH4-26.2	TGGCTGGGGCACAGTTACTA	VH4-26.2 R	TGAGTCCCCTTGGTGGTCTT	253
VHII-26.2+3100	VHII-26.2+3100_F	CCTCTCAACCAACTGGCCTAGT	VHII-26.2+3100_R	GATGGGATGGGGTTCCTCAC	220
VHII-26.2+6200	VHII-26.2+6200_F	CTGAGATTGGTGCAACCAGCTAC	VHII-26.2+6200_R	GGAAGCCAAACAAAGTGGAGAA	338
VHII-26.2+9400	VHII-26.2+9400_F	GAGGATGGAGCTACTTGGTTGC	VHII-26.2+9400_R	CTTCTTTCCCAGCTCATTGGAG	303
VH2-26	VH2-26_F	CTTGGCTGAGAGGCATGGTC	VH2-26_R	GCAGCAGGAGTGTGTAGCAAAGT	273
VH3-25	VH3-25_F	CTTGGCAAAGCCTGCGTG	VH3-25_R	CACACTGACCTCCCCTCACT	290
VH1-24	VH1-24_F	GAAGGTCTCCTGCAAGGTTTC	VH1-24_R	GGTTTCTGACGCTCTCAGGAT	282
VHII-22.1	VHII-22.1_F	ATTTGCTCTCGGGGACCTA	VHII-22.1_R	TTGAGGAGTGGGTTGGATTGTGC	300
VHII-20.1 upstream	VHII-20.1_upstream_F	ATCCCAGATCACAGGAGGTCAGG	VHII-20.1_upstream_R	GGTTGTGATGAGGAATCCAGCAT	283
VHII-20.1+3800	VHII-20.1+3800 F	TTTTCCCCATACTGAGTGTCTG	VHII-20.1+3800 R	GGAGTGAGGTAAATCCTGCTTAG	145
VHII-20.1+7500	VHII-20.1+7500 F	GTGGAGAACATTCACCTGATTGGA	VHII-20.1+7500 R	GAACCCACCACCAGTTCAGTTTAG	117
VHII-20.1+11000	VHII-20.1+11000 F	TTGTCTTCCCAGAGTGTCCTACCT	VHII-20.1+11000 R	CCCGATATCCAAAGACGACAAT	165
VH3-19	VH3-19 F	CTTAACTCACCTGGGTACAACAGT	VH3-19 R	TAAGTCCTCCTATCATATTACAATG	584
VHII-15.1	VHII-15.1 NEW	TCACGAGGAAACACACAGTGC	VHII-15.1 R	GATGTGTCCTGAGAGCCCTT	238
VH6-1	VH6-1 FR2 (Biomed 2)	TGGATCAGGCAGTCCCCATCGAGAG	VH6-1 ds_R	GGTAGTGCTGGGTGCCTTCTG	278
VH6-1+2700	VH6-1+2700 F	TAGATGTTTGCAGTGGCAGTGG	VH6-1+2700 R	TTGAAGGGAGACTGACGGAGTG	198

Appendix 7. *IGHM* gross deletion primers for deletion-scanning and breakpoint PCR (continued).

Target	Forward primer	Primer sequence (5'-3')	Reverse primer	Primer sequence (5'-3')	Size (bp)
deletion-scanning PCR					
VH6-1+4900	VH6-1+4900 F	ACCCAATGAAACTATCCTTCAG	VH6-1+4900 R	TGAAGAACTTCCTTTAGCCAGT	120
VH6-1+7100	VH6-1+7100 F	GCACTGGTCCCAGAAGTCTGAT	VH6-1+7100 R	GGACTAGCCAGTGGCAGAGTGT	216
KIAA0125	KIAA0125_F	GACTCCTCCCATCAGCAGAAACTT	KIAA0125_R	CCCTCACCTCCATCCCATCAAA	149
IGH_895300	IGH_895300_F	CTGTGGCGTCACTCTGGTA	IGH_895300_R	GGTTATCCCCTTGGACTG	302
IGH_897200	IGH_897200_F	GCACCTGGGCAAAGCCTG	IGH_897200_R	TGCTGTGGCCGGGGACCT	326
IGH_899400	IGH_899400_F	CCCAAAAGGTTCCATAGCAG	IGH_899400_R	TGCCAAAATGCAAGTTAGCC	539
IGH_901600	IGH_901600_F	ACAGTCACAGGACGCATCATTT	IGH_901600_R	GTCCCCAGGCTCTGCACC	143
DH1-1	DH1-1F	GGCGGAATGTGTACAGGC	DH1-1R	CCAGTTGTACCCACCGTGACTC	205
DH2-2 [a]	DH2-2 sense	GATATTGTAGTAGTACCAGCTGC	DH2-2 antisense	GTAGGATAGCGGCTGCTTCTC	421
DH2-8	DH2-8_F	AGATGGGGCTCAGTGACACC	DH2-8_R	TTTGGGCATTGATGGACACC	208
DH3-10 [a]	DH3-10 sense	GTATTACTATGGTTCGGGGAG	DH3-10 antisense	GAAATGACCCAGGAGAGGCTC	566
DH3-10+2000	DH3-10+2000_F	CCAGGGCAGGGATGATTGAG	DH3-10+2000_R	CTGTGGGATGGGACAGGCTTC	188
DH3-10+3500	DH3-10+3500_F	CTGACACCCAGGCCAGCAGAC	DH3-10+3500_R	ACTCTCTGGGCAGGTCAGC	180
DH3-10+5000	DH3-10+5000_F	GCCTCAAGCCCAGGTTCC	DH3-10+5000_R	CTGTGTGTGGGTGGCTCCTGC	178
DH3-10+10000	DH3-10+10000_F	ACCCTCCAGCCAGACATCACCT	DH3-10+10000_R	TGCATGGTCCCTGGGTGTCC	318
DH3-10+20000	DH3-10+20000_F	GCCTGACTCCTGCACACCCTC	DH3-10+20000_R	AGTTACCACCGTAGTCACACAGC	285
DH3-10+29000	DH3-10+29000_F	CCTCAAAAGGCCAAATCTAA	DH3-10+29000_R	TCTGCTGCCAGACATATTGGAG	315
IGHM	Cμ5U	CCACCCCAGCCTGACCCTATT	Cμ5L	CTTCTGTTGGGATCATTTCACCTAC	354
IGHM	Cμ6U	CGCCTCCACCTTCATCGTC	Cμ6L	ATTTCAGGGCAACAAGCGTAT	485
IGHM-IGHD intron	Cμ+6000_F	GGGGCTGCTGGGTACAAGT	Cμ+6300_nested_R	GGCAGTGTGGGGCTCCTCA	288
IGHM-IGHD intron	Cμ+7000_F	AGTCCACAATCTCAGCGAAAT	Cμ+7000_R	GCTGGGTTTGGGCTTGCTT	248
IGHM-IGHD intron	Cμ+9700_F	ACCCAGGTCTCCAGGAAAGTG	Cμ+9700_R	GCGTCTGGCCCTCTGAGTG	242
IGHM-IGHD intron	Cμ+12500_F	AACAACATGGGGTGCAGAC	Cμ+12500_R	GCATCTGGCCTGTGGGTCA	236
IgD	IgD c5'	ACCCGTCAAGCTTTCTCTGAACC	IgD c3'	CAGCTGAATTCTAGGCTCCGGC	315
Cγ3	Cγ3U	ATGCCCACGGTGCCCAGGTA	Cγ3L	GGATCTGGGATGGAGGTGGA	321
Cγ1	Cγ1U	GCCCTGAGAGTGACCGCTGTA	Cγ1L	GGCCGTCGCACTCATTTA	374
Cα1	Cα1 F	CCTCAACTCCACCTACCCCAT	Cα CONS R	GGCAGGACACTGGACACGCT	235

Appendix 7. *IGHM* gross deletion primers for deletion-scanning and breakpoint PCR (continued).

Target	Forward primer	Primer sequence (5'-3')	Reverse primer	Primer sequence (5'-3')	Size (bp)			
deletion-scanning PCR								
IGHC_1158400	IGHC_1158400_F	GGCTCTGCTCTCCATGTCCAG	IGHC_1158400_R	GTCTGCTTCAGGTCCACCACC	267			
IGHC_1162500	IGHC_1162500_F	GTGATGTGGTGTTGGGCTAC	IGHC_1162500_R	GGATGTCACACCTTGTCCCATTA	177			
IGHC_1165000	IGHC_1165000_F	TACCTGCGTAGGCTGGTGTG	IGHC_1165000_R	GATGCTGATGCCATGCTTCC	376			
Cγ2	γ2H-γ2CH2 sense	GCGCAAATGTTGTGTCGAGT	γ2H-γ2CH2 antisense	CGGTCCTGCCACAGGTGG	173			
Cγ cons (Cγ4) [a]	γCH2-γCH3 sense	GGAGGTGCATAATGCCAAGACAAAG	γCH2-γCH3 antisense NEW	CTGGTTCTTGGTCATCTCCTCCT	338			
IGHE	IGHE F	GCTGACCCCACGTCTGGC	IGHE R	GGGCTCTGAGCAGGCACA	485			
LM-PCR					DraI	HincII	PvuII	StuI
VH3-42+19600	VH3-42+19600_F	GACCAGTGGTGAGAAGCACAG	AP1	GTAATACGACTCACTATAGGGC				
	VH3-42+19600_nested_F	ATGTGAGTGCTGGGGACATTGTA	AP2	ACTATAGGGCACGCGTGGT	~200	~1,700	~500	/
VHII-20.1	VHII-20.1_NEW_F	GCGGGTTCACAGGACAGCAT	AP1	GTAATACGACTCACTATAGGGC				
	VHII-20.1_nested_F	ACAGAGGTGGGCACTCAGGAAC	AP2	ACTATAGGGCACGCGTGGT	~1,440	~850	~1,000	/
VHII-20.1+11000	VHII-20.1+11000 F	TTGTCTTCCCAGAGTGTCCTACCT	AP1	GTAATACGACTCACTATAGGGC				
	VHII-20.1+11000 nested F	GGAAAGAAGCCATCCTGCGAAT	AP2	ACTATAGGGCACGCGTGGT	580	2,528	765	/
VHII-20.1+11800	VHII-20.1+11800 F	CAGATGTGTTGTGGGTTCAGGC	AP1	GTAATACGACTCACTATAGGGC				
	VHII-20.1+11800 nested F	TTTACCGGCAAGCACCTCA	AP2	ACTATAGGGCACGCGTGGT	~2,400	~1,700	~2,900	/
VH6-1	VH6-1 FR2 (Biomed 2)	TGGATCAGGCAGTCCCCATCGAGAG	AP1	GTAATACGACTCACTATAGGGC				
	VH6-1 FR3 (Biomed 2)	GTTCTCCCTGCAGCTGAACTCTGTG	AP2	ACTATAGGGCACGCGTGGT	3,481	1,238	141	23
VH6-1+2700	VH6-1+2700 F	TAGATGTTTGCAGTGGCAGTGG	AP1	GTAATACGACTCACTATAGGGC				
	VH6-1+2700 nested F	GGAAGTCCCTCCTGTGTCCAGACC	AP2	ACTATAGGGCACGCGTGGT	1,033	1,293	3,221	1,132
VH6-1+4100	VH6-1+4100 F	TGCTTTGAACTTGGTGGTTCCTT	AP1	GTAATACGACTCACTATAGGGC				
	VH6-1+4100 nested F	CTCCATGCTGTTAGTGAAGAGGGAC	AP2	ACTATAGGGCACGCGTGGT	~2,000	/	~1,800	~2,300
DH3-10	DH3-10_LM_F	TGTGGGCATTTGAGAGTGCTGT	AP1	GTAATACGACTCACTATAGGGC				
	DH3-10_LM_nested_F	GAAGCCTGTCCCATCCCACAGC	AP2	ACTATAGGGCACGCGTGGT	/	~2,000	~1,100	~1,600
DH3-10+2700	DH3-10+2700_LM_F	CTAAGACAGGGTGGGCAGAC	AP1	GTAATACGACTCACTATAGGGC				
	DH3-10+2700_LM_nested_F	GACGGCTGTGGACGGGACAT	AP2	ACTATAGGGCACGCGTGGT	/	~1,550	~600	~1,100
Cμ+6300	AP1	GTAATACGACTCACTATAGGGC	Cμ+6300_R	CCCAGTGCAGACTCCCCAGC				
	AP2	ACTATAGGGCACGCGTGGT	Cμ+6300_nested_R	GGCAGTGTGGGGCTCCTCA	/	~330	/	~1,800

Appendix 7. *IGHM* gross deletion primers for deletion-scanning and breakpoint PCR (continued).

Target	Forward primer	Primer sequence (5'-3')	Reverse primer	Primer sequence (5'-3')	Size (bp)			
deletion-scanning PCR								
Cμ+6500	AP1	GTAATACGACTCACTATAGGGC	Cμ+6500_R	CCAGCAGCAGGGAGAGAGC				
	AP2	ACTATAGGGCACGCGTGGT	Cμ+6500_nested_R	CCTGGAACAGAGTCTCCCG	~300	/	~150	/
Cμ+6800	AP1	GTAATACGACTCACTATAGGGC	Cμ+6800_R	GAACGGAGGGCAAACTGTCA				
	AP2	ACTATAGGGCACGCGTGGT	Cμ+6800_nested_R	GTTAAAATGTGTCCCCTGTAAAT	/	~400	~100	~300
Cμ+6900	AP1	GTAATACGACTCACTATAGGGC	Cμ+6900_R	CCCACACTTGATGCAGGTGATATTT				
	AP2	ACTATAGGGCACGCGTGGT	Cμ+6900_nested_R	GAAGGGGTCAGCCGTTCT	/	~500	~200	~450
Cμ+7500	AP1	GTAATACGACTCACTATAGGGC	Cμ+7500_R	CAGAGCTTGGGCTGGCTTT				
	AP2	ACTATAGGGCACGCGTGGT	Cμ+7500_nested_R	CCGGGTGTGGACATTGAATCT	/	~1,050	~720	~910
Cμ+7700	AP1	GTAATACGACTCACTATAGGGC	Cμ+7700_R	GAAAAGTGTTGAAGGCACAACCTG				
	AP2	ACTATAGGGCACGCGTGGT	Cμ+7700_nested_R	GGCACAACCTGATTTCTCCTGACT		~1,200	~100	~1,100
Cμ+8700	AP1	GTAATACGACTCACTATAGGGC	Cμ+8700_R	GCGGAGGAGCTGTTTCCAA				
	AP2	ACTATAGGGCACGCGTGGT	Cμ+8700_nested_R	TGGAACAGATTCTCCCTGCCAAC	/	~2,400	~1,200	~2,300
Cμ+9500	AP1	GTAATACGACTCACTATAGGGC	Cμ+9500_R	CCCCAGGTGCTCCGTCTTAG				
	AP2	ACTATAGGGCACGCGTGGT	Cμ+9500_nested_R	TTCTACTCACCCCCTGCCTC	/	~300	~2,000	~450
breakpoint PCR								
BN polymorphism	VH3-42+20000_F	CCAGGACGCTCTGACATGCTCT	IGH_422700_R	GAAAACCCAGCTCAGCACAAACTG			239	
BN	IGH_423800_F	CAAGGCTCAGGCTTGGCTACTC	IGHC_1156200_R	CAGTTGCAGCTCTGGACAGGAA			273	
ID393, ID394	IGHBP2-5'_F	CCAGGGTTCGATGTTGGAACT	IGHBP2-3'_R	ACATCCTGGTTCCTCTGGCTC			242	
ID392, F7	IGHBP1-5'_F	ATGACCAGGGATGATTTAAGG	IGHBP1-3'_R	GTTCTCATTTCCCCACGTT			213	
F2	IGH_901600_F	ACAGTCACAGGACGCATCATTT	Cμ+6300_R	CCCAGTGCAGACTCCCCAGC			1,179	
SIOC	BP-SIOC 5'F	CAGATCCCAGGCAGGCTGTA	BP-SIOC 3'R	TTTTCCCTTTCCAGGTCTCTTC			293	

a. Milili *et al.* (2002) *Eur J Pediatr* **161**:479-484

Appendix 8. *BTK* gross deletion primers for deletion-scanning and breakpoint PCR.

Target	Forward primer	Primer sequence (5'-3')	Reverse primer	Primer sequence (5'-3')	Size (bp)
deletion-scanning PCR					
BTK_400	BTK_400_F	CGGAATTGTATTGATGGCTCTA	BTK_400_R	GCTCAGGGGTAATCCTCTAA	317
BTK_2900	BTK_2900_F	TCTCAGCTTCTGGTAACCATCCTT	BTK_2900_R	TGGGTGGAACTGGAGATCGTTAT	141
BTK_5800	BTK_5800_F	ATTATGTGGCAGGCACTGTTC	BTK_5800_R	AAGGGAAGAGCCGCTAGGTA	239
BTK_6800	BTK_6800_F	AGCAATCTCACTGGCCTTTAG	BTK_6800_R	CCCACAAAGCCACATAACACC	128
BTK_7500	BTK_7500_F	CTTTGCAGAGTAAGTGGCATGT	BTK_7500_R	GTGCAAGGCTTTGTGTGCTAGA	344
exon 1	BTK 1UII	GCAGCAGGGAACCAGATAGC	BTK 1LII	CCCTTTCGGCCTCTCTAATCAC	423
BTK_12000	BTK_12000_F	CAGTTATGGGGGATTGGTG	BTK_12000_R	GGTGTAGAAGGGGGTATCTGA	197
BTK_16200	BTK_16200_F	TAGGGACAGAGAAGGGTGGAT	BTK_16200_R	TCTTAGTCTTGCCAGGTTTTC	172
BTK_18200	BTK_18200_F	CACCAACCTTCTGGAGGACTCA	BTK_18200_R	GGTCTACCAGAATTGCCATGAT	183
BTK_19600	BTK_19600_F	AATGGAGGGCTGGTGGTT	BTK_19600_R	CTAATTCCGAACTTTCATTGG	416
exon 2	BTK 2U	TGCACATTTTTGTTCCATTTGAAAC	BTK 2LII	GACTTTCCGGCTTTAGCTAGTTA	448
exon 3	BTK 3U	TGGTTGCTTAATCCCTCTTAATCTTTC	BTK 3L	TGCACAGCATCACCAGTCTATTTACAG	249
BTK_22600	BTK_22600_F	GCATAATCCTGTTACCCAGTTAAA	BTK_22600_R	TACATCCATAATAGGGAATACCATAC	259
BTK_24000	BTK_24000_F	GTGAATTGCCCATTAATTTTTG	BTK_24000_R	GGGAGGAGAGGAGTTAGGTAGC	400
exon 4	BTK 4U	CTAAGGAAATTCAGAAAAGAGCAATGC	BTK 4L	CACCACCCCTTCTAATTGTGTTACAG	194
exon 5	BTK 5U	ACATAACCTGAACACCATTGCTGACT	BTK 5L	CTTCCTTCCTTTCCTTCTTTCTTTG	288
exon 6	BTK 6U	ATGCAAATGGTGGCTTCTCCTC	BTK 6L	GTACAACCTTATGCTATGACCCTGG	286
exon 7	BTK 7U	AACCTGTCCTTGGCCCCTTTC	BTK 7L	AAGTCCCAGGGCTAATTCTAAGACTCT	225
exon 8	BTK 8U	ATTTTAGTATGCCTGATTTGGGTGTCT	BTK 8L	ATGGGAATTATGCCGCAGCA	377
exon 9	BTK 9UII	ATACCTCCATTTGAGATTTGTTC	BTK 9L	ATATAGAGAGAGAGAGAGTTCCTCC TGG	218
exon 10	BTK 10U	GGGTGCTGTAACCTCCAATCTGC	BTK 10L	GTTCAAGATCCTCACTTATGCAAGG	176
exon 11	BTK 11U	AGCACCACTTCCTCCTACAGACAGC	BTK 11L	GGGACGGGCACAGCATCAA	204
exon 12	BTK 12U	TTGATGCTGTGCCCGTCCC	BTK 12L	TGCATTGCTTATCCTGGTGTCTGTA	301
exon 13	BTK 13UII	AGTGTGAGGCTTTAAGTGAGGA	BTK 13LII	ACGACAAAGCCCTTTATGAAT	373
exon 14	BTK 14UII	CCCGTTTCTGAGATTGACTTA	BTK 14LII	CCTCAGTGGGTTGTTGTGT	401
exon 15	BTK 15UII	GAGATTTAGGGTGGGAAAGAGGA	BTK 15LII	TGAAATGATGGCACCAGCAG	474
exon 16	BTK 16U	GCTTCCCTGGCTTCATTCTACTG	BTK 16L	TACTGTGCTATTTTTACTTCTGGAGGG	210
exon 17	BTK 17U	AATATCAAGCCTCCAAATCCTAATGC	BTK 17L	TGTCCAGTTTAGTTTTCATCCTCAGA	324
exon 18	BTK 18U	ACCCCTGCTATCCAAAAAGACTGC	BTK 18L	AGCTAAATGGGCAAGTAGATTCAAGG	273
exon 19	BTK 19U	TTTTTCCTTTTGTAATCGACAGGGC	BTK 19L	CAGCTTTCTGGAATACTGAGATAGGG	712
breakpoint PCR					
ID440	BTK_5800_F	ATTATGTGGCAGGCACTGTTC	BTK_18200_R	GGTCTACCAGAATTGCCATGAT	1,169
ID113	BTK_18200_F	CACCAACCTTCTGGAGGACTCA	BTK 4L	CACCACCCCTTCTAATTGTGTTACAG	3,857
ID434	BTK_34800_F	CACGTACAAGTTTAGTGCCAAG	BTK_43400_R	GCTACTCGGGAGGCTAAGG	462

Appendix 9. *Artemis* gross deletion primers for deletion-scanning and breakpoint PCR.

Target	Forward primer	Primer sequence (5'-3')	Reverse primer	Primer sequence (5'-3')	Size (bp)
Deletion-scanning PCR					
ART_min40000	ART_min40000_F	CCAGGCAGGCACTTCTGACAC	ART_min40000_R	CCACGGAGGCTGGTCATGTT	107
ART_min25000	ART_min25000_F	CACAGGGTGTACAAACATGGTGT	ART_min25000_R	ATGGAAGGCAGAGATTGGACAAC	293
ART_min23100	ART_min23100_F	TTCACTTCGGCCACGGTT	ART_min23100_R	CATCACAGTCAGGAGCATGCATTC	139
ART_min21500	ART_min21500_F	CACAGGTGGTGTACACACACGATGT	ART_min21500_R	ATATGATCGGTTATATCCTTGG	224
ART_min10000	ART_min10000_F	AAGTCCAGCACAATTATACGAGAG	ART_min10000_R	CAGTTGCATGACCTTGAGCTAACA	187
ART_min8400	ART_min8400_F	GCCAACGATCTACCTGCACA	ART_min8400_R	ACAGGCATGAGCTACTGCACTC	237
ART_min6000	ART_min6000_F	TTGCCTGTGGTGGAATGGAA	ART_min6000_R	AGTGGCATTTGCAGTGGGTCT	204
ART_min1300	ART_min1300_F	ATTCACCGAACTTCCATAGCA	ART_min1300_R	CTTTAATGTGTTTCAGCCAATAGA	185
ART_1600	ART_1600_F	AGGCTTGACTTTTTGCTCATATT	ART_1600_R	GCCAGCAAAGAAACCACTCAG	499
ART_4800	ART_4800_F	AGCCAATGTAGGTCCCAGGTAG	ART_4800_R	ATTCCTCGCAAATGAATAGATAG	188
exon 1	Art 1F1	CTCCGGACTCCTCTGATTGG	Art 1R1	GGGACAAGGCGTGTGCT	390
exon 2	Art 2U	ACTGTTTCCCTTGGGCTAA	Art 2L	CAAAGCTAACCCATGATGACAA	349
ART_15600_F	ART_15600_F	GATGTGAGCCACTGGCACCTG	ART_15600_R	GTAGGCAATGGCTGCAAGAGAA	234
ART_16300_F	ART_16300_F	CTCAAGTGATCTGCCTGCCTCA	ART_16300_R	GCAGAGCCTTGCATCTGTGTGG	160
ART_17400_F	ART_17400_F	AGGTGTGGTGGCATGTGCT	ART_17400_R	GGCCATCCTGACAGACTCAC	140
ART_18500	ART_18500_F	GTGGCGGTATCTTTCCAGAC	ART_18500_R	TCAGAAATTACGCTGGCACAA	197
exon 3	Art 3U	TTGGTGGCACTGAAATGGT	Art 3L	CAAAGTGCTGGGATTATAGGTG	528
ART_19400	ART_19400_F	CTTGAACCCAGGAAGCAATG	ART_19400_R	AGCAAGAGGGTCAATGAGATTAAGC	243
ART_20000	ART_20000_F	AGGCTGTCATAAGGAAGGGATG	ART_20000_R	TTTCTGCCTTCTTTGCATTTTC	148
ART_21300	ART_21300_F	CGGGCTTCTGTTTGCTCTTA	ART_21300_R	GAGCTGGGGACAGAAGTGTT	209
exon 4	Art 4U	ATGCATTTACAAGGCGTTTAC	Art 4L	CCCCTAAAACCCCACTCA	470
exon 5	Art 5F1	AAGGATTCCACTTGTTTCTA	Art 5R1	TGCAGCCTCCAACTC	396
exon 6	Art 6U	CTGAGTGCCTGGTGTGGTT	Art 6L	GTGGTGGCACGTACCTGTT	397
exon 7	Art 7U	CCAGATAGTTGGGAGGCTGAG	Art 7L	GCTTTGCAGTTCAGCCACACA	523
exon 8	Art 8U	AGCATTAAGCCATCCACCA	Art 8L	GGAGGGAGGGAGGGACCTG	438
exon 9	Art 9U	GGTGTTTTGCCCTGAGTAA	Art 9L	ACGAAAAGTGTGGTCATGG	367
exon 10	Art 10U	AAGAGTCCAGCCCAGTTGT	Art 10L	AGTAGCATCCCCTCCATTT	501
exon 11	Art 11F1	CACACGCGGTCTACAA	Art 11R1	GGGGACTACCTGTCAACTAC	342
exon 12	Art 12U	AGCATTGAGAACCACAGCCATA	Art 12L	ACCCAGGAGAAGCTGACAACAC	421
exon 13	Art 13U	ACCGGGTTGCATTTAATTG	Art 13L	GGCACCGTTTTCACCACTT	574
exon 14	Art 14U	AGTGGCCTCCCTATTCTTT	Art 14L	GTACATCCCCATCAGCCTT	411
exon 14	Art 14UI	CTGCAAGGAGATCTGGGC	Art 14LI	CAAACTTTTTTCCTCGGGTA	504
exon 14	Art 14UII	TTCCTGCCTCTCTCATGGAACA	Art 14LII	TGTGAGCCACCACACCCAAC	451
breakpoint PCR					
ID389	Art 9U	GGTGTTTTGCCCTGAGTAA	ART_43400_R	TTTCAGCTATTCTCCTGGGTTA	1,376

Appendix 10. PCR primers for amplification and sequencing of *CD19* coding regions from genomic DNA.

Exon(s)	Forward primer	Forward primer sequence	Reverse primer	Reverse primer sequence
1	CD19exonF1	CCTGCCTGCCCCAGCATC	CD19exonR1	GGACACCCAGCCACAACTCAAG
2	CD19exonF2	CATGCCCACACCTCTCTCCCTC	CD19exonR2	GTGGCCTCCCTTCTCTCCTCCT
3	CD19exonF3	GGGTGTCTCTGCATTTGGTTCT	CD19exonR3	GTTTCCCCTCCCCTCTGTCTCT
4	CD19exonF4	CAGAGACAGAGGGGAGGGGAAA	CD19exonR4	GGAGTCAGTGTTCTGTGTGGAAGT
5	CD19exonF5	CCTGGGGTCCTCGTCTCATA	CD19exonR5	GGGAAGATGGAGTAGGGTTT
6	CD19exonF6	CCTGCCAAGCAGCCCCTC	CD19exonR6	CAAGGCATAAAGTGAAGGTGGA
7	CD19exonF7	GGCCTGGACAGAATCTTGAGTG	CD19exonR7	CACAGCGGATTGGAGGGATGA
8	CD19exonF8	CATCTCTTCTGACCCCGTCTTA	CD19exonR8	ATGGGGATGGGGAACCGAGT
9+10	CD19exonF9+10	GTCTGTGGCCCGAATAGTG	CD19exonR9+10	CGCTGCAATTACACCTAAGTCTC
11+12	CD19exonF11+12	GCGCTGTGACACTCCTAGAA	CD19exonR11+12	GGGTCAGGGTCAGGTCAAAGTT
13+14	CD19exonF13+14	CCTCTGCCCCTCAGTCTTCCTC	CD19exonR13+14	CTGCCATCCGTGCCTATCT
15	CD19exonF15	GTTCCTAGGGCCTGGTGG	CD19exonR15	ATTCAAGACGCCTTTCCTTCC

Appendix 11. PCR primers and TaqMan probe for quantification of *CD19* transcripts.

Primer	Name	Sequence
Forward primer	CD19_RQ_F	CGTGGCAACCTGACCATGT
Reverse primer	CD19_RQ_R	CCACAAGGGAACACAGGCAG
TaqMan FAM-TAMRA probe	Tr-CD19	ACCAGTCCTCAGCAGCCAGT

Appendix 12. PCR primers and TaqMan probes for KREC and IgκREHMA assays.

Target	Forward primer	Primer sequence (5'-3')	Reverse primer(s)	Primer sequence (5'-3')	TaqMan probe	Probe sequence (5'-3')
Generation of human KREC signal joint construct				*Generation of human KREC signal joint construct*		
intronRSS-Kde KREC	KREC_BamHI_F	GGATCCTATCAGCCCAGGATAATGCTTAT	KREC_EcoRI_R	CTGGCTCCTGAATTCCCACC		
intronRSS-Kde KREC	KREC_EcoRI_F2	GTGGGAATTCAGGAGCCAGCTCT	KREC_XhoI_R	CTCGAGGGGTAAACGACAGCAACACA		
KREC assay				*KREC assay*		
IRS1-RS	5'IRS1_F	CGCTAAGGGCCATGTGAAC	3'RS_R	TTTAAAGCTACATTAGGGCTCAAATCT	T-3'RS	TGGCAGCCCAGGGTGGATCTCC
IRS1-RS KREC	5'RS_F	GGAGTGGATTCAGGACACTGCT	3'IRS1_R	CTCCAATAAGTCACCCTTTCCTTGT	T-5'RS	CCAGTTTCTGCACGGGCAGTCAGTTAG
pGEMT-easy	pgemtez_F	GTCACCTAAATAGCTTGGCGTAATC	pgemtez_R	CACGACAGGTTTCCCGACTG	T-pgemtez	CCACACAACATACGAGCCGGAAGCATAA
intronRSS-Kde	Intron-F1-TM	CCCGATTAATGCTGCCGTAG	Kde-R1-TM	CCTAGGGAGCAGGGAGGCTT	T-Kde-cons1	AGCTGCATTTTTGCCATATCCACTATTTGGAGT
intronRSS-Kde KREC	Int-Kde BREC F	TCAGCGCCCATTACGTTTCT	Int-Kde BREC R	GTGAGGGACACGCAGCC	T-Kde-RSS_2	CCAGCTCTTACCCTAGAGTTTCTGCACGG
human albumin [a]	F-ALB	TGAAACATACGTTCCCAAAGAGTTT	R-ALB	CTCTCCTTCTCAGAAAGTGTGCATAT	T-ALB	TGCTGAAACATTCACCTTCCATGCAGA
IgκREHMA				*IgκREHMA*		
Vκ3-20 - Jκ	Vκ3-20_intron_F	CTCAAGTCCTGTTACCTGGCAA	Jκ1-4 [b]	CTTACGTTTGATCTCCACCTTGGTCCC		
			Jκ5 [b]	CTTACGTTTAATCTCCAGTCGTGTCCC		

a. Pongers-Willemse *et al.* (1998) *Leukemia* **12**:2006-2014.
b. Van Dongen *et al.* (2003) *Leukemia* **17**:2257-2317.

ABBREVIATIONS

AID	activation-induced cytidine deaminase
APRIL	a proliferation-inducing ligand
BAFF	B-cell activating factor
BCMA	B-cell maturation antigen
BCR	B-cell antigen receptor
BER	base-excision repair
bHLH	basic helix-loop-helix
BM	bone marrow
C	constant
CDR	complementary-determining region
CLP	common lymphoid progenitor
CMP	common myeloid progenitor
CSR	class switch recombination
CVID	common variable immunodeficiency
CyIg	cytoplasmic Ig
D	diversity
DNA-PKcs	catalytic subunit of DNA-dependent protein kinase
ds	double-stranded
EBF	early B cell factor
ERV	endogenous retrovirus
FDC	follicular dendritic cell
FR	framework
GO	Gene Ontology
GVHD	graft-versus-host disease
HGMD	human gene mutation database
HMG	high mobility group
HSC	Hematopoietic stem cell
ICOS	inducible co-stimulor
IDbase	immunodeficiency mutation database
Ig	immunoglobulin
IgH	Ig heavy chain
Igκ	Ig kappa light chain
IgκREHMA	Igκ restriction enzyme hot-spot mutation assay
Igλ	Ig lambda light chain
ITAM	immunoreceptor tyrosine-based activation motif
J	joining
kb	kilobases
Kde	kappa-deleting element
KLF	Krüppel-like factor
KLH	keyhole limpet hemocyanin
KREC	kappa-deleting recombination excision circle
LIG4	DNA ligase IV

LINE	long interspersed element
LM-PCR	ligation-mediated PCR
LR-PCR	long-range PCR
MAR	matrix associating region
Mb	megabases
MLP	multilineage progenitor cell
MLS	multi-loop subcompartment
MMR	mismatch repair
MZ	marginal zone
NHEJ	nonhomologous end joining
NIP	4-hydroxy-3-iodo-5-nitrophenylacetate
NK	natural killer
N-nucleotide	random nucleotide
OMIM	online mendelian inheritance in man
ORF	open reading frame
PID	primary immunodeficiency
P-nucleotide	palindromic nucleotide
Pol	polymerase
RAG	recombinase activating gene
RQ-PCR	real-time quantitative PCR
RSS	recombination signal sequence
RW/GL	random-walk/giant-loop
S	Ig switch region
SCID	severe combined immunodeficiency
SHM	somatic hypermutation
SINE	short interspersed element
SmIg	surface membrane Ig
SNP	single nucleotide polymorphism
TACI	transmembrane activator and CAML interactor
TCR	T-cell receptor
TD	T-cell dependent
TdT	terminal deoxynucleotidyl transferase
TI	T-cell independent
TE	transposable element
TLR	Toll-like receptor
TNF	tumor necrosis factor
TREC	T-cell receptor excision circle
UCB	umbilical cord blood
UNG	uracil-N-glycosylase
V	variable
XLP	X-linked lymphoproliferative disease

SUMMARY

B lymphocytes are generated throughout life from hematopoietic stem cells in bone marrow, and contribute to the immune system by the production of antigen-specific antibodies (immunoglobulins; Ig). Two distinct phase of B-cell development can be distinguished: 1) antigen-independent precursor-B-cell differentiation in bone marrow, and 2) antigen-dependent B-cell maturation in peripheral lymphoid organs. The aim of precursor-B-cell differentiation is to generate a functional Ig receptor by V(D)J recombination of the genes encoding the Ig heavy (IgH) and Ig light (Igκ or Igλ) chains. When a precursor-B-cell succeeds in creating a functional Ig receptor, it will migrate to the periphery and become part of the naive B-cell pool. Because every precursor-B-cell creates a unique Ig receptor, the peripheral B-cell pool bears a diverse repertoire of specific receptors for antigen.

Naive B lymphocytes are short-lived cells and because new cells are continuously generated, there is a high turnover. However, once a naive mature B-lymphocyte recognizes antigen with its specific Ig receptor, it will undergo clonal proliferation and differentiation, thereby generating a large number of plasma cells that produce and secrete antigen-specific Igs. During this response, the B-cell initiates additional molecular mechanisms to adapt and optimize the antigen-binding affinity and the effector function of its Ig molecule.

Generation of a large repertoire of B lymphocytes and the response of one of these B lymphocytes to antigen is a multi-step process for which multiple proteins are required. The lack of one of these proteins leads to a problem with the humoral immunity of the individual, which is seen in patients with primary antibody deficiency diseases. These children or young adults carry rare inherited disorders and are subject to multiple, recurrent (mainly) bacterial infections. Several genetic defects have been identified in patients with a primary antibody deficiency. In many patients, however, the B-cell defect is not well understood. The studies in this thesis address several aspects of B-cell development and defects that lead to primary antibody deficiency diseases.

Most primary antibody-deficient patients with extremely low serum Ig levels carry a defect in precursor-B-cell differentiation in bone marrow. Several gene defects (*BTK*, *IGHM*, *BLNK*, *CD79A*, *L14.1*) have been identified that result in a block in precursor-B-cell differentiation, but the nature of the defect in (5 - 10 %) of the patients remains unknown. In **Chapter II**, we studied normal precursor-B-cell differentiation in childhood bone marrow. We purified precursor-B-cell subsets with membrane markers only, thus enabling the extraction of both DNA for analysis of the Ig gene rearrangements using real-time quantitative (RQ-)PCR and GeneScan assays, and RNA for genome-wide gene expression profiling using Affymetrix GeneChip arrays. Based on the Ig gene rearrangement patterns we divided human precursor-B-cell differentiation into five functional stages. Pro-B cells already massively initiate D_H-J_H rearrangements, which are completed with V_H-DJ_H rearrangements in pre-B-I cells. Large cycling pre-B-II cells are selected for in-frame *IGH* gene rearrangements. The first *IGK/IGL* gene rearrangements were initiated in pre-B-I cells, but their frequency increased strongly in small pre-B-II cells, and in-frame selection was found in immature B cells. Transcripts of the *RAG1* and *RAG2* genes, and earlier defined transcription factors, such as *E2A*, *EBF*, *E2-2*, *PAX5* and *IRF4*, were specifically up-regulated at stages undergoing Ig gene rearrangements. Careful analysis of genome-wide expression profiles of these five subsets enabled us to

identify 16 novel candidate genes for initiation and regulation of *IGH* and *IGK/IGL* gene rearrangements, thereby providing insight in the mechanism of precursor-B-cell differentiation. Furthermore, this detailed analysis represents an excellent template for studies on B-cell differentiation blocks in patients with a primary antibody deficiency.

The Ig gene rearrangements status can be used to determine the differentiation stage of a precursor-B cell. The induction of Ig gene rearrangements appears to be tightly controlled, but the molecular mechanisms underlying the hierarchical order found in **Chapter II** are largely unknown. In **Chapter III**, we studied how a large region spanning three megabases allows efficient joining of V and D region gene segments. Using a 3D fluorescence *in situ* hybridization approach we labeled eleven regions of the *IGH* locus with small DNA probes in uncommitted E2A-deficient pre-pro-B and committed RAG-deficient pro-B cells. Subsequently, we measured the end-to-end physical distances separating these markers to determine the structural organization of the murine *IGH* locus. Our observations show that in pre-pro-B and pro-B cells, the distance distribution within the *IGH* locus is most consistent with a 30 nm fiber that is organized into two rosettes that contain multiple loops (two 'flowers'). The data also indicate that in pre-pro-B cells V_H regions positioned in proximal and distal genomic regions are localized at similar spatial distances from the D_H and J_H elements. Furthermore, we show that it is the geometry of the *IGH* locus in interphase pro-B cell nuclei that allows V_H and D_HJ_H gene segments, as well as V_H promoters and the intron enhancer, to find each other over large genomic distances with relatively high frequency. This is the first study to demonstrate the structural organization of a genetic locus.

The majority of gene disruptions underlying human inherited diseases are micro-lesions smaller than 20 bp. Whereas this is true for *BTK* gene disruptions, about half of the mutated *IGHM* and *Artemis* alleles are disrupted by gross deletions. A mechanism for this striking difference has never been reported. In **Chapter IV**, we studied the gross deletion breakpoints involving the *IGHM*, *BTK* and *Artemis*. The *IGH* locus is complex and in developing B cells, large genomic regions are deleted during V(D)J recombination and Ig class switch recombination. We did not find involvement of RSS or Ig switch regions in the breakpoints of five unique *IGHM* gross deletions. In contrast, most gross deletion breakpoints were located in or near transposon-derived elements (TE). Similar observations were made for gross deletions in *BTK* and *Artemis*. However, whereas the *IGH* gross deletion breakpoints showed involvement of long interspersed elements (LINE) and LTR retrotransposons (ERV), *BTK* and *Artemis* gross deletion breakpoints were mainly located in or near short interspersed elements (SINE). This difference could be explained by the fact that LINE and ERV elements are specifically overrepresented in the *IGH* locus as compared to the average of the human genome, and SINEs are overrepresented in the *BTK* and *Artemis* genes. The high frequency of gross deletions in *IGH* and *Artemis* was associated with a high total TE content. We performed similar analyses on literature data of extensively studied genes in human genetic disorders. The observations made in *IGHM*, *BTK* and *Artemis* were confirmed in a total set of 20 genes, thereby supporting a model in which the TE content of a genomic region influences the gross deletion incidence. Thus our data shed some light on the mechanism underlying gene disruption by gross deletion.

In **Chapter V**, we described four antibody-deficient patients from two unrelated families with strongly reduced serum Igs. These patients had normal numbers of B cells in their

peripheral blood, but these cells lacked membrane expression of CD19. Sequence analysis revealed homozygous mutations in the *CD19* gene that resulted in premature stop codons. The CD19 protein forms a complex with CD21, CD81, and CD225 on the membrane of mature B cells. Together with the B-cell antigen receptor, this complex signals the B cell to decrease its threshold for activation by the antigen. Expression of CD21 was reduced, whereas the other two CD19-complex members CD81 and CD225 were found to be normally expressed on patients' B cells. All four patients showed reduced numbers of $CD27^+$ memory B cells. Small-to-moderate size germinal centers were found in secondary lymphoid organs, and Ig class switched *IGH* transcripts contained somatic mutations. However, stimulation of B cells with anti-IgM resulted in impaired Ca^{2+} flux and proliferation. Furthermore, the antibody responses upon rabies vaccination were quantitatively and qualitatively impaired. This study shows that the disruption of CD19 signaling results in a primary antibody deficiency, mainly characterized by a poor antigen-specific response. On the basis of the crucial role of CD19 in signaling by the B-cell receptor on antigen recognition, it is likely that defects in other members of the CD19-complex (CD21, CD81, and CD225) also lead to antibody deficiency in humans.

The contribution of proliferation to B-lymphocyte homeostasis and antigen responses is largely unknown. In **Chapter VI**, we quantified the replication history of mouse and human B-lymphocyte subsets by calculating the ratio between genomic coding joints and signal joints on kappa-deleting recombination excision circles (KREC) of the *IGK*-deleting rearrangement. We demonstrate that naive mature B lymphocytes, but not transitional B lymphocytes, undergo *in vivo* homeostatic proliferation in absence of somatic mutations in the periphery. T-cell dependent B-cell proliferation was substantially higher and showed higher frequencies of somatic hypermutation than T-cell independent responses, fitting with the robustness and high affinity of T-cell dependent antibody responses. More extensive proliferation and somatic hypermutation in antigen-experienced B lymphocytes from human adults compared to children indicated consecutive responses upon additional antigen exposures. Our combined observations unravel the contribution of proliferation to both B-lymphocyte homeostasis and antigen-induced B-cell expansion. We propose an important role for both processes in humoral immunity. These new insights will support the understanding and the identification of defects in homeostatic or antigen-induced B-cell proliferation in patients with common variable immunodeficiency or another antibody deficiency.

The studies described in this thesis support the understanding of important processes in B-cell development and the molecular basis of primary antibody deficiencies. Still, in many patients suffering from an antibody deficiency, the B-cell deficiency and the possible molecular basis are not yet understood. Extensive characterization of the B-cell compartment in these patients should facilitate the identification of B-cell differentiation blocks in these patients. Furthermore, new approaches in DNA chip technology can support molecular diagnostics by linkage analysis in large families or the direct identification of gene disruptions in likely candidate genes.

SAMENVATTING

B cellen zijn witte bloedcellen, die een belangrijke rol spelen in het immuunsysteem, omdat ze antistoffen produceren. Antistoffen kunnen een ziekteverwekker of ander lichaamsvreemd bestandsdeel (antigeen) herkennen en door er stevig aan te binden speelt het een belangrijke rol in het onschadelijk maken van ziekteverwekkers. B cellen worden gedurende het hele leven aangemaakt in het beenmerg, waar ze ontwikkelen uit bloedstamcellen. Tijdens de voorloper-B-cel ontwikkeling in het beenmerg creëert elke voorloper-B cel een uniek antistofmolecuul dat als receptor op het celoppervlak wordt gezet om antigeen te herkennen. Dit antistofmolecuul, ook wel immuunglobuline (Ig) genoemd, bestaat uit twee Ig zware eiwitketens (IgH) en twee Ig lichte eiwitketens (Igκ of Igλ).

Een antistof is specifiek voor één antigeen, en aangezien er enorm veel verschillende antigenen zijn, moeten er dus minstens zo veel verschillende antistoffen worden aangemaakt om het lichaam te beschermen. Om dit te kunnen doen worden in elke voorloper-B cel de stukken DNA die coderen voor de IgH, Igκ en Igλ eiwitketens herschikt. Het *IGH* gen bevat 66 variable (V), 27 diversity (D) en 6 joining (J) gensegmenten. Een speciaal herschikkingsmechanisme zorgt ervoor dat in een voorloper-B cel een V, een D en een J gensegment aan elkaar worden gekoppeld. Dit zogenoemde V(D)J recombinatie proces vindt ook plaats in de *IGK* en *IGL* genen, die uitsluitend V en J segmenten bevatten. De V(D)J koppelingen zijn niet precies, doordat DNA bouwstenen (nucleotiden) op de koppelplaats kunnen worden verwijderd en willekeurig worden ingevoerd; dit zorgt voor extra diversiteit. Er zijn drie factoren die bijdragen aan de diversiteit van het antistofrepertoire: de combinatie van V(D)J gensegmenten, de V(D)J koppelplaats, en de combinatie van Ig zware en Ig lichte ketens. Gezamenlijk kan dit een repertoire van meer dan duizend miljard ($>10^{12}$) verschillende antistoffen opleveren.

Elke voorloper-B cel maakt dus een eigen unieke Ig molecuul en brengt hiervan duizenden exemplaren op het celoppervlak tot expressie. Vervolgens verlaat deze B cel het beenmerg via de bloedbaan en migreert tussen de verschillende lymfoïde organen, zoals de milt en lymfklieren, op zoek naar antigeen. Wanneer een B cel antigeen herkent, zal deze worden geactiveerd en meerdere celdelingen doormaken, zodat er veel B cellen ontstaan met dezelfde Ig moleculen. Deze dochtercellen brengen willekeurige veranderingen aan in het DNA van hun Ig zware en Ig lichte keten. Deze veranderingen kunnen de bindingscapaciteit aan antigeen beïnvloeden. Doordat alle dochtercellen geselecteerd worden op goede binding met antigeen, vindt er selectie plaats van B cellen met Ig moleculen die het sterkst aan het antigeen binden. De geselecteerde dochtercellen rijpen vervolgens uit tot plasmacellen, die antistoffen produceren, of tot geheugencellen, die zeer snel in actie kunnen komen zodra hetzelfde antigeen nog een keer het lichaam binnenkomt.

Patiënten met een primaire antistofdeficiëntie kunnen geen goede antistoffen maken. Dit zijn vaak jonge kinderen, die daardoor meerdere zware bacteriële infecties per jaar doormaken. De enige behandeling voor deze patiënten is om elke drie weken antistoffen van gezonde donoren te krijgen. Bovendien hebben ze regelmatig antibiotica kuren nodig om infecties snel en tijdig te bestrijden. Deze patiënten hebben veelal een aangeboren afwijking, die wordt veroorzaakt door mutatie van een gen dat belangrijk is voor een goede B-celontwikkeling. In dit proefschrift is onderzoek gedaan naar de B-celontwikkeling en de

moleculaire processen nodig zijn om goede antistoffen aan te maken, maar die mogelijk defect kunnen zijn bij patiënten met een antistofdeficiëntie.

In **Hoofdstuk II** hebben we de voorloper-B-celontwikkeling bestudeerd in het beenmerg van gezonde kinderen. Wij konden deze ontwikkeling opsplitsen in 5 verschillende stadia, en cellen van elk stadium isoleren. Pro-B cellen (1) ondergingen D–J genherschikkingen voor de Ig zware keten, terwijl V–DJ genherschikkingen pas werden gevonden in pre-B-I cellen (2). Vervolgens vond er in large pre-B-II cellen (3) selectie plaats op functionele Ig zware ketens, waarna in small pre-B-II cellen (4) V–J genherschikkingen plaatsvonden voor de Ig lichte ketens. Tot slot werden immature B cellen (5) geselecteerd voor functionele Ig lichte ketens. Vervolgens hebben we de voorloper-B cellen van deze 5 stadia onderzocht om vast te stellen met DNA microchips welke genen tot expressie kwamen. Dit verschafte ons meer inzicht in de processen die plaatsvonden in deze 5 stadia van de normale voorloper-B-celontwikkeling. Deze inzichten kunnen we in de toekomst gebruiken voor verder onderzoek bij antistofdeficiënte patiënten met een probleem in de voorloper-B-celontwikkeling in het beenmerg. Het zou interessant zijn om de geblokkeerde B-celpopulatie van deze patiënten te vergelijken met de normale tegenpool van deze studie.

In **Hoofdstuk III** hebben we de structuur van het *IGH* gen bestudeerd in muizenmodellen. Het *IGH* gen is bijzonder groot en tijdens voorloper-B-celontwikkeling vindt er V(D)J recombinatie plaats tussen gensegmenten die ver uit elkaar liggen. Het is nog onduidelijk hoe deze gensegmenten bij elkaar worden gebracht. Om dit te bestuderen hebben we stamcelachtige pre-pro-B cellen uit E2A-deficiënte muizen en pro-B cellen uit RAG-deficiënte muizen geïsoleerd. Beide muizensoorten hebben een blokkade in de voorloper-B-celdifferentiatie en zijn niet in staat om V(D)J recombinatie te ondergaan, zodat het recombinatieproces niet de structuur van het *IGH* gen kan beïnvloeden. In deze cellen hebben we met fluorescerende DNA markers het *IGH* gen aangekleurd en in 3D opnames de onderlinge afstanden bepaald. In totaal hebben we de afstanden tussen 11 markers in het IgH gen en een marker net buiten het *IGH* gen bepaald. Op grond van deze data hebben we een model kunnen maken van de structuur van het *IGH* gen. Het *IGH* gen bestaat uit 1 of 2 subcompartimenten die het best kunnen worden beschreven als bundels van lussen, gescheiden door een tussenliggende verbinding. Hieruit bleek dat in pro-B cellen, de V_H gensegmenten dichter bij de D_H en J_H gensegmenten liggen dan in de onrijpere pre-pro-B cellen. Ook de V_H gensegmenten die theoretisch verder weg lagen, bleken in 3D ruimte net zo dichtbij als de andere V_H gensegmenten. Bovendien bleek de structuur van het *IGH* locus zodanig dat specifiek in pro-B cellen de V_H gensegmenten de D_H en J_H gensegmenten elkaar over grote genomische afstanden vinden en V(D)J recombinatie mogelijk is. Dit onderzoek toont voor het eerst de structuur van een gen aan en hoe interactie tussen ver uiteen gelegen elementen mogelijk wordt gemaakt. In vervolgstudies zal worden onderzocht welke DNA elementen een belangrijke rol spelen in de regulatie van de structurele organisatie van een gen.

Het grootste deel (>90 %) van de mutaties in genen die leiden tot een erfelijke ziekte zijn puntmutaties, kleine deleties, of kleine inserties (<20 nucleotiden). Grote veranderingen komen minder frequent voor en grote deleties vormen hierin de grootste groep (~6 % van de mutaties). In **Hoofdstuk IV** hebben we grote deleties van *IGHM*, *BTK* en *Artemis* bestudeerd om meer te weten te komen over de onderliggende mechanismen. Grote deleties komen in de *IGHM* en *Artemis* genen komen frequent voor vergeleken met grote deleties in

het *BTK* gen (~50 % t.o.v. 6 % van de mutaties). Daarnaast bevat het *IGH* locus vele DNA motieven die bij herschikkingsprocessen tijdens B-celontwikkeling grote deleties mediëren. Met PCR technieken hebben we de breukpunten van 5 verschillende *IGHM* deleties onderzocht. Na uitgebreide analyse vonden we geen aanwijzing voor DNA motieven die betrokken konden zijn bij het optreden van de grote deleties. De meeste breukpunten waren echter wel gepositioneerd in of vlakbij DNA sequenties die zijn afgeleid van transposon elementen. Dergelijke elementen komen frequent voor in het humane genoom en beslaan ongeveer 37 % van het totale genoom. Analyse van de breukpunten van de grote deleties in de *BTK* en *Artemis* genen liet ook co-localisatie met transposon elementen zien. De hoge frequentie van grote deleties in *IGHM* en *Artemis* ging gepaard met veel transposon elementen in deze genen (41 % en 45 % t.o.v. 29 % voor *BTK*). Bovendien bleek dat juist het meest voorkomende type transposon element in het betreffende gen betrokken was bij de grote deleties. Deze waarneming werd bevestigd in een grote serie van 20 genen, die gemuteerd zijn bij bepaalde erfelijke ziekten. Daarom concluderen wij dat transposon elementen waarschijnlijk een belangrijke rol spelen in het optreden van grote deleties, en dan met name als ze relatief oververtegenwoordigd zijn in een bepaald gen. Hoe deze elementen grote deleties mediëren is vooralsnog onbekend, maar ze zouden een rol kunnen spelen in de structuur van het gen of extra gevoelig zijn voor het optreden van DNA breuken. Deze studie verschaft meer inzicht in het optreden van mutaties als oorzaak van erfelijke ziektes.

Bij een groot deel van de patiënten met een antistofdeficiëntie is (nog) geen genetische afwijking gevonden. In **Hoofdstuk V** beschrijven wij voor het eerst een genetische afwijking in het *CD19* gen bij 4 patiënten van 2 niet-gerelateerde families. Deze patiënten hebben normale aantallen B cellen in het bloed, maar deze cellen zijn niet in staat om voldoende te reageren op antigeen. Dit komt omdat het CD19 eiwit een belangrijke rol speelt bij het doorgeven van het activatie signaal aan de cel op het moment van de herkenning van antigeen met het Ig molecuul op het celoppervlak van de B cel. Wij hebben laten zien dat geïsoleerde B cellen van de CD19-deficiënte patiënten nauwelijks reageren op signalen via de Ig moleculen. Als gevolg hiervan hebben deze patiënten sterk verlaagde antistofconcentraties in het bloed en bleken ze niet in staat om voldoende specifieke antistoffen te maken na rabiës vaccinatie. Deze studie draagt bij aan verder begrip van antistofdeficiënties en in het onderscheiden van specifieke B-cel defecten in deze groep patiënten. Aangezien CD19 in een complex met CD21, CD81 en CD225 functioneert, zouden mutaties in die genen ook kunnen leiden tot een zelfde type antistofdeficiëntie.

In de reactie van een B cel op antigeen speelt proliferatie een belangrijke rol. Hoeveel celdelingen een B cel ondergaat in deze reactie is echter onbekend. Ook is de bijdrage van proliferatie van naïeve B cellen, die nog geen antigeen reactie hebben doorgemaakt, aan de omvang van de perifere B-cel populatie onbekend. In **Hoofdstuk VI** hebben we de mate van proliferatie van perifere B-cel subpopulaties onderzocht met een nieuwe DNA techniek. Met deze methode vonden we dat B cellen die recent uit het beenmerg zijn gekomen (transitionele B cellen) nog geen proliferatie hebben ondergaan. Rijpe naïeve B cellen hebben wel al 1 à 2 celdelingen doorgemaakt. Deze proliferatie is niet het gevolg van een reactie op antigeen, want deze cellen hebben nog geen aanvullende veranderingen in hun Ig genen aangemaakt. Marginale zone B cellen in de milt reageren zonder T-cel hulp op antigeen. Deze cellen kunnen ook in lymfklieren en bloed worden gevonden als “natural effector

B cellen", en hebben 4 celdelingen doorgemaakt. Deze cellen bevatten wel veranderingen in hun Ig genen. De mate van proliferatie en Ig veranderingen waren echter nog groter in B cellen die een antigeen reactie met T-cel hulp hadden doorgemaakt (memory B cellen; 8 celdelingen). De T-cel afhankelijke B-cel reactie lijkt dus beduidend krachtiger te zijn dan de T-cel onafhankelijke B-cel reactie. De gemiddelde proliferatie van natural effector B cellen en memory B cellen was hoger bij volwassenen dan bij kinderen. Dit komt waarschijnlijk doordat een deel van deze cellen bij volwassenen meerdere keren een antigeen reactie heeft doorgemaakt. Deze studie verschaft nieuw inzicht in de bijdrage van proliferatie aan de opbouw van de perifere B-cel populatie en aan de antistof reactie tegen antigeen.

Dit proefschrift illustreert dat onderzoek naar fundamentele aspecten van B-celontwikkeling nieuwe inzichten oplevert betreffende het ontstaan van antistofdeficiënties. Aan de andere kant is gedetailleerd onderzoek van het B-celsysteem van patiënten met een antistofdeficiëntie buitengewoon belangrijk om nieuwe inzichten te verkrijgen in B-celontwikkeling. Zodoende benadrukt parallel onderzoek naar B-celontwikkeling en B-celdeficiënties de meerwaarde van kruisbestuiving tussen basaal en patiëntgebonden onderzoek.

DANKWOORD

De laatste loodjes wegen het zwaarst, en daar hoort in dit geval ook het dankwoord bij. Meer dan alle andere secties van een proefschrift is dit erg persoonlijk, en meer dan bij alle andere secties flitsen er allerlei gedachten en herinneringen door mijn hoofd. Bijna vijf jaar geleden begon ik aan dit onderzoek. Met dank aan Frank Staal kwam ik in contact met de afdeling Immunologie, waar ik nog voor mijn afstuderen een mijns inziens prachtig project kreeg voorgeschoteld binnen de groep van prof.dr. J.J.M. van Dongen, onder begeleiding van dr. M. van der Burg en in samenwerking met prof.dr. R. de Groot van de afdeling Kindergeneeskunde.

Als ik nu terug kijk denk ik dat hoe leuk het project mij toen ook al leek, de ervaringen die ik later heb opgedaan en de behaalde resultaten boven mijn verwachtingen zijn geweest. Dit was uiteraard te danken aan een gezamenlijke inspanning van vele mensen die ik bij deze graag wil bedanken voor hun onmisbare steun.

Mirjam, nog maar net gepromoveerd en begonnen als werkgroepleider van de primaire immuundeficiënties kreeg je mij onder je hoede. Vanaf het moment dat ik de afdeling bezocht stond je te popelen om te beginnen en dit enthousiasme duurt nog altijd voort. Bovendien kon ik altijd terecht met kleine en grotere problemen om deze samen op te lossen. Ik ben er zeker van dat jij nog vele promovendi zult begeleiden. Ik ben er dan ook nu al trots op de eerste te zijn geweest!

Beste Jacques, je was als mijn promotor enorm bevlogen en betrokken. Bovendien ben je een bijzonder kritische begeleider, maar ook zeer positief (met krul!) als je iets goed vindt. Hoewel je als hoofd van een grote unit met diverse onderwerpen altijd druk was in binnen- of buitenland, kon ik toch wanneer het nodig was direct contact met je opnemen, zelfs toen ik in San Diego zat. Ik heb dan ook veel van je mogen leren en hoop dat de komende jaren nog verder te kunnen doen.

Beste Ronald, de goede relatie en de gezamenlijke interesse van Jacques en jou in primaire immuundeficiënties hebben dit werk mogelijk gemaakt. De afgelopen jaren ben jij met name indirect bij mijn promotieonderzoek betrokken geweest. Het klinkt misschien vreemd, maar ik ben je daar juist heel erg dankbaar voor. Door de directe begeleiding in handen van Mirjam en Jacques te leggen, was de aansturing vanuit mijn oogpunt helder en heb ik nooit tussen twee vuren hoeven te zitten.

Tijdens mijn promotieonderzoek heb ik de mogelijkheid gekregen om een werkbezoek af te leggen in het lab van prof.dr. C. Murre aan de University of California, San Diego. Dit was voor mij een geweldige ervaring. Beste Kees, ik wil je graag hartelijk bedanken voor deze kans en voor het vertrouwen dat je me hebt gegeven. Hoewel je een goede groep leidt, ben je als een echte onderzoeker het liefst nog continue in het lab aan het werk. Dit gaf mij de kans om zowel conceptueel als praktisch gezien heel veel van je te leren. Het is dan ook een eer voor mij dat je in de kleine commissie zitting wilt nemen. In addition, I would like to thank Suchit Jhunjhunwala. With mixed feelings I look back to all those hours we spent in the basement behind the microscope (‘the beach’). It was tough, but thanks to your accuracy and great teamwork, we got the 10 kb probes to work.

Wim, hoewel we nooit aan een gezamenlijk project hebben gewerkt, heb ik zoveel van je geleerd. Bijna vier jaar hebben we een kamer gedeeld en back-to-back ons werk gedaan.

Tussendoor was er echter tijd om te reflecteren over ons werk, de immunologie, onderzoek in het algemeen en voetbal. Dankjewel hiervoor.

Helaas kan ik niet iedereen persoonlijk op deze plek bedanken. Wel wil ik graag iedereen van de afdeling Immunologie en in het bijzonder allen binnen de Unit Moleculaire Immunologie bedanken. In, addition, I would like to thank the people from the Murre lab for their kindness and willingness to help me out when needed. Barbara en Sandra, jullie hebben me vanaf het begin met raad en daad bijgestaan in het lab. Hetzelfde geldt voor Edwin, 'the sorter-guy': echt leuk dat we al die sorts én de Ca^{2+}-flux aan de praat hebben gekregen! Dick de Ridder was onmisbaar voor de microarray analyse en scherpe commentaren op manuscripten. Ik hoop dat je ook wat van ons, biomedici, geleerd hebt. De secretariële ondersteuning van Bibi, van Wendy voor de lay-out van het proefschrift en van Marieke voor de figuren was onmisbaar. Ik vond het heel leuk (en leerzaam) om met Nicole en Corinne samen te werken. Corinne, dankjewel dat je nu ook mijn paranimf wilt zijn.

Ik wil graag prof.dr. H. Hooijkaas en prof.dr. C.J.M. van Noesel danken voor het kritisch doorlezen van mijn proefschrift als leden van de kleine commissie. Ook dank ik prof. dr. R. Benner, hoofd van de afdeling Immunologie, voor het mogelijk maken van dit onderzoek. Ik dank Rob, prof.dr. E.A.M. Sanders, dr. N.G. Hartwig en dr. A. Lankester voor de toezegging zitting te nemen in de grote commissie; ik kijk erg uit naar de verdediging.

Tot slot wil ik het thuisfront bedanken. Papa en Mama, jullie zijn er altijd voor me geweest en hebben me altijd gesteund, hoe abstract de onderzoeksonderwerpen ook werden. Zonder jullie steun was ik nooit zover gekomen en ik vind het erg fijn dat jullie er nog altijd voor me zijn. Rosalie, al jaren zijn we nu uit huis en volgens mij hebben die ons contact er alleen maar beter op gemaakt. Ik ben er dan ook trots op mijn zusje als paranimf naast me te hebben.

Menno

CURRICULUM VITAE

Menno Cornelis van Zelm

Geboren: 24 juni 1979 te Utrecht

Opleiding

1991 – 1997	Christelijk Gymnasium Utrecht
1997 – 1998	Propedeuse Biologie, Universiteit Utrecht
1998 – 2000	Bachelor of Science (*cum laude*), University College Utrecht, Universiteit Utrecht
1999	Semester studie aan de University of California, San Diego, USA
2000 – 2002	Master of Science, Biologie, Universiteit Utrecht
2000	Stage 'Inhibition of Degranulation in Permeabilized RBL-2H3' Pharmacologie and Pathophysiologie, Utrecht Institute for Pharmaceutical Sciences (UIPS), Universiteit Utrecht Supervisie: Dr. F.A.M. Redegeld
2000 – 2001	Stage 'The effect of one point mutation on the structure and stability of a protein' Medicinal Chemistry, UIPS, Universiteit Utrecht Supervisie: Dr. E.E. Moret
2001 – 2002	Stage 'Molecular genetic analysis of a family of β-catenin interacting proteins involved in hypodermal morphogenesis' Hubrecht laboratory, Netherlands Institute for Developmental Biology, KNAW (Royal Netherlands Academy of Arts Sciences) Supervisie: Dr. H.C. Korswagen
2002	Scriptie 'Regulation of Wnt signaling by cell-cell interactions in *C. elegans*' Hubrecht Laboratory, Netherlands Institute for Developmental Biology, KNAW Supervisie: Dr. H.C. Korswagen

2002 – 2007 Promotie-onderzoek bij de afdelingen Immunologie en Kindergeneeskunde van het Erasmus MC/Erasmus Universiteit Rotterdam
'B-cell development and Primary Antibody Deficiencies'
Promotoren: Prof.dr. J.J.M. van Dongen, Prof.dr. R. de Groot
Copromotor: Dr. M. van der Burg

2005 – 2006 Werkbezoek
'Visualization of Ig Gene Rearrangement Processes in Precursor-B-cell Differentiation'
Division of Biological Sciences, University of California, San Diego, USA
Supervisie: Prof.dr. C. Murre

PUBLICATIONS

1. van Zelm MC, M van der Burg, D de Ridder, BH Barendregt, EFE de Haas, MJT Reinders, AC Lankester, T Révész, FJT Staal, and JJM van Dongen. Ig Gene Rearrangement Steps are initiated in Early Human Precursor-B-cell Subsets and Correlate with Specific Transcription Factor Expression. *J Immunol 2005;175(9):5912-22.*

2. van Zelm MC, I Reisli, M van der Burg, D Castaño, CJM van Noesel, MJD van Tol, C Woellner, B Grimbacher, PJ Patiño, JJM van Dongen, and JL Franco. An Antibody-Deficiency Syndrome Due to Mutations in the *CD19* Gene. *N Engl J Med 2006;354(18): 1901-12.*

3. van Zelm MC, T Szczepański, M van der Burg, and JJM van Dongen. Replication History of B-lymphocytes Reveals Homeostatic Proliferation and Extensive Antigen-induced B-cell Expansion. *J Exp Med 2007;204(3):645-55.*

4. Jhunjhunwala S, MC van Zelm, M Peak, R Riblet, JJM van Dongen, FG Grosveld, TA Knoch and C Murre. 3D-Architecture of the Immunoglobulin Heavy Chain Locus. *Submitted.*

5. van Zelm MC, C Geertsema, N Nieuwenhuis, D de Ridder, ME Conley, C Schiff, I Tezcan, E Bernatowska, NG Hartwig, EAM Sanders, J Litzman, I Kondratenko, JJM van Dongen, and M van der Burg. Gross Deletion of *IGHM*, *BTK* or *Artemis*: A model for Genomic Lesions Mediated by Transposable Elements. *Submitted.*