Technische Universität München

Lehrstuhl für Experimentelle Genetik

Diversification of the immunoglobulin genes: analysis of the molecular mechanisms in the chicken B cell line DT40

Ulrike B. Schötz

Vollständiger Abdruck der von der Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr. E. Grill

Prüfer der Dissertation:

1. apl. Prof. Dr. J. Adamski
2. Univ.-Prof. Dr. A. Gierl
3. Univ.-Prof. Dr. M.J. Atkinson

Die Dissertation wurde am 13.5.2009 bei der Technischen Universität München eingereicht und durch die Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt am 02.09.2009 angenommen.

Für Achim
Für meine Eltern

If we falter in resolve
Just because the task is hard,
No accomplishment can follow:
It is the world´s way.
- Emperor Meiji -

Cut, if need be, through thick briars,
Knots of brambles, tangled thorns,
For the path that´s yours to follow
Must be trodden to the end.
- Empress Shōken -

Table of contents

ZUSAMMENFASSUNG **1**

SUMMARY **3**

1 INTRODUCTION **5**

1.1 THE INNATE AND ADAPTIVE IMMUNE SYSTEMS ACT TOGETHER FOR A COMPLETE IMMUNE RESPONSE 5
1.2 B AND T CELLS RECOGNIZE FOREIGN PATHOGEN SUBSTANCES BY THEIR HIGHLY VARIABLE RECEPTORS 7
1.3 COMPOSITION OF A B CELL ANTIBODY MOLECULE 7
1.4 DIVERSIFICATION IS GAINED BY SEVERAL DISTINCT PROCESSES 8
1.4.1 V(D)J recombination is adding diversity to the gene loci by gene rearrangement 8
1.4.2 Hypermutation increases specificity of the antibody after encounter of antigen 10
1.4.3 Gene Conversion needs pseudogene templates to diversify the antibody genes 11
1.4.4 Class Switch Recombination changes the effector functions of the antibody molecule 12
1.5 DIVERSIFICATION IS REGULATED BY THE B CELL SPECIFIC ENZYME ACTIVATION-INDUCED CYTIDINE DEAMINASE (AID) 12
1.5.1 AID has high homology to the RNA-editing enzyme APOBEC-1 13
1.5.2 Regulation of AID is necessary to avoid genomic instability 14
1.6 URACIL DNA GLYCOSYLASE (UNG) IS INVOLVED IN THE PROCESSING OF THE AID-INDUCED DNA LESIONS 15
1.7 THE THREE DIVERSIFICATION PROCESSES INVOLVE DIFFERENT FACTORS OF DNA REPAIR 15
1.7.1 Error-prone repair in HM ends up with a diversified antibody gene after replication 16
1.7.2 Repair pathways that lead to GCV and CSR 17
1.8 SPECIFICITIES OF AID-DEPENDENT DIVERSIFICATION PROCESSES 18
1.8.1 Mutation frequency in hypermutating cells is strongly enhanced 18
1.8.2 The role of primary Ig sequences – AID can also target other sequences 19
1.8.3 Strand bias 19
1.8.4 Strong transcription is a prerequisite for AID-mediated diversification processes 19
1.8.5 Chromatin modifications seem to play a minor role in AID targeting 21
1.9 TRANS-ACTING FACTORS AND CIS-ACTING ELEMENTS IN IG GENE DIVERSIFICATION 21
1.9.1 AID is expressed only in B cells 22
1.9.2 AID recruitment needs additional factors 22
1.9.3 The role of Ig enhancers and matrix attachment regions (MARs) is not completely clarified 23
1.9.4 E-box motifs seem to be important for HM 24
1.9.5 E2A transcription factors bind to E-box motifs 24
1.10 DT40 CELLS AS MODEL SYSTEM 25
1.11 EFFECT OF E2A KNOCKOUT ON HM 26
1.12 AIMS 29

2 MATERIALS 31

2.1 BACTERIAL STRAIN 31
2.2 BUFFERS AND SOLUTIONS 31
2.3 CELL CULTURE 33
2.4 CELL LINES 33
2.5 CHEMICALS 34
2.6 CONSUMABLES 35
2.7 DNA SIZE MARKER 35
2.8 ENZYMES AND dNTPs 35
2.9 EXPERIMENTAL KITS 36
2.10 IMMUNO-STAINING ANTIBODIES & ANTI-ANTIBODIES 36
2.11 INSTRUMENTS 36
2.12 MEDIA 37
2.13 NUCLEOTIDE SEQUENCES 38
2.14 OLIGONUCLEOTIDES 38
2.15 PLASMIDS 38
2.16 SOFTWARE 38

3 METHODS 39

3.1 VECTOR DESIGN 39
3.1.1 Databases 39
3.1.2 E2A complementation vector 40
3.1.3 IgL targeting vectors for cis-element study 40
3.2 MOLECULAR BIOLOGY 44
3.2.1 Culture of E.coli 44
3.2.2 DNA ligation 44
3.2.3 Transformation 45
3.2.4 E.coli DH5α competent cell preparation 45
3.2.5 PCR amplification 46
3.2.6 Analysis of DNA by electrophoresis 46
3.2.7 PCR purification & Gel purification of DNA 47
3.2.8 DNA Purification by phenol/chloroform extraction 47
3.2.9 Ethanol precipitation of DNA 47
3.2.10 Topo Cloning 47
3.2.11 Plasmid preparation 48
3.2.12 Restriction enzyme digestion 48
3.2.13 Determination of DNA and RNA concentration 49
3.2.14 Genomic DNA isolation 49
3.2.15 Total RNA isolation 49
3.2.16 First strand cDNA synthesis 50
3.2.17 Sequencing 50
3.3 CELL CULTURE 51
3.3.1 Basic Cell Culture Conditions 51
3.3.2 Freeze down of cells 51
3.3.3 Thawing of cells 51
3.3.4 Transfection 51
3.3.5 Drug Resistance Marker Recycling 53
3.3.6 Subcloning of DT40 cells 54
3.3.7 Analysis of IgM expression by FACS 54
3.4 SOFTWARE FOR STATISTICS 56

4 RESULTS **57**

4.1 INACTIVATION OF THE E2A GENE 57
4.1.1 Complementation of the knockout using E2A cDNA expression cassettes *57*
4.1.2 mRNA levels of E2A do not correlate with IgL and AID transcription *61*
4.1.3 Ig mutation assay - E2A negative clones display strongly decreased rates of sIg loss *62*
4.1.4 Mutation pattern – Ig HM is reduced in the absence of E2A *64*
4.2 DEFINING CIS-ELEMENTS 67
4.2.1 A hypermutation reporter based on GFP expression *68*
4.2.2 How is AID-mediated diversification activated *70*
4.2.3 Fine mapping of 'W' *72*
4.2.4 Fine Mapping of 'S' fragment *75*
4.2.5 '0-2' series *76*
4.2.6 '2-4' series *77*
4.2.7 '0-4' series 1 kb deletions *78*
4.2.8 '0-4' series 200 bp internal deletions *79*
4.2.9 '0-4' series 400 bp internal deletions *81*
4.3 A NEW SERIES OF DELETIONS TO IDENTIFY SPECIFIC ACTIVE MOTIFS 83
4.3.1 '2-3' series 50 bp deletions 5'end *83*
4.3.2 '2-3' series 50 bp deletions 3'end *84*
4.3.3 '2-3' series 50 bp internal deletions *86*
4.3.4 Reconstitution and multimerization of '2.2-2.4' *87*
4.3.5 The 'HyCorE' sequence '2.2-2.4' starts HM at non-Ig loci *88*
4.3.6 The 'HyCorE' sequence '2.2-2.4' is conserved in other species *89*
4.3.7 GFP expression level – variations in HM of the GFP2 transgene are independent from GFP transcription *90*
4.4 EVALUATION OF CIS-ELEMENTS USING BOINFORMATICAL TOOLS 93

5 DISCUSSION **95**

5.1 E2A IS INVOLVED IN AID-MEDIATED DIVERSIFICATION 95
5.2 E2A TRANSCRIPTION FACTORS BIND TO THEIR DNA RECOGNITION MOTIF TO STIMULATE HM 97
5.3 E2A TRANSCRIPTION FACTORS REGULATE GENE CONVERSION AND CHROMATIN REMODELING 98
5.4 A GFP REPORTER REVEALED CIS-TRANS INTERACTIONS BALANCING AID-MEDIATED HM 99
5.4.1 The 'HyCorE' acts independently from transcription *100*
5.4.2 A 9.8 kb fragment of the IgL locus is able to start diversification at non-hypermutating loci *101*
5.5 THE 9.8 KB FRAGMENT 'W' WAS SHORTENED TO THE 4 KB SEQUENCE '0-4' 101
5.5.1 The 200 bp DNA element '2.2-2.4' is a true 'HyCorE' *102*
5.5.2 The 'HyCorE' starts HM at non-Ig loci *104*
5.5.3 The 'HyCorE' can be identified in other species *104*
5.5.4 The 'HyCorE' should also be relevant for gene conversion and class switch recombination *104*
5.5.5 Bioinformatical analysis – a theoretical discussion of putative motifs involved in the recruitment of HM *105*
5.6 A MODEL FOR AID RECRUITMENT TO HYPERMUTATING LOCI 108
5.7 OUTLOOK 109

6 LIST OF ABBREVIATIONS **111**

7 REFERENCES **113**

8 SUPPLEMENTARY INFORMATION ... 132

8.1 LIST OF PRIMERS ... 132
8.2 SEQUENCING ... 141
8.2.1 Sequencing of AID^R $\psi V^- E2A^{-/-}$ ... 141
8.2.2 Sequencing of $AID^R \psi V^- E2A^{RtE12}$ ... 147
8.2.3 Sequencing of $AID^R \psi V^- E2A^{RtE47}$ ... 154
8.3 SEQUENCE INFORMATION ... 165
8.3.1 Complete sequence of 'W' ... 165
8.3.2 Transcription factor binding sites of 'W' ... 168
8.4 TREATMENT OF DT40 MUTANTS WITH TRICHOSTATIN A ... 188

9 ACKNOWLEDGEMENTS ... 189

OWN PUBLICATIONS AND RESEARCH ACHIEVEMENTS ... 190

LEBENSLAUF ... 192

ERKLÄRUNG ... 193

Zusammenfassung

Die Affinitätsreifung der humoralen Immunantwort wird vermittelt durch Hypermutation (HM) der Immunglobulingene sowie Selektion von B-Zellklonen mit höherer Affinität nach Antigenkontakt. Das Enzym AID (activation-induced cytidine deaminase) desaminiert Deoxycytidinreste in transkribierter einzelsträngiger DNA. Das dabei entstehende Uracil ist der Initiator für die HM. AID ist stark mutagen und um die genomische Integrität zu gewährleisten ist es erforderlich, die HM gezielt auf die Immunglobulingene zu beschränken. Im Rahmen meiner Promotion habe ich untersucht, welche Rolle trans-agierende Faktoren und die korrespondierenden cis-DNA Elemente in der Rekrutierung von AID zu Orten der HM spielen. Es ist bekannt, dass das E-Box DNA Element CAGGTG die HM in einem Maus Igκ Transgen erhöht. Um die Rolle von E2A Transkriptionsfaktoren, die an dieses Element binden, näher zu untersuchen, habe ich das E2A Gen in der Hühner B Zelllinie DT40 inaktiviert. Der Knockout führte zu einem starken Abfall der Mutationsrate in der variablen Region des Immunglobulin Leichtkettenlokus (IgL Lokus) und dieser Effekt war unabhängig von der Transkriptionsrate des Immunglobulingens und der AID Expression. Eine Überexpression der E12 und E47 cDNA, welches die beiden Spleißvarianten des E2A Gens sind, konnte den Defekt komplimentieren. Dies führt zu der Schlussfolgerung, dass E2A-codierte Proteine die HM verstärken, indem sie AID rekrutieren. Um die exakte Sequenz und Position des korrespondierenden E-Box Elements und anderer involvierter Cis-Elemente zu ermitteln, habe ich ein 4 kb großes Fragment des IgL Lokus, welches den Enhancer enthält, einer Deletionsstudie unterzogen. Um die HM unabhängig von der Transkription untersuchen zu können wurde für die Studie ein GFP Reporter Konstrukt verwendet. Das GFP wird dabei Ziel für Mutationen sein, wenn es in der Position des hypermutierenden IgL Lokus inseriert wird. Mutationen, die die grüne Fluoreszenz des Proteins beeinträchtigen, können im Durchflusszytometer analysiert werden. Schrittweise Deletionen und Insertionen von Teilen des 4 kb Fragments anstelle des endogenen IgL Lokus ermöglichten es, ein 200 bp kleines Fragment zu identifizieren, welches in der Lage ist, die HM zu aktivieren. Dieses Fragment ist Teil des transkriptionellen IgL Enhancers. Es ist sowohl notwendig als auch allein ausreichend, um die HM in Gang zu setzen. Multimerisierung des Fragments erhöhte die Mutationsrate, was auf redundante DNA-Elemente innerhalb des 200 bp Fragments schließen lässt, die sich additiv auf die HM auswirken. Des Weiteren ist das Element in der Lage, HM auch an nicht hypermutierenden Loci auszulösen, wenn es dort inseriert wird.

Die korrespondierenden Sequenzen in Truthahn und Ente lösen ebenfalls HM aus, wenn sie in DT40 transfiziert werden. Dies ist die erste Studie, die eine Sequenz von nur 200 bp als Aktivator der HM ermitteln konnte.

Im Rahmen einer bioinformatischen Analyse konnte innerhalb dieser Sequenz das E-Box Element CAGCTG, ein NFκB Bindungsmotiv und ein ISRE (interferon-stimulated response element) als mögliche Kandidaten für eine Rekrutierung von AID identifiziert werden.

Summary

Affinity maturation of the humoral response is mediated by hypermutation (HM) of the immunoglobulin genes and selection of higher-affinity B cell clones after antigen encounter. The enzyme activation-induced cytidine deaminase (AID) is the only B cell-specific factor initiating HM by the deamination of deoxycytidine residues to generate uridine in transcribed single-stranded DNA. To ensure genomic integrity, AID-mediated HM needs to be targeted specifically to the immunoglobulin genes. During the course of my thesis, I shall identify trans-acting factors and their corresponding cis-binding DNA elements being involved in the recruitment of AID to sites of HM. Because the E Box motif CAGGTG was found to increase HM in a mouse Igκ transgene without affecting transcription, I tested the relevance of E2A transcription factors for HM by disruption of the E2A gene in a DT40 variant diversifying its immunoglobulin light chain (IgL) gene solely by HM and compared the mutation rates of E2A negative and E2A positive cells. The inactivation of the E2A gene strongly reduced the rate of mutations in the variable region of the IgL locus and this effect was not mediated by changes in the mRNA expression levels of surface immunoglobulin or AID. The defect is complemented by the expression of cDNAs corresponding to either of the two E2A splice variants E12 or E47. The results suggest that E2A-encoded proteins enhance immunoglobulin HM by recruitment of AID to the immunoglobulin loci. To find out more about the exact sequence and positioning of the E box and other cis-elements involved in AID recruitment, I started a serial deletion analysis of a 4 kb fragment of the IgL locus, which includes the transcriptional enhancer. I used a GFP reporter assay to evaluate HM independently from transcription. The GFP will be target for mutations when inserted in the position of the hypermutating IgL locus. Loss-of-function mutations can be monitored by FACS analysis. Step-wise deletions and insertions of the 4 kb fragment instead of the endogenous IgL locus enabled me to minimize the HM activating DNA-element to a 200-bp sequence being part of the IgL enhancer. It revealed to be both necessary and sufficient to confer the HM activity. Multimerisation of the fragment increased HM activity suggesting the existence of redundant motifs in the sequence that act in an additive manner. Beyond this, the 200 bp sequence, when inserted at non-immunoglobulin loci, is able to start HM there as well and the corresponding sequences of duck and turkey are able to start HM when transfected into DT40. A theoretical analysis using bioinformatical

tools identified an E-box element CAGCTG, an NFκB binding motif and an ISRE (interferon-stimulated response element) as probable candidates for a recruitment of AID.
The results significantly extend previously reported findings on AID mediated gene diversification. They show both by deletion and insertion analysis that cis-acting sequences and their trans-acting factors predispose neighbouring transcription units to HM. For the first time, a HM enhancing element could be identified which is small enough to make a statement on transcription factors which could be involved in AID recruitment and which can be used in further studies to elucidate the interplay of AID and possible cofactors in more detail.

1 Introduction

1.1 The innate and adaptive immune systems act together for a complete immune response

The immune system is an intrinsic defense system of the body to protect against diseases caused by viruses, bacteria, microbes and pathogen foreign substances. We distinguish between an innate and an adaptive immune system.

The innate immune system is congenital and mounts the first immune response to many common pathogens. Cells of the innate immune system are the phagocytes, which take up and destroy invading microorganisms by a process called phagocytosis.

Phagocytes are grouped into macrophages, which are located in the epithelials, and neutrophilic cells, which are located in the blood. Both carry a limited number of unalterable surface receptors which recognize a restricted set of common structures such as bacterial surface proteins. The receptors can distinguish between surface molecules of the body and pathogenic ones. One group of receptors recognizes foreign surface structures and induces phagocytosis of the foreign organism (Janeway et al., 2002). Another group, the toll-like receptors, does not bind the pathogen directly but imparts signals which activate specific reactions to a various set of antigens. Finally, the signalling cascade ends in the deletion of the pathogen. For example, toll-like receptor 4 recognizes specifically gram negative bacteria through association with the membrane protein CD14, a receptor for bacterial lipopolysaccharides. Toll-like receptor 2 and 4 activate NFκB, a factor that activates the expression of antimicrobial enzymes and molecules which can induce an adaptive immune response (Gay et al., 2007).

Macrophages can cause inflammation reactions in the body due to the production of antimicrobial enzymes like cytokines. The inflammation attracts neutrophilic cells and antimicrobial plasma proteins which support the immune response. Recognition of various pathogens induces different cytokines, which attract and activate different effector cells (Svanborg et al., 1999).

Another component of the innate immune system is the complement system, which interacts with the invading microorgansims thereby supporting whose elimination by phagocytosis. The complement consists of plasma proteins of the blood which are activated one after another in a cascade. The activated components build a complex on the surface of the pathogen and thereby mark it for the uptake and elimination by phagocytes or

enhance the opsonising of the pathogen by antibodies. If the proteins of the complement polymerize, they are able to introduce pores in the surface of the pathogen to destroy the organism (Tomlinson, 1993).
Pathogenic organisms are often able to mask themselves thereby circumventing the innate immune system. Therefore, innate responses can only be an initial response to an infection. At the second stage the adaptive immune response comes into effect, but this may take 4-7 days as it has to develop specifically against each single pathogen. During this time, the innate immune system has to limit the infection.
In the course of phylogenesis of the vertebrates the adaptive immune system developed from the innate immune system. The adaptive immune system is highly specific and consists of a pool of antibody molecules, each one binding a specific epitope on one specific foreign substance. This tremendous diversity is not encoded directly in the genome, and has its origin in gene rearrangements and processes of antibody gene diversification (Eason et al., 2004).
Beneath phagocytes and antigen presenting cells like dendritic cells, B- and T lymphocytes form the main part of the adaptive immune system.
Premature lymphocytes pass through the secondary peripheral lymphoid organs, where they come into contact with antigen presenting cells. The antibody is presented on the surface of the cell in form of a receptor. If the receptor has a high affinity towards the presented antigen, the antibody presenting cell remains in the peripheral lymphoid organ and starts proliferating. During this phase of maturation antibody specificity to the antigen is further increased by immunoglobulin (Ig) gene diversification. The progeny of this cell differentiate into effector cells which directly destroy the antigen or activate procedures leading in the destruction of the antigen. The highly specialized effector cell contributes to the adaptive immune response as long as there is contact to an antigen. Is the antigen gone, the pathogen is deemed to be killed. The effector cell is no more needed and undergoes apoptosis. A few cells are kept and further differentiated into memory cells to ensure an accelerated secondary immune response (Picker and Butcher, 1993).

1.2 B and T cells recognize foreign pathogen substances by their highly variable receptors

B cells mediate so-called humoral immunity via the production of antibodies. They combat pathogenic substances in blood and extracellular humor by marking invaders with secreted Ig antibodies. The antibodies neutralize the antigen or the opsonizing attracts phagocytes which destroy the foreign substances.
T cells are mainly responsible for the cell mediated immunity, which recognizes and contributes to the elimination of pathogens within a cell (e.g. viral proteins). They are classified into cytotoxic T cells and T Helper cells (T_H). Cytotoxic T cells bind antigenic epitopes in the form of pathogenic peptide fragments through their surface T cell receptors (TCR) and then directly kill the antigen presenting cell. T_H cells contribute to the

destruction of pathogens by recognizing peptides with their TCR. These peptides are bound by MHC-class II molecules and presented to the T cells by macrophages and B cells.

The antigen recognizing molecules presented on B and T cells are encoded by the Ig genes in B cells and the TCR genes in T cells. Each B and each T cell carries one of 10^9 different, specific and unique antibodies. Each antibody recognizes a specific antigen unique for the foreign substance (Janeway et al., 2002). This tremendous diversity is very important to our body as a restricted quantity results in low immunity (Revy et al., 2000; Imai et al., 2003). However, the need for diversity can not be met by the genome alone, as it would need too many genes and there exists only a restricted quantity of DNA. This places the idea, that a system of diversification is needed.

1.3 *Composition of a B cell antibody molecule*

Antibodies are heterodimeric proteins encoded by two separate genes for the heavy (H) and light (L) chains. Antibody molecules consist of two identical L chains and two identical H chains with a variable (V) region at the N-terminal portion and a constant (C) region at the C-terminal end. They associate by non-covalent interactions and disulfide bonding (Figure 1.1). Classification is done according to the type of C region of the H chain into the isotypes IgM, IgD, IgG, IgA and IgE (Tonegawa et al., 1983). The isotypes can be presented in a membrane bound form (receptor) on the surface of the cell or, by differential splicing they can be secreted as soluble antibodies.

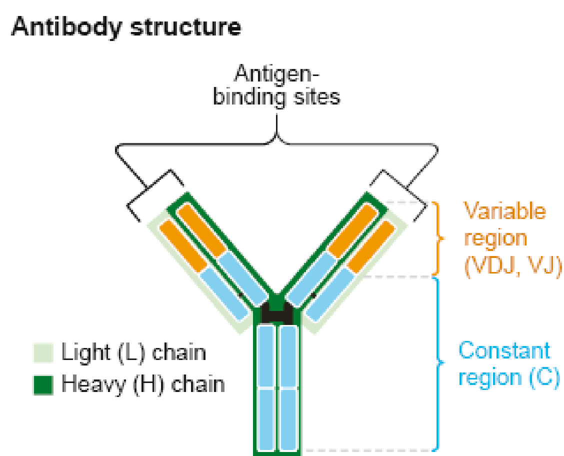


Figure 1.1 Structure of an antibody molecule.
Adapted from Peled et al. (2008). The two light chains are pastel green boxes and the two heavy chains are green boxes. The variable region of the light (VJ) and heavy (VDJ) chain are indicated in orange and their constant regions are marked in blue. The antigen binding sites are located at the N-terminal part of the variable regions of the light and heavy chains.

The enormous diversity of the B cell receptor (BCR) is generated in two different sections of the protein. In a first step, the diversification processes recombination and hypermutation vary the genetic code of the V region. Three hypervariable regions named complementary determining regions (CDR) which form loops of about 10 amino

acids each, build the antigen binding surface for specific recognition. These binding sites are identical in all antibodies presented or secreted by the same B cell clone irrespective of the isotype but vary from one clone to the next. In a second step the diversification process of subtype switching changes the C region. This is important for the different effector functions of the molecules like Fc-receptor recognition, activation of the complement cascade or active transport of the antibody molecule (Tonegawa et al., 1983).

1.4 Diversification is gained by several distinct processes

The immune system has evolved four different strategies to generate this enormous amount of antibody diversity. Starting with V(D)J recombination, in a second step the cell uses somatic hypermutation (HM) and gene conversion (GCV) for affinity maturation of the recombined V(D)J region of the L and H chains. Class switch recombination (CSR) is the final mechanism to change the type of the C region of the H chains (Figure 1.2).

1.4.1 V(D)J recombination is adding diversity to the gene loci by gene rearrangement

A study performing heteroduplex formation of mouse Igλ chain DNA revealed a diversification of Ig genes by somatic rearrangement (Brack et al., 1978). This gave a first direct evidence for V(D)J recombination giving rise to a primary immune repertoire.

The genomic locus encoding for B and T cell receptor molecules contains three different clusters of exons: the V (variable) segments, the J (joining) segments and the D (diversity) segments. The IgH, TCR β and δ chain gene carry all three kinds of segments. The IgL μ, κ and λ, TCR α and γ chain gene carry only V and J segments.

In human, for the L chain there are 40 different coding VLκ and 30 VLλ segments, 5 JLκ and 4JLλ segments scattered along the chromosome (Barbié et al., 1998; Pallarès et al., 1998). On the H chain, there are 39-45 VH segments, 6 JH and additional 27 DH segments (Pallarès et al., 1998; Ruiz et al., 1999). To form a functional rearranged antibody encoding gene, one of each segment is joined to another in a process called V(D)J recombination (Figure 1.2.A; Weigert et al., 1978; Kurosawa et al., 1981). Different combinations of the single segments lead to different specificities of the generated antibody. Due to the number of individual segments, there are many combinatorial possibilities, which in turn lead to a high number of potential antibody specificities (Tonegawa et al., 1983).

The process of V(D)J recombination does not require the presence of antigen. Recombination is an inherent component of B cell development. It starts before birth in the fetal liver and proceeds in adult bone marrow (Hardy et al., 2001).

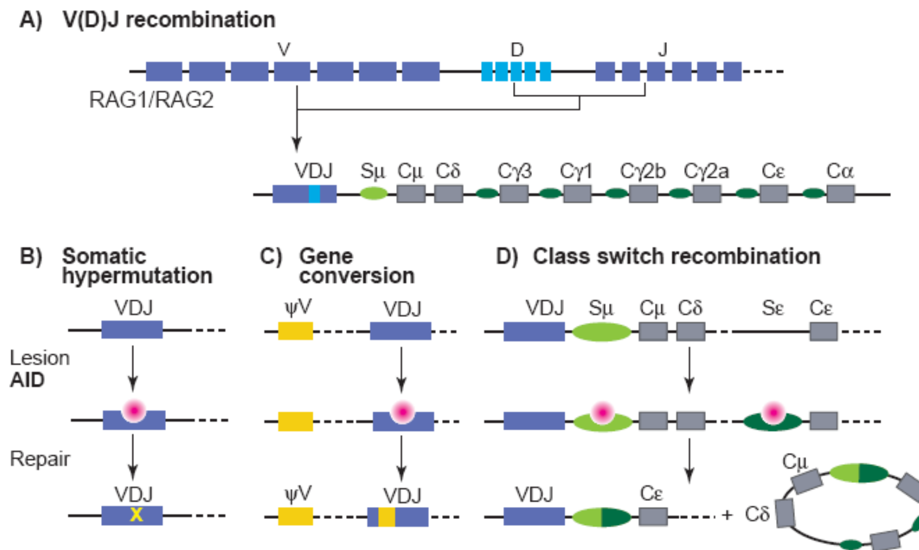


Figure 1.2 Diversification strategies of the immune system.
Figure adapted from Fugmann and Schatz (2002). A. RAG1/RAG2 mediated recombination of the variable (V), diversity (D) and joining (J) segments. One of each segment is rearranged with another. Segments in between are lost. B. After rearrangement HM introduces AID-mediated non templated mutations. The red spot is an AID induced lesion and the yellow X is the resulting mutation. C. AID-mediated Gene Conversion occurs by copying of a donor sequence into the V(D)J segment. Donor sequences are pseudo V genes (V) located upstream of the V region cluster. D. CSR joins together different switch regions thereby changing the type of the C region. The intervening sequence is excised.

V(D)J recombination is mediated by the activity of two recombination activating enzymes (RAG1 and RAG2) and involves proteins of the non homologous end joining (NHEJ) pathway (Lieber et al, 2004). RAG1/2 gene products are exclusively expressed in early stages of B and T cell development in the thymus and the bone marrow (Schatz et al., 1989: Oettinger et al., 1990) where V(D)J recombination takes place.

All germ line V, D and J segments are flanked by unique conserved recombination signal sequences (RSS) which are located 3' to the V segments, 5' and 3' to the D segments and 5' to the J segments. Recombination is accomplished by recognition of the RSS by RAG1 and RAG2. The recombinases join the V-D, V-J or D-J segments tail to head deleting intermediate DNA sequences. The RSS sequence consists of a highly conserved palindromic heptamer and an AT-rich nonamer sequence which are separated by a linker of either 12 bp or 23 bp length. Efficient V(D)J recombination occurs only between a 12-RSS and a 23-RSS, known as 12/23 rule (Eastman et al., 1996). RAG1 binds directly to the RSS and RAG2 stabilizes the formed complex together with HMG1 proteins. The sequences that are going to be recombined are thus brought into close proximity by the RAG complex. DNA is cleaved between the last nucleotide of the V, D or J gene coding sequence and the first nucleotide of the flanking RSS. This leads in two DNA double strand breaks (DSB) and creates 4 DNA ends. The two ends of the coding sequence which shall be recombined build up a hairpin structure, whereas the two signal ends which contain the RSS and the DNA that is excised from the genome are blunt ended (Fugmann et al., 2000). The NHEJ apparatus initiator Ku is supposed to bind the DNA ends thereby displacing the RAG complex. The hairpin structure is opened by the Artemis:DNA-PKcs complex (Ma et al., 2002) and the ends of the coding

sequences are joined together by the XRCC4/DNA ligase IV complex to generate a functionally rearranged Ig gene (Grawunder et al., 1997; Li et al., 1995). The DNA fragment containing the RSS sequences is ligated by the XRCC4/DNA ligase IV complex as well.

Joining of DNA ends by NHEJ is inherently imprecise as the endonucleolytic cleavage activity of the Artemis:DNA-PKcs complex can cause nucleotide loss (Ma et al., 2002). Consequently this may change the open reading frame and further contributes to antibody diversity. The joined ends of the V and J segments can vary by several nucleotides. The same imprecision was observed for D-J joining (Tonegawa et al., 1983). Another additional source for junctional diversity is created by the action of the terminal deoxynucleotidyl transferase (TdT), a DNA polymerase adding template independent nucleotides (Landau et al., 1987; Gilfillan et al., 1993).

Joining together different V, D and J segments in multiple ways, is the first diversification step in lymphocytes. TCR receptors are diversified solely by V(D)J recombination. Diversity of B cell receptors is further enhanced by the introduction of mutations into the V region of the rearranged immunoglobulin gene via hypermutation or gene conversion. To go through this antigen induced diversification processes, B cells migrate into the secondary lymphatic organs like spleen and lymph nodes.

1.4.2 Hypermutation increases specificity of the antibody after encounter of antigen

In 1966, Brenner and Milstein postulated a model of antibody diversification where nucleotide point mutations are introduced as a consequence of DNA synthesis errors, generated during the repair of DNA single strand nicks that are caused by enzymatic action. They suggested that these mutations manifest during DNA replication (Brenner and Milstein, 1966). Four years later there was evidence for Ig genes being somatically diversified by a process of spontaneous mutation, a phenomenon now called hypermutation (HM; Weigert et al., 1970).

HM is initiated only during the phase of antigen stimulation that takes place during affinity maturation of the B cell within specialized lymphoid structures namely the germinal centers (GC) of lymph nodes.

Activated B cells that have undergone V(D)J recombination migrate into the follicle centers of the secondary lymphatic organs in the environment of follicular dendritic cells after T cell dependent antigen stimulation. B cells start monoclonal expansion there, afterwards mutate the Ig genes for increased diversity and cells with higher affinity surface Ig (sIg) receptors are selected for differentiation. The GC is compartmentalized into dark zones that are built up from predominantly proliferating B cells and light zones that contain the B cells in the state of selection (MacLennan, 1994).

Lymphocyte centroblasts, that express sIg only at a low level, accumulate in the dark zone where they undergo proliferation and HM. Non-templated single point mutations and to a lower extent also deletion and insertions, are introduced into the recombined Ig V region gene (Figure 1.2.B; Berek and Ziegner, 1993; Kinoshita et al., 2001). Hallmarks of HM are the increased frequency of transition over transversion mutations (Golding et al.,

1987), hotspot focusing (Betz et al., 1993; Rogozin and Kolchanov, 1992) and the specific targeting of only the V region.

After re-expression of sIg, centroblasts give rise to centrocytes and migrate into the light zone where they are exposed to antigen. The presentation of the antigen by T cells and its recognition are necessary to select B cells which show the highest affinity in antibody-antigen reaction. B cells with low affinity receptors undergo negative selection and apoptosis (Janeway et al., 2002). Recent studies have shown that compartmentalization between dark and light zones is not absolute in terms of proliferation and migration of the B cells, as centroblasts and centrocytes share common features (Allen et al., 2007).

The complete process of proliferation, HM and selection is called affinity maturation and will be continued until antibody and antigen are coordinated perfectly (Jacob et al., 1993). Affinity maturation plays a major role in the development of memory B cells and contributes essentially to immunity by accelerating the immune response to the pathogen after a second infection (Berek, 1993).

1.4.3 Gene Conversion needs pseudogene templates to diversify the antibody genes

Domesticated animals like chicken and also rabbit or cattle use another mechanism for generating diversification called gene conversion (GCV; Knight and Barrington, 1998; Parng et al., 1996). HM plays only a minor role.

B cell development in chicken takes place in the follicles of the Bursa of Fabricius. The genome exhibits only one rearranged IgL and IgH gene due to allelic exclusion (Reynaud et al., 1989). Both loci carry only one V and one J segment, the IgH locus carries additional 16 D segments which are nearly identical. Rearrangement only takes place to express a functional antibody gene. Diversification is not gained during this step, because there is no choice between different gene segments and also a TdT does not exist (McCormack et al., 1989). This lack of diversification is compensated by GCV between the V region and a cluster of pseudo (ψ) V genes located upstream of the V region (Figure 1.2.C). The light chain locus carries 25 and the heavy chain locus 80 pseudogenes with high similarity to the VL and VHDH segments respectively. All pseudogenes are free from promoter, leader exon and V(D)J recombination signal sequences (Reynaud et al., 1987; Reynaud et al., 1989).

GCV is carried out according to a copy and paste like mechanism, where the pseudogenes serve as donor sequences and are copied into the V region of the Ig locus. The original pseudogene sequence is kept at its origin (Carlson et al., 1990, Reynaud et al. 1987). Sequences of 8 bp to 200 bp in length are copied from the pseudogene into the V region. Only pseudogenes located at the same chromosome as the V region serve as donors. Those with highest homology, closer proximity or located in an oppositional direction to the rearranged V segment are preferred (Mc Cormack and Thompson, 1990; Sayegh et al., 1999).

During the bursal stage of chicken B cell development GCV is initiated around embryonic day 15 (Mansikka et al., 1990). GCV and to a smaller content also HM create a pool of diversified preimmune Ig receptors (Thompson et al., 1987i; Ratcliffe, 2006; Parvari et al., 1990).

During the post bursal stages in the secondary lymphoid organs of the chicken, mature B cells undergo further gene diversification after antigen encounter. Here, predominantly HM is used to mutate the V region and GCV is down-regulated during this phase (Arakawa et al., 1996; Arakawa et al., 1998).

Mouse and human diversify their antibodies solely by V(D)J recombination and HM. There were some recombination events discussed as to be GCV (Krawinkel et al., 1983; Tsai et al., 2002; D'Avirro et al., 2005). But controversial studies disproved the results by showing that no essential donor sequences exist (Chien et al., 1988).

1.4.4 Class Switch Recombination changes the effector functions of the antibody molecule

The V and C region of the Ig receptors are encoded by two separate genes on the same chromosome and are joined together in a process called class witch recombination (CSR). CSR enables expression of antibodies with the same antigen specificity but different isotypes of the Ig receptor which means different effector functions (Chaudhuri and Alt, 2004) by changing the C region part of the antibody molecule (Figure 1.2.D).

The IgH gene carries several alternate C region exons. The µ region is in closest proximity to the V region. Antibodies with µ H chains can be expressed without a class switch. Therefore, the intron between the V(D)J exon and the first µ exon is removed from a long primary transcript by splicing. This leads to the isotype IgM. There exist several other C regions further downstream, in human for example eight (C_{δ}, $C_{\gamma3}$, $C_{\gamma1}$, $C_{\alpha1}$, $C_{\gamma2}$, $C_{\gamma4}$, $C_{\varepsilon1}$ and $C_{\alpha2}$). Each of these C regions is preceded by a switch region that enables recombination. Switch regions are highly repetitive GC-rich sequences of 1-10 kb length (Manis et al., 2002a). An interplay of sIgM, CD40 signaling and cytokines (like IL-4 or IFNβ) induces isotype switching by initiating DNA double strand breaks (DSB; Wuerfel et al., 1997). The DNA strands are joined together by the proteins of the NHEJ pathway as in V(D)J recombination (Kinoshita et al., 2001). Switch regions can also be affected by HM (Petersen et al., 2001; Dudley et al., 2002) indicating that both processes share a common mechanism.

1.5 Diversification is regulated by the B cell specific enzyme Activation-Induced Cytidine Deaminase (AID)

A few years ago, the three processes HM, GCV and CSR were thought to be independent of each other. Identification of an enzyme called activation-induced cytidine deaminase (AID) revealed that it is involved in HM and CSR in mouse and human as well as in GCV in chicken (Revy et al., 2000; Muramatsu et al., 2000; Muramatsu et al., 1999; Arakawa et al., 2002; Harris et al., 2002). In 1999 a study using cDNA screening in murine B lymphoma CH12F3-2 cells induced and uninduced for CSR first identified a novel mRNA coding for the protein named AID and showed that this protein is involved in CSR (Muramatsu et al., 1999). Northern blot

analysis of different mouse tissues confirmed that AID mRNA expression is restricted to lymphoid tissues (Muto et al., 2000). The expression is concentrated in germinal centers, where HM and CSR occur to diversify Ig gene information. In a later study the same group found the enzyme AID to be responsible for HM in mouse (Muramatsu et al., 2000). The results were confirmed by the fact that an immunological disorder "hyper IgM syndrome" results from deficiency in CSR and HM, which is caused by a mutation of the AID gene (Revy et al., 2000; Durandy and Honjo, 2001). Studies using the chicken B cell line DT40 showed an involvement of AID in GCV (Arakawa et al., 2002; Harris et al., 2002). Overexpression of AID in fibroblasts (Yoshikawa et al., 2002) leads to HM of an actively transcribed transgene and in hybridomas HM of the V region is restarted again (Martin et al., 2002). This indicates that AID is the only B cell specific factor that is needed to induce HM.

According to the current model, AID initiates all three diversification processes by the deamination of deoxycytidine to uracil. Uracil is further processed by the uracil DNA glycosylase (UNG) thereby creating an abasic site. The DNA lesions are further processed by repair pathways which are typical for HM, GCV or CSR (Neuberger et al., 2003) and involve enzymes of the DNA repair. This confirmed the model of a two phase process of antibody diversification, where in a first step DNA cleavage takes place followed by error prone repair. The model was suggested already 40 years ago by Brenner and Milstein (Brenner and Milstein, 1966).

1.5.1 AID has high homology to the RNA-editing enzyme APOBEC-1

AID is a small protein of 200 amino acids and 24 kDa. When it was first identified, it was suggested that AID is an RNA editing enzyme, because of its high homology with APOBEC-1, a mammalian RNA editing deaminase (Muramatsu et al., 2000). Several studies later revealed that AID actually acts as a DNA deaminase (Ramiro et al., 2003, Pham et al., 2003; Eto et al., 2003). In an E. coli study overexpression of AID led to mutation of cytidine residues suggesting an AID-mediated DNA deamination mechanism converting cytosine to uracil in DNA (Petersen-Mahrt et al., 2002). Ectopic and ubiquitous expression of AID led to transition mutations at cytosine and guanine in Ig and non-Ig genes in various eukaryotic cells (Martin et al., 2002; Okazaki et al., 2003; Poltoratsky et al., 2004).

Biochemical assays showed a stronger DNA deaminase activity towards ssDNA than to RNA or dsDNA (Bransteitter et al., 2003; Chaudhuri et al., 2003; Dickerson et al., 2003). In another E. coli study, AID-mediated HM targeted the nontemplate DNA strand, which is exposed as ssDNA during transcription (Ramiro et al., 2003). The lab of Reynaud derives similar results from a BL2 cell line stimulated for HM by simultaneous aggregation of three surface receptors (Faili et al., 2002). The assay is thought to serve as a model for the in vivo mechanism of AID activity, suggesting that ssDNA is the physiological substrate for AID.

1.5.2 Regulation of AID is necessary to avoid genomic instability

AID has the ability to mutate highly transcribed genes independent from sequence and position in the genome (Wang et al., 2004; Liu et al., 2008). To avoid genome wide DNA damage it has to be controlled tightly.

AID gene expression in mouse and human is induced by factors that mediate germinal center B-cell activation like lipopolysaccharide, Interleukin-4 (IL-4) or CD40 ligand (CD40L; Dedeoglu et al., 2004; Zhou et al., 2003). IL-4 and CD40L act synergistically, presumably through activation of specific signal transduction pathways and activator of transcription 6 (STAT6) and nuclear factor-κB (NFκB). IL-4 induces STAT6 binding to a site upstream of the promoter of the AID gene and CD40L induces binding of NFκB to two promoter sites located in the same region. Further studies confirm a regulation of AID expression by the E-protein E47 and its inhibitors the Id proteins, IRF-8, Bach2 and Pax5 transcription factors (Gonda et al., 2003; Sayegh et al., 2003; Yadav et al., 2006; Muto et al., 2004). The AID genomic locus in mouse contains a cis-regulatory element that is activated by E-proteins. This E-box element has been reported to contribute to positive and negative regulation of AID transcription through the binding of the basic helix-loop-helix (bHLH) transcription factor E47 and the inhibitor of DNA binding HLH protein Id3, respectively (Sayegh et al., 2003). Expression of E47 is, like AID expression, highly abundant in dark zone germinal center B cells, which undergo HM. Paired box gene 5 (Pax5) binds in the AID promoter region to increase transcription of AID in mouse and can be inhibited by Id2 (Gonda et al., 2003). Thus, AID is tightly regulated in an activation dependent manner through transcription factors whose expression is specific for B cells.

AID remains restricted to the cytoplasm of lymphocytes where it has no possibility to influence target genes until activation (Rada et al., 2002; Ito et al, 2004, Mc Bride et al., 2004). Probably AID is kept in the cytoplasm by chaperons and enters the nucleus after B cell receptor stimulation (Reynaud et al., 2003). A nuclear localization signal (NLS) is located at the N-terminus of the protein and a C-terminal nuclear export signal (NES) shuttles it from the nucleus to the cytoplasm (Figure 1.4). However, nuclear localization of AID is not sufficient for HM activity (Shinkura et al., 2004). A phosphorylated form of AID associates with chromatin, suggesting that a balanced localization of AID contributes to diversification activity (Brar et al., 2004; McBride et al., 2006).

Most likely the various AID-mediated diversification processes need different cofactors. This is supported by the finding that the C-terminal part of AID is important for CSR, whereas the N-terminal part is important to HM in E.coli, DT40 cells and mouse (Figure 1.3; Ta et al., 2003; Barreto et al., 2003; Bransteitter et al., 2004; Shinkura et al., 2004). Replication protein A (RPA) builds a complex with phosphorylated AID in an in-vitro assay (Chaudhuri et al., 2004), and Protein kinase A (PKA) is responsible for the phosphorylation of AID (Basu et al., 2005). In which way phosphorylation is important for diversification, remains unclear.

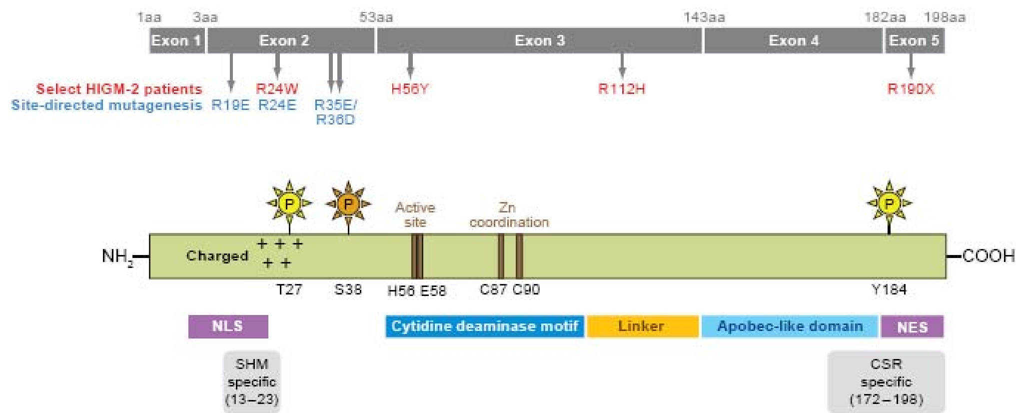


Figure 1.3 Model of AID exon and relevant functional domains.
Adapted from Peled et al. (2008). Inactivating mutations in the exons that result in amino acid changes are indicated in red and blue. Known phosphorylation sites on the AID protein are depicted with yellow and orange sunbursts.

1.6 Uracil DNA glycosylase (UNG) is involved in the processing of the AID-induced DNA lesions

The uracil introduced into the DNA by AID-mediated deamination is a strange base and needs to be excised. This is done by uracil DNA gylcosylase (UNG) and the created abasic site can be repaired in an error-free or error prone way giving rise to transition as well as transversion mutations (Figure 1.4). In HM, UNG deficiency leads to a transition bias at C and G bases in human, mouse and chicken (Di Noia and Neuberger, 2002; Di Noia and Neuberger, 2004; Imai et al., 2003; Petersen-Mahrt et al., 2002; Rada et al., 2002; Saribasak et al, 2006). These findings led to a model whereby uracils remain in the DNA and during replication lead to the incorporation of an adenine instead of a guanine on the complementary DNA strand. As a result mutations from C-to-T and G-to-A can be observed (Figure 1.4).

Moreover, the frequency of CSR is significantly reduced in $UNG^{-/-}$ mice (Rada et al., 2002). In chicken GCV is strongly reduced, suggesting UNG is the enzyme in charge to further process AID-induced lesions (Saribasak et al, 2006). HM is strongly enhanced in $UNG^{-/-}$ DT40 cells. Hence, AID-induced and UNG-processed lesions seem to be repaired mainly in an error free fashion and only a small part of the lesions leads to HM.

1.7 The three diversification processes involve different factors of DNA repair

The uracil introduced by AID gives two problems to the cell. First, uracil is a strange base in DNA and has to be removed. And second, the U:G pair is a mismatch which has to be repaired. To establish a diversification event, it needs an interplay with other factors that differ between HM, GCV and CSR. But all of them seem to be part of

DNA repair processes. Interestingly, in this context they are used preferably in an error prone fashion as described in more detail in the following chapters.

1.7.1 Error-prone repair in HM ends up with a diversified antibody gene after replication

Due to genetic studies the following model for HM was postulated. AID deaminates deoxycytidine to uracil in single stranded transcribed DNA. For the processing of the uracil there are three possibilities (Figure 1.4):

i. Replication over uracil

If the uracil is not excised from the DNA, replication results in a transition mutation at the deaminated position as oppositional to the uracil an adenine will be integrated into the DNA.

ii. Deletion of uracil

UNG, an enzyme from base excision repair (BER), excises the uracil from the DNA and leads to an abasic site. During replication, the lesion stops the movement of the replication fork. The cell uses the sister chromatid for an error free repair or it involves translesion DNA polymerases to incorporate any of the four bases. This extends the mutation spectrum to transversion mutations.

Studies reveal an apparent initiation of HM by a DSB (Bross et al., 2000, Papavasiliou and Schatz, 2000) or in combination with a SSB (Sale et al., 1998), but these observations remain controversial (Bross et al 2002; Chua et al., 2002). Mutations are initiated actively by error-prone processing of the DNA breaks or passively by the absence of repair machineries (Papavasiliou et al. 2000). DNA nicking is needed and error prone repair of the lesion is further supported by the involvement of translesion DNA polymerases. It is not completely clear which enzyme is responsible for DNA-nicking, although APE1 and MRE11-RAD50-NBS1 complex were discussed (Rada et al., 2002; Bardwell et al., 2004; Larson et al., 2005). Pol ι seems to be important for HM in the mouse B cell line BL2 (Poltoratsky et al., 2001), but has no effect in mouse (Shimizu et al., 2005). Other studies showed an involment of Pol η in HM. Pol $\eta^{-/-}$ mice exhibited a strong decrease in mutations at A and T bases (Faili et al., 2004) and in DT40 GCV and HM are reduced (Kawamoto et al., 2005). Pol $\theta^{-/-}$ mice exhibit a transition bias at C and G (Zan et al., 2005). REV1 knockout affects mainly C-to-G transversions on the non transcribed strand (Jansen et al., 2006). In DT40 both strands are affected (Ross and Sale, 2006). Ubiquitinated PCNA is involved in the recruitment of translesion polymerases. Studies in DT40 revealed, that the PCNAK164R mutant which can no more be ubiquitinated, has also a defect in HM (Arakawa et al., 2006). It appears to be likely, that multiple polymerases are involved in generating the full spectrum of HM events.

iii. Uracil is recognized by MMR proteins

MMR generates mutations outside of the initial lesion at A:T base pairs (Figure 1.4), but the mechanism is not understood in full detail up to now (Martin and Scharff, 2002i; Neuberger et al., 2005). MSH2$^{-/-}$ mice have reduced HM and mutations at adenine and thymine compared to the heterozygous knockout (Rada et al., 1998). HM in UNG$^{-/-}$MSH2$^{-/-}$ is absent at adenine and thymine residues but still acts on C and G bases (Rada et al.,

2004). The lack of mutations at adenine and thymine led to a model in which UNG and MMR proteins are responsible for an expansion of the mutation spectrum to other bases (Figure 1.4).

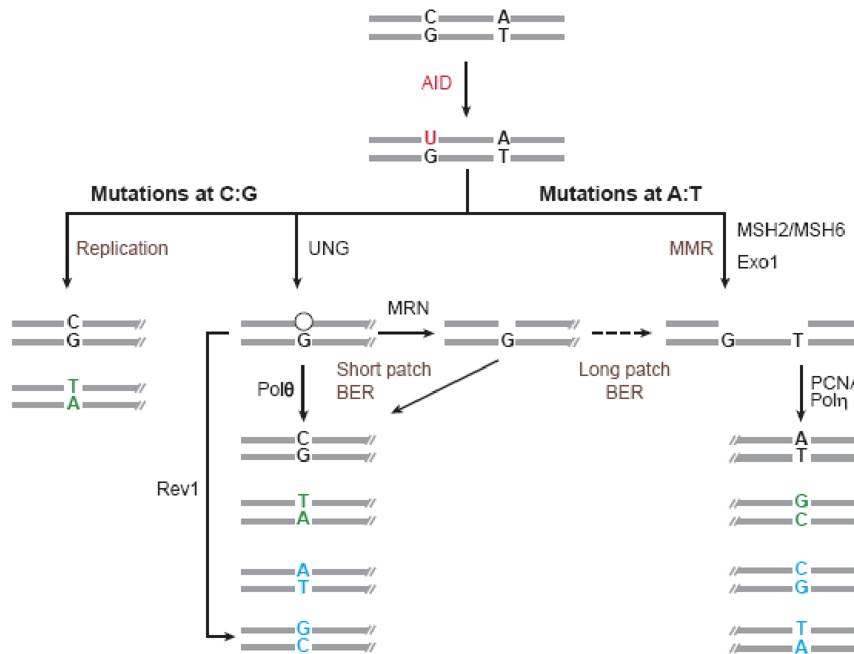


Figure 1.4 A model for HM in Ig genes.
Adapted from Teng and Papavasiliou (2007). AID initiates a lesion into the DNA by deaminating cytosine to uracil. The repair can lead to a mutation. The introduced uracil is coloured in red, transition mutations are indicated in green and transversion mutations in blue. The used pathways are written in brown. Enzymes involved in the repair are noted.

1.7.2 Repair pathways that lead to GCV and CSR

Also for GCV in a first step AID induces uracil into ssDNA. The uracil will be removed by UNG creating an abasic site (Saribasak et al., 2006). GCV now needs introduction of a single strand break (SSB) into the rearranged V region followed by homology-based repair involving the pseudogenes for a repair of the SSB. The detailed steps leading to GCV are unknown. Presumably the DNA is nicked at the abasic site by an endonuclease and proteins of the Rad52 pathway of homologous repair were shown to be involved (Sale et al., 2001; Bezzubova et al., 1997). The translesion DNA polymerase Pol η takes part in DNA synthesis in GCV (Kawamoto et al., 2005).

In CSR, recombination takes place between switch (S) regions which antedate the accompanying C region. Each S region has its own promoter. A prerequisite is the active transcription of the two S regions between those CSR takes place (Daniels and Lieber, 1995). AID will introduce uracil which is further processed to an abasic site by UNG. Transcription-induced R-loops form pieces of ssDNA which could be target for AID-mediated deamination (Yu et al., 2003). However, the repeats are not essential, as deletion did not stop CSR (Shinkura et al., 2003; Luby et al., 2001).

To switch now from one C region to the other, in both S regions accompanying the two C regions to be switched, a DSB has to be induced at abasic sites. The nicking involves members of the BER and MMR (Stavnezer and Schrader, 2006). UNG deficiency leads to a strongly reduced CSR level in mice (Rada et al., 2002) and humans (Imai et al, 2003). In MSH2$^{-/-}$ mice switch frequencies are significantly reduced (Ehrenstein and Neuberger, 1999). CSR in UNG$^{-/-}$MSH2$^{-/-}$ is around background level (Rada et al., 2004). UNG and MSH2 seem to be part of two complementary pathways, which process AID-initiated DNA lesions by producing CSR products.
Also the nucleases EXO1 seems to be involved in CSR in mice (Bardwell et al., 2004), most likely in blunting in the staggered ends of the DSBs, as only blunt-ended DSBs can be religated. The DSB is repaired by generating class-switched products (Rada et al., 2004; Catalan et al., 2003; Rush et al., 2004) with the help of NHEJ. Deficiencies in NHEJ components inhibit CSR essentially, as shown by knockouts of DNA-PKcs, Ku70 and Ku80 (Manis et al., 2002; Casellas et al., 1998; Manis et al., 1998).

1.8 Specificities of AID-dependent diversification processes

1.8.1 Mutation frequency in hypermutating cells is strongly enhanced

Based on genealogical trees of mutations found in mouse germinal center B cells it is probable that they are clonally related to one another. The rate of HM was estimated to be 10^{-3} – 10^{-4} mutations per bp and cell division (McKean et al., 1984; Sablitzky et al., 1985; Kleinstein et al., 2003), which is 10^{5}-10^{6} fold higher than the spontaneous mutation rate for somatic cells (Neel et al., 1986; Martin et al., 2002). In contrast, non-Ig genes in hypermutating B cells mutate with a low frequency. Tests of housekeeping genes in a mouse myeloma B cell line and a human Burkitt´s lymphoma B cell line showed a 10^{3}-10^{4} fold lower mutation rate compared to Ig genes (Baumal et al., 1973; Sale and Neuberger, 1998). Nevertheless, there are some non-Ig genes like Bcl-6, B29, mb1 and the CD95/Fas gene with an elevated mutation level in hypermutating B cells or B cell lymphoma compared to other non-Ig genes or their counterparts in non hypermutating B cells (Pasqualucci et al., 1998; Shen et al., 1998; Gordon et al., 2003; Landowski et al., 1997; Müschen et al., 2000). The mutation frequency in such genes is more than 50 times lower than in Ig genes supporting the idea of HM being mainly restricted to the Ig loci. However, it means that targeting of HM on non-Ig genes is possible.
AID-mediated diversification processes have the potential to introduce mutations and if HM targets proto-oncogenes this contributes to tumorigenesis (Pasqualucci et al., 2001). AID-initiated IgH locus breaks can lead to chromosomal translocations resulting in B-cell malignancies (Ramiro et al., 2006; Franco et al, 2006). This makes it essential to understand the molecular basis of the targeting specificity of AID.

1.8.2 The role of primary Ig sequences – AID can also target other sequences

A plausible mechanism to restrict AID activity to the Ig loci would be an intrinsic feature of the genomic sequence itself to catch AID and fix it to its sequence.

For HM the primary Ig sequence itself seems to be dispensable for the action of AID, as non-Ig transgenes like a GFP-provirus are diversified, too, and diversification occurs independently from Ig enhancer cis-elements (Parsa et al., 2007). During deamination there is a preference for hot spot motifs RGYW/WRCY and WA. About 30% of all mutations in mouse Ig genes, 50% in mouse Ig transgenes and 50% of mutations in chicken DT40 cells occur at a hotspot (Rogozin and Kolchanov, 1992; Michael et al., 2002; Arakawa et al., 2004). Also in vitro studies identified a similar motif (WRC) being a preferential target sequence (Pham et al., 2003). As the motif is rather degenerate, it alone can not explain the specificity of AID action for the Ig locus.

It is unknown if GCV needs specific motifs, but a heterologous GCV substrate being composed of a GFP transgene as donor and BFP transgene as acceptor sequence is able to undergo AID-dependent GCV in the chicken IgL locus (Kanayama et al., 2006).

1.8.3 Strand bias

In mouse and human HM a strand bias was observed in the positioning of the mutations. Thus the different bases mutate to a different frequency: cytidines and guanidines mutate to the same frequency, but adenines mutate twice as much as thymidines (Spencer et al., 1999). This led to the suggestion that the same base is targeted differently on the both DNA strands (Sohail et al., 2003). According to in vitro assays AID deamination occurs only on the non-template strand, but in vivo both strands seem to be a target for deamination as the same amount of guanine and cytosine is going to be mutated. However, it remains unclear how AID gains access to the template strands. A mechanism of negative supercoiling was suggested (Shen and Storb, 2004).

1.8.4 Strong transcription is a prerequisite for AID-mediated diversification processes

HM starts 100-200 bp downstream of the transcription initiation site of the Ig promoter and peaks around 400-500 bp, what is correlating with the position of the V(D)J region. The mutation rate declines with increasing distance from the promoter until it is reduced to background level around 1.5-2 kb downstream of the promoter (Winter et al., 1997; Lebecque and Gearhart, 1990). The exact circumstances causing this mutation profile are unknown, but studies in mice show that AID does not gain access to the 5' promoter region and the 3' C region of the Ig genes (Longerich et al., 2005). This assures the main part of the mutations being concentrated around the V region of the Ig genes sparing the C region.

The position of the HM profile within the target gene sequences is determined by the position of the promoter. If 2 kb of bacteriophage λ DNA are inserted between the Ig promoter and the leader exon in an Igκ transgene, HM

now targets the bacteriophage DNA (Winter et al., 1997). Insertion of an Ig promoter in front of the C region in a transgene erased a new HM window directly after the inserted promoter sequence targeting now the C region (Peters and Storb, 1996). Thus the sequence targeted by HM seems to have no intrinsic features essential for HM to occur. The precondition for HM action is an active promoter. This was shown in mouse, where the endogenous IgH promoter had been deleted (Fukita et al., 1998) and in DT40 cells where the endogenous IgL promoter was substituted by a bacteriophage T7 promoter, which is not able to initiate transcription in eukaryotic cells (Yang et al., 2006). In both experiments HM failed to target the Ig genes. In contrast, the sequence of the active promoter transcribing the hypermutating locus appears to be irrelevant. For example the B29 promoter, the human EF1α promoter or the chicken ß-globin and the CMV promoter can all substitute for endogenous Ig promoters (Betz et al., 1994; Tumas-Brundage and Manser, 1997; Papavasiliou and Schatz, 2000; Yang et al., 2006).

An active promoter is necessary for diversification to occur, presumably in its function to start transcription. As HM is mainly restricted to the promoter proximal end, initiation of transcription seems to play an important role, especially the changeover from initiation to elongation, where the RNA Pol II is pausing (Raschke et al., 1999). In transcribed DNA, ssDNA is present in the transcription bubble and could be a target for AID deamination. Most likely to conduct deamination the diversification machinery, including AID, is loaded onto the elongating transcription complex (Odegard and Schatz, 2006). This idea is supported by the verification of an interaction between RNA Pol II and AID. It was possible to coimmunoprecipitate AID and RNA Pol II (Nambu et al., 2003), and UV-crosslinking during an in vitro AID deamination assay with E. coli proteins captured AID bound to Pol II (Besmer et al., 2006).

Biochemical assays with T7 RNA Pol and RPA or using E.coli RNA Pol revealed AID-mediated deamination of a plasmid (Chaudhuri et al., 2003: Besmer et al., 2006). Those results as well as other studies involving non-Ig transgenes (Ruckerl et al., 2004; Parsa et al., 2007) led to the idea, that a high transcription level is a key feature for HM to occur. Several studies confirmed transcription of the target locus as being essential for HM to occur (Peters and Storb, 1996; Fukita et al., 1998). Studies in E. coli and the B cell line 18-81 verified that the rate of transcription correlates with the rate of HM (Bachl et al., 2001; Ramiro et al., 2003).

The assumption that highly transcribed genes are mutated as well, including non-Ig transcription cassettes randomly integrated at non-Ig loci, built the base line for GFP reporter assays to study HM. Often eGFP expression cassettes including a premature stop codon were used in combination with non-Ig promoters like CMV, thymidine kinase promoter and retroviral 5' LTR. HM mutates the stop codon and a normal protein is expressed which can be detected by FACS analysis (Yoshikawa et al., 2002; Bachl and Olsson, 1999; McBride et

al., 2004). However, the mutation frequency of the GFP reversion constructs varies. The major part exhibits a very low GFP positive population of 0.02-0.1% after 5-30 days. Sequencing to confirm the results often was not done. One study revealed a mutation level of the GFP reversion construct comparable to Ig genes and provided also sequencing data. Similar results were obtained for a non-Ig construct consisting of a strong RNA Pol II promoter and downstream AID expression cassette (Yoshikawa et al., 2002; Martin and Scharff, 2002i). Constitutive expression of AID in mice led to increased T cell lymphoma in combination with a clustering of mutations, whereas conditional overexpression of AID in B cells causes only minor perturbations of HM and CSR (Okazaki et al., 2003; Muto et al., 2006). Taken together the data provide evidence for AID being a potent mutator of highly transcribed non-Ig genes, but there exist additional mechanisms, presumably cis-acting elements, which tightly regulate targeting of AID to the specific substrates in B cells.

Similar to HM, CSR depends on germline transcription, too, and also in this context a heterologous promoter is able to initiate CSR (Bottaro et al., 1998; Okazaki et al., 2002). The molecular mechanism should be the same concerning the needs of transcription. If GCV needs an active promoter was not tested, but GCV is still active under the control of a CMV promoter (Kanayama et al., 2006).

1.8.5 Chromatin modifications seem to play a minor role in AID targeting

A study in human BL2 cells was able to show hyperacetylation of the histones H3 and H4 at the IgH V region and not C region upon stimulation for HM, which could be a sign for a correlation between acetylated histones and HM (Woo et al., 2003). Experiments using primary B cells of transgenic mice verified the results for the IgH and the IgL λ locus (Odegard et al., 2005). There is no difference between acetylation of naïve and activated B cells, which argues against a theory of modified chromatin being a marker for hypermutating regions. Methylation is also unaffected by the induction of HM, which is in contrast to a transgenic mouse study reporting that HM acts on demethylated transcribed Ig transgenes but not on methylated (Jolly and Neuberger, 2001). However, phosphorylated H2B could play a role in HM after recruitment of AID (Odegard et al., 2005).

1.9 Trans-acting factors and cis-acting elements in Ig gene diversification

The unique specificity of the AID mediated changes to Ig genes raises two actual issues:

i) how is diversification restricted to germinal-center B cells and

ii) how is it limited to the Ig loci in HM active cells.

1.9.1 AID is expressed only in B cells

The restriction to B cells can be explained by the fact that AID is expressed only in germinal-center B cells and in a few extrafollicular B cells, which undergo class switch and are part of the T cell independent immune response (Cattoretti et al., 2006).
AID can induce SHM in hybridoma cells which represent a fusion between a B cell at the stage of a plasma cell, which usually stopped all HM and an immature preB splenocyte (Martin et al., 2002). It indicated that AID works either alone or together with cofactors being expressed throughout B cell development. The lab of Honjo (Okazaki et al., 2002) induced CSR in an artificial switch construct in the murine fibroblast cell line NIH3T3 solely by overexpression of AID. It suggests that expression of all other components which may be required for CSR are not restricted to activated B lymphocytes. AID did not induce CSR in the endogenous Ig loci of NIH3T3, as these loci are not actively transcribed and therefore not accessible for AID. The same observations were made for HM in NIH3T3 cells, where ectopic expression of AID also induced HM in an actively transcribed, artificial GFP substrate (Yoshikawa et al., 2002). Distribution and pattern of mutations was similar to those in the endogenous Ig loci in B lymphocytes. It is a proof for AID being the only B-cell specific factor involved in HM and CSR and suggests regulation of diversification by restricted enzymes and not by B cell permissive factors.

1.9.2 AID recruitment needs additional factors

How AID is recruited to the DNA and specifically to the Ig locus, is not understood and can not be explained by the above mentioned observations. They showed that AID causes CSR and HM in non lymphoid cells, but AID might act unspecifically on all actively transcribed genes, because of a lack of specific cofactors targeting AID to a specific locus. The Honjo group did not examine endogenous actively transcribed genes, therefore this question can not be addressed. Martin (et al., 2002) reported about an AID transgene being mutated by hypermutation like events in B cells as well as non B-cells. They expressed human AID and an Ig transgene in Chinese hamster ovary cells and found both to be hypermutated. The same AID was expressed in Burkitt´s lymphoma Ramos and hybridoma P1-5 cells. In both cell lines the expression of the AID transgene caused HM of the endogenous Ig sequences and the AID transgene, too.
Ramos cells express surface markers that suggest that their normal cellular counterpart is a germinal-center centroblast and they undergo HM constitutively (Sale et al., 1998). Hence, this cell line might demonstrate that AID targeting by specific cofactors is not needed for AID induced HM. However, the selected Ramos clones are not mutating naturally, as the endogenous AID level is too low (Martin et al., 2002). In the hybridoma P1-5 B cell endogenous AID expression is downregulated, because naturally they stopped Ig HM after completing maturation. In the same way, specific cofactors targeting AID to the Ig locus could be downregulated, too. In this case, overexpression of AID can cause unspecific HM like events, as the targeting pathway is discontinued.

The question still remains how AID deamination activity is specifically targeted to the Ig loci and spares other regions of the genome. As AID activity has intrinsic mutagenic potential, it needs other factors, which restrict AID activity to a specific area to avoid DNA lesions throughout the genome.

1.9.3 The role of Ig enhancers and matrix attachment regions (MARs) is not completely clarified

Results on the role of enhancer sequences for regulating AID-mediated diversification are controversial. Analysis from the B cells of chimeric mice with the IgH intronic enhancer (iEμ) being deleted show the iEμ to be not sufficient for HM and CSR (Perlot et al., 2005). Deleting the two most 5' enhancers of the IgH enhancer cluster located downstream after the most 3' C region had no effect on CSR (Manis et al., 1998i). However, in both studies still remain other enhancers in the IgH genomic locus and if the enhancers share similar functions and carry overlapping redundant motifs, the experimental approach is not suitable to reveal a relevance of the enhancer sequences for HM or CSR. In contrast, the IgH 3'enhancer together with DNase I-sensitive regions (HS) 3b and 4 supports HM of an IgH transgene in mice (Terauchi et al., 2001).
Experiments with transgenic mice carrying randomly inserted Igλ transgenes demonstrate that the iEμ but not the κ3' enhancer can substitute the $\lambda_{2\text{-}4}$ enhancer in supporting HM of the transgene (Klotz and Storb, 1996; Kong et al., 1998). The correlation between transcription and HM was not examined in these two studies and also controls of an Igλ transgene without any enhancer sequence are missing. Therefore, it is not possible to say whether the effect on HM is due to diversification activating elements or due to an enhanced transcription of the Igλ transgene.
In mouse, 3'κ enhancer and the matrix attachment region/intronic enhancer (MAR/iEκ) of the IgL locus supported HM (Betz et al., 1994; Stoep et al., 1998; Klix et al., 1998). A deletion of the 3'κ enhancer inhibited expression of the transgene. In this context it is important to point out a dual role of the enhancer. On the one hand it could support high levels of transcription necessary for the efficient HM. On the other hand, the enhancer could be able to recruit other factors involved in targeting and triggering HM. Hence, it is not possible to say whether the 3'κ enhancer is involved in targeting HM (Goyenechea et al., 1997).
Although Igκ transgenes are used successfully for the study of HM, the frequency of HM varies tremendously from study to study and in most cases the frequency is strongly decreased compared to the endogenous Ig genes (Peters and Storb, 1996; Michael et al., 2003; Bachl et al., 2001; Papavasiliou and Schatz, 2000; Ruckerl et al., 2006; Parsa et al., 2007). In other studies targeted deletion of the MAR/iEκ does not reduce IgLκ locus transcription or HM, but deletion of the IgL 3'κ enhancer leads to a moderate reduction of expression and HM frequency (Stoep et al., 1998; Inlay et al., 2006). HM of the IgH gene in hybridoma cells seems to be negatively regulated by the IgH Eμ enhancer or the MARs if they are individually absent (Ronai et al., 2005).

The inconsistency in results may be due to the connection of transcription and HM targeting in these model systems. Furthermore, the transgenes in most cases are integrated randomly and the insertion at different non-Ig loci is prone to cause positional effects.

The question if Ig enhancers contain cis-elements to enable HM is still not solved. According to the results up to now, they are neither necessary nor sufficient. But still there is the possibility of cis-elements being distributed over the Ig locus and its enhancer. To clarify this issue needs a study which is eliminating redundant elements in the regulatory region.

1.9.4 E-box motifs seem to be important for HM

E-box motifs are transcription regulating DNA motifs identified in promoter and enhancer elements that regulate immunoglobulin-, muscle-, neuron-, and pancreas-specific gene expression. The common sequence motif is CANNTG, where the first two and the last two nucleotides are highly conserved and the two nucleotides in the middle can be any bases (Massari and Murre, 2000).

A study in transgenic mice using a reporter construct with two of the E-box motifs CAGGTG in an Igκ transgene revealed an enhancement of HM without influencing the transcription level (Michael et al., 2003). Additionally, E-box sequence motifs are present in the enhancers of Ig and actively transcribed hypermutating non-Ig genes (Kotani et al., 2005). This suggests E-box motifs being involved in Ig gene diversification.

1.9.5 E2A transcription factors bind to E-box motifs

The E-box motif CANNTG is the binding site for the basic helix-loop-helix proteins E12 und E47 which arise by alternative splicing of the E2A gene (Massari and Murre, 2000).

The motif was first identified within the enhancers and promoters of Ig genes (Ephrussi et al., 1985). The two proteins E12 and E47 were shown to bind to these E box sites by a flexible loop structure named basic helix-loop-helix (bHLH) motif (Henthorn et al., 1990; Murre et al., 1989).

By alternative splicing of a single exon the E2A gene encodes the two proteins E12 and E47, which differ only in their highly homologous DNA binding and dimerization domains. Both proteins are members of the class I bHLH proteins, also named E-proteins. The HLH motif is a conserved carboxy-terminal protein dimerization domain consisting of two amphipathic α-helices separated by a loop structure (Massari and Murre, 2000). DNA binding efficiency of E12 is reduced by an inhibitory domain amino-terminal to the HLH motif (Sun and Baltimore, 1991). E12 builds heterodimers with E47 or other bHLH proteins, whereas E47 is able to build homodimers, too. The E47 homodimers appear to be B cell specific which might support their role in Ig repertoire development (Goebel et al., 2001). The two conserved activation domains ADI and ADII are located at the N-terminal half of both proteins. Both are essential for activation of transcription of B cell lineage specific

genes (Kee and Murre, 1998). ADII contains a leucine zipper region which has been shown to play a role in stabilizing HLH-mediated dimerization and it is supposed to play an important role in directing transcription activation, perhaps involving interaction with other zipper proteins (Aronheim et al., 1993).
E2A transcription is greatest in areas of high proliferation like the bone marrow, in thymus and splenic germinal centers, where Ig gene rearrangement and diversification takes place (Xin et al., 1993).
The E2A transcription factors are essential for early B cell development, because E2A$^{-/-}$ mice lack pre-B and mature B lymphocytes. B cell development is blocked at a stage prior to Ig gene rearrangement (Bain et al., 1994; Zhuang et al., 1994), most likely because expression of genes involved in Ig gene recombination is impaired. Studies with E2A$^{+/-}$ mice show that the E2A proteins regulate gene expression of the recombination activating proteins RAG-1 and RAG-2 (Quong et al., 2004). Overexpression of E47 in a pre-T cell line leads to transcription of the Ig genes (Schlissel et al., 1991). Overexpression of E12 in a macrophage cell line leads to expression of the B cell specific genes RAG-1, EBF, Pax-5 and the surrogate light chain λ5, which is needed to express the pre B cell receptor (Kee and Murre, 1998). Beyond their involvment in B cell specific gene activation, E12 and E47 were shown to induce Ig gene rearrangements, too. Overexpression of the proteins in a pre-T cell line and various non-lymphoid cell lines promotes V(D)J recombination (Schlissel et al., 1991; Romanow et al., 2000; Choi et al., 1996) and they are also required for rearrangement of the T cell receptor γ and δ loci (Bain et al., 1999).
In later stages of Ig gene rearrangement, during CSR, E2A proteins become important regulators again. They are required for proper isotype switching in cell lines and primary B-lineage cells (Goldfarb et al., 1996). Repression of E2A by Id proteins, dominant negative regulators of bHLH proteins, inhibits CSR (Quong et al., 1999). Recent studies on senescent mice with reduced CSR show, that this reduction is due to decreased E47 and AID, which is regulated by E47 (Frasca et al., 2008; Sayegh et al., 2003).
E2A is found in chicken, too. The E2A gene is encoded on chromosome 28. It consists of 19 exons spanning a region of 38.7 kb. Like in mammals, chicken E47 and E12 are identical except for their bHLH domain. Both splicing variants possess the same activities like the mammalian homologues. They stimulate transcription, but E12 possesses only 80% of the activity of E47. Both E2A encoded proteins can be inhibited by Id proteins (Conlon and Meyer, 2004).

1.10 DT40 cells as model system

The DT40 cell line is derived from an avian leucosis virus (ALV) infected bursal B lymphoma cell of chickens of the SC strain with deregulation of the myc gene (Baba et al., 1985, Thompson et al., 1987). About 85% of all ALV induced tumors have integrated the 3' end of the viral long term repeats (LTR) upstream of the c-myc gene (Hayward et al., 1981; Buerstedde et al., 1990). C-myc induces synthesis of proteins important for tumor

progression. DT40 can be described as being arrested at the developmental stage of a bursal stem cell (McCormack et al., 1991).
The DT40 undergo permanent GCV and at later stages of development HM of the IgH and IgL loci, with GCV being more prominent than HM. Diversification in this cell line is comparable to a bursal B cell line, although the rate of GCV and the conversion track length is decreased (Buerstedde et al., 1990; Kim et al., 1990).
Due to its high rate of homologous recombination, targeted integration is strongly enhanced and manipulation of endogenous DNA elements is simplified (Buerstedde and Takeda, 1991). Moreover DT40 expresses a relatively stable karyotype and phenotype over long periods of cultivation (Sale, 2004) making the cell line an attractive model for multiple gene targeting. The growth rate is fast with a doubling time of about 10h and observation can be obtained rapidly after manipulation. Beyond this, cloning efficiency is very high and the stable transfectants can be selected and subcloned very easily (Yamazoe et al., 2004). This is not only a great advantage for immunologists, but enables also the examination of cell cycle, apoptosis, histone acetylation and DNA repair (Winding et al., 2001).

1.11 Effect of E2A knockout on HM

During my diploma thesis "The Role of Bach-2 and E2A for Immunoglobulin Gene Transcription and Repertoire Development" (Schoetz, 2005) I knocked out the E2A gene in the DT40 variant AID$^{R}\psi$V^{-}. This clone was chosen as progenitor clone to study the effect of gene disruption on HM. The excision of the pseudogenes of the rearranged IgL chain gene abolishes gene conversion completely and forces the cell line to undergo antibody diversification at the rearranged IgL locus gene solely by AID-mediated HM (Arakawa et al., 2004). The cell line expresses IgM on its surface (sIgM). HM will introduce random mutations into the IgL chain locus during replication. Some of these mutations, for example resulting missense mutations or introduction of an early stop codon, can be expected to disturb the expression of sIgM. The predominantly sIgM(+) cell line then shifts to sIgM(-) (Figure 1.5.).
If E2A transcription factors are necessary for HM to take place, a homozygous knockout of E2A will reduce the mutation rate at the V region of the rearranged IgL locus. The cell population will remain sIgM(+).
Within this cell line AID is expressed constitutively as a cDNA expression cassette under the control of a β-actin promoter. This is advantegous, because the splice variant E47 of the E2A gene regulates transcription of AID by binding the E-box in the enhancer of the AID gene (Sayegh et al., 2003). This study was done in mouse and might be transferred to chicken, too. By expressing transgenic AID, I expect to rule out a side effect of the E2A knockout on AID expression. Changes in AID expression would result in changes of hypermutation rate and an effect of E2A gene disruption would not be detectable.

Another advantage of the cell line is the easy inducibility by Cre recombinase. That enables a recycle of the used marker cassettes (described in chapter 3.3.5).

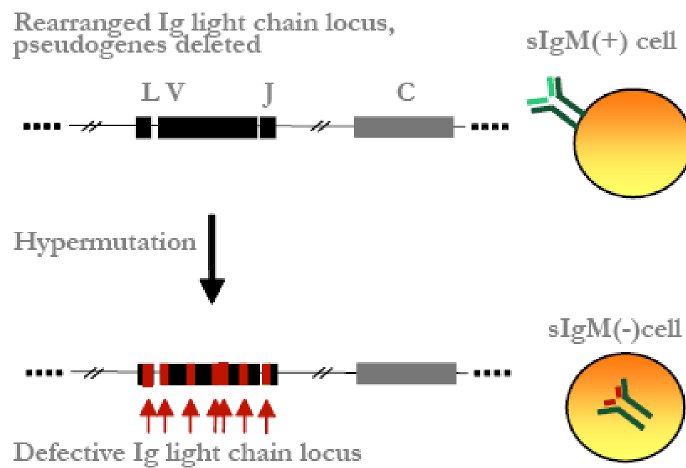


Figure 1.5 Measurement of deleterious Ig mutations in DT40.
Hypermutation introduces point mutations into the rearranged IgL chain locus. The pseudogenes are deleted and the cell has to undergo diversification solely by HM. Not all of the mutations are beneficial. Missense or nonsense mutations can interfere with the expression of sIgM. The appearance of sIgM(-) cells in a predominantly sIgM(+) population can be measured after antibody staining in flow cytometer.

To knock out the E2A gene I designed two different targeting constructs to delete both alleles. The vectors contain the blasticidin (bsr) or mycophenolic acid (gpt) drug resistance marker respectively, controlled by the β-actin promoter. The loxP sites surrounding the transgene enable a recycle of the marker. The marker cassette is flanked by arms sequences of the E2A genomic locus which mark the 5' and 3' border of the deletion. The 5' and 3' arms of the knockout constructs were designed to inactivate the E2A gene by the deletion of the exons encoding codons 142 – 463 in the case of the vector pE2Absr and codons 172 – 353 in the case of pE2Agpt (Figure 1.6.). The deleted region includes the putative NLS and the highly conserved activation domain II (ADII) which is essential for the function of the transcription factor (Kee and Murre, 1998; Aronheim et al., 1993).
Homologous recombination between the arms sequences of the construct and the corresponding genomic sequence of the E2A locus leads to a targeted integration of the construct. The genomic sequence flanked by the arms will be lost and substituted by the drug resistance transgene.
After transfection the cells were cultured in medium containing these drugs. This is to select for clones which integrated the drug resistance transgene into their genome. Although DT40 cells have a high ratio of targeted integration, still many clones integrate the construct randomly. Identification of a targeted clone is done by PCR using a forward primer specific for E2A genomic locus upstream of the 5' arm and a reverse primer specific for the drug resistance marker cassette. After identification of a targeted clone, the drug resistance can be removed by tamoxifen-induced cre recombination.

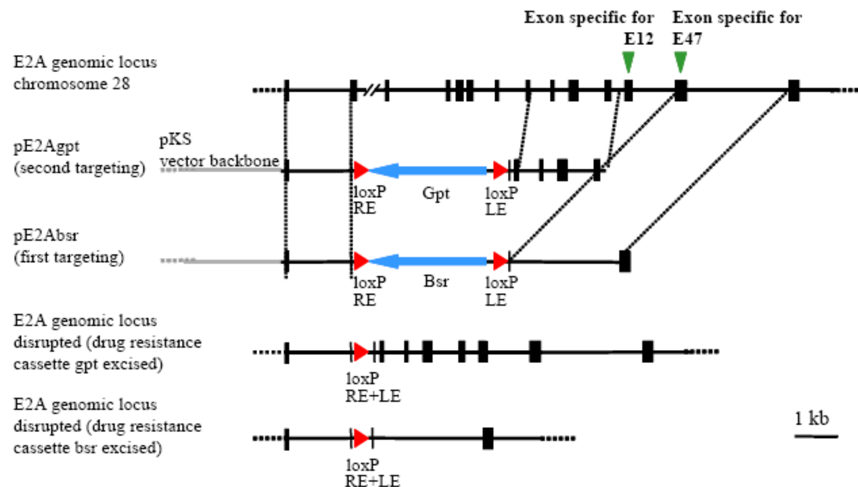


Figure 1.6 E2A gene disruption.
Aligned maps of the chicken E2A locus, the targeting constructs and the disrupted locus after targeted integration of the constructs and marker excision. Exons of the E2A gene are black boxes. E12 and E47 specific exons are marked by green triangles. LoxP sites are indicated by red triangles and the drug resistance cassettes by blue arrows. The drug resistance is integrated into the genome via the flanking arms sequences (sequences between the dotted lines). The genomic sequence flanked by the arms sequence is deleted. After Cre-mediated excision of the marker, only the loxP sequence remains in the genome.

The first allele was knocked out via transfection of the construct pE2Absr and yielded in the cell line AID$^R\psi$V-E2A+/-. Targeted transfectants were identified by PCR screening. One clone was chosen and grown up for a second transfection with the construct pE2Agpt to knock out the second allele. Targeted transfectants yielding in a homozygous knockout were identified in the same way like for the heterozygous knockout. One clone was chosen for further studies. The marker cassettes bsr and gpt were removed by Cre recombinase induction with Tetrahydroxytamoxifen and the resulting cell line AID$^R\psi$V$^-$E2A$^{-/-}$ was analyzed by FACS.

sIgM was detectable by PE antibody staining. AID cDNA expression is coupled to GFP by an IRES and enables a correlation of AID level and GFP expression. Cells expressing no or low GFP lost the floxed AID transgene cassette presumably due to a leaky Cre recombinase expression. Because AID-deficient cells stop hypermutation completely, GFP-negative cells were excluded from the study. A first FACS analysis of the primary clone showed a strong decrease of the sIgM(-) population from 28.9% in the progenitor clone AID$^R\psi$V$^-$ (Figure 1.7.A) to 3.3% in the homozygous knockout mutant AID$^R\psi$V$^-$E2A$^{-/-}$ (Figure 1.7.B).

This was a first hint for E2A transcription factors playing a role in AID-mediated HM. During my PhD thesis, I studied in more detail at which point E2A is engaged in the diversification process.

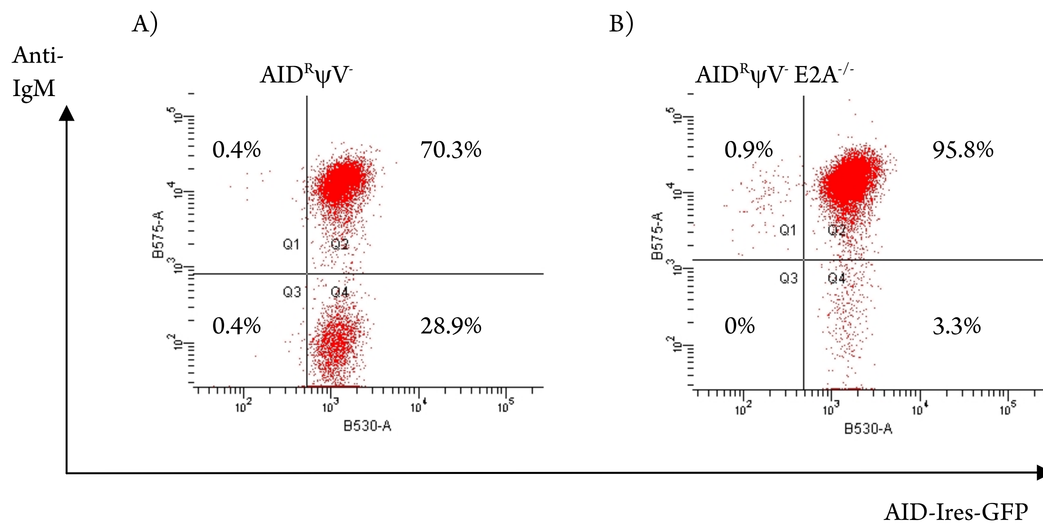


Figure 1.7 FACS analysis of the E2A homozygous knockout.
A) Hypermutation pattern of the progenitor cell line AID^{R}ψV^{-}. The cell line is predominantly sIgM(+) but exhibits a sIgM(-) subpopulation due to ongoing HM. B) Dot plot of the homozygous E2A knockout mutant AID^{R}ψV^{-}E2A$^{-/-}$. Compared to the progenitor clone, cells shift to sIgM(+) thereby reflecting a decrease of HM.

1.12 Aims

To ensure genomic integrity, HM needs to be targeted specifically to the Ig genes. The rare mistargeting of HM can result in mutations in oncogenes and is thought to contribute to the development of B-cell malignancies (Pasqualucci et al., 2007; Pérez-Durán et al., 2007; Okazaki et al., 2007). The question still remains how AID deamination activity is predominantly restriced to the Ig loci. Beyond the need of a transcriptionally active promoter, AID deaminates only ssDNA with high efficiency and the level of deamination correlates with the level of transcription of the target gene. Transcription is a regulated DNA process which uses platforms consisting of a number of cis-acting elements for transcription-regulating factors to assemble. These assemblies of protein factors have various functions, most notably mediating enhancer and promoter interactions, and mediating the recruitment of chromatin modifying enzymes and the transcription machinery. AID is assumed to physically interact with another factor which recruits AID to the Ig gene, as overexpression of AID in non- hypermutating B and non-B cells leads to unspecific deamination of strong transcribed genes and transgenes (Kotani et al., 2005; Martin et al., 2002; Yoshikawa et al., 2002). It is appealing to think that a unique assembly of protein factors on a DNA platform (in the following referred to as *'HyCorE'* for 'Hypermutation Core Element') could recruit the HM machinery specifically to Ig loci. However, no groups successfully identified the putative factors or their binding sites at the Ig loci and therefore in my study I want to use a new system to elucidate AID-targeting to hypermutating loci.

In mouse model, attempts to define putative targeting elements in Ig gene enhancer sequences were largely uninformative, because any observable decrease in HM was accompanied by a decrease in transcription. The Ig

loci contain multiple enhancer sequences which can be far distant from each other. This makes it impossible to study redundant elements which might function as *'HyCorE'*. DT40 is the only system which allows an extended study of DNA elements of the Ig locus. The rearranged IgL locus including the pseudo V gene sequences is only around 30 kb in contrast to the loci in mouse and human, which are several Mb in size. This advantage of DT40 enables to screen by deletion and insertion of parts of the IgL locus not only for cooperative and additive acting DNA elements, but also for redundant DNA elements supporting HM, and this is the first study revealing also redundant elements.

To examine the issue, the strategy is to use a GFP transgene as a reporter to test the conditions for AID deamination of the GFP gene substrate. A systemic analysis of the IgL locus of DT40 by extensive genetic manipulation will give insights into which cis-elements are involved in the targeting of diversification and help to identify the relevant proteins. Deleting and reinserting parts of the Ig locus together with the GFP transgene into the position of the genomic Ig locus will help to identify cooperative as well as redundant motifs.

Enhancer elements of the Ig locus contain binding sites for several trans-acting factors, including NFκB, octamer-binding transcription factor proteins, E12 and E47 and ETS family proteins. Targeted disruption of any of the transcription factor genes in transgenic mouse model typically has pleiotropic effects on transcription and early lymphocyte development, making it difficult to determine the role of a given factor in HM. Thus, to study the function of proteins for Ig transcription and hypermutation, DT40 cells can be an advantegous system. I used it in my thesis to study the role of the transcription factor E2A, which is an important regulator in B cells and binding to Ig promoters and enhancers.

2 MATERIALS

2.1 Bacterial Strain

E. coli-DH5α	cells were used as chemically competent cells F^-, Δ80dlacZΔM15, Δ(lacZYA-argF)U169, deoR, recA1, endA1, hsdR17(r_K^-, m_K^+), phoA, supEE44, λ-, thi-1, gyrA96, relA1
TOPO TA Cloning® Kit	TOP10 One Shot® Cells [Invitrogen Corporation]

2.2 Buffers and solutions

0.5 M EDTA pH 8.0	EDTA•2 H_2O	186.1 g
	milliQ H_2O	fill up to 1 l
	NaOH	to pH 8.0
10x PBS	NaCl	40 g (0.68M)
	KCl	1 g (13.5mM)
	Na_2HPO_4 (358.14 g/mol)	7.2 g (20mM)
	KH_2PO_4 (136.09 g/mol)	1.2 g (9mM)
	1x distilled H_2O	to 500 ml
	NaOH	to pH 7.4
50x TAE	1M TRIS•HCl (pH 8.0)	50 ml (0.5M)
	0,5M EDTA	10 ml (50mM)
	milliQ H_2O	40 ml
50x TE	1M TRIS•HCl (pH 8.0)	242 g (1M)
	Glacial Acetic Acid (100%)	57.1 ml (1 M)
	0.5M EDTA (pH 8.0)	100 ml (50 mM)
	milliQ H_2O	to 1000 ml

Agarose gel	0.8% (for a DNA size smaller than 600 bp: 2%) Agarose Serva Premium [Serva Electrophoresis GmbH] in 1x TAE, 0.25µg/ml ethidium bromide [Sigma-Aldrich Chemie GmbH]	
Binding Solution	Potassium Iodide	100 g
	Distilled Water	70 ml
FTB	0,5 M PIPES	20 ml (100 mM)
	$CaCl_2 \bullet 2H_2O$	2.2 g (15 mM)
	Distilled Water	to 1000 ml
	KCl	18.64 g (250 mM)
	$MnCl_2 \bullet 4H_2O$	10.88 g (55 mM)
	The pH was adjusted to 6.7 after KCl and $MnCl_2 \bullet 4H_2O$ treatments respectively. The solution is sterilized by filtration.	
K Buffer	Expand Long Template PCR System Buffer 2	10 ml (10%)
	Distilled water	90 ml
	Tween 20	0.5%
	Proteinase K	0.1 mg/ml
	Tween 20 and Proteinase K are added directly before use.	
Loading buffer 6x	Bromphenol Blue [Sigma-Aldrich Chemie GmbH]	0.25 g (0.25%)
	Xylene Cyanol FF [Sigma-Aldrich Chemie GmbH]	0.25 g (0.25%)
	Glycerol [Merck KGaA]	30 ml (30%)
	milliQ H_2O	to 100 ml
PCR loading buffer 10x Cresol Red	Cresol Red [Sigma-Aldrich Chemie GmbH]	150 mg (0.15%)
	Ficoll 400 [Fluka, Sigma-Aldrich Chemie GmbH]	10 g (10%)
	milliQ H_2O	to 100 ml
Proteinase K Buffer	5 M NaCl	200 µl (100 mM)
	2 M Tris-HCl (pH 8,0)	50 µl (10 mM)
	0.5 M EDTA	500 µl (25 mM)
	1xdistilled water	up to 10 ml
	Add 0.5% SDS and 0.1 mg/ml proteinase K prior to use.	
Solution I	50x TE	1 ml
	Distilled Water	49 ml
	RNaseA (100mg/ml)	30 µl
Solution II	NaOH 10 N	10 ml
	Distilled Water	500 ml
	20 % SDS	2,5 µl

Solution III	Glacial acetic acid	70 ml
	Distilled Water	70 ml
	Potassium Acetate 5 M	360 ml

2.3 Cell Culture

Chicken medium	DMEM/F-12 Medium [Gibco™ Invitrogen Corporation]	500 ml
	FBS [Biochrom AG]	50 ml (10%)
	Chicken Serum [Pan GmbH]	5 ml (1%)
	Penicillin/Streptomycin (10,000 $^{U}/_{ml}$ / 10,000 $^{\mu g}/_{ml}$) [Gibco™ Invitrogen Corporation]	10 ml (200 $^{U}/_{ml}$ /200 $^{\mu g}/_{ml}$))
	β-Mercaptoethanol 1M [Sigma-Aldrich Chemie GmbH]	50 µl (0.5 mM)

Drug selection

Blasticidin [Gibco™ Invitrogen Corporation]: dissolved in chicken medium to a final concentration of 12.5 µg/ml.

Mycophenolic acid [Sigma-Aldrich Chemie GmbH]: dissolved in chicken medium to a final concentration of 0.5 µg/ml.

Puromycin [Sigma-Aldrich Chemie GmbH]: dissolved in chicken medium to a final concentration of 1µg/ml.

Removal of drug selection marker

Drug resistance markers can be removed by incubation with tetrahydroxy-tamoxifen (4-HT):

4-HT (387.5 g/ml) 5mg [Sigma-Aldrich Chemie GmbH] in 64.5 ml chicken medium for a concentration of 0,2 mM. For use in cell culture 4-HT is diluted to a final working concentration of 100nM.

FACS staining buffer	1x PBS	450 ml
	Fetal Bovine Serum	50 ml (10%)
	Sodium Azide	2.5 ml (7.5 mM)

Freezing medium	Chicken medium	70 ml (70%)
	FBS [Biochrom AG]	20 ml (20%)
	DMSO [Sigma-Aldrich Chemie GmbH]	10 ml (10%)

2.4 Cell Lines

DT40

Mutants of the chicken B cell line DT40 were used to transfect constructs. AID expression at a constant high level is ensured by an AID cDNA expression transgene.

AID$^{R}\psi V^{-}$ — The pseudogenes of the rearranged IgL locus were deleted thereby forcing the clone to diversify its VJ region solely by HM.

The endogenous AID was deleted and reconstituted by random transfection using a dicistronic expression cassette that links AID cDNA and GFP expression by an internal ribosomal entry site (IRES) and is controlled by the chicken β-actin promoter. The expression cassette is followed by the puromycin resistance gene which is controlled by an additional β-actin promoter (Arakawa et al., 2002 and 2004).

Therefore, AID expression can be verified in two ways:

i) FACS analysis of GFP protein fluorescence and

ii) maintaining the cell clone in chicken medium containing puromycin allows the selection of AID-positive cells.

AID$^{R1}\psi V^{-}$ — The entire 20 kb of the ψV locus of the rearranged IgL gene are deleted thereby forcing the clone to diversify its VJ region solely by HM.

The endogenous AID within this clone was deleted and AID was reconstituted by random transfection of a dicistronic expression cassette that links AID cDNA and the mycophenolic acid resistance gene by an IRES and is controlled by the chicken β -actin promoter (Arakawa et al., 2008).

Maintaining the cell clone in chicken medium containing mycophenolic acid allows the selection of AID-positive cells.

AID^{R1}IgL^{-} — The progenitor clone of this cell line is AID$^{R1}\psi V^{-}$.

Additionally to the ψV locus, the entire rearranged IgL locus is deleted and a puromycin resistance transgene is located at its position (Blagodatski et al., 2009).

AID$^{R}\psi V^{-}$E2A$^{+/-}$, AID$^{R}\psi V^{-}$E2A$^{-/-}$ — These cell lines are derived from the precursor clone AID$^{R}\psi V^{-}$. One or both alleles of the E2A gene are deleted respectively (Schoetz et al., 2006).

Nomenclature of the cell lines generated during this study

AID$^{R}\psi V^{-}$E2A^{RtE12}, AID$^{R}\psi V^{-}$E2A^{RtE47} — These cell lines are derived from the precursor clone AID$^{R}\psi V^{-}$E2A$^{-/-}$. The two splice variants E12 and E47 of the E2A gene are inserted randomly as a transgenic cDNA cassette into the genome.

AID^{R1}IgL$^{n,\ GFP2}$ — These cell lines are derived from the precursor clone AID^{R1}IgL^{-}, where the GFP2 reporter transgene together with a fragment of the IgL locus was inserted at the position of the deleted rearranged IgL locus. *'n'* describes the name of the inserted IgL fragment according to the deletion vectors on p.39 and the PCR amplified fragments described in the list of primers chapter 8.1., 3.

2.5 Chemicals

All chemicals — [Merck KGaA] or [Sigma-Aldrich Chemie GmbH]

2.6 Consumables

Bacterial culture	96 wellplate [Greiner bio-one GmbH]
	Petri dish Falcon Easy Grip™ [Becton Dickinson Labware]
Cell culture	96 wellplate NuncTM, [Nalge Nunc International]
	24 and 6 wellplate Falcon Multiwell™ [Becton Dickinson Labware]
	Falcon Tissue Culture Flask [Becton Dickinson Labware]
Cuvettes	Electroporation Bio Rad Gene Pulser® Cuvette [Bio-Rad Lab.]
	Photometry UV-Vis cuvette [Eppendorf AG]
PCR	96 wellplate [ABgene Inc.]
	Cap strips [ABgene Inc.]
Pipet tips	[Eppendorf AG]
	filtered tips Tip One [StarLab GmbH]
Tubes	1.5ml and 2ml [Eppendorf AG]
	10ml [Greiner bio-one GmbH]
	15ml and 50ml Falcon Blue Max™ [Becton Dickinson Labware]
	FACS Falcon Round Bottom Tube [Becton Dickinson Labware]
	Kryo vials Cryo.S [Greiner bio-one GmbH]
	Vi-Cell [Beckmann Coulter GmbH]

2.7 DNA Size Marker

λ DNA-*Hin*dIII	(500 ng/μl) [New England Biolabs GmbH]	400 μl
ϕX174 DNA-*Hae*III	(1.000 ng/μl) [New England Biolabs GmbH]	100 μl
6x loading buffer		334 μl
MiliQ water		1166 μl
	The mixture was incubated for 15 min at 70°C and for 5 min on ice.	

2.8 Enzymes and dNTPs

Calf Intestinal Phosphatase	[New England Biolabs GmbH]
DNA Polymerases	Expand Long Template PCR System [Roche Diagnostics GmbH]
	Pfu Ultra Hotstart [Stratagene]
	Taq Polymerase [New England Biolabs GmbH]
dNTPs	dNTP Mix 10 mM [Fermentas GmbH]
	dATP [Fermentas GmbH]

Proteinase K	[Qiagen]
Restriction Enzymes	[New England Biolabs GmbH]
RNaseA	[Qiagen]

2.9 Experimental Kits

DNA Ligation	Version 2.1 [Takara Bio Inc.]
First strand cDNA Synthesis	Super Script IIITM [Invitrogen GmbH]
Gel Extraction	[Qiagen GmbH]
PCR Purification	[Qiagen GmbH]
Plasmid Isolation	Maxi [Qiagen GmbH]
	Mini [Fermentas GmbH]
Sequencing	BigDye Terminator v3.1 Cycle Sequencing Kit [Applied Biosystems Inc.]
Topo TA Cloning® Kit	[Invitrogen GmbH]

2.10 Immuno-staining antibodies & anti-antibodies

First antibody	Mouse Anti Chicken IgM-UNLB (Clone M- 1)
Second antibody	Goat Anti Mouse IgG (H+L)-RPE Human Adsorbed
	[Southern Biotech. Ass. Inc., Biozol]

2.11 Instruments

Analytical FACS	BD™ LSRII Flow Cytometer [Becton Dickinson]
Cell Viability Analyzer	Vi-Cell™ [Beckman Coulter GmbH]
Centrifuges	Heraeus [Kendro Lab. Products]
Culture Shaker	Innova 4430 [New Brunswick Scientific]
DNA Sequencer	ABI 3730 DNA Analyzer [Applied Biosystems & Hitachi; Hitachi High Tech. Corp.].
Electrophoresis	Chambers [Bio-Rad Lab.]
	Power supply Power Pac 300 [Bio-Rad Lab.]
Electroporator	Gene Pulser Xcell™ [Bio-Rad Lab.]
Gel Visualization	Gel Doc 2000™ [Bio-Rad Lab]

Incubators	Heraeus [Kendro Lab. Products]
Laminar Work Bank	Heraeus Herasafe KS [Kendro Lab. Products]
Light Microscope	Axiovert 25 [Zeiss]
Magnetic stirrer	MR 3001 [Heidolph Instruments GmbH]
PCR machines	GeneAmp® PCR System 9700 [Applied Biosystems]
pH-Meter	Inolab Level 1 [WTW Wissenschaftlich-Technische Werkstätten GmbH]
Phosphoimager	Fuji, FLA-3000 [Fuji Photo Film Corp.]
Thermo mixer	Thermomixer comfort [Eppendorf GmbH]
Tube rotator	Blood Tube Rotator SB1Stuart Scientific [Barloworld Scientific]
UV Spectrophotometer	BioPhotometer [Eppendorf GmbH]
Vacuum Pump	[Vacuubrand GmbH]
Vacuum Manifold	[Millipore GmbH]
Vortex Mixer	[neoLab GmbH]

2.12 Media

2YT Broth	Bacto tryptone [Becton, Dickinson and Company Inc.]	16 g
	Yeast extract [Sigma-Aldrich Chemie GmbH]	10 g
	NaCl	5 g
	10N NaOH	250 µl
	Distilled water	up to 1000 ml
	Autoclave and add 250mg/l ampicillin [Sigma-Aldrich Chemie GmbH] before use.	
LB Agar Plates	LB Broth Base [Sigma-Aldrich Chemie GmbH]	20 g
	Bacto™ Agar [Becton, Dickinson and Company Inc.]	15 g
	Distilled water	up to 1000 ml
	Autoclave and add 250mg/l ampicillin and 0.4% X-gal [Sigma-Aldrich Chemie GmbH] before pouring onto petri plates.	
SOB Broth	Bacto tryptone [Becton, Dickinson and Company Inc.]	20 g
	Yeast extract [Sigma-Aldrich Chemie GmbH]	5 g
	NaCl	0.585 g
	KCl	0.186 g
	Distilled water	up to 1000 ml
	$MgCl_2$	10 mM
	$MgSO_4$	10 mM
	Add $MgCl_2$ and $MgSO_4$ after autoclaving.	

2.13 Nucleotide sequences

Abbreviation	Description	Genebank accession number
Bsr	blasticidin resistance gene	P19997
E12	mRNA	AJ579995
E47	mRNA	AJ579996
eGFP	green fluorescent protein	AJC06700.1
gpt	guanine phosphoribosyl transferase; resistance to mycophenolic acid	AAA23928
puro	puromycin resistance gene	

2.14 Oligonucleotides

Oligomers [Sigma-Genosys]

The sequences are summarized in Supplementary information, List of primers, chapter 8.1.

2.15 Plasmids

pBluescript II KS (+) phagemid vector [Stratagene, California; USA]

Drug resistance marker

ß-actin promoter-blasticidin* flanked by mutated loxP sites was used for drug selection with blasticidin [Invitrogen GmbH]

ß-actin promoter-mycophenolic acid* flanked by mutated loxP sites was used for drug selection with mycophenolic acid [Sigma-Aldrich Chemie GmbH]

ß-actin promoter-puromycin* flanked by mutated loxP sites was used for drug selection with puromycin [Sigma-Aldrich Chemie GmbH]

The marker cassettes are described on the website: http://pheasant.gsf.de/DEPARTMENT/dt40.html

2.16 Software

Bioinformatical tools MatInspector software [Genomatix Software GmbH] MatInspector is a tool to screen a DNA sequence for transcription factor binding sites (Quandt et al., 1995; Cartharius et al., 2005).

Software for statistics R : http://www.r-project.org/ R is a freely available software for statistical computing and graphics. I used R to perform a test of significance according to the Wilcoxon U-Test.

3 METHODS

“Gene disruption in a cell line is an alternative to knockouts in murine ES cells, if the mutant phenotype can be studied in cell culture. The chicken B cell line DT40 is popular for these studies due to unusual high ratios of targeted to random integration.” (Buerstedde and Takeda, Cell 1991)

3.1 Vector Design

The vector backbone for all constructs is *pBluescript II KS(+)*. All vectors used in this study are mutant loxP vectors containing loxP flanked drug resistance marker genes. LoxP are DNA sites, which enable the excision of the DNA located between two loxP sites by a Cre-recombinase induced loxP recombination (Arakawa et al., 2001). It is a useful approach to recycle the drug resistance gene and use it again for additional transfections in the same cell line. The vectors for cDNA complementation of knockout phenotypes are *pIresSfGpt* and *pIresSfBsr*. Both contain loxP flanked expression cassettes to clone in the cDNA of interest (Arakawa et al., 2001). The expression cassettes consist of a chicken β -actin promoter and a SV40 polyA signal followed by an IRES and a drug resistance gene.

3.1.1 Databases

The design of targeting vectors for knockout studies in DT40 via homologous recombination needs accurate sequence information of the targeted genomic region. More than 90% of chicken genomic sequence is released (International Chicken Genome Sequencing Consortium, 2004) and accessible at public databases like *The National Center for Biotechnology Information* (NCBI) or *The European Molecular Biology Laboratory* (EMBL). I used the databases to obtain the E2A genomic sequence for the E2A knockout study and the IgL locus sequence for the mapping of the cis-elements.

Cloned full-length cDNA’s are neccessary to complement mutant phenotypes and artificially express the encoded protein. The DT40 Web site presents a cDNA library called ‘riken1’ and ESTs (expressed sequences

tags) expressed in chicken bursal lymphocytes. Moreover, this resource contains a collection of SAGE (Serial Analysis of Gene Expression) tags from both the chicken bursal lymphocytes and DT40 cells (Wahl and Caldwell, 2004). Large EST and full length cDNA sequences (Abdrakhmanov et al., 2000; Caldwell et al., 2005) are available at the public databases NCBI and EMBL.

I used the databases to identify E12 and E47 cDNAs. The cDNA can easily be obtained by a BLAST search. The cDNA page lists the full-length cDNA sequence, the 5' and 3'UTR sequences and the translated ORF. By following the link to ENSEMBL one can access information about predicted intron-exon structure of the gene and its genomic context.

3.1.2 E2A complementation vector

After inactivation of the E2A gene in DT40, the two splice products E12 and E47 are reconstituted respectively. Therefore I used the vectors *pIresSfGpt* and *pIresSfBsr* to clone in the cDNA of the proteins. The two constructs contain a loxp flanked expression cassette which consists of a chicken β-actin promoter followed by an IRES sequence and a gpt or bsr gene respectively. A multiple cloning site between promoter and IRES facilitates the cloning of a cDNA of interest. E12 and E47 full-length cDNAs isolated from the riken1 bursal cDNA library (Caldwell et al., 2005) and their entire sequence was verified by primer walks (primers according to table S.1). The E12 and E47 expression vectors were made by cloning the corresponding cDNAs downstream of the β-actin promoter and upstream of the IRES using the SfiI restriction sites resulting in the two final constructs *pE12Gpt* and *pE47Bsr*.

3.1.3 IgL targeting vectors for cis-element study

For the mapping of the cis-elements, the entire rearranged IgL locus of the DT40 had been deleted. The resulting cell line is $AID^{R1}\psi V^{-}$. Hereafter, fragments of the IgL locus are inserted again in the position of the deleted region and analyzed for their potential to activate HM. This was done by using a reporter assay containing a GFP transgene which was cloned and transfected together with the extended deletions of the IgL locus.

In the following, I describe the produced targeting vectors together with the cloning strategy.

All clonings are verified by miniprep and enzymatical digest hereafter. For transfection, a maxiprep was done and linearization with the unique restriction enzyme NotI.

GFP2 The GFP2 construct was made by H. Arakawa and R.B. Caldwell. It combines a GFP open reading frame and the bsr gene via an IRES under the control of a RSV

promoter. GFP2 was flanked by unique BamHI restriction sites (Blagodatski et al., 2009).

pIgL$^{-,GFP2}$

The targeting construct was made by cloning the arm sequences (List of primers chapter 8.1) for targeting of the rearranged IgL locus into *pBluescriptKS+* and then inserting GFP2 into unique BamHI sites. Upstream of the RSV promoter, there are unique NheI and SpeI sites located which can be used for cloning of the Ig locus fragments.

Transfection of this vector results in a deletion of the entire rearranged IgL locus. This construct is the precursor of all further vectors designed for the mapping of the cis-elements.

pIgL^{W,GFP2}

The *'W'* fragment of 9784 nucleotides starts at position -7 relative to the first base of the IgL start codon and corresponds to the chicken genome coordinates chr15:8165070-8176699 but lacks the VJ intervening sequence. The *'W'* fragment was inserted into unique NheI/SpeI sites of *pIgL*$^{-,GFP2}$. This plasmid served as precursor DNA for PCR amplification of all further IgL deletion fragments.

IgL fragment deletion vectors

All fragments were amplified using primers according to the list of primers (chapter 8.1). The amplicons were PCR purified and digested with SpeI/NheI. The inserts were cloned into *pIgL*$^{-,GFP2}$ which was digested with SpeI/NheI, too.

Nomenclature

The first series of constructs was designed to examine the *'W'* fragment in more detail. Each deletion fragment is named with an alphabetic character.

The following deletion constructs were designed to analyze the *'S'* fragment, in the following referred to as *'0-4'*, in more detail. *'0-4'* has 4 kb in size and was divided in 4 parts: *'0-1'*, *'1-2'*, *'2-3'* and *'3-4'*. According to this, for example *'0-2'* are the first 2 kb of *'0-4'*, *'2-4'* are the last 2 kb of *'0-4'* and so on.

For the serial end deletions, the name describes the inserted fragment. For example, *'0.2-2'* has the first 200 bp of the *'0-2'* fragment deleted, *'0-1.8'* has the last 200 bp of the *'0-2'* fragment deleted.

For the internal deletions, the name describes the inserted fragment with the position of the deletion within the fragment assigned behind the Δ. For example, *'0-4Δ2.2-2.4'* is the fragment *'0-4'* with the part of the DNA at position *'2.2-2.4'* being

deleted. A '*series*' of constructs consists of all deletions performed on one specific part of the '*0-4*'.

'*W*' series	$pIgL^{A,GFP2}$, $pIgL^{B,GFP2}$, $pIgL^{C,GFP2}$, $pIgL^{D,GFP2}$, $pIgL^{E,GFP2}$, $pIgL^{F,GFP2}$, $pIgL^{G,GFP2}$, $pIgL^{I,GFP2}$, $pIgL^{K,GFP2}$, $pIgL^{L,GFP2}$, $pIgL^{M,GFP2}$, $pIgL^{N,GFP2}$, $pIgL^{P,GFP2}$, $pIgL^{S,GFP2}$
'*0-2*' series 200bp end deletion	$pIgL^{0-2,GFP2}$, $pIgL^{0.2-2,GFP2}$, $pIgL^{0.4-.2,GFP2}$, $pIgL^{0.6-2,GFP2}$, $pIgL^{0.8-2,GFP2}$, $pIgL^{0-1.2,GFP2}$, $pIgL^{0-1.4,GFP2}$, $pIgL^{0-1.6,GFP2}$, $pIgL^{0-1.8,GFP2}$
'*2-4*' series 200bp end deletion	$pIgL^{2-4,GFP2}$, $pIgL^{2.2-4,GFP2}$, $pIgL^{2.4-.4,GFP2}$, $pIgL^{2.6-4,GFP2}$, $pIgL^{2.8-4,GFP2}$, $pIgL^{2-3.2,GFP2}$, $pIgL^{2-3.4,GFP2}$, $pIgL^{2-3.6,GFP2}$, $pIgL^{2-3.8,GFP2}$
'*0-4*' series 1kb serial deletion	$pIgL^{0-1,GFP2}$, $pIgL^{0-2,GFP2}$, $pIgL^{0-3,GFP2}$, $pIgL^{0-4,GFP2}$, $pIgL^{1-2,GFP2}$, $pIgL^{2-3,GFP2}$, $pIgL^{3-4,GFP2}$, $pIgL^{2-4,GFP2}$, $pIgL^{1-4,GFP2}$
'*0-4*' series 200bp internal deletion	$pIgL^{0-3.8,GFP2}$, , $pIgL^{0.2-4,GFP2}$
'*0-4*' series 400bp internal deletion	$pIgL^{0-3.6,GFP2}$, , $pIgL^{0.4-4,GFP2}$
'*2-3*' series 50bp end deletion	$pIgL^{2-2.05,GFP2}$, $pIgL^{2-2.10,GFP2}$, $pIgL^{2-2.15,GFP2}$, $pIgL^{2-2.20,GFP2}$, $pIgL^{2-2.25,GFP2}$, $pIgL^{2-2.30,GFP2}$, $pIgL^{2-2.35,GFP2}$, $pIgL^{2-2.40,GFP2}$, $pIgL^{2-2.45,GFP2}$, $pIgL^{2-2.50,GFP2}$, $pIgL^{2-2.55,GFP2}$, $pIgL^{2-2.60,GFP2}$, $pIgL^{2-2.65,GFP2}$, $pIgL^{2-2.70,GFP2}$, $pIgL^{2-2.75,GFP2}$, $pIgL^{2-2.80,GFP2}$, $pIgL^{2-2.85,GFP2}$, $pIgL^{2-2.90,GFP2}$, $pIgL^{2-2.95,GFP2}$, $pIgL^{2-2.20,GFP2}$, $pIgL^{2.05-3,GFP2}$, $pIgL^{2.10-3,GFP2}$, $pIgL^{2.15-3,GFP2}$, $pIgL^{2.20-3,GFP2}$, $pIgL^{2.25-3,GFP2}$, $pIgL^{2.30-3,GFP2}$, $pIgL^{2.35-3,GFP2}$, $pIgL^{2.40-3,GFP2}$, $pIgL^{2.45-3,GFP2}$, $pIgL^{2.50-3,GFP2}$, $pIgL^{2.55-3,GFP2}$, $pIgL^{2.60-3,GFP2}$, $pIgL^{2.65-3,GFP2}$, $pIgL^{2.70-3,GFP2}$, $pIgL^{2.75-3,GFP2}$, $pIgL^{2.80-3,GFP2}$, $pIgL^{2.85-3,GFP2}$, $pIgL^{2.90-3,GFP2}$, $pIgL^{2.95-3,GFP2}$ The fragments were PCR amplified using a forward primer with a NheI site and a reverse primer with a SpeI site. The digested and purified fragments were inserted into unique NheI/SpeI sites of $pIgL^{-,GFP2}$.
'*0-4*' series 200bp internal deletion	$pIgL^{0-4\Delta0.2-0.4,GFP2}$, $pIgL^{0-4\Delta0.4-0.6,GFP2}$, $pIgL^{0-4\Delta0.6-0.8,GFP2}$, $pIgL^{0-4\Delta0.8-1.0,GFP2}$, $pIgL^{0-4\Delta1.0-1.2,GFP2}$, $pIgL^{0-4\Delta1.2-1.4,GFP2}$, $pIgL^{0-4\Delta1.4-1.6,GFP2}$, $pIgL^{0-4\Delta1.6-1.8,GFP2}$, $pIgL^{0-4\Delta1.8-2.0,GFP2}$, $pIgL^{0-4\Delta2.0-2.2,GFP2}$, $pIgL^{0-4\Delta2.2-2.4GFP2}$, $pIgL^{0-4\Delta2\cdot4-2.6,GFP2}$, $pIgL^{0-4\Delta2\cdot6-2.8,GFP2}$, $pIgL^{0-4\Delta2.8-3.0,GFP2}$, $pIgL^{0-4\Delta3.0-3.2,GFP2}$, $pIgL^{0-4\Delta3.2-3.4,GFP2}$, $pIgL^{0-4\Delta3.4-3.6,GFP2}$, $pIgL^{0-4\Delta3.6-3.8,GFP2}$

‘0-4’ series 400bp internal deletion

$pIgL^{0-4\Delta0.4-0.8,GFP2}$, $pIgL^{0-4\Delta0.8-1.2,GFP2}$, $pIgL^{0-4\Delta1\cdot2-1.6,GFP2}$, $pIgL^{0-4\Delta2.6-2.0,GFP2}$, $pIgL^{0-4\Delta2.0-2.4,GFP2}$, $pIgL^{0-4\Delta2.4-2.8,GFP2}$, $pIgL^{0-4\Delta2.8-3.2,GFP2}$, $pIgL^{0-4\Delta3.2-3.6,GFP2}$

‘2-3’ series 50bp internal deletion

$pIgL^{2-3\Delta2.05-2.10,GFP2}$, $pIgL^{2-3\Delta2.10-2.15,GFP2}$, $pIgL^{2-3\Delta2.15-2.20,GFP2}$, $pIgL^{2-3\Delta2.20-2.25,GFP2}$, $pIgL^{2-3\Delta2.25-2.30,GFP2}$, $pIgL^{2-3\Delta2.30-2.35,GFP2}$, $pIgL^{2-3\Delta2.35-2.40,GFP2}$, $pIgL^{2-3\Delta2.40-2.45,GFP2}$, $pIgL^{2-3\Delta2.45-2.50,GFP2}$, $pIgL^{2-3\Delta2.50-2.55,GFP2}$, $pIgL^{2-3\Delta2.60-2.65,GFP2}$, $pIgL^{2-3\Delta2.65-2.70,GFP2}$, $pIgL^{2-3\Delta2.70-2.75,GFP2}$, $pIgL^{2-3\Delta2.75-2.80,GFP2}$, $pIgL^{2-3\Delta2.80-2.85,GFP2}$, $pIgL^{2-3\Delta2.85-2.90,GFP2}$, $pIgL^{2-3\Delta2.90-2.95,GFP2}$

The fragments were amplified in a two step PCR. In a first step a 5’ fragment (DNA1) and a 3’ fragment (DNA2) were amplified separately. These two fragments flank the region to be deleted. The forward primer of DNA1 and the reverse primer of DNA2 contain unique restriction sites NheI and SpeI respectively for later cloning into the vector $pIgL^{-,GFP2}$. The reverse primer of DNA1 contains a unique NdeI restriction site and a 9bp DNA sequence complementary to the 5’end of DNA2. The forward primer of DNA2 contains vice versa a unique NdeI restriction site and a 9bp DNA sequence complementary to the 3’end of DNA1. This overlap enables the two PCR products to anneal in a subsequent PCR, where both DNA1 and DNA2 are mixed and used for template. The IgL fragment is then amplified using the forward primer of the DNA1 and the reverse primer of the DNA2. The digested and purified fragments were inserted into unique NheI/SpeI sites of $pIgL^{-,GFP2}$.

‘0-4’ series 400bp internal deletion

$pIgL^{0-4\Delta2.0-2.4,GFP2}$, $pIgL^{0-4\Delta3.2-3.6,GFP2}$

Since the two step hybrid PCR was not successful for these two fragments, a two step cloning was performed. A 5’ fragment (DNA1) and a 3’ fragment (DNA2) were PCR amplified separately and cloned into the TOPO vector. These two fragments surround the desired region to be deleted. The forward primer of DNA1 and the reverse primer of DNA2 contain restriction sites NheI and SpeI respectively for later cloning into the vector $pIgL^{-,GFP2}$. The reverse primer of DNA1 contains a unique AatII site followed by a SpeI site and the forward primer of DNA2 contains an AatII site, too. Fragments were excised from the TOPO vector using the indicated restriction sites. In a first step DNA1 was inserted into $pIgL^{-,GFP2}$ using the unique NheI/SpeI restriction sites. The resulting plasmid was digested with SpeI/AatII and DNA2 was inserted.

Multimerization of '2.2-2.4'	*$pIgL^{2.2\text{-}2.4,GFP2}$, $pIgL^{2x2.2\text{-}2.4,GFP2}$, $pIgL^{4x2.2\text{-}2.4,GFP2}$, $pIgL^{14x2.2\text{-}2.4,GFP2}$* Fragments were cloned to the TOPO vector and after excision by the *NheI/SpeI* restriction sites repeatedly cloned into *$pIgL^{-,GFP2}$*. Clones carrying the described amount of repeats were selected.
Multimerization of '2-3'	*$pIgL^{2x2\text{-}3,GFP2}$* A second '2-3' fragment was cloned into the SpeI site of *$pIgL^{2\text{-}3,GFP2}$*.
Bach2 locus	*$pBach2^{2.2\text{-}2.4,\ GFP2}$* The '2.2-2.4' fragment was incorporated into the Bach2 targeting vector (Blagodatski et al., 2009).
Homologues turkey and duck	*$pIgL^{2.2\text{-}2.4turkey,GFP2}$, $pIgL^{2.2\text{-}2.4duck,GFP2}$* The '2.2-2.4' homologue of turkey and duck were inserted into the unique NheI/SpeI sites of *$pIgL^{-,GFP2}$*.

3.2 *Molecular Biology*

3.2.1 Culture of E.coli

Luria Broth (LB) agar plates and 2YT Broth medium were used for culture of *E.coli.* The plates are incubated at 37°C over night. The medium is incubated under the same conditions while shaking.
The plates were stored at 4°C. For long term storage glycerol stocks were prepared in the following way: 3 volumes of cells + 1 volume of 50% glycerol. Stocks were frozen at -80°C deep freezer.
Inoculation was always done starting from a single clone.

3.2.2 DNA ligation

Ligation of plasmid DNA and insert DNA either purified from plasmid or PCR amplified and purified was done by using the Takara kit according to the manual. In principle, 1 volume (0.5µl) of vector (100ng/µl), 9 volume (4.5µl) of insert (100ng/µl) and 10 volume (5µl) of ligation kit nr.1 were used. The ligation mix was incubated 30 min at 16°C.

3.2.3 Transformation

Transformation of plasmid DNA and E. coli DH5α cells was done using following heat shock protocol:
The ligation mix (see DNA ligation) or the plasmid (>1ng) was added to 40-100 µl of competent cells on ice. The tube was incubated on ice for 20 min, followed by a heat shock at 42°C for 45 sec, and finally chilled on ice for 2 min. After addition of 100 µl 2YT Broth without antibiotics, the tube was incubated at 37°C for at least 15 min. Depending on the efficiency of the cloning, the mix was plated completely or in a distribution of 1/10 and 9/10 volume on each half of the LB plate containing ampicillin as selection drug. The plates were incubated overnight at 37°C. To verify the success of the transformation, 6-12 colonies were picked and prepared for miniprep according to chapter 3.2.11.

3.2.4 E.coli DH5α competent cell preparation

A streak of bacterial strain DH5α was made on a LB agar plate without antibiotics for single colony isolation. The plate was incubated over night at 37°C.
Next day a single colony was picked and cultured in 5 ml SOB Broth medium at 37°C over night. SOB Broth medium is kept at 25°C and FTB at 4°C.
The third day 1 ml of SOB broth was taken and the background was measured at OD600. 2.5 ml of overnight culture were transferred into 500 ml of SOB Broth and bacteria were cultured at 25°C. OD600 was measured every 30 min to every 1 h. When the OD600 was 0.4, which indicates that the bacteria entered the logarithmic growth phase, the bacterial culture was cooled down immediately on ice and further kept on ice for 10 min.
The culture was centrifuged at 2,000 xg at 4°C for 10 min. The supernatant was discarded and the bacterial pellet was resuspended (without vortexing) in 330 ml of ice-cold FTB.
The culture was kept on ice for 10 min, and was centrifuged at 2,000 xg at 4°C for 10 min. Supernatant was discarded and the competent cell pellet was resuspended (without vortexing) in 50 ml of ice-cold FTB. 3.5 ml of DMSO was added and the suspension mixed gently.
Aliquots of 400 µl were dispensed into sterile 1.5 ml tubes. The tubes were then frozen in liquid nitrogen and the competent cells were kept at -80°C.
The efficiency of competent cells was tested by using 1 ng of *pBluescript* plasmid for transformation to one aliquot of the competent cells (400 µl). Cells are plated on 3 LB plates: on the first 1/100, the second 1/10 and the rest of the cells on the third plate.
Next day the titer of effectively transfected competent cells was calculated. Titer of competent cells is the number of colonies which would be produced by transfection of 1 µg plasmid. Competent cells have to have a titer of at least 10^6 cells to guarantee efficient transformation.

3.2.5 PCR amplification

Amplification of genomic DNA sequences and cDNA	Aqua dest.	40,5 µl
	Buffer Pfu 10x	5 µl
	dNTP 10mM	1 µl
	Pfu Polymerase	1 µl
	Primer forward (25mM)	1 µl
	Primer reverse (25mM)	1 µl
	DNA (0.3 µg/µl)	0,5 µl

Genomic DNA was extracted from the DT40 clone AID$^{R}\psi V^{-}$ and the cDNA was plasmid DNA prepared from the riken1 cDNA library.

Amplification of Ig end deletion mutants	Aqua dest.	94,5 µl
	Expand Long Range PCR System Buffer1	15 µl
	Cresol Red 10x	15 µl
	dNTP 10mM	3 µl
	Expand Long Range PCR System Polymerase Mix	1,5 µl
	Primer forward (25mM)	3 µl
	Primer reverse (25mM)	3 µl
	DNA (1 ng/µl)	15 µl

Amplification was done using plasmid DNA *pIgL*^{W,GFP2}.

Amplification of Ig internal deletion mutants	Aqua dest.	108,9 µl
	Expand Long Range PCR System Buffer1	15 µl
	Cresol Red 10x	15 µl
	dNTP 10mM	3 µl
	Expand Long Range PCR System Polymerase Mix	1,5 µl
	Primer forward (25mM)	3 µl
	Primer reverse (25mM)	3 µl
	DNA 1	3 µl
	DNA 2	3 µl

Amplification was done using PCR amplified DNA.

PCR protocol				
	1)	93°	2 min	
	2)	93°	30 sec	2)-4) was repeated 34 times
	3)	65°	30 sec	for a total of 35 cycles
	4)	68°	5 min*	*time increases: 20 seconds each cycle
	5)	68°	7 min	
	6)	4°	∞	

3.2.6 Analysis of DNA by electrophoresis

0.8% (2% for DNA fragments smaller than 600 bp) agarose gel in 1x TAE buffer and 0.1 µg/ml EtBr were used to visualize DNA. The gel was run for 40 min at 110 V. For gel purification, time was increased up to 90 min.

3.2.7 PCR purification & Gel purification of DNA

To purify PCR amplified DNA products the Qiagen PCR Purification kit was used according to manufacturer's instructions.

After restriction enzyme digest of plasmids or PCR amplicons for DNA ligation, the digest mix was applied to a gel and the DNA excised afterwards. The DNA was purified from the gel using the Qiagen Gel Purification Kit according to manufacturer's instructions.

3.2.8 DNA Purification by phenol/chloroform extraction

Phenol/chloroform extraction was carried out to remove salts and protein and to isolate the DNA in a reaction mix for example after enzymatical digest.

1 volume of phenol/chloroform (1:1) was added to the sample. The sample was mixed and centrifuged at 15,000 xg and 4°C for 5 min. The upper aqueous phase was transferred to a new tube containing the same volume of chloroform. After mixing another centrifugation step at 15,000 xg, 4°C for 5 min followed. The aqueous phase was mixed with an equal amount of isopropanol and 0.1 volume of 3M NaOAc. After thorough mixing the sample was centrifuged at 15,000 xg and 4°C for 30 min. The pellet was washed with 70% ethanol and centrifuged at 15,000 xg and 4°C for 5 min. The pellet was air-dried for 30-60 min and dissolved in milliQ water or 1x TE.

3.2.9 Ethanol precipitation of DNA

The DNA sample was mixed with a 2.5 fold volume of 100% ethanol and kept on ice for 45 min. After centrifugation at 15,000 xg and 4°C for 30 min the pellet was washed with 250 µl of 70% ethanol and centrifuged again for 5 min. The pellet was air-dried for 30-60 min and dissolved in milliQ water or 1x TE.

3.2.10 Topo Cloning

The method was used if direct cloning of PCR amplified DNA fragments into the vector $pIgL^{-,GFP2}$ failed. Cloning of the PCR fragment into the Topo vector, excision and cloning into $pIgL^{-,GFP2}$ afterwards enhanced the cloning efficiency. Cloning was performed according to manufacturer's instructions. To enhance cloning efficiency of the insert into the Topo vector, the PCR insert was incubated prior to cloning with Taq Polymerase and dATP at 72°C for 20 min to increase the amount of A-overhangs on the insert DNA. The cloning was verified by miniprep and enzymatical digest hereafter. The inserted fragment was digested from the Topo vector by the unique restriction sites SpeI/NheI, gel purified and ligated into the vector $pIgL^{-,GFP2}$.

3.2.11 Plasmid preparation

To confirm the results of cloning, plasmid extraction was done by using the Fermentas GeneJet™ miniprep kit according to manufacturer's protocol and the DNA was used for enzyme digestion.
After enzymatical confirmation of constructs, a maxiprep was done using Qiagen Maxiprep kit according to manufacturer's protocol to gain a higher amount of plasmid DNA for transfection into the cell line.

During cloning of Ig gene PCR products for sequencing, large-scale miniprep preparation was done without using a purchased kit as follows (preparation of the solutions is described in chapter 2.2):
Bacterial colonies were picked into a 96 well U-Bottom plate each well containing 70 µl of 2YT Broth with ampicillin. 25 µl of each individual sample were transferred into two 96 well blocks containing 1.3 ml 2YT Broth per well. The blocks were incubated 20-24 h at 37°C at 220 rpm. Following day, the blocks were spin down at 2,600 xg for 10 min. Supernatant was discarded and 80 µl of Solution-I was added into each well. The pellet was resuspended by vortexing and transferred to the second block to resuspend this pellet as well. 80 µl of Solution-II was added and vortexed immediately. The samples were incubated for additional 5 min at room temperature. 80 µl Solution-III was added and vortexed, followed by centrifugation at 2,600 xg for 5 min. Millipore FB plate was placed in the bottom, and a NA lysate clearing plate was put on top of a Vacuum Manifold. 200 µl of supernatants were transferred onto the clearing plate. Vacuum was applied for 5 min until the lysate is drawn to the FB plate. The FB plate was placed on top of the empty Vacuum Manifold. 150 µl of Binding Solution was added to each well and mixed. Vacuum was applied at full strength for 2 min. 200 µl of 80% ethanol was added to each well. The vacuum was applied at full strength for 2 min. The FB plate was centrifuged at 230 xg for 30 sec (this step was repeated two times). The FB plate was put on top of a new 96-well microtiter plate and 100 µl of 1x TE Buffer was added to each well. The plate was incubated for 5 min and centrifuged at 2,800 xg for 5 min to elute the DNA. The plasmid was kept at -20°C.

3.2.12 Restriction enzyme digestion

Analytical digests were performed with amounts of plasmid DNA between 0.25 µg and 1 µg and a sample volume of 20 µl. Enzyme and buffers were used as per manufacturer's protocol.
DNA samples were incubated at 37°C for at least 3 h. Electrophoresis was done at 100-120 V for 35-45 min.
Preparative digests for cloning were performed with at least 10 µg of plasmid DNA or purified PCR product. Sample volume was 40-50 µl. The samples were run on a gel at 100-120 V for 60-90 min and gel purified afterwards.

3.2.13 Determination of DNA and RNA concentration

Purity and concentration of the DNA were checked by a spectrophotometer. The intensity of absorbance at the wavelength 260 nm was used for calculating concentration and the ratio of 260/280 nm was used to estimate the purity.

3.2.14 Genomic DNA isolation

50 ml of DT40 cells (~ 50×10^6 cells) were pelleted from culture medium at 500 xg and 4°C for 5 min. The supernatant was discarded, the pellet was washed with 1-2 ml 1x PBS and again was centrifuged at 1500 rpm and 4°C for 5 min.

The supernatant was discarded, the pellet was resuspended in 500 µl proteinase K buffer containing 0.1 mg/ml of proteinase K. The mixture was transferred into a 2 ml tube and 12.5 µl of SDS (20%) was added and mixed by inverting. Protein degradation was carried out overnight at 56°C.

The next day, 1 volume of phenol was added onto the DNA extract and mixed. The tube was centrifuged at 15,000 xg for 5 min.

The upper phase was transferred into a new tube and 1 volume of phenol/ chloroform was added and mixed. Again it was centrifuged at 15,000 xg and 4°C for 5 min to remove cell debris and proteins.

The upper phase was transferred to a new tube and 1 volume of chloroform was added and mixed. It was centrifuged again at 15,000 xg and 4°C for 5 min.

The upper phase was transferred to a new tube and 2 µl of RNAse A (100mg/ml) was added and incubated for 2 h at 37°C to digest RNA.

The DNA was transferred into a dialysis membrane and the ends were closed with clamps. The membrane was put into a 2 liter glass filled with cold 1x TE and kept at 4°C for 2 h. TE was changed at least three times with 2-4 h intervals. DNA was transferred to a new tube and stored at -20°C after the DNA concentration was measured.

3.2.15 Total RNA isolation

20 ml of DT40 cells (~ 20^*10^6 cells) were centrifuged at 500 xg and 4°C for 5 min.

The supernatant was discarded and the pellet was lysed in 1.5 ml of TRIzol reagent. After that, it was transferred into a 2 ml eppendorf tube and was incubated for 5 min at room temperature to permit the complete dissociation of nucleoprotein complexes.

0.3 ml of chloroform was added to the mix. The tube was mixed by inverting for 15 sec and incubated for 3 min at room temperature, centrifuged at 15,000 xg, 4°C for 15 min.

The upper phase was transferred to a new 2 ml eppendorf tube 750 µl of propanol was added and centrifuged at 15,000 xg, 4°C for 10 min.

RNA pellet was washed with 1 ml of RNAse free 75% ethanol, and centrifuged at 15,000 xg for 5 min. The supernatant was discarded, the sample was air dried for ~10 min and dissolved in 100 µl of DEPC water. Total RNA was stored at -80°C. The quality was checked by a 1% agarose gel after 50 min run under RNAse free conditions.

3.2.16 First strand cDNA synthesis

Super Script III (Invitrogen) was used to synthesize first strand cDNA with the included random primers according to manufacturer's intructions. The cDNA was stored at -20°C.

3.2.17 Sequencing

Sequencing reaction	The BigDye Terminator v3.1 Cycle Sequencing Kit (BDT) is used.	
	DMSO	0.5 µl
	DEPC H_2O	4.1 µl
	5x buffer BDT	1 µl
	BDT	2 µl
	Primer (0.25µM)	0.4 µl
	Template (≥100ng/µl)	2 µl

PCR protocol	1)	96°	4 min
	2)	95°	30 sec
	3)	50°	20 sec
	4)	60°	4 min
	5)	4°	∞

DNA purification

To precipitate the DNA from the sequencing mix, 2.5 µl of 125 mM EDTA were pipetted into each well of the 96-well PCR plate. A spin down at 900 xg for 2 min was performed. 30 µl of 100% ethanol were pipetted into each well and the mix was incubated for 15 min at room temperature. The plate was centrifuged at 2,000 xg and 4°C for 30 min. Afterwards the supernatant was removed by short spinning the plate bottom up at no more than 9 xg. To wash the pellet, 50 µl of 70% ethanol was pipetted into each well and again the plate was short spinned bottom up at no more than 9 xg. The pellet was dried at 70°C for 10 min. The DNA was resuspended in 30 µl LiSolv water and transferred to a sequencing plate.

3.3 *Cell Culture*

3.3.1 Basic Cell Culture Conditions

DT40 cells were cultured in culture flasks, petri dishes, or in 24 and 6 well plates.

Microtiter plates were suitable for transfection or subcloning. The optimum culture condition for the cells was 41°C with 5 % CO_2 because chicken has higher body temperature than human and mouse. Chicken medium was used to culture DT40 cells and freezing medium was used to freeze down. To keep them in healthy condition which correlates with the exponential growth phase, cell density should be between 0.5 and $1.5*10^6$ cells per ml.

3.3.2 Freeze down of cells

The cells were pelleted at 500 xg for 5 min and the pellet was resuspended with 1 ml freezing medium for 3-$10*10^6$ cells. The freezing medium contains a high amount of serum and DMSO to avoid ice crystals which destroy the cell. After storage in – 80°C for up to 1 week the cells were transferred to the nitrogen tank.

3.3.3 Thawing of cells

The tube which was taken from nitrogen tank was thawed at 41°C and centrifuged at 500 xg for 5 min. The supernatant was discarded; the pellet was resuspended with chicken medium and transferred to a flask or plate.

3.3.4 Transfection

3.3.4.1 Plasmid preparation

300 µg of plasmid DNA was linearized by an appropriate restriction enzyme within 400 µl total reaction volume over night at 37°C. Next day, completeness of the digest was checked by electrophoresis of 0.5 µl of the reaction mix. The DNA was purified by phenol/chloroform extraction or, the DNA was ethanol precipitated. Under sterile conditions, ethanol was removed and the pellet was air-dried for 20-60 min and resuspended in 300 µl sterile distilled water for a final concentration of 1 µg/µl.

3.3.4.2 Electroporation

The concentration of DT40 cells in exponential growth phase was adjusted. 10^7 cells and 40 µg DNA were used per electroporation. The calculated amount of cell suspension was transferred to a tube and centrifuged for 5 min at 500 xg, 4°C.

After removal of the supernatant, the pellet was resuspended in 800 µl of chicken medium and transferred to an electroporation cuvette together with DNA.

Electroporation was done using 25 µF and 700 V.

The mixture was transferred into 9.5 ml of chicken medium and the solution was distributed into the wells (100 µl) of a flat-bottom microtiter plate.

The following day (12-24 h after electroporation), 100 µl of selection medium (containing twice the final concentration of the selection drug) was added to each well.

The plates were left for seven to ten days in the incubator without changing the medium. Drug resistant colonies were picked for evaluation by PCR and drug check as described in the following.

3.3.4.3 Identifying targeted events by PCR

Seven to ten days after electroporation, the visible colonies were picked with a 10 µl pipette into 300 µl of chicken medium in flat-bottom 96-well plate and incubated for 3 days.

Crude extract preparation for DT40 DNA	300 µl of DT40 cell suspension was transferred to a 96 well PCR plate, washed with PBS and centrifuged 5 min at 500 xg. The supernatant was discarded, the pellet was resuspended into 10 µl of K buffer [1x PCR buffer, 0.1 mg/ml proteinase K and 0.5% Tween 20] for proteinase K-mediated proteolysis. The cells were incubated for 45 min at 56°C and to inactivate the proteinase K, the sample was incubated at 95°C for 10 min. 1 µl of the crude extract was used for PCR screening.
PCR screening	A primer located upstream of the 5' targeting arm in the genome can be used for this PCR together with a primer located within the transfected reporter cassette. For the cis-element study, primers amplifying the VJ intervening sequence are used for to control a targeting of the rearranged locus. Only the unrearranged allele still has the intervening sequence. If transfection targeted the unrearranged instead of the rearranged allele, the PCR band will disappear.

PCR reaction		
	Aqua dest.	6.3 µl
	Expand Long Range PCR System Buffer1	1 µl
	Cresol Red 10x	1 µl
	dNTP 10mM	0.2 µl
	Expand Long Range PCR System Polymerase Mix	0.07 µl
	Primer forward (25mM)	0.2 µl
	Primer reverse (25mM)	0.2 µl
	Crude extract	1 µl

PCR protocol				
	1)	93°	2 min	
	2)	93°	30 sec	2)-4) was repeated 34 times

3)	65°	30 sec	for a total of 35 cycles
4)	68°	5 min*	*time increases: 20 seconds each cycle
5)	68°	7 min	
6)	4°	∞	

The samples were run on an electrophoresis gel at 120V for 30-40min.

3.3.4.4 Identifying targeted events by drug check

Drug resistance marker genes are a feature of constructs used for transfection. Stable integration of the constructs into the genome or their deletion can be verified by culturing the cells in drug resistance marker containing media. In this condition, only cells expressing the respective drug resistance marker gene are able to survive.

The cell line $AID^{R1}IgL^{-}$ has a puromycin resistance transgene inserted at the position of the rearranged IgL locus. Transfection of the cell line with the IgL deletion fragment constructs leads to a loss of the puromycin cassette if the targeted integration of the construct was successful. Hence, in a drug test the loss of puromycin resistance indicates a positive transfectant. Additionally, AID expression in this cell line is linked to mycophenolic acid resistance. To validate expression of the AID cDNA, the mutants can be screened for mycophenolic resistant clones.

Seven to ten days after electroporation, the visible colonies were picked up into 300 µl of chicken medium in flat-bottom 96-well plates. 100 µl of the cell mix are transferred to 100 µl puromycin medium (final concentration 1 µg/ml) and another 100 µl of the cell mix are transferred to 100 µl mycophenolic acid medium (final concentration 0.5 µg/ml). Positive clones (puromycin sensitive and mycophenolic acid resistant) are selected from mycophenolic acid medium after 3 days.

3.3.5 Drug Resistance Marker Recycling

Sequential modifications of the genome by transfection are limited by the number of available drug resistance genes. This problem can be solved by the excision of the drug resistance marker after effective recombination using a site-specific recombination system such as Cre/loxP.

DT40Cre1 cell line, which is transgenic with an inducible Cre recombinase (MerCreMer) is useful for transient induction of Cre/loxP recombination. MerCreMer is active in the presence of the Mer ligand, 4-hydroxy tamoxifen, but sequestered in an inactive form by heat shock proteins in the absence of the ligand. The vectors used in this study carry floxed drug-resistance marker cassettes which can be deleted in a Cre-recombinase mediated process (Arakawa et al., 2001). Efficiency of Cre/loxP recombination depends on the distance of two loxP sites. Duration of tamoxifen induction needs to be optimized depending on the distance between two loxP sites.

DT40Cre1 is the precursor cell line of all mutants used witin this study and the system can be applied to all clones.

Protocol for recycle using induced Cre/loxP recombination:
Around 10^5 cells were cultured with fresh chicken medium containing 0.01 mM 4- hydroxytamoxifen. The cells were subcloned by limiting dilution for final concentration of 10, 3, 1 and 0.3 cells per well in 96-well flat-bottom plates filled with 200 µl chicken medium. For selective removal of the floxed cassettes, the cells were cultured in selective-drug culture medium.
Seven days after subcloning, 10 µl of the visible subclones were transferred to chicken medium as well as drug medium, in order to test the excision of the drug-resistance genes. The cells deficient for markers were selected for further studies. For example, to excise bsr cassette but to keep puro cassette, tamoxifen induced cells were cultured in puromycin containing chicken medium and blasticidin containing chicken medium. Cells sensitive for blasticidin but resistant to puromycin were selected.

3.3.6 Subcloning of DT40 cells

Subcloning by limited dilution can be used to derive clonal cell populations from a heterogeneous DT40 cell culture.
Cells in exponential growth phase were diluted to a concentration of 3, 1 and 0.3 cells per well in 200 µl chicken medium and distributed to 96 well flat bottom microtiter plates.
The plates were incubated for seven days without changing medium.
10 µl of colonies were picked from the lowest dilution showing growth into 1 ml of chicken medium in 24-well flat-bottom plates. Altogether 24 colonies were picked.
All two days, medium was changed in order to avoid over-growth of the cells.
14 days after subcloning, FACS analysis is carried out.

3.3.7 Analysis of IgM expression by FACS

IgM is presented as receptor on the surface of B cells. It is an indicator for Ig expression and can be used to quantify HM and GCV.

3.3.7.1 E2A transcription factor knockout

Dying or dead cells can cause fluorescence pseudo-positive signals because of auto-fluorescence or non-specific binding of antibodies. In order to avoid such background, the cells need to be carefully cultured to avoid overgrowth.

Predominantly surface IgM (sIgM) (-) subclones (more than 80%) need to be excluded from the analysis, since they most likely originated from cells that were already sIgM(-) at the time of subcloning. In order to quantify the amounts of sIgM expression as well as sIgM, PE-antibody staining for FACS analysis was done.
300 µl of cells were transferred into round-bottom 96-well plates and centrifuged at 500 xg and 4°C for 5 min. The medium was discarded and the cell pellet washed with 200 µl PBS per well. After another centrifugation step at 500 xg and 4°C for 5 min the PBS was discarded. For primary antibody (anti-IgM) mouse staining, the cells were incubated with anti-chicken Cµ monoclonal antibody-M1 (dilution 1/500 in FACS staining buffer, 30 µl per sample) on ice for 30 min. Afterwards 200 µl PBS per well was added and the plate centrifuged at 500 xg and 4°C for 5 min. The supernatant was discarded. For secondary antibody staining, the cells were incubated with R-PE-conjugated goat anti-mouse IgG polyclonal antibody (dilution 1/100 in FACS staining buffer, 30µl per sample) on ice and in the dark for 30 min. Afterwards 200 µl PBS per well was added and the plate centrifuged at 500 xg and 4°C for 5 min. The supernatant was discarded and each sample was resuspended in 250 µl PBS.
The sIgM expression of cell lines was measured and analyzed by flow cytometer. Cells were gated, to separate dead and living cells, whereas at least 5,000 events have to fall into the living gate. To separate stable GFP expressing cells from cells with decreased or lost GFP expression, and to separate PE-high from PE-low stained cells a gate was set to divide the dot plot in four parts according to the four possible clouds: 1) GFP-high/PE-low, 2) GFP-high/PE-high, 3) GFP-low/PE-low and 4) GFP-low/PE-high.The gate was set according to the clouds of the control cell lines $AID^{-/-}\psi V^{-}$ (GFP-low/PE-low, because it does not express GFP-coupled AID cDNA and has no surface IgM) and $AID^{R}\psi V^{-}$ (GFP-high/PE-high and GFP-high/PE-low, because it expresses GFP-coupled AID cDNA and has predominantly surface IgM-positive cells with a surface IgM-negative subpopulation due to HM).

3.3.7.2 Cis-element study

Predominantly GFP(-) subclones (more than 50%) need to be excluded from the analysis, since they most likely originated from cells that were already GFP(-) at the time of subcloning.
For GFP expression analysis, aliquots of 300 µl – 1ml cells depending on their density were centrifuged at 500 xg and 4°C for 5 min. The medium was discarded and the cell pellet washed with 200 µl PBS per well. After another centrifugation step at 500 xg and 4°C for 5 min the cell pellet is resuspended in 250 µl PBS.
GFP expression of cell lines were measured and analyzed by flow cytometer. Cells were gated, to separate dead and living cells, whereas at least 5000 events have to fall into the living gate. To separate stable GFP expressing cells from cells with decreased or lost GFP expression, a gate was set directly underneath the main cloud.

3.4 Software for statistics

R is freely available software for statistical computing and graphics. R was used to perform a test of significance according to the Wilcoxon U-Test.

Concerning the deletion studies of the cis-elements, I had to take care if there is a difference between the precursor clone of a series and the single deletion mutants generated based on the precursor.

The hypothesis is that GFP green fluorescence is decreased, because a *'HyCorE'* cis-element helps to introduce AID-induced mutations. The assumption is always, that there exists a significant difference between the precursor clone and the mutant.

For example, the *'0-4'* series has 1 kb serial end deletions. The deletion mutants of this series contain a *'HyCorE'* cis-element, if their GFP expression profile is not significantly different from the one of the *'0-4'* mutant, means the GFP decrease is similar in the mutant and the precursor clone. Therefore, the hypothesis has to be rejected.

The deletion mutants of the *'0-4'* series 400 bp internal deletions contain a *'HyCorE'* cis-element, if the GFP expression profile is significantly different from the one of the *'0-4'* clone, means the GFP decrease is significantly lower in the mutant compared to the precursor clone. Therefore, the hypothesis has to be accepted.

To verify the hypothesis, the GFP expression of the 24 subclones of the two clones for the precursor cell line ($\psi V^- IgL^{0\text{-}4,GFP2}$ Clone 1 and 2) was compared with the GFP expression of the 24 subclones of the two clones for each mutant cell line ($\psi V^- IgL^{0\text{-}4\Delta 0.2\text{-}0.4GFP2}$ Clone 1 and 2, $\psi V^- IgL^{0\text{-}4\Delta 0.4\text{-}0.8GFP2}$ Clone 1 and 2, and so on). Because this is a test taking random samples, it is possible to make an incorrect decision. That means the hypothesis is verified although it is wrong. To minimize this error and to get highly significant values, the p-value was set at $p<0.0001$. The p-value is the statistical probability for the hypothesis to be true, means for the significant difference between the precursor clone and the mutant. Only clones showing this p-value were considered to contain relevant cis-elements.

4 Results

4.1 Inactivation of the E2A gene

Based on my diploma thesis 'The Role of Bach-2 and E2A for Immunoglobulin Gene Transcription and Repertoire Development' (Schoetz, 2005), I found the E2A gene having a relevant influence on HM activity in DT40 and I conducted the work for my PhD thesis and studied in more detail at which point E2A is engaged in the diversification process.

4.1.1 Complementation of the knockout using E2A cDNA expression cassettes

The knockout was performed by targeted integration of the knockout construct into the E2A locus via arms sequences, which are identical with parts of the E2A genomic locus. The DT40 cell line has a high ratio of targeted to random integration (Buerstedde et al, 1991), but it can not be ruled out completely, that additional to the targeted integration also a random integration took place in the same cell. Then, effects on hypermutation could be due to the random integration of the knockout construct, for example if a for the diversification process important factor was destroyed by the integration. To rule out this possibility, E2A transcription factors can be overexpressed as a cDNA expression cassette. The overexpression should resconstitute the phenotype.

E2A cDNA clones for the two splice variants E12 and E47 were obtained by running a Blast Search (http://blast.ncbi.nlm.nih.gov/) using E12 mRNA or E47 mRNA sequences respectively (Conlon et al., 2004). The ESTs of the riken1 cDNA clones 8e2, 13d10, 24c5 and 34l23 (Caldwell et al., 2005) exhibit high similarity (Table 4.1) to the E-protein cDNAs and each of them includes the start codon within the first EST. Alignment of the riken1 8e2 ESTs displayed identity of the clone with the E47 mRNA. However, for other clones only the sequence for the first EST (primer r1) was available. Hence, I performed sequencing analysis on all clones and found 34l23 to be a cDNA clone of E12 mRNA (Figure 4.1.).

	Sequences producing significant alignments:	Values for Blast with E12 / E47 mRNA				
Accession	**Description**	**Max score**	**Total score**	**Query coverage**	**E value**	**Max ident**
AJ456337.1	AJ456337 riken1 Gallus gallus cDNA clone 8e2r1, mRNA sequence	913	913	20%	0.0	99%
AJ456338.1	AJ456338 riken1 Gallus gallus cDNA clone 8e2r2, mRNA sequence	713	713	16%	0.0	98%
AJ456339.1	AJ456339 riken1 Gallus gallus cDNA clone 8e2r3, mRNA sequence	892	892	19%	0.0	100%
AJ456340.1	AJ456340 riken1 Gallus gallus cDNA clone 8e2r4, mRNA sequence	1200 / 1299	1200 / 1299	27% / 29%	0.0	99%
AJ456341.1	AJ456341 riken1 Gallus gallus cDNA clone 8e2r5, mRNA sequence	985 / 1278	985 / 1278	30% / 29%	0.0	89% / 98%
AJ456342.1	AJ456342 riken1 Gallus gallus cDNA clone 8e2r6, mRNA sequence	239	239	6%	9e-61	93%
AJ735483.1	AJ735483 riken1 Gallus gallus cDNA clone 8e2s9, mRNA sequence	847	847	20%	0.0	97%
AJ735475.1	AJ735475 riken1 Gallus gallus cDNA clone 8e2s10, mRNA sequence	686 / 978	686 / 978	23%	0.0	97% / 97%
AJ735476.1	AJ735476 riken1 Gallus gallus cDNA clone 8e2s11, mRNA sequence	953	953	21%	0.0	99%
AJ735477.1	AJ735477 riken1 Gallus gallus cDNA clone 8e2s12, mRNA sequence	1371	1371	31%	0.0	99%
AJ446292.1	AJ446292 riken1 Gallus gallus cDNA clone 13d10r1, mRNA sequence	1072	1072	24%	0.0	99%
AJ450099.1	AJ450099 riken1 Gallus gallus cDNA clone 24c5r1, mRNA sequence	908	908	20%	0.0	99%
AJ453527.1	AJ453527 riken1 Gallus gallus cDNA clone 34l23r1, mRNA sequence	1018	1018	22%	0.0	99%

Table 4.1. Results for a Blast search using expressed sequence tags (ESTs) database of gallus gallus (Taxid: 9031) and E12 (AJ579995) or E47 (AJ579996) mRNA respectively. The riken clone 8e2 is already sequenced completely and corresponds to E47. For the clones 13d10, 24c5 and 34l23, only the first primer run was sequenced. These clones were chosen for sequencing by primer walk using the riken1 primers.

The riken1 clone 8e2 had additional amino acids inserted at position 75 (glutamine) and 173 (asparagine), a deletion at 544 (serine) and an amino acid change from methionine to valine at 443 (Figure 4.1.A). The riken1 clone 34l23 has an amino acid deletion at position 173 (asparagine; Figure 4.1.B). These changes do not affect important functional domains and might reflect polymorphisms between various chicken breeds used for the sequencing experiments.

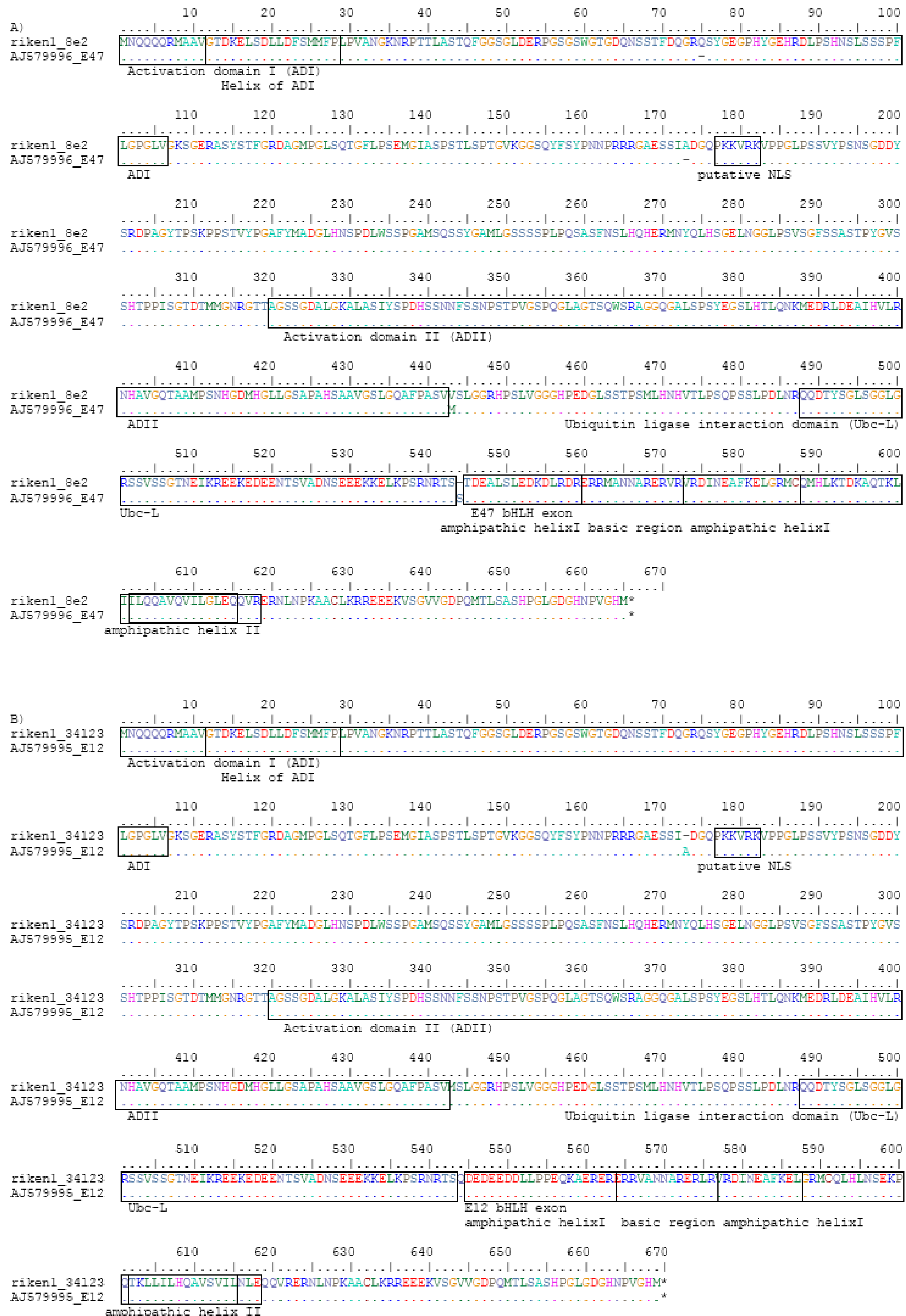


Figure 4.1 Alignment of the sequenced riken1 clones reflecting E47 and E12 cDNA.
'.' denotes identical residues. '-' gaps and '*' the stop codon. Important functional regions are described in black boxes (Aronheim et al., 1993; Kho et al., 1997; Quong et al., 1993). A) Alignment of the riken1 clone 8e2 and E47 mRNA. B) Alignment of the riken1 clone 34l23 and E12 mRNA.

The full-length cDNA inserts of the pKS plasmid-derived cDNA library riken1 are flanked by SfiI restriction sites. These sites were used to extract and insert the cDNAs into the vectors *pIRESSfBsr* and *pIRESSfGpt*. Both vectors are suitable for the cloning of riken1 cDNA libraries, as they contain an expression system created by a β-actin promoter which is followed by a SfiI site, IRES, drug resistance marker gene and polyA site. The SfiI site is used to clone in the E12 or E47 cDNA respectively and the IRES connects expression to the gpt or bsr marker cassettes, respectively. Transcription is terminated by the artificial SV40 polyA signal. The whole expression cassette is flanked by loxP sites for easy removal (Figure 4.2.A). For transfection the constructs were linearized with ScaI and purified by phenol/chloroform extraction.

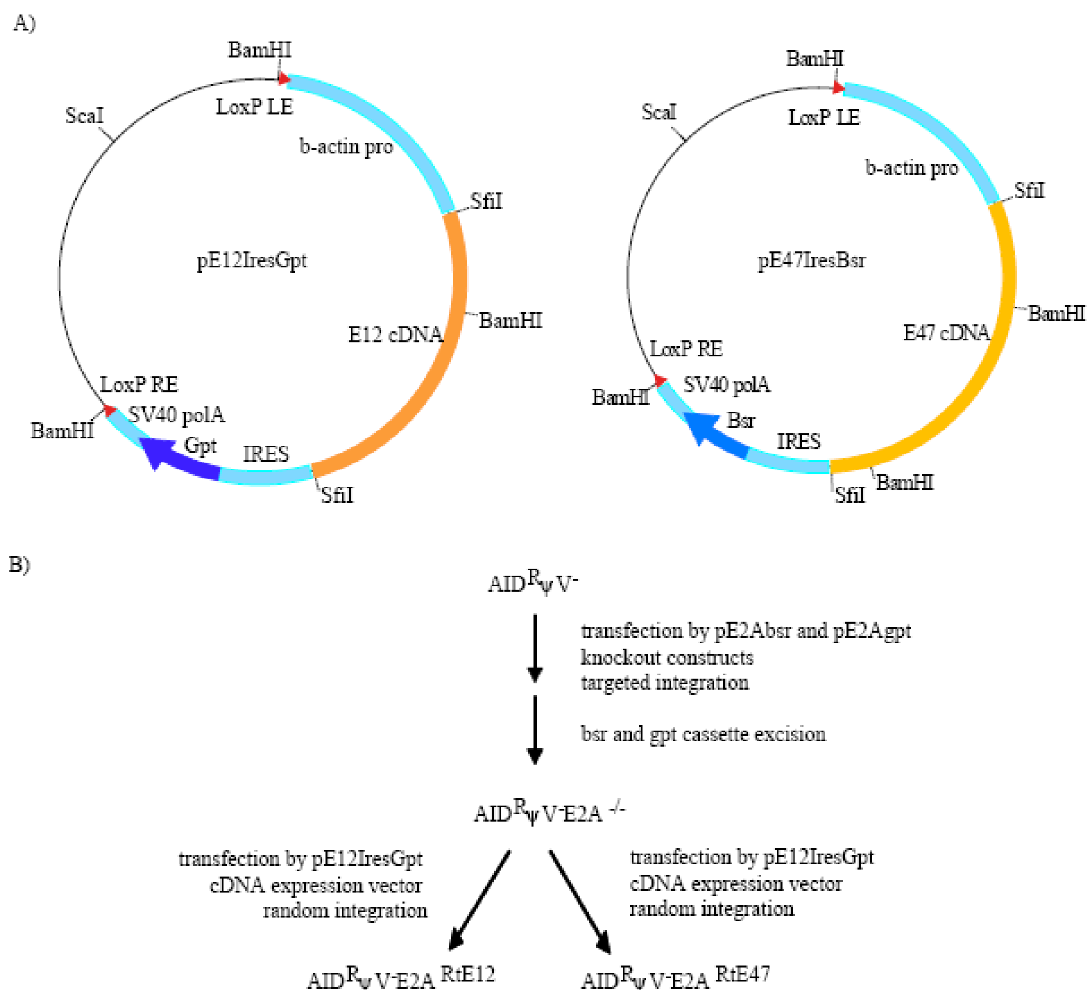


Figure 4.2 E2A complementation
A) Chart of the two cDNA expression vectors *pE12IresGpt* and *pE47IresBsr*. Important transcriptional regulatory regions like the β-actin promoter, IRES and SV40 polyA signal are annotated with light blue bars, E12 and E47 cDNA expression cassettes are orange and the gene for the selection markers are dark blue arrows. LoxP sites are red triangles. Restriction sites used for cloning and linearization are indicated. B) Flow chart showing the generation of the E2A knockout and recomplemented clones. The precursor cell line AID$^{R}\psi$V^{-} was transfected two times to generate the clone AID$^{R}\psi$V^{-}E2A$^{-/-}$. After removal of the drug resistance genes, the cDNA expression cassette E12 and E47 were transfected respectively yielding the clones AID$^{R}\psi$V^{-}E2A^{RtE12} and AID$^{R}\psi$V^{-}E2A^{RtE47}.

For reconstitution the cell line AID$^{R}\psi$V^{-}E2A$^{-/-}$ was transfected randomly either by *pE12IresGpt* or *pE47IresBsr* yielding the clones AID$^{R}\psi$V^{-}E2A^{RtE12} and AID$^{R}\psi$V^{-}E2A^{RtE47} (Figure 4.2.B).

4.1.2 mRNA levels of E2A do not correlate with IgL and AID transcription

The status of E2A expression in the wild-type, E2A mutant and E2A complemented clones was confirmed by semi-quantitative RT-PCR using primers specific for the exons coding the splice variants E12 (E2A9/E2A10) and E47 (E2A9/E2A11) respectively. As an internal standard the housekeeping gene elongation factor 1α (EF1α) was amplified (List of primers, chapter 8.1). EF1α showed similar expression levels throughout all analyzed cell lines (Figure 4.3).

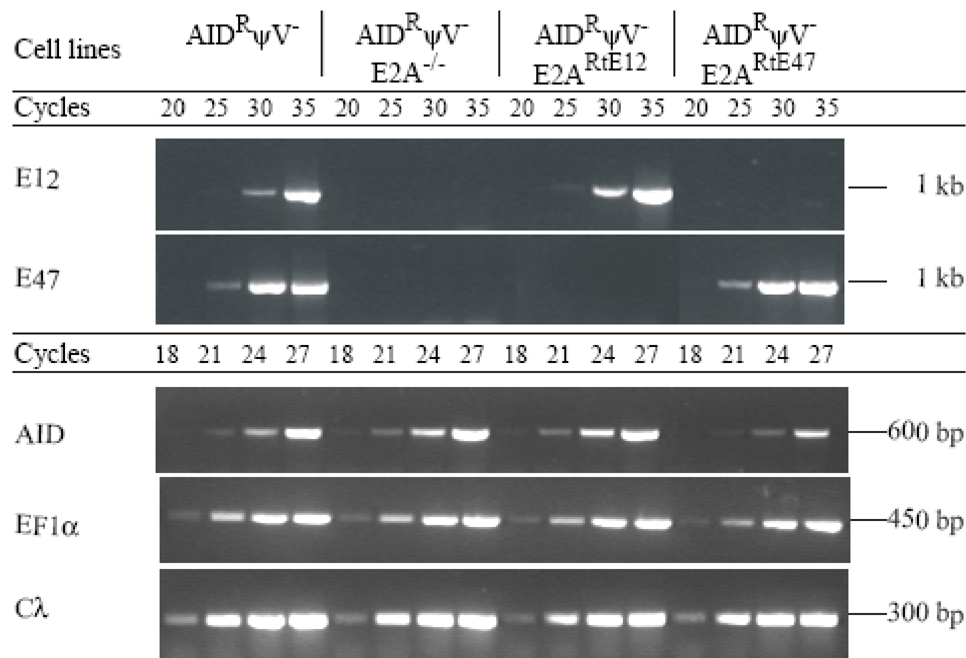


Figure 4.3 E12, E47, AID and Cλ mRNA levels measured by semi-quantitative RT-PCR.
The housekeeping gene EF1α was amplified as an internal standard. Analysis was done for the progenitor cell line AID$^{R}\psi$V^{-}, the homozygous knockout AID$^{R}\psi$V^{-}E2A$^{-/-}$ and the complemented clones AID$^{R}\psi$V^{-}E2A^{RtE12} and AID$^{R}\psi$V^{-}E2A^{RtE47}. The number of PCR cycles is indicated.

The progenitor cell line AID$^{R}\psi$V^{-} expressed slightly more E47 than E12 mRNA. The E2A knockout cell line was negative for E12 and E47 mRNA and the E12 and E47 cDNA rescued cell lines were positive for E12 or E47 mRNA, respectively. Consequently, E2A transcription factors were knocked out in the cell line AID$^{R}\psi$V^{-}E2A$^{-/-}$ and E12 and E47 reconstitution was successful.

E2A proteins are B cell transcription factors which may potentially act on the IgL locus. As HM is suspected to be strongly connected to transcription (Bachl et al., 2001), the deserved effect for E2A could be due to

changing transcription level of the IgL chain locus instead of directly influencing diversification of the Ig gene. Amplification of Ig light chain transcripts using primers for the C region (Table S.1, Figure 4.3 Cλ) established that the lack of E2A expression per se did not alter the steady-state level of the Ig light chain mRNA, suggesting that E2A proteins are not required for Ig light chain transcription in the analyzed DT40 clones.

RT-PCR using AID cDNA specific primers demonstrated furthermore that the clones AID$^{R}\psi$V^{-}, AID$^{R}\psi$VE2A$^{-/-}$ and AID$^{R}\psi$V^{-}E2A^{RtE12} expressed roughly equivalent amounts of AID mRNA, whereas equal levels of the clone AID$^{R}\psi$V^{-}E2A^{RtE47} seems to support reduced levels of AID mRNA (Figure 4.3 AID). AID is expressed as an artificial cDNA under the control of a β-actin promoter in the AID$^{R}\psi$V^{-} progenitor clone and all mutant clones derived from this clone, making it unlikely but not impossible that its expression levels are influenced by E2A. The promoter has an enhancer-like sequence (Fregien and Davidson, 1986) and E2A transcription factors could bind to this enhancer. On the other hand, AID cDNA is transfected randomly and it is not known how many copies have integrated into the genome. Possibly this cell line lost one of the copies and therefore expresses less AID. However, subsequent experiments showed a high HM rate for the AID$^{R}\psi$V^{-}E2A^{RtE47} clone so that the slight reduction of AID is not a concern for our purposes (Figures 4.3 and 4.4, see the following paragraphs and discussion).

AID expression was present in all clones (Figure 4.3) and thought to be strong enough to keep HM activity at a high level. mRNA data suggest the reduced HM in the AID$^{R}\psi$V^{-}E2A$^{-/-}$ mutant to have its origin in the loss of E2A expression which needs to be examined in more detail.

4.1.3 Ig mutation assay - E2A negative clones display strongly decreased rates of sIg loss

To examine the Ig HM rate in the E2A cell mutants, deleterious Ig light chain mutations were analyzed by measuring the appearance of sIg(-) populations in the predominantly sIgM(+) clones (Arakawa et al., 2004). 14 days after subcloning at least 24 subclones per cell line of the clones AID$^{-/-}\psi$V^{-}, AID$^{R}\psi$V^{-}, AID$^{R}\psi$V^{-}E2A$^{+/-}$, AID$^{R}\psi$V^{-}E2A$^{-/-}$, AID$^{R}\psi$V^{-}E2A^{RtE12} and AID$^{R}\psi$V^{-}E2A^{RtE47} were analyzed by FACS (Figure 4.6.A and B). The average percentages of sIg(-) cells for the mutant and control cell lines were calculated as mean values of all clonally derived subclones to account for fluctuation effects (Table 4.2).

FACS analysis after antibody staining indicated that the majority of cells in all isolated clones remained sIg(+) and that differences in E2A expression did not measurably influence the sIg expression levels of sIg(+) cells (Figure 4.4.A).

A low number of cells and a few subclones had apparently lost AID expression as indicated by the loss of the coupled GFP expression. This reflects most likely the excision of the floxed AID-IRES-GFP transgene cassette

by leaky Cre recombinase expression. Because a lack of AID expression stops Ig HM (Arakawa et al., 2002), GFP-negative cells were excluded from the analysis.

AID$^{-/-}\psi$V^{-}		AID$^{R}\psi$V^{-}		AID$^{R}\psi$V^{-}E2A$^{+/-}$	AID$^{R}\psi$V^{-}E2A$^{-/-}$		AID$^{R}\psi$V^{-}E2A^{RtE12}	AID$^{R}\psi$V^{-}E2A^{RtE47}
0.1 %	0.2 %	32.5 %	18.9 %	3.9 %	3.5 %	7.1 %	49.8 %	9.7 %
0.1 %	1.1 %	88.3 %	47.6 %	3.8 %	5.0 %	6.4 %	27.5 %	16.2 %
0.1 %	0.2 %	25.6 %	5.2 %	10.2 %	4.0 %	8.5 %	22.2 %	18.0 %
0 %	0.2 %	66.3 %	22.1 %	2.9 %	3.9 %	9.2 %	28.6 %	35.8 %
0.1 %	0.1 %	16.3 %	15.7 %	3.3 %	2.3 %	7.6 %	15.2 %	33.5 %
0.3 %	0.1 %	9.4 %	4.9 %	5.6 %	3.2 %	5.0 %	39.0 %	36.4 %
2.3 %	0.1 %	11.5 %	5.4 %	6.9 %	2.9 %	8.2 %	39.7 %	31.4 %
0.2 %	0.1 %	7.9 %	17.6 %	2.6 %	4.7 %	13.6 %	15.9 %	17.9 %
0.6 %	0.2 %	6.4 %	21.2 %	57 %	2.7 %	5.7 %	42.7 %	79.9 %
0.2 %	0.2 %	3.5 %	16.7 %	10.3 %	2.5 %	8.1 %	43.2 %	19.7 %
0.1 %	0.3 %	52.1 %	14.0 %	5.5 %	12.8 %	6.7 %	16.5 %	48.1 %
0.1 %	0.2 %	6.8 %	21.2 %	7.0 %	1.6 %	4.1 %	20.3 %	50.2 %
0 %	0.5 %	5.2 %	7.3 %	3.4 %	2.1 %	7.5 %	28.0 %	10.2 %
0.5 %	0.3 %	8.6 %	5.7 %	3.0 %	2.5 %	12.9 %	15.9 %	47.8 %
0.1 %	0.2 %	29.3 %	30.2 %	4.9 %	1.8 %	3.1 %	52.0 %	18.0 %
0.1 %	0.2 %	4.4 %	44.5 %	4.6 %	1.4 %	12.4 %	53.5 %	15.5 %
0.5 %	0.5 %	39.9 %	12.0 %		0.9 %	9.6 %	21.7 %	14.0 %
0.9 %		26.0 %	4.6 %		0.2 %	10.1 %	43.1 %	19.2 %
0.2 %		11.2 %	9.2 %		1.0 %	8.2 %	24.0 %	12.0 %
0.3 %		5.6 %	41.8 %		1.1 %	6.0 %	14.2 %	25.8 %
0.5 %		21.6 %			1.8 %	5.0 %	44.1 %	19.3 %
0.6 %		16.7 %			0.8 %	15.9 %	48.7 %	44.4 %
0.4 %		12.8 %			1.2 %	5.5 %	58.1 %	12.9 %
0.2 %		9.9 %			8.9 %	4.4 %	16.6 %	81.0 %

Table 4.2 Percentages of events falling into sIgM(-) gates for individual subclones of the precursor clone AID$^{R}\psi$V^{-}, the E2A knockout clones AID$^{R}\psi$V^{-}E2A$^{+/-}$ and AID$^{R}\psi$V^{-}E2A$^{-/-}$, the complemented clones AID$^{R}\psi$V^{-}E2A^{RtE12} and AID$^{R}\psi$V^{-}E2A^{RtE47} and the AID-negative control clone AID$^{-/-}\psi$V^{-}. This sIgM(-) subpopulation of cells is a indicator for HM activity.

The mutation assay revealed a significant reduction in the percentages of sIgM(-) cells for the heterozygous E2A mutant AID$^{R}\psi$V^{-}E2A$^{+/-}$ (8.1%) and the homozygous E2A mutant AID$^{R}\psi$V^{-}E2A$^{-/-}$ (5.5%) compared to the wild-type AID$^{R}\psi$V^{-} (19.7%). The complemented E2A mutant clones AID$^{R}\psi$V^{-}E2A^{RtE12} (32.5%) and AID$^{R}\psi$V^{-}E2A^{RtE47} (29.9 %) could restore the phenotype beyond the level of the progenitor cell line (19.7%; Figure 4.4.B). Although the cell line AID$^{R}\psi$V^{-}E2A^{RtE47} expressed reduced levels of AID mRNA (Figure 4.2), it was hypermutating at a high rate. This suggests, although AID cDNA expression might be influenced somehow by E47, the activity of AID is present. There might be a saturation level for AID in AID-mediated HM which is reached already in the AID$^{R}\psi$V^{-}E2A^{RtE47} clone, so there is still plenty of AID to induce strong HM.

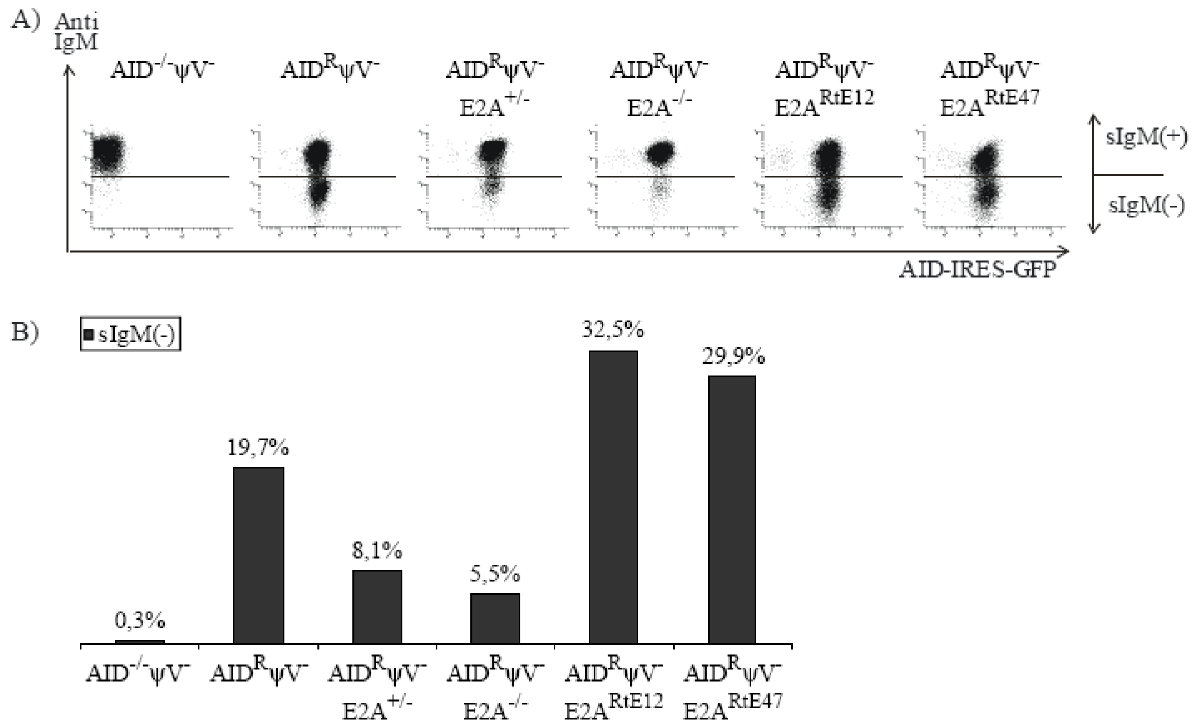


Figure 4.4 sIgM expression analysis of E2A$^{-/-}$ and control clones.
A) FACS anti-IgM staining profiles of representative subclones derived from initially sIgM(+) clones. B) Average percentages of events falling into sIgM(-) gates based on the measurement of at least 24 subclones of the cell lines AID$^{-/-}$ψV$^-$, AID^{R}ψV$^-$, AID^{R}ψV$^-$ E2A$^{+/-}$, AID^{R}ψV$^-$E2A$^{-/-}$, AID^{R}ψV$^-$E2A^{RtE12} and AID^{R}ψV$^-$E2A^{RtE47}.

AID$^{-/-}$ψV$^-$ which has stopped all Ig hypermutation after the deletion of the AID cDNA expression cassette was used as a control for a stable sIgM(+) clone. Only very few events (0.3%) accumulated in the sIgM(-) gate of this clone (Fig.4.4.B).

The result suggests a direct involvement of E2A transcription factors in diversification processes at the IgL chain locus in the DT40 cell line.

4.1.4 Mutation pattern – Ig HM is reduced in the absence of E2A

For sequencing, the cell lines AID$^{-/-}$ψV$^-$, AID^{R}ψV$^-$E2A$^{-/-}$, AID^{R}ψV$^-$E2A^{E47} and AID^{R}ψV$^-$E2A^{E12} were subcloned, for each cell line one subclone derived from a single cell was chosen and cultured for 6 weeks. Then, crude extract of DNA was prepared and the rearranged VJ region of the IgL chain locus, the preferential HM target, was amplified by PCR using primer pair VL1/VL2 (List of primers, chapter 8.1) and cloned into pUC19 vector after digestion with HindIII and XbaI. At least 96 bacterial clones for each of the cell lines were prepared for sequencing using the primer VL3. The mutation pattern was compared to the sequence in the AID^{R}ψV$^-$ (Saribasak et al., 2006) progenitor clone. The consensus sequence of all sequences from each subclone was taken as the likely sequence of the precursor cell of the subclone. Differences in the subclone sequences in comparison to the precursor cell sequence are regarded as mutations.

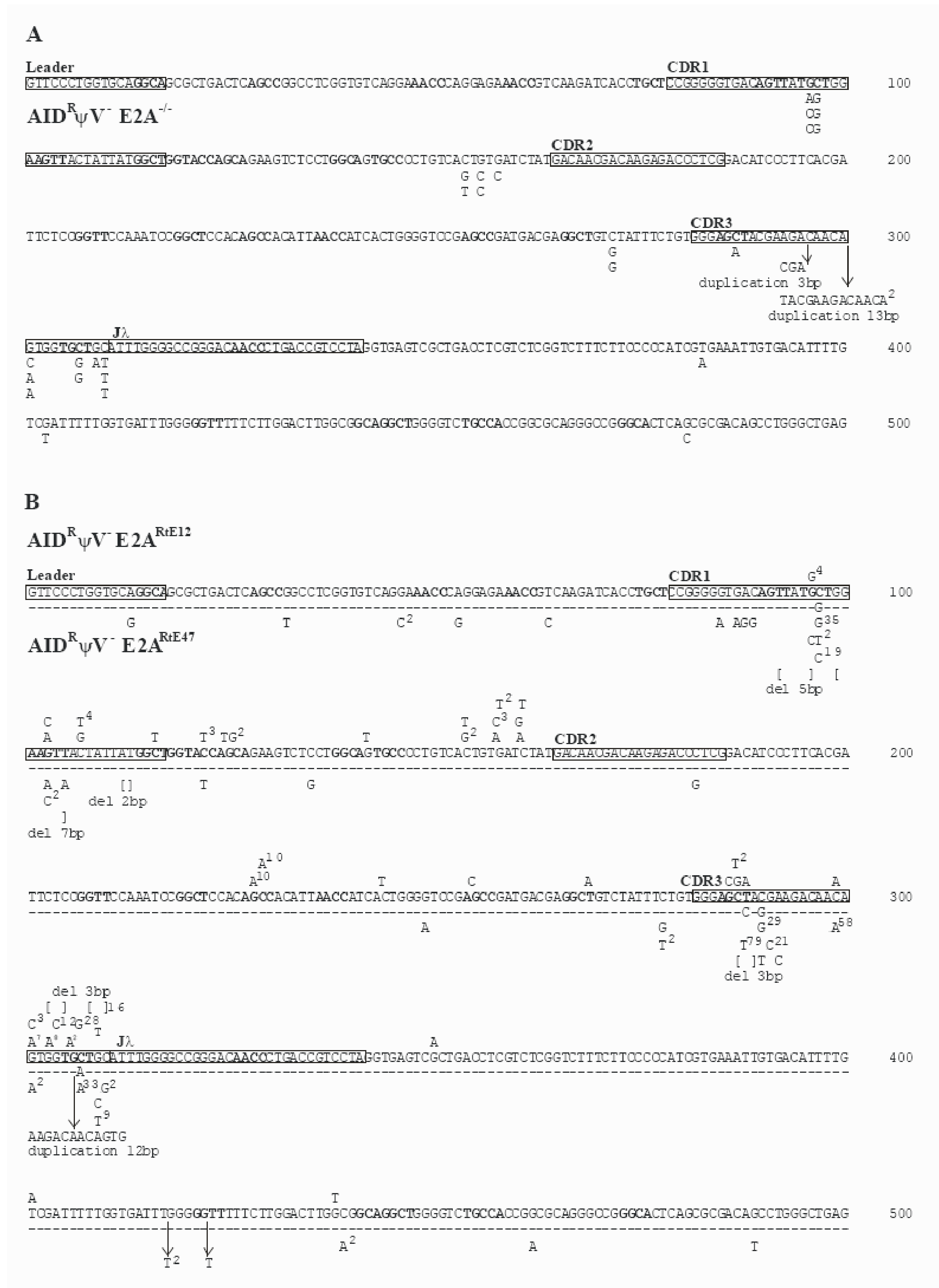


Figure 4.5 Comparison of mutations from AID^{R}ψV^{-}E2A$^{-/-}$, AID^{R}ψV^{-}E2A^{RtE12}, AID^{R}ψV^{-}E2A^{RtE47} and AID^{R}ψV^{-} cells.
Mutations within the rearranged VJ light chain segments 6 weeks after subcloning. The mutations of AID^{R}ψV^{-} cells have been described previously (Saribasak et al., 2006). A) Single nucleotide substitutions identified in the AID^{R}ψV^{-}E2A$^{-/-}$ cells are mapped onto the likely sequence of the precursor cell for the subclone. Occasional deletions and insertions are indicated. Hot spot motifs (RGYW and its complement WRCY) are highlighted in bold letters. B) Single nucleotide substitutions identified in the AID^{R}ψV^{-}E2A^{RtE12} and AID^{R}ψV^{-}E2A^{RtE47} cells are shown above and below the likely sequence of the precursor cell for the subclone, respectively. Differences

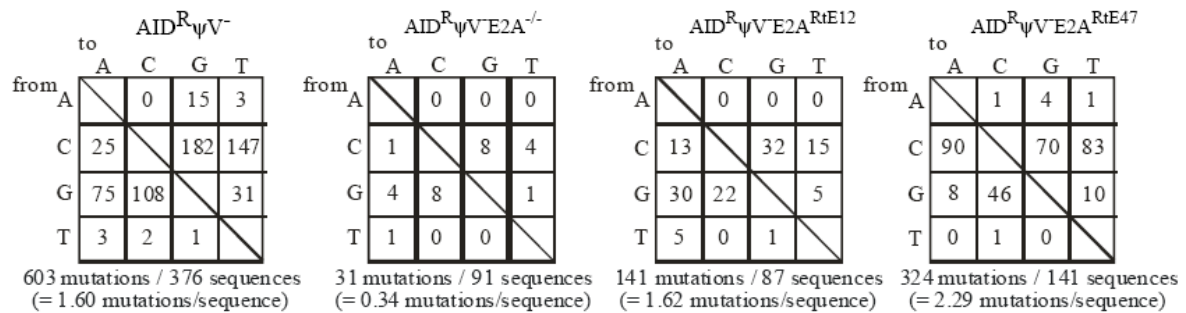
between the progenitor sequences of $AID^R\psi V^-E2A^{RtE12}$, $AID^R\psi V^-E2A^{RtE47}$ are indicated in the line below the sequence. Occasional deletions and insertions are indicated. Hot spot motifs (RGYW and its complement WRCY) are highlighted in bold letters.

Mutations cluster primarily in the three V subregions referred to as complementarity-determining regions (CDR1, CDR2 and CDR3; Tonegawa et al., 1983). For E2A knockout and E12/E47 overexpressed clones, mutations are increased in CDR1 and CDR3 and slightly in front of CDR2 (Figure 4.5.A and B). 55% of all mutations occur at RGYW (R for purines A/G, Y for pyrimidines C/T, and W for A or T) motifs and its complement WRCY, known as hotspots of Ig HM in mouse and chicken (Michael et al., 2002; Arakawa et al., 2004).

The $AID^R\psi V^-E2A^{-/-}$ clone with 0.34 mutations/sequence accumulated mutations at an about five fold reduced rate compared to the $AID^R\psi V^-$ progenitor clone with 1.6 mutations/sequence (Figure 4.6).
The decrease in mutation confirms the results of the Ig mutation assay, where the sIgM(-) population, which reflects the HM activity of the clone, was shown to be reduced from 19.7% in the progenitor clone to 5.5% in the $AID^R\psi V^-E2A^{-/-}$ clone. The distribution of the mutations and the mutation spectrum of the $AID^R\psi V^-E2A^{-/-}$ clone were very similar to previous reports for the $AID^R\psi V^-$ clone (Arakawa et al., 2004). C-to-T and C-to-G are the most frequent mutations in the progenitor clone, followed by G-to-C and G-to-A. The mutation pattern was kept in the E2A double knockout clone.
In addition, VJ regions of the complemented clones $AID^R\psi V^-E2A^{RtE12}$ and $AID^R\psi V^-E2A^{RtE47}$ were sequenced. The clones accumulated 1.62 mutations/sequence and 2.29 mutations/sequence, which is again consistent with the FACS data, where the clones had accumulated sIgM(-) populations at a rate similar or higher (29.9% and 32.5%) than that observed for the $AID^R\psi V^-$ clone (19.7%; Figure 4.4.B). The analysis of the mutation spectrum of the complemented clones showed some deviations from the mutation spectrum of the $AID^R\psi V^-$ progenitor clone. In the $AID^R\psi V^-E2A^{RtE12}$ clone mutations from C-to-A were increased and C-to-T mutations were approximately two fold decreased. The $AID^R\psi V^-E2A^{RtE47}$ clone had a relatively high frequency of C-to-A mutations (nine fold increase compared to the progenitor clone). G-to-A mutations within this clone were two fold decreased. However, the observed differences were not consistent between the two complemented clones and within different subclones of the same clone indicating that they reflected most likely fluctuation and different timing of mutations during the expansion of the sequenced subclones.

To knock out a factor downstream of AID, like UNG, will change the mutation pattern at the IgL chain locus in the cell (Saribasak et al., 2006). The AID-generated intermediate uracil will be further processed by UNG resulting in an abasic site. Inactivation of the UNG gene blocks the processing of uracils and leads exclusively to transition mutations (C-to-T and G-to-A) when uracil pairs with adenine in the next replication cycle (see

also Figure 1.4). Knockout of the E2A gene and overexpression of the E12 and E47 cDNAs did not lead in a similar perturbation of the mutation pattern.



AIDRψV- from \ to	A	C	G	T
A		0	15	3
C	25		182	147
G	75	108		31
T	3	2	1	

AIDRψV-E2A-/- from \ to	A	C	G	T
A		0	0	0
C	1		8	4
G	4	8		1
T	1	0	0	

AIDRψV-E2ARtE12 from \ to	A	C	G	T
A		0	0	0
C	13		32	15
G	30	22		5
T	5	0	1	

AIDRψV-E2ARtE47 from \ to	A	C	G	T
A		1	4	1
C	90		70	83
G	8	46		10
T	0	1	0	

Figure 4.6 Comparison of mutations from AID^{R}ψV^{-}, AID^{R}ψV^{-}E2A$^{-/-}$, AID^{R}ψV^{-}E2A^{RtE12}, and AID^{R}ψV^{-}E2A^{RtE47} cells.
Mutation pattern within the rearranged VJ segment of the IgL locus 6 weeks after subcloning. The mutations of AIDRψV^{-} cells have been described previously (Saribasak et al., 2006). Only single nucleotide substitutions are included. Mutation frequency is calculated as per sequence.

4.2 *Defining cis-elements*

E12 and E47 are known to regulate transcription by binding as homo- or heterodimers to their DNA-binding motifs known as E-boxes. The E-box is a 6 bp motif CANNTG, where N can be any base, with a highly conserved core of 4 bp flanking the NN. The motif is spread all over the rearranged IgL chain locus (Figure 4.7.A), although most of them cluster around the previously described enhancer region (Bulfone-Paus et al., 1995). The most frequent motif is CAGCTG (Figure 4.7.B). E-box motifs in the enhancer sequences of Ig genes are known to influence Ig gene recombination (Henthorn et al., 1990). To find out the exact sequence and exact position of the E-box motif involved in Ig HM needs a mapping of the complete IgL locus with a focus on the enhancer.

Thinking about a transcriptional enhancer, many cis-elements and DNA binding factors are involved in triggering transcription. In parallel, also a true *'HyCorE'* could involve a cluster of cis- and trans-acting factors to recruit the AID-mediated HM specifically to the Ig loci.

The work described in the following chapters is focused on this issue. All experiments were done in collaboration with Sabine Schmidl.

For a description of the nomenclature of the used plasmids and generated cell lines see chapter 2.4 and 3.1.3.

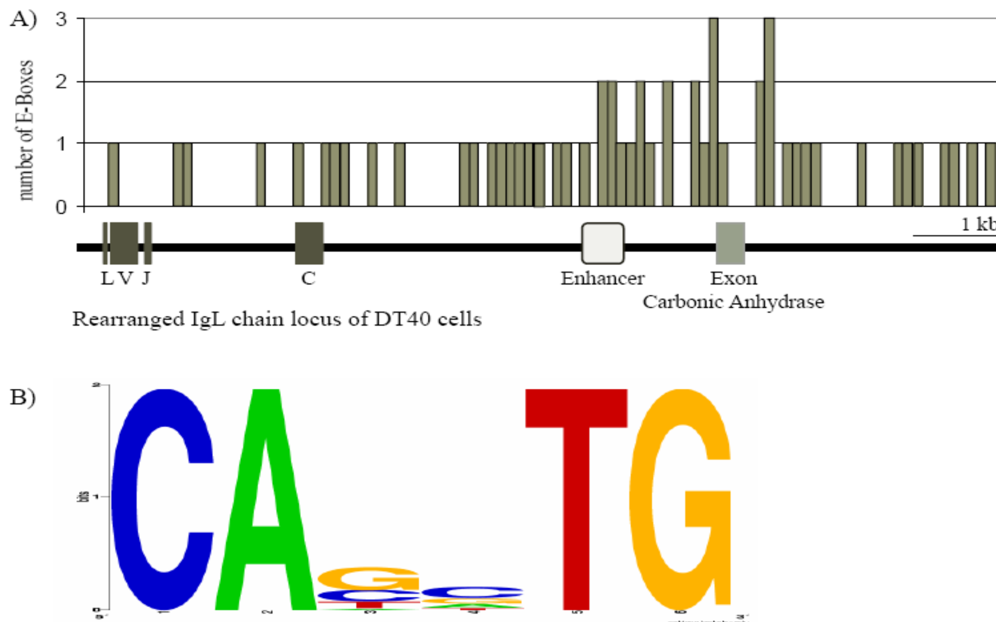


Figure 4.7 Frequency of E Box motifs (CANNTG) on the rearranged IgL chain locus.
A) Distribution along the locus. Counts per 100 bp. A total of 55 E-boxes in 9.8 kb with a clustering around the enhancer region. B) WebLogo (http://weblogo.berkeley.edu/; Schneider et al., 1990) displaying the frequency of the motifs with respect to the two optional nucleotides.

4.2.1 A hypermutation reporter based on GFP expression

In a previous study from our lab we demonstrated that a GFP transgene is a target for mutations, if it is inserted in the rearranged IgL gene of a DT40 cell line and expressed by the endogenous IgL promoter (Arakawa et al., 2008). The gene conversion donors in this cell line were deleted to cause a shift towards hypermutation and AID was expressed constitutively (Arakawa et al., 2004). Within this cell conditions, the GFP fluorescence was influenced by HM mediated mutations, from which some were frame shift/missense or nonsense mutations or an early stop codon. This led to two subpopulations of cells that show decreased and increased green fluorescence respectively when measured by FACS analysis. HM targeted the GFP and led to mutations in the cromophor, causing higher fluorescence in the one population and decreased fluorescence in the other population. This artificial evolution system was used as a reporter assay for HM activity. The new expression cassette named GFP2 consists of the RSV promoter followed by the GFP coding region, an IRES, the bsr gene and the SV40 polyadenylation signal (Figure 4.8). The Rous Sarcoma virus long terminal repeat (RSV) is known to be a very strong promoter yielding in a high expression level (Gorman et al., 1982). The IRES was used to drive expression of a second gene within one promoter. The blasticidin resistance was necessary for later drug selection of the transfected clones. To avoid side effects like interference of transcriptional and post-transcriptional regulation systems of the GFP2 transgene and the IgL gene, the GFP2 reporter was incorporated into the targeting construct *pIgL*^{GFP2} in the opposite transcriptional orientation of the IgL gene and the internal IgL promoter was deleted. The RSV promoter is a one-site directed promoter

acting only upon the GFP2 gene. IgL gene expression was stopped completely.

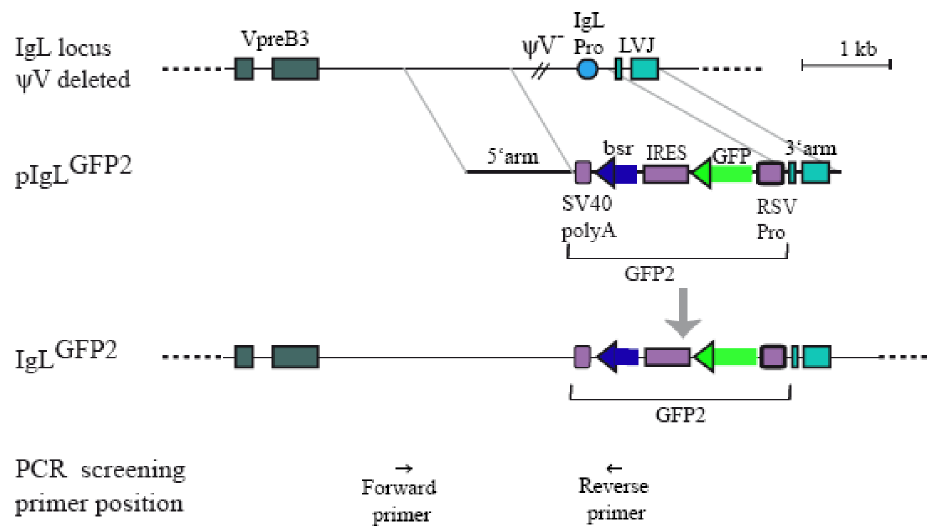


Figure 4.8 Configuration of the GFP2 reporter and targeting strategy of the construct to the rearranged IgL locus.
GFP2 consists of eGFP driven by the RSV promoter. A dicistronic expression cassette controls expression of the GFP and the blasticidin resistance (bsr) gene via an IRES under the same promoter. The polyadenylation signal is from the SV40 virus. The whole reporter is inserted in the opposite direction of the IgL gene at the position of the IgL promoter using convenient arms sequences (sequence between the grey lines). PCR screening is done using a forward primer upstream of the 5' arm and a reverse primer within the bsr gene.

pIgL^{GFP2} was transfected into the DT40 clone AID^{R1}ψV^{-}. After drug selection of bsr resistant clones a number of targeted transfectants named ψV^{-}IgL^{GFP2} were identified by PCR screening using a forward primer upstream of the 5'arm sequence of the construct and a reverse primer within the bsr gene (Figure 4.8; List of primers chapter 8.1). In this primary clones targeted integration has substituted the IgL promoter by the GFP2 transgene. Two independent primary transfectants were subcloned and after two weeks 24 subclones of each clone were analyzed by FACS. The median decrease of green fluorescence was 5.2% and 7.5% (Figure 4.9.B) respectively and confirmed the results of the previous study (Arakawa et al., 2008). The HM machinery is targeting the GFP2 transgene and mutates GFP at a high rate. The cell subpopulation of decreased green fluorescence correlates with the HM activity and can be used for a quantification assay.

HM depends on the expression of the AID gene (Arakawa et al., 2002). As a control, the floxed AID cDNA expression cassette of the clone ψV^{-}IgL^{GFP2} was excised by Cre-mediated recombination induced by tetrahydroxytamoxifen. AID cDNA expression is linked to the expression of the gpt gene and AID negative clones were obtained by negative drug selection using mycophenolic acid. At the same time a positive selection of blasticidin resistant clones was carried out to keep the GFP2 reporter. Two independent ψV^{-}IgL^{GFP2}AID$^{-/-}$ clones were selected for subcloning and 24 subclones of each were analysed by FACS 14 days hereafter. Medians of decreased green fluorescence dropped significantly within both clones to 0.001% (Figure 4.9.A and B) indicating the absence of HM. Thus, the mutations in the ψV^{-}IgL^{GFP2} were caused by AID-mediated diversification. This study was done in cooperation with Hiroshi Arakawa.

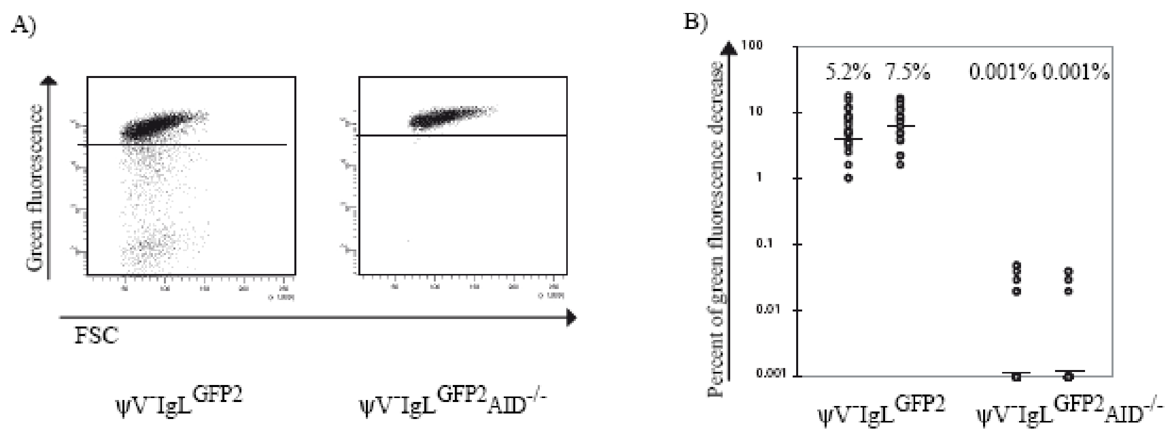


Figure 4.9 Loss of green fluorescence in the GFP2 reporter transfected cell line $\psi V^{-}IgL^{GFP2}$.
A) Representative dot plot of FACS analysis for 24 subclones of the cell mutants $\psi V^{-}IgL^{GFP2}$ and $\psi V\text{-}IgL^{GFP2}AID^{-/-}$. B) Diagram showing the median of decreased green fluorescence (cells below the black gate in the dot plot A) for 24 subclones for two primary clones of the cell mutants $\psi V^{-}IgL^{GFP2}$ and $\psi V\text{-}IgL^{GFP2}AID^{-/-}$.

4.2.2 How is AID-mediated diversification activated

Independently from the primary sequence of the promoter HM can target a transgene inserted at the position normally occupied by the IgL chain gene. As reported previously, the endogenous IgL promoter does not contain elements to specifically recruit HM (Yang et al., 2006). The pseudogenes are deleted in the clone $\psi V^{-}IgL^{GFP2}$ and therefore do not contribute to HM activity, but the locus still contains a 9.8 kb fragment beginning with the transcription start site of the IgL gene and continuing until the 3' end of the next gene (Carbonic Anhydrase (CA)). To test whether the IgL genomic locus includes essential cis-elements, the complete 9.8 kb region, referred to in the following as fragment *'W'*, was deleted. The GFP2 construct for transfection was identical to *pIgL*^{GFP2}, but the downstream 3'arm was shifted from the VJ-region into the region of the next gene CA (Figure 4.10.A and B) thus deleting the entire IgL locus. The construct is referred to as *pIgL*$^{-,GFP2}$ in the following.

Transfection was carried out within the clone $\psi V^{-}IgL^{-}$ in which the entire rearranged IgL locus had been replaced by a puromycin resistance gene. After targeted integration the puromycin cassette will be replaced by the GFP2 reporter at the position of the deleted IgL locus (Figure 4.10.A). Positive drug selection with blasticidin as well as negative drug selection with puromycin identified the resulting transfectants $\psi V^{-}IgL^{-,GFP2}$. PCR screening was done to confirm the targeting. Two independent primary clones were subcloned and 24 subclones of each were analyzed by FACS 14 days after subcloning. Subclones of $\psi V^{-}IgL^{-,GFP2}$ showed medians of only 0.01% and 0.02 % decreased green fluorescence (Figure 4.10.C and D), more than 100 fold lower than the medians of $\psi V^{-}IgL^{GFP2}$ subclones (5.2% and 7.5%). This result clearly indicates the relevance of the *'W'* fragment for HM activity. $\psi V^{-}IgL^{-,GFP2}$ which does not contain *'W'* stopped HM of the GFP2 transgene compared to the $\psi V^{-}IgL^{GFP2}$ clone in which *'W'* is present.

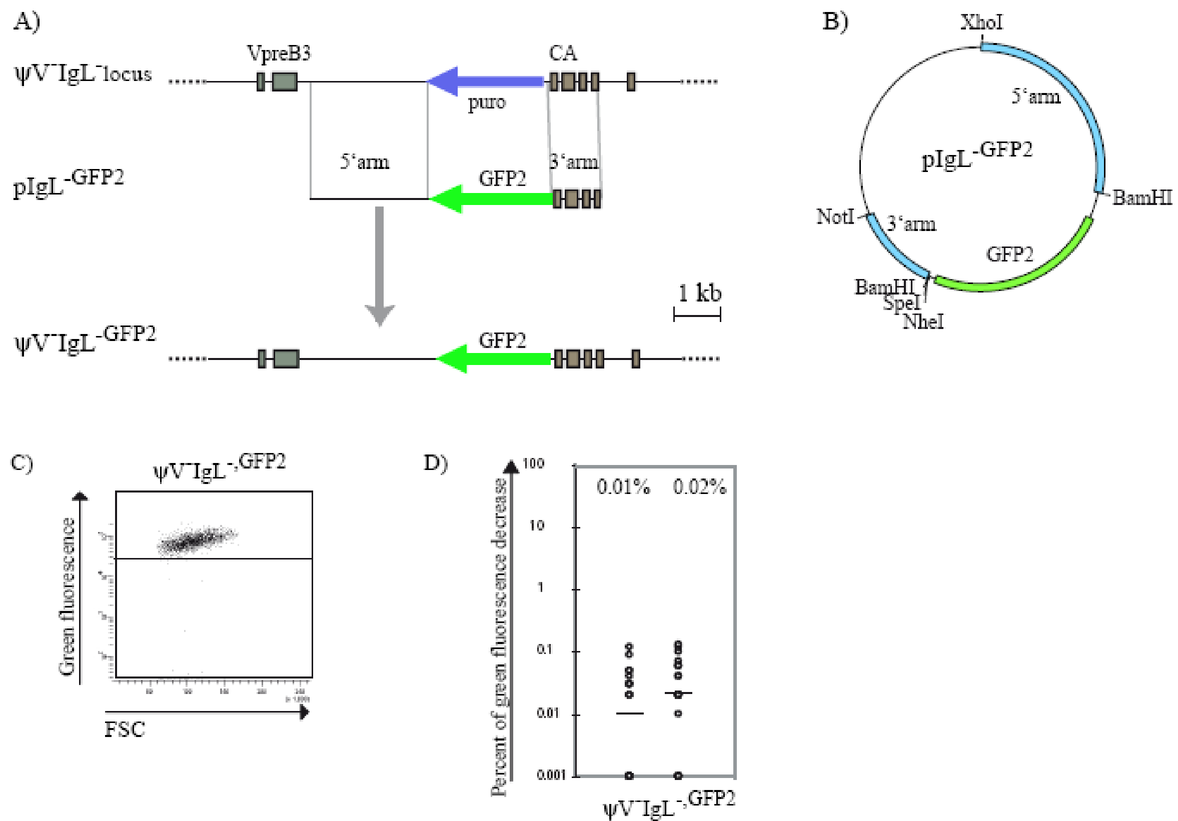


Figure 4.10 GFP2 reporter construct for a knockout of the entire IgL locus.
A) A physical map of the rearranged IgL locus which has been deleted and resubstituted by a puromycin cassette (puro, blue arrow), a targeting construct including the GFP2 reporter (green arrow) and the IgL locus after targeted insertion of the GFP2 reporter. The position of the neighbouring genes VpreB and carbonic anhydrase (CA) is indicated. The arms sequences are framed by grey lines. B) Vector map of the *pIgL*$^{-,GFP2}$ construct. This construct is the basic vector for all further deletion constructs. Basic restriction sites are indicated. SpeI and NheI sites can be used for insertions of IgL deletion fragments. C) Dot plot from a representative clone from FACS analysis. D) Diagram of 24 subclones from two primary clones. The median of decreased green fluorescence is indicated above the bars.

Reinsertion of the *'W'* fragment into the position of the deleted IgL locus should confirm the results. If *'W'* contains all elements relevant for HM, diversification of the transgene should be activated again.

Using the unique SpeI/NheI sites between the 3'arm and the GFP2 reporter (Figure 4.10.A and B), *'W'* was cloned into the plasmid *pIgL*$^{-,GFP2}$.The orientation of the fragment is the same like for the endogenous IgL locus. The construct was then transfected into the ψV^-IgL^- cells and drug selection for blasticidin resistant and puromycin sensitive cell clones yielded in the transfectants $\psi V^-IgL^{W,GFP2}$. The GFP2 followed by the *'W'* fragment resubstituted the puromycin cassette in the position of the IgL gene locus. Subcloning was performed with two independent primary clones and 24 subclones of each were examined by FACS analysis after 14 days. Subclones showed medians of 7.4% and 7.5% decreased green fluorescence (Figure 4.11.B and C). The clone ψV^-IgL^{GFP2} which carries the endogenous 9.8 kb big fragment, has similar medians of decreased green fluorescence (5.2% and 7.5%). Thus, the *'W'* fragment efficiently activates hypermutation after reinsertion into the IgL locus.

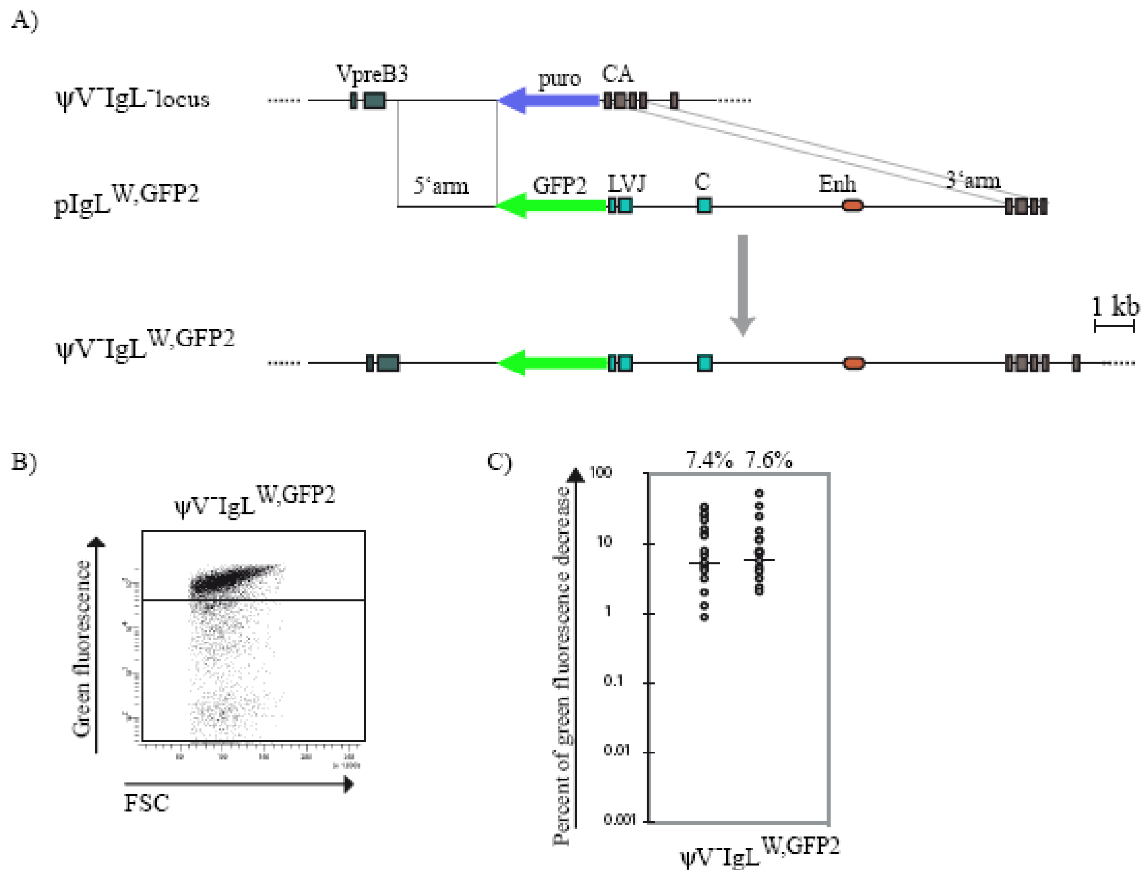


Figure 4.11 GFP2 reporter construct for a knockin of the 9.8 kb IgL fragment referred to as *'W'*.
A) A physical map of the rearranged IgL locus of the cell line ψV⁻IgL⁻, a targeting construct including the GFP2 reporter together with the *'W'* fragment and the IgL locus after targeted insertion of the construct. Puromycin resistance (puro) is indicated by a blue arrow, GFP2 by a green arrow, the neighboring genes are VpreB3 and carbonic anhydrase (CA). *'W'* extends from the rearranged LVJ gene to the 5'end of CA. The arm sequences are framed in grey lines. B) Dot plot from a representative clone from FACS analysis. C) Diagram of all 24 subclones for two primary clones. The median of decreased green fluorescence is indicated above the bars.

4.2.3 Fine mapping of *'W'*

To define putative regulatory sequences, a fine mapping of the gene fragment *'W'* was carried out by combining the GFP2 reporter with serial deletions of the IgL locus. The precursor plasmid for this and all subsequent deletion series was *pIgL$^{-,GFP2}$* (Figure 4.13.B). The SpeI/NheI sites within the plasmid were used to clone in the IgL fragments adjacent to the GFP2 reporter. For transfection, the plasmids were linearized with NotI.

A series of targeting constructs was designed to introduce 1 kb serial deletions from both ends of the *'W'* fragment (Figure 4.12). This resulted in 13 constructs. Additionally, a construct was designed which contains only the middle part of the IgL locus surrounding the region where the enhancer is located.

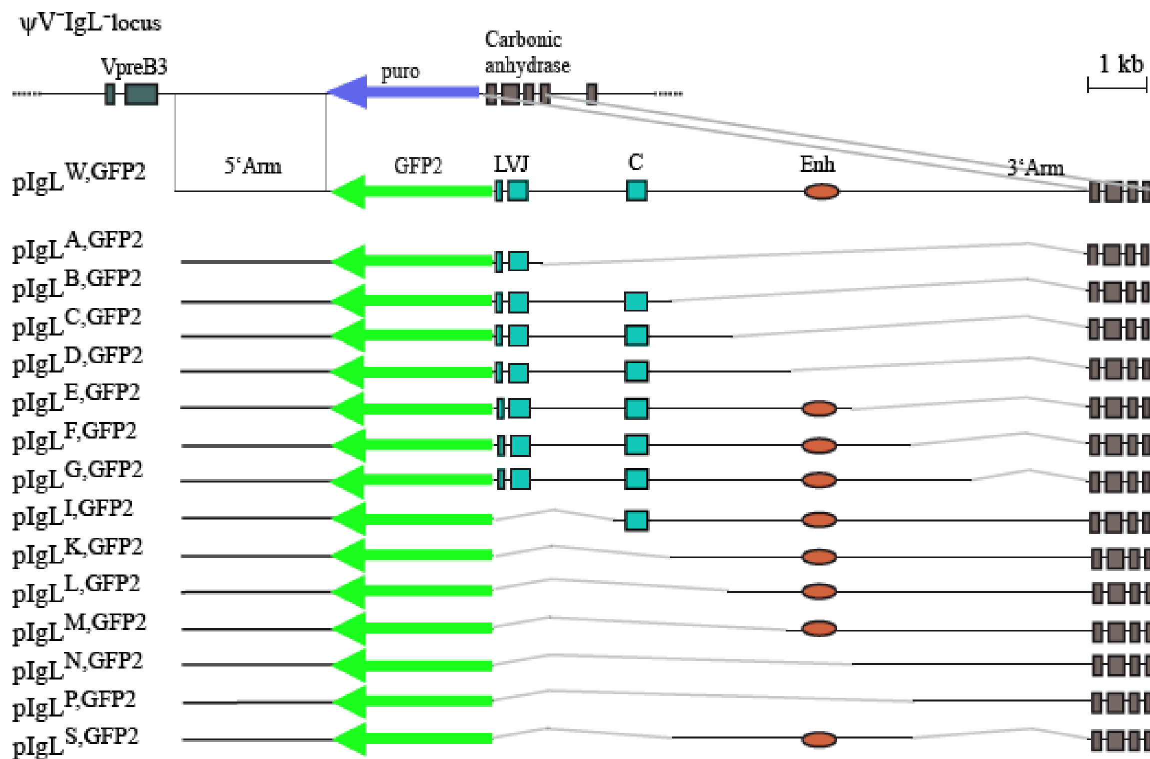


Figure 4.12 *'W'* fragment deletion series.
A physical map of the IgL locus of the cell line ψV⁻IgL⁻ and the aligned targeting constructs leading to the insertion of the GFP2 reporter (green arrow) together with parts of the *'W'* fragment. The puromycin cassette (puro) is indicated by a blue arrow. *'W'* extends from the rearranged LVJ gene to the beginning of the next gene carbonic anhydrase.

The constructs were transfected into ψV⁻IgL⁻ and for each resulting cell mutant FACS analysis was done on subclones of two individual clones isolated 14 days after subcloning (Figure 4.13.A). Shortening the *'W'* fragment from both ends resulted in a preceding decrease of HM activity. Medians of decreased fluorescence from the *'A'*, *'B'* and *'P'* fragments are still increased compared to those of the ψV⁻IgL$^{-,GFP2}$ (0.01% and 0.02%), but are low compared to the full length *'W'* fragment (Figure 4.13.B). None of the fragments alone are able to induce strong active HM.

The *'D'* fragment had medians of 0.5% and 0.3% decreased fluorescence. Extension of *'D'* to include the additional 1 kb fragment that contains the putative enhancer resulted in the *'E'* fragment and increased HM activity 5.5 fold. Interestingly, all fragments (*'E','F','G','I','K','L','M','S'*; Figure 4.12 and 4.13.B) which contain the previously identified IgL enhancer (Bulfone-Paus et al., 1997) showed clearly enhanced HM levels. However, the *'N'* fragment, without the putative enhancer sequence had medians of 4.4% and 9% decreased green fluorescence. *'N'* contains a 1 kb fragment (referred to as *'3-4'*) more sequence than *'P'* but induced HM activity 136 fold compared to the *'P'*. Adding the same 1 kb fragment to the *'E'* fragment resulted in the *'F'* fragment which induced HM only 2.5 fold. However, there are several possibilities that might explain this divergence. First, the sequence *'E'* might not only contain enhancing, but also silencing DNA elements which are able to act on the HM level. Second, *'3-4'* contains essential elements which are repeats of motifs contained

within *'E'* or third, HM has a saturation level and can not be increased beyond this.

The *'S'* fragment, which is the fragment between the *'B'* and the *'P'* fragment, produced a median decreased fluorescence of 1.7% and 2.7%. It contains the enhancer and the 1 kb fragment *'3-4'*. However, the frequency was lower compared to the other fragments containing both of these 1 kb fragments. 5' elongation of the *'S'* fragment by adding the *'B'* fragment, which had no strong HM activity on its own, resulted in the *'F'* fragment with medians of 4.3% and 6.8%. This should not be due to a distance effect from the promoter, because the *'K'* fragment, which consists of *'S'* and the 3' elongation *'P'*, had also significantly increased medians of 6% and 10.4%, which was four fold higher than the *'S'* fragment alone. Presumably, the *'S'* fragment contains core elements which are absolutely necessary for HM. These core elements seem to be supported by flanking redundant elements, which have minor activity on their own for HM, and are able to enhance the overall activity slightly when in combination.

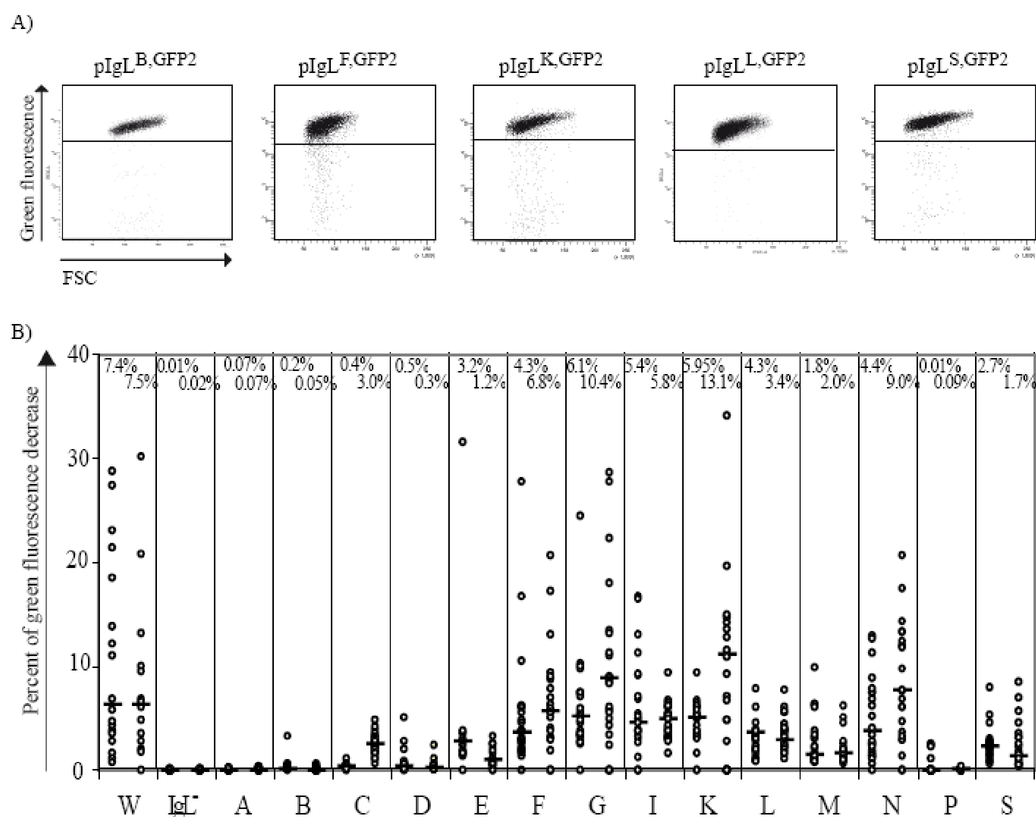


Figure 4.13 Analysis of the 'W' fragment deletion series.

A) FACS analysis of representative primary transfectants. B) Fluctuation analysis of subclones. Only the name of the reconstituted fragment and not the full name of the transfectant is indicated for clarity. The median of decreased green fluorescence is indicated above the bars.

Taken together, the observed effects are hard to interpret as the variation between the different deletion constructs is high. To make a more precise conclusion about involved motifs, it is necessary to narrow the area of the hypothetic HM activator.

The *'S'* fragment with 4 kb in size is compact and it contains both areas which were earlier identified as being important (the fragment containing the enhancer and the following 1 kb *'3-4'*). Hence, we decided to use the *'S'* fragment for further deletion series and started a fine mapping of the locus.

4.2.4 Fine Mapping of *'S'* fragment

Thinking about motifs of a HM activator, one must consider different situations. First, there could be a core element which is supported by different redundant elements. A second possibility would be that cooperative elements act together. One deletion alone may not be suitable to show up a complex element structure. Thus several independent deletion series were planned. To define shorter elements responsible for activation of diversification, the *'S'* fragment, in the following referred to as *'0-4'*, was mapped into smaller fragments.

The 4 kb *'0-4'* fragment spans 2 kb sequence upstream and 1.6 kb sequence downstream of the already identified 0.4 kb enhancer. This element might not only effect the transcription of the IgL locus through the enhancer but may also involve the hypermutation mechanism. Hence it was decided to perform a detailed deletion analysis of the 4 kb region.

In the following the five different deletion series are described:

1) *'0-2'* series and *'2-4'* series:

The 5' starting point of the *'0-4'* was defined as zero point and the sequence was divided into two 2 kb regions *'0-2'* and *'2-4'*. Staggered 200 bp end deletions were introduced into these two fragments to define redundant motifs.

2) *'0-4'* series 1kb serial deletions:

The 5' starting point of the *'0-4'* was defined as zero point and the sequence was divided into four 1 kb regions. The deletion series consisted of 1 kb stepwise deletions starting from the 5' as well as from the 3' end of the sequence, and of four 1 kb single fragments of the *'0-4'* region to define a smaller core region containing the putative HM active motifs.

3) *'0-4'* series 200 bp and 400 bp internal deletions:

The 5' starting point of the *'0-4'* was defined as zero point and the sequence was divided into 20 200 bp and 10 400 bp regions respectively. The deletions were introduced individually into the 4 kb fragment. If an essential region is deleted, the fragment will loose the ability to hypermutate the GFP2 transgene. Cooperative elements would show up even if they are far distant from each other.

The exact sequence of all deletion fragments in correlation to the sequence of *'W'* can be obtained by aligning the primer sequence (list of primers, chapter 8.1) and the sequence of *'W'* (chapter 8.3.1).

All in the following targeting constructs were transfected into the DT40 mutant $\psi V^{-}IgL^{-}$. This cell line has a puromycin drug resistance cassette inserted at the position of the deleted IgL chain locus which will be deleted after targeted integration of the GFP2 reporter and the cell line will become puromycin sensitive. Furthermore, the endogenous AID of this mutant is knocked out and resubstituted by a randomly integrated dicistronic AID-IRES-gpt expression cassette conferring a resistance to mycophenolic acid. To keep AID, drug selection with mycophenolic acid was performed. This is an important issue, since a decrease in HM activity could also be due to a decreased AID expression.
After blasticidin selection screening for successful integration of the GFP2 reporter into the genome, the resulting cell mutants were obtained by further drug selection for puromycin-negative and mycophenolic acid-positive primary clones. A PCR screening additionally confirmed insertion of the GFP2 reporter at the position of the IgL locus. Subcloning was then performed on two individual primary clones which had been cultured in mycophenolic acid for 3 days. Under this condition AID-deficient cells will be killed. The subclones were analysed by FACS 14 days after subcloning.

4.2.5 *'0-2'* series

Targeting constructs were designed for a staggered deletion of 200 bp from both ends of the *'0-2'* fragment (Figure 4.14.A). Medians of decreased green fluorescence were compared to the *'0-4'* fragment. All fragments show significantly decreased medians of green fluorescence of around 0.1% (Figure 4.14.B). The decay in HM was not due to a decrease in AID expression, as AID activity was assured by culturing the primary clones in mycophenolic acid. GFP transgene diversification is very low in these mutants and therefore we could not identify a HM active motif. The complete *'0-2'* fragment is not able to support diversification of the GFP transgene. This is in accordance with the results of the previous study, where the *'D'* fragment, which contains the *'0-2'*, but not the *'2-4'* fragment, has only very low levels of 0.5% and 0.3% decreased green fluorescence (Fig.4.12 and 13).

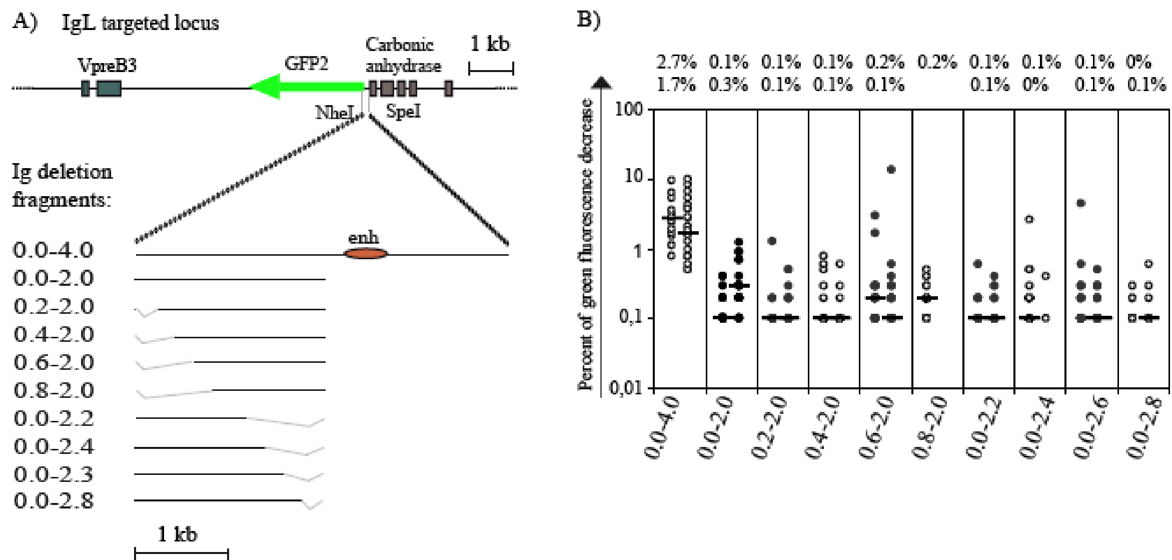


Figure 4.14 *'0-2'* Series: 200 bp end deletions.
A) A physical map showing the generated deletion fragments. The position of the enhancer (enh; red oval) is indicated. The fragments are cloned adjacent to the GFP2 reporter (green arrow) via the NheI and SpeI sites. The reporter constructs are integrated into the position of the rearranged IgL locus of the cell line $\psi V^{-} IgL^{-}$. B) Fluctuation analysis of subclones. Only the name of the reconstituted fragment and not the full name of the transfectants is indicated for clarity. The median of decreased green fluorescence is indicated above the bars.

4.2.6 *'2-4'* series

Targeting constructs were designed for a staggered deletion of 200 bp from both ends of the *'2-4'* fragment (Figure 4.15.A). Medians of decreased green fluorescence were compared to the *'0-4'* fragment. The *'2-4'* fragment still has similar HM activity to the *'0-4'*. Removal of the first 200 bp led to a 3.5 fold increase of the mutation rate (fragment *'2.2-4.0'*, Figure 4.15.B). Probably a silencing element was removed or this was due to a distance effect. However, after deletion of another 200 bp HM dropped back to the level of the full fragment *'2-4'*. In the *'2.6-4'* fragment, a first decrease of mutation activity was visible. The medians are 0.2% and 1.4%. Also the *'2.8-4'* fragment with 0.4% and 0.5% has low medians of decreased green fluorescence. These results led to the suggestion, that the 600 bp fragment *'2.0-2.6'* contains important motifs that activate diversification. Deletions from the 3' end of *'2-4'* gave no indication of any motifs within this region. HM was still present in all of the deletion mutants at a level comparable to the *'2-4'* fragment. This supports the idea of *'2.0-2.6'* triggering diversification. Interestingly, this fragment also contains the main part of the IgL enhancer.

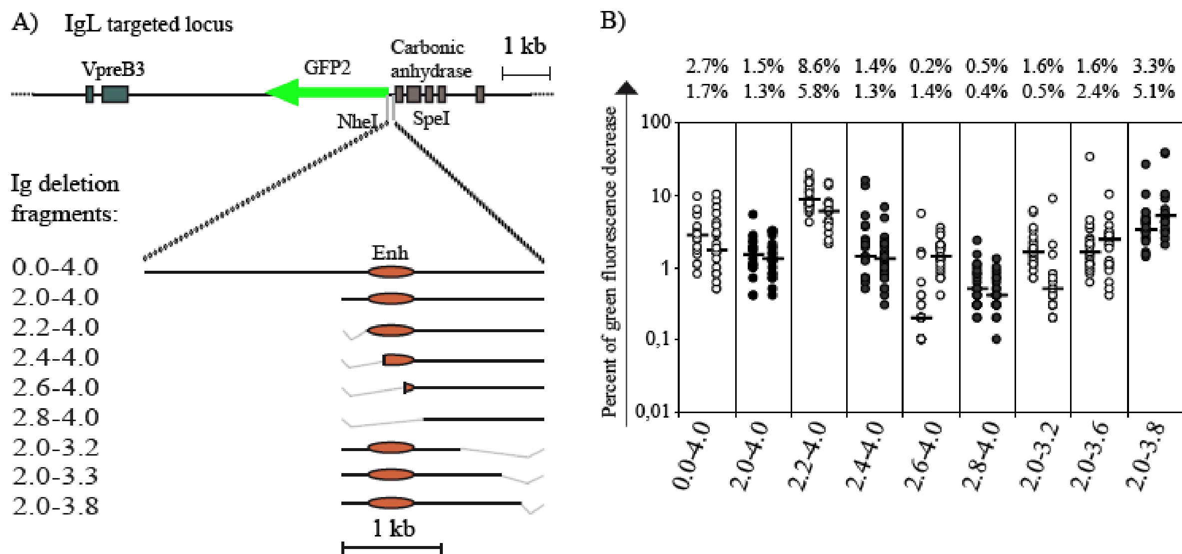


Figure 4.15 '2-4' Series: 200 bp end deletions.
A) A physical map showing the generated deletion fragments. The position of the enhancer (enh; red oval) is indicated. The fragments are cloned adjacent to the GFP2 reporter (green arrow) via the NheI and SpeI sites. The reporter constructs are integrated into the position of the rearranged IgL locus of the cell line ψV⁻IgL⁻. B) Fluctuation analysis of subclones. Only the name of the reconstituted fragment and not the full name of the transfectants is indicated for clarity. The median of decreased green fluorescence is indicated above the bars.

4.2.7 *'0-4'* series 1 kb deletions

Targeting constructs producing 1 kb end deletions from both ends of *'0-4'* and 1 kb partial insertions were transfected into $\psi V^- IgL^-$ (Figure 4.16.A). The resulting mutants were subcloned and analysed by FACS after 14 days. Figure 4.16.B shows the decrease in green fluorescence of the subclones in comparison to the *'0-4'* fragment and the $\psi V^- IgL^{-,GFP2}$. The significance of the difference in HM activity between *'0-4'* and the deletion mutants was tested by the Mann-Whitney U-test using the statistical program *R*.

The *'0-2'* with medians of 0.1% and 0.3% which corresponds to an eleven fold decrease, showed strongly diminished HM activity. In accordance with this result, the 1 kb mutants *'0-1'* and *'1-2'* exhibited a significant drop in diversification of the GFP transgene. But compared to *'IgL-'*, the green fluorescence decrease of *'0-1'* (0.1% for both clones) and *'0-2'* (0.1% and 0.3%) was still significantly above background level. Interestingly the *'0-3'* (which includes the *'2-3'* in addition to the *'0-2'*) with a median of 3.7% and 1.3% was not significantly different from *'0-4'* and hypermutated actively.

'2-4' was not significantly different from the *'0-4'* and therefore HM active. Although the *'2-3'* fragment is only 1 kb in size, it still had a median of 0.9% and 1.3%. Compared to *'0-4'* HM is decreased two fold, but the U-test indicated that this value was not significant. This indicates cis-elements responsible for inducing HM in *'2-3'*. Diversification activity in *'3-4'* was diminished to 0.4% and 0.1%.

'2-3' was the only 1kb fragment that showed still increased levels of HM. Interestingly, the fragments *'0-3'*, *'2-4'* and *'1-4'*, which all contain the *'2-3'*, exhibit no significant difference compared to *'0-4'* in Mann-Whitney test, whereas the fragments which do not contain the sequence of *'2-3'*, are free from HM activity.

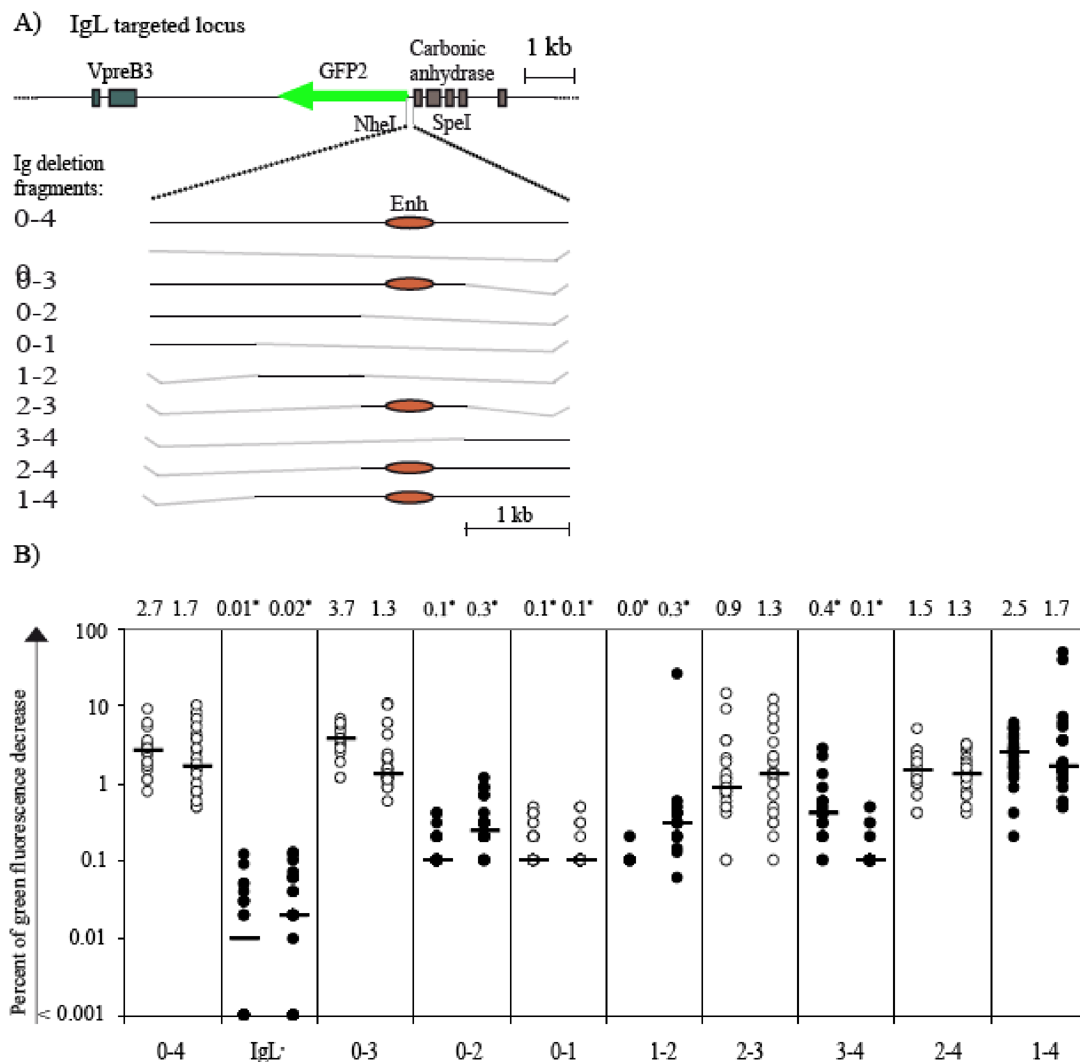


Figure 4.16 *'0-4'* series: 1 kb deletions.
A) A physical map showing the generated deletion fragments. The position of the enhancer (enh; red oval) is indicated. The fragments are cloned adjacent to the GFP2 reporter via the NheI and SpeI sites. The reporter constructs are integrated into the position of the rearranged IgL locus of the cell line ψV⁻IgL⁻. B) Fluctuation analysis of subclones. Only the name of the reconstituted fragment and not the full name of the transfectants is indicated for clarity. The median of decreased green fluorescence is displayed above the bars. Highly significant p-values ($p<0.001$) are marked with an asterisk.

4.2.8 *'0-4'* series 200 bp internal deletions

A set of 20 constructs was designed to introduce internal deletions of 200 bp (Figure 4.17.A) into the *'0-4'* fragment. The constructs were transfected into ψV⁻IgL⁻ (Figure 4.16.A). Two individual primary clones of each mutant were subcloned and analysed by FACS after 14 days.

The majority of the deletions did not exhibit significant changes in HM frequency compared to the *'0-4'* fragment (Figure 4.17.B). For some mutants there were differences between the two individual primary clones (*'0-4Δ0.0-0.2'*, *'0-4Δ0.8-1.0'*, *'0-4Δ3.8-4.0'*). Due to an early mutation and the expansion of the subclones one clone might show increased HM activity compared to the other. To check the mRNA level or prepare a sequencing of the GFP2 transgene could clarify the situation. However, the activity of the clones is in the same

range like *'0-4'* or even higher, which means there will be no relevant active diversification motifs within the deleted 200 bp piece and consequently the clones were not further examined.

The medians of decreased green fluorescence of the fragments *'0-4Δ0.8-1.0'* (2.8% and 7.5%) and *'0-4Δ1.4-1.6'* (5.0% and 7.9%) were highly significantly increased compared to *'0-4'*. Presumably, the deletion of the 200 bp in these fragments had removed an inhibitory element. *'0-4Δ0.8-1.0'* was also deleted in *'1-4'* and *'0-4Δ1.4-1.6'* was deleted in *'2-4'*. Both clones with a median decreased green fluorescence of 2.5% and 1.7%, and 1.5% and 1.3%, respectively, could not confirm a role for an inhibitory element. However, silencing motifs do not play a role for a *'HyCorE'* and this phenotype was not further analyzed.

Two of the 200 bp internal deletions exhibit highly significant reduced medians of green fluorescence decrease (Figure 4.17.B, red boxes). *'0-4Δ0.4-0.6'* is decreased to 0.2%, and *'0-4Δ2.2-2.4'* has only 0.4% and 0.5% decrease of green fluorescence (Figure 4.17.B). The 200 bp fragment *'0.4-0.6'* is within the 1 kb fragment *'0-1'* which has no activity on its own (Figure 4.16.B). To make sure if *'0.4-0.6'* contains relevant cooperative motifs further deletion studies are necessary.

The 200 bp fragment *'2.2-2.4'* contains the first part of the IgL enhancer and is included in the 1 kb fragment *'2-3'* which was shown to be sufficient to induce AID-mediated diversification of the GFP2 reporter (Figure 4.16).

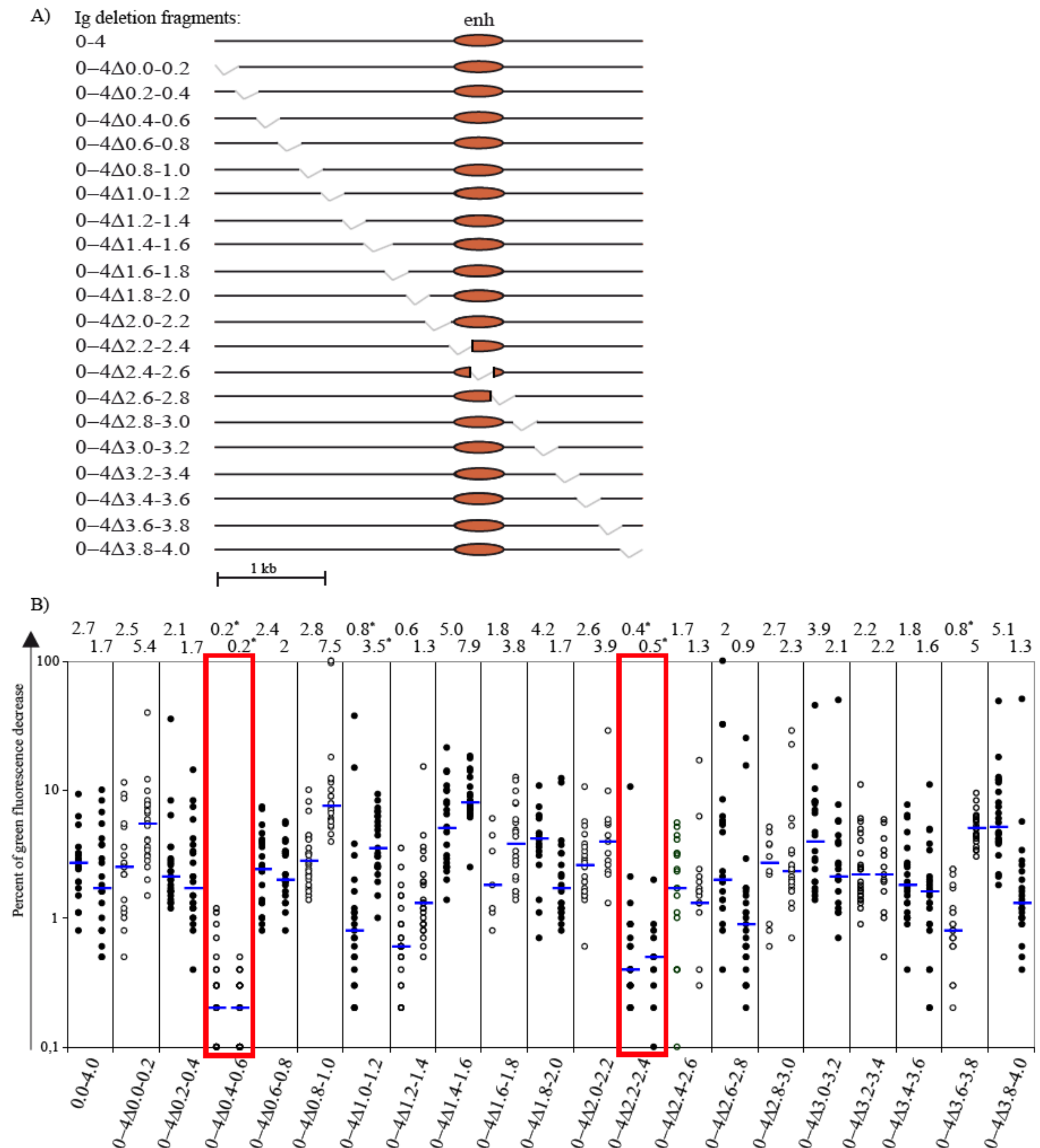


Figure 4.17 ‘0-4‘ series: 200 bp internal deletions.
A) A physical map showing the generated deletion fragments. The position of the enhancer (enh; red oval) is indicated. B) Fluctuation analysis of subclones. Only the name of the reconstituted fragment and not the full name of the transfectants is indicated for clarity. Median of green fluorescence decrease for each clone is written above the bars. Highly significant p-values ($p<0.001$) are indicated by an asterisk. Clones with a highly significant decrease of mutations are highlighted with a red box.

4.2.9 *‘0-4’* series 400 bp internal deletions

The 400 bp internal deletions were thought to help confirm a role for active motifs found within the 200 bp internal deletions. Targeting constructs were designed and transfected into the cell line ψV⁻IgL⁻ (Figure 4.18.A). Two individual primary clones of each mutant were subcloned and analyzed by FACS after 14 days.

All cell clones were mutating the GFP transgene to a similar or higher frequency than the clone with the *'0-4'* fragment (Figure 4.18.B). There was no significant drop in HM activity observable. The mutants *'0-4Δ0.0-0.4'*, *'0-4Δ1.2-1.6'* and *'0-4Δ1.6-2.0'* had a highly significant p-value. As *'0-4Δ0.0-0.4'* and *'0-4Δ1.6-2.0'* are deleted in *'1-4'* and *'2-4'* respectively and the two latter ones show similar levels of diversification like *'0-4'*, a possible role for silencing elements within the two deletion mutants could not be confirmed. Also the 200 bp internal deletions can not confirm the result. The increased medians of *'0-4Δ1.2-1.6'* are in agreement with the *'0-4Δ1.4-1.6'* but could not be confirmed by other deletion studies. To define silencing elements, another approach would be necessary.

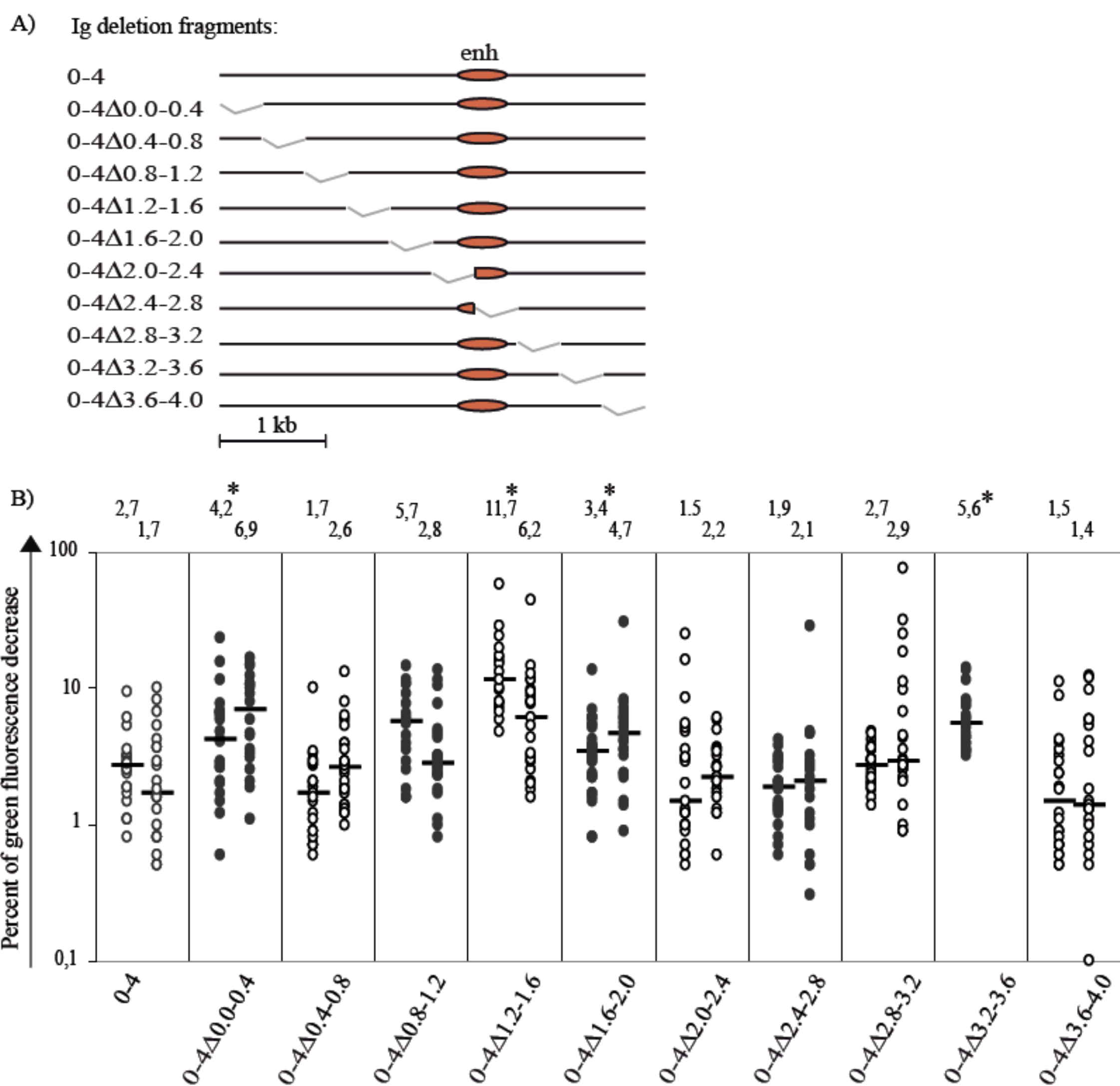


Figure 4.18 '0-4' series: 400 bp end deletions.
A) A physical map showing the generated deletion fragments. The position of the enhancer (enh; red oval) is indicated. **B)** Fluctuation analysis of subclones. Only the name of the reconstituted fragment and not the full name of the transfectants is indicated for clarity. Median of GFP decrease for each clone is written above the bars. Highly significant p-values ($p<0.001$) are indicated by an asterisk.

The putative active motif *'0.4-0.6'* found with the 200 bp deletion series could not be confirmed within the 400 bp internal deletions (Figure 4.18.B). The median decreased green fluorescence of the corresponding fragment *'0-4Δ0.4-0.8'* (1.7% and 2.6%) is not significantly different from *'0-4'*. It was surprising that also deletion of the fragment *'2.0-2.4'* had no effect on HM level. The cell line with the fragment *'0-4Δ2.0-2.4'* has in addition to the *'2.2-2.4'* another 200 bp upstream of the enhancer deleted (Figure 4.18.A, *'0-4Δ2.0-2.4'*). Also *'0-4Δ2.4-2.8'* which has the 3' part of the enhancer deleted was not significantly different from *'0-4'* (Figure 4.18.A and B). An active role of the enhancer for diversification of the GFP2 reporter could not be confirmed within this deletion series.

4.3 *A new series of deletions to identify specific active motifs*

Taken together, results until now support the idea of the enhancer being also involved in activating HM. The fragment *'2-3'* was sufficient for HM, as insertion of it into the position of the deleted IgL locus induced diversification of the GFP2 transgene. The element was also necessary to start HM, as the 1 kb mutants of *'0-4'* which did not contain the critical *'2-3'* fragment showed no HM activity. Within the enhancer, the *'2.2-2.4'* fragment is most interesting, because it was necessary to start HM in the 200 bp internal deletions fragments of *'0-4'*.

Typical cis-elements which serve as binding platforms for transcription factors have a length of 10-20 bp. To minimize the size and define a more precise active motif, serial deletions of 50 bp in the *'2-3'* fragment were performed. End deletions from both ends are thought to remove redundant elements and internal deletions are useful to remove a core motif. Only one primary clone for each construct was examined.

4.3.1 '2-3' series 50 bp deletions 5'end

19 constructs were designed for a stepwise deletion of the *'2-3'* fragment starting from the 5' end of the fragment (Figure 4.19.A). All deletion mutants were compared to the *'2-3'* fragment and the p-values were calculated ($p<0.001$). High significance was indicated by a drop in the medians of decreased green fluorescence in the deletion mutants which means HM had stopped due to removal of an active motif. Deletion of the first 200 bp did not change HM activity. A decay in HM activity was seen first with the *'2.25-3.00'* fragment. The median of 0.2% is five times lower and highly significant different from the 0.9% median of *'2-3'* (Figure 4.19.B). All following mutants show a highly significant difference in diversification activity compared to *'2-3'*. These results suggest that there is a HM active motif within the 250 bp fragment *'2.00-2.25'*. Compared to the *'2.2-3'* fragment with a median of 0.6%, the median of *'2.25-3.00'* is three fold lower.

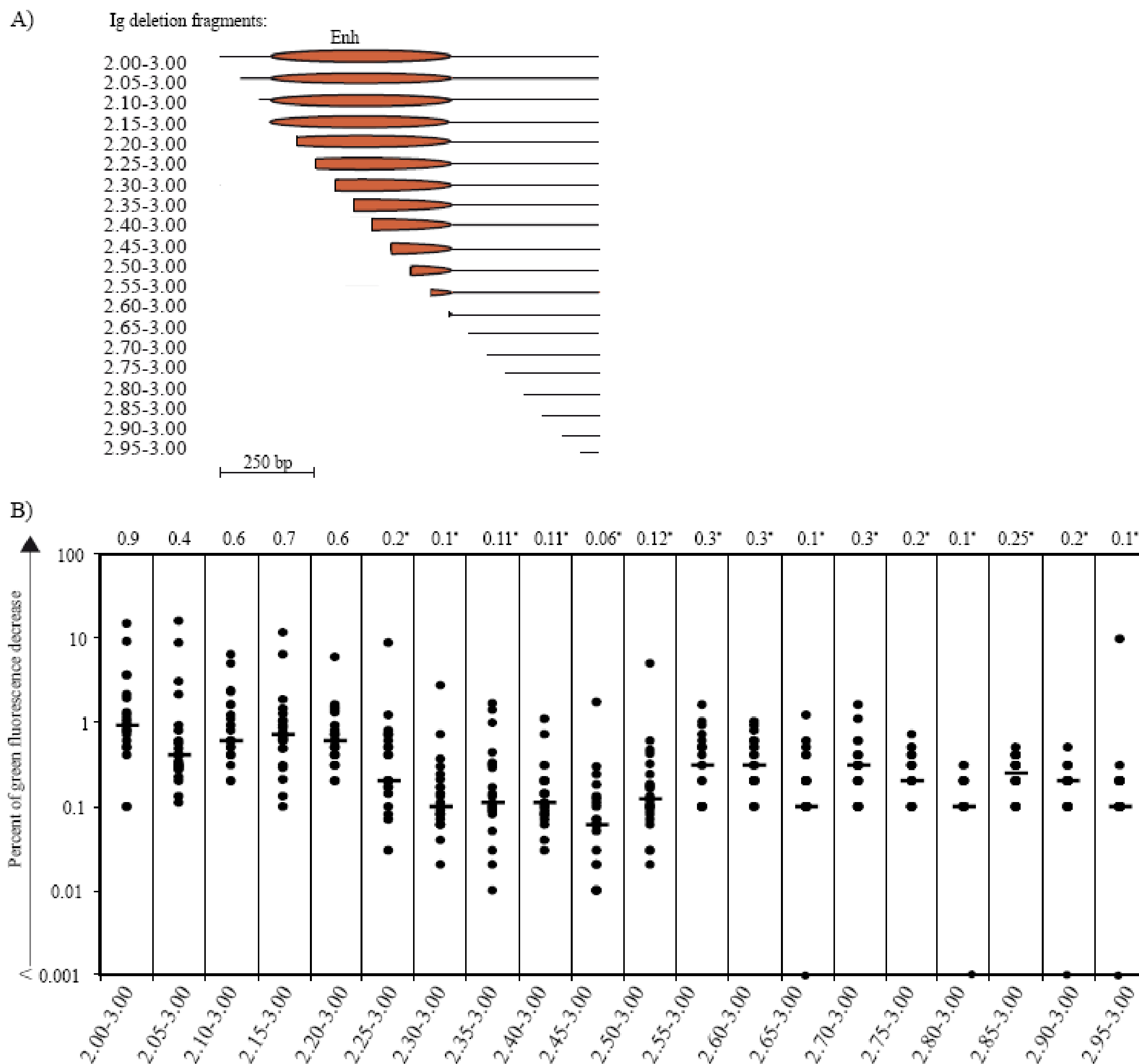


Figure 4.19 '2-3' series: 50 bp 5'end deletions.
A) A physical map showing the generated deletion fragments. The position of the enhancer (enh; red oval) is indicated. B) Fluctuation analysis of subclones. Only the name of the reconstituted fragment and not the full name of the transfectants is indicated for clarity. Median of GFP decrease for each clone is written above the bars. Highly significant p-values ($p<0.001$) are indicated by an asterisk.

4.3.2 '2-3' series 50 bp deletions 3'end

An additional 19 constructs were designed to delete the *'2-3'* fragment from the 3' end (Figure 4.20.A). A true HM motif should show up in both deletion series as it would work independently from other motifs. Deletion of the first 700 bp did not give a hint on HM activating sequences. The fragments *'2.00-2.80'* with 0.2% and *'2.00-2.75'* with 0.3% showed a highly significant decreased median of green fluorescence, but HM restarted again when further deleting parts of the sequence (Figure 4.20.B). Therefore, the sequence *'2.75-2.85'* can not contain a diversification activating motif. The fragment *'2.00-2.25'*, which has 750 bp of the 3'end deleted, had a highly significant decreased median of 0.2% and also all following fragments were no more able to mediate

HM. Conclusively, the *'2.00-2.30'* fragment contains relevant motifs, where the last 50 bp *'2.25-2.30'* seem to be most important.

Compared to the study of the 5' end deletions, the results are similar. The *'2.00-2.25'* fragment was seen to have a strong influence on HM with the *'2.20-2.25'* being most important. The region *'2.20-2.30'* was also confirmed by the 200 bp internal deletions *'2.2-2.4'* of the *'0-4'* fragment (*'0-4Δ2.2-2.4'*, Figure 4.17).

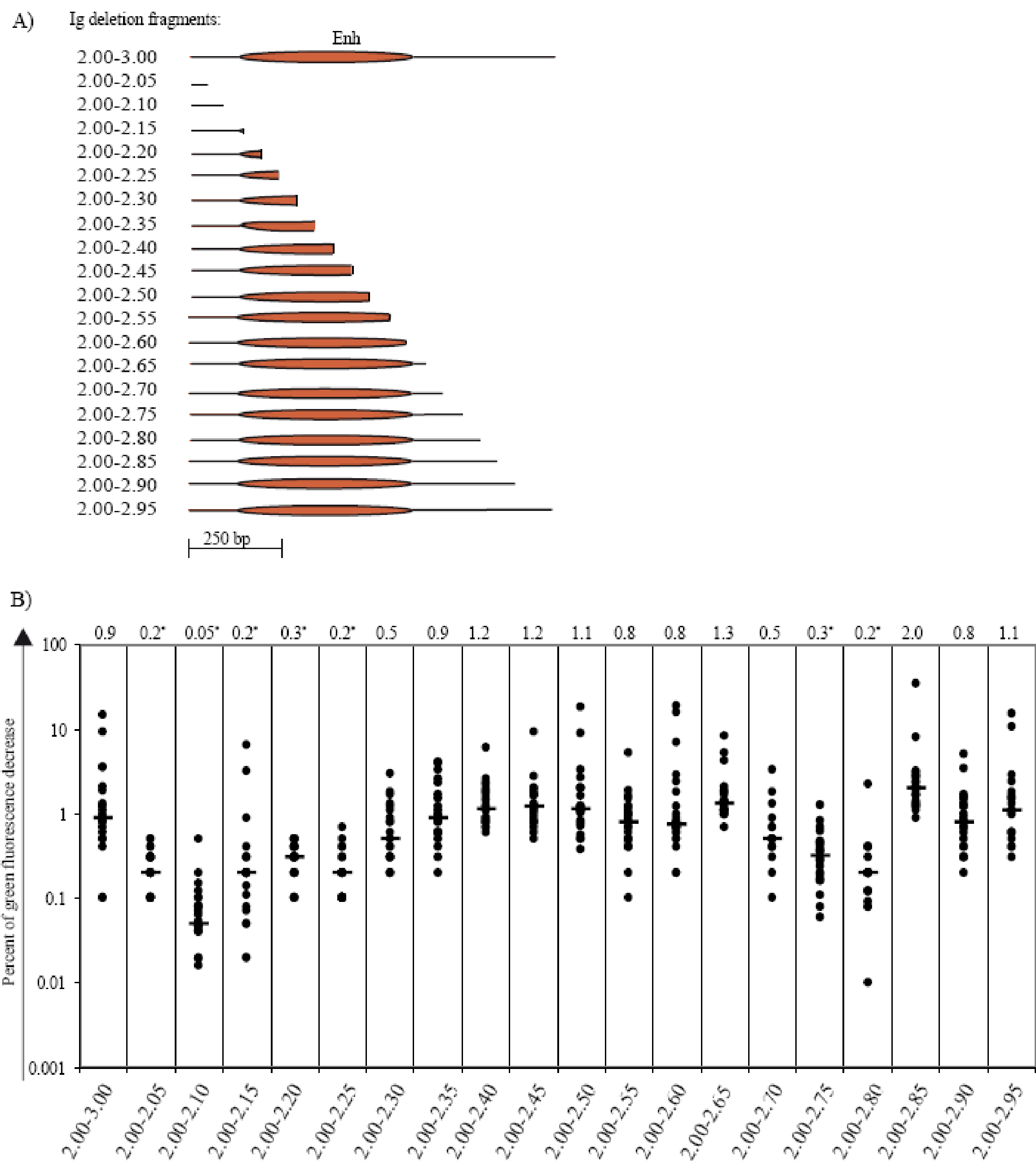


Figure 4.20 '2-3' Series: 50 bp 3'end deletions.
A) A physical map showing the generated deletion fragments. The position of the enhancer (enh; red oval) is indicated. B) Fluctuation analysis of subclones. Only the name of the reconstituted fragment and not the full name of the transfectants is indicated for clarity. Median of GFP decrease for each clone is written above the bars. Highly significant p-values ($p<0.001$) are indicated by an asterisk.

4.3.3 '2-3' series 50 bp internal deletions

20 constructs were designed each with another 50 bp of the '2-3' fragment being deleted (Figure 4.21.A). The medians of the resulting mutants were compared to the median of '2-3'. The mutants '2-3Δ2.35-2.4' and '2-3Δ2.60-2.65' show a highly significant increase of their medians. All other mutants are not significantly different from '2-3' (Figure 4.21.B). Finally, with this deletion series I could not identify a DNA element triggering diversification. This strengthens a theory of redundant motifs being responsible for HM activation.

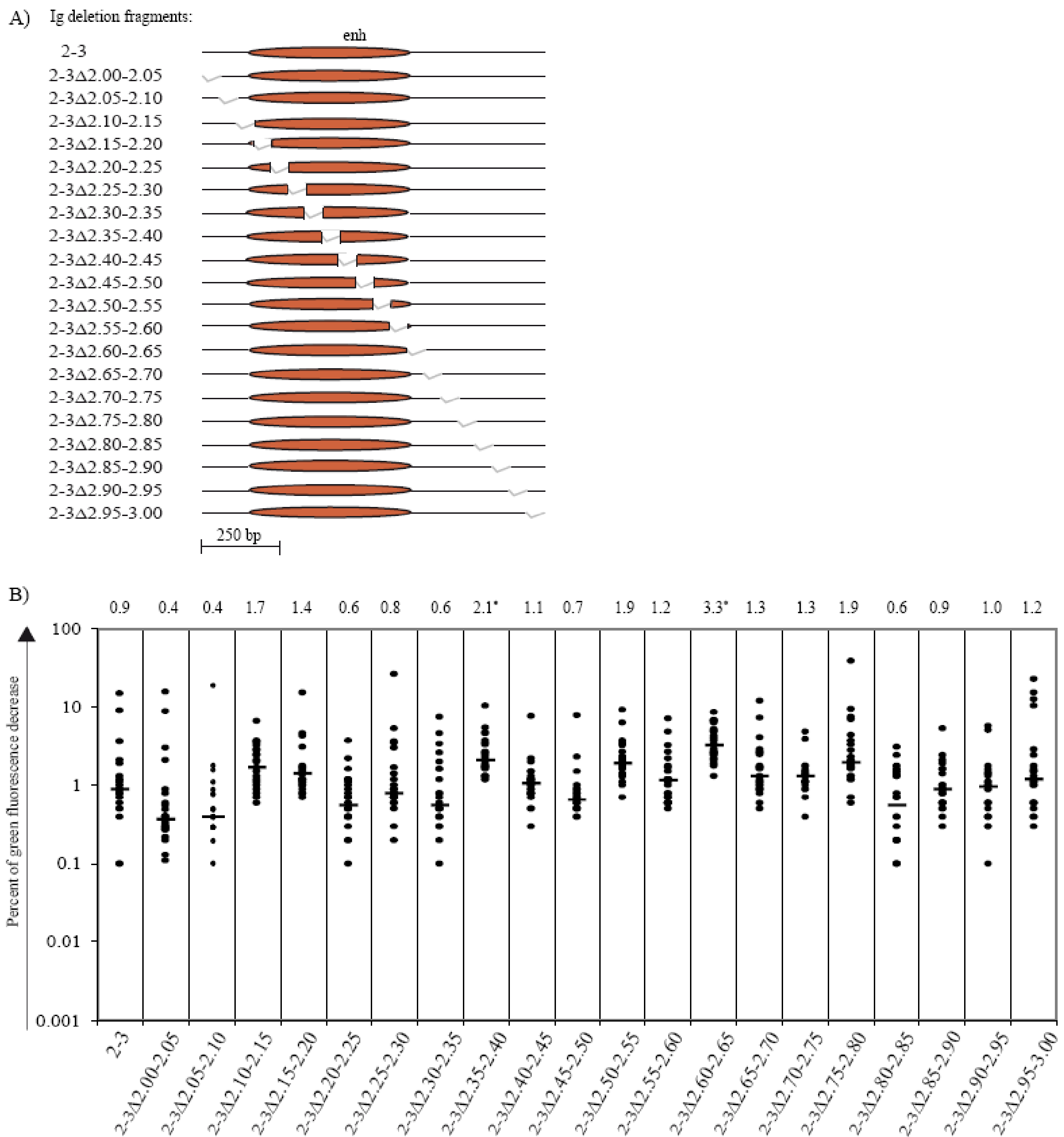


Figure 4.21'2-3' Series: 50 bp internal deletions.
A) A physical map showing the generated deletion fragments. The position of the enhancer (enh; red oval) is indicated. B) Fluctuation analysis of subclones. Only the name of the reconstituted fragment and not the full name of the transfectants is indicated for clarity. Median of GFP decrease for each clone is written above the bars. Highly significant p-values ($p<0.001$) are indicated by an asterisk.

4.3.4 Reconstitution and multimerization of *'2.2-2.4'*

Deleting *'2.2-2.4'* in the *'0-4'* fragment led to a five fold reduction of median decreased green fluorescence (Figure 4.17). Deleting *'2.25-2.3'* and *'2.2-2.25'* in the *'2-3'* end deletion series led to a four fold reduction of median decreased green fluorescence (Figure 4.19 and 20) and confirmed the essential role of the *'2.2-2.4'* enhancer fragment in the process of hypermutation.

In a next step, the 200 bp fragment *'2.2-2.4'* was inserted together with the GFP2 reporter into the position of the deleted IgL locus of the cell line $\psi V^{-}IgL^{-}$ to see whether the fragment is also sufficient to induce HM.

The single insertion *'1x [2.2-2.4]'* showed a median of 0.3% and 0.6% decreased green fluorescence (Figure 4.22.A and B). This is 30 times more activity than compared to the clone *'IgL-'*, which does not contain any HM supporting sequence. However, the level is 2.4 fold lower than that of *'2-3'* with 1.3% and 0.9%, and 4.8 fold lower than with *'0-4'* (1.7% and 2.7%). To define whether this low activity is due to additive elements which are not contained within the 200 bp or due to repetitive motifs, which just multimerize the effect, several constructs with multimerizations of *'2.2-2.4'* (Figure 4.22.A) were designed and transfected into the cell line $\psi V^{-}IgL^{-}$. After identification of targeted transfectants, two individual primary clones of each mutant were subcloned and analyzed by FACS 14 days hereafter.

Doubling the sequence already enhanced the median decreased green fluorescence to 1.6% and 2.2%. A quadruplication of the fragment resulted in a median of 3.6% and 4%, which is even higher than observed for the *'2-3'* and *'0-4'*. Repeating the sequence 14 times did not further contribute to enhance HM.

Doubling of the *'2-3'* fragment led to a 1.5 fold increased diversification of the GFP (Figure 4.22.B) compared to the single fragment. The 2 kb fragment *'2x [2-3]'* with 1.8% and 4.6% has less HM activity than the 800 bp fragment *'4x [2.2-2.4]'* with 4% and 3.6%.

'2.2-2.4' is the only sequence within *'2-3'* important for HM. The sequence does not need additional elements, but repeating the sequence can also increase HM. Probably there is a saturation level, and HM can not be increased beyond this limit.

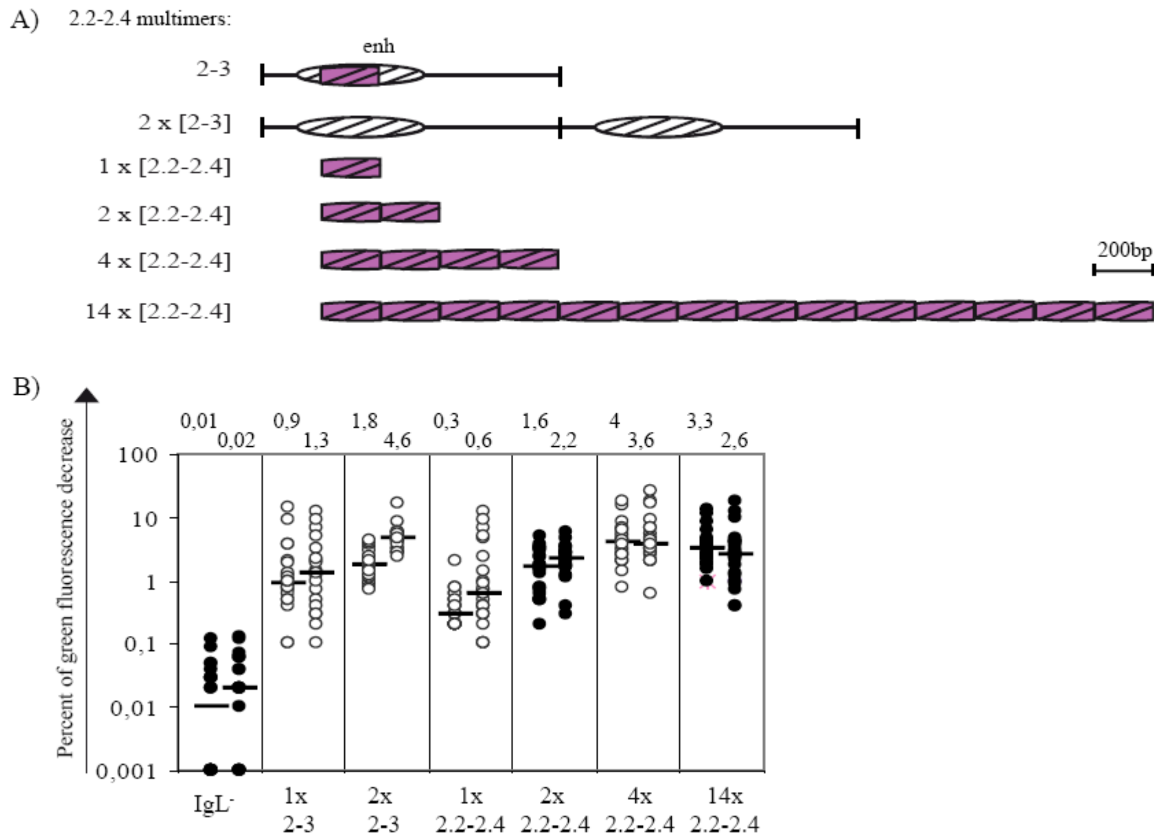


Figure 4.22 Multimerization of '2-3' and '2.2-2.4'.
A) A physical map showing the position of '2.2-2.4' within the ‚2-3'fragment and the generated multimers. The enhancer (enh) is indicated as a striped oval. The position of the '2.2-2.4'within the enhancer is indicated with a violet striped box. B) Fluctuation analysis of subclones. Only the name of the reconstituted fragment and not the full name of the transfectants is indicated for clarity. The median of green fluorescence decrease for each clone is written above the bars.

4.3.5 The *'HyCorE'* sequence *'2.2-2.4'* starts HM at non-Ig loci

Results until now show that '2.2.-2.4' acts like a true *'HyCorE'* as this DNA element is able to start HM and a repeat of the motifs can also enhance HM. Another feature of a supposed *'HyCorE'* is the ability to target HM to a specific locus. A following experiment should reveal, if *'2.2.-2.4'* is able to start HM also at a non-Ig loci. The GFP2 reporter together with '2.2.-2.4' was targeted to the Bach2 locus (Figure 4.23.A). The resulting clone $AID^{R1}IgL^{-}Bach2^{+/2.2\text{-}2.4,\ GFP2}$ was subcloned and analyzed by FACS. The median of decreased green fluorescence was compared to that of the clone $AID^{R1}IgL^{2.2\text{-}2.4,\ GFP2}$. Insertion of the GFP2 reporter alone at the position of the Bach2 locus ($AID^{R1}IgL^{-}Bach2^{+/GFP2}$) did not lead in mutations in the GFP transgene (personal communication with A. Blagodatski, Blagodatski et al., 2009). Results in Figure 4.23.B ('*Bach2 2.2-2.4'*) show that the GFP was a target for mutation, if it was inserted together with the fragment *'2.2-2.4'*. With 1% decrease in green fluorescence, HM activity is higher than in *'IgL 2.2-2.4'*, where the fragment *'2.2-2.4'* is inserted at the position of the deleted IgL locus. Both values are

significantly different from the *'IgL-'*, a cell line in which the GFP2 reporter is not hypermutating due to the deletion of the entire IgL locus (Figure 4.23.B).

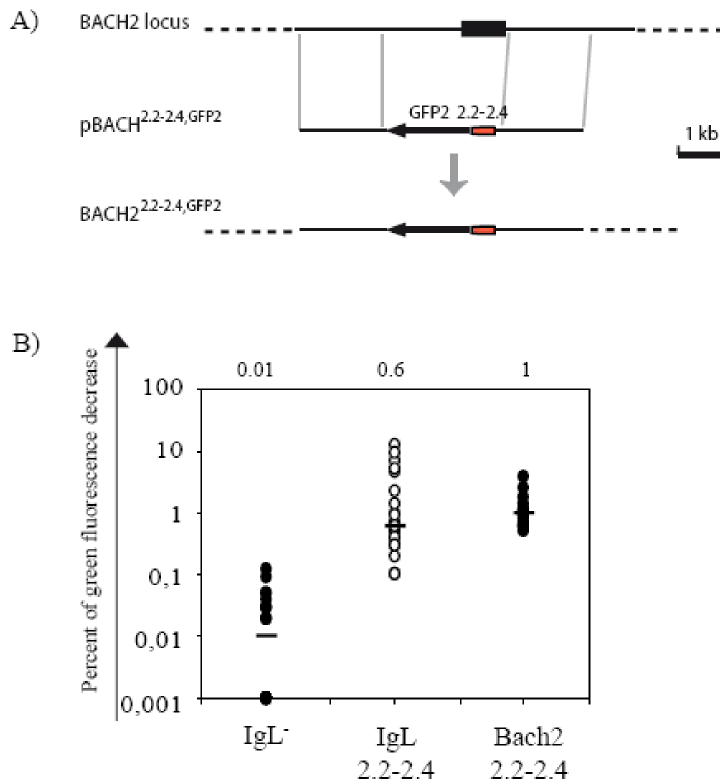


Figure 4.23 Insertion of *'2.2-2.4'* at the non-hypermutating locus Bach2.
A) A physical map showing the position of the GFP2 reporter and the *'2.2-2.4'* fragment targeted to the Bach2 locus. The GFP2 is indicated by an arrow and *'2.2-2.4'* is marked by a red box. B) Fluctuation analysis of subclones. Only the letter and position of the reconstituted fragment and not the full name of the transfectants is indicated for clarity. The median of GFP decrease for each clone is written above the bars.

4.3.6 The *'HyCorE'* sequence *'2.2-2.4'* is conserved in other species

Given the conservation of AID mediated *Ig* gene diversification during vertebrate evolution, the identification of the chicken IgL *'HyCorE'* should also be of relevance for mammals. Searches for cis-acting HM regulatory sequences in transgenic mice showed that a λ 3'enhancer (Klotz and Storb, 1996; Kong et al., 1998) and multiple sequences from the Igκ locus including E_i/MAR, Cκ and its flanks and the E3' enhancer (Klix et al., 1998) conferred HM activity. Indeed, the location and functional characteristics of these regions appear to be similar to what we observed for the chicken IgL enhancer. They are able to start HM in transgenic sequences which are not related to Ig genes, diversification occurs independently from the integration site of the transgene in the genome and interestingly, the HM enhancing sequences function even if they are transposed or inverted (Klix et al., 1998). The latter feature goes along with findings in our lab, where we could see that

the *'W'* fragment confers HM activity of the GFP reporter transgene even over long distances in both directions (Blagodatski et al., 2009).

However, results on the Igκ and Igλ loci in mouse are conflicting, as the studies show significant differences in mutation frequencies and deletion of the Igκ enhancers could not prevent HM (Inlay et al., 2006; van der Stoep et al., 1998). The difficulty to prove unambiguously the existence of HM activator sequences may relate to the large size of the murine Ig loci (several Megabases) and the fact that enhancers of HM seem to be composed of multiple interacting regions. Each of the murine Ig loci possesses at least two enhancers at different positions, therefore a murine *'HyCorE'* may be composed of multiple discontinuous sequences. Beyond this, during evolution the sequence motifs could be degenerated.

However, it was possible to identify the *'2.2-2.4'* in the strongly related species duck and turkey (personal communication with Arundhathi Sriharshan; Fig.4.24.A).

Strikingly, transfection of the turkey and duck sequences together with the GFP2 reporter into $\psi V^{-}IgL^{-}$ at the position of the IgL locus yielded in an activation of HM. The median decrease of green fluorescence with 1% for turkey *'2.2-2.4'* and 0.9% for duck *'2.2-2.4'* is similar to that of the chicken *'2.2-2.4'* with 0.6% (Fig.4.24.B).

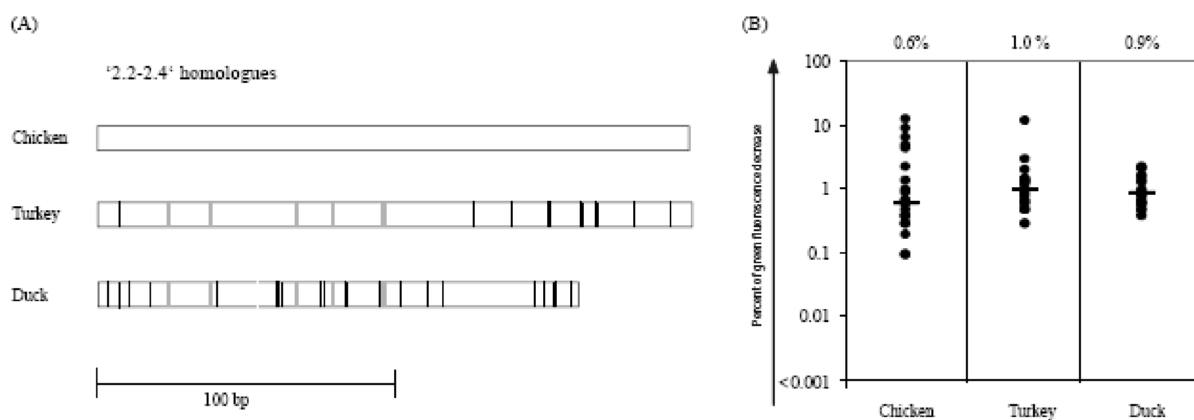


Figure 4.24 Analysis of the '2.2-2.4' homologues in turkey and duck.
A) A physical map of the *'2.2-2.4'* fragment in chicken and its homologous counterparts in turkey and duck. The white area displays the sequence conserved between the species. The black bars indicate the sequence differences between the species. Grey bars are conserved sequences in duck and turkey but differ from chicken. B) Fluctuation analysis of the subclones (n=24) of one primary clone. The median of decreased green fluorescence for each clone is written above the bars.

4.3.7 GFP expression level – variations in HM of the GFP2 transgene are independent from GFP transcription

As there is a strong correlation between transcription and HM, I have to check whether the HM activities of the produced clones are really due to a role of the cis-elements, or if it is only due to changes in transcription of the GFP2 reporter.

Duplication of the *'2-3'* led to a 1.9 fold increase of green fluorescence compared to IgL^{-} (Figure 4.25). This fragment contains the complete IgL enhancer and presumably the RSV promoter got additional stimulation

due to the doubling. Hence, changes in diversification within this clone could be due to changes in transcription. $\psi V^{-}IgL^{14x2.2\text{-}2.4,GFP2}$ was 1.5 fold decreased. This might explain why no further stimulation of HM in this clone was reached (see Figure 4.22), but it is difficult to say why the GFP expression level is decreased if the sequence *'2.2-2.4'* is repeated 14 times. Probably the targeted integration of the construct led to changes in chromatin structure which might interfere with the diversification activity. The construct also can not be sequenced because of the many repeats. Changes in AID expression level are unlikely, as the clones are cultured in mycophenolic acid medium and resistance to drug is an indicator for positive AID expression. The average green fluorescence in the main population of $\psi V^{-}IgL^{2.2\text{-}2.4,GFP2}$, $\psi V^{-}IgL^{2x2.2\text{-}2.4,GFP2}$ and $\psi V^{-}IgL^{4x2.2\text{-}2.4,GFP2}$ were all in the same range and not significantly different from $\psi V^{-}IgL^{-,GFP2}$ (Figure 4.25). A correlation of transcription and HM can not be found within these clones. This suggests, the increase of HM level indeed was due to a true *'HyCorE'* which contains cis-elements that stimulate solely diversification.

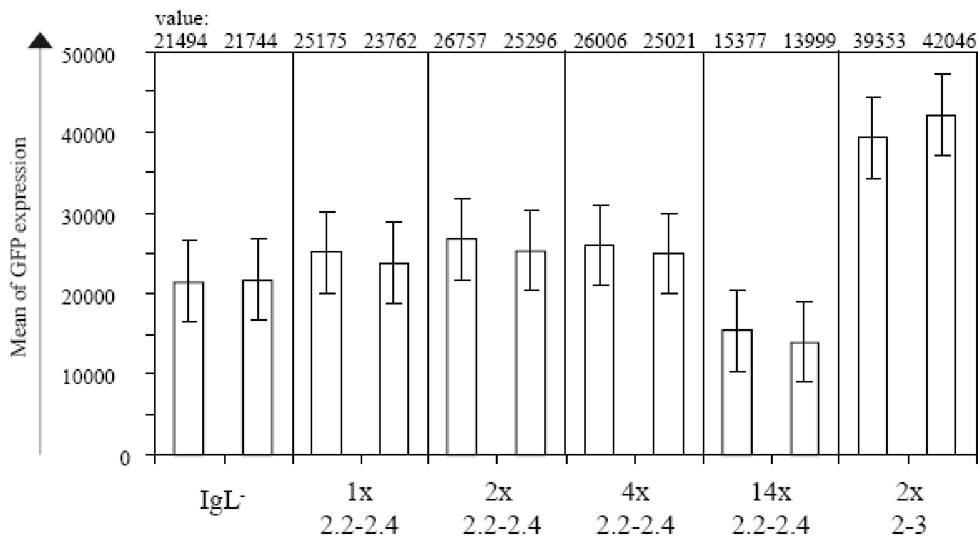


Figure 4.25 GFP expression levels for selected clones.
Two primary clones of each mutant were examined. The mean value was calculated for 24 subclones for each clone and is described above the bars. The standard deviation is indicated.

GFP levels were shown to be not significantly different when analyzed by FACS. These observations were confirmed independently by Hiroshi Arakawa for the clones $AID^{R1}IgL^{A,\,GFP2}$, $AID^{R1}IgL^{B,\,GFP2}$, $AID^{R1}IgL^{C,\,GFP2}$, $AID^{R1}IgL^{D,\,GFP2}$, $AID^{R1}IgL^{E,\,GFP2}$, $AID^{R1}IgL^{F,\,GFP2}$, $AID^{R1}IgL^{G,\,GFP2}$, $AID^{R1}IgL^{I,\,GFP2}$, $AID^{R1}IgL^{K,\,GFP2}$, $AID^{R1}IgL^{L,\,GFP2}$, $AID^{R1}IgL^{M,\,GFP2}$, $AID^{R1}IgL^{N,\,GFP2}$ and $AID^{R1}IgL^{S,\,GFP2}$ for a semiquantitative RT-PCR (Blagodatski et al., 2009). Similar levels of steady-state GFP2 transcripts were confirmed for the ten clones by RT-PCR (Figure 4.26), but the levels of hypermutation differ from clone to clone (Figure 4.13.B). There was no correlation between transcription and diversification of the GFP transgene.

Hence, changes in variation of HM activity in the different clones were independently from transcription of

the GFP2 reporter.

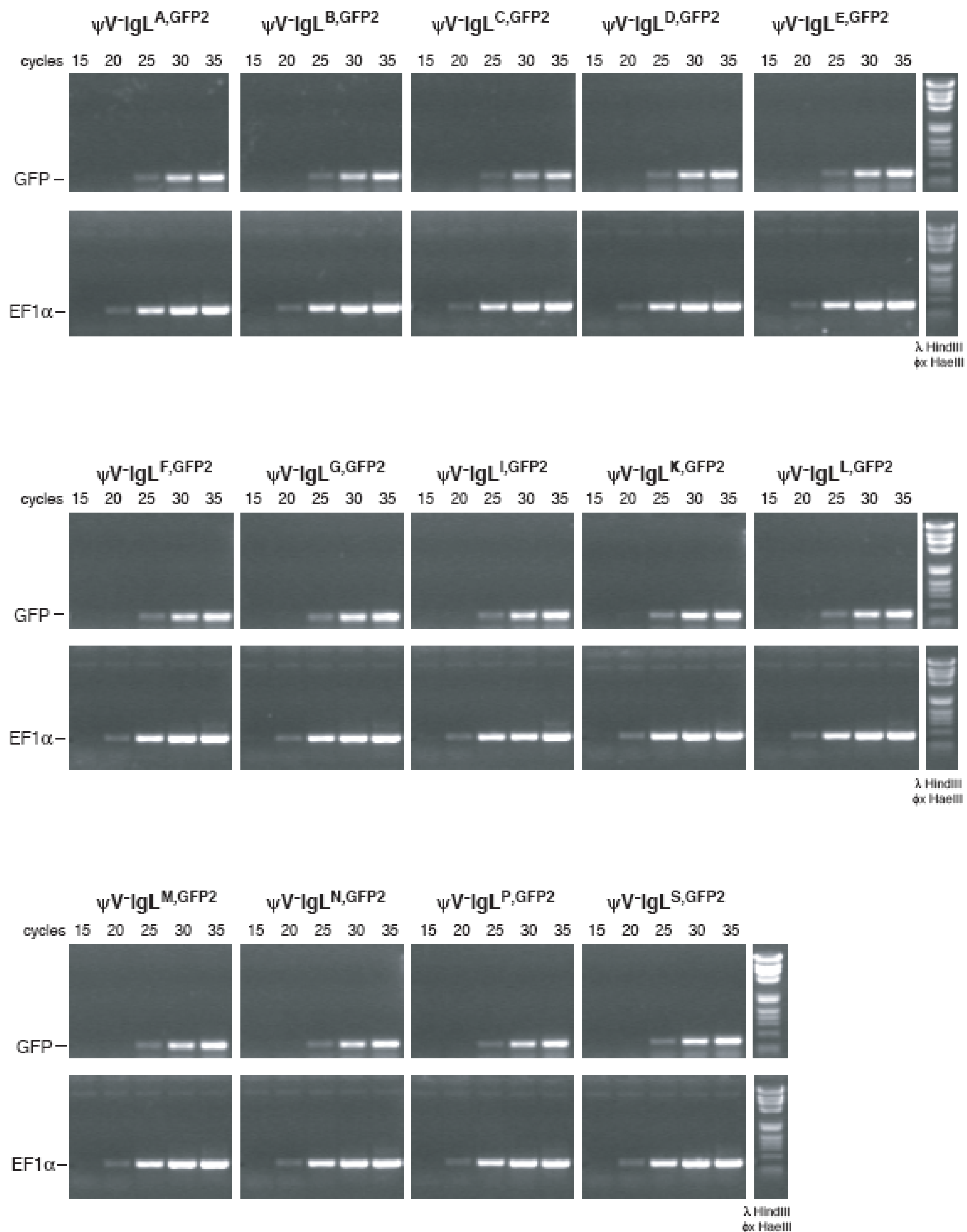
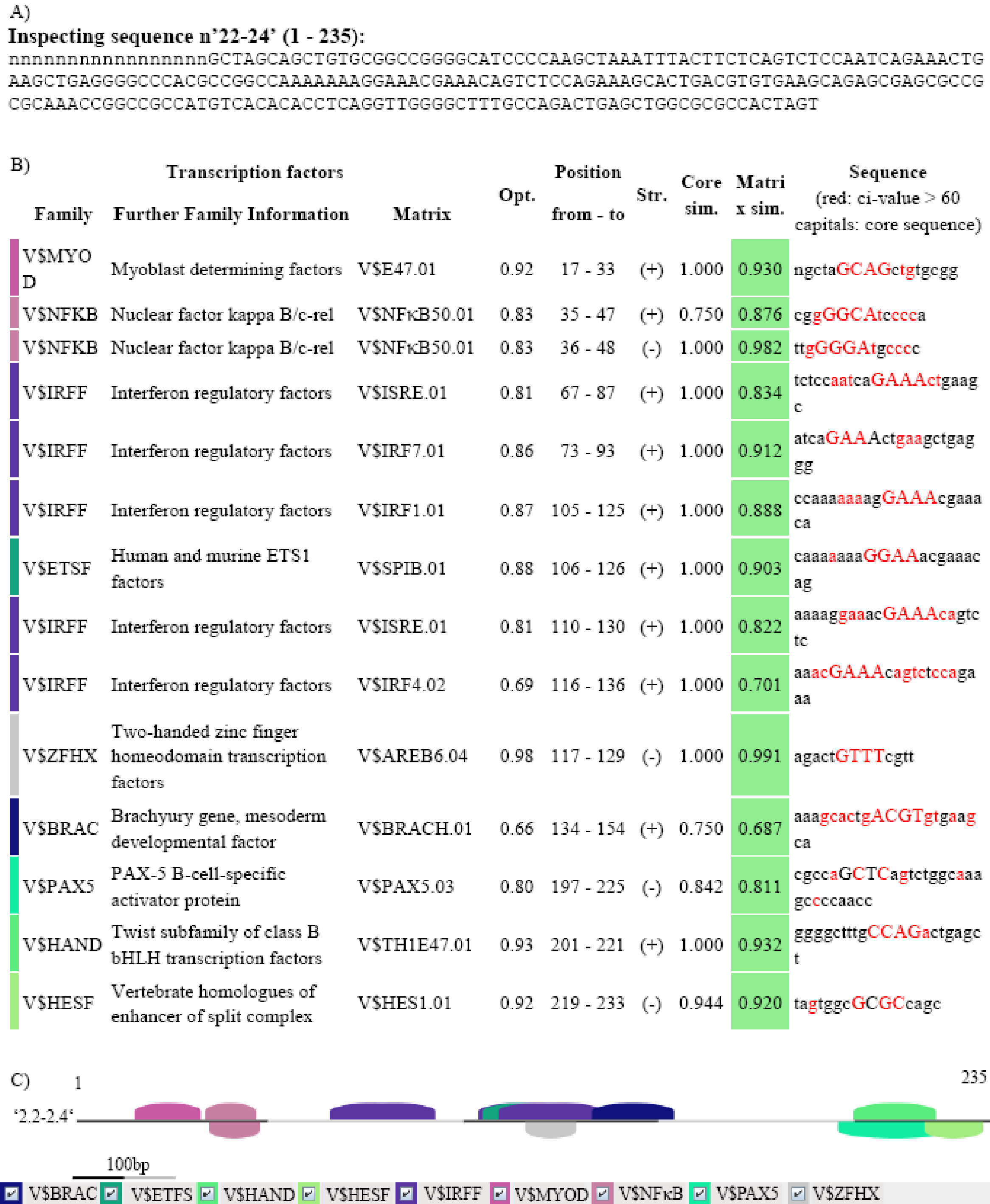


Figure 4.26 Semiquantitative RT-PCR for selected Ig deletion clones.
Adapted from Blagodatski et al. (2009). Measurement of steady state transcript levels of the GFP gene. The house keeping gene EF1α was chosen as for control. The number of PCR cycles is indicated.

4.4 *Evaluation of cis-elements using boinformatical tools*

To search for putative cis-elements (matrices) within the *'2.2-2.4'*, I used MatInspector at Genomatix (www.genomatix.de; Quandt et al., 1995; Cartharius et al., 2005). This is a search module for new transcription factor binding sites using TRANSFAC matrices, a database on eukaryotic cis-acting regulatory DNA elements and trans-acting factors. For my study, I used the vertebrate group of the matrix family library which contains 519 matrices in 154 families. 4712 transcription factors are known to bind to the indicated matrices. Only matrices known to have regulatory functions in the immune system were taken into account to narrow the search to factors which are expressed in B cells.

The database was screened with the sequence of *'2.2-2.4'* (Fig.4.27.A). Results show in the middle of the sequence a cluster of cis-elements binding Interferon regulatory factors (IRF) and a DNA binding site for SpiB (Fig.4.27.B and C). A single binding motif for the p50 subunit of NFκB is found at the 5'end of the sequence, and a Pax5 binding site can be found at the 3'end. Interestingly, there are also two putative E47 binding motifs, one E-box motif binding E47 homodimers at the very 5' end with the sequence CAGCTG and a more degenerated motif at the 3' end (CAGACTG) which is thought to bind Thing1/E47 heterodimers. The E-box CAGCTG could be a candidate to bind the E2A transcription factors.



A)

Inspecting sequence n'22-24' (1 - 235):

nnnnnnnnnnnnnnnnnGCTAGCAGCTGTGCGGCCGGGGCATCCCCAAGCTAAATTTACTTCTCAGTCTCCAATCAGAAACTG
AAGCTGAGGGGCCCACGCCGGCCAAAAAAAGGAAACGAAACAGTCTCCAGAAAGCACTGACGTGTGAAGCAGAGCGAGCGCCG
CGCAAACCGGCCGCCATGTCACACACCTCAGGTTGGGGCTTTGCCAGACTGAGCTGGCGCGCCACTAGT

B)

Transcription factors		Matrix	Opt.	Position	Str.	Core sim.	Matrix sim.	Sequence (red: ci-value > 60 capitals: core sequence)
Family	Further Family Information			from - to				
V$MYOD	Myoblast determining factors	V$E47.01	0.92	17 - 33	(+)	1.000	0.930	ngctaGCAGctgtgcgg
V$NFKB	Nuclear factor kappa B/c-rel	V$NFκB50.01	0.83	35 - 47	(+)	0.750	0.876	cggGGCAtcccca
V$NFKB	Nuclear factor kappa B/c-rel	V$NFκB50.01	0.83	36 - 48	(-)	1.000	0.982	ttgGGGAtgcccc
V$IRFF	Interferon regulatory factors	V$ISRE.01	0.81	67 - 87	(+)	1.000	0.834	tctccaatcaGAAActgaagc
V$IRFF	Interferon regulatory factors	V$IRF7.01	0.86	73 - 93	(+)	1.000	0.912	atcaGAAActgaagctgagg
V$IRFF	Interferon regulatory factors	V$IRF1.01	0.87	105 - 125	(+)	1.000	0.888	ccaaaaaaagGAAAcgaaaca
V$ETSF	Human and murine ETS1 factors	V$SPIB.01	0.88	106 - 126	(+)	1.000	0.903	caaaaaaaGGAAacgaaacag
V$IRFF	Interferon regulatory factors	V$ISRE.01	0.81	110 - 130	(+)	1.000	0.822	aaaaggaaacGAAAcagtctc
V$IRFF	Interferon regulatory factors	V$IRF4.02	0.69	116 - 136	(+)	1.000	0.701	aaacGAAAcagtctccagaaa
V$ZFHX	Two-handed zinc finger homeodomain transcription factors	V$AREB6.04	0.98	117 - 129	(-)	1.000	0.991	agactGTTTcgtt
V$BRAC	Brachyury gene, mesoderm developmental factor	V$BRACH.01	0.66	134 - 154	(+)	0.750	0.687	aaagcactgACGTgtgaagca
V$PAX5	PAX-5 B-cell-specific activator protein	V$PAX5.03	0.80	197 - 225	(-)	0.842	0.811	cgccaGCTCagtctggcaaagccccaacc
V$HAND	Twist subfamily of class B bHLH transcription factors	V$TH1E47.01	0.93	201 - 221	(+)	1.000	0.932	ggggctttgCCAGactgagct
V$HESF	Vertebrate homologues of enhancer of split complex	V$HES1.01	0.92	219 - 233	(-)	0.944	0.920	tagtggcGCGCcagc

C)

Figure 4.27 Results for the identification of putative cis-elements using MatInspector. A) Sequence of '2.2-2.4' used for the database screen. B) Results of putative cis-elements and their binding transcription factors (TF). TF families are depicted in a color code. The matrix name has an identifier that indicates the matrix family group (V$=Vertebrates), followed by an acronym for the TF the matrix refers to, and a consecutive number discriminating between different matrices for the same TF. The Ci-vector (consensus index vector) for the matrix represents the degree of conservation of each position within the matrix with the maximum Ci-value of 100 for a total conservation. C) Representation of the putative binding sites plotted onto the sequence.

5 Discussion

5.1 *E2A is involved in AID-mediated diversification*

In 2003 a study could identify a role for E-box motifs in HM (Michael et al., 2003). The motif CAGGTG was found to enhance HM of a mouse Igκ transgene without enhancing transcription. Using the method of EMSA (electromobility shift assay) the transcription factor E47 was shown to bind this motif. E2A proteins are necessary for early B cell commitment and B cell development. B cell development in E2A null mice arrests prior to Ig gene activation and hence onset of HM (Bain et al., 1994). Therefore, in vivo mouse studies are not suitable to examine a role of E2A in HM.

The DT40 cell model offers the possibility to directly study the role of E2A in HM. The results of the inactivation of E2A clearly showed that the products of the gene, E12 and E47, are necessary for strong HM to occur. In the homozygous knockout the mutation rate is significantly reduced, and overexpression of either E12 or E47 could rescue the phenotype. The phenotype was independent from AID and Ig expression and did not alter the mutation pattern. Taken together, the results suggest a role for E2A transcription factors in the recruitment of AID.

E-proteins are known to function mainly as activators of transcription in B cells (Massari et al., 2000). Therefore downregulation of the HM level in an E2A knockout clone could be due to a decrease in IgL transcription. The results confirm that the involvement of E2A is not in regulation of transcription of the IgL locus. Thus RT-PCR analysis of expression of C_L region transcripts show similar levels for all cell lines, the progenitor cell line, the E2A knockout and also the E12 and E47 overexpressing clones. Similar results have been produced from studies in mature B cells, where Id3, a potential antagonist of E2A transcription factors, has no influence on IgM production (Quong et al., 1999). Presumably E2A is not needed for Ig gene transcription.

Downregulation of AID gene expression itself is another mechanism which may be responsible for regulating HM. RT-PCR of the AID transgene revealed similar transcript levels for the progenitor cell line and both E2A

knockout and E12 overexpressing clones. Surprisingly, AID was downregulated in the E47 overexpressing clone. E47 is known to regulate transcription of the AID gene in mouse (Sayegh et al., 2003), but it is acting as an activator rather than inhibiting transcription. It is unlikely that AID is regulated by E47 in the DT40 cell line used for this study, as it was expressed constitutively under the strong β-actin promoter. The differences may be due to clonal variations arising due to transfection, or there is an unknown regulation by E47 in repressing or of E12 in enhancing via binding to the enhancer sequence contained within the β-actin promoter. Although AID mRNA is decreased in the $AID^{R}\psi V^{-}E2A^{RtE47}$ clone, mutations are still occurring at a high rate as shown in the Ig mutation assay and by sequencing. This could be due to a stable level of the AID protein. Presumably, there is a threshold level for AID expression which is saturated in all cell lines. Or the HM enhancing effect of E47 overexpression may overcompensate for the low expression level of AID. As it is a negative correlation of AID level and HM level, this result does not interfere with the main conclusion of the study, that E2A is needed for strong HM and this is not mediated through elevating IgL transcription or the level of AID expression.

The Ig mutation assay was done to compare the HM activity of the E2A knockout mutants and the E12 and E47 overexpressing clones to the precursor cell line and a non-hypermutating cell line. The study showed that expression of either the E12 or the E47 cDNA can complement the Ig HM defect in E2A-disrupted cells to wild-type levels. This indicates that the decreased HM activity is indeed due to the lack of E2A expression.

Ig HM is only initiated by AID-mediated cytidine deamination and depends on further processing of the resulting uracils by uracil DNA glycosylase (UNG; Saribasak et al., 2006; Di Noia and Neuberger, 2002) and error-prone polymerases (Simpson and Sale, 2003; Ross and Sale, 2006). It cannot be ruled out that the reduction of HM in E2A-negative cells is caused by effects downstream of AID, such as reduced activity of UNG or RNA polymerases. To elucidate the phenotype of the E2A mutants in more detail an analysis of the mutation pattern was necessary. Sequencing showed no significant difference in the mutation spectrum or hot spot preferences between the progenitor clone $AID^{R}\psi V^{-}$ and the mutants $AID^{R}\psi V^{-}E2A^{-/-}$, $AID^{R}\psi V^{-}E2A^{RtE12}$ or $AID^{R}\psi V^{-}E2A^{RtE47}$. The spectrum for $AID^{R}\psi V^{-}E2A^{RtE12}$ cells exhibited some differences in the frequency of mutations at A bases compared to the progenitor. $AID^{R}\psi V^{-}E2A^{RtE12}$ had no mutations at A bases, whereas $AID^{R}\psi V^{-}$ had mutations from A-to-G and A-to-T, although these mutations were less prominent. Mutations from G-to-A were slightly elevated. For the precursor cell line $AID^{R}\psi V^{-}$ were 4.5 times more sequences screened than for the $AID^{R}\psi V^{-}E2A^{RtE12}$ clone. Additional sequencing for the latter clone might diminish the numerical differences between the two clones. In the $AID^{R}\psi V^{-}E2A^{RtE47}$ clone are mutations from C-to-A elevated compared to the progenitor clone. This could be due to an early HM event during expansion of the clone, as most of the C-to-A mutations occurred at the same position. However, one has to take into account that the overexpression of E12 or E47 alone does not completely recreate the physiological situation of the

progenitor clone, where both proteins are expressed at the same time, and can therefore serve only as a partial reconstitution. This might be also reflected in the mutation pattern of the clones.

The absence of consistent changes in the mutation spectrum of the E2A-deficient or -complemented clones compared with the wild-type E2A progenitor clone argues against the possibility of E2A being involved in regulation of HM downstream of AID. Other DT40 mutants that interfere with the conversion of AID-induced uracils into hypermutations showed clear changes in the mutation spectrum. These include a transitions bias in UNG-deficient cells (Saribasak et al., 2006) and selectively reduced C-to-G and G-to-C mutations after disruption of the DNA polymerase REV1 (Ross and Sale, 2006). Compared to UNG$^{-/-}$ clones, where the mutation spectrum is completely perturbed, the E2A knockout and reconstituted clones reflect the preferences of the progenitor clone. This suggests the more likely scenario of an influence of E2A on Ig HM upstream of AID action.

E47 homodimers are known to be B-cell specific (Shen et al., 1995). However, the results suggest that E12 can resubstitute the phenotype, too. To become active, the E-proteins need to dimerize on their DNA motifs known as E-boxes. The E-box motif CANNTG is found in Ig enhancer and promoter sequences of human and mouse (Staudt et al., 1991; Nelsen et al., 1992) and also in the previously described enhancer region of DT40 (Bulfone-Paus et al., 1995). To find out the exact sequence and exact position of the E-box motif would help to clarify the role of E2A transcription factors.

5.2 E2A transcription factors bind to their DNA recognition motif to stimulate HM

An influence of E2A transcriptional regulators on HM upstream of AID is consistent with the known role of the proteins E12 and E47 as transcription factors which recognize E-box consensus motifs in the enhancers of their target genes. Because E-box sequence motifs are present in the enhancers of Ig and hypermutating non-Ig genes (Kotani et al., 2005), it is probable that E12 and E47 influence Ig HM by binding to the Ig enhancers. This is consistent with the observation that the introduction of an E-box sequence stimulated the mutation rate of an Igk mouse transgene (Michael et al., 2003). It remains unknown how the binding of the E2A encoded proteins to the Ig enhancers might stimulate Ig HM, but the effect is apparently not due to the stimulation of Ig transcription activity, because no decrease of IgL transcription was observed in the E2A-negative mutant.

Nonphysiologic expression of AID in T cells increased the mutation rate in the same subset of non-Ig genes (the oncogenes c-myc and pim1) that were previously shown to be mutated in B cells (Kotani et al., 2005). HM highly selective targeted genes in T cells, because oncogenes which are not mutated in B cells did also not mutate in T cells. Thus, the targeting of HM was similar like in B cells thereby suggesting a regulation through

cis-elements contained in the genomic sequence of the mutated genes. The results suggest that the regulatory machinery for HM may not be B cell-specific but shared by T cells. Interestingly, all mutated genes share the E47 binding motif somewhere in the genes and their flanks. Moreover, many B lineage-specific genes regulated by E47, like the proto-oncogenes BCL6, MYC, RhoH/TTF (ARHH), PAX5 and genes encoding the B cell receptor (BCR) accessory proteins B29 (Igα, CD79b) and mb1 (Igβ, CD79a), are known to be target of HM (Migliazza et al., 1995; Shen et al., 1998; Pasqualucci et al., 2001; Gordon et al., 2003). Beyond this, also T lineage-specific genes (cd4 and cd5) regulated by E47 accumulate HM-like mutations in the T cell lymphoma of AID transgenic mice (Kotani et al., 2005). In our study, the targeting of the Ig locus by HM is strongly diminished in E2A knockout mutants. Taken these results together, E2A transcription factors seem to be potent regulators in the targeting of HM as not only the Ig genes, but also other genes which are regulated by E47 are prone to HM.

5.3 *E2A transcription factors regulate gene conversion and chromatin remodeling*

In wild-type DT40 cells, which in contrast to the $AID^{R}\psi V^{-}$ clone harbor pseudo-V gene conversion donors upstream of the rearranged IgL chain VJ segment, AID activity predominantly leads to Ig gene conversion. If Ig hypermutation is enhanced by E2A at a step before AID-mediated cytosine deamination, one would expect that E2A expression will stimulate not only Ig hypermutation but also Ig gene conversion. This prediction has been confirmed by reported increased rate of gene conversion in DT40 cells that overexpress the E47 cDNA and diminished gene conversion in cells that have the endogenous E2A gene knocked out (Conlon and Meyer, 2006; Kitao et al., 2008). Steady state levels of Ig transcription and AID expression remained unaffected in these cells. This confirms our study on the E2A knockout and the E12 and E47 reconstituted clones.

AID expression level in the E47 reconstituted clone was reduced as shown by RT-PCR. However, Ig mutation assay and sequencing revealed that the clone is still hypmutating at a high rate. Most likely, the AID level is reduced due to a loss of a randomly inserted AID cDNA cassette and not due to an influence of E47 on AID expression.

One of the most intriguing possibilities would be that the E2A-encoded proteins specifically recruit AID and the HM machinery to the Ig loci.

E-proteins are involved in chromatin remodeling. E47 associates with coactivators such as histone acetyltransferases and collaboratively regulates B cell development (Bradney et al., 2003). It is known to interact directly with the hyperacetylating SAGA-complex (Massari et al., 1999). Hyperacetylation of histones in the Ig variable region gene locus increased the frequency of HM in a B cell line (Woo et al, 2003). E47 and

E12 are likely involved in opening of chromatin around Ig genes (Choi et al., 1996). Hence, a hypothesis would be that E12 and E47 make the locus more accessible for AID by opening the chromatin structure.

Histone acetylation also regulates expression of genes in a way that it changes chromatin structure from transcriptionally inactive heterochromatin to transcriptionally active euchromatin, or it affects the DNA binding of transcription factors.

Recently, a group showed that E2A regulated gene conversion by H4 acetylation (Kitao et al., 2008). Knockout of the E2A gene in the DT40 variant Cre1 showed strongly diminished gene conversion (GCV) rates and impaired H4 acetylation. Treatment of the cells with the histone deacetylase inhibitor Trichostatin A (TSA) was able to reverse the defect in GCV and H4 acetylation. The group suggests a role for E2A in targeting AID to the Ig locus by maintining histone acetylation.

In our lab, I used the cell line $AID^{R}IgM^{-}$ (which has the same features like DT40Cre1 but the endogenous AID is deleted and reconstituted by a cDNA expression cassette) for treatment with TSA and could observe a strong increase in GCV activity. However, treating $AID^{R}\psi V^{-}$, which is the precursor clone of my E2A study, and $AID^{R}\psi V^{-}E2A^{-/-}$ with TSA did not lead in increased HM (Supplemental Figure chapter 8.4). GCV and HM are both AID-regulated processes and thus the effect of E2A-induced H4 acetylation in the study of Kitao is unlikely to be directly related to AID activity. It seems more likely, that the acetylation accelerates the process of homologous recombination, which is necessary for GCV to take place but not HM. Probably in that way, that it makes the locus more accessible for DNA repair proteins (Murr et al., 2006). As E2A is not only regulating GCV but also HM, this type of acetylation alone can not explain the effect of E2A proteins on AID-mediated diversification processes. Other mechanisms beyond H4 acetylation could target AID to the Ig locus, E2A proteins could directly recruit AID or they play an indirect role for example by modifying DNA binding sites of factors involved in AID recruitment.

5.4 A GFP reporter revealed cis-trans interactions balancing AID-mediated HM

The study of Kotani et al. (2005) revealed not only the E-box motif, but also other cis-elements being shared between mutated genes in B cell and T cell lymphoma.

Understanding the nature and function of these elements will be of great interest for elucidating the targeting mechanism of HM.

The mapping of the IgL locus of DT40 was used to identify cis-acting HM elements and their potential binding factors. To uncouple effects on HM from effects on transcription, a reporter using GFP as HM target was selected. For this study, the GFP2 reporter construct from Arakawa et al. (2008) was modified. A disadvantage of this construct might be, that it functions as a loss-of-activity screen scoring mutations from GFP(+) to GFP(-) as the read-out of HM activity. Theoretically a silencing of the GFP expression itself can

occur, and needed to be excluded. Whilst a GFP reversion forward assay (Bachl et al., 2001; McBride et al., 2004) insensitive to silencing or other transcriptional effects exists, the GFP2 assay has the great advantage that it covers a far greater spectrum of mutations, including insertions, deletions and duplications. These are mutations that are naturally apart of the HM process occurring, whereas a GFP reversion forward assay measures only mutations at a predefined artificial stop codon. Both assays are not able to detect null mutations without sequencing, because null mutations do not lead in a change in phenotype and thus they can not be detected by FACS analysis.

To exclude transcriptional silencing of the construct, an additional experiment was done in our lab (Blagodatski et al., 2009). High (GFP-high) and low (GFP-low) green fluorescence cells were sorted from one of the $\psi V^{-}IgL^{GFP2}$ primary transfectants and their GFP mRNA levels were analyzed by semi-quantitative RT-PCR. Although GFP-low cells showed much lower green fluorescence than GFP-high cells, the levels of GFP mRNA were comparable between sorted GFP-low, GFP-high and non-sorted cells. The results demonstrated that the decrease of green fluorescence is highly unlikely to be caused by the silencing of GFP gene expression.

5.4.1 The *'HyCorE'* acts independently from transcription

Changes in HM level can be due to changes in transcription level (Bachl et al., 2001). But a true *'HyCorE'*, would elevate diversification levels independently from transcription.

GFP expression levels of the clones $\psi V^{-}IgL^{-,GFP2}$, $\psi V^{-}IgL^{2.2\text{-}2.4,GFP2}$, $\psi V^{-}IgL^{2x2.2\text{-}2.4,GFP2}$ and $\psi V^{-}IgL^{4x2.2\text{-}2.4,GFP2}$ were shown to be not significantly different when analyzed by FACS.

These observations were confirmed for the clones $AID^{R1}IgL^{A, GFP2}$, $AID^{R1}IgL^{B, GFP2}$, $AID^{R1}IgL^{C, GFP2}$, $AID^{R1}IgL^{D, GFP2}$, $AID^{R1}IgL^{E, GFP2}$, $AID^{R1}IgL^{F, GFP2}$, $AID^{R1}IgL^{G, GFP2}$, $AID^{R1}IgL^{I, GFP2}$, $AID^{R1}IgL^{K, GFP2}$, $AID^{R1}IgL^{L, GFP2}$, $AID^{R1}IgL^{M, GFP2}$, $AID^{R1}IgL^{N, GFP2}$ and $AID^{R1}IgL^{S, GFP2}$ by a semiquantitative RT-PCR in our lab.

All of the clones showed various mutation levels of the GFP transgene, but the transcription levels were not significantly different. The result suggests, that GFP expression level and GFP hypermutation level are two independent events.

Fragments *'A'*, *'B'*, *'C'*, *'D'*, *'N'* and *'P'* contain no enhancer sequence, whereas the other fragments *'E'*, *'F'* and *'G'* contain the enhancer sequence in a genuine distance like in the complete 9.8 kb *'W'* fragment and the fragments *'I'*, *'K'*, *'L'*, *'M'* and *'S'* contain also the enhancer, but the distance to the promoter is shortened as several kb of the 5' end of the IgL locus are missing in these clones. Independently from the presence and the distance of the enhancer, all clones showed similar GFP transcript levels. Distance effects and the enhancer seem not to influence RSV promoter mediated transcription of the GFP transgene to a significant content.

5.4.2 A 9.8 kb fragment of the IgL locus is able to start diversification at non-hypermutating loci

Before starting the fine-mapping of the IgL locus, some experiments were done to underline the importance of the genomic sequence. Deletion and reinsertion of a 9.8 kb fragment referred to as *'W'*, revealed this IgL fragment as being necessary and sufficient for HM to occur. In the absence of *'W'*, the GFP2 reporter did not mutate at all. In the presence of *'W'* the transgene was prone to mutations. The mutations were AID dependent, as they did not occur when AID was deleted.

The results give rise to the conclusion, that *'W'* induces AID-mediated HM. The results could be explained by the presence of a cis-acting sequence which activated HM in a distance dependent manner. Without *'W'* HM does not occur at a detectable level, which means additional DNA elements upstream or downstream of *'W'* are not involved to an important degree. HM is activated and depends on DNA elements contained in *'W'*.

Other studies in our lab, showed that the *'W'* fragment is able to start HM also at other non hypermutating loci (Blagodatski et al., 2009). The GFP2 reporter construct was targeted integrated to the Bach2 locus. Integration of GFP2 only did not lead to mutation events. If GFP2 was integrated together with the *'W'* fragment, the transgene was target for HM in the presence of AID. These results demonstrated that the '*W*' fragment was able to activate *AID* mediated HM at loci which otherwise did not support HM.

Most likely, the *'W'* fragment contains cis-elements which serve as platform for the assembly of a HM enhancing complex which recruits AID to the transcription initiation site. Without this platform, it is no more possible to establish the complex and HM is inhibited. Alternatively, insertion of the platform at non-hypermutating loci will activate AID-mediated diversification there.

5.5 The 9.8 kb fragment 'W' was shortened to the 4 kb sequence '0-4'

How could a platform for an assembly of HM activating protein factors look like? Does it contain redundant elements or cooperative ones? To address this question, for the deletion study a strategy of deleting and inserting parts of the IgL locus was used.

A first deletion series was done to narrow the 9.8 kb fragment. The fragment *'W'* extends from the IgL transcription start site towards the next downstream gene and seems to be composed of multiple interacting regions. Whereas the 4 kb core sequence *'0-4'*, which includes the *IgL* enhancer, activates HM more than 100 fold above background level, the flanking regions possess less activity on their own, but stimulate HM when combined with the core sequence. *'0-4'* alone is sufficient to stimulate HM activity and therefore must contain all relevant DNA-elements. The flanking regions seem to contain redundant elements which possess only low activity on their own, but are able to trigger HM in combination with the DNA-elements contained in *'0-4'*.

The IgL enhancer stimulates transcription of the IgL locus and HM could be activated due to a stimulation of transcription of the GFP reporter by the IgL enhancer. RT PCR reveals similar GFP transcript levels for hypermutating and non hypermutating cell mutants, making it unlikely that the RSV promoter gets stimulated through the IgL enhancer. Thus, *'0-4'* stimulates HM activity independently from GFP transcription.

5.5.1 The 200 bp DNA element *'2.2-2.4'* is a true *'HyCorE'*

To use the fragment *'0-4'* for further deletion studies, enabled to narrow the *'HyCorE'* to even smaller sequences. The fragment *'0-2'*, which presents the first two kb of *'0-4'*, was not able to induce HM when inserted together with the GFP2 reporter at the position of the rearranged IgL locus. This might be due to silencing elements, but introducing 200 bp extending end deletions into the '0-2', did not lead to AID-mediated HM as well. Conclusively, the first 2kb of *'0-4'* are not at all relevant for AID-mediated HM.
In contrast, the 2 kb fragment *'2-4'* exhibited a similar HM activity like the complete *'0-4'*. A series of constructs introducing 200 bp extending end deletions into *'2-4'*, revealed a significant drop in HM after deletion of the first 800 bp (*'0.0-0.8'*). This supported the idea of *'0.0-0.8'* triggering diversification. Interestingly, this fragment also contains the sequence of the IgL enhancer.
Introduction of 1 kb deletions into the *'0-4'* showed, that the median decrease of green fluorescence in the 1 kb fragment *'2-3'* was two fold decreased compared to *'0-4'* and approximately 100 fold increased in comparison to *'IgL-'*. The p-value for this mutant was not significant, a hint that *'2-3'* contains important elements that support HM. This result goes along together well with the fact that all other Ig deletion fragments of this series which contained the sequence of *'2-3'*, exhibited HM levels similar to that of *'0-4'*, whereas the fragments which had the *'2-3'* deleted, supported HM to a significantly lower content. These fragments might contain redundant elements, which are not fully active on their own, but support the activity of a core element which seems to be located in the IgL enhancer.
A series of constructs introducing 200 bp internal deletions into the *'0-4'* was able to confirm the importance of the *'2-3'*. The 200 bp fragment *'2.2-2.4'* contains the first part of the IgL enhancer and its deletion in *'0-4'* strongly impaired the HM activity. The result is highly significant and hence the *'HyCorE'* sequence could be reduced to a 200 bp fragment presumably containing the active motif.
To transfect constructs with a larger internal deletion of 400 bp was thought to be able to confirm the results of the 200 bp internal deletion series. Surprisingly this was not the case. Deletion of *'2.0-2.4'* in the *'0-4'* fragment did not result in a drop of HM like it was the case for *'2.2-2.4'*.
Presumably, there exists a redundancy of the motifs and I suggest that deletion of another 200 bp created a new platform serving as *'HyCorE'*. Another possibility is suggested by insertion of the fragment *'2.2-4.0'* into position of the rearranged IgL locus. Compared to '*2.0-4.0'*, HM targeted the GFP2 reporter more efficiently and the median of decreased green fluorescence increases 3.5 fold from 2.2% to 7.7%. This argues for a

silencing motif in the *'2.0-2.4'*. However, removal of the *'2.0-2.2'* in the *'0-4'* fragment and also in the '2-3' did not increase HM activity.

Although the *'2.2-2.4'* in the 400 bp deletion study did not affect AID-mediated diversification, its role for HM could be confirmed in a study of 50 bp extending end deletions of the *'2-3'* fragment. HM dropped significantly after deletion of the *'2.0-2.25'*. Deleting only 50 bp less, left HM unaltered. Thus, the first 200 bp *'2.0-2.2'* are not important, but it may tempt to speculate that *'2.2-2.25'* contains a motif which is necessary for enhanced HM. However, inserting *'2.0-2.25'* alone did not induce HM to a significant level compared to the *'2-3'* fragment. Adding only 50 bp more resulted in the fragment *'2.0-2.3'*, and supported HM to a level similar like that observed for *'2-3'*.

To screen for cooperative *'HyCorE'* elements, also 50 bp internal deletions were introduced into the fragment *'2-3'*. Unfortunately, this deletion series did not give conclusive results. Is it possible that I could not identify an active motif? Let´s have another look on the end deletions of *'2-3'*. The 3' end deletion constructs *'2.00-2.30'* and *'2.00-2.25'* identified the 50 bp sequence *'2.25-2.30'* as important element. The 5'end constructs '2.25-3.00' and '2.20-3.00' identified the 50 bp sequence *'2.20-2.25'* as important element. Interestingly, the fragments *'2.00-2.25'* and *'2.25-3.00'*, which contain only one of the two fragments respectively, are clearly repressed in their ability to mutate the transgene. The fragments *'2.00-2.30'* and *'2.20-3.00'* which contain both identified sequence motifs, are actively mutating the transgene. This is a strong hint for both elements working together to induce HM. This could also be the reason why I could not identify a single 50 bp motif within the study of the internal deletions of *'2-3'*.

Again, this series proofed the importance of the *'2.2-2.4'* fragment for HM. It was also interesting if the fragment alone was sufficient to induce HM and it was inserted alone and in multiple copies into the position of the rearranged IgL locus. Beyond induction of HM, the duplication and quadruplication of the *'2.2-2.4'* led to an increase of HM activity. The element has an additive effect on HM, but repeating the sequencce 14 times did not further increase HM. Presumably there is a saturation level for AID recruitment to the locus, as other intrinsic factors like availability of the other involved proteins or the speed of transcription/recruitment of RNA polymerase II can limit the process.

Taken together the results of all deletion series, it was possible to identify a *'HyCorE'* of 200 bp named *'2.2-2.4'* with the likely core sequence being the *'2.2-2.3'*. Deletion of *'2.2-2.4'* within the *'0-4'*, the *'2-4'* and the *'2-3'* fragment reduced diversification activity significantly and insertion of the *'2.2-2.4'* alone was able to start HM. Hence, the element is both necessary and sufficient for HM to occur. The element is part of the IgL enhancer. It seems to contain redundant motifs which start diversification and act additively, as multimerization of the fragment was able to enhance HM activity. This is the first time that a true HM activator element could be identified.

5.5.2 The *'HyCorE'* starts HM at non-Ig loci

Similar to the *'W'* fragment, the *'2.2-2.4'* was inserted together with the GFP2 reporter construct into the Bach2 locus leading in the cell line $AID^{R1}IgL^{-}Bach2^{+/2.2\text{-}2.4,\ GFP2}$. In FACS analysis, the median of decreased fluorescence was compared to that of the cell line $AID^{R1}IgL^{2.2\text{-}2.4,\ GFP2}$, where *'2.2-2.4'* was integrated at the position of the rearranged IgL locus.

Integration of GFP2 only did not lead to mutation events (Blagodatski et al., 2009), which means the Bach2 locus is not hypermutating. If GFP2 was integrated together with the *'2.2-2.4'* fragment, the transgene was target for HM. The median of decreased green fluorescence was similar to that of the clone $AID^{R1}IgL^{2.2\text{-}2.4,\ GFP2}$. These results demonstrated that the *'2.2-2.4'* fragment was able to activate AID-mediated HM at loci which otherwise did not support HM. Consequently, the fragment contains all relevant cis-elements involved in the targeting of AID.

5.5.3 The *'HyCorE'* can be identified in other species

The *'HyCorE' '2.2-2.4'* was identified in the strongly related species duck and turkey. Transfection of the duck and turkey sequences into DT40 activated HM at a level similar to the chicken sequence.

This result fortifies the importance of the 200 bp sequence *'2.2-2.4'* for HM and further shows the conservation during evolution. Transgenic expression of the 200 bp fragment should be able to start HM also in other species like human and mouse.

The mechanism of HM seems to be highly conserved over species since experiments with an Igκ transgene in mouse show similar results to this study in chicken B cells (Michael et al., 2003). The transgene included an E box binding site which was able to enhance HM without enhancing transcription of the transgene.

5.5.4 The *'HyCorE'* should also be relevant for gene conversion and class switch recombination

Bursal B cells and DT40 diversify their rearranged IgL loci by gene conversion in the presence of ψV genes upstream of the IgL locus. The AID-dependence of this process suggests that one of the physiological roles of the IgL *'HyCorE'* may be also activation of gene conversion. Recently, other groups (Kothapalli et al., 2008) were able to identify diversification activating sequences of the DT40 IgL chain locus. Deletion of the enhancer had no effect in their study, but they indicated the importance of a sequence named 3'RR for Ig gene conversion. The sequence was 4.1 kb in length and did not contain the enhancer. This observation is not in conflict with our current findings, as the 3'RR has high similarity to the *'N'* fragment, which also exhibits high HM activity. The 3'RR corresponds to nt 5745-9826 and the *'N'* to nt 6131-9808 of GenBank entry FJ482234 (*'W'* fragment). We did not further examine the *'N'* fragment, because at that time point of the experiments *'N'*

was not sequenced completely. A bioinformatical discussion of *'N'* follows in the next chapter. Presumably, *'N'* and the 3'RR contain the same motifs as the *'2.2-2.4'*.

In general, it is a disadvantage of the study of Kothapalli et al. (2008) that they carried out only deletion experiments. To identify redundant motifs or multiple interacting sequences which can resubstitute for each other, it needs also reinsertion of parts of the IgL sequence into position of the deleted IgL locus. This might also be the reason why they could not identify the enhancer as being a potent diversification activator. They deleted the enhancer, but the 3'RR, which also contains HM-activating motifs, was still present.

Another lab which used a similar experimental setup to knock out the chicken IgL enhancer (Yang et al., 2006), was not able to identify HM relevant cis-element within the enhancer sequence, too. Also in this case, the sequence corresponding to *'N'* is still present and can resubstitute for the deleted enhancer.

It also seems likely that *'HyCorE'* play a role for switch recombination which is accompanied by HM of the recombining switch regions (Nagaoka et al., 2002). Possibly a dedicated activation sequence near the switch regions activates switch recombination. As the chicken IgL *'W'* fragment can activate HM in both directions over large distances (Blagodatski et al., 2009), it is also conceivable that a similar activation sequence in the heavy chain loci regulates both HM and switch recombination.

5.5.5 Bioinformatical analysis – a theoretical discussion of putative motifs involved in the recruitment of HM

To learn more about how AID is restricted to Ig loci in B cells, we have to find out which cofactors are needed to recruit and activate AID. The *'HyCorE'* gives the great advantage to study DNA-binding transcription factors involved in the recruitment of HM.

Using bioinformatical tools I could identify several cis-elements within the *'HyCorE'* *'2.2-2.4'*. The most prominent binding sites are the E-box motif CAGCTG and an NFκB binding site for its subunit p50 in the first 100 bp of the *'HyCorE'*. A cluster of interferon regulatory elements and a SpiB binding site accumulate in the middle part. The 3'end of the sequence contains binding sites for Pax5 and Thing1/E47 heterodimers.

Interferons (IFN) are a family of pleiotropic secreted proteins that play a key role in mediating antiviral and apoptotic responses, and in immune modulation. Interferons induce a large number of genes through activating the janus tyrosine kinase (JAK)–signal transducers and activators of transcription proteins (STAT) pathway, and the binding of transcription factors at specific DNA regulatory elements known as interferon-stimulated response element (ISRE). A role for diversification processes in B cells has not been examined yet.

NFκB is a heterodimeric transcription factor known to play an important role in the regulation of genes involved in immune modulation. Knockout studies in mice reveal immunological defects. The mice are more

prone to infections and despite normal Ig levels they fail to produce antigen-specific antibodies (Schwarz et al., 1997). B cell levels in spleen, bone marrow, and lymph nodes are reduced markedly and formation of GC is impaired (Caamano et al., 1998).
The binding motif found within the *'2.2-2.4'*, is specific for the p50 subunit of NFκB. Mice lacking this subunit show non-specific responses to infections and defects in germline C_H transcription and Ig class switching (Snapper et al., 1995; Sha et al., 1995).
The predominant form of NFκB heterodimers are from the Rel/NFκB family. They also play a critical role during B cell maturation and survival (Gugasyan et al., 2000). In general, NFκB proteins stay in the cytoplasm and are activated and translocated to the nucleus by NFκB essential modulator (NEMO). Interestingly, studies with B cells from patients with hyper-IgM syndrome who have deficient expression of NEMO show insufficient HM (Jain et al., 2004; Souto-Carneiro et al., 2008). Therefore, the NFκB binding site seems to be an very interesting candidate for a motif of the *'HyCorE'*.

Thinking about the '2-3' series of 50 bp end deletions, the fragments *'2.20-2.25'* and *'2.25-2.3'* seem to interact to support HM. Interestingly, the *'2.20-2.25'* contains the NFκB binding site and the *'2.25-2.3'* contains an ISRE. ISRE are known to cooperate with the NFκB motif to stimulate gene expression (Naschberger et al., 2004; Wu et al., 1994). This suggests, that both motifs act together to stimulate AID-mediated HM. To proof this idea, further knockout experiments or a coimmunoprecipitation of the *'HyCorE'* with its DNA-binding factors will be necessary.

The *'HyCorE'* element contains the E-box motif CAGCTG and most likely E12 and E47 are dimerizing on this sequence thereby activating AID-mediated diversification. This idea is supported by the results of the E2A knockout studies in DT40, which lead in decreased HM and GCV (Kitao et al., 2008; Schoetz et al., 2006). However, one E-box alone can not have this strong effect on the targeting of HM, as E-boxes 'CANNTG' are clustering all over the Ig locus with the motif CAGCTG being the most prominent one (see chapter Results, Figure 4.7). It is more likely that it needs an interplay of the E-box and other motifs of the *'HyCorE'* which still need to be examined experimentally.

Pax5 is a known regulator of AID expression (Gonda et al., 2003) and required for the maintenance of B cell identity during later development (Horcher et al., 2001). The knockout of Pax5 in DT40 led to plasma cell differentiation (Nera et al., 2006). A role for Pax5 in HM and GCV still needs to be discussed, but involvement in enhancer activity seems to be rather in a negative regulation.

Spi-B is a haematopoietic-specific transcription factor activating gene expression of target genes in lymphoid lineages. A homozygous knockout in mice leads in defective T-cell dependent response of B cells and the formation of only small germinal centers (Su et al., 1997). However, HM of Peyer's patch (a gut-associated lymphoid tissue which is collecting and presenting antigens) B lymphocytes is unaffected (Kim et al., 2003). Therefore, the Spi-B binding motif can only play a minor role in AID-mediated HM.

Most likely, the *'W'* fragment is composed of multiple redundant motifs with high similarity to the *'HyCorE'* element *'2.2-2.4'*. Results of multimerization militate in favor of this idea. As doubling and quadruplication of *'2.2-2.4'* did enhance HM, other elements seem not to be necessary. To evaluate this statement theoretically, I used again the MatInspector transcription factor database (Supplementary information chapter 8.3; www.genomatix.de; Quandt et al., 1995; Cartharius et al., 2005).
In the *'2.2-2.4'* the three motifs E Box, NFκB and ISRE stand out especially as discussed before. So I screened the sequence of *'W'* for these three motifs (Figure 5.1).

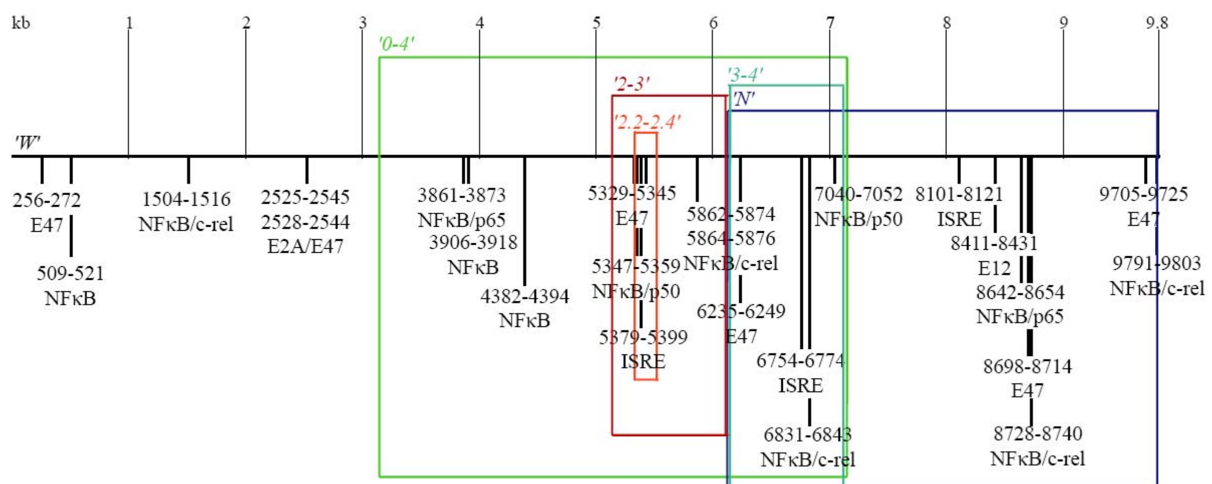


Figure 5.1Position of binding motifs relative to the 9826 bp fragment *'W'*.
All binding motifs for E2A transcription factors, NFκB factors, and ISRE (interferon-stimulated response element) which is binding interferon regulatory factors, are shown. The motifs are annotated with their name and their exact nt position. The position of important deletion fragments within 'W' is described, too: '0-4' (green box), '2-3' (red box), '2.2-2.4' (orange box), '3-4' (light blue box), 'N' (dark blue box).

In the first 5 kb of the sequence, which do not have a significant HM activity, no clustering of the three motifs is present. The tightest cluster of the three motifs is found within the *'2.2-2.4'* (Figure 5.1, orange box). Another cluster is between 6-7.1 kb (Figure 5.1, light blue box). This part of the sequence is contained within *'0-4'* and *'N'*, both fragments, which are actively hypermutating. One more cluster is at 8-9 kb and part of *'N'*.
If these motifs are really involved in HM remains elusive, but Figure 5.1 illustrates, that they cluster especially in that part of the sequence of *'W'* which was proven within this thesis to be relevant for the induction of HM.

This hypothetical analysis does not evaluate all putative motifs and therefore can not resubstitute experiments, but it points further analysis in direction of a knockout of factors binding to the NFkB and ISRE motifs.

5.6 A model for AID recruitment to hypermutating loci

The mechanism of how a cis-regulatory sequence can activate HM in adjacent transcription units remains speculative. Intriguingly, the chicken IgL *'HyCorE'* not only includes a part of the IgL enhancer, but also seems to act as a HM enhancer by activating HM at normally non-hypermutating loci and activating HM over long distances in upstream or downstream target genes when placed within *'W'* (Blagodatski et al., 2009). A plausible hypothesis may be that the *'HyCorE'* promotes the formation of protein complexes which first bind AID and then hand it over to the neighboring transcription initiation complex formed at the enhancer (Figure 5.2). Candidates for proteins involved in building such an AID docking station would be DNA binding factors which recognize sequence motifs within a *'HyCorE'*. The best candidate identified up to now is E47 as it also can bind to the CAGCTG binding motif within the *'HyCorE'* fragment *'2.2-2.4'*.

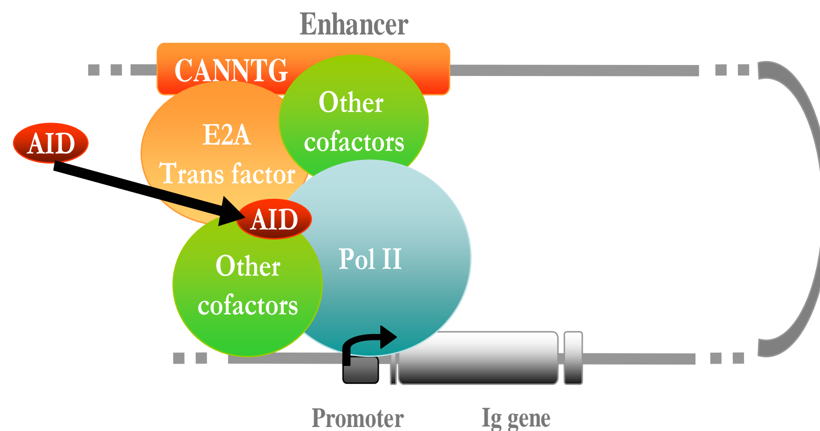


Figure 5.2 Proposed model for AID recruitment. E2A transcription factors (light orange oval) dimerize on their DNA binding site CANNTG on the enhancer (dark orange rectangle) and help to recruit AID (red oval) presumably in a complex with other factors (green ovals) to the transcription initiation site. Associated with RNA Polymerase II (blue oval), AID might start to run along the Ig gene.

Recent findings (Liu et al., 2008) come to an additional conclusion: AID is selectively targeted, but not the whole genome is protected. They analyzed 118 expressed genes from hypermutating B cells extracted from Peyer's patch of ~6 month old littermate mice and found about 25% of these genes being mutated. Beyond this, they found numerous genes which show no HM above background level in wild-type mice but a high mutation rate, if the high fidelity repair system in these mice is knocked out. This led the group to suggest that specificity is regulated by a balanced regulation of high-fidelity repair factors and error-prone DNA

polymerases. Consistent with the hypothesis of E2A transcription factors having a role in recruiting AID to hypermutating loci, Liu et al. found the E-box motif CAG($^{G}/_{C}$)TG was enriched in mutated genes of Peyer´s patch B cells of the knock-out mice compared to the wild-type. This is consistent with the idea that E2A transcription factors and the E-box CAGCTG in the 200 bp *'2.2-2.4' 'HyCorE'* are responsible for the targeting of AID specificity.

To use this 200 bp element and its binding factors in an assay to induce abasic sites, will contribute considerably to elucidate how error-free and error-prone repair are regulating the process of antibody diversification.

5.7 *Outlook*

AID was first identified in the lab of Honjo in 1999. The precise mechanism of how it is causing HM only at the Ig loci still remains unclear.

With the identification of cis-trans acting factors and elements involved in of AID-mediated diversification in my thesis, I wanted to clarify the locus specificity of AID. A recent study in another lab found a 4 kb DNA stretch of the IgL locus of DT40 to recruit AID-mediated diversification to the IgL locus (Kothapalli et al., 2008). In this study, for the first time it was possible to show an endogenous DNA element referred to as *'HyCorE'* that is small enough (200 bp) to make a tangible statement on trans-acting factors that bind to its sequence. The *'HyCorE'* is necessary and sufficient to induce HM and in an additive manner, it is also able to enhance HM. Insertion in other positions of the genome, will start HM there, too. The *'HyCorE'* is conserved during evolution, as it could be identified in the related sequences of duck and turkey. Both sequences were also able to induce HM. This gives rise to the conclusion, that the DNA element will be able to start diversification also in human or mouse.

Beyond this I showed that E2A transcription factors, which are able to bind to this *'HyCorE'*, are involved in HM as well. To clarify the mechanism of AID-targeting completely is of great importance, because the mistargeting of AID can lead to lymphoma. The results will have great impact on further elucidating the other proteins involved in recruiting AID. The *'HyCorE'* together with E12 and E47 proteins can be used for example in SILAC experiments for the coprecipitation of other complexing factors.

Furthermore, the data obtained during my thesis are very valuable on developing a new phenotypic assay to induce abasic sites at a specific locus. Having developed a method to specifically induce double-strand breaks (DSBs)and to monitor subsequent step-wise repair reactions has greatly contributed to understanding the role of individual repair factors in double-strand break repair. Berkovich et al. (2008) describes a method to induce DSBs at defined target sequences within the human genome using an endonuclease. This enables to analyze proteins bound to these DSBs by Chromatin Immunoprecipitation (ChIP) and PCR analysis. Similarly,

inducing abasic sites and thereby block replication at these sites will allow to investigate how cells re-start replication with accompanying mutagenesis. The *'HyCorE'* can be used to introduce abasic sites at a specific target sequence. Following ChIP and PCR analysis of this target sequence allow to monitor the distribution of DNA damage repair proteins around the abasic site at different stages of the cell cycle.

This study may have strong impact on the study of cancer chemotherapy, because a number of chemotherapeutic agents, such as cisplatin, kill cycling cells by blocking replication at cisplatin induced lesions, causing point mutations and inhibiting DNA repair (Comis, 1994).

Also a biotechnological application of the *'HyCorE'* is conceivable. Because of their short generation time (12 h), accumulation of many mutations, easy gene modification due to a high rate of homologous recombination and easy subcloning by limiting dilution, DT40 cells offer an advantageous system for the generation of monoclonal antibodies. Not only HM, but also gene conversion was already used to generate monoclonal antibodies with high specificity against a various selection of antigens including human IgG, Protein A and Streptavidin (Cumbers et al., 2002; Seo et al., 2005).

Beyond this, the *'HyCorE'* now offers a great tool for protein evolution. In our lab, a GFP with strong enhanced green fluorescence compared to commercial available GFP was already created (Arakawa et al., 2008). The GFP cDNA cassette was inserted 5' to the hypermutation targeting IgL locus. In a step forward, our vectors designed for this study, for example $pIgL^{4x2.2\text{-}2.4,\ GFP2}$, which has a mutation load of 3.6-4% (Figure 4.22.B) can be used to clone in other fluorescent proteins or other proteins instead of the GFP. As the sequence of the *'HyCorE'* is very short and easy to clone, these vectors can be used for random transfection into DT40 or other cell systems. Rapid generation of proteins, for example proteases or other enzymes used for industrial purposes is an ambitious market and the *'HyCorE'* could contribute to establish a new and improved technology.

6 List of abbreviations

A .... adenine
ab........ antibody
AD .... activation domain
AID .... Activation induced cytidine deaminase
ag........ antigen
ALV........ avian leucosis virus
BCR .... B cell receptor
BER .... base excision repair
bHLH........ basic helix loop helix
bsr........ blasticidin resistance
C........ constant; cytosine
CDR .... complementary determining regions
CSR .... class switch recombination
D........ diversity
DH........ diverse segment of the heavy chain
ds .... double stranded
DSB .... double strand break
EF1α........ elongation factor 1α
EST........ expressed sequence tag
FACS .... fluorescence activated cell sorting
G .... guanine
GC .... germinal center
GCV........ gene conversion
gpt .... guanine phosphoribosyl transferase: resistance to mycophenolic acid
H........ heavy
HM........ hypermutation
HyCorE........ hypermutation core element
iEμ .... IgH intronic enhancer
Ig........ Immunoglobulin
IgL........ Immunoglobulin light chain

IL .... Interleukin
IRES .... internal ribosomal entry site
ISRE .... interferon-stimulated response element
J .... joining
JH, JL .... joining segment of the heavy and light chain respectively
L .... light
LTR .... (viral) long term repeats
MAR .... matrix attachment region
MMR .... Mismatch repair
NES .... nuclear export signal
NFkB .... nuclear factor kappa B
NHEJ .... non homologous end joining
NLS .... nuclear localization signal
ORF .... open reading frame
PCR .... polymerase chain reaction
Pol .... polymerase
ψ .... pseudo
PKA .... Protein kinase A
RAG 1 and 2 .... recombination activating enzyme 1 and 2
RPA .... replication protein A
RSS .... recombination signal sequences
RSV .... Rous Sarcoma virus long terminal repeat
sIg .... surface immunoglobulin
S .... switch
SAGE .... Serial Analysis of Gene Expression
SSB .... single strand break
ss .... single stranded
STAT6 .... signal transducer and activator of transcription 6
T .... thymine
T_H .... T helper cell
TCR .... T cell receptor
U .... uracil
UNG .... Uracil DNA glycosylase
UTR .... untranslated region
V .... variable
VH, VL .... variable segment of the heavy or light chain respectively

7 References

A

Abdrakhmanov I, Lodygin D, Geroth P, Arakawa H, Law A, Plachy J, Korn B, Buerstedde JM. A large database of chicken bursal ESTs as a resource for the analysis of vertebrate gene function. Genome Res. 2000 Dec;10(12):2062-9.

Allen CD, Okada T, Cyster JG. Germinal-center organization and cellular dynamics. Immunity. 2007 Aug;27(2):190-202.

Arakawa H, Kudo H, Batrak V, Caldwell RB, Rieger MA, Ellwart JW, Buerstedde JM. Protein evolution by hypermutation and selection in the B cell line DT40. Nucleic Acids Res. 2008 Jan;36(1):e1.

Arakawa H, Moldovan GL, Saribasak H, Saribasak NN, Jentsch S, Buerstedde JM. A role for PCNA ubiquitination in immunoglobulin hypermutation. PLoS Biol. 2006 Nov;4(11):e366.

Arakawa H, Saribasak H, Buerstedde JM. Activation-induced cytidine deaminase initiates immunoglobulin gene conversion and hypermutation by a common intermediate. PLoS Biol. 2004 Jul;2(7):E179.

Arakawa H, Hauschild J, Buerstedde JM. Requirement of the activation-induced deaminase (AID) gene for immunoglobulin gene conversion. Science. 2002 Feb 15;295(5558):1301-6.

Arakawa H, Lodygin D, Buerstedde JM. Mutant loxP vectors for selectable marker recycle and conditional knock-outs. BMC Biotechnol. 2001;1:7.

Arakawa H, Kuma K, Yasuda M, Furusawa S, Ekino S, Yamagishi H. Oligoclonal development of B cells bearing discrete Ig chains in chicken single germinal centers. J Immunol. 1998 May 1;160(9):4232-41.

Arakawa H, Furusawa S, Ekino S, Yamagishi H. Immunoglobulin gene hyperconversion ongoing in chicken splenic germinal centers. EMBO J. 1996 May 15;15(10):2540-6.

Aronheim A, Shiran R, Rosen A, Walker MD. The E2A gene product contains two separable and functionally distinct transcription activation domains. Proc Natl Acad Sci U S A. 1993 Sep 1;90(17):8063-7.

Aronheim A, Shiran R, Rosen A, Walker MD. Cell-specific expression of helix-loop-helix transcription factors encoded by the E2A gene. Nucleic Acids Res. 1993 Apr 11;21(7):1601-6.

B

Baba TW, Humphries EH. Formation of a transformed follicle is necessary but not sufficient for development of an avian leukosis virus-induced lymphoma. Proc Natl Acad Sci U S A. 1985 Jan;82(1):213-6.

Bachl J, Carlson C, Gray-Schopfer V, Dessing M, Olsson C. Increased transcription levels induce higher mutation rates in a hypermutating cell line. J Immunol. 2001 Apr 15;166(8):5051-7.

Bachl J, Olsson C. Hypermutation targets a green fluorescent protein-encoding transgene in the presence of immunoglobulin enhancers. Eur J Immunol. 1999 Apr;29(4):1383-9.

Bain G, Romanow WJ, Albers K, Havran WL, Murre C. Positive and negative regulation of V(D)J recombination by the E2A proteins. J Exp Med. 1999 Jan 18;189(2):289-300.

Bain G, Maandag EC, Izon DJ, Amsen D, Kruisbeek AM, Weintraub BC, Krop I, Schlissel MS, Feeney AJ, van Roon M, et al. E2A proteins are required for proper B cell development and initiation of immunoglobulin gene rearrangements. Cell. 1994 Dec 2;79(5):885-92.

Barbié V, Lefranc MP. The human immunoglobulin kappa variable (IGKV) genes and joining (IGKJ) segments. Exp Clin Immunogenet. 1998;15(3):171-83.

Bardwell PD, Woo CJ, Wei K, Li Z, Martin A, Sack SZ, Parris T, Edelmann W, Scharff MD. Altered somatic hypermutation and reduced class-switch recombination in exonuclease 1-mutant mice. Nat Immunol. 2004 Feb;5(2):224-9.

Barreto V, Reina-San-Martin B, Ramiro AR, McBride KM, Nussenzweig MC. C-terminal deletion of AID uncouples class switch recombination from somatic hypermutation and gene conversion. Mol Cell. 2003 Aug;12(2):501-8.

Basu U, Chaudhuri J, Alpert C, Dutt S, Ranganath S, Li G, Schrum JP, Manis JP, Alt FW. The AID antibody diversification enzyme is regulated by protein kinase A phosphorylation. Nature. 2005 Nov 24;438(7067):508-11.

Baumal R, Birshtein BK, Coffino P, Scharff MD. Mutations in immunoglobulin-producing mouse myeloma cells. Science. 1973 Oct 12;182(108):164-6.

Berek C. Somatic mutation and memory. Curr Opin Immunol. 1993 Apr;5(2):218-22.

Berek C, Ziegner M. The maturation of the immune response. Immunol Today. 1993 Aug;14(8):400-4. Review. Erratum in: Immunol Today 1993 Oct;14(10):479.

Besmer E, Market E, Papavasiliou FN. The transcription elongation complex directs activation-induced cytidine deaminase-mediated DNA deamination. Mol Cell Biol. 2006 Jun;26(11):4378-85.

Betz AG, Milstein C, González-Fernández A, Pannell R, Larson T, Neuberger MS. Elements regulating somatic hypermutation of an immunoglobulin kappa gene: critical role for the intron enhancer/matrix attachment region. Cell. 1994 Apr 22;77(2):239-48.

Betz AG, Rada C, Pannell R, Milstein C, Neuberger MS. Passenger transgenes reveal intrinsic specificity of the antibody hypermutation mechanism: clustering, polarity, and specific hot spots. Proc Natl Acad Sci U S A. 1993 Mar 15;90(6):2385-8.

Bezzubova O, Silbergleit A, Yamaguchi-Iwai Y, Takeda S, Buerstedde JM. Reduced X-ray resistance and homologous recombination frequencies in a RAD54-/- mutant of the chicken DT40 cell line. Cell. 1997 Apr 18;89(2):185-93.

Blagodatski A, Batrak V, Schmidl S, Schoetz U, Caldwell RB, Arakawa H, Buerstedde JM. A cis-acting diversification activator both necessary and sufficient for AID-mediated hypermutation. PLoS Genet. 2009 Jan;5(1):e1000332. Epub 2009 Jan 9.

Bottaro A, Young F, Chen J, Serwe M, Sablitzky F, Alt FW. Deletion of the IgH intronic enhancer and associated matrix-attachment regions decreases, but does not abolish, class switching at the mu locus. Int Immunol. 1998 Jun;10(6):799-806.

Brack C, Hirama M, Lenhard-Schuller R, Tonegawa S. A complete immunoglobulin gene is created by somatic recombination. Cell. 1978 Sep;15(1):1-14.

Bradney C, Hjelmeland M, Komatsu Y, Yoshida M, Yao TP, Zhuang Y. Regulation of E2A activities by histone acetyltransferases in B lymphocyte development. J Biol Chem. 2003 Jan 24;278(4):2370-6.

Bransteitter R, Pham P, Scharff MD, Goodman MF. Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. Proc Natl Acad Sci U S A. 2003 Apr 1;100(7):4102-7.

Bransteitter R, Pham P, Calabrese P, Goodman MF. Biochemical analysis of hypermutational targeting by wild type and mutant activation-induced cytidine deaminase. J Biol Chem. 2004 Dec 3;279(49):51612-21.

Brar SS, Watson M, Diaz M. Activation-induced cytosine deaminase (AID) is actively exported out of the nucleus but retained by the induction of DNA breaks. J Biol Chem. 2004 Jun 18;279(25):26395-401.

Brenner S, Milstein C. Origin of antibody variation. Nature. 1966 Jul 16;211(5046):242-3.

Bross L, Fukita Y, McBlane F, Démollière C, Rajewsky K, Jacobs H. DNA double-strand breaks in immunoglobulin genes undergoing somatic hypermutation. Immunity. 2000 Nov;13(5):589-97.

Bross L, Muramatsu M, Kinoshita K, Honjo T, Jacobs H. DNA double-strand breaks: prior to but not sufficient in targeting hypermutation. J Exp Med. 2002 May 6;195(9):1187-92.

Buerstedde JM, Takeda S. Increased ratio of targeted to random integration after transfection of chicken B cell lines. Cell. 1991 Oct 4;67(1):179-88.

Buerstedde JM, Reynaud CA, Humphries EH, Olson W, Ewert DL, Weill JC. Light chain gene conversion continues at high rate in an ALV-induced cell line. EMBO J. 1990 Mar;9(3):921-7.

Bulfone-Paus S, Reiners-Schramm L, Lauster R. The chicken immunoglobulin lambda light chain gene is transcriptionally controlled by a modularly organized enhancer and an octamer-dependent silencer. Nucleic Acids Res. 1995 Jun 11;23(11):1997-2005.

C

Caamaño JH, Rizzo CA, Durham SK, Barton DS, Raventós-Suárez C, Snapper CM, Bravo R. Nuclear factor (NF)-kappa B2 (p100/p52) is required for normal splenic microarchitecture and B cell-mediated immune responses. J Exp Med. 1998 Jan 19;187(2):185-96.

Caldwell RB, Kierzek AM, Arakawa H, Bezzubov Y, Zaim J, Fiedler P, Kutter S, Blagodatski A, Kostovska D, Koter M, Plachy J, Carninci P, Hayashizaki Y, Buerstedde JM. Full-length cDNAs from chicken bursal lymphocytes to facilitate gene function analysis. Genome Biol. 2005;6(1):R6.

Catalan N, Selz F, Imai K, Revy P, Fischer A, Durandy A. The block in immunoglobulin class switch recombination caused by activation-induced cytidine deaminase deficiency occurs prior to the generation of DNA double strand breaks in switch mu region.J Immunol. 2003 Sep 1;171(5):2504-9.

Carlson LM, McCormack WT, Postema CE, Humphries EH, Thompson CB. Templated insertions in the rearranged chicken IgL V gene segment arise by intrachromosomal gene conversion. Genes Dev. 1990 Apr;4(4):536-47.

Cartharius K, Frech K, Grote K, Klocke B, Haltmeier M, Klingenhoff A, Frisch M, Bayerlein M, Werner T. MatInspector and beyond: promoter analysis based on transcription factor binding sites. Bioinformatics. 2005 Jul 1;21(13):2933-42.

Casellas R, Nussenzweig A, Wuerffel R, Pelanda R, Reichlin A, Suh H, Qin XF, Besmer E, Kenter A, Rajewsky K, Nussenzweig MC. Ku80 is required for immunoglobulin isotype switching. EMBO J. 1998 Apr 15;17(8):2404-11.

Cattoretti G, Büttner M, Shaknovich R, Kremmer E, Alobeid B, Niedobitek G. Nuclear and cytoplasmic AID in extrafollicular and germinal center B cells. Blood. 2006 May 15;107(10):3967-75.

Chaudhuri J, Alt FW. Class-switch recombination: interplay of transcription, DNA deamination and DNA repair. Nat Rev Immunol. 2004 Jul;4(7):541-52. Erratum in: Nat Rev Immunol. 2004 Aug;4(8): 655.

Chaudhuri J, Khuong C, Alt FW. Replication protein A interacts with AID to promote deamination of somatic hypermutation targets. Nature. 2004 Aug 26;430(7003):992-8.

Chaudhuri J, Tian M, Khuong C, Chua K, Pinaud E, Alt FW.Transcription-targeted DNA deamination by the AID antibody diversification enzyme. Nature. 2003 Apr 17;422(6933):726-30.

Chien NC, Pollock RR, Desaymard C, Scharff MD. Point mutations cause the somatic diversification of IgM and IgG2a antiphosphorylcholine antibodies. J Exp Med. 1988 Mar 1;167(3):954-73. Erratum in: J Exp Med 1990 Aug 1;172(2):669.

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 Sep 16;15(18):5014-21.

Chua KF, Alt FW, Manis JP. The function of AID in somatic mutation and class switch recombination: upstream or downstream of DNA breaks. J Exp Med. 2002 May 6;195(9):F37-41.

Conlon TM, Meyer KB. The chicken Ig light chain 3'-enhancer is essential for gene expression and regulates gene conversion via the transcription factor E2A. Eur J Immunol. 2006 Jan;36(1):139-48. Erratum in: Eur J Immunol. 2006 Sep;36(9):2556.

Conlon TM, Meyer KB. Cloning and functional characterisation of avian transcription factor E2A. BMC Immunol. 2004 Jun 14;5:11.

Cumbers SJ, Williams GT, Davies SL, Grenfell RL, Takeda S, Batista FD, Sale JE, Neuberger MS. Generation and iterative affinity maturation of antibodies in vitro using hypermutating B-cell lines. Nat Biotechnol. 2002 Nov;20(11):1129-34. Epub 2002 Oct 15.

D

Daniels GA, Lieber MR. Transcription targets recombination at immunoglobulin switch sequences in a strand-specific manner. Curr Top Microbiol Immunol. 1996;217:171-89.

D'Avirro N, Truong D, Xu B, Selsing E. Sequence transfers between variable regions in a mouse antibody transgene can occur by gene conversion. J Immunol. 2005 Dec 15;175(12):8133-7.

Dedeoglu F, Horwitz B, Chaudhuri J, Alt FW, Geha RS. Induction of activation-induced cytidine deaminase gene expression by IL-4 and CD40 ligation is dependent on STAT6 and NFkappaB. Int Immunol. 2004 Mar;16(3):395-404.

Dickerson SK, Market E, Besmer E, Papavasiliou FN. AID mediates hypermutation by deaminating single stranded DNA. J Exp Med. 2003 May 19;197(10):1291-6.

Di Noia JM, Neuberger MS. Immunoglobulin gene conversion in chicken DT40 cells largely proceeds through an abasic site intermediate generated by excision of the uracil produced by AID-mediated deoxycytidine deamination. Eur J Immunol. 2004 Feb;34(2):504-8.

Di Noia J, Neuberger MS. Altering the pathway of immunoglobulin hypermutation by inhibiting uracil-DNA glycosylase. Nature. 2002 Sep 5;419(6902):43-8.

Dudley DD, Manis JP, Zarrin AA, Kaylor L, Tian M, Alt FW. Internal IgH class switch region deletions are position-independent and enhanced by AID expression. Proc Natl Acad Sci U S A. 2002 Jul 23;99(15):9984-9.

Durandy A, Honjo T. Human genetic defects in class-switch recombination (hyper-IgM syndromes).
Curr Opin Immunol. 2001 Oct;13(5):543-8.

E

Eason DD, Cannon JP, Haire RN, Rast JP, Ostrov DA, Litman GW. Mechanisms of antigen receptor evolution. Semin Immunol. 2004 Aug;16(4):215-26.

Eastman QM, Leu TM, Schatz DG. Initiation of V(D)J recombination in vitro obeying the 12/23 rule.
Nature. 1996 Mar 7;380(6569):85-8.

Ehrenstein MR, Neuberger MS. Deficiency in Msh2 affects the efficiency and local sequence specificity of immunoglobulin class-switch recombination: parallels with somatic hypermutation. EMBO J. 1999 Jun 15;18(12):3484-90.

Ephrussi A, Church GM, Tonegawa S, Gilbert W. B lineage--specific interactions of an immunoglobulin enhancer with cellular factors in vivo. Science. 1985 Jan 11;227(4683):134-40.

Eto T, Kinoshita K, Yoshikawa K, Muramatsu M, Honjo T. RNA-editing cytidine deaminase Apobec-1 is unable to induce somatic hypermutation in mammalian cells. Proc Natl Acad Sci U S A. 2003 Oct 28;100(22):12895-8.

F

Faili A, Aoufouchi S, Weller S, Vuillier F, Stary A, Sarasin A, Reynaud CA, Weill JC. DNA polymerase eta is involved in hypermutation occurring during immunoglobulin class switch recombination. J Exp Med. 2004 Jan 19;199(2):265-70.

Faili A, Aoufouchi S, Guéranger Q, Zober C, Léon A, Bertocci B, Weill JC, Reynaud CA. AID-dependent somatic hypermutation occurs as a DNA single-strand event in the BL2 cell line. Nat Immunol. 2002 Sep;3(9):815-21. Epub 2002 Jul 29.

Franco S, Gostissa M, Zha S, Lombard DB, Murphy MM, Zarrin AA, Yan C, Tepsuporn S, Morales JC, Adams MM, Lou Z, Bassing CH, Manis JP, Chen J, Carpenter PB, Alt FW. H2AX prevents DNA breaks from progressing to chromosome breaks and translocations. Mol Cell. 2006 Jan 20;21(2):201-14.

Frasca D, Landin AM, Lechner SC, Ryan JG, Schwartz R, Riley RL, Blomberg BB. Aging down-regulates the transcription factor E2A, activation-induced cytidine deaminase, and Ig class switch in human B cells. J Immunol. 2008 Apr 15;180(8):5283-90.

Fregien N, Davidson N. Activating elements in the promoter region of the chicken beta-actin gene. Gene. 1986;48(1):1-11.

Fugmann SD, Schatz DG. Immunology. One AID to unite them all. Science. 2002 Feb 15;295(5558):1244-5.

Fugmann SD, Lee AI, Shockett PE, Villey IJ, Schatz DG. The RAG proteins and V(D)J recombination: complexes, ends, and transposition. Annu Rev Immunol. 2000;18:495-527.

Fukita Y, Jacobs H, Rajewsky K. Somatic hypermutation in the heavy chain locus correlates with transcription. Immunity. 1998 Jul;9(1):105-14.

G

Gay NJ, Gangloff M. Structure and function of Toll receptors and their ligands. Annu Rev Biochem. 2007;76:141-65.

Gilfillan S, Dierich A, Lemeur M, Benoist C, Mathis D. Mice lacking TdT: mature animals with an immature lymphocyte repertoire.Science. 1993 Aug 27;261(5125):1175-8. Erratum in: Science 1993 Dec 24;262(5142):1957.

Goebel P, Janney N, Valenzuela JR, Romanow WJ, Murre C, Feeney AJ.Localized gene-specific induction of accessibility to V(D)J recombination induced by E2A and early B cell factor in nonlymphoid cells. J Exp Med. 2001 Sep 3;194(5):645-56.

Goldfarb AN, Flores JP, Lewandowska K. Involvement of the E2A basic helix-loop-helix protein in immunoglobulin heavy chain class switching. Mol Immunol. 1996 Aug;33(11-12):947-56.

Golding GB, Gearhart PJ, Glickman BW. Patterns of somatic mutations in immunoglobulin variable genes. Genetics. 1987 Jan;115(1):169-76.

Gonda H, Sugai M, Nambu Y, Katakai T, Agata Y, Mori KJ, Yokota Y, Shimizu A. The balance between Pax5 and Id2 activities is the key to AID gene expression. J Exp Med. 2003 Nov 3;198(9):1427-37.

Gordon MS, Kanegai CM, Doerr JR, Wall R. nSomatic hypermutation of the B cell receptor genes B29 (Igbeta, CD79b) and mb1 (Igalpha, CD79a). Proc Natl Acad Sci U S A. 2003 Apr 1;100(7):4126-31.

Gorman CM, Merlino GT, Willingham MC, Pastan I, Howard BH. The Rous sarcoma virus long terminal repeat is a strong promoter when introduced into a variety of eukaryotic cells by DNA-mediated transfection. Proc Natl Acad Sci U S A. 1982 Nov;79(22):6777-81.

Goyenechea B, Klix N, Yélamos J, Williams GT, Riddell A, Neuberger MS, Milstein C. Cells strongly expressing Ig(kappa) transgenes show clonal recruitment of hypermutation: a role for both MAR and the enhancers. EMBO J. 1997 Jul 1;16(13):3987-94.

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 Jul 31;388(6641):492-5.

Gugasyan R, Grumont R, Grossmann M, Nakamura Y, Pohl T, Nesic D, Gerondakis S. Rel/NF-kappaB transcription factors: key mediators of B-cell activation. Immunol Rev. 2000 Aug;176:134-40.

H

Hardy RR, Hayakawa K. B cell development pathways. Annu Rev Immunol. 2001;19:595-621.

Harris RS, Sale JE, Petersen-Mahrt SK, Neuberger MS. AID is essential for immunoglobulin V gene conversion in a cultured B cell line. Curr Biol. 2002 Mar 5;12(5):435-8.

Harris LJ, Larson SB, Hasel KW, Day J, Greenwood A, McPherson A. The three-dimensional structure of an intact monoclonal antibody for canine lymphoma. Nature. 1992 Nov 26;360(6402):369-72.

Hayward WS, Neel BG, Astrin SM. Activation of a cellular onc gene by promoter insertion in ALV-induced lymphoid leukosis. Nature. 1981 Apr 9;290(5806):475-80.

Henthorn P, Kiledjian M, Kadesch T. Two distinct transcription factors that bind the immunoglobulin enhancer microE5/kappa 2 motif. Science. 1990 Jan 26;247(4941):467-70.

Horcher M, Souabni A, Busslinger M. Pax5/BSAP maintains the identity of B cells in late B lymphopoiesis. Immunity. 2001 Jun;14(6):779-90.

I

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 Oct;4(10):1023-8.

Inlay MA, Gao HH, Odegard VH, Lin T, Schatz DG, Xu Y. Roles of the Ig kappa light chain intronic and 3' enhancers in Igk somatic hypermutation. J Immunol. 2006 Jul 15;177(2):1146-51.

International Chicken Genome Sequencing Consortium. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature. 2004 Dec 9;432(7018):695-716. Erratum in: Nature. 2005 Feb 17;433(7027):777.

Ito S, Nagaoka H, Shinkura R, Begum N, Muramatsu M, Nakata M, Honjo T. Activation-induced cytidine deaminase shuttles between nucleus and cytoplasm like apolipoprotein B mRNA editing catalytic polypeptide 1. Proc Natl Acad Sci U S A. 2004 Feb 17;101(7):1975-80.

J

Jacob J, Przylepa J, Miller C, Kelsoe G. In situ studies of the primary immune response to (4-hydroxy-3-nitrophenyl)acetyl. III. The kinetics of V region mutation and selection in germinal center B cells. J Exp Med. 1993 Oct 1;178(4):1293-307.

Jain A, Ma CA, Lopez-Granados E, Means G, Brady W, Orange JS, Liu S, Holland S, Derry JM. Specific NEMO mutations impair CD40-mediated c-Rel activation and B cell terminal differentiation. J Clin Invest. 2004 Dec;114(11):1593-602.

Janeway C.A., Travers P., Walport M., Shlomchik M. Immunologie. 5. Auflage Spektrum, Akad. Verl. 2002 ISBN 3-8274-1079-7

Jansen JG, Langerak P, Tsaalbi-Shtylik A, van den Berk P, Jacobs H, de Wind N. Strand-biased defect in C/G transversions in hypermutating immunoglobulin genes in Rev1-deficient mice. J Exp Med. 2006 Feb 20;203(2):319-23.

Jolly CJ, Neuberger MS. Somatic hypermutation of immunoglobulin kappa transgenes: association of mutability with demethylation. Immunol Cell Biol. 2001 Feb;79(1):18-22.

K

Kanayama N, Todo K, Takahashi S, Magari M, Ohmori H. Genetic manipulation of an exogenous non-immunoglobulin protein by gene conversion machinery in a chicken B cell line. Nucleic Acids Res. 2006 Jan 18;34(2):e10.

Kawamoto T, Araki K, Sonoda E, Yamashita YM, Harada K, Kikuchi K, Masutani C, Hanaoka F, Nozaki K, Hashimoto N, Takeda S. Dual roles for DNA polymerase eta in homologous DNA recombination and translesion DNA synthesis. Mol Cell. 2005 Dec 9;20(5):793-9.

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 Aug 17;188(4):699-713.

Kho CJ, Huggins GS, Endege WO, Hsieh CM, Lee ME, Haber E. Degradation of E2A proteins through a ubiquitin-conjugating enzyme, UbcE2A. J Biol Chem. 1997 Feb 7;272(6):3845-51.

Kim N, Martin TE, Simon MC, Storb U. The transcription factor Spi-B is not required for somatic hypermutation. Mol Immunol. 2003 Jan;39(10):577-83.

Kim S, Humphries EH, Tjoelker L, Carlson L, Thompson CB. Ongoing diversification of the rearranged immunoglobulin light-chain gene in a bursal lymphoma cell line. Mol Cell Biol. 1990 Jun;10(6):3224-31.

Kinoshita K, Honjo T. Linking class-switch recombination with somatic hypermutation. Nat Rev Mol Cell Biol. 2001 Jul;2(7):493-503.

Kitao H, Kimura M, Yamamoto K, Seo H, Namikoshi K, Agata Y, Ohta K, Takata M. Regulation of histone H4 acetylation by transcription factor E2A in Ig gene conversion. Int Immunol. 2008 Feb;20(2):277-84.

Kleinstein SH, Louzoun Y, Shlomchik MJ. Estimating hypermutation rates from clonal tree data.
J Immunol. 2003 Nov 1;171(9):4639-49.

Klix N, Jolly CJ, Davies SL, Brüggemann M, Williams GT, Neuberger MS. Multiple sequences from downstream of the J kappa cluster can combine to recruit somatic hypermutation to a heterologous, upstream mutation domain. Eur J Immunol. 1998 Jan;28(1):317-26.

Klotz EL, Storb U. Somatic hypermutation of a lambda 2 transgene under the control of the lambda enhancer or the heavy chain intron enhancer. J Immunol. 1996 Nov 15;157(10):4458-63.

Knight KL, Barrington RA. Somatic diversification of IgH genes in rabbit. Immunol Rev. 1998 Apr;162:37-47.

Kong Q, Zhao L, Subbaiah S, Maizels N. A lambda 3' enhancer drives active and untemplated somatic hypermutation of a lambda 1 transgene. J Immunol. 1998 Jul 1;161(1):294-301.

Kotani A, Okazaki IM, Muramatsu M, Kinoshita K, Begum NA, Nakajima T, Saito H, Honjo T.

A target selection of somatic hypermutations is regulated similarly between T and B cells upon activation-induced cytidine deaminase expression. Proc Natl Acad Sci U S A. 2005 Mar 22;102(12):4506-11.

Kothapalli N, Norton DD, Fugmann SD. Cutting edge: a cis-acting DNA element targets AID-mediated sequence diversification to the chicken Ig light chain gene locus. J Immunol. 2008 Feb 15;180(4):2019-23.

Krawinkel U, Zoebelein G, Brüggemann M, Radbruch A, Rajewsky K. Recombination between antibody heavy chain variable-region genes: evidence for gene conversion. Proc Natl Acad Sci U S A. 1983 Aug;80(16):4997-5001.

Kurosawa Y, von Boehmer H, Haas W, Sakano H, Trauneker A, Tonegawa S. Identification of D segments of immunoglobulin heavy-chain genes and their rearrangement in T lymphocytes. Nature. 1981 Apr 16;290(5807):565-70.

L

Landau NR, Schatz DG, Rosa M, Baltimore D. Increased frequency of N-region insertion in a murine pre-B-cell line infected with a terminal deoxynucleotidyl transferase retroviral expression vector. Mol Cell Biol. 1987 Sep;7(9):3237-43.

Landowski TH, Qu N, Buyuksal I, Painter JS, Dalton WS. Mutations in the Fas antigen in patients with multiple myeloma. Blood. 1997 Dec 1;90(11):4266-70.

Larson ED, Cummings WJ, Bednarski DW, Maizels N. MRE11/RAD50 cleaves DNA in the AID/UNG-dependent pathway of immunoglobulin gene diversification. Mol Cell. 2005 Nov 11;20(3):367-75.

Lebecque SG, Gearhart PJ. Boundaries of somatic mutation in rearranged immunoglobulin genes: 5' boundary is near the promoter, and 3' boundary is approximately 1 kb from V(D)J gene. J Exp Med. 1990 Dec 1;172(6):1717-27.

Li Z, Zhao C, Iglesias-Ussel MD, Polonskaya Z, Zhuang M, Yang G, Luo Z, Edelmann W, Scharff MD. The mismatch repair protein Msh6 influences the in vivo AID targeting to the Ig locus. Immunity. 2006 Apr;24(4):393-403.

Li Z, Otevrel T, Gao Y, Cheng HL, Seed B, Stamato TD, Taccioli GE, Alt FW. The XRCC4 gene encodes a novel protein involved in DNA double-strand break repair and V(D)J recombination.
Cell. 1995 Dec 29;83(7):1079-89.

Lieber MR, Ma Y, Pannicke U, Schwarz K. The mechanism of vertebrate nonhomologous DNA end joining and its role in V(D)J recombination. DNA Repair (Amst). 2004 Aug-Sep;3(8-9):817-26.

Liu M, Duke JL, Richter DJ, Vinuesa CG, Goodnow CC, Kleinstein SH, Schatz DG. Two levels of protection for the B cell genome during somatic hypermutation. Nature. 2008 Feb 14;451(7180):841-5.

Longerich S, Tanaka A, Bozek G, Nicolae D, Storb U. The very 5' end and the constant region of Ig genes are spared from somatic mutation because AID does not access these regions. J Exp Med. 2005 Nov 21;202(10):1443-54.

Luby TM, Schrader CE, Stavnezer J, Selsing E. The mu switch region tandem repeats are important, but not required, for antibody class switch recombination. J Exp Med. 2001 Jan 15;193(2):159-68.

M

Ma Y, Pannicke U, Schwarz K, Lieber MR. Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination. Cell. 2002 Mar 22;108(6):781-94.

MacLennan IC. Germinal centers. Annu Rev Immunol. 1994;12:117-39.

Manis JP, Dudley D, Kaylor L, Alt FW. IgH class switch recombination to IgG1 in DNA-PKcs-deficient B cells. Immunity. 2002 Apr;16(4):607-17.

Manis JP, Tian M, Alt FW. Mechanism and control of class-switch recombination. Trends Immunol. 2002 Jan;23(1):31-9.

Manis JP, van der Stoep N, Tian M, Ferrini R, Davidson L, Bottaro A, Alt FW.i Class switching in B cells lacking 3' immunoglobulin heavy chain enhancers. J Exp Med. 1998 Oct 19;188(8):1421-31.

Manis JP, Gu Y, Lansford R, Sonoda E, Ferrini R, Davidson L, Rajewsky K, Alt FW. Ku70 is required for late B cell development and immunoglobulin heavy chain class switching. J Exp Med. 1998 Jun 15;187(12):2081-9.

Mansikka A, Sandberg M, Lassila O, Toivanen P. Rearrangement of immunoglobulin light chain genes in the chicken occurs prior to colonization of the embryonic bursa of Fabricius. Proc Natl Acad Sci U S A. 1990 Dec;87(23):9416-20.

Martin A, Scharff MD.i Somatic hypermutation of the AID transgene in B and non-B cells. Proc Natl Acad Sci U S A. 2002 Sep 17;99(19):12304-8.

Martin A, Scharff MD. AID and mismatch repair in antibody diversification. Nat Rev Immunol. 2002 Aug;2(8):605-14.

Martin A, Bardwell PD, Woo CJ, Fan M, Shulman MJ, Scharff MD. Activation-induced cytidine deaminase turns on somatic hypermutation in hybridomas. Nature. 2002 Feb 14;415(6873):802-6.

Martomo SA, Yang WW, Gearhart PJ. A role for Msh6 but not Msh3 in somatic hypermutation and class switch recombination. J Exp Med. 2004 Jul 5;200(1):61-8.

Massari ME, Murre C. Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms.
Mol Cell Biol. 2000 Jan;20(2):429-40.

Massari ME, Grant PA, Pray-Grant MG, Berger SL, Workman JL, Murre C. A conserved motif present in a class of helix-loop-helix proteins activates transcription by direct recruitment of the SAGA complex. Mol Cell. 1999 Jul;4(1):63-73.

McBride KM, Gazumyan A, Woo EM, Barreto VM, Robbiani DF, Chait BT, Nussenzweig MC. Regulation of hypermutation by activation-induced cytidine deaminase phosphorylation.
Proc Natl Acad Sci U S A. 2006 Jun 6;103(23):8798-803.

McBride KM, Barreto V, Ramiro AR, Stavropoulos P, Nussenzweig MC. Somatic hypermutation is limited by CRM1-dependent nuclear export of activation-induced deaminase. J Exp Med. 2004 May 3;199(9):1235-44.

McCormack WT, Tjoelker LW, Thompson CB. Avian B-cell development: generation of an immunoglobulin repertoire by gene conversion. Annu Rev Immunol. 1991;9:219-41.

McCormack WT, Thompson CB. Chicken IgL variable region gene conversions display pseudogene donor preference and 5' to 3' polarity. Genes Dev. 1990 Apr;4(4):548-58.

McCormack WT, Tjoelker LW, Carlson LM, Petryniak B, Barth CF, Humphries EH, Thompson CB.
Chicken IgL gene rearrangement involves deletion of a circular episome and addition of single nonrandom nucleotides to both coding segments. Cell. 1989 Mar 10;56(5):785-91.

McKean D, Huppi K, Bell M, Staudt L, Gerhard W, Weigert M. Generation of antibody diversity in the immune response of BALB/c mice to influenza virus hemagglutinin. Proc Natl Acad Sci U S A. 1984 May;81(10):3180-4.

Michael N, Shen HM, Longerich S, Kim N, Longacre A, Storb U. The E box motif CAGGTG enhances somatic hypermutation without enhancing transcription. Immunity. 2003 Aug;19(2):235-42.

Michael N, Martin TE, Nicolae D, Kim N, Padjen K, Zhan P, Nguyen H, Pinkert C, Storb U.
Effects of sequence and structure on the hypermutability of immunoglobulin genes. Immunity. 2002 Jan;16(1):123-34.

Migliazza A, Martinotti S, Chen W, Fusco C, Ye BH, Knowles DM, Offit K, Chaganti RS, Dalla-Favera R. Frequent somatic hypermutation of the 5' noncoding region of the BCL6 gene in B-cell lymphoma. Proc Natl Acad Sci U S A. 1995 Dec 19;92(26):12520-4.

Muramatsu M, Kinoshita K, Fagarasan S, Yamada S, Shinkai Y, Honjo T. Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. Cell. 2000 Sep 1;102(5):553-63.

Muramatsu M, Sankaranand VS, Anant S, Sugai M, Kinoshita K, Davidson NO, Honjo T. Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. J Biol Chem. 1999 Jun 25;274(26):18470-6.

Murr R, Loizou JI, Yang YG, Cuenin C, Li H, Wang ZQ, Herceg Z. Histone acetylation by Trrap-Tip60 modulates loading of repair proteins and repair of DNA double-strand breaks. Nat Cell Biol. 2006 Jan;8(1):91-9.

Murre C, McCaw PS, Baltimore D. A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. Cell. 1989 Mar 10;56(5):777-83.

Müschen M, Re D, Jungnickel B, Diehl V, Rajewsky K, Küppers R. Somatic mutation of the CD95 gene in human B cells as a side-effect of the germinal center reaction. J Exp Med. 2000 Dec 18;192(12):1833-40.

Muto T, Okazaki IM, Yamada S, Tanaka Y, Kinoshita K, Muramatsu M, Nagaoka H, Honjo T. Negative regulation of activation-induced cytidine deaminase in B cells. Proc Natl Acad Sci U S A. 2006 Feb 21;103(8):2752-7. Epub 2006 Feb 13.

Muto A, Tashiro S, Nakajima O, Hoshino H, Takahashi S, Sakoda E, Ikebe D, Yamamoto M, Igarashi K. The transcriptional programme of antibody class switching involves the repressor Bach2. Nature. 2004 Jun 3;429(6991):566-71.

Muto T, Muramatsu M, Taniwaki M, Kinoshita K, Honjo T. Isolation, tissue distribution, and chromosomal localization of the human activation-induced cytidine deaminase (AID) gene. Genomics. 2000 Aug 15;68(1):85-8.

N

Nagaoka H, Muramatsu M, Yamamura N, Kinoshita K, Honjo T. Activation-induced deaminase (AID)-directed hypermutation in the immunoglobulin Smu region: implication of AID involvement in a common step of class switch recombination and somatic hypermutation. J Exp Med. 2002 Feb 18;195(4):529-34.

Nambu Y, Sugai M, Gonda H, Lee CG, Katakai T, Agata Y, Yokota Y, Shimizu A. Transcription-coupled events associating with immunoglobulin switch region chromatin. Science. 2003 Dec 19;302(5653):2137-40.

Naschberger E, Werner T, Vicente AB, Guenzi E, Töpolt K, Leubert R, Lubeseder-Martellato C, Nelson PJ, Stürzl M. Nuclear factor-kappaB motif and interferon-alpha-stimulated response element co-operate in the activation of guanylate-binding protein-1 expression by inflammatory cytokines in endothelial cells. Biochem J. 2004 Apr 15;379(Pt 2):409-20.

Neel JV, Satoh C, Goriki K, Fujita M, Takahashi N, Asakawa J, Hazama R. The rate with which spontaneous mutation alters the electrophoretic mobility of polypeptides. Proc Natl Acad Sci U S A. 1986 Jan;83(2):389-93.

Nelsen B, Sen R. Regulation of immunoglobulin gene transcription. Int Rev Cytol. 1992;133:121-49.

Nera KP, Kohonen P, Narvi E, Peippo A, Mustonen L, Terho P, Koskela K, Buerstedde JM, Lassila O.
Loss of Pax5 promotes plasma cell differentiation. Immunity. 2006 Mar;24(3):283-93.

Neuberger MS, Di Noia JM, Beale RC, Williams GT, Yang Z, Rada C. Somatic hypermutation at A.T pairs: polymerase error versus dUTP incorporation. Nat Rev Immunol. 2005 Feb;5(2):171-8.

Neuberger MS, Harris RS, Di Noia J, Petersen-Mahrt SK. Immunity through DNA deamination.
Trends Biochem Sci. 2003 Jun;28(6):305-12.

O

Odegard VH, Schatz DG. Targeting of somatic hypermutation. Nat Rev Immunol. 2006 Aug;6(8):573-83.

Odegard VH, Kim ST, Anderson SM, Shlomchik MJ, Schatz DG. Histone modifications associated with somatic hypermutation. Immunity. 2005 Jul;23(1):101-10.

Oettinger MA, Schatz DG, Gorka C, Baltimore D. RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. Science. 1990 Jun 22;248(4962):1517-23.

Okazaki IM, Kotani A, Honjo T. Role of AID in tumorigenesis. Adv Immunol. 2007;94:245-73.

Okazaki IM, Hiai H, Kakazu N, Yamada S, Muramatsu M, Kinoshita K, Honjo T. Constitutive expression of AID leads to tumorigenesis. J Exp Med. 2003 May 5;197(9):1173-81.

Okazaki IM, Kinoshita K, Muramatsu M, Yoshikawa K, Honjo T. The AID enzyme induces class switch recombination in fibroblasts. Nature. 2002 Mar 21;416(6878):340-5. Epub 2002 Mar 3.

P

Pallarès N, Lefebvre S, Contet V, Matsuda F, Lefranc MP. The human immunoglobulin heavy variable genes. Exp Clin Immunogenet. 1999;16(1):36-60..

Pallarès N, Frippiat JP, Giudicelli V, Lefranc MP. The human immunoglobulin lambda variable (IGLV) genes and joining (IGLJ) segments. Exp Clin Immunogenet. 1998;15(1):8-18.

Papavasiliou FN, Schatz DG. Somatic hypermutation of immunoglobulin genes: merging mechanisms for genetic diversity. Cell. 2002 Apr;109 Suppl:S35-44.

Papavasiliou FN, Schatz DG. Cell-cycle-regulated DNA double-stranded breaks in somatic hypermutation of immunoglobulin genes. Nature. 2000 Nov 9;408(6809):216-21.

Parng CL, Hansal S, Goldsby RA, Osborne BA. Gene conversion contributes to Ig light chain diversity in cattle. J Immunol. 1996 Dec 15;157(12):5478-86.

Parsa JY, Basit W, Wang CL, Gommerman JL, Carlyle JR, Martin A. AID mutates a non-immunoglobulin transgene independent of chromosomal position. Mol Immunol. 2007 Jan;44(4):567-75.

Parvari R, Ziv E, Lantner F, Heller D, Schechter I. Somatic diversification of chicken immunoglobulin light chains by point mutations. Proc Natl Acad Sci U S A. 1990 Apr;87(8):3072-6.

Pasqualucci L, Bhagat G, Jankovic M, Compagno M, Smith P, Muramatsu M, Honjo T, Morse HC 3rd, Nussenzweig MC, Dalla-Favera R. AID is required for germinal center-derived lymphomagenesis. Nat Genet. 2008 Jan;40(1):108-12.

Pasqualucci L, Neumeister P, Goossens T, Nanjangud G, Chaganti RS, Küppers R, Dalla-Favera R. Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. Nature. 2001 Jul 19;412(6844):341-6.

Pasqualucci L, Migliazza A, Fracchiolla N, William C, Neri A, Baldini L, Chaganti RS, Klein U, Küppers R, Rajewsky K, Dalla-Favera R. BCL-6 mutations in normal germinal center B cells: evidence of somatic hypermutation acting outside Ig loci. Proc Natl Acad Sci U S A. 1998 Sep 29;95(20):11816-21.

Peled JU, Kuang FL, Iglesias-Ussel MD, Roa S, Kalis SL, Goodman MF, Scharff MD. The biochemistry of somatic hypermutation. Annu Rev Immunol. 2008;26:481-511.

Petersen S, Casellas R, Reina-San-Martin B, Chen HT, Difilippantonio MJ, Wilson PC, Hanitsch L, Celeste A, Muramatsu M, Pilch DR, Redon C, Ried T, Bonner WM, Honjo T, Nussenzweig MC, Nussenzweig A. AID is required to initiate Nbs1/gamma-H2AX focus formation and mutations at sites of class switching. Nature. 2001 Dec 6;414(6864):660-5.

Petersen-Mahrt SK, Harris RS, Neuberger MS. AID mutates E. coli suggesting a DNA deamination mechanism for antibody diversification. Nature. 2002 Jul 4;418(6893):99-103.

Peters A, Storb U. Somatic hypermutation of immunoglobulin genes is linked to transcription initiation. Immunity. 1996 Jan;4(1):57-65.

Pérez-Durán P, de Yebenes VG, Ramiro AR. Oncogenic events triggered by AID, the adverse effect of antibody diversification. Carcinogenesis. 2007 Dec;28(12):2427-33.

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 Oct 4;102(40):14362-7.

Pham P, Bransteitter R, Petruska J, Goodman MF. Processive AID-catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation. Nature. 2003 Jul 3;424(6944):103-7.

Picker LJ, Butcher EC. Physiological and molecular mechanisms of lymphocyte homing. Annu Rev Immunol. 1992;10:561-91.

Poltoratsky VP, Wilson SH, Kunkel TA, Pavlov YI. Recombinogenic phenotype of human activation-induced cytosine deaminase. J Immunol. 2004 Apr 1;172(7):4308-13.

Poltoratsky V, Woo CJ, Tippin B, Martin A, Goodman MF, Scharff MD. Expression of error-prone polymerases in BL2 cells activated for Ig somatic hypermutation. Proc Natl Acad Sci U S A. 2001 Jul 3;98(14):7976-81.

Q

Quandt K, Frech K, Karas H, Wingender E, Werner T. MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. Nucleic Acids Res. 1995 Dec 11;23(23):4878-84.

Quong MW, Martensson A, Langerak AW, Rivera RR, Nemazee D, Murre C. Receptor editing and marginal zone B cell development are regulated by the helix-loop-helix protein, E2A. J Exp Med. 2004 Apr 19;199(8):1101-12.

Quong MW, Harris DP, Swain SL, Murre C. E2A activity is induced during B-cell activation to promote immunoglobulin class switch recombination. EMBO J. 1999 Nov 15;18(22):6307-18.

Quong MW, Massari ME, Zwart R, Murre C. A new transcriptional-activation motif restricted to a class of helix-loop-helix proteins is functionally conserved in both yeast and mammalian cells.
Mol Cell Biol. 1993 Feb;13(2):792-800.

R

Rada C, Di Noia JM, Neuberger MS. Mismatch recognition and uracil excision provide complementary paths to both Ig switching and the A/T-focused phase of somatic mutation.
Mol Cell. 2004 Oct 22;16(2):163-71.

Rada C, Williams GT, Nilsen H, Barnes DE, Lindahl T, Neuberger MS. Immunoglobulin isotype switching is inhibited and somatic hypermutation perturbed in UNG-deficient mice. Curr Biol. 2002 Oct 15;12(20):1748-55.

Rada C, Jarvis JM, Milstein C. AID-GFP chimeric protein increases hypermutation of Ig genes with no evidence of nuclear localization. Proc Natl Acad Sci U S A. 2002 May 14;99(10):7003-8.

Rada C, Ehrenstein MR, Neuberger MS, Milstein C. Hot spot focusing of somatic hypermutation in MSH2-deficient mice suggests two stages of mutational targeting. Immunity. 1998 Jul;9(1):135-41.

Ramiro AR, Jankovic M, Callen E, Difilippantonio S, Chen HT, McBride KM, Eisenreich TR, Chen J, Dickins RA, Lowe SW, Nussenzweig A, Nussenzweig MC. Role of genomic instability and p53 in AID-induced c-myc-Igh translocations. Nature. 2006 Mar 2;440(7080):105-9. Epub 2006 Jan 8.

Ramiro AR, Stavropoulos P, Jankovic M, Nussenzweig MC. Transcription enhances AID-mediated cytidine deamination by exposing single-stranded DNA on the nontemplate strand. Nat Immunol. 2003 May;4(5):452-6.

Raschke EE, Albert T, Eick D. Transcriptional regulation of the Ig kappa gene by promoter-proximal pausing of RNA polymerase II. J Immunol. 1999 Oct 15;163(8):4375-82.

Ratcliffe MJ. Antibodies, immunoglobulin genes and the bursa of Fabricius in chicken B cell development. Dev Comp Immunol. 2006;30(1-2):101-18.

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 Sep 1;102(5):565-75.

Reynaud CA, Aoufouchi S, Faili A, Weill JC. What role for AID: mutator, or assembler of the immunoglobulin mutasome? Nat Immunol. 2003 Jul;4(7):631-8.

Reynaud CA, Dahan A, Anquez V, Weill JC. Somatic hyperconversion diversifies the single Vh gene of the chicken with a high incidence in the D region. Cell. 1989 Oct 6;59(1):171-83.

Reynaud CA, Anquez V, Grimal H, Weill JC. A hyperconversion mechanism generates the chicken light chain preimmune repertoire. Cell. 1987 Feb 13;48(3):379-88.

Rogozin IB, Kolchanov NA. Somatic hypermutagenesis in immunoglobulin genes. II. Influence of neighbouring base sequences on mutagenesis. Biochim Biophys Acta. 1992 Nov 15;1171(1):11-8.

Romanow WJ, Langerak AW, Goebel P, Wolvers-Tettero IL, van Dongen JJ, 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 Feb;5(2):343-53.

Ronai D, Iglesias-Ussel MD, Fan M, Shulman MJ, Scharff MD. Complex regulation of somatic hypermutation by cis-acting sequences in the endogenous IgH gene in hybridoma cells. Proc Natl Acad Sci U S A. 2005 Aug 16;102(33):11829-34. Epub 2005 Aug 8.

Ross AL, Sale JE. The catalytic activity of REV1 is employed during immunoglobulin gene diversification in DT40. Mol Immunol. 2006 Apr;43(10):1587-94. Epub 2005 Nov 2.

Rückerl F, Busse B, Bachl J. Episomal vectors to monitor and induce somatic hypermutation in human Burkitt-Lymphoma cell lines. Mol Immunol. 2006 Apr;43(10):1645-52. Epub 2005 Nov 28.

Rückerl F, Mailhammer R, Bachl J. Dual reporter system to dissect cis- and trans-effects influencing the mutation rate in a hypermutating cell line. Mol Immunol. 2004 Nov;41(12):1135-43.

Ruiz M, Pallarès N, Contet V, Barbi V, Lefranc MP. The human immunoglobulin heavy diversity (IGHD) and joining (IGHJ) segments. Exp Clin Immunogenet. 1999;16(3):173-84.

Rush JS, Fugmann SD, Schatz DG. Staggered AID-dependent DNA double strand breaks are the predominant DNA lesions targeted to S mu in Ig class switch recombination. Int Immunol. 2004 Apr;16(4):549-57.

S
Sablitzky F, Wildner G, Rajewsky K. Somatic mutation and clonal expansion of B cells in an antigen-driven immune response. EMBO J. 1985 Feb;4(2):345-50.

Sale JE. Immunoglobulin diversification in DT40: a model for vertebrate DNA damage tolerance.
DNA Repair (Amst). 2004 Jul 2;3(7):693-702.

Sale JE, Calandrini DM, Takata M, Takeda S, Neuberger MS. Ablation of XRCC2/3 transforms immunoglobulin V gene conversion into somatic hypermutation. Nature. 2001 Aug 30;412(6850):921-6.

Sale JE, Neuberger MS. TdT-accessible breaks are scattered over the immunoglobulin V domain in a constitutively hypermutating B cell line. Immunity. 1998 Dec;9(6):859-69.

Saribasak H, Saribasak NN, Ipek FM, Ellwart JW, Arakawa H, Buerstedde JM. Uracil DNA glycosylase disruption blocks Ig gene conversion and induces transition mutations. J Immunol. 2006 Jan 1;176(1):365-71.

Sayegh CE, Quong MW, Agata Y, Murre C. E-proteins directly regulate expression of activation-induced deaminase in mature B cells. Nat Immunol. 2003 Jun;4(6):586-93.

Sayegh CE, Drury G, Ratcliffe MJ. Efficient antibody diversification by gene conversion in vivo in the absence of selection for V(D)J-encoded determinants. EMBO J. 1999 Nov 15;18(22):6319-28.

Schatz DG, Oettinger MA, Baltimore D. The V(D)J recombination activating gene, RAG-1.
Cell. 1989 Dec 22;59(6):1035-48.

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 Aug;5(8):1367-76.

Schneider TD, Stephens RM. Sequence logos: a new way to display consensus sequences. Nucleic Acids Res. 1990 Oct 25;18(20):6097-100.

Schoetz U. The role of Bach-2 and E2A for Immunoglobulin Gene Transcription and Repertoire Development. Diploma Thesis, Helmholtz Center Munich, Institute of Molecular Radiationbiology, 2005

Schoetz U, Cervelli M, Wang YD, Fiedler P, Buerstedde JM. E2A expression stimulates Ig hypermutation. J Immunol. 2006 Jul 1;177(1):395-400.

Schwarz EM, Krimpenfort P, Berns A, Verma IM. Immunological defects in mice with a targeted disruption in Bcl-3. Genes Dev. 1997 Jan 15;11(2):187-97.

Seo H, Masuoka M, Murofushi H, Takeda S, Shibata T, Ohta K. Rapid generation of specific antibodies by enhanced homologous recombination.

Nat Biotechnol. 2005 Jun;23(6):731-5.

Sha WC, Liou HC, Tuomanen EI, Baltimore D. Targeted disruption of the p50 subunit of NF-kappa B leads to multifocal defects in immune responses. Cell. 1995 Jan 27;80(2):321-30.

Shen HM, Ratnam S, Storb U. Targeting of the activation-induced cytosine deaminase is strongly influenced by the sequence and structure of the targeted DNA. Mol Cell Biol. 2005 Dec;25(24):10815-21.

Shen HM, Storb U. Activation-induced cytidine deaminase (AID) can target both DNA strands when the DNA is supercoiled. Proc Natl Acad Sci U S A. 2004 Aug 31;101(35):12997-3002.

Shen HM, Peters A, Baron B, Zhu X, Storb U. Mutation of BCL-6 gene in normal B cells by the process of somatic hypermutation of Ig genes. Science. 1998 Jun 12;280(5370):1750-2.

Shen CP, Kadesch T. B-cell-specific DNA binding by an E47 homodimer. Mol Cell Biol. 1995 Aug;15(8):4518-24.

Shimizu T, Azuma T, Ishiguro M, Kanjo N, Yamada S, Ohmori H. Normal immunoglobulin gene somatic hypermutation in Pol kappa-Pol iota double-deficient mice. Immunol Lett. 2005 May 15;98(2):259-64.

Shinkura R, Ito S, Begum NA, Nagaoka H, Muramatsu M, Kinoshita K, Sakakibara Y, Hijikata H, Honjo T. Separate domains of AID are required for somatic hypermutation and class-switch recombination. Nat Immunol. 2004 Jul;5(7):707-12.

Shinkura R, Tian M, Smith M, Chua K, Fujiwara Y, Alt FW. The influence of transcriptional orientation on endogenous switch region function. Nat Immunol. 2003 May;4(5):435-41.

Sigvardsson M, O'Riordan M, Grosschedl R. EBF and E47 collaborate to induce expression of the endogenous immunoglobulin surrogate light chain genes. Immunity. 1997 Jul;7(1):25-36.

Snapper CM, Zelazowski P, Rosas FR, Kehry MR, Tian M, Baltimore D, Sha WC. B cells from p50/NF-kappa B knockout mice have selective defects in proliferation, differentiation, germ-line CH transcription, and Ig class switching. J Immunol. 1996 Jan 1;156(1):183-91.

Sohail A, Klapacz J, Samaranayake M, Ullah A, Bhagwat AS. Human activation-induced cytidine deaminase causes transcription-dependent, strand-biased C to U deaminations. Nucleic Acids Res. 2003 Jun 15;31(12):2990-4.

Souto-Carneiro MM, Fritsch R, Sepúlveda N, Lagareiro MJ, Morgado N, Longo NS, Lipsky PE.
The NF-kappaB canonical pathway is involved in the control of the exonucleolytic processing of coding ends during V(D)J recombination. J Immunol. 2008 Jan 15;180(2):1040-9.

Spencer J, Dunn M, Dunn-Walters DK. Characteristics of sequences around individual nucleotide substitutions in IgVH genes suggest different GC and AT mutators. J Immunol. 1999 Jun 1;162(11):6596-601.

Stavnezer J, Schrader CE. Mismatch repair converts AID-instigated nicks to double-strand breaks for antibody class-switch recombination. Trends Genet. 2006 Jan;22(1):23-8.

Staudt LM, Lenardo MJ. Immunoglobulin gene transcription. Annu Rev Immunol. 1991;9:373-98.

van der Stoep N, Gorman JR, Alt FW. Reevaluation of 3'Ekappa function in stage- and lineage-specific rearrangement and somatic hypermutation. Immunity. 1998 Jun;8(6):743-50.

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 Dec 1;16(23):7118-29.

Sun XH, Baltimore D. An inhibitory domain of E12 transcription factor prevents DNA binding in E12 homodimers but not in E12 heterodimers. Cell. 1991 Jan 25;64(2):459-70. Erratum in: Cell. 1991 Aug 9;66(3):423.

Svanborg C, Godaly G, Hedlund M. Cytokine responses during mucosal infections: role in disease pathogenesis and host defence. Curr Opin Microbiol. 1999 Feb;2(1):99-105.

T
Ta VT, Nagaoka H, Catalan N, Durandy A, Fischer A, Imai K, Nonoyama S, Tashiro J, Ikegawa M, Ito S, Kinoshita K, Muramatsu M, Honjo T. AID mutant analyses indicate requirement for class-switch-specific cofactors. Nat Immunol. 2003 Sep;4(9):843-8.

Teng G, Papavasiliou FN. Immunoglobulin somatic hypermutation. Annu Rev Genet. 2007;41:107-20.

Terauchi A, Hayashi K, Kitamura D, Kozono Y, Motoyama N, Azuma T. A pivotal role for DNase I-sensitive regions 3b and/or 4 in the induction of somatic hypermutation of IgH genes. J Immunol. 2001 Jul 15;167(2):811-20.

Thompson CB, Neiman PE. Somatic diversification of the chicken immunoglobulin light chain gene is limited to the rearranged variable gene segment. Cell. 1987i Feb 13;48(3):369-78.

Thompson CB, Humphries EH, Carlson LM, Chen CL, Neiman PE. The effect of alterations in myc gene expression on B cell development in the bursa of Fabricius. Cell. 1987 Nov 6;51(3):371-81.

Tomlinson S. Complement defense mechanisms. Curr Opin Immunol. 1993 Feb;5(1):83-9.

Tonegawa S. Somatic generation of antibody diversity. Nature. 1983 Apr 14;302(5909):575-81.

Tumas-Brundage K, Manser T. The transcriptional promoter regulates hypermutation of the antibody heavy chain locus. J Exp Med. 1997 Jan 20;185(2):239-50.

Tsai HF, D'Avirro N, Selsing E. Gene conversion-like sequence transfers in a mouse antibody transgene: antigen selection allows sensitive detection of V region interactions based on homology.
Int Immunol. 2002 Jan;14(1):55-64.

W

Wahl MB, Caldwell RB, Kierzek AM, Arakawa H, Eyras E, Hubner N, Jung C, Soeldenwagner M, Cervelli M, Wang YD, Liebscher V, Buerstedde JM. Evaluation of the chicken transcriptome by SAGE of B cells and the DT40 cell line. BMC Genomics. 2004 Dec 21;5(1):98.

Wang CL, Harper RA, Wabl M. Genome-wide somatic hypermutation. Proc Natl Acad Sci U S A. 2004 May 11;101(19):7352-6. Epub 2004 Apr 29.

Weigert M, Gatmaitan L, Loh E, Schilling J, Hood L. Rearrangement of genetic information may produce immunoglobulin diversity. Nature. 1978 Dec 21-28;276(5690):785-90.

Weigert MG, Cesari IM, Yonkovich SJ, Cohn M. Variability in the lambda light chain sequences of mouse antibody. Nature. 1970 Dec 12;228(5276):1045-7.

Winding P, Berchtold MW. The chicken B cell line DT40: a novel tool for gene disruption experiments. J Immunol Methods. 2001 Mar 1;249(1-2):1-16.

Winter DB, Sattar N, Mai JJ, Gearhart PJ. Insertion of 2 kb of bacteriophage DNA between an immunoglobulin promoter and leader exon stops somatic hypermutation in a kappa transgene.
Mol Immunol. 1997 Apr;34(5):359-66.

Woo CJ, Martin A, Scharff MD. Induction of somatic hypermutation is associated with modifications in immunoglobulin variable region chromatin. Immunity. 2003 Oct;19(4):479-89.

Wu C, Ohmori Y, Bandyopadhyay S, Sen G, Hamilton T. Interferon-stimulated response element and NF kappa B sites cooperate to regulate double-stranded RNA-induced transcription of the IP-10 gene.
J Interferon Res. 1994 Dec;14(6):357-63.

Wuerffel RA, Du J, Thompson RJ, Kenter AL. Ig Sgamma3 DNA-specifc double strand breaks are induced in mitogen-activated B cells and are implicated in switch recombination. J Immunol. 1997 Nov 1;159(9):4139-44.

X

Xin XQ, Nelson C, Collins L, Dorshkind K. Kinetics of E2A basic helix-loop-helix-protein expression during myelopoiesis and primary B cell differentiation. J Immunol. 1993 Nov 15;151(10):5398-407. Erratum in: J Immunol 1994 Mar 1;152(5):1635.

Y

Yadav A, Olaru A, Saltis M, Setren A, Cerny J, Livák F. Identification of a ubiquitously active promoter of the murine activation-induced cytidine deaminase (AICDA) gene. Mol Immunol. 2006 Feb;43(6):529-41.

Yamazoe M, Sonoda E, Hochegger H, Takeda S. Reverse genetic studies of the DNA damage response in the chicken B lymphocyte line DT40. DNA Repair (Amst). 2004 Aug-Sep;3(8-9):1175-85.

Yoshikawa K, Okazaki IM, Eto T, Kinoshita K, Muramatsu M, Nagaoka H, Honjo T. AID enzyme-induced hypermutation in an actively transcribed gene in fibroblasts. Science. 2002 Jun 14;296(5575):2033-6.

Yu K, Chedin F, Hsieh CL, Wilson TE, Lieber MR. R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. Nat Immunol. 2003 May;4(5):442-51.

Z

Zan H, Shima N, Xu Z, Al-Qahtani A, Evinger Iii AJ, Zhong Y, Schimenti JC, Casali P. The translesion DNA polymerase theta plays a dominant role in immunoglobulin gene somatic hypermutation. EMBO J. 2005 Nov 2;24(21):3757-69.

Zhou C, Saxon A, Zhang K. Human activation-induced cytidine deaminase is induced by IL-4 and negatively regulated by CD45: implication of CD45 as a Janus kinase phosphatase in antibody diversification. J Immunol. 2003 Feb 15;170(4):1887-93.

Zhuang Y, Soriano P, Weintraub H. The helix-loop-helix gene E2A is required for B cell formation.
Cell. 1994 Dec 2;79(5):875-84.

8 Supplementary Information

8.1 *List of primers*

All primers are designated in 5' -> 3' direction

Primer	Sequence	Amplicon
1. Universal Primers binding the vector backbone of pKS(+):		
UC1	AGCGGATAACAATTTCACACAGGA	
UC2	CGCCAGGGTTTTCCCAGTCACGAC	
2. E2A transcription factor knockout		
PCR screening		
KO1/2f	GGCGGCTTGCGCACTGAGGGACCT	
KO1r	CGATTGAAGAACTCATTCCACTCAAATATACCC	
KO2r	TGTTGATATCCCGCAAGATACCTGGATTGA	
Controlf	CTCTGTTCTCCCCGAGGGACCTCAATGC	
Controlr	CCACTTCCAGGTGGATTCCAACAATG	
Primer for cDNA sequencing		
r2	TTTGGAGGATCAGGTTTGGA	
r3	GAGAGCAGCATAGCAGATGG	
r4	GTAACAGAGGAACCACGGC	
r5	CGGGAACCAACGAGATAAAG	
r6	TGTCTGAAACGAAAGCAAAC	
s10	GAGACCCCCAGATGACGCTT	
s11	CCCCCAGCATGCTCCACAAC	
s12	TACGGCGTCTCCAGCCACAC	
s13	GACCGGGATCGGGTTCCTGG	
ML13	GAATTAATTCAAGCTTGGGCCACCGGCTAG	
IR7	GGCCTTATTCCAAGCGGCTTCGGCCAGTAA	
Primer for RT-PCR		
AI1	CCCGCTAGCGCCACCATGGACAGCCTCTTGATGAAGAGGA	AID
AI2	CCCAGATCTTGCTTGTGAAGTCTTCTTATTGCTG	AID
E2A9	GGATCCGG CTGGATACACCCCCTCGAAACCTCCC	E12/ E47
E2A10	GGAAGAAGATCG TCCTCCTCATCCTCATCT	E12
E2A11	TCCTCCAGTGATAAGGCCTCGTCTGTACTG	E47
Cl1	AGTGGATCCGCCAGCCCAAGGTGGCC	IgL C region
Cl2	ATCGAATTCTATTAGCACTCGGACCT	IgL C region
EF6	GGGAAGCTTCGGAAGAAAGAAGCTAAAGACCATC	EF-1α
EF7	GGGACTAGTAGAAGAGCGTGCTCACGGGTCTGCC	EF-1α
Primer for sequencing		
VL1	GGGTCTAGACCTCTCAGCTTTTTCAGCAGAATAACCTCC	IgL VJ region

Primer	Sequence	Amplicon
VL2	GGGAAGCTTTGGGAAATACTGGTGATAGGTGGAT	IgL VJ region
VL3	GTGCGTGCGGGGCCGTCACTGATTGCCGTT	IgL VJ region

Primer	Sequence
3. Mapping of cis-elements	
Primer for PCR screening:	
Forward:	AGCTTGGAATTTAACCTCTCCTGTAAA
Reverse:	CCCACCGACTCTAGAGGATCATAATCAGCC
VJ intervening sequence of unrearranged IgL locus:	
Forward:	GGGGGATCCAGATCTGTGACCGGTGCAAGTGATAGAAAACT
Reverse:	TACAAAAACCTCCTGCCACTGCAAGGAGCGAGCTGATGGTTTTTACTGTCT
Targeting Arms for plasmid *pIgL$^{-,GFP2}$*	
5'-arm Forward:	GGGCTCGAGGGTACTGCGTTTTCCACAAAATTCTCACAG
5'-arm Reverse:	GGGAGATCTCTGCACTCTGGCACCGTTAAGCACCATCAC
3'-arm Forward:	GGGTGATCAAGATCTGCTAGCACTAGTGGATCCGTCGA
3'-arm Reverse:	GGGGAAAAGCGGCCGCCACTGGAAGGAGCTGAAGGCCAC

Fragment	Primer	Sequence
DNA fragments for plasmids belonging to the *'W'* series:		
***'A'*:**	Forward:	GAAGCTAGCTTCCGCCATGGCCTGGGCTCCTCTCC
	Reverse:	GAAACTAGTATTTTTTGACAGCACTTACCTGGACAGCTGAAAAACTGAA
***'B'*:**	Forward:	GGGGCTAGCGGTGGATGTGTTTGTTTTACAGAGG
	Reverse:	GAAGCTAGCGCAAATCTCTGCTAGGGACCTGGCG
***'C'*:**	Forward:	GGGGCTAGCGGTGGATGTGTTTGTTTTACAGAGG
	Reverse:	GAAGCTAGCGTGTGGCAGAGAGTCTACACATGGC
***'D'*:**	Forward:	GGGGCTAGCGGTGGATGTGTTTGTTTTACAGAGG
	Reverse:	GAAGCTAGCATGGAGCTGTACCATGCGGCCTGCT
***'E'*:**	Forward:	GGGGCTAGCGGTGGATGTGTTTGTTTTACAGAGG
	Reverse:	GAAGCTAGCAAGCTCAGGGTCTCAGTTTGGAGCT
***'F'*:**	Forward:	GGGGCTAGCGGTGGATGTGTTTGTTTTACAGAGG
	Reverse:	GAAGCTAGCATTGCTGCAGTGCAAACGCCCTGGT
***'G'*:**	Forward:	GGGGCTAGCGGTGGATGTGTTTGTTTTACAGAGG
	Reverse:	GGGACTAGTTGTTCAGATGGAACTTCTTATGTTC
***'I'*:**	Forward:	GGGGCTAGCGGTGGATGTGTTTGTTTTACAGAGG
	Reverse:	GAAGCTAGCATGGGATGGAAGGGCCCGTCTGGCC
***'K'*:**	Forward:	GAAGCTAGCTTTATGCTGGGAACAGGGGGAGTTC
	Reverse:	GAAGCTAGCATGGGATGGAAGGGCCCGTCTGGCC
***'L'*:**	Forward:	GAAGCTAGCAGGACTGTGCTGTCCTCATGCCCCT
	Reverse:	GAAGCTAGCATGGGATGGAAGGGCCCGTCTGGCC
***'M'*:**	Forward:	GAAGCTAGCCACGACAGCTGGGGCCACACAAAGA
	Reverse:	GAAGCTAGCATGGGATGGAAGGGCCCGTCTGGCC
***'N'*:**	Forward:	GAAGCTAGCGTCACAGGTTGTAACAGGCTGACAT
	Reverse:	GAAGCTAGCATGGGATGGAAGGGCCCGTCTGGCC
***'P'*:**	Forward:	GGGGCTAGCTCACAGAAACATTGAAATGGCTCCT
	Reverse:	GAAGCTAGCATGGGATGGAAGGGCCCGTCTGGCC
***'S'*:**	Forward:	GAAGCTAGCTTTATGCTGGGAACAGGGGGAGTTC
	Reverse:	GAAACTAGTATTGCTGCAGTGCAAACGCCCTGGT
DNA fragments for plasmids belonging to the *'0-2'* series 200bp end deletion:		
***'0-2'*:**	Forward:	GAAGCTAGCTTTATGCTGGGAACAGGGGGAGTTC
	Reverse:	GAAACTAGTATGGAGCTGTACCATGCGGCCTGCT
***'0.2-2'*:**	Forward:	GAGGAAGCTAGCGCCCCTGTAGGGCAGCTTTTAGCAC
	Reverse:	GAAACTAGTATGGAGCTGTACCATGCGGCCTGCT
***'0.4-2'*:**	Forward:	GAGGAAGCTAGCATTCATCCCACCCACCCAAACATGT
	Reverse:	GAAACTAGTATGGAGCTGTACCATGCGGCCTGCT
***'0.6-2'*:**	Forward:	GAGGAAGCTAGCCTCCATCCAGCCACTGGTGGGGTGC
	Reverse:	GAAACTAGTATGGAGCTGTACCATGCGGCCTGCT
***'0.8-2'*:**	Forward:	GAGGAAGCTAGCCCATGGCAGGCATTCATGACACTGG

Fragment	Primer	Sequence
	Reverse:	GAAACTAGTATGGAGCTGTACCATGCGGCCTGCT
'0-1.2':	Forward:	GAAGCTAGCTTTATGCTGGGAACAGGGGGAGTTC
	Reverse:	GAAACTAGTGAGCAGCTGCCTCAGGGCACAGTTTGGT
'0-1.4':	Forward:	GAAGCTAGCTTTATGCTGGGAACAGGGGGAGTTC
	Reverse:	GAAACTAGTAGCACATGCTCCTCAGTGGGTTGTTG
'0-1.6':	Forward:	GAAGCTAGCTTTATGCTGGGAACAGGGGGAGTTC
	Reverse:	GAAACTAGTCCGGAGCCCTGGGGTTGCTTTCTATC
'0-1.8':	Forward:	GAAGCTAGCTTTATGCTGGGAACAGGGGGAGTTC
	Reverse:	GAAACTAGTCCCCTGTGCGATCCAGGCCCCACGAT

DNA fragments for plasmids belonging to the *'2-4'* series 200bp end deletion

'2-4':	Forward:	GAAGCTAGCCACGACAGCTGGGGCCACACAAAGA
	Reverse:	GAAACTAGTATTGCTGCAGTGCAAACGCCCTGGT
'2.2-4':	Forward:	GAGGAAGCTAGCAGCTGTGCGGCCGGGGCATCCCCAAGC
	Reverse:	GAAACTAGTATTGCTGCAGTGCAAACGCCCTGGT
'2.4-4':	Forward:	GAGGAAGCTAGCCTTTGCTGCTGCTCGGGGTGGG
	Reverse:	GAAACTAGTATTGCTGCAGTGCAAACGCCCTGGT
'2.6-4':	Forward:	GAGGAAGCTAGCCGCTCCCACCACGCGTCAACCCAAATCC
	Reverse:	GAAACTAGTATTGCTGCAGTGCAAACGCCCTGGT
'2.8-4':	Forward:	GGGGCTAGCGAGGAATTAAATTAAATCAATAAAT
	Reverse:	GAAACTAGTATTGCTGCAGTGCAAACGCCCTGGT
'2-3.2':	Forward:	GAAGCTAGCCACGACAGCTGGGGCCACACAAAGA
	Reverse:	GAGACTAGTGCTTTGTGGCCCTCCTGCATCGGG
'2-3.4':	Forward:	GAAGCTAGCCACGACAGCTGGGGCCACACAAAGA
	Reverse:	GAGACTAGTCATCCCCCCACCTCCTGTGTGTGCT
'2-3.6':	Forward:	GAAGCTAGCCACGACAGCTGGGGCCACACAAAGA
	Reverse:	GAGACTAGTAGTGCAGGCTCAGCTGTGGGGCTGG
'2-3.8':	Forward:	GAAGCTAGCCACGACAGCTGGGGCCACACAAAGA
	Reverse:	GAGACTAGTCTTGCTGCTGTGCCGGGCAGCGGCG

DNA fragments for plasmids belonging to the *'0-4'* 1kb fragment deletion series

'0-1':	Forward:	GAAGCTAGCTTTATGCTGGGAACAGGGGGAGTTC
	Reverse:	GAAACTAGTGTGTGGCAGAGAGTCTACACATGGC
'0-2':	Forward:	GAAGCTAGCTTTATGCTGGGAACAGGGGGAGTTC
	Reverse:	GAAACTAGTATGGAGCTGTACCATGCGGCCTGCT
'0-3':	Forward:	GAAGCTAGCTTTATGCTGGGAACAGGGGGAGTTC
	Reverse:	GAAACTAGTAAGCTCAGGGTCTCAGTTTGGAGCT
'0-4':	Forward:	GAAGCTAGCTTTATGCTGGGAACAGGGGGAGTTC
	Reverse:	GAAACTAGTATTGCTGCAGTGCAAACGCCCTGGT
'1-2':	Forward:	GAAGCTAGCAGGACTGTGCTGTCCTCATGCCCCT
	Reverse:	GAAACTAGTATGGAGCTGTACCATGCGGCCTGCT
'2-3':	Forward:	GAAGCTAGCCACGACAGCTGGGGCCACACAAAGA
	Reverse:	GAAACTAGTAAGCTCAGGGTCTCAGTTTGGAGCT
'3-4':	Forward:	GAAGCTAGCGTCACAGGTTGTAACAGGCTGACAT
	Reverse:	GAAACTAGTATTGCTGCAGTGCAAACGCCCTGGT
'2-4':	Forward:	GAAGCTAGCCACGACAGCTGGGGCCACACAAAGA
	Reverse:	GAAACTAGTATTGCTGCAGTGCAAACGCCCTGGT
'1-4':	Forward:	GAAGCTAGCAGGACTGTGCTGTCCTCATGCCCCT
	Reverse:	GAAACTAGTATTGCTGCAGTGCAAACGCCCTGGT

DNA fragments for plasmids belonging to the *'0-4'* series 200 bp internal deletion

Forward primer for hybrid PCR: GAAGCTAGCTTTATGCTGGGAACAGGGGGAGTTC
Reverse primer for hybrid PCR: GAAACTAGTATTGCTGCAGTGCAAACGCCCTGGT

'0.2-4.0':	Forward:	GAAGCTAGCTGGCCCCTGTAGGAGCTTTTAGCAC
	Reverse:	GAAACTAGTATTGCTGCAGTGCAAACGCCCTGGT
'0-4Δ0.2-0.4':		
0.0-0.2	Forward:	GAAGCTAGCTTTATGCTGGGAACAGGGGGAGTTC
	Reverse:	GGATGAATGCATATGCTGCAAAATCTCCCTGGGGAGCGGG

Fragment	Primer	Sequence
0.4-4.0	Forward:	ATTTTGCAGCATATGCATTCATCCCACCCACCCAAACATG
	Reverse:	GAAACTAGTATTGCTGCAGTGCAAACGCCCTGGT
***‘0-4Δ0.4-0.6’*:**		
0.0-0.4	Forward:	GAAGCTAGCTTTATGCTGGGAACAGGGGGAGTTC
	Reverse:	CTGGATGGACATATGGAGATGTTGGTGTAGATGGAATAG
0.6-4.0	Forward:	CAACATCTCCATATGTCCATCCAGCCACTGGTGGGGTGCA
	Reverse:	GAAACTAGTATTGCTGCAGTGCAAACGCCCTGGT
***‘0-4Δ0.6-0.8’*:**		
0.0-0.6	Forward:	GAAGCTAGCTTTATGCTGGGAACAGGGGGAGTTC
	Reverse:	TGCCTGCCACATATGAGACATAGGGTGGGTGGGATGGCTG
0.8-4.0	Forward:	CCTATGTCTCATATGTGGCAGGCATTCATGACACTGGGTT
	Reverse:	GAAACTAGTATTGCTGCAGTGCAAACGCCCTGGT
***‘0-4Δ0.8-1.0’*:**		
0.0-0.8	Forward:	GAAGCTAGCTTTATGCTGGGAACAGGGGGAGTTC
	Reverse:	CACAGTCCTCATATGTGGAGCTGGGAGATCCCAGCCCATCT
1.0-4.0	Forward:	CCCAGCTCCACATATGAGGACTGTGCTGTCCTCATGCCCCT
	Reverse:	GAAACTAGTATTGCTGCAGTGCAAACGCCCTGGT
***‘0-4Δ1.0-1.2’*:**		
0.0-1.0	Forward:	GAAGCTAGCTTTATGCTGGGAACAGGGGGAGTTC
	Reverse:	TGAGCAGCTCATATGGTGTGGCAGAGAGTCTACACATGGC
1.2-4.0	Forward:	CTGCCACACCATATGAGCTGCTCATGCTGGATAAAGTCAC
	Reverse:	GAAACTAGTATTGCTGCAGTGCAAACGCCCTGGT
***‘0-4Δ1.2-1.4’*:**		
0.0-1.2	Forward:	GAAGCTAGCTTTATGCTGGGAACAGGGGGAGTTC
	Reverse:	ATTCAGCACCATATGGCCTCAGGGACAGTTTGGTAAATCC
1.4-4.0	Forward:	CCCTGAGGCCATATGGTGCTGAATTATACATCACAGCTCC
	Reverse:	GAAACTAGTATTGCTGCAGTGCAAACGCCCTGGT
***‘0-4Δ1.4-1.6’*:**		
0.0-1.4	Forward:	GAAGCTAGCTTTATGCTGGGAACAGGGGGAGTTC
	Reverse:	GCTGCAGCTCATATGATGCTCCTCAGTGGGTTGTTGCTTC
1.6-4.0	Forward:	GAGGAGCATCATATGAGCTGCAGCTCTTGCTCTGTCGTGT
	Reverse:	GAAACTAGTATTGCTGCAGTGCAAACGCCCTGGT
***‘0-4Δ1.6-1.8’*:**		
0.0-1.6	Forward:	GAAGCTAGCTTTATGCTGGGAACAGGGGGAGTTC
	Reverse:	GGATATCAGCATATGGGGTAGCCTGGGGTTGCTTTCTATC
1.8-4.0	Forward:	AGGCTACCCCATATGCTGATATCCTCACTAGCAGATACAC
	Reverse:	GAAACTAGTATTGCTGCAGTGCAAACGCCCTGGT
***‘0-4Δ1.8-2.0’*:**		
0.0-1.8	Forward:	GAAGCTAGCTTTATGCTGGGAACAGGGGGAGTTC
	Reverse:	GCTGTCGTGCATATGCTCCCTGTGCGATCCAGGCCCCACG
2.0-4.0	Forward:	CACAGGGAGCATATGCACGACAGCTGGGGCCACACAAAGA
	Reverse:	GAAACTAGTATTGCTGCAGTGCAAACGCCCTGGT
***‘0-4Δ2.0-2.2’*:**		
0.0-2.0	Forward:	GAAGCTAGCTTTATGCTGGGAACAGGGGGAGTTC
	Reverse:	CTCACAGCTCATATGATGGAGCTGTACCATGCGGCCTGCT
2.2-4.0	Forward:	CAGCTCCATCATATGAGCTGTGAGGCCGGGGCATCCCCAA
	Reverse:	GAAACTAGTATTGCTGCAGTGCAAACGCCCTGGT
***‘0-4Δ2.2-2.4’*:**		
0.0-2.2	Forward:	GAAGCTAGCTTTATGCTGGGAACAGGGGGAGTTC
	Reverse:	GCAGCAAAGCATATGGCATGGTGGGGCTGAGCGTGCTGCA
2.4-4.0	Forward:	CCACCATGCCATATGCTTTGCTGCTGCTGCTCGGGGTGGG
	Reverse:	GAAACTAGTATTGCTGCAGTGCAAACGCCCTGGT
***‘0-4Δ2.4-2.6’*:**		
0.0-2.4	Forward:	GAAGCTAGCTTTATGCTGGGAACAGGGGGAGTTC
	Reverse:	TGGGAGCGGCATATGACTCAGTCTGCAAAGGCCCCAACCT
2.6-4.0	Forward:	AGACTGAGTCATATGCCGCTCCCACCACGCGTCAACCCAA
	Reverse:	GAAACTAGTATTGCTGCAGTGCAAACGCCCTGGT
***‘0-4Δ2.6-2.8’*:**		

Fragment	Primer	Sequence
0.0-2.6	Forward:	GAAGCTAGCTTTATGCTGGGAACAGGGGGAGTTC
	Reverse:	TAATTCCTCCATATGGCAGGGGTGGCACATGGGGACAGAG
2.8-4.0	Forward:	CCACCCCTGCCATATGGAGGAATTAAATTAAATCAATAAAT
	Reverse:	GAAACTAGTATTGCTGCAGTGCAAACGCCCTGGT
***'0-4Δ2.8-3.0'*:**		
0.0-2.8	Forward:	GAAGCTAGCTTTATGCTGGGAACAGGGGGAGTTC
	Reverse:	ACCTGTGACCATATGTTTAATTGCTGTGTGATGGCTCTGA
3.0-4.0	Forward:	GCAATTAAACATATGGTCACAGGTTGTAACAGGCTGACAT
	Reverse:	GAAACTAGTATTGCTGCAGTGCAAACGCCCTGGT
***'0-4Δ3.0-3.2'*:**		
0.0-3.0	Forward:	GAAGCTAGCTTTATGCTGGGAACAGGGGGAGTTC
	Reverse:	TGTCAGTGTCATATGAAGCTCAGGGTCTCAGTTTGGAGCT
3.2-4.0	Forward:	CCTGAGCTTCATATGACACTGACAACACAATGTGAGCTGA
	Reverse:	GAAACTAGTATTGCTGCAGTGCAAACGCCCTGGT
***'0-4Δ3.2-3.4'*:**		
0.0-3.2	Forward:	GAAGCTAGCTTTATGCTGGGAACAGGGGGAGTTC
	Reverse:	TGCTGGCTGCATATGGCATCTTGTGGCCCTCCTGCATCGG
3.4-4.0	Forward:	ACAAGATGCCATATGCAGCCAGCAGCTGCCCTGCACTAAG
	Reverse:	GAAACTAGTATTGCTGCAGTGCAAACGCCCTGGT
***'0-4Δ3.4-3.6'*:**		
0.0-3.4	Forward:	GAAGCTAGCTTTATGCTGGGAACAGGGGGAGTTC
	Reverse:	ATCTGGGAGCATATGCATCCCCCCACCTCCTCTGTGTGTG
3.6-4.0	Forward:	GGGGGGATGCATATGCTCCCAGATGTGCTGACCGCAGCCA
	Reverse:	GAAACTAGTATTGCTGCAGTGCAAACGCCCTGGT
***'0-4Δ3.6-3.8'*:**		
0.0-3.6	Forward:	GAAGCTAGCTTTATGCTGGGAACAGGGGGAGTTC
	Reverse:	CTGGCTCTGCATATGAGTGCAGGCTCAGCTGTGGGGCTGG
3.8-4.0	Forward:	GCCTGCACTCATATGCAGAGCCAGGAGCAGGAAATGCTGA
	Reverse:	GAAACTAGTATTGCTGCAGTGCAAACGCCCTGGT
***'0.0-3.8'*: Forward:**		GAAGCTAGCTTTATGCTGGGAACAGGGGGAGTTC
	Reverse:	GAAACTAGTGTTGCTGCTGTGCCGGGCAGCGGCG

DNA fragments for plasmids belonging to the *'0-4'* series 400 bp internal deletions

Forward primer for hybrid PCR: GAAGCTAGCTTTATGCTGGGAACAGGGGGAGTTC

Reverse primer for hybrid PCR: GAAACTAGTATTGCTGCAGTGCAAACGCCCTGGT

'0.4-4': Forward:		GGGGCTAGCCATTCATCCCACCCACCCAAACATG
	Reverse:	GAAACTAGTATTGCTGCAGTGCAAACGCCCTGGT
***'0-4Δ0.4-0.8'*:**		
0-0.4	Forward:	GAAGCTAGCTTTATGCTGGGAACAGGGGGAGTTC
	Reverse:	TGCCTGCCACATATGGAGATGTTGGTGTAGGATGGAATAG
0.8-4	Forward:	CAACATCTCCATATGTGGCAGGCATTCATGACACTGGGTT
	Reverse:	GAAACTAGTATTGCTGCAGTGCAAACGCCCTGGT
***'0-4Δ0.8-1.2'*:**		
0-0.8	Forward:	GAAGCTAGCTTTATGCTGGGAACAGGGGGAGTTC
	Reverse:	TGAGCAGCTCATATGTGGAGCTGGGAGATCCCAGCCCATCT
1.2-4	Forward:	CCCAGCTCCACATATGAGCTGCTCATGCTGGATAAAGTCAC
	Reverse:	GAAACTAGTATTGCTGCAGTGCAAACGCCCTGGT
***'0-4Δ1.2-1.6'*:**		
0-1.2	Forward:	GAAGCTAGCTTTATGCTGGGAACAGGGGGAGTTC
	Reverse:	GCTGCAGCTCATATGGCCTCAGGGACAGTTTGGTAAATCC
1.6-4	forward:	CCCTGAGGCCATATGAGCTGCAGCTCTTGCTCTGTCGTGT
	Reverse:	GAAACTAGTATTGCTGCAGTGCAAACGCCCTGGT
***'0-4Δ1.6-2.0'*:**		
0-1.6	Forward:	GAAGCTAGCTTTATGCTGGGAACAGGGGGAGTTC
0-1.6	Reverse:	GCTGTCGTGCATATGGGGTAGCCTGGGGTTGCTTTCTATC
2.0-4	Forward:	AGGCTACCCCATATGCACGACAGCTGGGGCCACACAAAGA
2.0-4	Reverse:	GAAACTAGTATTGCTGCAGTGCAAACGCCCTGGT
***'0-4Δ2.0-2.4'*:**		

0-2.0 Forward:	GAAGCTAGCTTTATGCTGGGAACAGGGGGAGTTC
0-2.0 Reverse:	GAGACTAGTGACGTCATGGAGCTGTACCATGCGGCCTGCT
2.4-4 Forward:	GAGGACGTCCTTTGCTGCTGCTGCTCGGGGTGGG
2.4-4 Reverse:	GAAACTAGTATTGCTGCAGTGCAAACGCCCTGGT
***'0-4Δ2.4-2.8'*:**	
0-2.4 Forward:	GAAGCTAGCTTTATGCTGGGAACAGGGGGAGTTC
0-2.4 Reverse:	GTGATGGCTCTGCATATGACTCAGTCTGCAAAGGCCCCAACCT
2.8-4 Forward:	CGAGACTGAGTCATATGCAGAGCCATCACACAGCAATTAAAGAG
2.8-4 Reverse:	GAAACTAGTATTGCTGCAGTGCAAACGCCCTGGT
***'0-4Δ2.8-3.2'*:**	
0-2.8 Forward:	GAAGCTAGCTTTATGCTGGGAACAGGGGGAGTTC
0-2.8 Reverse:	TGTCAGTGTCATATGTTTAATTGCTGTGTGATGGCTCTGA
3.2-4 Forward:	GCAATTAAACATATGACACTGACAACACAATGTGAGCTGA
3.2-4 Reverse:	GAAACTAGTATTGCTGCAGTGCAAACGCCCTGGT
***'0-4Δ3.2-3.6'*:**	
0-3.2 Forward:	GAAGCTAGCTTTATGCTGGGAACAGGGGGAGTTC
0-3.2 Reverse:	GAGACTAGTGACGTCGCATCTTGTGGCCCTCCTGCATCGG
3.6-4 Forward:	GAGGACGTCCTCCCAGATGTGCTGACCGCAGCCA
3.6-4 Reverse:	GAAACTAGTATTGCTGCAGTGCAAACGCCCTGGT
***'0-3.6'*:**	
Forward:	GAAGCTAGCTTTATGCTGGGAACAGGGGGAGTTC
Reverse:	GGGACTAGTAGTGCAGGCTCAGCTGTGGGGCTGG

DNA fragments for plasmids belonging to the '2-3' 50 bp fragment end deletion series

***'2.00-2.05'*:** Forward:	GAAGCTAGCCACGACAGCTGGGGCCACACAAAGA
Reverse:	GGGACTAGTCTGCCCACAGTAACCCAGCTCTTTG
***'2.00-2.10'*:** Forward:	GAAGCTAGCCACGACAGCTGGGGCCACACAAAGA
Reverse:	GGGACTAGTGCTCCCGGGGCTATTCTGAGCCCCCA
***'2.00-2.15'*:** Forward:	GAAGCTAGCCACGACAGCTGGGGCCACACAAAGA
Reverse:	GGGACTAGGCTGTGTGGCCCTGGCCTGGTGCT
***'2.00-2.20'*:** Forward:	GAAGCTAGCCACGACAGCTGGGGCCACACAAAGA
Reverse:	GGGACTAGTGGTGGGGCTGAGCGTGCTGCACCCT
***'2.00-2.25'*:** Forward:	GAAGCTAGCCACGACAGCTGGGGCCACACAAAGA
Reverse:	GGGACTAGAGACTGAGAAGTAAATTTAGCTTGG
***'2.00-2.30'*:** Forward:	GAAGCTAGCCACGACAGCTGGGGCCACACAAAGA
Reverse:	GGGACTAGCGTTTCCTTTTTTTGGCCGGCGTGGG
***'2.00-2.35'*:** Forward:	GAAGCTAGCCACGACAGCTGGGGCCACACAAAGA
Reverse:	GGGACTAGGCGCGGCGCTCGCTCTGCTTCACAC
***'2.00-2.40'*:** Forward:	GAAGCTAGCCACGACAGCTGGGGCCACACAAAGA
Reverse:	GGGACTAGGCTCAGTCTGGCAAAGCCCCAACCTG
***'2.00-2.45'*:** Forward:	GAAGCTAGCCACGACAGCTGGGGCCACACAAAGA
Reverse:	GGGACTAGCCCATCCCATGTGCCCAGGCCGTGG
***'2.00-2.50'*:** Forward:	GAAGCTAGCCACGACAGCTGGGGCCACACAAAGA
Reverse:	GGGACTAGTCACCTGAAGTGTTGGGGTGTGGGTG
***'2.00-2.55'*:** Forward:	GAAGCTAGCCACGACAGCTGGGGCCACACAAAGA
Reverse:	GGGACTAGAGGCCCATGCGTGGGGGGGTCAGC
***'2.00-2.60'*:** Forward:	GAAGCTAGCCACGACAGCTGGGGCCACACAAAGA
Reverse:	GGGACTAGGGCAGGGGTGGCACATGGGGACAGAG
***'2.00-2.65'*:** Forward:	GAAGCTAGCCACGACAGCTGGGGCCACACAAAGA
Reverse:	GGGACTAGCAGGAGTCGTGGGATTAACTCAGG
***'2.00-2.70'*:** Forward:	GAAGCTAGCCACGACAGCTGGGGCCACACAAAGA
Reverse:	GGGACTAGCTGCTCTGCATTTTGGGCATCTCCAG
***'2.00-2.75'*:** Forward:	GAAGCTAGCCACGACAGCTGGGGCCACACAAAGA
Reverse:	GGGACTAGTCCGCAGGGGAACACCCATAACTCC
***'2.00-2.80'*:** Forward:	GAAGCTAGCCACGACAGCTGGGGCCACACAAAGA
Reverse:	GGGACTAGCTCTTTAATTGCTGTGTGATGGCTCTG
***'2.00-2.85'*:** Forward:	GAAGCTAGCCACGACAGCTGGGGCCACACAAAGA
Reverse:	GGGACTAGGGTGCTGGCAGCTGAGCCCGCCTAAAAC
***'2.00-2.90'*:** Forward:	GAAGCTAGCCACGACAGCTGGGGCCACACAAAGA

Reverse:	GGGACTAGACAGCCATGCAAATGCTCTCCTCTTTGC
'2.00-2.95': Forward:	GAAGCTAGCCACGACAGCTGGGGCCACACAAAGA
Reverse:	GGGACTAGTTTCCAGATAACACCATCCCAGCTGC
'2.05-3.00': Forward:	GACGCTAGCGAAACCCGAAAACAAGAGCTGGGGGCTC
Reverse:	GAAACTAGTAAGCTCAGGGTCTCAGTTTGGAGCT
'2.10-3.00': Forward:	GACGCTAGCCACCAGGCCCAGGGCCACACAGCCC
Reverse:	GAAACTAGTAAGCTCAGGGTCTCAGTTTGGAGCT
'2.15-3.00': Forward:	GACGCTAGCAGGAAGGCACAGCGCTGTCAGGGTGC
Reverse:	GAAACTAGTAAGCTCAGGGTCTCAGTTTGGAGCT
'2.20-3.00': Forward:	GACGCTAGCCTGTGCGGCCGGGGCATCCCCAAGC
Reverse:	GAAACTAGTAAGCTCAGGGTCTCAGTTTGGAGCT
'2.25-3.00': Forward:	GACGCTAGCAGAAACTGAAGCTGAGGGGCCCACG
Reverse:	GAAACTAGTAAGCTCAGGGTCTCAGTTTGGAGCT
'2.30-3.00': Forward:	GACGCTAGCGTCTCCAGAAAGCACTGACGTGTGA
Reverse:	GAAACTAGTAAGCTCAGGGTCTCAGTTTGGAGCT
'2.35-3.00': Forward:	GACGCTAGCGCCGCCATGTCACACACCTCAGGTT
Reverse:	GAAACTAGTAAGCTCAGGGTCTCAGTTTGGAGCT
'2.40-3.00': Forward:	GACGCTAGCTGCTGCTCGGGGTGGGTGCCCACGG
Reverse:	GAAACTAGTAAGCTCAGGGTCTCAGTTTGGAGCT
'2.45-3.00': Forward:	GACGCTAGCCACGTACACACACTTGCACACCCACACC
Reverse:	GAAACTAGTAAGCTCAGGGTCTCAGTTTGGAGCT
'2.50-3.00': Forward:	GACGCTAGCGCAGATGGGTGCCCCCCAGGCTGACC
Reverse:	GAAACTAGTAAGCTCAGGGTCTCAGTTTGGAGCT
'2.55-3.00': Forward:	GACGCTAGC CACTGCTCCATCCGTGTCTCTGTCC
Reverse:	GAAACTAGTAAGCTCAGGGTCTCAGTTTGGAGCT
'2.60-3.00': Forward:	GACGCTAGC ACCACGCGTCAACCCAAATCCTGAG
Reverse:	GAAACTAGTAAGCTCAGGGTCTCAGTTTGGAGCT
'2.65-3.00': Forward:	GACGCTAGC CCAGCGTCCATGGCAGACTGGAGAT
Reverse:	GAAACTAGTAAGCTCAGGGTCTCAGTTTGGAGCT
'2.70-3.00': Forward:	GACGCTAGC CTGAATCTGAGAGATGAAATGGAGT
Reverse:	GAAACTAGTAAGCTCAGGGTCTCAGTTTGGAGCT
'2.75-3.00': Forward:	GACGCTAGC CCAGCTGTAGGAAGCTCAGAGCCATC
Reverse:	GAAACTAGTAAGCTCAGGGTCTCAGTTTGGAGCT
'2.80-3.00': Forward:	GACGCTAGC ATTAAATCAATAAATGTTTTAGGCG
Reverse:	GAAACTAGTAAGCTCAGGGTCTCAGTTTGGAGCT
'2.85-3.00': Forward:	GACGCTAGC CGAAACAGCCCGCTTGCAAAGAGG
Reverse:	GAAACTAGTAAGCTCAGGGTCTCAGTTTGGAGCT
'2.90-3.00': Forward:	GACGCTAGC CAGCAACCGCCTGTTGTGCAGCTGG
Reverse:	GAAACTAGTAAGCTCAGGGTCTCAGTTTGGAGCT
'2.95-3.00': Forward:	GACGCTAGC AGCCCAGGAGGGGTAAACAGCTCC
Reverse:	GAAACTAGTAAGCTCAGGGTCTCAGTTTGGAGCT

DNA fragments for plasmids belonging to the '2-3' 50 bp fragment inner deletion series
Forward primer for hybrid PCR: GAAGCTAGCCACGACAGCTGGGGCCACACAAAGA
Reverse primer for hybrid PCR: GAAACTAGTAAGCTCAGGGTCTCAGTTTGGAGCT

'2-3Δ2.05-2.10':	
2.00-2.05 Forward:	GAAGCTAGCCACGACAGCTGGGGCCACACAAAGA
Reverse:	CATATG AGCCTGCTGCCCACAGTAACCCAGC
2.10-3.00 Forward:	CAGCAGGCT CATATG ACCAGGCCCAGGGCCACACAGCCCT
Reverse:	GAAACTAGTAAGCTCAGGGTCTCAGTTTGGAGCT
'2-3Δ2.10-2.15':	
2.00-2.10 Forward:	GAAGCTAGCCACGACAGCTGGGGCCACACAAAGA
Reverse:	TCTGAGCCCCATATGGCTTGCCCTCTCCCCCAGCCCTGCC
2.15-3.00 Forward:	AGGGCAAGCCATATGGGGCTCAGAATAGCTGTCAGGGTGC
Reverse:	GAAACTAGTAAGCTCAGGGTCTCAGTTTGGAGCT
'2-3Δ2.15-2.20':	
2.00-2.15 Forward:	GAAGCTAGCCACGACAGCTGGGGCCACACAAAGA
Reverse:	GGCCGCACACATATGCCAGCTCTTGCTGTGCCTTCCTGGA
2.20-3.00 Forward:	AAGAGCTGGCATATGTGTGCGGCCGGGGCATCCCCAAGCT

2.20-3.00 Reverse:	GAAACTAGTAAGCTCAGGGTCTCAGTTTGGAGCT
***'2-3Δ2.20-2.25'*:**	
2.00-2.20 Forward:	GAAGCTAGCCACGACAGCTGGGGCCACACAAAGA
Reverse:	TTCAGTTTCCATATGGCTGCATGGTGGGGCTGAGCGTGCT
2.25-3.00 Forward:	CCATGCAGCCATATGGAAACTGAAGCTGAGGGGCCCACGC
Reverse:	GAAACTAGTAAGCTCAGGGTCTCAGTTTGGAGCT
***'2-3Δ2.25-2.30'*:**	
2.00-2.25 Forward:	GAAGCTAGCCACGACAGCTGGGGCCACACAAAGA
Reverse:	TTCTGGAGACATATGTGATTGGAGACTGAGAAGTAAATTT
2.30-3.00 Forward:	CTCCAATCACATATGTCTCCAGAAAGCACTGACGTGTGAA
Reverse:	GAAACTAGTAAGCTCAGGGTCTCAGTTTGGAGCT
***'2-3Δ2.30-2.35'*:**	
2.00-2.30 Forward:	GAAGCTAGCCACGACAGCTGGGGCCACACAAAGA
Reverse:	ACATGGCGGCATATGCTGTTTCGTTTCCTTTTTTGGCCG
2.35-3.00 Forward:	ACGAAACAGCATATGCCGCCATGTCACACACCTCAGGTTG
Reverse:	GAAACTAGTAAGCTCAGGGTCTCAGTTTGGAGCT
***'2-3Δ2.35-2.40'*:**	
2.00-2.35 Forward:	GAAGCTAGCCACGACAGCTGGGGCCACACAAAGA
Reverse:	CCGAGCAGCCATATGCCGGTTTGCGCGGCGCTCGCTCTGC
2.40-3.00 Forward:	GCAAACCGGCATATGGCTGCTCGGGGTGGGTGCCCACGGC
Reverse:	GAAACTAGTAAGCTCAGGGTCTCAGTTTGGAGCT
***'2-3Δ2.40-2.45'*:**	
2.00-2.40 Forward:	GAAGCTAGCCACGACAGCTGGGGCCACACAAAGA
Reverse:	GTGTGTACGCATATGAGCAAAGCTCAGTCTGGCAAAGCCC
2.45-3.00 Forward:	AGCTTTGCTCATATGCGTACACACACTTGCACACCCACAC
Reverse:	GAAACTAGTAAGCTCAGGGTCTCAGTTTGGAGCT
***'2-3Δ2.45-2.50'*:**	
2.00-2.45 Forward:	GAAGCTAGCCACGACAGCTGGGGCCACACAAAGA
Reverse:	ACCCATCTGCATATGTGTGTACCCCATCCCATGTGCCCAG
2.50-3.00 Forward:	GGGTACACACATATGCAGATGGGTGCCCCCCAGGCTGACC
Reverse:	GAAACTAGTAAGCTCAGGGTCTCAGTTTGGAGCT
***'2-3Δ2.50-2.55'*:**	
2.00-2.50 Forward:	GAAGCTAGCCACGACAGCTGGGGCCACACAAAGA
Reverse:	TGGAGCAGTCATATGCACCAGCATCACCTGAAGTGTTGGG
2.55-3.00 Forward:	ATGCTGGTGCATATGACTGCTCCATCCGTGTCTCTGTCCC
Reverse:	GAAACTAGTAAGCTCAGGGTCTCAGTTTGGAGCT
***'2-3Δ2.55-2.60'*:**	
2.00-2.55 Forward:	GAAGCTAGCCACGACAGCTGGGGCCACACAAAGA
Reverse:	GACGCGTGGCATATGGTGGGGCCAGGCCCATGCGTGGGGG
2.60-3.00 Forward:	TGGCCCCACCATATGCCACGCGTCAACCCAAATCCTGAGT
Reverse:	GAAACTAGTAAGCTCAGGGTCTCAGTTTGGAGCT
***'2-3Δ2.60-2.65'*:**	
2.00-2.60 Forward:	GAAGCTAGCCACGACAGCTGGGGCCACACAAAGA
Reverse:	TGGACGCTGCATATGTGGGAGCGGGCAGGGGTGGCACATG
2.65-3.00 Forward:	CCGCTCCCACATATGCAGCGTCCATGGCAGACTGGAGATG
Reverse:	GAAACTAGTAAGCTCAGGGTCTCAGTTTGGAGCT
***'2-3Δ2.65-2.70'*:**	
2.00-2.65 Forward:	GAAGCTAGCCACGACAGCTGGGGCCACACAAAGA
Reverse:	CTCAGATTCCATATGGAAGCAGGCAGGAGTCGTGGGATTA
2.70-3.00 Forward:	GCCTGCTTCCATATGGAATCTGAGAGATGAAATGGAGTTA
Reverse:	GAAACTAGTAAGCTCAGGGTCTCAGTTTGGAGCT
***'2-3Δ2.70-2.75'*:**	
2.00-2.70 Forward:	GAAGCTAGCCACGACAGCTGGGGCCACACAAAGA
Reverse:	CTACAGCTGCATATGAGGGAAACCTGCTCTGCATTTTGGG
2.75-3.00 Forward:	GGTTTCCCTCATATGCAGCTGTAGGAAGCTCAGAGCCATC
Reverse:	GAAACTAGTAAGCTCAGGGTCTCAGTTTGGAGCT
***'2-3Δ2.75-2.80'*:**	
2.00-2.75 Forward:	GAAGCTAGCCACGACAGCTGGGGCCACACAAAGA
Reverse:	TTGATTTAACATATGGGGCTCCGCCGCAGGGGAACACCCA

2.80-3.00 Forward:	GCGGAGCCCCATATGTTAAATCAATAAATGTTTTAGGCGG
Reverse:	GAAACTAGTAAGCTCAGGGTCTCAGTTTGGAGCT
'2-3Δ2.80-2.85':	
2.00-2.80 Forward:	GAAGCTAGCCACGACAGCTGGGGCCACACAAAGA
Reverse:	GGCTGTTTCCATATGTTTAATTCCTCTTTAATTGCTGTGT
2.85-3.00 Forward:	GGAATTAAACATATGGAAACAGCCCGCTTGCAAAGAGGAG
Reverse:	GAAACTAGTAAGCTCAGGGTCTCAGTTTGGAGCT
'2-3Δ2.85-2.90':	
2.00-2.85 Forward:	GAAGCTAGCCACGACAGCTGGGGCCACACAAAGA
Reverse:	GCGGTTGCTCATATGGGTCAGGTGGTGCTGGCAGCTGAGC
2.90-3.00 Forward:	CACCTGACCCATATGAGCAACCGCCTGTTGTGCAGCTGGG
Reverse:	GAAACTAGTAAGCTCAGGGTCTCAGTTTGGAGCT
'2-3Δ2.90-2.95':	
2.00-2.90 Forward:	GAAGCTAGCCACGACAGCTGGGGCCACACAAAGA
Reverse:	CTCCTGGGCCATATGGTTTTGCCACAGCCATGCAAATGCT
2.95-3.00 Forward:	TGGCAAAACCATATGGCCCAGGAGGGGTAAACAGCTCCAA
Reverse:	GAAACTAGTAAGCTCAGGGTCTCAGTTTGGAGCT

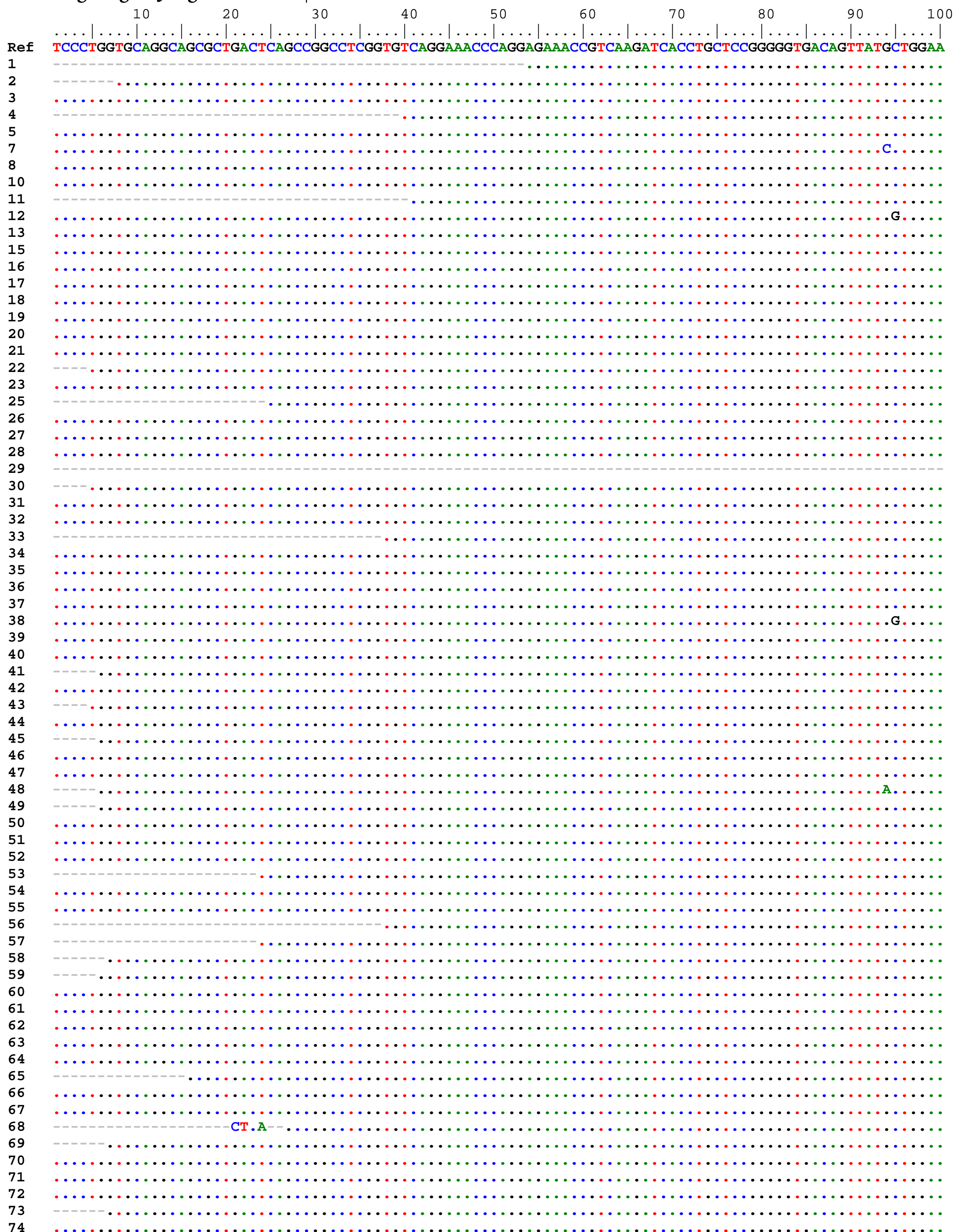
Reconstitution and multimerization of *'2.2-2.4'*

'2.2-2.4': Forward:	GAGGCTAGCAGCTGTGCGGCCGGGGCATCCCCAA
Reverse:	GAGACTAGTGGCGCGCCAGCTCAGTCTGGCAAAGCCCCAACCT

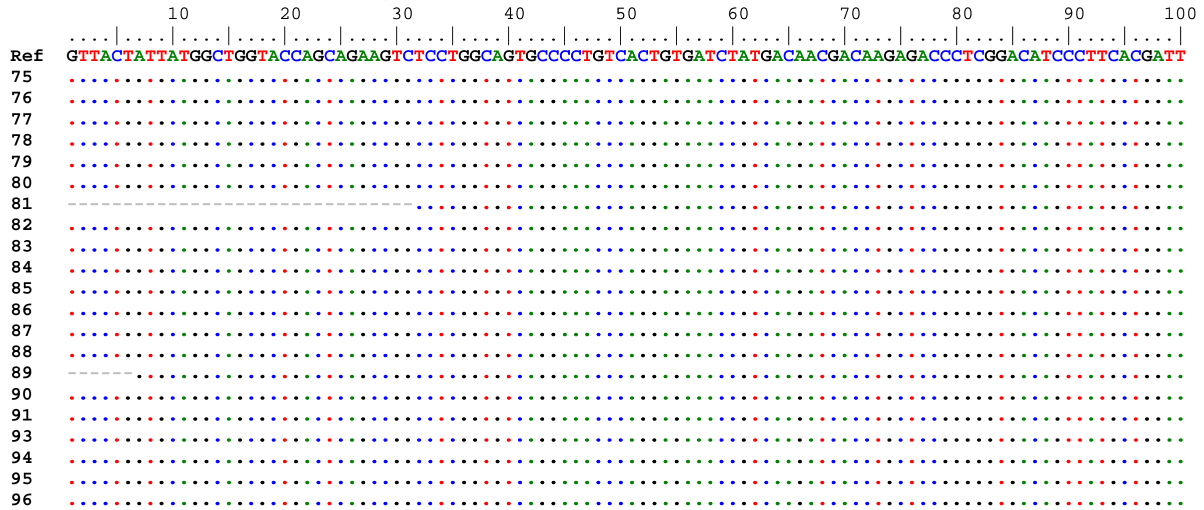
8.2 *Sequencing*

8.2.1 Sequencing of AID^{R} ψV^{-}E2A$^{-/-}$

Rearranged IgL VJ region of AID^{R} ψV^{-}E2A$^{-/-}$



Rearranged IgL VJ region of AID^{R} ψV^{-}E2A$^{-/-}$

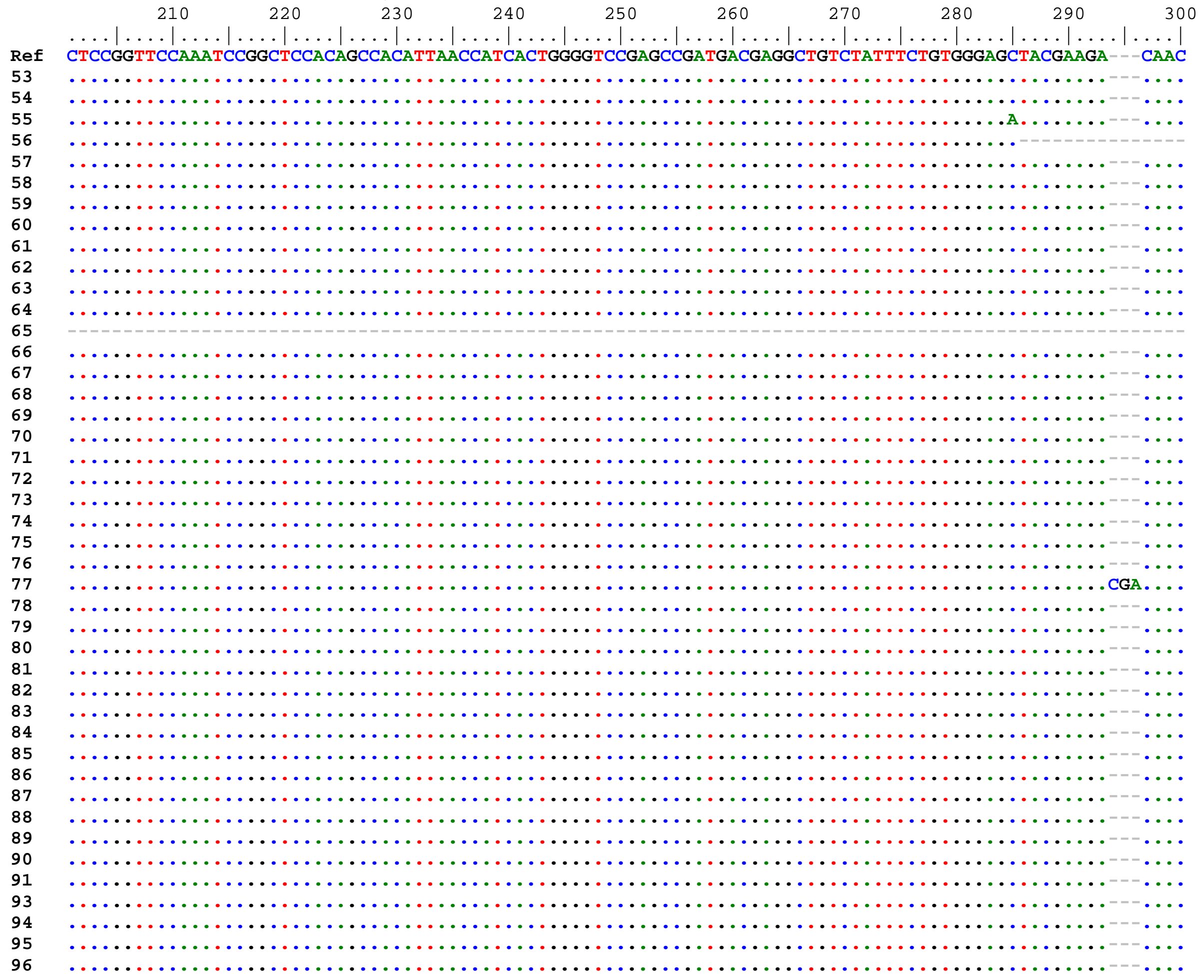


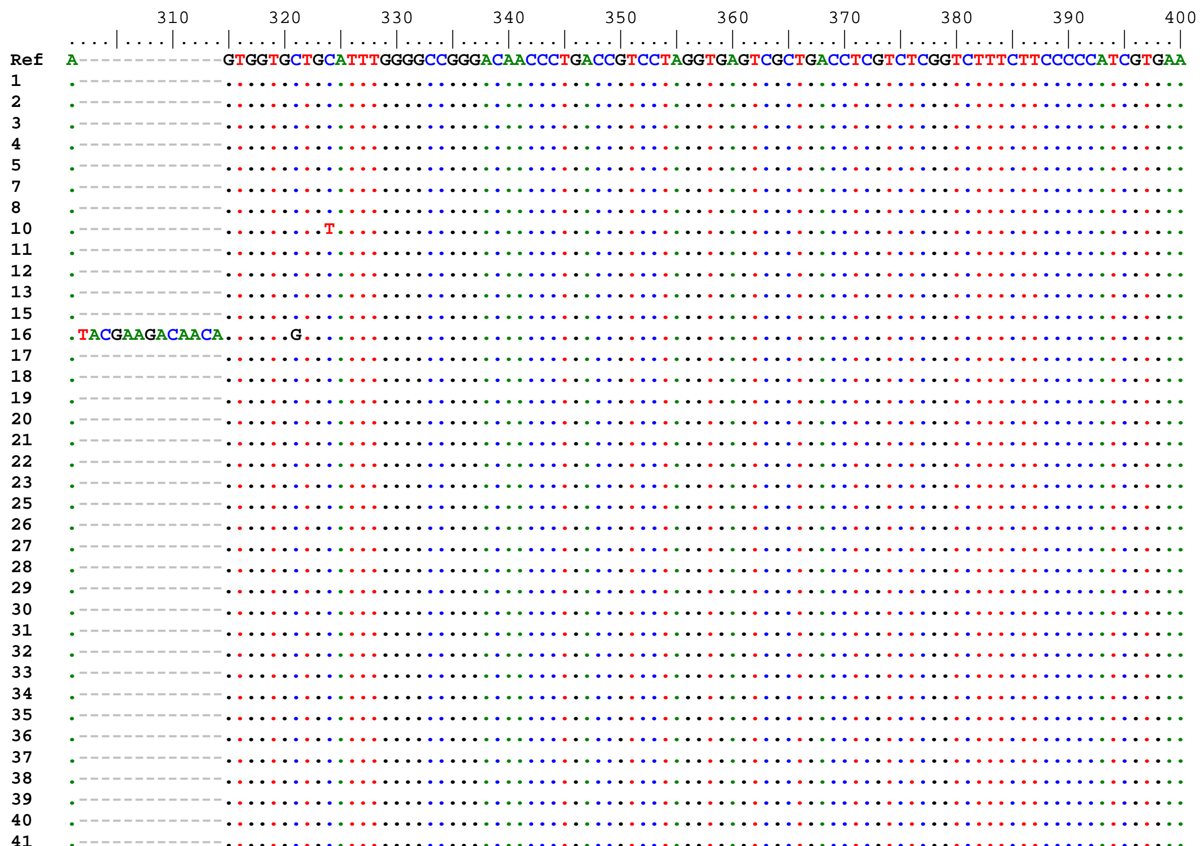
```
                 110       120       130       140       150       160       170       180       190       200
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Ref       GTTACTATTATGGCTGGTACCAGCAGAAGTCTCCTGGCAGTGCCCCTGTCACTGTGATCTATGACAACGACAAGAGACCCTCGGACATCCCTTCACGATT
1         ....................................................................................................
2         ....................................................................................................
3         ....................................................................................................
4         ....................................................................................................
5         ....................................................................................................
7         ....................................................................................................
8         ....................................................................................................
10        ....................................................................................................
11        ....................................................................................................
12        ....................................................................................................
13        ....................................................................................................
15        ....................................................................................................
16        ....................................................................................................
17        ....................................................................................................
18        ....................................................................................................
19        ....................................................................................................
20        ....................................................................................................
21        ....................................................................................................
22        ....................................................................................................
23        ...................................................G................................................
25        ....................................................................................................
26        ....................................................................................................
27        ....................................................................................................
28        ....................................................................................................
29        -----------------------------------------------------------------------------------.................
30        ....................................................................................................
31        ....................................................................................................
32        ....................................................................................................
33        ....................................................................................................
34        .....................................................C..............................................
35        ....................................................................................................
36        ....................................................................................................
37        ....................................................................................................
38        ....................................................................................................
39        ....................................................................................................
40        ....................................................................................................
41        ....................................................................................................
42        ....................................................................................................
43        ....................................................................................................
44        C......................................................C............................................
45        ....................................................................................................
46        ....................................................................................................
47        ....................................................................................................
48        ....................................................................................................
49        ....................................................................................................
50        ....................................................................................................
51        ....................................................................................................
52        ....................................................................................................
53        ....................................................................................................
54        ....................................................................................................
55        ....................................................................................................
56        ....................................................................................................
57        ....................................................................................................
58        ....................................................................................................
59        ....................................................................................................
60        ....................................................................................................
61        ....................................................................................................
62        ....................................................................................................
63        ....................................................................................................
```

Rearranged IgL VJ region of AID^{R} ψV^{-}E2A$^{-/-}$

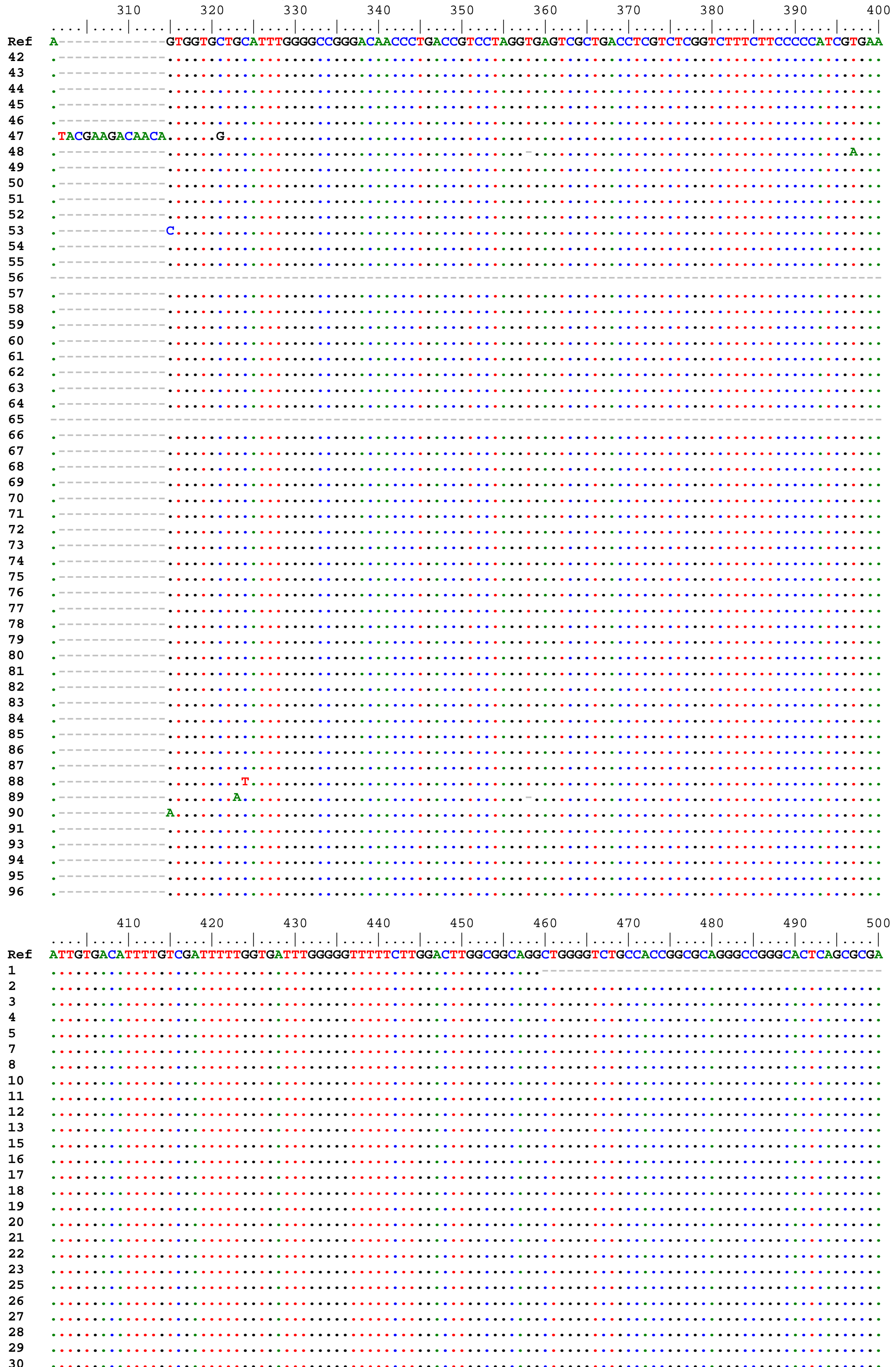
```
              210       220       230       240       250       260       270       280       290       300
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Ref      CTCCGGTTCCAAATCCGGCTCCACAGCCACATTAACCATCACTGGGGTCCGAGCCGATGACGAGGCTGTCTATTTCTGTGGGAGCTACGAAGA---CAAC
1        .............................................................................................---....
2        .............................................................................................---....
3        .............................................................................................---....
4        .............................................................................................---....
5        .............................................................................................---....
7        .............................................................................................---....
8        .............................................................................................---....
10       .............................................................................................---....
11       .............................................................................................---....
12       .............................................................................................---....
13       ....................................................................................G........---....
15       .............................................................................................---....
16       ...........................................................................G.................---....
17       .............................................................................................---....
18       .............................................................................................---....
19       .............................................................................................---....
20       .............................................................................................---....
21       .............................................................................................---....
22       .............................................................................................---....
23       .............................................................................................---....
25       .............................................................................................---....
26       .............................................................................................---....
27       .............................................................................................---....
28       .............................................................................................---....
29       .............................................................................................---....
30       .............................................................................................---....
31       .............................................................................................---....
32       .............................................................................................---....
33       .............................................................................................---....
34       .............................................................................................---....
35       .............................................................................................---....
36       .............................................................................................---....
37       .............................................................................................---....
38       .............................................................................................---....
39       .............................................................................................---....
40       .............................................................................................---....
41       .............................................................................................---....
42       .............................................................................................---....
43       .............................................................................................---....
44       .............................................................................................---....
45       .............................................................................................---....
46       .............................................................................................---....
47       ...........................................................................G.................---....
48       .............................................................................................---....
49       .............................................................................................---....
50       .............................................................................................---....
51       .............................................................................................---....
52       .............................................................................................---....
```

Rearranged IgL VJ region of AID^{R} ψV^{-}E2A$^{-/-}$

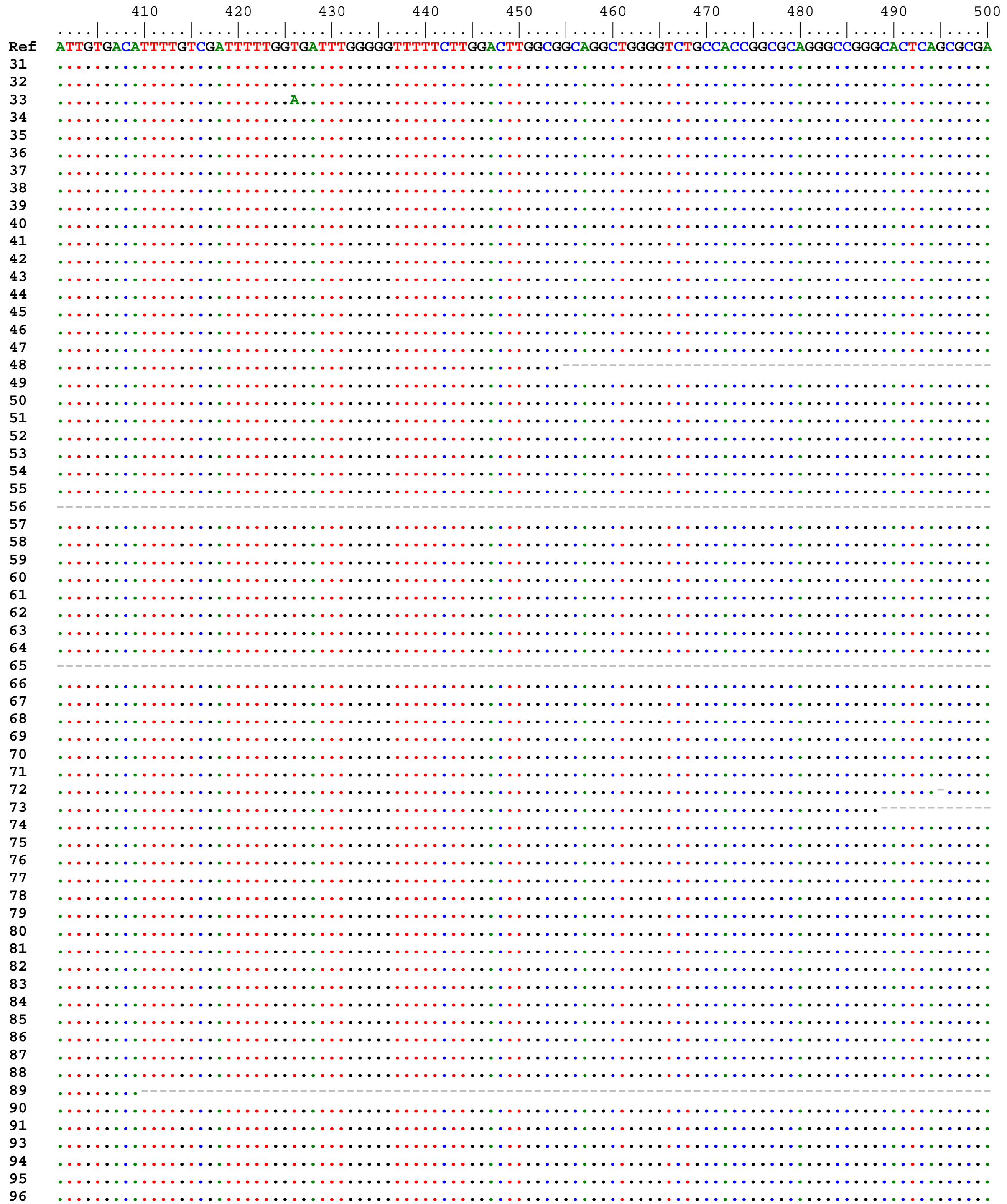




Rearranged IgL VJ region of AID$^{R}\psi$V^{-}E2A$^{-/-}$

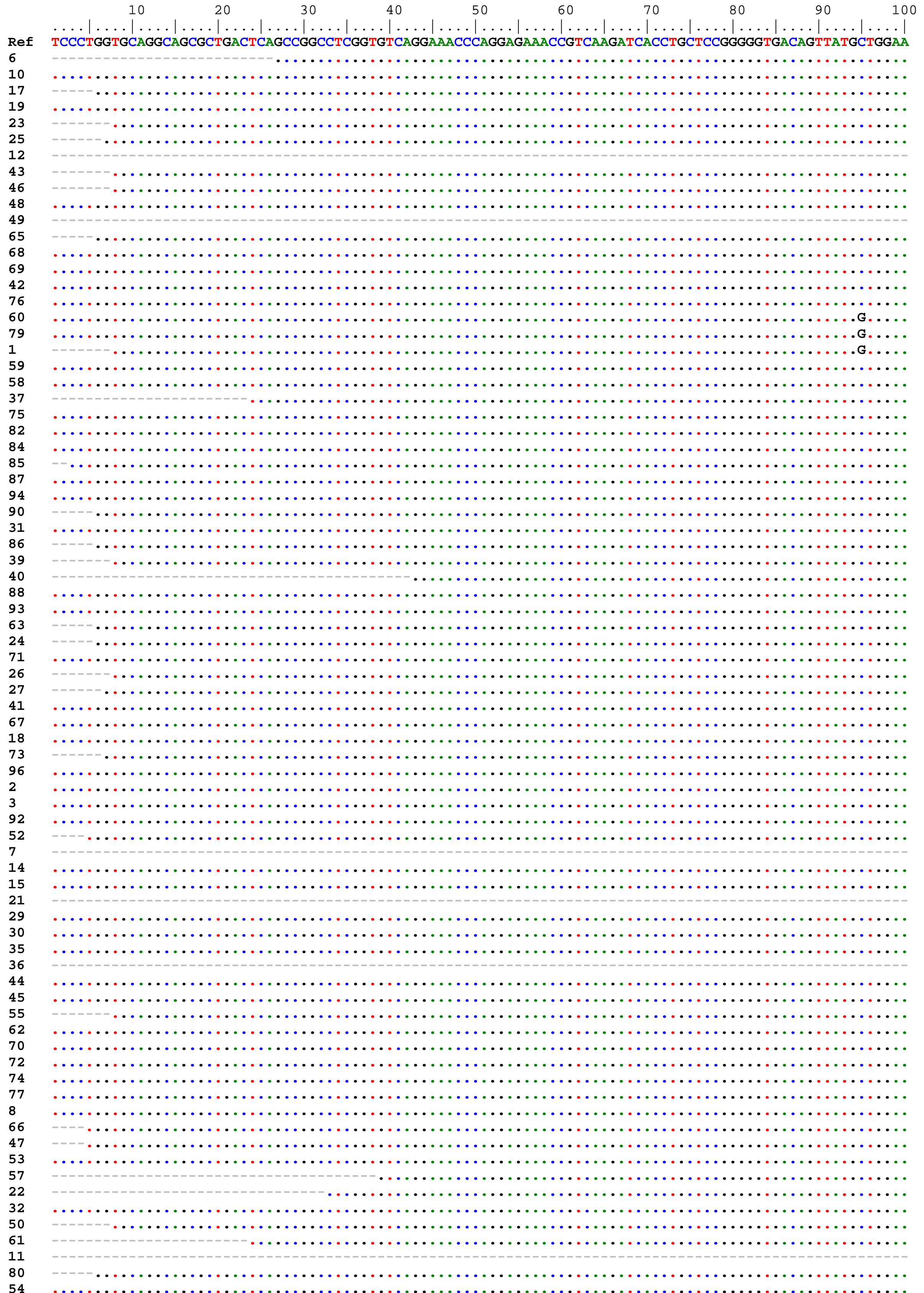


Rearranged IgL VJ region of AID$^R\psi$V$^-$E2A$^{-/-}$



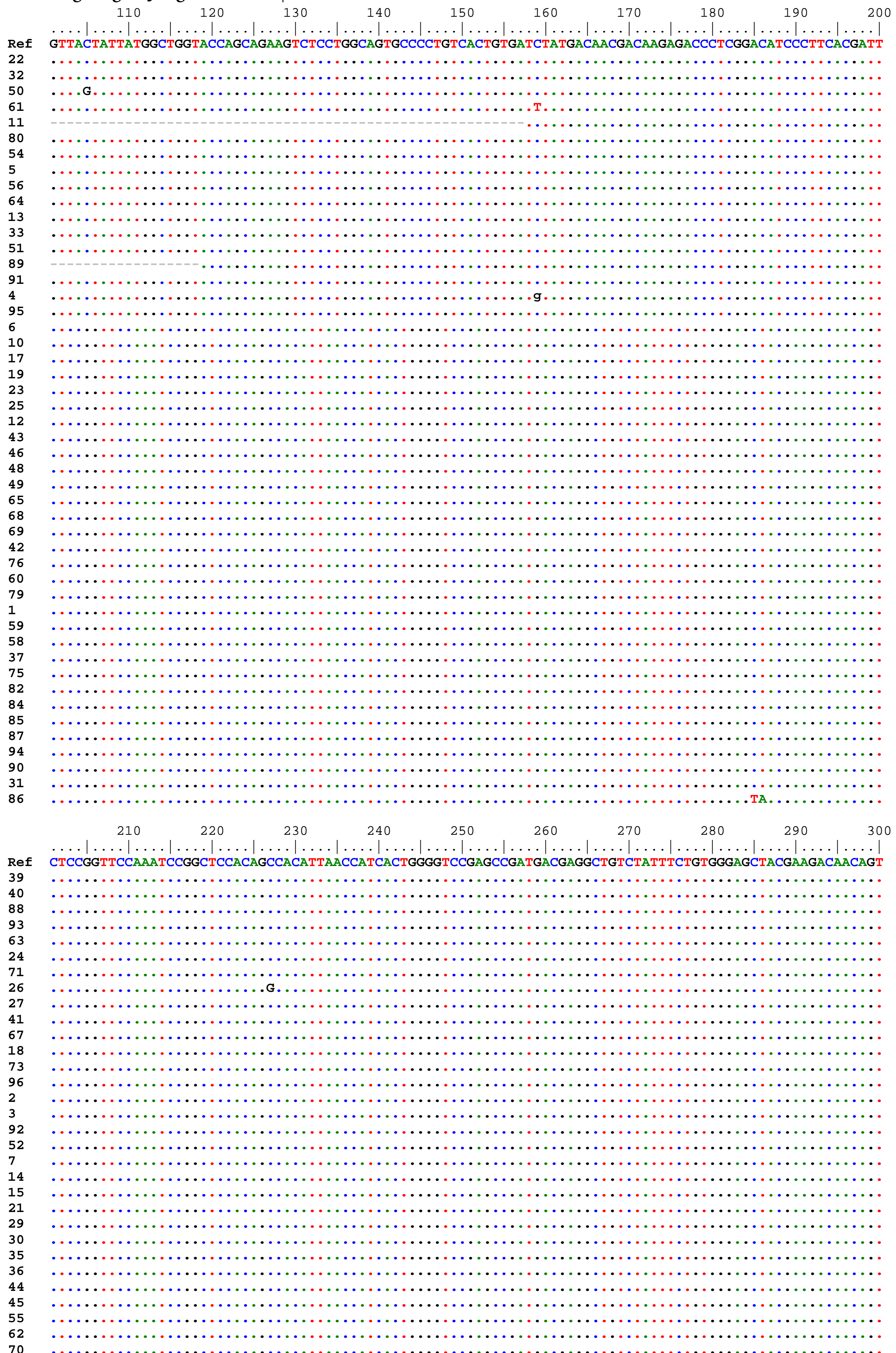
8.2.2 Sequencing of AID$^{R}\psi$V^{-}E2A^{RtE12}

Rearranged IgL VJ region of AID$^{R}\psi$V^{-}E2A^{RtE12}

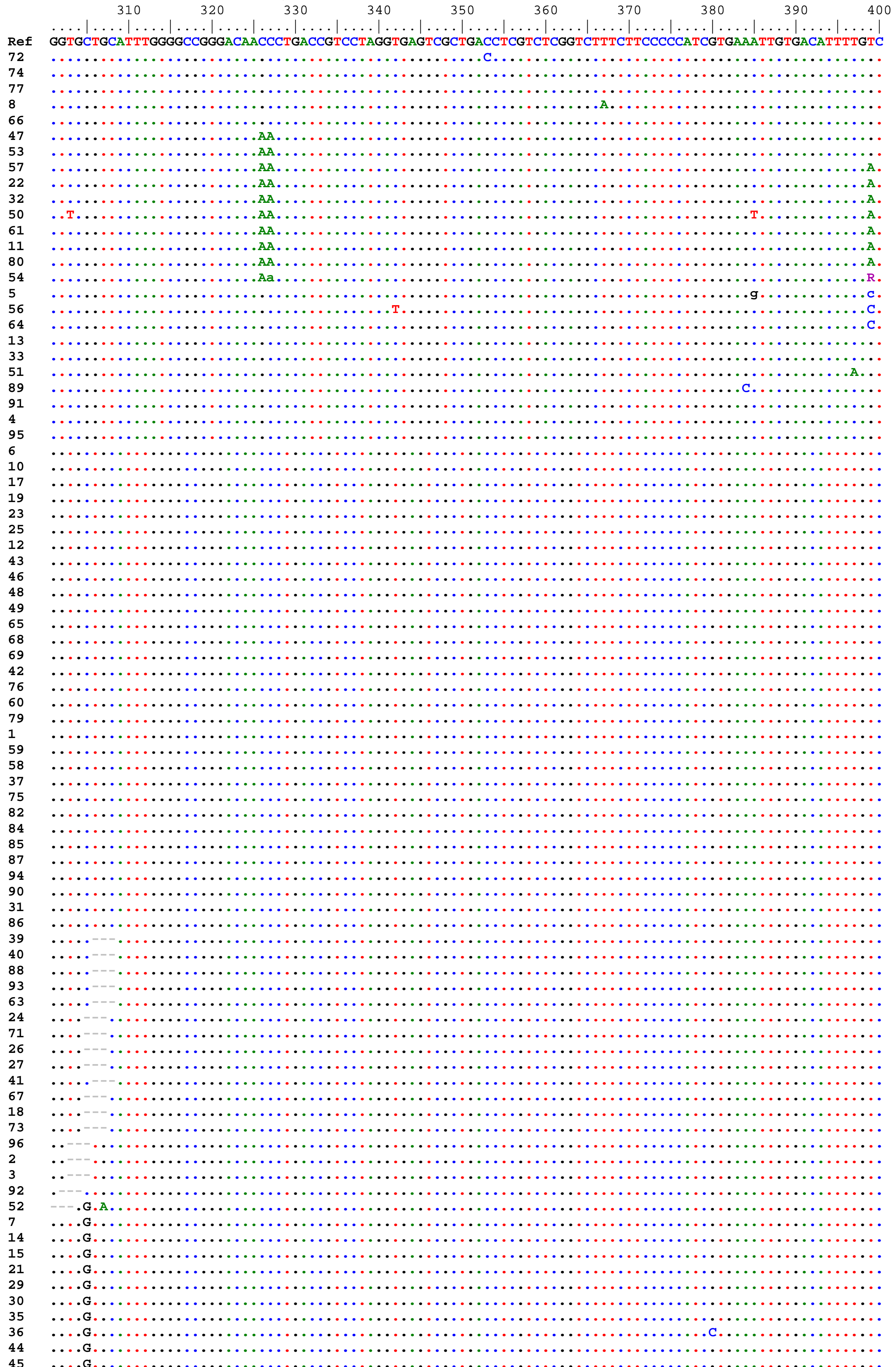


Rearranged IgL VJ region of AID^{R}ψV^{-}E2A^{RtE12}

Rearranged IgL VJ region of AID$^{R}\psi V^{-}$E2A^{RtE12}



Rearranged IgL VJ region of AID$^{R}\psi V^{-}$E2A^{RtE12}



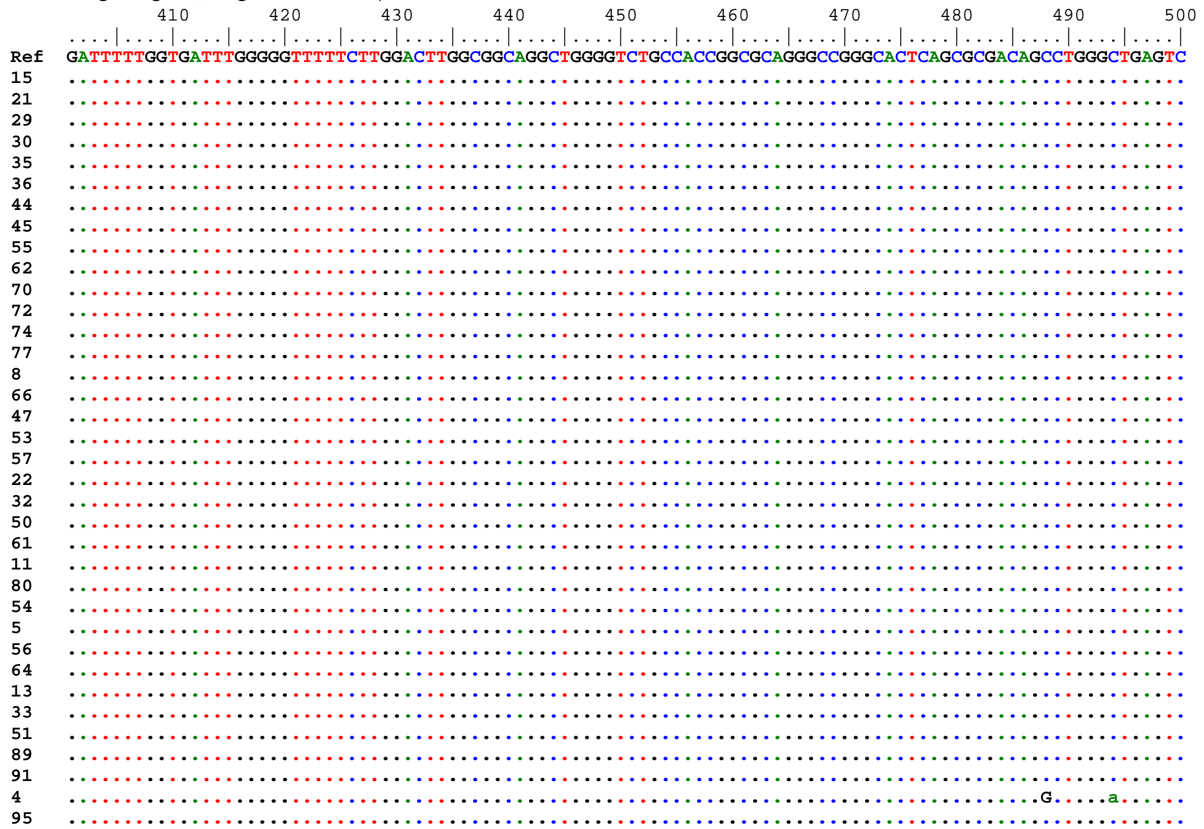
Rearranged IgL VJ region of AID^R ψV^- E2A^RtE12

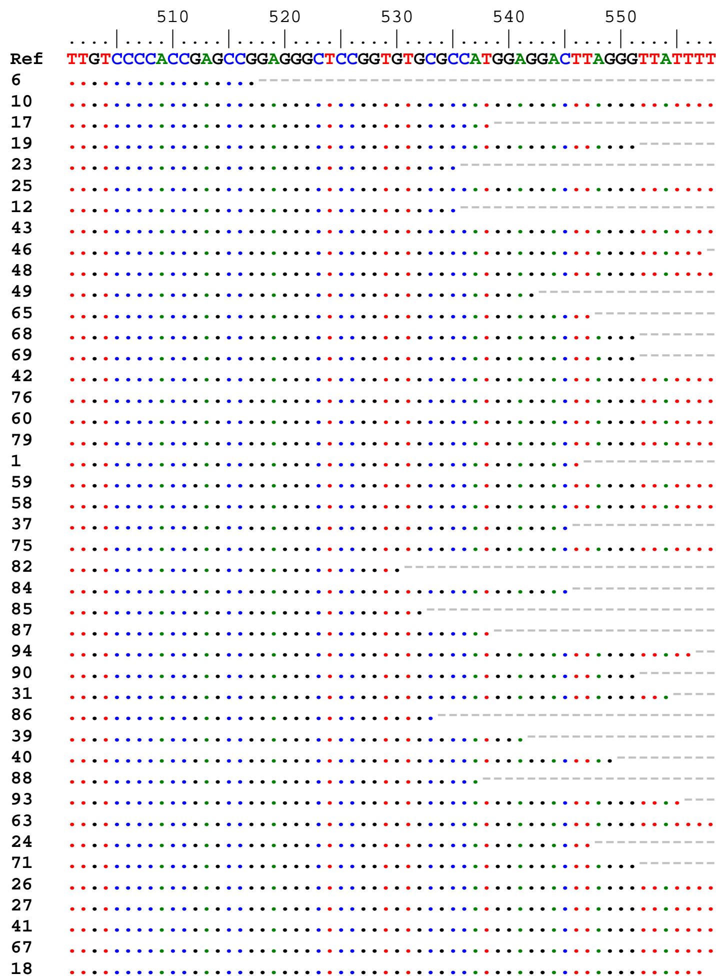
```
             310       320       330       340       350       360       370       380       390       400
      ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Ref   GGTGCTGCATTTGGGGCCGGGACAACCCTGACCGTCCTAGGTGAGTCGCTGACCTCGTCTCGGTCTTTCTTCCCCCATCGTGAAATTGTGACATTTTGTC
55    ....G...............................................................................................
62    ....G..........................................................................C....................
70    ....G...............................................................................................
72    ....G...............................................................................................
74    ....G...............................................................................................
77    ....G...............................................................................................
8     ....G...............................................................................................
66    ....G...............................................................................................
47    AC..................................................................................................
53    AC..................................................................................................
57    AC............................................................................................G.....
22    AC..................................................................................................
32    aC..................................................................................................
50    AC..................................................................................................
61    AC..................................................................................................
11    .c..................................................................................................
80    AC..................................................................................................
54    Rc..................................................................................................
5     ....................................................................................................
56    ....................................................................................................
64    ....................................................................................................
13    .c..................................................................................................
33    .c..................................................................................................
51    ....................................................................................................
89    ......A.............................................................................................
91    .......T............................................................................................
4     .t......W.......................................a...............................A.................A.
95    ....g...............................................................................................
```

Rearranged IgL VJ region of AID^R ψV^- E2A^RtE12

```
             410       420       430       440       450       460       470       480       490       500
      ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Ref   GATTTTTGGTGATTTGGGGGTTTTTCTTGGACTTGGCGGCAGGCTGGGGTCTGCCACCGGCGCAGGGCCGGGCACTCAGCGCGACAGCCTGGGCTGAGTC
6     .....................................................................K..............................
10    ....................................................................................................
17    ....................................................................................................
19    ....................................................................................................
23    ....................................................................................................
25    ....................................................................................................
12    ....................................................................................................
43    ....................................................................................................
46    ....................................................................................................
48    ....................................................................................................
49    ....................................................................................................
65    ....................................................................................................
68    ....................................................................................................
69    ....................................................................................................
42    ....................................................................................................
76    ....................................................................................................
60    ....................................................................................................
79    ....................................................................................................
1     ....................................................................................................
59    ....................................................................................................
58    ....................................................................................................
37    ....................................................................................................
75    ....................................................................................................
82    ....................................................................................................
84    ....................................................................................................
85    ....................................................................................................
87    ....................................................................................................
94    ....................................................................................................
90    ....................................................................................................
31    ....................................................................................................
86    ....................................................................................................
39    ....................................................................................................
40    ....................................................................................................
88    ....................................................................................................
93    ....................................................................................................
63    ....................................................................................................
24    ....................................................................................................
71    ....................................................................................................
26    ....................................................................................................
27    ....................................................................................................
41    ....................................................................................................
67    ....................................................................................................
18    ....................................................................................................
73    ....................................................................................................
96    ....................................................................................................
2     ....................................................................................................
3     ....................................................................................................
92    ....................................................................................................
52    ...................................T................................................................
7     ....................................................................................................
14    ....................................................................................................
```

Rearranged IgL VJ region of AID$^{R}\psi$V^{-}E2A^{RtE12}



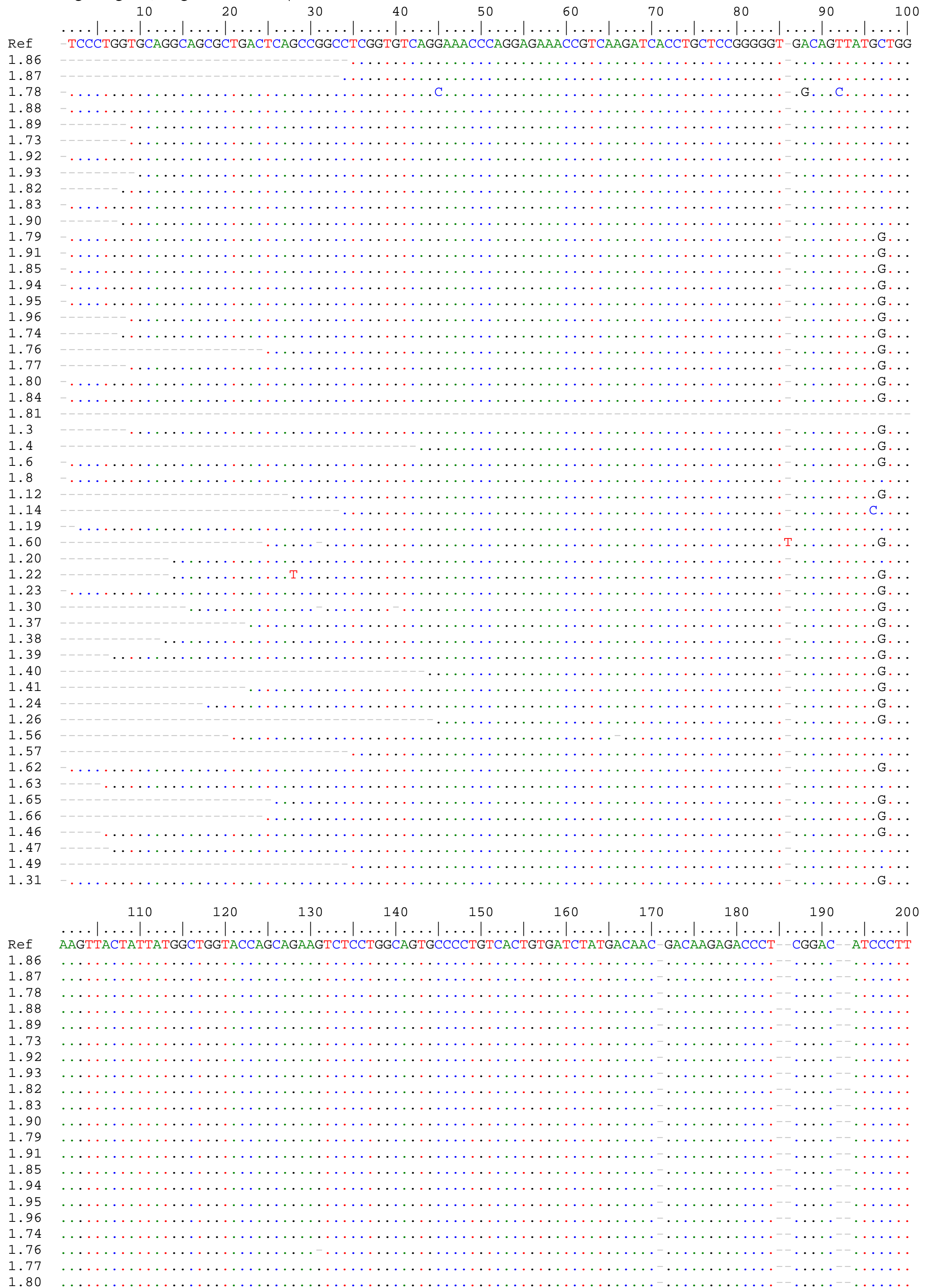


Rearranged IgL VJ region of AID$^{R}\psi$V^{-}E2A^{RtE12}

```
               510       520       530       540       550
      ....|....|....|....|....|....|....|....|....|....|....|...
Ref   TTGTCCCCACCGAGCCGGAGGGCTCCGGTGTGCGCCATGGAGGACTTAGGGTTATTTT
73    ..........................................................
96    ..........................................................
2     ..........................................................
3     .............................................------------
92    ..........................................................
52    .........................................................-
7     ......................................--------------------
14    .....................................................-----
15    .......................................................---
21    ......................................--------------------
29    ..........................................................
30    ..........................................................
35    ............................................--------------
36    .........................................-----------------
44    ..........................................................
45    ..........................................................
55    .......................................------------------
62    ..........................................................
70    ......................................................----
72    ..........................................----------------
74    ..........................................................
77    ..........................................................
8     ................................--------------------------
66    .........................................-----------------
47    ..........................................................
53    .........................................................-
57    ..........................................................
22    ............................................--------------
32    ..........................................................
50    ................................................----------
61    ..........................................................
11    .....................................A....................
80    ..........................................................
54    .....................................................-----
5     ...........................-------------------------------
56    ......................................--------------------
64    ..........................................................
13    ...................................................-------
33    ...................................................-------
51    ........................................................--
89    .............................................-------------
91    ...........................................---------------
4     .................................-------------------------
95    ....................................................------
```

8.2.3 Sequencing of AID[R]ψV⁻E2A[RtE4]

Rearranged IgL VJ region of AID[R]ψV⁻E2A[RtE47]Cl.1



Rearranged IgL VJ region of AID^{R}ψV^{-}E2A^{RtE47}Cl.1

```
              110       120       130       140       150       160       170       180       190       200
      ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Ref   AAGTTACTATTATGGCTGGTACCAGCAGAAGTCTCCTGGCAGTGCCCCTGTCACTGTGATCTATGACAAC-GACAAGAGACCCT--CGGAC--ATCCCTT
1.84  ......................................................................-.............--.....--.......
1.81  -------------------------------------------...........................-.............--.....--.......
1.3   ......................................................................-.............--.....--.......
1.4   ......................................................................-.............--.....--.......
1.6   ......................................................................-.............--.....--.......
1.8   ......................................................................-.............--.....--.......
1.12  ......................................................................-.............--.....--.......
1.14  ......................................................................-.............--.....--.......
1.19  ......................................................................-.............--.....--.......
1.60  ......................................................................-.............--.....--.......
1.20  ......................................................................-.............--.....--.......
1.22  ......................................................................-.............--.....--.......
1.23  ......................................................................-.............--.....--.......
1.30  ......................................................................-.............--.....--.......
1.37  ......................................................................-.............--.....--.......
1.38  ......................................................................-.............--.....--.......
1.39  ......................................................................-.............--.....--.......
1.40  ......................................................................-.............--.....--.......
1.41  ......................................................................-.............--.....--.......
1.24  ......................................................................-.............--.....--.......
1.26  ......................................................................-.............--.....--.......
1.56  ......................................................................-.............--.....--.......
1.57  ......................................................................-.............--.....--.......
1.62  ......................................................................-.............--.....--.......
1.63  ......................................................................-.............--.....--.......
1.65  ......................................................................-.............--.....--.......
1.66  ......................................................................-.............--.....--.......
1.46  ......................................................................-.............--.....--.......
1.47  ......................................................................-.............--.....--.......
1.49  ......................................................................-.............--.....--.......
1.31  ......................................................................-.............--.....--.......

              210       220       230       240       250       260       270       280       290       300
      ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Ref   --CACGATTCTCCGGTTCCAAATCCGGCTCCACAGCCACATT-AACCAT-CACT-GGGGTCCGAGCCGAT-GACGAGGCTGTC-TATTTCT-GTGGGAGC
1.86  --........................................-......-....-...............-............-.....T.-........
1.87  --........................................-......-....-...............-............-.....T.-........
1.78  --........................................-......-....-...............-............-.......-.......-
1.88  --........................................-......-....-...............-............-.......-........
1.89  --........................................-......-....-...............-............-.......-........
1.73  --........................................-......-....-...............-............-.......-........
1.92  --........................................-......-....-...............-............-.......-........
1.93  --........................................-......-....-...............-............-.......-........
1.82  --........................................-......-....-...............-............-.......-........
1.83  --........................................-......-....-...............-............-.......-........
1.90  --........................................-......-....-...............-............-.......-........
1.79  --........................................-......-....-...............-............-.......-........
1.91  --........................................-......-....-...............-............-.......-........
1.85  --........................................-......-....-...............-............-.......-........
1.94  --........................................-......-....-...............-............-.......-........
1.95  --........................................-......-....-...............-............-.......-........
1.96  --........................................-......-....-...............-............-.......-........
1.74  --........................................-......-....-...............-............-.......-........
1.76  --........................................-......-....-...............-............-.......-........
1.77  --........................................-......-....-...............-............-.......-........
1.80  --........................................-......-....-...............-............-.......-........
1.84  --........................................-......-....-...............-............-.......-........
1.81  --........................................-......-....-...............-............-.......-........
1.3   --........................................-......-....-...............-............-.......-........
1.4   --........................................-......-....-...............-............-.......-........
1.6   --........................................-......-....-...............-............-.......-........
1.8   --........................................-......-....-...............-............-.......-........
1.12  --........................................-......-....-...............-............-.......-........
1.14  --........................................-......-....-...............-............-.......-........
1.19  --........................................-......-....-...............-............-.......-........
1.60  --........................................-......-....-...............-............-.......-........
1.20  --........................................-......-....-...............-............-.......-........
1.22  --........................................-......-....-...............-............-.......-........
1.23  --........................................-......-....-...............-............-.......-........
1.30  --........................................-......-....-...............-............-.......-........
1.37  --........................................-......-....-...............-............-.......-........
1.38  --........................................-......-....-...............-............-.......-........
1.39  --........................................-......-....-...............-............-.......-........
1.40  --........................................-......-....-...............-............-.......-........
1.41  --........................................-......-....-...............-............-.......-........
1.24  --........................................-......-....-...............-............-.......-........
1.26  --........................................-......-....-...............-............-.......-........
1.56  --........................................-......-....-...............-............-.......-........
1.57  --........................................-......-....-...............-............-.......-........
1.62  --........................................-......-....-...............-............-.......-........
1.63  --........................................-......-....-...............-............-.......-........
1.65  --........................................-......-....-...............-............-.......-........
1.66  --........................................-......-....-...............-............-.......-........
1.46  --........................................-......-....-...............-............-.......-........
```

Rearranged IgL VJ region of AID$^{R}\psi$V^{-}E2A^{RtE47}Cl.1

```
                  210       220       230       240       250       260       270       280       290       300
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Ref      --CACGATTCTCCGGTTCCAAATCCGGCTCCACAGCCACATT-AACCAT-CACT-GGGGTCCGAGCCGAT-GACGAGGCTGTC-TATTTCT-GTGGGAGC
1.47     --........................................-......-....-...............-............-.......-........
1.49     --........................................-......-....-...............-............-.......-........
1.31     --........................................-......-....-...............-............-.......-........

                  310       320       330       340       350       360       370       380       390       400
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Ref      TACGAAGACAACAGTGGTGCTGCATTTGGGGCCGGGACAACCCTGACCGTCCTAGGTGAGTCGCTGACCTCGTCTCGGTCTTTCTTCCCCCATCGTGAAA
1.86     ......................G.............................................................................
1.87     ......................G.............................................................................
1.78     --T.C...............................................................................................
1.88     ....................................................................................................
1.89     ....................................................................................................
1.73     ....................................................................................................
1.92     ....................................................................................................
1.93     ....................................................................................................
1.82     ....................................................................................................
1.83     ....................................................................................................
1.90     ....................................................................................................
1.79     ...................A................................................................................
1.91     C.G................A................................................................................
1.85     C.G................A................................................................................
1.94     C.G................A................................................................................
1.95     C.G................A................................................................................
1.96     C.G................A................................................................................
1.74     C.G................A................................................................................
1.76     C.G................A................................................................................
1.77     C.G................A................................................................................
1.80     C.G................A................................................................................
1.84     C.G................A................................................................................
1.81     C.G................A................................................................................
1.3      C.G................A................................................................................
1.4      C.G................A................................................................................
1.6      C.G................A................................................................................
1.8      ....................................................................................................
1.12     C.G................A................................................................................
1.14     ....................................................................................................
1.19     ....................................................................................................
1.60     C.G................A................................................................................
1.20     ....................................................................................................
1.22     C.G................A................................................................................
1.23     C.G................A................................................................................
1.30     C.G................A................................................................................
1.37     C.G................A................................................................................
1.38     C.G................A................................................................................
1.39     C.G................A................................................................................
1.40     C.G................A................................................................................
1.41     ...........A.......A................................................................................
1.24     C.G................A................................................................................
1.26     C.G................A................................................................................
1.56     ....................................................................................................
1.57     ....................................................................................................
1.62     C.G................A................................................................................
1.63     ....................................................................................................
1.65     ...........A.......A................................................................................
1.66     C.G................A................................................................................
1.46     C.G................A................................................................................
1.47     .............A......................................................................................
1.49     ....................................................................................................
1.31     C.G................A................................................................................
```

Rearranged IgL VJ region of AID$^{R}\psi$V^{-}E2A^{RtE47}Cl.1

```
                 410       420       430       440       450       460       470       480       490       500
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Ref      TTGTGACATTTT-GTCGATTTTT-GGTGATTTGGGGG-TTTTTCTTGGACTTGGCGG-CAGGCT-GGGGTCTGCCACCGGCGCAGGGCCGGGCACTCAGC
1.86     ............-..........-.............-.............-------------------------------------------------
1.87     ............-..........-.............-...................-.G.C-.-..........-...............-........
1.78     ............-..........-.............-...................-......-...................................
1.88     ............-..........-.............-...................-......-...................................
1.89     ............-..........-.............-...................-......-...................................
1.73     ............-..........-.............-...................-......-...................................
1.92     ............-..........-.............-...................-......-...................................
1.93     ............-..........-.............-...................-......-...................................
1.82     ............-..........-.............-...................-......-...................................
1.83     ............-..........-.............-...................-......-...................................
1.90     ............-..........-.............-...................-......-...................................
1.79     ............-..........-.............-...................-......-...................................
1.91     ............-..........-.............-...................-......-...................................
1.85     ............-..........-.............-...................-......-...................................
1.94     ............-..........-.............-...................-......-..............-G.................--
1.95     ............-..........-.............-...................-......-...................................
1.96     ............-..........-.............-...................-......-...................................
1.74     ............-..........-.............-...................-......-...................................
1.76     ............-..........-.............-...................-......-...................................
1.77     ............-..........-.............-...................-......-...................................
```

Rearranged IgL VJ region of AID$^{R}\psi$V^{-}E2A^{RtE47}Cl.1

```
              410       420       430       440       450       460       470       480       490       500
      ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Ref   TTGTGACATTTT-GTCGATTTTT-GGTGATTTGGGGG-TTTTTCTTGGACTTGGCGG-CAGGCT-GGGGTCTGCCACCGGCGCAGGGCCGGGCACTCAGC
1.80  ............-..........-.............-...................-......-...................................
1.84  ............-..........-.............-...................-......-...................................
1.81  ............-..........-.............-...................-......-....-..............................
1.3   ............-..........-........-....-...................-......-...................................
1.4   ............-..........-.............-...................-......-..........................-........
1.6   ............-..........-.............-...................-......-...................................
1.8   ............-..........-.............-...................-......-...................................
1.12  ............-..........-.............-...................-......-...................................
1.14  ............-..........-.............-...................-......-...............-...................
1.19  ............-..........-.............-...................-......-...................................
1.60  ............-..........-.............-...................-......-...................................
1.20  ............-..........-.............-...................-......-...................................
1.22  ............-..........-.............T...................-......-...................................
1.23  ............-..........-.............-...................-......-...................................
1.30  ............-..........T.............T...................-......-.................................A.
1.37  ............-..........-.............-...................-......-...................................
1.38  ............-..........-.............-...................-......-.....................C.............
1.39  ............-..........-.............-...................-......-...................................
1.40  ............-..........-.............-...................-......-...............-...................
1.41  ............-..........T.............-...................-......-...................................
1.24  ............-..........-.......G.....-...................-G..T.G...........C....-C.....G....A.....--
1.26  ............-..........-.............-...................-......-....C.....C..........-.............
1.56  ............-..........-.............-...............T.GC.G......-..................................-
1.57  ............-..........--.G...........-...................-......-....................-.............
1.62  ............-..........-.............-...................-......-...................................
1.63  ............-..........T.............-...................-......-...................................
1.65  ............-..........-.............-...................-......-...................................
1.66  ............-..........-.............G...................-......-....................-............-
1.46  ............-..........-.............-...................-......-...................................
1.47  ............-..........-.............-...................-......-...................................
1.49  ............-..........-.............-...................-......-.....................-.............
1.31  ............-..........-.............-...................-......G...-...............................

               510       520
      ....|....|....|....|.
Ref   GCGACAGCCTGGGCTGAGTCT
1.86  ---------------------
1.87  ........-............
1.78  .....................
1.88  .....................
1.89  .....................
1.73  .....................
1.92  .....................
1.93  .....................
1.82  .....................
1.83  .....................
1.90  .....................
1.79  .....-...............
1.91  .....................
1.85  .....................
1.94  .....-...........--..
1.95  .....................
1.96  .....................
1.74  .............-.......
1.76  .....................
1.77  .....................
1.80  .....................
1.84  .....................
1.81  .....................
1.3   .....................
1.4   .....................
1.6   .....................
1.8   .....................
1.12  .....................
1.14  ........-........
1.19  .....................
1.60  .....................
1.20  .....................
1.22  .....................
1.23  .....................
1.30  ....................
1.37  .....................
1.38  .....................
1.39  .....................
1.40  .....................
1.41  .....................
1.24  .....-......
1.26  .................
1.56  .............
1.57  .............
1.62  .....................
1.63  .....................
1.65  .....................
1.66  .....-...........-...
```

Rearranged IgL VJ region of AID^{R}ψV^{-}E2A^{RtE47}Cl.1

```
              510       520
      ....|....|....|....|.
Ref   GCGACAGCCTGGGCTGAGTCT
1.46  .....................
1.47  .....................
1.49  .........
1.31  .....................
```

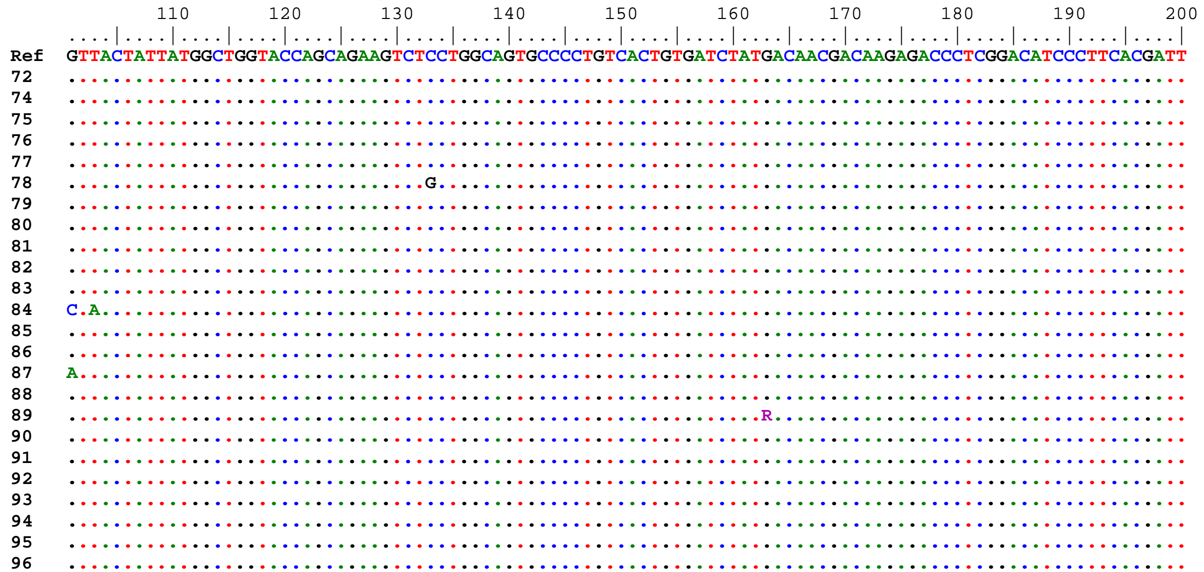
Rearranged IgL VJ region of AID^{R}ψV^{-}E2A^{RtE47}Cl.51

```
              10        20        30        40        50        60        70        80        90       100
      ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Ref   TCCCTGGTGCAGGCAGCGCTGACTCAGCCGGCCTCGGTGTCAGGAAACCCAGGAGAAACCGTCAAGATCACCTGCTCCGGGGGTGACAGTTATGCTGGAA
1     ....................................................................................................
2     ..............................................................................................G.....
3     ..............................................................................................G.....
4     ....................................................................................................
5     ....................................................................................................
6     ..............................................................................................G.....
7     ..............................................................................................G.....
9     ....................................................................................................
11    ....................................................................................................
12    ..............................................................................................G.....
13    ..............................................................................................G.....
14    ..............................................................................................G.....
15    ..............................................................................................G.....
16    -----...............................................................................................
17    ..............................................................................................G.....
18    ....................................................................................................
19    ....................................................................................................
20    ....................................................................................................
21    ....................................................................................................
22    ..............................................................................................G.....
23    ..............................................................................................G.....
24    ....................................................................................................
25    ....................................................................................................
26    ....................................................................................................
27    ....................................................................................................
28    ....................................................................................................
29    ..............................................................................................G.....
30    ....................................................................................................
31    ....................................................................................................
32    ..............................................................................................G.....
33    ..............................................................................................G.....
34    ....................................................................................................
35    ..............................T............C......G...............................A.AGG..-----..----
36    ....................................................................................................
37    ....................................................................................................
38    ..............................................................................................G.....
39    ..............................................................................................G.....
40    ..............................................................................................G.....
41    ....................................................................................................
42    ..............................................................................................G.....
43    ....................................................................................................
44    ....................................................................................................
45    ....................................................................................................
46    ..............................................................................................G.....
47    ..............................................................................................G.....
48    ..............................................................................................G.....
50    ....................................................................................................
51    ....................................................................................................
52    ..............................................................................................G.....
53    ....................................................................................................
55    ..............................................................................................G.....
56    ....................................................................................................
57    ....................................................................................................
58    ....................................................................................................
60    ....................................................................................................
61    ....................................................................................................
62    ....................................................................................................
63    ..............................................................................................G.....
65    ..............................................................................................T.....
66    .........G..........................................................................................
67    ..............................................................................................G.....
68    ....................................................................................................
69    ..............................................................................................G.....
70    ....................................................................................................
71    ....................................................................................................
72    ....................................................................................................
74    ..............................................................................................G.....
75    ....................................................................................................
76    ..............................................................................................G.....
77    ....................................................................................................
78    ..............................................................................................G.....
79    ..............................................................................................G.....
80    ..............................................................................................G.....
81    ..............................................................................................G.....
```

Rearranged IgL VJ region of AID^R^ψV^-^E2A^RtE47^Cl.51

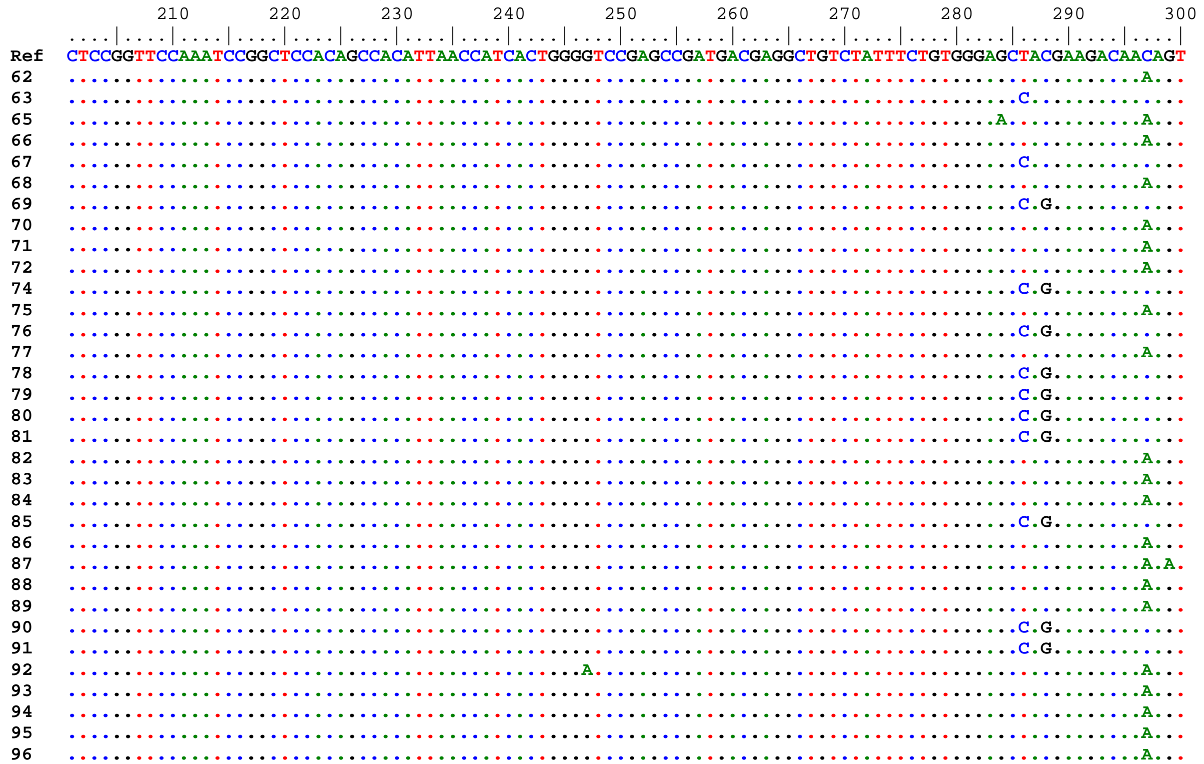
```
                10        20        30        40        50        60        70        80        90       100
       ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Ref    TCCCTGGTGCAGGCAGCGCTGACTCAGCCGGCCTCGGTGTCAGGAAACCCAGGAGAAACCGTCAAGATCACCTGCTCCGGGGGTGACAGTTATGCTGGAA
82     ....................................................................................................
83     ....................................................................................................
84     ....................................................................................................
85     ..............................................................................................G.....
86     ....................................................................................................
87     ....................................................................................................
88     ..............................................................................................T.....
89     ....................................................................................................
90     ..............................................................................................G.....
91     ..............................................................................................G.....
92     ....................................................................................................
93     ....................................................................................................
94     ....................................................................................................
95     ....................................................................................................
96     ....................................................................................................

               110       120       130       140       150       160       170       180       190       200
       ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Ref    GTTACTATTATGGCTGGTACCAGCAGAAGTCTCCTGGCAGTGCCCCTGTCACTGTGATCTATGACAACGACAAGAGACCCTCGGACATCCCTTCACGATT
1      ....................................................................................................
2      ....................................................................................................
3      ....................................................................................................
4      ....................................................................................................
5      ....................................................................................................
6      ....................................................................................................
7      ....................................................................................................
9      ....................................................................................................
11     ....................................................................................................
12     ....................................................................................................
13     ....................................................................................................
14     ....................................................................................................
15     ....................................................................................................
16     ....................................................................................................
17     ....................................................................................................
18     ...............................................................................G....................
19     ...................T................................................................................
20     ....................................................................................................
21     ....................................................................................................
22     ....................................................................................................
23     ....................................................................................................
24     ....................................................................................................
25     ....................................................................................................
26     ....................................................................................................
27     ....................................................................................................
28     ....................................................................................................
29     ....................................................................................................
30     ....................................................................................................
31     C...................................................................................................
32     ....................................................................................................
33     ....................................................................................................
34     ....................................................................................................
35     ---...--............................................................................................
36     ....................................................................................................
37     ....................................................................................................
38     ....................................................................................................
39     ....................................................................................................
40     ....................................................................................................
41     ....................................................................................................
42     ....................................................................................................
43     ....................................................................................................
44     ....................................................................................................
45     ....................................................................................................
46     ....................................................................................................
47     ....................................................................................................
48     ....................................................................................................
50     ....................................................................................................
51     ....................................................................................................
52     ....................................................................................................
53     ....................................................................................................
55     ....................................................................................................
56     ....................................................................................................
57     ....................................................................................................
58     ....................................................................................................
60     ....................................................................................................
61     ....................................................................................................
62     ....................................................................................................
63     ....................................................................................................
65     ....................................................................................................
66     ....................................................................................................
67     ....................................................................................................
68     ....................................................................................................
69     ....................................................................................................
70     ....................................................................................................
71     ....................................................................................................
```

Rearranged IgL VJ region of AID^{R}ψV^{-}E2A^{RtE47}Cl.51



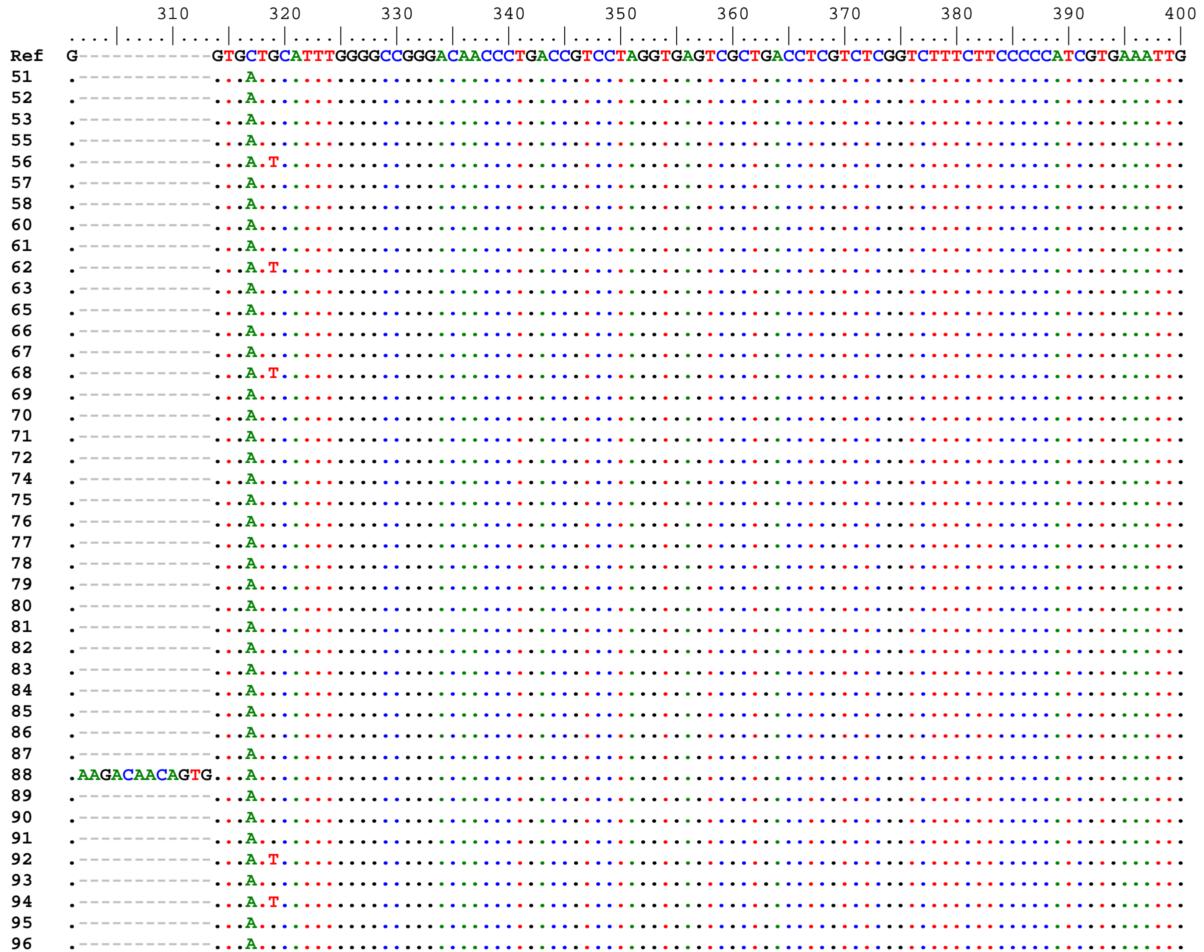
```
              210       220       230       240       250       260       270       280       290       300
      ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Ref   CTCCGGTTCCAAATCCGGCTCCACAGCCACATTAACCATCACTGGGGTCCGAGCCGATGACGAGGCTGTCTATTTCTGTGGGAGCTACGAAGACAACAGT
1     ..............................................................................................A...
2     .....................................................................................C..............
3     .....................................................................................C.G............
4     ..............................................................................................A...
5     ..............................................................................................A...
6     .....................................................................................C.G............
7     .....................................................................................C.G............
9     ..............................................................................................A...
11    ..............................................................................................A...
12    .....................................................................................C.G............
13    .....................................................................................C.G............
14    ....................................................................................................
15    .....................................................................................C.G............
16    ..............................................................................................A...
17    .....................................................................................C.G............
18    ..............................................................................................A...
19    ..............................................................................................A...
20    ..............................................................................................A...
21    ..............................................................................................A...
22    .....................................................................................C.G............
23    .....................................................................................C.G............
24    ..............................................................................................A...
25    ..............................................................................................A...
26    ..............................................................................................A...
27    ..............................................................................................A...
28    ..............................................................................................A...
29    ..............................................................................................A...
30    ...........................................................................G....................A...
31    ..............................................................................................A...
32    .....................................................................................C.G............
33    .....................................................................................C.G............
34    ..............................................................................................A...
35    ..............................................................................................A...
36    ..............................................................................................A...
37    ..............................................................................................A...
38    .....................................................................................C.G............
39    .....................................................................................C.G............
40    .....................................................................................C.G............
41    ..............................................................................................A...
42    .....................................................................................C.G........A...
43    ..............................................................................................A...
44    ..............................................................................................A...
45    ..............................................................................................A...
46    .....................................................................................C.G............
47    ....................................................................................................
48    .....................................................................................C.G............
50    ..............................................................................................A...
51    ..............................................................................................A...
52    .....................................................................................C.G............
53    ..............................................................................................A...
55    .....................................................................................C.G............
56    ..............................................................................................A...
57    ..............................................................................................A...
58    ..............................................................................................A...
60    ..............................................................................................A...
61    ..............................................................................................A...
```

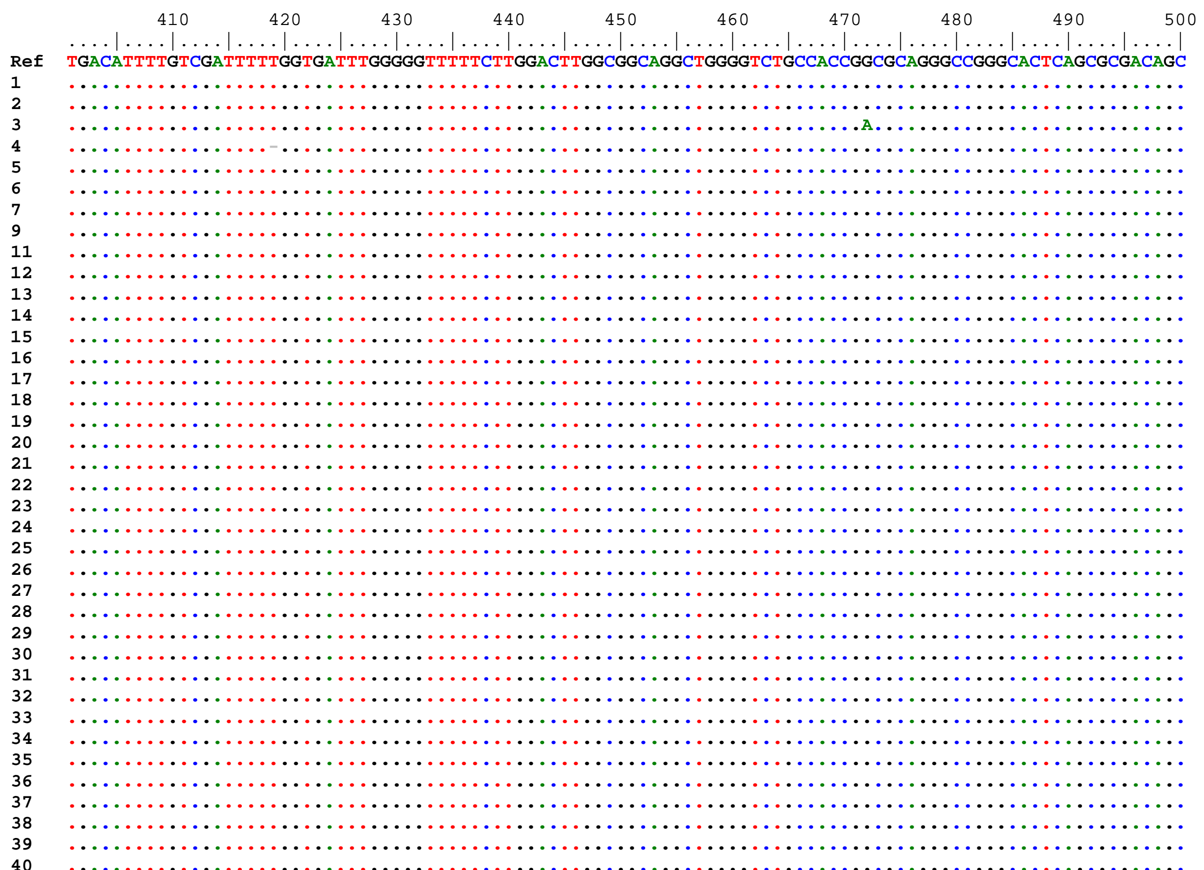
Rearranged IgL VJ region of AID$^{R}\psi$V^{-}E2A^{RtE47}Cl.51



```
              310       320       330       340       350       360       370       380       390       400
     ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Ref  G------------GTGCTGCATTTGGGGCCGGGACAACCCTGACCGTCCTAGGTGAGTCGCTGACCTCGTCTCGGTCTTTCTTCCCCCATCGTGAAATTG
1    .------------...A...................................................................................
2    .------------...A...................................................................................
3    .------------...A...................................................................................
4    .------------...A.T.................................................................................
5    .------------...A...................................................................................
6    .------------...A...................................................................................
7    .------------...A...................................................................................
9    .------------...A...................................................................................
11   .------------...A...................................................................................
12   .------------...A...................................................................................
13   .------------...A...................................................................................
14   .------------...A...................................................................................
15   .------------...A...................................................................................
16   .------------...A...................................................................................
17   .------------...A...................................................................................
18   .------------...A...................................................................................
19   .------------...A...................................................................................
20   .------------...A...................................................................................
21   .------------...A...................................................................................
22   .------------...A...................................................................................
23   .------------...A...................................................................................
24   .------------...A...................................................................................
25   .------------...A.C.................................................................................
26   .------------...A...................................................................................
27   .------------...A...................................................................................
28   .------------...A...................................................................................
29   .------------...A...................................................................................
30   .------------...A...................................................................................
31   .------------...A...................................................................................
32   .------------...A...................................................................................
33   .------------...A...................................................................................
34   .------------...A...................................................................................
35   .------------...A.T.................................................................................
36   .------------...A.T.................................................................................
37   .------------...A...................................................................................
38   .------------...A...................................................................................
39   .------------...A...................................................................................
40   .------------...A...................................................................................
41   .------------...A...................................................................................
42   .------------...A...................................................................................
43   .------------...A...................................................................................
44   .------------...A...................................................................................
45   .------------...A.T.................................................................................
46   .------------...A...................................................................................
47   .------------...A...................................................................................
48   .------------...A...................................................................................
50   .------------...A...................................................................................
```

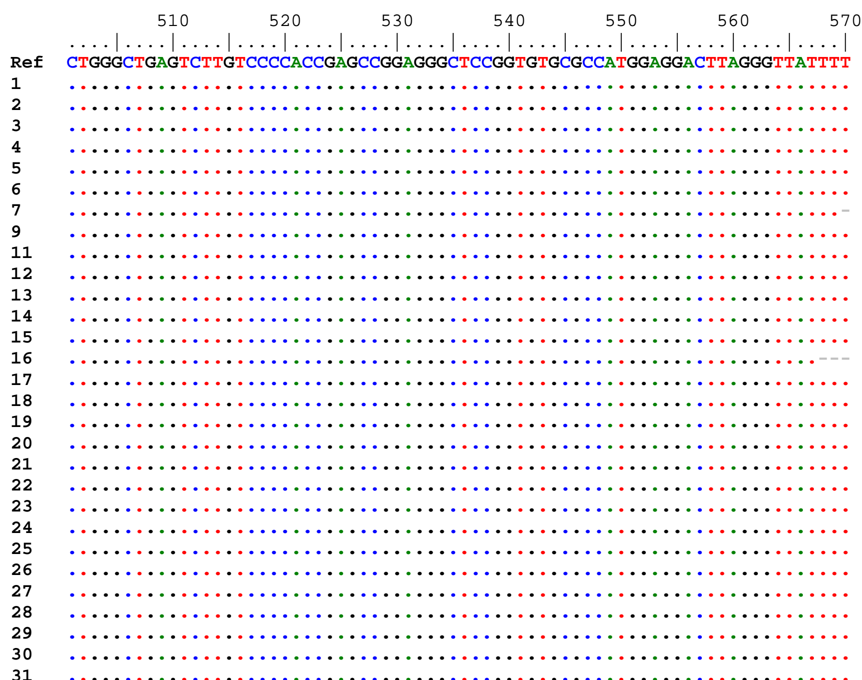
Rearranged IgL VJ region of AID$^{R}\psi V^{-}$E2A^{RtE47}Cl.51



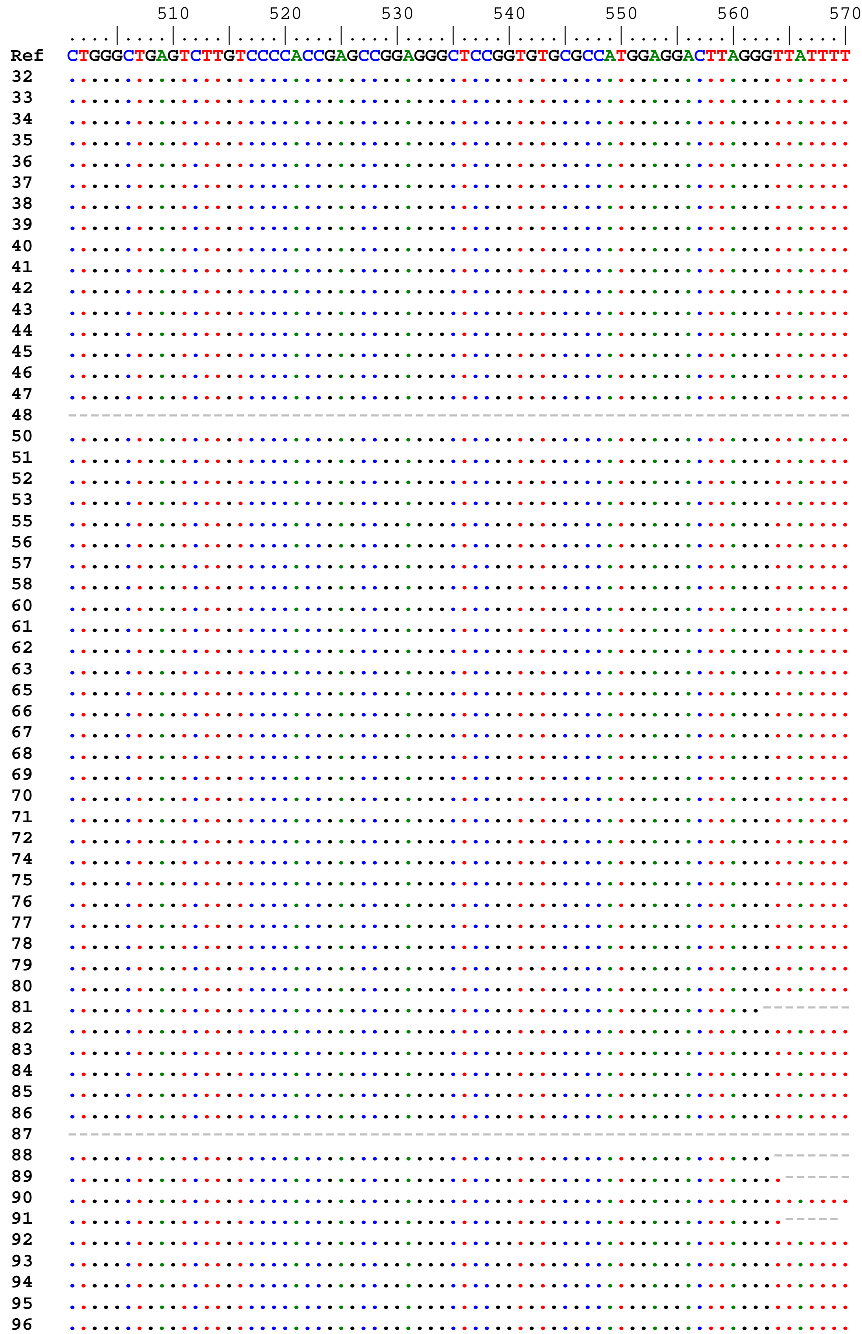


Rearranged IgL VJ region of AID^R ψV^- E2A^RtE47 Cl.51

```
              410       420       430       440       450       460       470       480       490       500
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Ref      TGACATTTTGTCGATTTTTGGTGATTTGGGGGTTTTTCTTGGACTTGGCGGCAGGCTGGGGTCTGCCACCGGCGCAGGGCCGGGCACTCAGCGCGACAGC
41       ....................................................................................................
42       ....................................................................................................
43       ....................................................................................................
44       ....................................................................................................
45       ....................................................................................................
46       ....................................................................................................
47       ....................................................................................................
48       .................................-------------------------------------------------------------------
50       ....................................................................................................
51       ....................................................................................................
52       ....................................................................................................
53       ....................................................................................................
55       ....................................................................................................
56       ....................................................................................................
57       ....................................................................................................
58       ....................................................................................................
60       ....................................................................................................
61       ....................................................................................................
62       ....................................................................................................
63       ....................................................................................................
65       ....................................................................................................
66       .................................................................................................T..
67       ....................................................................................................
68       ....................................................................................................
69       ....................................................................................................
70       ....................................................................................................
71       ....................................................................................................
72       ....................................................................................................
74       ....................................................................................................
75       ....................................................................................................
76       ....................................................................................................
77       ................................................A...................................................
78       ....................................................................................................
79       ....................................................................................................
80       ....................................................................................................
81       ....................................................................................................
82       ....................................................................................................
83       ................................................A...................................................
84       ....................................................................................................
85       ....................................................................................................
86       ....................................................................................................
87       ................................--------------------------------------------------------------------
88       ....................................................................................................
89       ....................................................................................................
90       ....................................................................................................
91       ....................................................................................................
92       ....................................................................................................
93       ....................................................................................................
94       ....................................................................................................
95       ....................................................................................................
96       ....................................................................................................
```



Rearranged IgL VJ region of AID[R]ψV[-]E2A[RtE47]Cl.51



8.3 Sequence Information

8.3.1 Complete sequence of 'W'

```
LOCUS       FJ482234            9826 bp    DNA     linear   VRT 07-JAN-2009
DEFINITION  Gallus gallus immunoglobulin light chain gene, complete sequence.
ACCESSION   FJ482234
VERSION     FJ482234.1  GI:218942844
KEYWORDS    .
SOURCE      Gallus gallus (chicken)
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Dinosauria; Saurischia; Theropoda; Coelurosauria;
            Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 9826)
  AUTHORS   Blagodatski,A., Batrak,V., Schmidl,S., Schoetz,U., Caldwell,R.B.,
            Arakawa,H. and Buerstedde,J.-M.
  TITLE     A cis-acting diversification activator both necessary and
            sufficient for AID mediated hypermutation
  JOURNAL   PLoS Genet. 5 (1), e1000332 (2009)
  REMARK    Publication Status: Online-Only
REFERENCE   2  (bases 1 to 9826)
  AUTHORS   Caldwell,R.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-NOV-2008) Institute of Molecular Radiobiology,
            Helmholtz Center Munich - German Research Center for Environmental
            Health (GmbH), Ingolstaedter Landstrasse 1, Neuherberg, Bayern
            85764, Germany
FEATURES             Location/Qualifiers
     source          1..9826
                     /organism="Gallus gallus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9031"
                     /chromosome="15"
                     /map="15:8165000-8176700"
                     /cell_line="DT40"
                     /rearranged
     gene            1..9826
                     /gene="immunoglobulin light chain"
     misc_feature    20..501
                     /gene="immunoglobulin light chain"
                     /note="contains V region"
     misc_feature    502..1024
                     /gene="immunoglobulin light chain"
                     /note="contains J segment"
     misc_feature    2211..2848
                     /gene="immunoglobulin light chain"
                     /note="contains C region"
     misc_feature    2849..9826
                     /gene="immunoglobulin light chain"
                     /note="diversification activator cis-region (DIVAC)"
ORIGIN
        1 gctagcacta gcttccgcca tggcctgggc tcctctcctc ctggcggtgc tcgcccacac
       61 ctcaggtact cgttgcgccc ggtcggggac tgtgggcacg gggctctgtc ccactgctgc
      121 gcgggcaggg ctgtgcgtgc ggggccgtca ctgattgccg ttttctcccc tctctcctct
      181 ccctctccag gttccctggt gcaggcagcg ctgactcagc cggcctcggt gtcagcaaat
      241 ccaggagaaa ccgtcaagat cacctgctcc gggggtggca gctatgctgg aagttactat
      301 tatggctggt accagcagaa gtctcctggc agtgcccctg tcactgtgat ctatgacaac
      361 gacaagagac cctcggacat cccttcacga ttttccggtt ccaaatccgg ctccacagcc
      421 acattaacca tcactgggtt ccgagccgat gacgaggctg tctatttctg tgggagctac
      481 gaagacaaca gtggtgctgc atttggggcc gggacaaccc tgaccgtcct aggtgagtcg
      541 ctgacctcgt ctcggtcttt cttcccccat cgtgaaattg tgacattttg tcgatttttg
      601 gtgatttggg ggtttttctt ggacttggcg gcaggctggg gtctgccacc ggcgcagggc
      661 cgggcactca gcgcggcagc ctgggctgag tcttgtcccc accgagccgg agggctccgg
      721 tgtgcgccat ggaggactta gggttatttt gtcaatggaa agttcttaaa atttgaccag
      781 aaaatgtgcc cgaggtctgt ctctgccaca cgatttcaga aattgtgtct aggtcgatga
      841 gaagacagtt tttgtctttg tcaggaaatt agttgtgagt tgttagtcct tccctcttag
      901 tcctaaggac taagaccttt gtccccggtc tggtctctca ctggggactc ttggctccag
      961 tgccatgggg agcccaagtg tcactgacac ggtgtccttg gggtgaaat tcagtttttc
     1021 agctgtccag gtaagtgctg tcaaaaaatg gaggttattc tgctgaaaaa gctgagagga
     1081 atattttgtc attttttcg gaaatatata tatatatata tatataaata tataaattat
     1141 ttatatattt tatatatata tatatataaa atatatgtat ttctctcttt ctttttatat
     1201 atatatataa atatatatat atttctccct ctttctttat atatatttat agggagagaa
     1261 tctatatttt tggccaattt ggccaatttc tctctctctc tctttatata tatatatata
```

1321 tatatatgta atttatatat atatatataa atttatattt atttatttat ttatttatat
1381 agattatttt tatatacata tatatatata aatttggcca atttaggaag aaattggcca
1441 aaaaaaaacc caatctgaac caaattactt agggtataag tccaaacaaa aaaagtgcag
1501 ttcaggaatt ccctcgctgg aaagtttta atgggggcca gttctttccc ccatttctg
1561 gtgtagatat gagtgaagat atagatatgt atgttagata tagatttggc tatagtttgt
1621 ttgatgaaca gtctgaggaa gagtaaatct tgccttcctc atggcaattt ttaattctac
1681 ttcttctgag acaagtctga acatgccttg aggtagaaag agcccgaaaa ttgagtgctt
1741 tttctttgaa attttaaata aaattaaaaa atagccctta cagcaggaaa gaaagggggg
1801 tgatggcagc taatcggagt tccaaagctg aaccacgttc acccaaacac gtggttgaaa
1861 ttttgtgtgt gtatatatat aaatatatat atgtatacaa aagaaaatgc ctatatataa
1921 aaaaattgtt atgtatatta tatagattta ttatagatat gtacataaga tataacacgt
1981 atacatatat atgtagtctg ctgtatgctt tgtataaatt tttatagata tgatgtatat
2041 tatatgcatg ggttatatat ctataatata ttttatctgt agaaggaact acagtataaa
2101 atgcatatat gcgtatatat acacatattt atatatatgt gtatatatat gtatacgcac
2161 acgtataaca tccatgaaaa tagatatgga tgggtggatg tgtttgtttt acagaggtgc
2221 atgtgtgtct gtacatacac agccatacat acgcgtgtgg ccgctctgcc tctctcttgc
2281 aggccagccc aaggtggccc ccaccatcac cctcttccca ccgtcaaagg aggagctgaa
2341 cgaagccacc aaggccaccc tggtgtgcct gataaacgac ttctacccca gcccagtgac
2401 tgtggattgg gtgatcgatg gctccacccg ctctggcgag accacagcac cacagcggca
2461 gagcaacagc cagtatatgg ccagcagcta cctgtcactg tctgccagcg actggtcaag
2521 ccacgagacc tacacctgca gggtcacaca caacggcacc tctatcacga agaccctgaa
2581 gaggtccgag tgctaatagt cccactgggg atgcaatgtg aggacagtgg ttcctcaccc
2641 tccctgtccc tctgggccgc tgctggtggc agcagccctc acttcccact cagatgtccc
2701 ccaccgtgcc cccatcaccc acctctgcct gtcgactcct cttgccctca tctctccagg
2761 tgtcacatta ataaacacga cactgaacta gtgctgactc tgcatccatg tctctgtgtc
2821 cttttgcgtg ctgtctgcat ctcacacagt ggggtcggcc ccagtatggg gaagggctgg
2881 ggggcacata cacacatatt ggtaatgttg ggggcggggg ggggagggcg ggggggtcaa
2941 cagatcagca ctggagacac tggtgtatac cctggcaaca ccaacatcta aggcagggtg
3001 ctttggggca attttggggc agtttaaggt ctgtgctggc actgagcaca tggctgtggc
3061 cgtgctgtcc tcgtctccca cccactacgg tctgtgcgcc aggtccctag cagagatttg
3121 ctttatgctg ggaacagggg gagttctggg tctgttccct tgcattcaga caccctggtg
3181 ccccctgggt gggatgtcag tgtgaatact cctttgtgcc ctgtgcctgc agcagcctga
3241 ccctccacac accacacgcc ttgtgtgcac cccacccctg tcactatccc tctccccgct
3301 ccccagggag attttgcagt ggcccctgta gggcagcttt tagcacagcc cccagcagca
3361 agcaagcaga aagcactgct gtgcacagct tgtcagctgt gtgtgtttgc tgaggaggat
3421 ctgtcttttg ctgaggccat cagtcttgtc ctgctcaacc tccatcgatg ctgcccacct
3481 caacacatct acccatctat tccatctaca ccaacatctc cattcatccc acccacccaa
3541 acatgtccat ccatcacaac acctccatcc aacccgcaca ctccagcacc tccaatcatt
3601 ccatctacac caccatgctg atctgctaca gccactccaa cgcaaccgtc cattccatct
3661 acaccaatgt ccatccatcc cagccactcc agcacctcca gccatcccac ccaccctatg
3721 tctccatcca gccactggtg gggtgcagga catggggcca gctctactgt caggactggg
3781 gtttttgcat ggccccatac cacttctgca gaagagacgc actgaaagtt tggctgacca
3841 ttttctccgc ggtagagctg tggaaattct gtaatttagg gtcttttatc cagtttggag
3901 atgggctggg atctcccagc tccatggcag gcattcatga cactgggttt agtatctgat
3961 gggtgggatg tggctgaact tcattttctt tccccagtga caaagttttt gcagttgaat
4021 atgaattcct gctttctgct ctatgagttg tttttttccc aggacgtaca cagggaatca
4081 gcagtcttca ttctccctct gccatgtgta gactctctgc cacacaggac tgtgctgtcc
4141 tcatgcccct gcgcccaaat tgctgccctc tgcccatgcc tgccaagctg agcccccctg
4201 caggctgcca tgctggattg acatgagccc tgagattggt acagaaatgg tgattttggg
4261 gttttctctg cactcaggaa gctgaaggct caatgctcag tgatggattt accaaactgt
4321 gccctgaggc agctgctcat gctggataaa gtcactggag cacagggaac caggcgctgg
4381 gcagggattt ctcatgggcc ccacttggaa agctgcaggc tgcaaacctg gctctgcctt
4441 cacgcctcac cctcatgagg acaacctcac taattattga ttaaaagatt ttgctaaacc
4501 atctccagaa gcaacaaccc actgaggagc atgtgctgaa ttatacatca cagctccacg
4561 gccctgccct catagcaggg ctgcatggca cccacagtgg cactcagcgg gaccacaggg
4621 ctgagacagc cgggtctggt ggtggggaca cagctgagca taggatgagc cccccgggca
4681 gtgctgggct ttgctaatga gcagaagtat ggatagaaag caaccccagg gctccggtcc
4741 cagctgcagc tcttgctctg tcgtgtcctt tggtgaaact ttaaacagtc gccttttttt
4801 ttctctttct tttctggctt gccattaatt tcaaaccgag agagacctaa tttagtaaat
4861 gagatgcttc aggaaggctt taattggctg cagatggagg caggcagtgc tatcgtgggg
4921 cctggatcgc acaggggggct gcatatcctc actagcagaa tacactcagg ctgggtccct
4981 cccacattca tgccccagac cagagggaat atgctctgtt ccccacacat ctctcccaat
5041 cttgcagctg ttgagcccca acatcccacc agcacacggg gctcagcacg cctggcgacg
5101 tggcatcagc agagcaggcc gcatggtaca gctccatcag cacagctggg gccacacaaa
5161 gagctgggtt actgtgggca gcaggctgaa acccgaaaac aagagctggg ggctcagaat
5221 agccccggga gcaggcaggg cctgggggag agggcaagca ccaggcccag ggccacacag
5281 cccttccagg aaggcacagc gctgtcaggg tgcagcacgc tcagccccac catgcagctg
5341 tgcggccggg gcatccccaa gctaaattta cttctcagtc tccaatcaga aactgaagct
5401 gaggggccca cgccggccaa aaaaaggaaa cgaaacagtc tccagaaagc actgacgtgt
5461 gaagcagagc gagcgccgcg caaaccggcc gccatgtcac acacctcagg ttggggcttt
5521 gccagactga gctttgctgc tgctcggggt gggtgcccac ggcctgggca catgggatgg
5581 ggtacacacg tacacacact tgcacaccca caccccaaca cttcaggtga tgctggtgca
5641 gatgggtgcc ccccaggctg acccccccac gcatgggcct ggccccacac tgctccatcc
5701 gtgtctctgt ccccatgtgc caccccctgcc cgctcccacc acgcgtcaac ccaaatcctg
5761 agttaatccc acgactcctg cctgcttcca gcgtccatgg cagactggag atgcccaaaa
5821 tgcagagcag gtttccctga atctgagaga tgaaatggag ttatgggtgt tcccctgcgg
5881 cggagcccca gctgtaggaa gctcagagcc atcacacagc aattaaagag gaattaaatt
5941 aaatcaataa atgttttagg cgggctcagc tgccagcacc acctgaccga aacagcccgc

```
     6001 ttgcaaagag gagagcattt gcatggctgt ggcaaaacag caaccgcctg ttgtgcagct
     6061 gggatggtgt tatctggaaa tgtacgcagc ccaggagggg taaacagctc caaactgaga
     6121 ccccgagctt gtccacaggt tgtaaacagg ctgacataaa cacctttgtg ccgtggaaaa
     6181 atatttatca cctcaaatat agcaggttaa taaaataaaa ctcccaacgg agctacacac
     6241 ctgctttgga agggaagcag acacttgttt tctgcttgat gttggctgta ggaagccatg
     6301 tttcccgatg caggagggcc acaaagcact gacaacacaa tgtgagctga gcttcgcccc
     6361 tgtttaagcc cccacctcag ggcttgtggc ctcggagcag gcaggacgca ggggtggcac
     6421 cgggctgggt gacatgggct ggtcctgggg tgtctcactg tgctctttgg ggaggggttg
     6481 gagccctggg gcaatcacag cacacacagg aggtgggggg atgcagccag cagctgccct
     6541 gcactaagaa aaccccatcc gtgggctttc agatggcctt cccatctctc tgcagcctct
     6601 gcatgggctg agcacaaggt ttaagtgttt ctgccatgtt tttgggcatg tttggagggg
     6661 cagcatgggc ccgggcatac gggtaccgcc acgtgccgcc agccccacag ctgagcctgc
     6721 actctcccag atgtgctgac cgcagccacg ggggcaacag tttctcttgc taaaaattgt
     6781 agccgggaag aaaacacgtg gcaacttcgg ccaaacagca gctggaggac aggaatagcc
     6841 gtggccacgg cacgctctgc ttcctcggca caaacattcc agtatgtggc accacgagcg
     6901 ccgctgcccg gcacagcagc aagcagagcc aggagcagga aatgctgatt tgggccccat
     6961 tttggccatg gctgagagaa gaggcttcca gggagctggt cagcttggtc cccaagctgt
     7021 ggcttgggga aatgatgggg aggggattgc cactgcccac cctgcagagc aggctctggt
     7081 cccatctcac tgcagggcac cagggcgttt gcactgcagc aattcacaga aacattgaaa
     7141 tggctcctgc cttgttcaac atcttcatca gtgacctgga tgaggggaca gcatccacca
     7201 tcagcgggtt cactgatcat atgaagtcgg gaggagtggc tgacgcacca caaggctgtg
     7261 ctgccattca acaggacgtg gacagactgg agagctggac agggaggaac ccaatgaggt
     7321 tcaacaatgg caagtgtagg atctacacct gggaaggaat aacagcatgc atcagttcag
     7381 gttaggggct gagctgctgc agatgagctc tgagagaagg acctgagcat cctgctggac
     7441 agcaggctgg ctgtgagcca ccggtgtgcc ctggtggcca agaaggccag tggtatcctg
     7501 gggagcaccg caatgagagt gggcagcagg gcgagggagg tgaggctgca tttggagcac
     7561 cgtgcccagt tctgggctcc tcagttcaag gcagacaggg aactgctgga gagagcccag
     7621 cagaggggct gcaatgatga tgagggtcct ggagcatcgc ctgtatgagg aaaggctgag
     7681 ggacctggga ttgttcagct tggagaagag aagactacag ggcaggagcc aagtggatag
     7741 ggccgggctc ttttcagcag tgcccagtga caggacaagg ggcagcaggc acaaagtgga
     7801 acataagaag ttccatctga acatgaggaa aaactgcctc gctttgaggg tgaccgagca
     7861 ctggaagaag ctgcccagag aggtggtgga gtctcctctg gagatattca gagcctggca
     7921 ggacactttt tgctgagtaa cctactgtag ggaacctgac gcagcagagg ggtcggactg
     7981 gaggatctcc ggaggtctct ttcaacccct acagttccat gaaatacctc aaacactgcc
     8041 aagcgcagtg ctaaggcaag ggtaacattt gtaaactgaa acagggtggg tttaagttag
     8101 atgtaaagaa gaaactcttc actcagaggg tggcgaggcc ctggcacagg ctgcccatgg
     8161 aggctgcggg tgccccatcc ctggcagtgc ccaaggcaag agcccagcag tgaccacagc
     8221 cccacaagga cgagcgtggc cctcgtatct cagctcaccc tgccccagct caactcccac
     8281 ctccggcaca gcgcgggcac acagccgggc cctgtgctta tggagccctt ggggcaggtc
     8341 agcactcaca ccctccaaac acagccgtgg ctcccaaccg gaggcagctg gatctcggca
     8401 gccataacca agcagggcca tgcgggggtg acaccggggt cccccacccc ctgtggggca
     8461 gcgtatgggc tgggcccctg ctccagtctg cagcgtgtgc atgggaacca ttgccagaca
     8521 ccgtctggac cacccgcagc cctaagctgc ctcacagcag ggattgctcc gtcacaccgt
     8581 gaccccgtgc ccttattcca tcacttatgg ggctgggagt gcctggacct tgggcacatt
     8641 aacgaggatt tcccgctctg ccctcgcttt gctccgagcc gtggggctgt gtagtgcaga
     8701 cacagctgca gcctaaaatt agtacctggg aaaggccccc atgctgcacc gcgcagggct
     8761 gagatgtgcc acgtccccat ggccggagct ggggaaggca acgtggccct gtgcgtgtgc
     8821 acgctgagca caaggacacg tgctgggcca ggatttgtct ccccggggct cacgctatgt
     8881 gtcacccggt gctgcgccat cccctcccgc agcccccagc tcccccacgg ccgcacgccg
     8941 cctgcatccc tgcaacggca ccgcacagag acacggagcc aggggccgca cacggggcca
     9001 ggagctcacc tttattgcag ccctgacagc cccacggccc agcccgcacc ggggctgcca
     9061 catcctcacc cggccgacgg cccccagctg ctccttacca tttcttcccc ccatcacccc
     9121 ataaaccaga agccgcctca ccgctacgcg gagcgggcag cagggaaccc gggccctaag
     9181 ggggagacga gagggggccg agcaggggca ggaggagcag cagggcgagg gggcagcggg
     9241 ggcacccaca gctggccgtg gcatcccggg aggagaagac cttgcggctg cggagcggtt
     9301 gtggcggacg gaaattgttg gtcatcttca ggggcgcagc gcccgaggcc gggaagtgca
     9361 cggtgctgac aaacgcctgc agctgcgggg agagcaccgc gggcgccgca gccgtgaggc
     9421 gtagggcgaa gcggggcaca cgcgtggctg ctgctgtctt tccccctacc ggcagcacac
     9481 ggctctgcac acaccgcgct tcgtgccgcc tcgcagccga cgctgcagga agcccagccg
     9541 agcgcttaca gagcggccgg gaaatgcatc tgctgaggtg cccgggcaat gcagaacttc
     9601 atccatcccc acatccattc accagtcccc tcccaaaccc ccaagcccat ccggcgaccc
     9661 acccaccctc ctcttggtgc ccctctcaag ctctccatcc ccacattcct acagatgtcc
     9721 cctttacttt gcctgcaagg tgcaagaaaa cgcacaggac cgggggtgct cacagcacgg
     9781 ctttggccag acgggccctt ccatcccagc tagtgctagt ggatcc
//
```

8.3.2 Transcription factor binding sites of *'W'*

The binding sites were determined using the program MatInspector by Genomatix (www.genomatix.de; Quandt et al., 1995; Cartharius et al., 2005).

Inspecting sequence *'W'* (1 - 9826):
Solution parameters:

Family matches: yes

MatInspector library: Matrix Family Library Version 8.0 (November 2008)

Selected groups (core/matrix sim)
- Vertebrates (0.75/Optimized)

Selected tissues:
- Immune System

Search Results (733 matches):

Transcription factors			Opt.	Position	Str.	Core sim.	Matrix sim.	Sequence (red: ci-value > 60 capitals: core sequence)
Family	Detailed Family Information	Matrix		from - to				
V$ETSF	Human and murine ETS1 factors	V$SPI1_PU1.02	0.96	4 - 24	(-)	1.000	0.962	gccatggcGGAAgctagtgct
V$MAZF	Myc associated zinc fingers	V$MAZ.01	0.90	32 - 44	(-)	1.000	0.903	ccagGAGGagagg
V$EGRF	EGR/nerve growth factor induced protein C & related factors	V$WT1.01	0.92	46 - 62	(-)	1.000	0.930	aggtgTGGGcgagcacc
V$EGRF	EGR/nerve growth factor induced protein C & related factors	V$EGR2.01	0.79	48 - 64	(-)	0.777	0.840	tgagGTGTgggcgagca
V$HESF	Vertebrate homologues of enhancer of split complex	V$HES1.01	0.92	92 - 106	(-)	1.000	0.926	gagccccGTGCccac
V$HESF	Vertebrate homologues of enhancer of split complex	V$HELT.01	0.91	93 - 107	(+)	1.000	0.945	tgggCACGgggctct
V$AP1R	MAF and AP1 related factors	V$MAFA.01	0.92	112 - 132	(+)	1.000	0.968	cacTGCTgcgcgggcagggct
V$AHRR	AHR-arnt heterodimers and AHR-related factors	V$AHRARNT.03	0.95	125 - 149	(+)	1.000	0.977	gcagggctgtGCGTgcggggccgtc
V$EGRF	EGR/nerve growth factor induced protein C & related factors	V$NGFIC.01	0.80	131 - 147	(+)	1.000	0.827	ctgtGCGTgcggggccg
V$MYBL	Cellular and viral myb-like transcriptional regulators	V$VMYB.05	0.90	153 - 165	(-)	1.000	0.944	gaaAACGgcaatc
V$KLFS	Krueppel like transcription factors	V$GKLF.01	0.86	161 - 179	(-)	0.825	0.879	gaggagagaGGGGagaaaa
V$MZF1	Myeloid zinc finger 1 factors	V$MZF1.01	0.99	162 - 172	(-)	1.000	0.991	gaGGGGagaaa
V$MAZF	Myc associated zinc fingers	V$MAZ.01	0.90	164 - 176	(-)	1.000	0.921	gagaGAGGggaga
V$MAZF	Myc associated zinc fingers	V$MAZ.01	0.90	171 - 183	(-)	1.000	0.918	gggaGAGGagaga
V$AP1R	MAF and AP1 related factors	V$BACH2.01	0.89	204 - 224	(-)	1.000	0.960	gccggcTGAGtcagcgctgcc
V$AP1R	MAF and AP1 related factors	V$NFE2.01	0.85	206 - 226	(+)	1.000	0.913	cagcgCTGActcagccggcct
V$AP1R	MAF and AP1 related factors	V$VMAF.01	0.82	220 - 240	(-)	1.000	0.853	atttgcTGACaccgaggccgg
V$AP1R	MAF and AP1 related factors	V$MARE.01	0.97	224 - 244	(-)	1.000	0.995	ctggatttGCTGacaccgagg
V$MYBL	Cellular and viral myb-like transcriptional regulators	V$CMYB.02	0.96	246 - 258	(+)	0.898	0.960	agAACcgtcaag
V$HAND	Twist subfamily of class B bHLH transcription factors	V$PARAXIS.01	0.86	254 - 274	(+)	0.823	0.882	tcaagATCAcctgctccgggg
V$MYOD	Myoblast determining factors	V$E47.02	0.93	256 - 272	(-)	1.000	0.933	ccggagcaGGTGatctt
V$ETSF	Human and murine ETS1 factors	V$CETS1P54.01	0.92	281 - 301	(+)	0.843	0.928	gctatgCTGGaagttactatt
V$FKHD	Fork head domain factors	V$FHXB.01	0.83	293 - 309	(-)	1.000	0.830	ccagccATAAtagtaac
V$GATA	GATA binding factors	V$GATA3.02	0.91	343 - 355	(-)	1.000	0.935	catAGATcacagt
V$HOXC	HOX - PBX complexes	V$PBX_HOXA9.01	0.79	343 - 359	(+)	1.000	0.848	actgTGATctatgacaa
V$OCT1	Octamer binding protein	V$OCT.01	0.78	350 - 366	(+)	0.857	0.781	tctATGAcaacgacaag
V$EVI1	EVI1-myeloid transforming protein	V$EVI1.02	0.83	359 - 375	(+)	1.000	0.849	acgacAAGAgaccctcg
V$ETSF	Human and murine ETS1 factors	V$ELK1.02	0.91	384 - 404	(-)	1.000	0.972	ttggaaccGGAAaatcgtgaa
V$IRFF	Interferon regulatory factors	V$IRF7.01	0.86	385 - 405	(-)	0.821	0.895	tttgGAACcggaaaatcgtga
V$XBBF	X-box binding factors	V$MIF1.01	0.76	407 - 425	(+)	0.800	0.761	ccggctccacaGCCAcatt

V$HOXF	Paralog hox genes 1-8 from the four hox clusters A, B, C, D	V$HOX1-3.01	0.82	413 - 431	(-)	1.000	0.854	atggtTAATgtggctgtgg
V$IRFF	Interferon regulatory factors	V$IRF3.01	0.85	452 - 472	(-)	0.758	0.852	cacagaaataGACAgcctcgt
V$NFKB	Nuclear factor kappa B/c-rel	V$NFKAPPAB .01	0.89	509 - 521	(+)	1.000	0.895	ccGGGAcaaccct
V$KLFS	Krueppel like transcription factors	V$EKLF.01	0.89	512 - 530	(-)	1.000	0.911	aggacggtcaGGGTtgtcc
V$MZF1	Myeloid zinc finger 1 factors	V$MZF1.02	0.99	559 - 569	(-)	1.000	0.994	tgGGGGaagaa
V$GCMF	Chorion-specific transcription factors with a GCM DNA binding domain	V$GCM1.01	0.85	562 - 572	(+)	0.789	0.859	ttCCCCcatcg
V$GFI1	Growth factor independence transcriptional repressor	V$GFI1B.01	0.86	595 - 609	(-)	1.000	0.883	ccaAATCaccaaaaa
V$EGRF	EGR/nerve growth factor induced protein C & related factors	V$NGFIC.01	0.80	643 - 659	(-)	0.762	0.809	ccctGCGCcggtggcag
V$AP1R	MAF and AP1 related factors	V$NFE2.01	0.85	681 - 701	(+)	1.000	0.854	ctgggCTGAgtcttgtcccca
V$MZF1	Myeloid zinc finger 1 factors	V$MZF1.01	0.99	692 - 702	(-)	1.000	1.000	gtGGGGacaag
V$EVI1	EVI1-myleoid transforming protein	V$EVI1.06	0.83	739 - 755	(-)	0.750	0.875	ttgacaAAATaacccta
V$STAT	Signal transducer and activator of transcription	V$STAT1.01	0.77	810 - 828	(+)	0.767	0.771	acgatttcaGAAAttgtgt
V$CEBP	Ccaat/Enhancer Binding Protein	V$CEBPB.01	0.94	811 - 825	(-)	1.000	0.946	caatttctGAAAtcg
V$EVI1	EVI1-myleoid transforming protein	V$MEL1.02	0.99	829 - 845	(+)	1.000	0.990	ctaggtcGATGagaaga
V$EVI1	EVI1-myleoid transforming protein	V$EVI1.02	0.83	837 - 853	(+)	1.000	0.855	atgagAAGAcagttttt
V$ETSF	Human and murine ETS1 factors	V$ETS2.01	0.84	856 - 876	(+)	1.000	0.916	ctttgtcAGGAaattagttgt
V$HOXF	Paralog hox genes 1-8 from the four hox clusters A, B, C, D	V$HOXA3.01	0.86	858 - 876	(-)	1.000	0.877	acaacTAATttcctgacaa
V$HOXF	Paralog hox genes 1-8 from the four hox clusters A, B, C, D	V$HOXA5.01	0.84	861 - 879	(+)	1.000	0.848	tcaggaAATTagttgtgag
V$NBRE	NGFI-B response elements, nur subfamily of nuclear receptors	V$NBRE.01	0.86	909 - 923	(-)	1.000	0.861	gacaAAGGtcttagt
V$NOLF	Neuron-specific-olfactory factor	V$OLF1.01	0.82	931 - 953	(-)	1.000	0.852	caagagTCCCcagtgagagacc a
V$PAX5	PAX-5 B-cell-specific activator protein	V$PAX5.03	0.80	940 - 968	(-)	1.000	0.824	ccatgGCACtggagccaagagt ccccagt
V$PAX5	PAX-5 B-cell-specific activator protein	V$PAX5.01	0.79	942 - 970	(-)	0.952	0.800	ccccatGGCActggagccaaga gtcccca
V$GREF	Glucocorticoid responsive and related elements	V$ARE.02	0.89	982 - 1000	(+)	0.959	0.901	cactgacacggtGTCCttg
V$GREF	Glucocorticoid responsive and related elements	V$GRE.02	0.82	982 - 1000	(-)	0.874	0.872	caaGGACaccgtgtcagtg
V$PRDF	Positive regulatory domain I binding factor	V$PRDM1.01	0.81	1000 - 1018	(+)	1.000	0.830	gggggtGAAAttcagtttt
V$MYBL	Cellular and viral myb-like transcriptional regulators	V$VMYB.01	0.88	1007 - 1019	(-)	0.865	0.901	aaaAACTgaattt
V$STAT	Signal transducer and activator of transcription	V$STAT6.01	0.84	1053 - 1071	(-)	0.862	0.842	ctttTTCAgcagaataacc
V$AP1R	MAF and AP1 related factors	V$MARE.02	0.88	1058 - 1078	(+)	1.000	0.925	ttcTGCTgaaaaagctgagag
V$EVI1	EVI1-myleoid transforming protein	V$EVI1.04	0.73	1081 - 1097	(-)	0.750	0.751	aaaaaatgacaaAATAt
V$STAT	Signal transducer and activator of transcription	V$STAT.01	0.87	1091 - 1109	(+)	1.000	0.886	attttttcGGAAatatat
V$FKHD	Fork head domain factors	V$FREAC7.01	0.96	1119 - 1135	(+)	1.000	0.969	tatataTAAAtatataa
V$HOXF	Paralog hox genes 1-8 from the four hox clusters A, B, C, D	V$HOXC6.01	0.85	1126 - 1144	(-)	1.000	0.855	ataaatAATTtatatattt
V$HOXC	HOX - PBX complexes	V$PBX_HOX A9.01	0.79	1127 - 1143	(-)	0.750	0.834	taaaTAATttatatatt
V$OCT1	Octamer binding protein	V$OCT1.06	0.81	1127 - 1143	(+)	1.000	0.860	aatatataAATTattta
V$HOXF	Paralog hox genes 1-8 from the four hox clusters A, B, C, D	V$HOXB7.01	0.84	1129 - 1147	(+)	1.000	0.842	tatataAATTatttatata

V$FKHD	Fork head domain factors	V$FREAC7.01	0.96	1133 - 1149	(-)	1.000	0.966	aatataTAAAtaattta
V$HOXC	HOX - PBX complexes	V$PBX_HOXA9.01	0.79	1133 - 1149	(+)	0.750	0.834	taaaTTATttatatatt
V$FKHD	Fork head domain factors	V$FREAC7.01	0.96	1202 - 1218	(+)	1.000	0.969	tatataTAAAtatatat
V$FKHD	Fork head domain factors	V$XFD2.01	0.89	1229 - 1245	(-)	1.000	0.898	tatataTAAAgaaagag
V$FKHD	Fork head domain factors	V$XFD2.01	0.89	1239 - 1255	(-)	1.000	0.907	tccctaTAAAtatatat
V$AIRE	Autoimmune regulatory element binding factor	V$AIRE.01	0.86	1263 - 1289	(+)	0.916	0.861	tatatttttggccaattTGGCcaattt
V$PPAR	Peroxisome proliferative activated receptor homodimers	V$PPARG.01	0.67	1267 - 1289	(+)	0.794	0.741	tttTTGGccaatttggccaattt
V$PPAR	Peroxisome proliferative activated receptor homodimers	V$PPARG.01	0.67	1267 - 1289	(-)	0.794	0.734	aaaTTGGccaaattggccaaaa a
V$PRDF	Positive regulatory domain I binding factor	V$BLIMP1.01	0.81	1278 - 1296	(-)	1.000	0.856	gagagaGAAAttggccaaa
V$FKHD	Fork head domain factors	V$XFD2.01	0.89	1296 - 1312	(-)	1.000	0.890	tatataTAAAgagagag
V$HOXF	Paralog hox genes 1-8 from the four hox clusters A, B, C, D	V$HOXC6.01	0.85	1321 - 1339	(-)	1.000	0.850	tatataAATTacatatata
V$OCT1	Octamer binding protein	V$OCT1.06	0.81	1323 - 1339	(+)	0.750	0.827	tatatgtaATTTatata
V$HOXF	Paralog hox genes 1-8 from the four hox clusters A, B, C, D	V$HOXC6.01	0.85	1324 - 1342	(+)	1.000	0.858	atatgtAATTtatatatat
V$HOXC	HOX - PBX complexes	V$PBX_HOXA9.01	0.79	1325 - 1341	(+)	0.750	0.797	tatgTAATttatatata
V$OCT1	Octamer binding protein	V$OCT1.06	0.81	1325 - 1341	(-)	1.000	0.862	tatatataAATTacata
V$OCT1	Octamer binding protein	V$OCT1.06	0.81	1342 - 1358	(+)	1.000	0.862	tatatataAATTtatat
V$FKHD	Fork head domain factors	V$HFH3.01	0.97	1351 - 1367	(-)	0.955	0.976	aataaatAAATataaat
V$FKHD	Fork head domain factors	V$HFH3.01	0.97	1355 - 1371	(-)	0.955	0.974	aataaatAAATaaatat
V$FKHD	Fork head domain factors	V$HFH3.01	0.97	1359 - 1375	(-)	0.955	0.974	aataaatAAATaaataa
V$FKHD	Fork head domain factors	V$HFH3.01	0.97	1363 - 1379	(-)	0.955	0.974	tataaatAAATaaataa
V$FKHD	Fork head domain factors	V$FREAC7.01	0.96	1367 - 1383	(-)	1.000	0.970	tctataTAAAtaaataa
V$FKHD	Fork head domain factors	V$XFD2.01	0.89	1382 - 1398	(-)	1.000	0.903	tgtataTAAAaatatc
V$OCT1	Octamer binding protein	V$POU3F3.01	0.81	1386 - 1402	(-)	0.762	0.838	tatatGTATataaaaat
V$OCT1	Octamer binding protein	V$POU3F3.01	0.81	1391 - 1407	(+)	0.754	0.823	tatatACATatatatat
V$OCT1	Octamer binding protein	V$OCT1.06	0.81	1403 - 1419	(+)	1.000	0.862	tatatataAATTtggcc
V$HOXF	Paralog hox genes 1-8 from the four hox clusters A, B, C, D	V$HOX1-3.01	0.82	1454 - 1472	(-)	1.000	0.871	ctaagTAATttggttcaga
V$BCL6	POZ domain zinc finger expressed in B-Cells	V$BCL6.01	0.76	1495 - 1511	(-)	1.000	0.767	gaaTTCCtgaactgcac
V$ETSF	Human and murine ETS1 factors	V$ETS2.01	0.84	1497 - 1517	(+)	1.000	0.875	gcagttcAGGAattccctcgc
V$IKRS	Ikaros zinc finger family	V$IK3.01	0.84	1500 - 1512	(+)	1.000	0.853	gttcaGGAAttcc
V$NFKB	Nuclear factor kappa B/c-rel	V$CREL.01	0.91	1504 - 1516	(-)	1.000	0.940	cgagggaaTTCCt
V$IKRS	Ikaros zinc finger family	V$IK1.01	0.92	1505 - 1517	(-)	1.000	0.940	gcgaGGGAattcc
V$HOXF	Paralog hox genes 1-8 from the four hox clusters A, B, C, D	V$NANOG.01	0.94	1523 - 1541	(+)	1.000	0.957	agtttttAATGggggccag
V$MZF1	Myeloid zinc finger 1 factors	V$MZF1.02	0.99	1543 - 1553	(-)	1.000	0.994	tgGGGGaaaga
V$GATA	GATA binding factors	V$GATA2.02	0.90	1562 - 1574	(+)	1.000	0.911	tgtaGATAtgagt
V$EVI1	EVI1-myleoid transforming protein	V$EVI1.04	0.73	1566 - 1582	(+)	1.000	0.733	gatatgagtgaaGATAt
V$FKHD	Fork head domain factors	V$XFD3.01	0.82	1568 - 1584	(+)	0.782	0.826	tatgagtgAAGAtatag
V$EVI1	EVI1-myleoid transforming protein	V$EVI1.05	0.81	1571 - 1587	(+)	1.000	0.819	gagtgaaGATAtagata
V$FKHD	Fork head domain factors	V$HNF3B.02	0.91	1613 - 1629	(-)	0.938	0.910	gttcatCAAAcaaacta
V$ETSF	Human and murine ETS1 factors	V$PU1.01	0.89	1629 - 1649	(+)	1.000	0.915	cagtctgaGGAAgagtaaatc
V$ETSF	Human and murine ETS1 factors	V$PEA3.01	0.94	1646 - 1666	(-)	1.000	0.945	tgccatgAGGAaggcaagatt
V$OCT1	Octamer binding protein	V$OCT1.05	0.89	1658 - 1674	(+)	1.000	0.899	ctCATGgcaatttttaa
V$RBIT	Regulator of B-Cell IgH transcription	V$BRIGHT.01	0.92	1668 - 1680	(-)	1.000	0.937	gtagaATTAaaaa

V$AHRR	AHR-arnt heterodimers and AHR-related factors	V$AHRARNT.02	0.77	1723 - 1747	(+)	0.750	0.793	cccgaaaattGAGTgctttttctt t
V$STAT	Signal transducer and activator of transcription	V$STAT5.01	0.89	1736 - 1754	(-)	0.845	0.949	aaatTTCAaagaaaaagca
V$STAT	Signal transducer and activator of transcription	V$STAT5.01	0.89	1738 - 1756	(+)	1.000	0.942	ctttTTCTtgaaatttta
V$BCL6	POZ domain zinc finger expressed in B-Cells	V$BCL6.02	0.77	1739 - 1755	(+)	0.771	0.853	tttttctTTGAaatttt
V$RBIT	Regulator of B-Cell IgH transcription	V$BRIGHT.01	0.92	1758 - 1770	(+)	1.000	0.938	ataaaATTAaaaa
V$KLFS	Krueppel like transcription factors	V$GKLF.01	0.86	1785 - 1803	(+)	1.000	0.884	aggaaagaaAGGGgggtga
V$KLFS	Krueppel like transcription factors	V$GKLF.01	0.86	1788 - 1806	(+)	0.825	0.862	aaagaaaggGGGGtgatgg
V$FKHD	Fork head domain factors	V$FREAC7.01	0.96	1874 - 1890	(+)	1.000	0.969	tatataTAAAtatatat
V$OCT1	Octamer binding protein	V$POU3F3.01	0.81	1883 - 1899	(-)	0.754	0.820	tgtatACATatatatat
V$FKHD	Fork head domain factors	V$FREAC4.01	0.78	1887 - 1903	(-)	0.750	0.795	cttttgtaTACAtatat
V$FKHD	Fork head domain factors	V$FKHRL1.01	0.83	1888 - 1904	(+)	0.750	0.848	tatatgtaTACAaaaga
V$OCT1	Octamer binding protein	V$POU3F3.01	0.81	1888 - 1904	(+)	0.762	0.822	tatatGTATacaaaaga
V$FKHD	Fork head domain factors	V$XFD2.01	0.89	1912 - 1928	(+)	1.000	0.894	tatataTAAAaaattg
V$FKHD	Fork head domain factors	V$FHXB.01	0.83	1919 - 1935	(-)	0.909	0.845	tacataACAAtttttt
V$OCT1	Octamer binding protein	V$OCT1.05	0.89	1919 - 1935	(-)	0.850	0.905	taCATAacaatttttt
V$FKHD	Fork head domain factors	V$FHXB.01	0.83	1922 - 1938	(-)	1.000	0.903	atatacATAAcaattt
V$OCT1	Octamer binding protein	V$OCT3_4.02	0.88	1923 - 1939	(-)	0.790	0.896	aatatACATaacaattt
V$OCT1	Octamer binding protein	V$POU3F3.01	0.81	1928 - 1944	(+)	0.762	0.851	gttatGTATattatata
V$SATB	Special AT-rich sequence binding protein	V$SATB1.01	0.94	1928 - 1942	(-)	1.000	0.958	tatAATAtacataac
V$HOXC	HOX - PBX complexes	V$PBX_HOXA9.01	0.79	1940 - 1956	(+)	0.750	0.791	atatAGATttattatag
V$SATB	Special AT-rich sequence binding protein	V$SATB1.01	0.94	1941 - 1955	(-)	1.000	0.966	tatAATAaatctata
V$OCT1	Octamer binding protein	V$OCT3_4.02	0.88	1951 - 1967	(-)	0.790	0.882	tatgtACATatctataa
V$OCT1	Octamer binding protein	V$OCT1.01	0.77	1953 - 1969	(-)	1.000	0.771	ctTATGtacatatctat
V$OCT1	Octamer binding protein	V$OCT.01	0.78	1956 - 1972	(+)	0.795	0.804	gatATGTacataagata
V$EVI1	EVI1-myleoid transforming protein	V$EVI1.05	0.81	1962 - 1978	(+)	1.000	0.876	tacataaGATAtaacac
V$GATA	GATA binding factors	V$GATA3.02	0.91	1965 - 1977	(+)	1.000	0.926	ataAGATataaca
V$OCT1	Octamer binding protein	V$POU3F3.01	0.81	1973 - 1989	(-)	0.762	0.822	tatatGTATacgtgtta
V$OCT1	Octamer binding protein	V$POU3F3.01	0.81	1978 - 1994	(+)	0.754	0.819	cgtatACATatatatgt
V$OCT1	Octamer binding protein	V$OCT1.06	0.81	2012 - 2028	(-)	1.000	0.851	tctataaaAATTtatac
V$FKHD	Fork head domain factors	V$XFD2.01	0.89	2014 - 2030	(-)	1.000	0.899	tatctaTAAAaatttat
V$GATA	GATA binding factors	V$GATA3.02	0.91	2023 - 2035	(+)	1.000	0.949	tatAGATatgatg
V$EVI1	EVI1-myleoid transforming protein	V$EVI1.05	0.81	2025 - 2041	(+)	0.750	0.816	tagatatGATGtatatt
V$HOXC	HOX - PBX complexes	V$PBX_HOXA9.01	0.79	2027 - 2043	(+)	1.000	0.838	gataTGATgtatattat
V$OCT1	Octamer binding protein	V$POU3F3.01	0.81	2036 - 2052	(-)	1.000	0.887	cccatGCATataatata
V$OCT1	Octamer binding protein	V$OCT3_4.02	0.88	2041 - 2057	(+)	1.000	0.902	tatatGCATgggttata
V$FKHD	Fork head domain factors	V$FHXB.01	0.83	2057 - 2073	(+)	1.000	0.852	atatctATAAtatattt
V$SATB	Special AT-rich sequence binding protein	V$SATB1.01	0.94	2062 - 2076	(+)	1.000	0.959	tatAATAtattttat
V$EVI1	EVI1-myleoid transforming protein	V$EVI1.05	0.81	2063 - 2079	(-)	0.750	0.876	cagataaAATAtattat
V$GATA	GATA binding factors	V$GATA1.05	0.88	2069 - 2081	(-)	1.000	0.984	tacaGATAaaata
V$PLZF	C2H2 zinc finger protein PLZF	V$PLZF.01	0.86	2087 - 2101	(+)	1.000	0.930	aacTACAgtataaaa
V$OCT1	Octamer binding protein	V$OCT3_4.02	0.88	2093 - 2109	(-)	1.000	0.928	tatatGCATtttatact
V$OCT1	Octamer binding protein	V$POU3F3.01	0.81	2098 - 2114	(+)	1.000	0.894	aaaatGCATatatgcgt
V$OCT1	Octamer binding protein	V$POU3F3.01	0.81	2101 - 2117	(-)	1.000	0.831	tatacGCATatatgcat

Family	Description	Matrix	Core sim.	Position	Strand	Core sim.	Matrix sim.	Sequence
V$OCT1	Octamer binding protein	V$POU2F3.01	0.80	2106 - 2122	(+)	1.000	0.822	tatATGCgtatatatac
V$FKHD	Fork head domain factors	V$FREAC7.01	0.96	2121 - 2137	(-)	1.000	0.971	tatataTAAAtatgtgt
V$OCT1	Octamer binding protein	V$POU3F3.01	0.81	2141 - 2157	(-)	0.754	0.819	cgtatACATatatatac
V$OCT1	Octamer binding protein	V$POU3F3.01	0.81	2146 - 2162	(+)	0.762	0.819	tatatGTATacgcacac
V$EGRF	EGR/nerve growth factor induced protein C & related factors	V$EGR3.01	0.77	2152 - 2168	(-)	0.758	0.779	ttatACGTgtgcgtata
V$OCT1	Octamer binding protein	V$POU2F3.01	0.80	2171 - 2187	(+)	0.762	0.828	tccATGAaaatagatat
V$GATA	GATA binding factors	V$GATA1.06	0.96	2179 - 2191	(+)	1.000	0.980	aataGATAtggat
V$FKHD	Fork head domain factors	V$HFH3.01	0.97	2196 - 2212	(-)	1.000	0.977	gtaaaacAAACacatcc
V$CABL	C-abl DNA binding sites	V$CABL.01	0.97	2200 - 2210	(-)	1.000	0.973	aaAACAaacac
V$MITF	Microphthalmia transcription factor	V$MIT.01	0.81	2213 - 2231	(+)	1.000	0.837	agaggtgCATGtgtgtctg
V$HAND	Twist subfamily of class B bHLH transcription factors	V$HAND2_E 12.01	0.75	2284 - 2304	(+)	1.000	0.819	ccagcccaaggTGGCccccac
V$EGRF	EGR/nerve growth factor induced protein C & related factors	V$WT1.01	0.92	2292 - 2308	(-)	1.000	0.933	gatggTGGGggccacct
V$ETSF	Human and murine ETS1 factors	V$GABP.01	0.86	2306 - 2326	(-)	1.000	0.869	tgacggtgGGAAgagggtgat
V$RBPF	RBPJ - kappa	V$RBPJK.02	0.94	2310 - 2324	(-)	1.000	0.948	acggTGGGaagaggg
V$KLFS	Krueppel like transcription factors	V$GKLF.02	0.96	2319 - 2337	(+)	1.000	0.986	caccgtcAAAGgaggagct
V$GATA	GATA binding factors	V$GATA.01	0.93	2367 - 2379	(+)	1.000	0.976	gcctGATAaacga
V$HAML	Human acute myelogenous leukemia factors	V$AML3.01	0.84	2398 - 2412	(+)	1.000	0.868	gactGTGGattgggt
V$HOXC	HOX - PBX complexes	V$HOX_PBX. 01	0.81	2408 - 2424	(+)	1.000	0.912	tgggTGATcgatggctc
V$DICE	Downstream Immunoglobulin Control Element, critical for B cell activity and specificity	V$DICE.01	0.80	2430 - 2444	(-)	0.756	0.822	tggtCTCGccagagc
V$HAML	Human acute myelogenous leukemia factors	V$AML1.01	0.93	2435 - 2449	(-)	1.000	0.971	tgctGTGGtctcgcc
V$HAML	Human acute myelogenous leukemia factors	V$AML1.01	0.93	2443 - 2457	(-)	1.000	0.932	cgctGTGGtgctgtg
V$XBBF	X-box binding factors	V$RFX1.01	0.89	2452 - 2470	(+)	1.000	0.914	acagcggcagaGCAAcagc
V$MYBL	Cellular and viral myb-like transcriptional regulators	V$CMYB.01	0.90	2462 - 2474	(+)	1.000	0.983	agCAACagccagt
V$PAX5	PAX-5 B-cell-specific activator protein	V$PAX5.03	0.80	2488 - 2516	(-)	0.789	0.815	accagTCGCtggcagacagtgacaggtag
V$PAX5	PAX-5 B-cell-specific activator protein	V$PAX5.03	0.80	2490 - 2518	(+)	0.947	0.876	acctgTCACtgtctgccagcgactggtca
V$AHRR	AHR-arnt heterodimers and AHR-related factors	V$AHRARNT. 02	0.77	2512 - 2536	(-)	0.750	0.772	ggtgtaggtcTCGTggcttgaccag
V$HESF	Vertebrate homologues of enhancer of split complex	V$HELT.01	0.91	2518 - 2532	(+)	1.000	0.932	aagcCACGagaccta
V$HAND	Twist subfamily of class B bHLH transcription factors	V$TAL1_E2A. 01	0.98	2525 - 2545	(-)	1.000	0.981	gaccctgCAGGtgtaggtctc
V$MYOD	Myoblast determining factors	V$E47.01	0.92	2528 - 2544	(-)	1.000	0.973	accctGCAGgtgtaggt
V$KLFS	Krueppel like transcription factors	V$EKLF.01	0.89	2531 - 2549	(+)	1.000	0.907	tacacctgcaGGGTcacac
V$AP1R	MAF and AP1 related factors	V$TCF11MAF G.01	0.81	2532 - 2552	(-)	1.000	0.866	tgtgtgTGACcctgcaggtgt
V$MYBL	Cellular and viral myb-like transcriptional regulators	V$VMYB.02	0.90	2549 - 2561	(+)	1.000	0.970	cacAACGgcacct
V$GATA	GATA binding factors	V$GATA1.01	0.96	2558 - 2570	(-)	1.000	0.960	tcgtGATAgaggt
V$DICE	Downstream Immunoglobulin Control Element, critical for B cell activity and specificity	V$DICE.01	0.80	2573 - 2587	(-)	1.000	0.815	ggacCTCTtcaggt
V$NOLF	Neuron-specific-olfactory factor	V$OLF1.02	0.88	2594 - 2616	(+)	1.000	0.882	taatagTCCCactggggatgcaa
V$ETSF	Human and murine ETS1 factors	V$PU1.01	0.89	2622 - 2642	(-)	1.000	0.905	gagggtgaGGAAccactgtcc

V$HAML	Human acute myelogenous leukemia factors	V$AML3.01	0.84	2623 - 2637	(+)	1.000	0.880	gacaGTGGttcctca
V$KLFS	Krueppel like transcription factors	V$EKLF.01	0.89	2632 - 2650	(-)	1.000	0.970	gggacagggaGGGTgagga
V$NOLF	Neuron-specific-olfactory factor	V$OLF1.02	0.88	2641 - 2663	(+)	1.000	0.890	tccctgTCCCtctgggccgctgc
V$ETSF	Human and murine ETS1 factors	V$GABP.01	0.86	2674 - 2694	(-)	1.000	0.863	tctgagtgGGAAgtgagggct
V$RBPF	RBPJ - kappa	V$RBPJK.02	0.94	2678 - 2692	(-)	1.000	0.959	tgagTGGGaagtgag
V$HAND	Twist subfamily of class B bHLH transcription factors	V$HAND2_E 12.01	0.75	2684 - 2704	(+)	0.751	0.820	tcccactcagaTGTCcccac
V$PAX5	PAX-5 B-cell-specific activator protein	V$PAX5.03	0.80	2686 - 2714	(-)	1.000	0.837	tggggGCACggtggggacatc tgagtgg
V$EGRF	EGR/nerve growth factor induced protein C & related factors	V$WT1.01	0.92	2692 - 2708	(-)	1.000	0.920	cacggTGGGggacatct
V$PAX5	PAX-5 B-cell-specific activator protein	V$PAX5.01	0.79	2707 - 2735	(-)	0.952	0.809	tcgacaGGCAgaggtgggtgat gggggca
V$EVI1	EVI1-myleoid transforming protein	V$MEL1.02	0.99	2743 - 2759	(-)	1.000	0.993	ctggagaGATGagggca
V$GCMF	Chorion-specific transcription factors with a GCM DNA binding domain	V$GCM1.01	0.85	2743 - 2753	(+)	1.000	0.904	tgCCCTcatct
V$BRAC	Brachyury gene, mesoderm developmental factor	V$TBX5.01	0.99	2749 - 2769	(+)	1.000	0.994	catctctccaGGTGtcacatt
V$MYOD	Myoblast determining factors	V$MYOD.01	0.88	2752 - 2768	(-)	0.931	0.912	atgtGACAcctggagag
V$HOXF	Paralog hox genes 1-8 from the four hox clusters A, B, C, D	V$HOXB4.01	0.84	2761 - 2779	(-)	0.756	0.849	cgtgttTATTaatgtgaca
V$HOXF	Paralog hox genes 1-8 from the four hox clusters A, B, C, D	V$HOXA3.01	0.86	2764 - 2782	(+)	1.000	0.864	cacatTAATaaacacgaca
V$FKHD	Fork head domain factors	V$HFH8.01	0.92	2766 - 2782	(+)	1.000	0.956	cattaatAAACacgaca
V$AP1R	MAF and AP1 related factors	V$MARE.01	0.97	2785 - 2805	(+)	1.000	0.994	gaactagtGCTGactctgcat
V$AP1R	MAF and AP1 related factors	V$MAFA.01	0.92	2787 - 2807	(-)	0.904	0.956	ggaTGCAgagtcagcactagt
V$AP1R	MAF and AP1 related factors	V$MARE.02	0.88	2789 - 2809	(+)	1.000	1.000	tagTGCTgactctgcatccat
V$PAX5	PAX-5 B-cell-specific activator protein	V$PAX5.01	0.79	2790 - 2818	(-)	0.761	0.791	cacagaGACAtggatgcagagt cagcact
V$OCT1	Octamer binding protein	V$OCT3_4.02	0.88	2797 - 2813	(+)	1.000	0.892	actctGCATccatgtct
V$GREF	Glucocorticoid responsive and related elements	V$ARE.02	0.89	2806 - 2824	(+)	0.959	0.901	ccatgtctctgtGTCCttt
V$NBRE	NGFI-B response elements, nur subfamily of nuclear receptors	V$NBRE.01	0.86	2813 - 2827	(-)	1.000	0.878	gcaaAAGGacacaga
V$AHRR	AHR-arnt heterodimers and AHR-related factors	V$AHRARNT.01	0.92	2816 - 2840	(+)	1.000	0.972	gtgtcctttgCGTGctgtctgca t
V$OCT1	Octamer binding protein	V$OCT3_4.02	0.88	2832 - 2848	(+)	1.000	0.936	tgtctGCATctcacaca
V$ETSF	Human and murine ETS1 factors	V$SPI1_PU1.0 2	0.96	2862 - 2882	(+)	1.000	0.963	cagtatggGGAAgggctgggg
V$MZF1	Myeloid zinc finger 1 factors	V$MZF1.02	0.99	2866 - 2876	(+)	1.000	0.994	atGGGGaaggg
V$HOXF	Paralog hox genes 1-8 from the four hox clusters A, B, C, D	V$HOX1-3.01	0.82	2898 - 2916	(+)	1.000	0.841	attggTAATgttgggggcg
V$EGRF	EGR/nerve growth factor induced protein C & related factors	V$EGR1.02	0.88	2904 - 2920	(+)	1.000	0.897	aatgttggGGGCggggg
V$KLFS	Krueppel like transcription factors	V$KKLF.01	0.91	2907 - 2925	(+)	1.000	0.953	gttgggggcGGGGgggga
V$EGRF	EGR/nerve growth factor induced protein C & related factors	V$WT1.01	0.92	2910 - 2926	(+)	0.953	0.984	gggggCGGGggggggag
V$KLFS	Krueppel like transcription factors	V$KKLF.01	0.91	2910 - 2928	(+)	1.000	0.940	gggggcgggGGGGggaggg
V$MAZF	Myc associated zinc fingers	V$MAZ.01	0.90	2910 - 2922	(+)	0.866	0.900	ggggGCGGggggg
V$EGRF	EGR/nerve growth factor induced protein C & related factors	V$WT1.01	0.92	2912 - 2928	(+)	0.837	0.945	gggcgGGGGggggaggg
V$KLFS	Krueppel like transcription factors	V$KKLF.01	0.91	2912 - 2930	(+)	1.000	0.969	gggcgggggGGGGagggcg
V$EGRF	EGR/nerve growth factor induced protein C & related factors	V$WT1.01	0.92	2914 - 2930	(+)	0.837	0.943	gcgggGGGGggagggcg

V$KLFS	Krueppel like transcription factors	V$KLF6.01	0.92	2914 - 2932	(+)	1.000	0.922	gcggggGGGGgagggcggg
V$MAZF	Myc associated zinc fingers	V$MAZR.01	0.88	2915 - 2927	(+)	1.000	0.959	cggggGGGGagg
V$EGRF	EGR/nerve growth factor induced protein C & related factors	V$CKROX.01	0.88	2918 - 2934	(+)	1.000	0.992	ggggGGGAgggcggggg
V$MZF1	Myeloid zinc finger 1 factors	V$MZF1.03	0.95	2919 - 2929	(+)	1.000	1.000	ggGGGGagggc
V$KLFS	Krueppel like transcription factors	V$KKLF.01	0.91	2921 - 2939	(+)	1.000	0.925	ggggagggcGGGGgggtca
V$EGRF	EGR/nerve growth factor induced protein C & related factors	V$WT1.01	0.92	2924 - 2940	(+)	0.953	0.984	gagggCGGGggggtcaa
V$EGRF	EGR/nerve growth factor induced protein C & related factors	V$WT1.01	0.92	2926 - 2942	(+)	0.837	0.932	gggcgGGGGggtcaaca
V$MAZF	Myc associated zinc fingers	V$MAZR.01	0.88	2927 - 2939	(+)	1.000	0.939	ggcgggGGGGtca
V$PAX5	PAX-5 B-cell-specific activator protein	V$PAX5.01	0.79	2950 - 2978	(+)	0.761	0.798	actggaGACActggtgtataccc tggcaa
V$XBBF	X-box binding factors	V$RFX1.01	0.89	2964 - 2982	(+)	1.000	0.913	tgtataccctgGCAAcacc
V$HAND	Twist subfamily of class B bHLH transcription factors	V$PARAXIS.0 1	0.86	3040 - 3060	(+)	0.882	0.896	cactgAGCAcatggctgtggc
V$MITF	Microphthalmia transcription factor	V$MIT.01	0.81	3041 - 3059	(-)	1.000	0.838	ccacagcCATGtgctcagt
V$GREF	Glucocorticoid responsive and related elements	V$ARE.02	0.89	3055 - 3073	(+)	0.959	0.914	tgtggccgtgctGTCCtcg
V$GREF	Glucocorticoid responsive and related elements	V$GRE.02	0.82	3055 - 3073	(-)	0.874	0.820	cgaGGACagcacggccaca
V$EGRF	EGR/nerve growth factor induced protein C & related factors	V$WT1.01	0.92	3069 - 3085	(-)	1.000	0.935	gtgggTGGGagacgagg
V$EGRF	EGR/nerve growth factor induced protein C & related factors	V$EGR2.01	0.79	3071 - 3087	(-)	0.766	0.827	tagtGGGTgggagacga
V$GFI1	Growth factor independence transcriptional repressor	V$GFI1.02	0.90	3107 - 3121	(-)	1.000	0.947	gcaAATCtctgctag
V$OCT1	Octamer binding protein	V$OCT1.02	0.85	3110 - 3126	(-)	0.750	0.855	ataAAGCaaatctctgc
V$HOXC	HOX - PBX complexes	V$PBX_HOX A9.01	0.79	3115 - 3131	(+)	0.750	0.821	gattTGCTttatgctgg
V$RBPF	RBPJ - kappa	V$RBPJK.01	0.84	3125 - 3139	(+)	1.000	0.848	atgcTGGGaacaggg
V$IKRS	Ikaros zinc finger family	V$IK3.01	0.84	3126 - 3138	(+)	1.000	0.843	tgctgGGAAcagg
V$OCT1	Octamer binding protein	V$OCT1.02	0.85	3152 - 3168	(-)	1.000	0.856	tgaATGCaagggaacag
V$BRAC	Brachyury gene, mesoderm developmental factor	V$BRACH.01	0.66	3178 - 3198	(+)	0.750	0.675	gtgcccctGGGTgggatgtc
V$HAML	Human acute myelogenous leukemia factors	V$AML3.01	0.84	3245 - 3259	(-)	1.000	0.854	gcgtGTGGtgtgg
V$KLFS	Krueppel like transcription factors	V$BKLF.01	0.95	3264 - 3282	(-)	1.000	0.966	gacagGGGTggggtgcaca
V$MAZF	Myc associated zinc fingers	V$MAZR.01	0.88	3267 - 3279	(-)	1.000	0.909	aggggtGGGGtgc
V$KLFS	Krueppel like transcription factors	V$KKLF.01	0.91	3288 - 3306	(-)	1.000	0.919	ctggggagcGGGGagaggg
V$MZF1	Myeloid zinc finger 1 factors	V$MZF1.01	0.99	3289 - 3299	(-)	1.000	0.991	gcGGGGagagg
V$NOLF	Neuron-specific-olfactory factor	V$OLF1.02	0.88	3293 - 3315	(-)	1.000	0.911	aaaatcTCCCtggggagcgggg a
V$NOLF	Neuron-specific-olfactory factor	V$OLF1.02	0.88	3294 - 3316	(+)	1.000	0.902	ccccgcTCCCcagggagatttt g
V$GREF	Glucocorticoid responsive and related elements	V$GRE.02	0.82	3380 - 3398	(+)	0.890	0.826	tgtGCACagcttgtcagct
V$FKHD	Fork head domain factors	V$HNF3.01	0.98	3399 - 3415	(-)	1.000	1.000	cctcagcAAACacacac
V$AIRE	Autoimmune regulatory element binding factor	V$AIRE.01	0.86	3443 - 3469	(-)	0.964	0.860	atcgatggaggttgagcAGGAc aagac
V$HOXF	Paralog hox genes 1-8 from the four hox clusters A, B, C, D	V$NANOG.01	0.94	3513 - 3531	(-)	1.000	0.947	tgggatgAATGgagatgtt
V$HOXC	HOX - PBX complexes	V$HOX_PBX.01	0.81	3517 - 3533	(-)	0.944	0.863	ggtgGGATgaatggaga
V$EGRF	EGR/nerve growth factor induced protein C & related factors	V$WT1.01	0.92	3524 - 3540	(-)	1.000	0.948	ttgggTGGGtggatga

V$EGRF	EGR/nerve growth factor induced protein C & related factors	V$NGFIC.01	0.80	3526 - 3542	(-)	0.754	0.881	gtttGGGTgggtgggat
V$HOXC	HOX - PBX complexes	V$HOX_PBX.01	0.81	3544 - 3560	(-)	1.000	0.908	gttgTGATggatggaca
V$FKHD	Fork head domain factors	V$FHXB.01	0.83	3550 - 3566	(+)	0.909	0.844	tccatcACAAcacctcc
V$EGRF	EGR/nerve growth factor induced protein C & related factors	V$EGR1.01	0.79	3631 - 3647	(-)	0.832	0.903	ggttgcgttGGAGtggc
V$MYBL	Cellular and viral myb-like transcriptional regulators	V$CMYB.01	0.90	3641 - 3653	(+)	1.000	0.988	cgCAACcgtccat
V$HOXF	Paralog hox genes 1-8 from the four hox clusters A, B, C, D	V$NANOG.01	0.94	3643 - 3661	(-)	1.000	0.942	tagatggAATGgacggttg
V$HOXC	HOX - PBX complexes	V$HOX_PBX.01	0.81	3668 - 3684	(-)	0.944	0.911	gctgGGATggatggaca
V$EGRF	EGR/nerve growth factor induced protein C & related factors	V$WT1.01	0.92	3702 - 3718	(-)	1.000	0.948	tagggTGGGtgggatgg
V$EGRF	EGR/nerve growth factor induced protein C & related factors	V$NGFIC.01	0.80	3704 - 3720	(-)	0.754	0.845	cataGGGTgggtgggat
V$PAX5	PAX-5 B-cell-specific activator protein	V$PAX5.01	0.79	3725 - 3753	(+)	0.809	0.796	catccaGCCActggtggggtgcaggacat
V$KLFS	Krueppel like transcription factors	V$EKLF.01	0.89	3731 - 3749	(+)	1.000	0.915	gccactggtgGGGTgcagg
V$HAML	Human acute myelogenous leukemia factors	V$AML3.01	0.84	3774 - 3788	(+)	0.767	0.863	gactGGGGttttgc
V$OCT1	Octamer binding protein	V$OCT1.02	0.85	3777 - 3793	(-)	1.000	0.951	gccATGCaaaaacccca
V$OCT1	Octamer binding protein	V$OCT3_4.02	0.88	3782 - 3798	(+)	1.000	0.914	ttttGCATggccccat
V$PRDF	Positive regulatory domain I binding factor	V$BLIMP1.01	0.81	3818 - 3836	(+)	1.000	0.811	cgcactGAAAgtttggctg
V$OCT1	Octamer binding protein	V$OCT1.06	0.81	3857 - 3873	(+)	1.000	0.824	gctgtggaAATTctgta
V$NFAT	Nuclear factor of activated T-cells	V$NFAT.01	0.95	3859 - 3877	(+)	1.000	0.960	tgtGGAAattctgtaattt
V$NFKB	Nuclear factor kappa B/c-rel	V$NFKAPPAB65.01	0.87	3861 - 3873	(-)	1.000	0.874	tacagaatTTCCa
V$CEBP	Ccaat/Enhancer Binding Protein	V$CEBPB.01	0.94	3863 - 3877	(+)	0.940	0.950	gaaattctGTAAttt
V$GATA	GATA binding factors	V$GATA1.05	0.88	3882 - 3894	(-)	1.000	0.918	actgGATAaaaga
V$NFKB	Nuclear factor kappa B/c-rel	V$NFKAPPAB.01	0.89	3906 - 3918	(+)	1.000	0.941	ctGGGAtctccca
V$NFKB	Nuclear factor kappa B/c-rel	V$NFKAPPAB.01	0.89	3907 - 3919	(-)	1.000	0.942	ctGGGAgatccca
V$HOXF	Paralog hox genes 1-8 from the four hox clusters A, B, C, D	V$NANOG.01	0.94	3924 - 3942	(-)	1.000	0.946	tgtcatgAATGcctgccat
V$GATA	GATA binding factors	V$GATA2.02	0.90	3948 - 3960	(-)	1.000	0.909	atcaGATActaaa
V$HOXC	HOX - PBX complexes	V$HOX_PBX.01	0.81	3953 - 3969	(+)	1.000	0.828	tatcTGATgggtgggat
V$IRFF	Interferon regulatory factors	V$IRF7.01	0.86	3972 - 3992	(-)	1.000	0.913	gaaaGAAAatgaagttcagcc
V$GREF	Glucocorticoid responsive and related elements	V$ARE.01	0.80	3973 - 3991	(+)	0.750	0.829	gctgaacttcaTTTTcttt
V$MYBL	Cellular and viral myb-like transcriptional regulators	V$CMYB.01	0.90	4007 - 4019	(-)	1.000	0.901	ttCAACtgcaaaa
V$IRFF	Interferon regulatory factors	V$IRF7.01	0.86	4013 - 4033	(+)	0.936	0.869	agttGAATatgaattcctgct
V$ETSF	Human and murine ETS1 factors	V$ETS2.01	0.84	4017 - 4037	(-)	1.000	0.863	agaaagcAGGAattcatattc
V$RBPF	RBPJ - kappa	V$RBPJK.02	0.94	4051 - 4065	(-)	1.000	0.947	gtccTGGGaaaaaaa
V$AP1R	MAF and AP1 related factors	V$MARE.03	0.82	4066 - 4086	(-)	1.000	0.837	gactGCTGattccctgtgtac
V$HOXC	HOX - PBX complexes	V$PBX1.01	0.78	4211 - 4227	(+)	1.000	0.819	tgctgGATTgacatgag
V$OCT1	Octamer binding protein	V$OCT1.05	0.89	4211 - 4227	(-)	1.000	0.914	ctCATGtcaatccagca
V$GFI1	Growth factor independence transcriptional repressor	V$GFI1B.01	0.86	4244 - 4258	(-)	1.000	0.880	caaAATCaccatttc
V$ETSF	Human and murine ETS1 factors	V$CETS1P54.01	0.92	4269 - 4289	(+)	0.901	0.932	tgcactCAGGaagctgaaggc

	V$HOXC	HOX - PBX complexes	V$HOX_PBX.01	0.81	4297 - 4313	(+)	1.000	0.870	tcagTGATggatttacc
	V$PAX5	PAX-5 B-cell-specific activator protein	V$PAX5.03	0.80	4299 - 4327	(-)	1.000	0.876	tcaggGCACagtttggtaaatcc atcact
	V$AP1R	MAF and AP1 related factors	V$MARE.03	0.82	4317 - 4337	(-)	1.000	0.834	agcaGCTGcctcagggcacag
	V$GATA	GATA binding factors	V$GATA1.05	0.88	4341 - 4353	(+)	1.000	0.896	gctgGATAaagtc
	V$XBBF	X-box binding factors	V$RFX1.02	0.90	4355 - 4373	(+)	0.881	0.924	ctggagcacagGGAAccag
	V$GFI1	Growth factor independence transcriptional repressor	V$GFI1.01	0.96	4378 - 4392	(-)	1.000	0.970	agaAATCcctgccca
	V$NFKB	Nuclear factor kappa B/c-rel	V$NFKAPPAB.02	0.82	4382 - 4394	(+)	1.000	0.840	caGGGAtttctca
	V$NFAT	Nuclear factor of activated T-cells	V$NFAT5.01	0.83	4404 - 4422	(+)	1.000	0.896	cttGGAAagctgcaggctg
	V$GCMF	Chorion-specific transcription factors with a GCM DNA binding domain	V$GCM1.01	0.85	4448 - 4458	(+)	1.000	0.940	caCCCTcatga
	V$HOXF	Paralog hox genes 1-8 from the four hox clusters A, B, C, D	V$HOXB7.01	0.84	4463 - 4481	(-)	1.000	0.973	atcaatAATTagtgaggtt
	V$OCT1	Octamer binding protein	V$OCT1.03	0.85	4465 - 4481	(+)	1.000	0.951	cctcactaATTAttgat
	V$HOXF	Paralog hox genes 1-8 from the four hox clusters A, B, C, D	V$HOXB7.01	0.84	4466 - 4484	(+)	1.000	0.980	ctcactAATTattgattaa
	V$OCT1	Octamer binding protein	V$OCT1.03	0.85	4466 - 4482	(-)	1.000	0.938	aatcaataATTAgtgag
	V$HOXF	Paralog hox genes 1-8 from the four hox clusters A, B, C, D	V$HOXB4.01	0.84	4470 - 4488	(-)	0.757	0.844	tcttttAATCaataattag
	V$OCT1	Octamer binding protein	V$OCT1.03	0.85	4472 - 4488	(+)	1.000	0.860	aattattgATTAaaaga
	V$HOXF	Paralog hox genes 1-8 from the four hox clusters A, B, C, D	V$HOXC4.01	0.84	4473 - 4491	(+)	0.754	0.847	attattGATTaaaagattt
	V$HAML	Human acute myelogenous leukemia factors	V$AML3.01	0.84	4512 - 4526	(-)	1.000	0.844	ctcaGTGGgttgttg
	V$HAND	Twist subfamily of class B bHLH transcription factors	V$PARAXIS.01	0.86	4522 - 4542	(-)	0.882	0.926	aattcAGCAcatgctcctcag
	V$MITF	Microphthalmia transcription factor	V$MIT.01	0.81	4523 - 4541	(+)	1.000	0.886	tgaggagCATGtgctgaat
	V$AP1R	MAF and AP1 related factors	V$MARE.03	0.82	4531 - 4551	(+)	1.000	0.841	atgtGCTGaattatacatcac
	V$OCT1	Octamer binding protein	V$POU3F3.01	0.81	4535 - 4551	(-)	0.762	0.817	gtgatGTATaattcagc
	V$HOXC	HOX - PBX complexes	V$HOX_PBX.01	0.81	4538 - 4554	(-)	1.000	0.823	gctgTGATgtataattc
	V$PAX5	PAX-5 B-cell-specific activator protein	V$PAX5.01	0.79	4546 - 4574	(-)	0.952	0.803	tatgagGGCAgggccgtggagc tgtgatg
	V$KLFS	Krueppel like transcription factors	V$GKLF.01	0.86	4556 - 4574	(-)	1.000	0.881	tatgagggcAGGGccgtgg
	V$GCMF	Chorion-specific transcription factors with a GCM DNA binding domain	V$GCM1.01	0.85	4565 - 4575	(+)	1.000	0.925	tgCCCTcatag
	V$PAX5	PAX-5 B-cell-specific activator protein	V$PAX5.01	0.79	4568 - 4596	(+)	0.904	0.844	cctcatAGCAgggctgcatggca cccaca
	V$PAX5	PAX-5 B-cell-specific activator protein	V$PAX5.03	0.80	4577 - 4605	(-)	0.894	0.804	gagtgCCACtgtgggtgccatgc agccct
	V$PAX5	PAX-5 B-cell-specific activator protein	V$PAX5.02	0.73	4592 - 4620	(+)	1.000	0.779	ccacagtggcactcAGCGggac cacaggg
	V$HAML	Human acute myelogenous leukemia factors	V$AML1.01	0.93	4606 - 4620	(-)	1.000	0.945	ccctGTGGtcccgct
	V$AP1R	MAF and AP1 related factors	V$NRL.01	0.85	4614 - 4634	(-)	1.000	0.864	cccgGCTGtctcagccctgtg
	V$AP1R	MAF and AP1 related factors	V$MARE.02	0.88	4616 - 4636	(+)	0.875	0.925	cagGGCTgagacagccgggtc
	V$MAZF	Myc associated zinc fingers	V$MAZR.01	0.88	4638 - 4650	(+)	1.000	0.889	ggtggtGGGGaca
	V$MZF1	Myeloid zinc finger 1 factors	V$MZF1.01	0.99	4642 - 4652	(+)	1.000	1.000	gtGGGGacaca
	V$HAND	Twist subfamily of class B bHLH transcription factors	V$HEN1.01	0.82	4643 - 4663	(-)	1.000	0.856	ctatgctCAGCtgtgtcccca
	V$HAND	Twist subfamily of class B bHLH transcription factors	V$HEN1.01	0.82	4644 - 4664	(+)	1.000	0.874	ggggacaCAGCtgagcatagg

V$EVI1	EVI1-myleoid transforming protein	V$MEL1.02	0.99	4657 - 4673	(+)	1.000	0.993	agcatagGATGagcccc
V$HOXF	Paralog hox genes 1-8 from the four hox clusters A, B, C, D	V$HOXB3.01	0.84	4687 - 4705	(-)	0.867	0.882	tctgcTCATtagcaaagcc
V$HOXF	Paralog hox genes 1-8 from the four hox clusters A, B, C, D	V$HOXB3.01	0.84	4690 - 4708	(+)	1.000	0.880	tttgcTAATgagcagaagt
V$OCT1	Octamer binding protein	V$POU3F3.01	0.81	4701 - 4717	(-)	0.750	0.812	tctatCCATacttctgc
V$OCT1	Octamer binding protein	V$POU3F3.01	0.81	4706 - 4722	(+)	0.761	0.817	agtatGGATagaaagca
V$IRFF	Interferon regulatory factors	V$IRF4.02	0.69	4712 - 4732	(+)	1.000	0.707	gataGAAAgcaaccccagggc
V$HAND	Twist subfamily of class B bHLH transcription factors	V$HEN1.02	0.81	4733 - 4753	(-)	1.000	0.857	agagctgcaGCTGggaccgga
V$HAND	Twist subfamily of class B bHLH transcription factors	V$HEN1.01	0.82	4734 - 4754	(+)	1.000	0.905	ccggtccCAGCtgcagctctt
V$MYOD	Myoblast determining factors	V$MYOGENIN.02	0.90	4735 - 4751	(+)	1.000	0.931	cggtccCAGCtgcagct
V$KLFS	Krueppel like transcription factors	V$GKLF.02	0.96	4760 - 4778	(-)	1.000	0.961	tttcaccAAAGgacacgac
V$NBRE	NGFI-B response elements, nur subfamily of nuclear receptors	V$NBRE.01	0.86	4760 - 4774	(-)	1.000	0.864	accaAAGGacacgac
V$FKHD	Fork head domain factors	V$XFD2.01	0.89	4776 - 4792	(+)	1.000	0.894	aaacttTAAAcagtcgc
V$IRFF	Interferon regulatory factors	V$IRF4.01	0.94	4793 - 4813	(-)	1.000	0.950	aaaagaaagaGAAAaaaaaag
V$EVI1	EVI1-myleoid transforming protein	V$EVI1.02	0.83	4800 - 4816	(-)	1.000	0.851	cagaaAAGAaagagaaa
V$HOXF	Paralog hox genes 1-8 from the four hox clusters A, B, C, D	V$NANOG.01	0.94	4815 - 4833	(-)	1.000	0.949	tgaaattAATGgcaagcca
V$HOXF	Paralog hox genes 1-8 from the four hox clusters A, B, C, D	V$HOXB8.01	0.83	4817 - 4835	(+)	1.000	0.880	gcttgccATTAatttcaaa
V$HOXF	Paralog hox genes 1-8 from the four hox clusters A, B, C, D	V$HOXB8.01	0.83	4818 - 4836	(-)	1.000	0.893	gtttgaaATTAatggcaag
V$OCT1	Octamer binding protein	V$OCT1.05	0.89	4852 - 4868	(-)	0.950	0.964	agCATCtcattactaa
V$ETSF	Human and murine ETS1 factors	V$ELF2.01	0.90	4864 - 4884	(+)	1.000	0.910	atgcttcaGGAAggctttaat
V$HOXF	Paralog hox genes 1-8 from the four hox clusters A, B, C, D	V$HOXD8.01	0.81	4873 - 4891	(-)	1.000	0.815	gcagccaATTAaagccttc
V$HOXF	Paralog hox genes 1-8 from the four hox clusters A, B, C, D	V$HOXD1.01	0.87	4876 - 4894	(+)	1.000	0.870	ggctttAATTggctgcaga
V$RBIT	Regulator of B-Cell IgH transcription	V$BRIGHT.01	0.92	4877 - 4889	(-)	1.000	0.938	agccaATTAaagc
V$NEUR	NeuroD, Beta2, HLH domain	V$NEUROG.01	0.92	4888 - 4900	(-)	1.000	0.932	cctCCATctgcag
V$OCT1	Octamer binding protein	V$POU3F3.01	0.81	4936 - 4952	(+)	1.000	0.815	gggctGCATatcctcac
V$EGRF	EGR/nerve growth factor induced protein C & related factors	V$CKROX.01	0.88	4971 - 4987	(-)	1.000	0.939	atgtGGGAgggacccag
V$HOXF	Paralog hox genes 1-8 from the four hox clusters A, B, C, D	V$NANOG.01	0.94	4977 - 4995	(-)	1.000	0.949	gggcatgAATGtgggaggg
V$GCMF	Chorion-specific transcription factors with a GCM DNA binding domain	V$GCM1.01	0.85	4979 - 4989	(+)	0.789	0.866	ctCCCAcattc
V$NOLF	Neuron-specific-olfactory factor	V$OLF1.01	0.82	4992 - 5014	(-)	1.000	0.836	gcatatTCCCtctggtctggggc
V$MZF1	Myeloid zinc finger 1 factors	V$MZF1.01	0.99	5016 - 5026	(-)	1.000	1.000	gtGGGGaacag
V$MYOD	Myoblast determining factors	V$MYOGENIN.02	0.90	5040 - 5056	(-)	1.000	0.940	gctcaaCAGCtgcaaga
V$MYBL	Cellular and viral myb-like transcriptional regulators	V$CMYB.01	0.90	5043 - 5055	(-)	1.000	0.903	ctCAACagctgca
V$GREF	Glucocorticoid responsive and related elements	V$GRE.01	0.85	5069 - 5087	(-)	0.833	0.851	gctgagccccgtGTGCtgg
V$GREF	Glucocorticoid responsive and related elements	V$GRE.02	0.82	5069 - 5087	(+)	0.890	0.840	ccaGCACacggggctcagc
V$AHRR	AHR-arnt heterodimers and AHR-related factors	V$AHRARNT.03	0.95	5077 - 5101	(-)	1.000	0.958	acgtcgccagGCGTgctgagccccg
V$AP1R	MAF and AP1 related factors	V$MAFA.01	0.92	5125 - 5145	(-)	1.000	0.937	ctgTGCTgatggagctgtacc
V$HAND	Twist subfamily of class B bHLH	V$HEN1.01	0.82	5135 - 5155	(-)	1.000	0.880	gtggcccCAGCtgtgctgatg

	transcription factors								
V$HAND	Twist subfamily of class B bHLH transcription factors	V$HEN1.01	0.82	5136 - 5156	(+)	1.000	0.832	atcagcaCAGCtggggccaca
V$GREF	Glucocorticoid responsive and related elements	V$GRE.01	0.85	5137 - 5155	(-)	0.833	0.878	gtggccccagctGTGCtga
V$MYOD	Myoblast determining factors	V$MYOGENIN.02	0.90	5138 - 5154	(-)	1.000	0.947	tggcccCAGCtgtgctg
V$FKHD	Fork head domain factors	V$FKHRL1.01	0.83	5190 - 5206	(+)	1.000	0.865	aacccgaaAACAagagc
V$NOLF	Neuron-specific-olfactory factor	V$OLF1.02	0.88	5233 - 5255	(-)	1.000	0.913	gccctcTCCCccaggccctgcct
V$ETSF	Human and murine ETS1 factors	V$GABP.01	0.86	5275 - 5295	(-)	1.000	0.909	gccttcctGGAAgggctgtgt
V$STAT	Signal transducer and activator of transcription	V$STAT3.02	0.94	5278 - 5296	(-)	1.000	0.972	tgccTTCCtggaagggctg
V$BCL6	POZ domain zinc finger expressed in B-Cells	V$BCL6.01	0.76	5279 - 5295	(-)	1.000	0.850	gccTTCCtggaagggct
V$STAT	Signal transducer and activator of transcription	V$STAT3.02	0.94	5280 - 5298	(+)	1.000	0.984	gcccTTCCaggaaggcaca
V$ETSF	Human and murine ETS1 factors	V$ELF2.01	0.90	5281 - 5301	(+)	1.000	0.922	cccttccaGGAAggcacagcg
V$KLFS	Krueppel like transcription factors	V$EKLF.01	0.89	5298 - 5316	(+)	1.000	0.925	agcgctgtcaGGGTgcagc
V$AHRR	AHR-arnt heterodimers and AHR-related factors	V$AHRARNT.01	0.92	5306 - 5330	(-)	1.000	0.923	gtggggctgagCGTGctgcaccctg
V$KLFS	Krueppel like transcription factors	V$KLF6.01	0.92	5316 - 5334	(-)	1.000	0.926	catggtGGGGctgagcgtg
V$PAX5	PAX-5 B-cell-specific activator protein	V$PAX5.01	0.79	5322 - 5350	(-)	1.000	0.804	cccggcCGCAcagctgcatggtggggctg
V$MYOD	Myoblast determining factors	V$E47.01	0.92	5329 - 5345	(+)	1.000	0.937	accatGCAGctgtgcgg
V$NFKB	Nuclear factor kappa B/c-rel	V$NFKAPPAB50.01	0.83	5347 - 5359	(+)	0.750	0.876	cggGGCAtcccca
V$NFKB	Nuclear factor kappa B/c-rel	V$NFKAPPAB50.01	0.83	5348 - 5360	(-)	1.000	0.982	ttgGGGAtgcccc
V$IRFF	Interferon regulatory factors	V$ISRE.01	0.81	5379 - 5399	(+)	1.000	0.834	tctccaatcaGAAActgaagc
V$IRFF	Interferon regulatory factors	V$IRF7.01	0.86	5385 - 5405	(+)	1.000	0.912	atcaGAAActgaagctgaggg
V$EGRF	EGR/nerve growth factor induced protein C & related factors	V$EGR3.01	0.77	5401 - 5417	(-)	1.000	0.818	gccgGCGTgggcccctc
V$IRFF	Interferon regulatory factors	V$IRF1.01	0.87	5417 - 5437	(+)	1.000	0.888	ccaaaaaaagGAAAcgaaaca
V$ETSF	Human and murine ETS1 factors	V$SPIB.01	0.88	5418 - 5438	(+)	1.000	0.903	caaaaaaaGGAAacgaaacag
V$IRFF	Interferon regulatory factors	V$ISRE.01	0.81	5422 - 5442	(+)	1.000	0.822	aaaaggaaacGAAAcagtctc
V$IRFF	Interferon regulatory factors	V$IRF4.02	0.69	5428 - 5448	(+)	1.000	0.701	aaacGAAAcagtctccagaaa
V$BRAC	Brachyury gene, mesoderm developmental factor	V$BRACH.01	0.66	5446 - 5466	(+)	0.750	0.687	aaagcactgACGTgtgaagca
V$PAX5	PAX-5 B-cell-specific activator protein	V$PAX5.03	0.80	5509 - 5537	(-)	0.842	0.817	gcaaaGCTCagtctggcaaagccccaacc
V$HAND	Twist subfamily of class B bHLH transcription factors	V$TH1E47.01	0.93	5513 - 5533	(+)	1.000	0.932	ggggctttgCCAGactgagct
V$KLFS	Krueppel like transcription factors	V$BKLF.01	0.95	5542 - 5560	(+)	1.000	0.957	gctcgGGGTgggtgcccac
V$EGRF	EGR/nerve growth factor induced protein C & related factors	V$WT1.01	0.92	5545 - 5561	(+)	1.000	0.933	cggggTGGGtgcccacg
V$HAND	Twist subfamily of class B bHLH transcription factors	V$SCX.01	0.91	5561 - 5581	(-)	1.000	0.921	cccatcccatgTGCCcaggcc
V$HESF	Vertebrate homologues of enhancer of split complex	V$DEC2.01	0.96	5565 - 5579	(+)	0.903	0.969	tgggcaCATGggatg
V$EGRF	EGR/nerve growth factor induced protein C & related factors	V$WT1.01	0.92	5599 - 5615	(-)	1.000	0.930	gggtgTGGGtgtgcaag
V$EGRF	EGR/nerve growth factor induced protein C & related factors	V$EGR2.01	0.79	5601 - 5617	(-)	0.777	0.811	tgggGTGTgggtgtgca
V$KLFS	Krueppel like transcription factors	V$BKLF.01	0.95	5602 - 5620	(-)	1.000	0.963	tgttgGGGTgtgggtgtgc
V$KLFS	Krueppel like transcription factors	V$KKLF.01	0.91	5607 - 5625	(-)	1.000	0.950	tgaagtgttGGGGtgtggg

V$ETSF	Human and murine ETS1 factors	V$FLI.01	0.81	5613 - 5633	(-)	0.750	0.845	gcatcaCCTGaagtgttgggg
V$OCT1	Octamer binding protein	V$OCT3_4.01	0.81	5627 - 5643	(+)	1.000	0.879	gtgATGCtggtgcagat
V$NEUR	NeuroD, Beta2, HLH domain	V$NEUROD1.01	0.83	5636 - 5648	(-)	1.000	0.887	caccCATCtgcac
V$HESF	Vertebrate homologues of enhancer of split complex	V$HES1.01	0.92	5639 - 5653	(+)	1.000	0.937	cagatggGTGCcccc
V$EGRF	EGR/nerve growth factor induced protein C & related factors	V$WT1.01	0.92	5658 - 5674	(-)	1.000	1.000	atgcgTGGGggggtcag
V$KLFS	Krueppel like transcription factors	V$KKLF.01	0.91	5658 - 5676	(-)	1.000	0.911	ccatgcgtgGGGGggtcag
V$MAZF	Myc associated zinc fingers	V$MAZR.01	0.88	5659 - 5671	(-)	1.000	0.928	cgtgggGGGGtca
V$EGRF	EGR/nerve growth factor induced protein C & related factors	V$NGFIC.01	0.80	5660 - 5676	(-)	1.000	0.972	ccatGCGTgggggggtc
V$PAX5	PAX-5 B-cell-specific activator protein	V$PAX5.03	0.80	5699 - 5727	(+)	0.789	0.830	ccgtgTCTCtgtccccatgtgcc accct
V$HESF	Vertebrate homologues of enhancer of split complex	V$DEC2.01	0.96	5709 - 5723	(-)	0.903	0.969	gtggcaCATGgggac
V$MYOD	Myoblast determining factors	V$MYOD.01	0.88	5709 - 5725	(-)	1.000	0.888	gggtGGCAcatggggac
V$NEUR	NeuroD, Beta2, HLH domain	V$NEUROG.01	0.92	5710 - 5722	(+)	1.000	0.922	tccCCATgtgcca
V$KLFS	Krueppel like transcription factors	V$KKLF.01	0.91	5717 - 5735	(-)	1.000	0.911	gagcgggcaGGGGtggcac
V$EGRF	EGR/nerve growth factor induced protein C & related factors	V$NGFIC.01	0.80	5732 - 5748	(-)	1.000	0.813	tgacGCGTggtgggagc
V$WHNF	Winged helix binding sites	V$WHN.01	0.95	5739 - 5749	(-)	1.000	0.954	ttgACGCgtgg
V$WHNF	Winged helix binding sites	V$WHN.01	0.95	5787 - 5797	(-)	1.000	0.965	tggACGCtgga
V$NFKB	Nuclear factor kappa B/c-rel	V$NFKAPPAB50.01	0.83	5805 - 5817	(+)	0.750	0.865	ctgGAGAtgccca
V$OCT1	Octamer binding protein	V$OCT3_4.02	0.88	5812 - 5828	(-)	1.000	0.896	gctctGCATtttgggca
V$STAT	Signal transducer and activator of transcription	V$STAT.01	0.87	5827 - 5845	(-)	1.000	0.960	cagattcagGGAAacctgc
V$XBBF	X-box binding factors	V$MIF1.01	0.76	5828 - 5846	(-)	0.850	0.812	tcagattcaggGAAAcctg
V$STAT	Signal transducer and activator of transcription	V$STAT5.01	0.89	5829 - 5847	(+)	0.945	0.893	aggtTTCCctgaatctgag
V$GATA	GATA binding factors	V$GATA3.02	0.91	5845 - 5857	(+)	1.000	0.915	gagAGATgaaatg
V$NFKB	Nuclear factor kappa B/c-rel	V$CREL.01	0.91	5862 - 5874	(+)	1.000	0.931	tatgggtgTTCCc
V$NFKB	Nuclear factor kappa B/c-rel	V$NFKAPPAB.01	0.89	5864 - 5876	(-)	1.000	0.894	agGGGAacaccca
V$IKRS	Ikaros zinc finger family	V$IK3.01	0.84	5866 - 5878	(-)	1.000	0.876	gcaggGGAAcacc
V$XBBF	X-box binding factors	V$RFX1.01	0.89	5866 - 5884	(-)	0.881	0.913	tccgccgcaggGGAAcacc
V$MYOD	Myoblast determining factors	V$MYOGENIN.02	0.90	5883 - 5899	(+)	1.000	0.957	gagcccCAGCtgtagga
V$MYOD	Myoblast determining factors	V$MYOGENIN.02	0.90	5884 - 5900	(-)	1.000	0.906	ttcctaCAGCtggggct
V$NEUR	NeuroD, Beta2, HLH domain	V$NEUROG.01	0.92	5885 - 5897	(+)	0.875	0.921	gccCCAGctgtag
V$ETSF	Human and murine ETS1 factors	V$PEA3.01	0.94	5889 - 5909	(+)	1.000	0.943	cagctgtAGGAagctcagagc
V$HOXF	Paralog hox genes 1-8 from the four hox clusters A, B, C, D	V$HOXC8.01	0.85	5912 - 5930	(-)	0.779	0.851	ctctttaATTGctgtgtga
V$HOXF	Paralog hox genes 1-8 from the four hox clusters A, B, C, D	V$HOXB8.01	0.83	5915 - 5933	(+)	1.000	0.926	cacagcaATTAaagaggaa
V$RBIT	Regulator of B-Cell IgH transcription	V$BRIGHT.01	0.92	5917 - 5929	(+)	1.000	0.938	cagcaATTAaaga
V$ETSF	Human and murine ETS1 factors	V$SPIB.01	0.88	5922 - 5942	(+)	1.000	0.947	attaaagaGGAAttaaattaa
V$HOXF	Paralog hox genes 1-8 from the four hox clusters A, B, C, D	V$HOXA6.01	0.82	5923 - 5941	(-)	1.000	0.820	taatttAATTcctctttaa
V$HOXF	Paralog hox genes 1-8 from the four hox clusters A, B, C, D	V$HOXB8.01	0.83	5926 - 5944	(+)	1.000	0.868	aagaggaATTAaattaaat
V$HOXF	Paralog hox genes 1-8 from the four hox clusters A, B, C, D	V$HOXC4.01	0.84	5928 - 5946	(-)	1.000	0.860	tgatttAATTtaattcctc

V$RBIT	Regulator of B-Cell IgH transcription	V$BRIGHT.01	0.92	5928 - 5940	(+)	1.000	0.953	gaggaATTAaatt
V$HOXF	Paralog hox genes 1-8 from the four hox clusters A, B, C, D	V$HOXD3.01	0.85	5931 - 5949	(+)	1.000	0.871	gaattaAATTaaatcaata
V$RBIT	Regulator of B-Cell IgH transcription	V$BRIGHT.01	0.92	5933 - 5945	(+)	1.000	0.946	attaaATTAaatc
V$HOXC	HOX - PBX complexes	V$PBX_HOX A9.01	0.79	5934 - 5950	(-)	1.000	0.811	ttatTGATttaatttaa
V$GFI1	Growth factor independence transcriptional repressor	V$GFI1B.01	0.86	5939 - 5953	(+)	1.000	0.867	ttaAATCaataaatg
V$HOXC	HOX - PBX complexes	V$PBX_HOX A9.01	0.79	5948 - 5964	(+)	0.750	0.793	taaaTGTTtaggcggg
V$AP1R	MAF and AP1 related factors	V$MARE.02	0.88	5954 - 5974	(-)	0.791	0.883	ggcAGCTgagcccgcctaaaa
V$HAND	Twist subfamily of class B bHLH transcription factors	V$HEN1.01	0.82	5959 - 5979	(-)	1.000	0.832	gtgctggCAGCtgagcccgcc
V$HAND	Twist subfamily of class B bHLH transcription factors	V$HEN1.01	0.82	5960 - 5980	(+)	1.000	0.877	gcgggctCAGCtgccagcacc
V$MYOD	Myoblast determining factors	V$MYOD.01	0.88	5975 - 5991	(-)	0.827	0.889	ttcgGTCAggtggtgct
V$PAX5	PAX-5 B-cell-specific activator protein	V$PAX5.03	0.80	6009 - 6037	(-)	0.894	0.844	ttttgCCACagccatgcaaatgct ctcct
V$OCT1	Octamer binding protein	V$OCT1.02	0.85	6011 - 6027	(-)	1.000	0.981	gccATGCaaatgctctc
V$OCT1	Octamer binding protein	V$OCT3_4.02	0.88	6016 - 6032	(+)	1.000	0.928	catttGCATggctgtgg
V$CEBP	Ccaat/Enhancer Binding Protein	V$CEBP.02	0.92	6024 - 6038	(+)	1.000	0.923	tggctgtgGCAAaac
V$FKHD	Fork head domain factors	V$ILF1.01	0.98	6028 - 6044	(+)	1.000	0.980	tgtggcaaAACAgcaac
V$MYBL	Cellular and viral myb-like transcriptional regulators	V$CMYB.01	0.90	6039 - 6051	(+)	1.000	0.999	agCAACcgcctgt
V$MYBL	Cellular and viral myb-like transcriptional regulators	V$CMYB.01	0.90	6043 - 6055	(-)	1.000	0.944	caCAACaggcggt
V$GATA	GATA binding factors	V$GATA1.04	0.91	6066 - 6078	(-)	1.000	0.993	tccaGATAacacc
V$NFAT	Nuclear factor of activated T-cells	V$NFAT5.01	0.83	6073 - 6091	(+)	1.000	0.850	tctGGAAatgtacgcagcc
V$FKHD	Fork head domain factors	V$ILF1.01	0.98	6095 - 6111	(+)	1.000	0.989	gaggggtaAACAgctcc
V$FKHD	Fork head domain factors	V$HNF3.01	0.98	6137 - 6153	(+)	1.000	0.982	aggttgtAAACaggctg
V$FKHD	Fork head domain factors	V$FREAC7.01	0.96	6151 - 6167	(+)	1.000	0.984	ctgacaTAAAcaccttt
V$BRAC	Brachyury gene, mesoderm developmental factor	V$BRACH.01	0.66	6154 - 6174	(-)	1.000	0.689	acggcacaaAGGTgttatgt
V$FKHD	Fork head domain factors	V$FHXB.01	0.83	6172 - 6188	(+)	0.818	0.840	cgtggaAAAAtatttat
V$NFAT	Nuclear factor of activated T-cells	V$NFAT.01	0.95	6172 - 6190	(+)	1.000	0.972	cgtGGAAaaatatttatca
V$GATA	GATA binding factors	V$GATA2.01	0.92	6181 - 6193	(-)	1.000	0.922	aggtGATAaatat
V$EVI1	EVI1-myleoid transforming protein	V$EVI1.04	0.73	6185 - 6201	(-)	1.000	0.737	tatatttgaggtGATAa
V$HOXF	Paralog hox genes 1-8 from the four hox clusters A, B, C, D	V$HOXC4.01	0.84	6200 - 6218	(-)	0.754	0.847	ttatttTATTaacctgcta
V$HOXF	Paralog hox genes 1-8 from the four hox clusters A, B, C, D	V$HOXA4.01	0.81	6203 - 6221	(+)	0.755	0.824	caggttAATAaaataaaac
V$OCT1	Octamer binding protein	V$OCT1.03	0.85	6203 - 6219	(-)	1.000	0.866	tttattttATTAacctg
V$SATB	Special AT-rich sequence binding protein	V$SATB1.01	0.94	6206 - 6220	(+)	1.000	0.946	gttAATAaaataaaa
V$NOLF	Neuron-specific-olfactory factor	V$OLF1.01	0.82	6216 - 6238	(+)	1.000	0.833	taaaacTCCCaacggagctaca c
V$MYBL	Cellular and viral myb-like transcriptional regulators	V$VMYB.02	0.90	6223 - 6235	(+)	1.000	0.980	cccAACGgagcta
V$MYOD	Myoblast determining factors	V$E47.01	0.92	6233 - 6249	(-)	1.000	0.966	ccaaaGCAGgtgtgtag
V$HAND	Twist subfamily of class B bHLH transcription factors	V$SCX.01	0.91	6254 - 6274	(-)	0.882	0.928	cagaaaacaagTGTCtgcttc
V$FKHD	Fork head domain factors	V$ILF1.01	0.98	6261 - 6277	(-)	1.000	0.980	aagcagaaAACAagtgt
V$ETSF	Human and murine ETS1 factors	V$PEA3.01	0.94	6283 - 6303	(+)	1.000	0.953	tggctgtAGGAagccatgttt
V$NFAT	Nuclear factor of activated T-cells	V$NFAT5.01	0.83	6290 - 6308	(-)	1.000	0.843	tcgGGAAacatggcttcct

V$FKHD	Fork head domain factors	V$FREAC2.01	0.84	6356 - 6372	(-)	1.000	0.886	ggggctTAAAcaggggc
V$XBBF	X-box binding factors	V$MIF1.01	0.76	6383 - 6401	(-)	0.800	0.786	cctgctccgagGCCAcaag
V$HESF	Vertebrate homologues of enhancer of split complex	V$HES1.01	0.92	6413 - 6427	(-)	1.000	0.937	cagcccgGTGCcacc
V$GREF	Glucocorticoid responsive and related elements	V$ARE.02	0.89	6448 - 6466	(+)	0.897	0.890	gggtgtctcactGTGCtct
V$NOLF	Neuron-specific-olfactory factor	V$OLF1.01	0.82	6457 - 6479	(-)	1.000	0.824	aaccccTCCCcaaagagcacag t
V$EGRF	EGR/nerve growth factor induced protein C & related factors	V$CKROX.01	0.88	6466 - 6482	(+)	1.000	0.895	tttgGGGAgggttgga
V$KLFS	Krueppel like transcription factors	V$KLF6.01	0.92	6468 - 6486	(+)	1.000	0.940	tggggaGGGGttggagccc
V$MAZF	Myc associated zinc fingers	V$MAZR.01	0.88	6468 - 6480	(+)	1.000	0.910	tggggaGGGGttg
V$AP1R	MAF and AP1 related factors	V$MAFA.01	0.92	6485 - 6505	(-)	1.000	0.927	gtgTGCTgtgattgccccagg
V$GFI1	Growth factor independence transcriptional repressor	V$GFI1.02	0.90	6490 - 6504	(+)	1.000	0.935	ggcAATCacagcaca
V$KLFS	Krueppel like transcription factors	V$KKLF.01	0.91	6506 - 6524	(+)	1.000	0.950	acaggaggtGGGGggatgc
V$EGRF	EGR/nerve growth factor induced protein C & related factors	V$WT1.01	0.92	6509 - 6525	(+)	1.000	0.926	ggaggTGGGggatgca
V$MZF1	Myeloid zinc finger 1 factors	V$MZF1.01	0.99	6515 - 6525	(+)	1.000	0.995	ggGGGGatgca
V$MYOD	Myoblast determining factors	V$MYF5.01	0.90	6522 - 6538	(+)	1.000	0.931	tgcagcCAGCagctgcc
V$MYOD	Myoblast determining factors	V$MYOGENI N.02	0.90	6525 - 6541	(+)	1.000	0.912	agccagCAGCtgccctg
V$MYOD	Myoblast determining factors	V$MYF5.01	0.90	6526 - 6542	(-)	1.000	0.960	gcagggCAGCtgctggc
V$EGRF	EGR/nerve growth factor induced protein C & related factors	V$EGR3.01	0.77	6555 - 6571	(+)	0.753	0.777	ccatCCGTgggctttca
V$HAND	Twist subfamily of class B bHLH transcription factors	V$TAL1BETA E47.01	0.87	6563 - 6583	(+)	1.000	0.876	gggctttCAGAtggccttccc
V$NEUR	NeuroD, Beta2, HLH domain	V$NEUROG.01	0.92	6567 - 6579	(-)	1.000	0.943	aggCCATctgaaa
V$IKRS	Ikaros zinc finger family	V$IK1.01	0.92	6575 - 6587	(-)	1.000	0.925	agatGGGAaggcc
V$AP1R	MAF and AP1 related factors	V$MAFA.01	0.92	6586 - 6606	(-)	0.904	0.921	ccaTGCAgaggctgcagagag
V$AP1R	MAF and AP1 related factors	V$MAFA.01	0.92	6588 - 6608	(+)	0.904	0.944	ctcTGCAgcctctgcatgggc
V$OCT1	Octamer binding protein	V$OCT3_4.02	0.88	6596 - 6612	(+)	1.000	0.891	cctctGCATgggctgag
V$KLFS	Krueppel like transcription factors	V$KLF6.01	0.92	6651 - 6669	(+)	1.000	0.940	tttggaGGGGcagcatggg
V$MAZF	Myc associated zinc fingers	V$MAZ.01	0.90	6651 - 6663	(+)	1.000	0.900	tttgGAGGggcag
V$HESF	Vertebrate homologues of enhancer of split complex	V$HES1.02	0.87	6685 - 6699	(-)	1.000	0.998	gcggCACGtggcggt
V$MYOD	Myoblast determining factors	V$MYOD.01	0.88	6685 - 6701	(-)	1.000	0.891	tggcGGCAcgtggcggt
V$HESF	Vertebrate homologues of enhancer of split complex	V$HES1.01	0.92	6686 - 6700	(+)	1.000	0.973	ccgccacGTGCcgcc
V$PTF1	Pancreas transcription factor 1, heterotrimeric transcription factor	V$PTF1.01	0.76	6688 - 6708	(+)	0.809	0.790	gccaCGTGccgccagccccac
V$HAND	Twist subfamily of class B bHLH transcription factors	V$HEN1.01	0.82	6700 - 6720	(-)	1.000	0.826	gcaggctCAGCtgtggggctg
V$MYOD	Myoblast determining factors	V$MYOGENI N.02	0.90	6702 - 6718	(+)	1.000	0.933	gccccaCAGCtgagcct
V$AP1R	MAF and AP1 related factors	V$MARE.02	0.88	6706 - 6726	(+)	0.791	0.919	cacAGCTgagcctgcactctc
V$IKRS	Ikaros zinc finger family	V$LYF1.01	0.98	6720 - 6732	(-)	1.000	1.000	atcTGGGagagtg
V$AP1R	MAF and AP1 related factors	V$MARE.01	0.97	6727 - 6747	(+)	1.000	0.995	ccagatgtGCTGaccgcagcc
V$HAML	Human acute myelogenous leukemia factors	V$AML1.01	0.93	6733 - 6747	(-)	0.909	0.956	ggctGCGGtcagcac
V$XBBF	X-box binding factors	V$RFX1.02	0.90	6743 - 6761	(+)	1.000	0.918	cagccacggggGCAAcagt
V$PRDF	Positive regulatory domain I binding factor	V$BLIMP1.01	0.81	6752 - 6770	(-)	1.000	0.876	gcaagaGAAActgttgccc
V$IRFF	Interferon regulatory factors	V$ISRE.01	0.81	6754 - 6774	(-)	1.000	0.905	tttagcaagaGAAActgttgc
V$DICE	Downstream Immunoglobulin	V$DICE.01	0.80	6782 - 6796	(-)	0.891	0.820	tgttTTCTtcccggc

		Control Element, critical for B cell activity and specificity							
	V$PTF1	Pancreas transcription factor 1, heterotrimeric transcription factor	V$PTF1.01	0.76	6782 - 6802	(-)	0.809	0.857	gccaCGTGttttcttcccggc
	V$FKHD	Fork head domain factors	V$ILF1.01	0.98	6785 - 6801	(+)	1.000	0.991	gggaagaaAACAcgtgg
	V$HAND	Twist subfamily of class B bHLH transcription factors	V$SCX.01	0.91	6788 - 6808	(+)	1.000	0.960	aagaaaacacgTGGCaacttc
	V$XBBF	X-box binding factors	V$RFX1.01	0.89	6790 - 6808	(+)	1.000	0.898	gaaaacacgtgGCAActtc
	V$PTF1	Pancreas transcription factor 1, heterotrimeric transcription factor	V$PTF1.01	0.76	6806 - 6826	(-)	0.857	0.850	tccaGCTGctgtttggccgaa
	V$MYOD	Myoblast determining factors	V$MYF5.01	0.90	6810 - 6826	(+)	1.000	0.950	gccaaaCAGCagctgga
	V$MYOD	Myoblast determining factors	V$MYF5.01	0.90	6813 - 6829	(+)	1.000	0.957	aaacagCAGCtggagga
	V$MYOD	Myoblast determining factors	V$MYOGENIN.02	0.90	6814 - 6830	(-)	1.000	0.945	gtcctcCAGCtgctgtt
	V$NFKB	Nuclear factor kappa B/c-rel	V$CREL.01	0.91	6831 - 6843	(-)	1.000	0.924	cacggctaTTCCt
	V$XBBF	X-box binding factors	V$MIF1.01	0.76	6838 - 6856	(-)	0.800	0.768	agcgtgccgtgGCCAcggc
	V$AHRR	AHR-arnt heterodimers and AHR-related factors	V$AHRARNT.03	0.95	6841 - 6865	(-)	1.000	0.955	aggaagcagaGCGTgccgtggccac
	V$ETSF	Human and murine ETS1 factors	V$PEA3.01	0.94	6852 - 6872	(-)	1.000	0.940	tgtgccgAGGAagcagagcgt
	V$AP1R	MAF and AP1 related factors	V$MARE.02	0.88	6853 - 6873	(-)	0.875	0.905	ttgTGCCgaggaagcagagcg
	V$OCT1	Octamer binding protein	V$OCT1.05	0.89	6873 - 6889	(+)	0.900	0.901	aaCATTccagtatgtgg
	V$ETSF	Human and murine ETS1 factors	V$ETS1.01	0.92	6930 - 6950	(+)	1.000	0.951	caggagcaGGAAatgctgatt
	V$OCT1	Octamer binding protein	V$OCT1.06	0.81	6939 - 6955	(+)	0.750	0.814	gaaatgctGATTtgggc
	V$STAT	Signal transducer and activator of transcription	V$STAT3.02	0.94	6982 - 7000	(+)	1.000	0.966	aggcTTCCagggagctggt
	V$HAML	Human acute myelogenous leukemia factors	V$AML3.01	0.84	7015 - 7029	(+)	1.000	0.874	agctGTGGcttgggg
	V$MZF1	Myeloid zinc finger 1 factors	V$MZF1.02	0.99	7024 - 7034	(+)	1.000	0.994	ttGGGGaaatg
	V$KLFS	Krueppel like transcription factors	V$KKLF.01	0.91	7028 - 7046	(+)	1.000	0.917	ggaaatgatGGGGaggga
	V$EGRF	EGR/nerve growth factor induced protein C & related factors	V$CKROX.01	0.88	7034 - 7050	(+)	1.000	0.882	gatgGGGAggggattgc
	V$MAZF	Myc associated zinc fingers	V$MAZ.01	0.90	7036 - 7048	(+)	1.000	0.959	tgggGAGGggatt
	V$MZF1	Myeloid zinc finger 1 factors	V$MZF1.02	0.99	7040 - 7050	(+)	1.000	0.992	gaGGGGattgc
	V$NFKB	Nuclear factor kappa B/c-rel	V$NFKAPPAB50.01	0.83	7040 - 7052	(+)	1.000	0.882	gagGGGAttgcca
	V$KLFS	Krueppel like transcription factors	V$BKLF.01	0.95	7049 - 7067	(-)	1.000	0.951	ctgcaGGGTgggcagtggc
	V$AP1R	MAF and AP1 related factors	V$MARE.02	0.88	7051 - 7071	(+)	0.875	0.929	cacTGCCcaccctgcagagca
	V$AP1R	MAF and AP1 related factors	V$MAFA.01	0.92	7060 - 7080	(+)	0.904	0.929	cccTGCAgagcaggctctggt
	V$PAX5	PAX-5 B-cell-specific activator protein	V$PAX5.01	0.79	7080 - 7108	(+)	0.904	0.790	tcccatCTCActgcagggcaccagggcgt
	V$STAT	Signal transducer and activator of transcription	V$STAT5.01	0.89	7117 - 7135	(-)	1.000	0.930	atgtTTCTgtgaattgctg
	V$STAT	Signal transducer and activator of transcription	V$STAT5.01	0.89	7119 - 7137	(+)	0.845	0.933	gcaaTTCAcagaaacattg
	V$EVI1	EVI1-myleoid transforming protein	V$MEL1.03	0.95	7154 - 7170	(-)	1.000	0.993	tgatgaaGATGttgaac
	V$EVI1	EVI1-myleoid transforming protein	V$MEL1.02	0.99	7172 - 7188	(+)	1.000	0.994	tgacctgGATGaggga
	V$GCMF	Chorion-specific transcription factors with a GCM DNA binding domain	V$GCM1.01	0.85	7178 - 7188	(-)	1.000	0.942	tcCCCTcatcc
	V$MZF1	Myeloid zinc finger 1 factors	V$MZF1.01	0.99	7182 - 7192	(+)	1.000	0.991	gaGGGGacagc
	V$GATA	GATA binding factors	V$GATA3.02	0.91	7211 - 7223	(+)	0.875	0.924	cacTGATcatatg
	V$AP1R	MAF and AP1 related factors	V$BACH2.01	0.89	7233 - 7253	(-)	0.868	0.946	ttgtggTGCGtcagccactcc
	V$HOXF	Paralog hox genes 1-8 from the four hox clusters A, B, C, D	V$NANOG.01	0.94	7257 - 7275	(-)	1.000	0.956	cctgttgAATGgcagcaca
	V$ETSF	Human and murine ETS1 factors	V$CETS1P54.	0.92	7266 - 7286	(+)	0.901	0.923	attcaaCAGGacgtggacaga

		01						
V$ETSF	Human and murine ETS1 factors	V$PU1.01	0.89	7298 - 7318	(+)	1.000	0.919	gacagggaGGAAcccaatgag
V$HAND	Twist subfamily of class B bHLH transcription factors	V$TAL1_E2A.01	0.98	7338 - 7358	(-)	1.000	0.988	tccttccCAGGtgtagatcct
V$MYBL	Cellular and viral myb-like transcriptional regulators	V$VMYB.01	0.88	7358 - 7370	(+)	0.817	0.910	aatAACAgcatgc
V$HOXC	HOX - PBX complexes	V$HOX_PBX.01	0.81	7362 - 7378	(-)	1.000	0.825	gaacTGATgcatgctgt
V$OCT1	Octamer binding protein	V$POU3F3.01	0.81	7364 - 7380	(+)	1.000	0.820	agcatGCATcagttcag
V$KLFS	Krueppel like transcription factors	V$KLF6.01	0.92	7379 - 7397	(+)	1.000	0.934	aggttaGGGGctgagctgc
V$EVI1	EVI1-myleoid transforming protein	V$MEL1.02	0.99	7395 - 7411	(+)	1.000	0.992	tgctgcaGATGagctct
V$ETSF	Human and murine ETS1 factors	V$PDEF.01	0.93	7420 - 7440	(-)	1.000	0.978	gtccagcaGGATgctcaggtc
V$PAX5	PAX-5 B-cell-specific activator protein	V$PAX5.01	0.79	7446 - 7474	(-)	0.809	0.810	ccagggCACAccggtggctcacagccagc
V$XBBF	X-box binding factors	V$XBOX.01	0.90	7490 - 7508	(-)	0.875	0.920	gtgctCCCCaggataccac
V$EGRF	EGR/nerve growth factor induced protein C & related factors	V$NGFIC.01	0.80	7527 - 7543	(+)	0.762	0.826	caggGCGAgggaggtga
V$MAZF	Myc associated zinc fingers	V$MAZ.01	0.90	7529 - 7541	(+)	1.000	0.900	gggcGAGGgaggt
V$MAZF	Myc associated zinc fingers	V$MAZ.01	0.90	7533 - 7545	(+)	1.000	0.922	gaggGAGGtgagg
V$AP1R	MAF and AP1 related factors	V$TCF11MAFG.01	0.81	7535 - 7555	(+)	0.814	0.814	gggaggTGAGgctgcatttgg
V$PAX5	PAX-5 B-cell-specific activator protein	V$PAX5.03	0.80	7542 - 7570	(-)	1.000	0.835	actggGCACggtgctccaaatgcagcctc
V$DICE	Downstream Immunoglobulin Control Element, critical for B cell activity and specificity	V$DICE.01	0.80	7603 - 7617	(-)	1.000	0.800	ggctCTCTccagcag
V$KLFS	Krueppel like transcription factors	V$KLF6.01	0.92	7619 - 7637	(+)	1.000	0.930	agcagaGGGGctgcaatga
V$EVI1	EVI1-myleoid transforming protein	V$MEL1.02	0.99	7632 - 7648	(+)	1.000	0.995	caatgatGATGagggtc
V$GCMF	Chorion-specific transcription factors with a GCM DNA binding domain	V$GCM1.01	0.85	7638 - 7648	(-)	1.000	0.927	gaCCCTcatca
V$ETSF	Human and murine ETS1 factors	V$PU1.01	0.89	7661 - 7681	(+)	1.000	0.908	ctgtatgaGGAAaggctgagg
V$CEBP	Ccaat/Enhancer Binding Protein	V$CEBP.02	0.92	7662 - 7676	(+)	0.885	0.958	tgtatgagGAAAggc
V$NFAT	Nuclear factor of activated T-cells	V$NFAT.01	0.95	7666 - 7684	(+)	1.000	0.976	tgaGGAAaggctgagggac
V$GCMF	Chorion-specific transcription factors with a GCM DNA binding domain	V$GCM1.01	0.85	7674 - 7684	(-)	1.000	0.895	gtCCCTcagcc
V$EVI1	EVI1-myleoid transforming protein	V$EVI1.02	0.83	7701 - 7717	(+)	1.000	0.882	tggagAAGAgaagacta
V$PTF1	Pancreas transcription factor 1, heterotrimeric transcription factor	V$PTF1.01	0.76	7717 - 7737	(-)	0.761	0.789	tccaCTTGgctcctgccctgt
V$HAND	Twist subfamily of class B bHLH transcription factors	V$HAND2_E12.01	0.75	7722 - 7742	(-)	1.000	0.750	ccctatccactTGGCtcctgc
V$GATA	GATA binding factors	V$GATA1.01	0.96	7732 - 7744	(+)	1.000	0.966	agtgGATAgggcc
V$PAX5	PAX-5 B-cell-specific activator protein	V$PAX5.03	0.80	7740 - 7768	(-)	1.000	0.830	actggGCACtgctgaaaagagcccggccc
V$PAX5	PAX-5 B-cell-specific activator protein	V$PAX5.01	0.79	7742 - 7770	(-)	0.952	0.863	tcactgGGCActgctgaaaagagcccggc
V$ETSF	Human and murine ETS1 factors	V$SPIB.01	0.88	7790 - 7810	(+)	1.000	0.918	cacaaagtGGAAcataagaag
V$FKHD	Fork head domain factors	V$FHXB.01	0.83	7797 - 7813	(+)	1.000	0.836	tggaacATAAgaagttc
V$XBBF	X-box binding factors	V$MIF1.01	0.76	7807 - 7825	(-)	0.850	0.764	catgttcagatGGAActtc
V$NFAT	Nuclear factor of activated T-cells	V$NFAT.01	0.95	7824 - 7842	(+)	1.000	1.000	tgaGGAAaaactgcctcgc
V$GCMF	Chorion-specific transcription factors with a GCM DNA binding domain	V$GCM1.01	0.85	7842 - 7852	(-)	1.000	0.861	caCCCTcaaag
V$AP1R	MAF and AP1 related factors	V$MARE.03	0.82	7843 - 7863	(-)	0.775	0.820	cagtGCTCggtcaccctcaaa
V$AP1R	MAF and AP1 related factors	V$TCF11MAF	0.81	7845 - 7865	(+)	1.000	0.869	tgagggTGACcgagcactgga

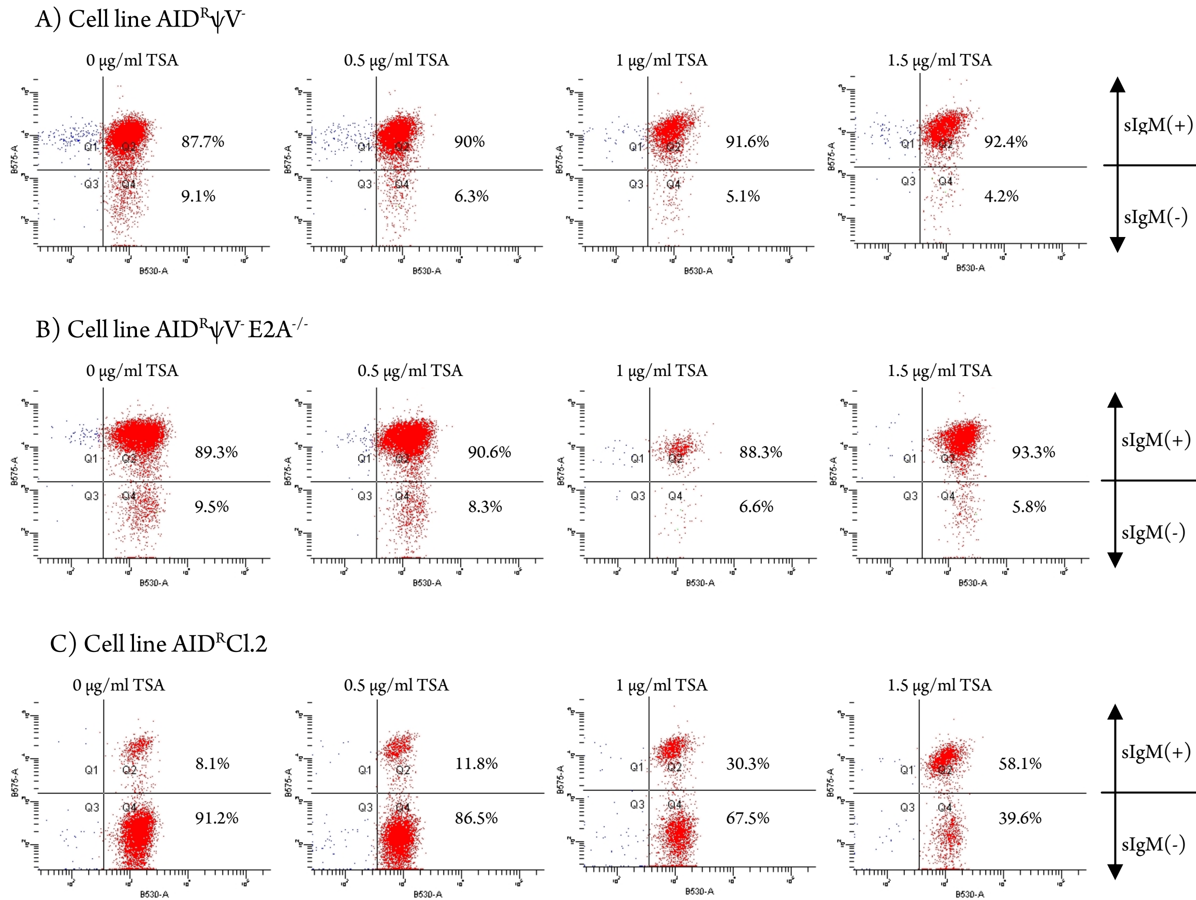
		G.01						
V$ETSF	Human and murine ETS1 factors	V$NRF2.01	0.86	7855 - 7875	(+)	1.000	0.869	cgagcactGGAAgaagctgcc
V$GREF	Glucocorticoid responsive and related elements	V$ARE.02	0.89	7855 - 7873	(-)	0.897	0.890	cagcttcttccaGTGCtcg
V$GREF	Glucocorticoid responsive and related elements	V$GRE.01	0.85	7918 - 7936	(-)	0.893	0.865	tcagcaaaaagtGTCCtgc
V$XBBF	X-box binding factors	V$MIF1.01	0.76	7926 - 7944	(-)	1.000	0.786	taggttactcaGCAAaaag
V$XBBF	X-box binding factors	V$RFX1.01	0.89	7926 - 7944	(+)	0.945	0.894	ctttttgctgaGTAAccta
V$AP1R	MAF and AP1 related factors	V$NFE2.01	0.85	7928 - 7948	(+)	1.000	0.876	ttttgCTGAgtaacctactgt
V$AP1R	MAF and AP1 related factors	V$BACH2.01	0.89	7949 - 7969	(-)	0.868	0.910	ctctgcTGCGtcaggttccct
V$AP1R	MAF and AP1 related factors	V$TCF11MAF G.01	0.81	7951 - 7971	(+)	1.000	0.867	ggaaccTGACgcagcagaggg
V$ETSF	Human and murine ETS1 factors	V$CETS1P54. 01	0.92	7983 - 8003	(+)	1.000	0.933	ggatctCCGGaggtctcttc
V$PLZF	C2H2 zinc finger protein PLZF	V$PLZF.01	0.86	8007 - 8021	(+)	1.000	0.908	cccTACAgttccatg
V$STAT	Signal transducer and activator of transcription	V$STAT5.01	0.89	8009 - 8027	(-)	0.845	0.922	gtatTTCAtggaactgtag
V$STAT	Signal transducer and activator of transcription	V$STAT5.01	0.89	8011 - 8029	(+)	0.945	0.931	acagTTCCatgaaatacct
V$IRFF	Interferon regulatory factors	V$IRF1.01	0.87	8068 - 8088	(+)	1.000	0.919	tttgtaaactGAAAcagggtg
V$KLFS	Krueppel like transcription factors	V$BKLF.01	0.95	8079 - 8097	(+)	1.000	0.956	aaacaGGGTgggtttaagt
V$OCT1	Octamer binding protein	V$OCT1.05	0.89	8089 - 8105	(-)	0.950	0.920	taCATCtaacttaaacc
V$FKHD	Fork head domain factors	V$FREAC2.01	0.84	8098 - 8114	(+)	1.000	0.870	tagatgTAAAgaagaaa
V$OCT1	Octamer binding protein	V$OCT.01	0.78	8098 - 8114	(+)	0.795	0.810	tagATGTaaagaagaaa
V$IRFF	Interferon regulatory factors	V$ISRE.01	0.81	8101 - 8121	(+)	1.000	0.858	atgtaaagaaGAAActcttca
V$KLFS	Krueppel like transcription factors	V$EKLF.01	0.89	8118 - 8136	(+)	1.000	0.940	ttcactcagaGGGTggcga
V$PAX5	PAX-5 B-cell-specific activator protein	V$PAX5.03	0.80	8133 - 8161	(+)	0.789	0.822	gcgagGCCCtggcacaggctgc ccatgga
V$GREF	Glucocorticoid responsive and related elements	V$GRE.02	0.82	8141 - 8159	(+)	0.890	0.830	ctgGCACaggctgcccatg
V$GCMF	Chorion-specific transcription factors with a GCM DNA binding domain	V$GCM1.01	0.85	8162 - 8172	(-)	0.842	0.850	caCCCGcagcc
V$XBBF	X-box binding factors	V$XBOX.01	0.90	8184 - 8202	(-)	1.000	0.906	ctcttGCCTtgggcactgc
V$XBBF	X-box binding factors	V$XBOX.01	0.90	8185 - 8203	(+)	0.875	0.937	cagtgCCCAaggcaagagc
V$HAML	Human acute myelogenous leukemia factors	V$AML1.01	0.93	8207 - 8221	(-)	1.000	0.991	ggctGTGGtcactgc
V$GATA	GATA binding factors	V$GATA1.01	0.96	8241 - 8253	(-)	1.000	0.962	ctgaGATAcgagg
V$PPAR	Peroxisome proliferative activated receptor homodimers	V$PPARG.01	0.67	8248 - 8270	(-)	0.862	0.673	agcTGGGgcagggtgagctga ga
V$AP1R	MAF and AP1 related factors	V$MARE.02	0.88	8249 - 8269	(+)	0.791	0.882	ctcAGCTcaccctgccccagc
V$ETSF	Human and murine ETS1 factors	V$CETS1P54. 01	0.92	8272 - 8292	(-)	1.000	0.935	gctgtgCCGGaggtgggagtt
V$HESF	Vertebrate homologues of enhancer of split complex	V$HES1.01	0.92	8293 - 8307	(-)	1.000	0.926	cggctgtGTGCccgc
V$AP1R	MAF and AP1 related factors	V$MAFA.01	0.92	8327 - 8347	(-)	1.000	0.924	gagTGCTgacctgccccaagg
V$AP1R	MAF and AP1 related factors	V$MARE.01	0.97	8331 - 8351	(-)	1.000	0.996	gtgtgagtGCTGacctgcccc
V$ETSF	Human and murine ETS1 factors	V$CETS1P54. 01	0.92	8372 - 8392	(+)	1.000	0.936	tcccaaCCGGaggcagctgga
V$HAND	Twist subfamily of class B bHLH transcription factors	V$HEN1.02	0.81	8377 - 8397	(-)	1.000	0.829	cgagatccaGCTGcctccggt
V$HAND	Twist subfamily of class B bHLH transcription factors	V$HEN1.02	0.81	8378 - 8398	(+)	1.000	0.852	ccggaggcaGCTGgatctcgg
V$MYOD	Myoblast determining factors	V$MYOD.01	0.88	8379 - 8395	(+)	1.000	0.946	cggaGGCAgctggatct
V$MYOD	Myoblast determining factors	V$MYOGENI N.02	0.90	8380 - 8396	(-)	1.000	0.916	gagatcCAGCtgcctcc

V$HAND	Twist subfamily of class B bHLH transcription factors	V$HAND2_E 12.01	0.75	8411 - 8431	(-)	1.000	0.792	tcaccccgcaTGGCctgct
V$EGRF	EGR/nerve growth factor induced protein C & related factors	V$NGFIC.01	0.80	8418 - 8434	(+)	0.785	0.859	ccatGCGGgggtgacac
V$GCMF	Chorion-specific transcription factors with a GCM DNA binding domain	V$GCM1.01	0.85	8418 - 8428	(-)	0.842	0.880	ccCCCGcatgg
V$EGRF	EGR/nerve growth factor induced protein C & related factors	V$WT1.01	0.92	8435 - 8451	(-)	1.000	0.987	gggggTGGGggaccccg
V$KLFS	Krueppel like transcription factors	V$BKLF.01	0.95	8436 - 8454	(-)	1.000	0.973	acaggGGGTggggacccc
V$KLFS	Krueppel like transcription factors	V$GKLF.03	0.98	8438 - 8456	(-)	1.000	0.980	ccacaggggGTGGgggacc
V$MAZF	Myc associated zinc fingers	V$MAZR.01	0.88	8439 - 8451	(-)	1.000	0.905	gggggtGGGGgac
V$KLFS	Krueppel like transcription factors	V$EKLF.01	0.89	8441 - 8459	(-)	1.000	0.901	gccccacaggGGGTggggg
V$AP1R	MAF and AP1 related factors	V$MARE.02	0.88	8484 - 8504	(-)	0.791	0.901	ccaTGCAcacgctgcagactg
V$AP1R	MAF and AP1 related factors	V$MAFA.01	0.92	8486 - 8506	(+)	0.904	0.935	gtcTGCAgcgtgtgcatggga
V$OCT1	Octamer binding protein	V$OCT3_4.02	0.88	8494 - 8510	(+)	1.000	0.942	cgtgtGCATgggaacca
V$HAND	Twist subfamily of class B bHLH transcription factors	V$TH1E47.01	0.93	8505 - 8525	(+)	1.000	0.979	gaaccattgCCAGacaccgtc
V$GCMF	Chorion-specific transcription factors with a GCM DNA binding domain	V$GCM1.01	0.85	8531 - 8541	(+)	0.842	0.850	caCCCGcagcc
V$AP1R	MAF and AP1 related factors	V$MARE.03	0.82	8542 - 8562	(+)	1.000	0.845	ctaaGCTGcctcacagcaggg
V$AP1R	MAF and AP1 related factors	V$TCF11MAF G.01	0.81	8560 - 8580	(-)	1.000	0.888	acggtgTGACggagcaatccc
V$HESF	Vertebrate homologues of enhancer of split complex	V$HELT.01	0.91	8579 - 8593	(-)	1.000	0.923	agggCACGgggtcac
V$HOXF	Paralog hox genes 1-8 from the four hox clusters A, B, C, D	V$HOX1-3.01	0.82	8600 - 8618	(+)	0.750	0.824	atcacTTATggggctggga
V$HOXF	Paralog hox genes 1-8 from the four hox clusters A, B, C, D	V$NANOG.01	0.94	8629 - 8647	(-)	1.000	0.942	cctcgttAATGtgcccaag
V$MYBL	Cellular and viral myb-like transcriptional regulators	V$VMYB.05	0.90	8638 - 8650	(+)	1.000	0.901	attAACGaggatt
V$NFKB	Nuclear factor kappa B/c-rel	V$NFKAPPAB 65.01	0.87	8642 - 8654	(+)	1.000	0.871	acgaggatTTCCc
V$AP1R	MAF and AP1 related factors	V$MARE.02	0.88	8692 - 8712	(+)	0.791	0.906	tagTGCAgacacagctgcagc
V$HAND	Twist subfamily of class B bHLH transcription factors	V$HEN1.02	0.81	8695 - 8715	(-)	1.000	0.880	taggctgcaGCTGtgtctgca
V$HAND	Twist subfamily of class B bHLH transcription factors	V$HEN1.02	0.81	8696 - 8716	(+)	1.000	0.866	gcagacacaGCTGcagcctaa
V$MYOD	Myoblast determining factors	V$MYF5.01	0.90	8697 - 8713	(+)	1.000	0.903	cagacaCAGCtgcagcc
V$MYOD	Myoblast determining factors	V$E47.01	0.92	8698 - 8714	(-)	1.000	0.932	aggctGCAGctgtgtct
V$STAT	Signal transducer and activator of transcription	V$STAT6.01	0.84	8719 - 8737	(+)	0.793	0.860	ttagTACCtgggaaaggcc
V$RBPF	RBPJ - kappa	V$RBPJK.02	0.94	8723 - 8737	(+)	1.000	0.945	taccTGGGaaaggcc
V$IKRS	Ikaros zinc finger family	V$IK1.01	0.92	8724 - 8736	(+)	1.000	0.926	acctGGGAaaggc
V$NFKB	Nuclear factor kappa B/c-rel	V$CREL.01	0.91	8728 - 8740	(-)	1.000	0.971	ggggccctTTCCc
V$GCMF	Chorion-specific transcription factors with a GCM DNA binding domain	V$GCM1.01	0.85	8734 - 8744	(+)	0.789	0.899	ggCCCCcatgc
V$PAX5	PAX-5 B-cell-specific activator protein	V$PAX5.01	0.79	8746 - 8774	(+)	1.000	0.813	gcaccgCGCAgggctgagatgt gccacgt
V$HESF	Vertebrate homologues of enhancer of split complex	V$HES1.02	0.87	8765 - 8779	(-)	0.750	0.870	tgggGACGtggcaca
V$ETSF	Human and murine ETS1 factors	V$SPI1_PU1.02	0.96	8785 - 8805	(+)	1.000	0.961	ggagctggGGAAggcaacgtg
V$AHRR	AHR-arnt heterodimers and AHR-related factors	V$AHRARNT. 03	0.95	8803 - 8827	(+)	1.000	0.960	gtggccctgtGCGTgtgcacgct ga
V$AHRR	AHR-arnt heterodimers and AHR-	V$AHRARNT.	0.95	8810 - 8834	(-)	1.000	0.969	cttgtgctcaGCGTgcacacgc

	related factors	03						aca
V$AP1R	MAF and AP1 related factors	V$MARE.02	0.88	8813 - 8833	(-)	1.000	0.891	ttgTGCTcagcgtgcacacgc
V$HESF	Vertebrate homologues of enhancer of split complex	V$DEC2.01	0.96	8832 - 8846	(-)	1.000	0.968	ccagcaCGTGtcctt
V$HESF	Vertebrate homologues of enhancer of split complex	V$HELT.01	0.91	8833 - 8847	(+)	1.000	0.910	aggaCACGtgctggg
V$NOLF	Neuron-specific-olfactory factor	V$OLF1.02	0.88	8854 - 8876	(+)	1.000	0.898	tttgtcTCCCcgggctcacgct
V$AP1R	MAF and AP1 related factors	V$BACH2.01	0.89	8872 - 8892	(+)	0.813	0.916	acgctaTGTGtcacccggtgc
V$KLFS	Krueppel like transcription factors	V$KLF6.01	0.92	8892 - 8910	(-)	1.000	0.946	gcgggaGGGGatggcgcag
V$EGRF	EGR/nerve growth factor induced protein C & related factors	V$CKROX.01	0.88	8896 - 8912	(-)	1.000	0.953	ctgcGGGAggggatggc
V$MZF1	Myeloid zinc finger 1 factors	V$MZF1.02	0.99	8896 - 8906	(-)	1.000	0.992	gaGGGGatggc
V$MAZF	Myc associated zinc fingers	V$MAZ.01	0.90	8898 - 8910	(-)	1.000	0.923	gcggGAGGggatg
V$AHRR	AHR-arnt heterodimers and AHR-related factors	V$NXF_ARNT.01	0.90	8916 - 8940	(-)	1.000	0.937	cggcgtgcggcCGTGggggagctgg
V$EGRF	EGR/nerve growth factor induced protein C & related factors	V$WT1.01	0.92	8916 - 8932	(-)	1.000	0.930	ggccgTGGGggagctgg
V$AHRR	AHR-arnt heterodimers and AHR-related factors	V$AHRARNT.03	0.95	8924 - 8948	(-)	1.000	0.966	gatgcaggcgGCGTgcggccgtggg
V$HESF	Vertebrate homologues of enhancer of split complex	V$HES1.01	0.92	8929 - 8943	(-)	1.000	0.939	aggcggcGTGCggcc
V$MYBL	Cellular and viral myb-like transcriptional regulators	V$VMYB.02	0.90	8951 - 8963	(+)	1.000	0.982	tgcAACGgcaccg
V$HOXF	Paralog hox genes 1-8 from the four hox clusters A, B, C, D	V$HOXB8.01	0.83	9004 - 9022	(-)	0.835	0.843	ggctgcaATAAaggtgagc
V$EVI1	EVI1-myleoid transforming protein	V$MEL1.03	0.95	9055 - 9071	(-)	1.000	0.982	gggtgagGATGtggcag
V$HAND	Twist subfamily of class B bHLH transcription factors	V$HEN1.02	0.81	9077 - 9097	(-)	1.000	0.835	taaggagcaGCTGggggccgt
V$HAND	Twist subfamily of class B bHLH transcription factors	V$HEN1.01	0.82	9078 - 9098	(+)	1.000	0.849	cggccccCAGCtgctccttac
V$MYOD	Myoblast determining factors	V$MYOGENIN.02	0.90	9079 - 9095	(+)	1.000	0.976	ggccccCAGCtgctcct
V$NEUR	NeuroD, Beta2, HLH domain	V$NEUROD1.01	0.83	9081 - 9093	(+)	0.790	0.833	ccccCAGCtgctc
V$OCT1	Octamer binding protein	V$OCT1.04	0.80	9088 - 9104	(-)	0.846	0.804	gaAATGgtaaggagcag
V$MZF1	Myeloid zinc finger 1 factors	V$MZF1.01	0.99	9102 - 9112	(-)	1.000	0.995	ggGGGGaagaa
V$HOXC	HOX - PBX complexes	V$PBX_HOXA9.01	0.79	9116 - 9132	(-)	0.750	0.828	cttcTGGTttatggggt
V$ETSF	Human and murine ETS1 factors	V$FLI.01	0.81	9120 - 9140	(+)	0.750	0.831	cataaaCCAGaagccgcctca
V$KLFS	Krueppel like transcription factors	V$KLF6.01	0.92	9199 - 9217	(+)	1.000	0.938	cgagcaGGGGcaggaggag
V$MYOD	Myoblast determining factors	V$MYOD.01	0.88	9203 - 9219	(+)	1.000	0.896	caggGGCAggaggagca
V$EGRF	EGR/nerve growth factor induced protein C & related factors	V$WT1.01	0.92	9223 - 9239	(+)	0.837	0.927	gggcgAGGGggcagcgg
V$MAZF	Myc associated zinc fingers	V$MAZ.01	0.90	9223 - 9235	(+)	1.000	0.909	gggcGAGGgggca
V$MYOD	Myoblast determining factors	V$MYOD.01	0.88	9228 - 9244	(+)	1.000	0.920	agggGGCAgcggggca
V$PTF1	Pancreas transcription factor 1, heterotrimeric transcription factor	V$PTF1.01	0.76	9236 - 9256	(-)	0.857	0.775	gccaGCTGtgggtgccccgc
V$MYOD	Myoblast determining factors	V$MYOGENIN.02	0.90	9243 - 9259	(+)	1.000	0.915	cacccaCAGCtggccgt
V$NEUR	NeuroD, Beta2, HLH domain	V$NEUROG.01	0.92	9246 - 9258	(-)	0.875	0.920	cggCCAGctgtgg
V$MAZF	Myc associated zinc fingers	V$MAZ.01	0.90	9266 - 9278	(+)	1.000	0.903	ccggGAGGagaag
V$MYBL	Cellular and viral myb-like transcriptional regulators	V$CMYB.01	0.90	9291 - 9303	(-)	1.000	0.919	caCAACcgctccg
V$ETSF	Human and murine ETS1 factors	V$ELK1.02	0.91	9302 - 9322	(+)	1.000	0.914	tggcggacGGAAattgttggt
V$EVI1	EVI1-myleoid transforming protein	V$MEL1.03	0.95	9317 - 9333	(-)	1.000	0.983	ccctgaaGATGaccaac
V$DICE	Downstream Immunoglobulin	V$DICE.01	0.80	9320 - 9334	(+)	0.837	0.816	ggtcATCTtcagggg

	Control Element, critical for B cell activity and specificity							
V$ETSF	Human and murine ETS1 factors	V$GABP.01	0.86	9344 - 9364	(+)	1.000	0.958	cgaggccgGGAAgtgcacggt
V$AP1R	MAF and AP1 related factors	V$MARE.01	0.97	9357 - 9377	(+)	1.000	0.992	tgcacggtGCTGacaaacgcc
V$MYOD	Myoblast determining factors	V$MYF5.01	0.90	9375 - 9391	(-)	1.000	0.931	tccccgCAGCtgcaggc
V$MZF1	Myeloid zinc finger 1 factors	V$MZF1.01	0.99	9385 - 9395	(+)	1.000	0.991	gcGGGGagagc
V$PAX5	PAX-5 B-cell-specific activator protein	V$PAX5.01	0.79	9396 - 9424	(-)	0.904	0.791	ctacgcCTCAcggctgcggcgcccgcggt
V$PAX5	PAX-5 B-cell-specific activator protein	V$PAX5.02	0.73	9416 - 9444	(+)	1.000	0.749	gaggcgtagggcgaAGCGgggcacacgcg
V$MYOD	Myoblast determining factors	V$MYF5.01	0.90	9446 - 9462	(-)	1.000	0.918	gaaagaCAGCagcagcc
V$PTF1	Pancreas transcription factor 1, heterotrimeric transcription factor	V$PTF1.01	0.76	9446 - 9466	(+)	0.857	0.764	ggctGCTGctgtctttccccc
V$MZF1	Myeloid zinc finger 1 factors	V$MZF1.02	0.99	9457 - 9467	(-)	1.000	0.994	agGGGGaaaga
V$HESF	Vertebrate homologues of enhancer of split complex	V$HES1.01	0.92	9471 - 9485	(-)	1.000	0.944	gagccgtGTGCtgcc
V$AP1R	MAF and AP1 related factors	V$MAFA.01	0.92	9510 - 9530	(-)	0.904	0.933	tccTGCAgcgtcggctgcgag
V$EGRF	EGR/nerve growth factor induced protein C & related factors	V$EGR3.01	0.77	9511 - 9527	(-)	1.000	0.794	tgcaGCGTcggctgcga
V$WHNF	Winged helix binding sites	V$WHN.01	0.95	9517 - 9527	(+)	1.000	0.966	ccgACGCtgca
V$ETSF	Human and murine ETS1 factors	V$PEA3.01	0.94	9520 - 9540	(+)	1.000	0.951	acgctgcAGGAagcccagccg
V$HAND	Twist subfamily of class B bHLH TFs	V$PARAXIS.01	0.86	9559 - 9579	(-)	0.882	0.867	acctcAGCAgatgcatttccc
V$MYOD	Myoblast determining factors	V$MYF5.01	0.90	9561 - 9577	(+)	0.836	0.901	gaaatgCATCtgctgag
V$OCT1	Octamer binding protein	V$OCT3_4.01	0.81	9561 - 9577	(+)	1.000	0.828	gaaATGCatctgctgag
V$PTF1	Pancreas transcription factor 1, heterotrimeric transcription factor	V$PTF1.01	0.76	9565 - 9585	(+)	0.857	0.779	tgcaTCTGctgaggtgcccgg
V$XBBF	X-box binding factors	V$XBOX.01	0.90	9575 - 9593	(+)	0.875	0.907	gaggtGCCCgggcaatgca
V$OCT1	Octamer binding protein	V$OCT3_4.02	0.88	9581 - 9597	(-)	1.000	0.913	gttctGCATtgcccggg
V$HOXC	HOX - PBX complexes	V$HOX_PBX.01	0.81	9596 - 9612	(-)	0.944	0.826	gtggGGATggatgaagt
V$MZF1	Myeloid zinc finger 1 factors	V$MZF1.01	0.99	9602 - 9612	(-)	1.000	1.000	gtGGGGatgga
V$GCMF	Chorion-specific TFs with a GCM DNA binding domain	V$GCM1.01	0.85	9606 - 9616	(+)	0.789	0.864	tcCCCAcatcc
V$HOXF	Paralog hox genes 1-8 from the four hox clusters A, B, C, D	V$NANOG.01	0.94	9608 - 9626	(-)	1.000	0.967	actggtgAATGgatgtggg
V$MZF1	Myeloid zinc finger 1 factors	V$MZF1.01	0.99	9622 - 9632	(-)	1.000	0.991	gaGGGGactgg
V$MAZF	Myc associated zinc fingers	V$MAZ.01	0.90	9624 - 9636	(-)	1.000	0.940	ttggGAGGggact
V$ETSF	Human and murine ETS1 factors	V$PDEF.01	0.93	9640 - 9660	(-)	1.000	0.935	gggtcgccGGATgggcttggg
V$KLFS	Krueppel like transcription factors	V$BKLF.01	0.95	9651 - 9669	(-)	1.000	0.959	agggtGGGTgggtcgccgg
V$EGRF	EGR/nerve growth factor induced protein C & related factors	V$WT1.01	0.92	9654 - 9670	(-)	1.000	0.948	gagggTGGGtgggtcgc
V$EGRF	EGR/nerve growth factor induced protein C & related factors	V$NGFIC.01	0.80	9656 - 9672	(-)	0.754	0.832	aggaGGGTgggtgggtc
V$MZF1	Myeloid zinc finger 1 factors	V$MZF1.01	0.99	9694 - 9704	(-)	1.000	1.000	gtGGGGatgga
V$GCMF	Chorion-specific TFs with a GCM DNA binding domain	V$GCM1.01	0.85	9698 - 9708	(+)	0.789	0.864	tcCCCAcattc
V$BCL6	POZ domain zinc finger expressed in B-Cells	V$BCL6.01	0.76	9703 - 9719	(+)	1.000	0.794	acaTTCCtacagatgtc
V$HAND	Twist subfamily of class B bHLH transcription factors	V$TAL1ALPHAE47.01	0.87	9705 - 9725	(+)	1.000	0.900	attcctaCAGAtgtccccttt
V$NEUR	NeuroD, Beta2, HLH domain	V$NGN_NEUROD.01	0.98	9708 - 9720	(+)	1.000	0.984	cctaCAGAtgtcc
V$NFKB	Nuclear factor kappa B/c-rel	V$CREL.01	0.91	9791 - 9803	(+)	1.000	0.922	acgggcccTTCCa

8.4 Treatment of DT40 mutants with Trichostatin A



Treatment of DT40 mutant cell lines with different concentrations of Trichostatin A (TSA) for seven days.

A) The cell line AID$^R\psi V^-$ diversifies ist rearranged Ig light chain gene solely by hypermutation (HM). Due to HM, the predominantly sIgM(+) cell line exhibits a sIgM(-) subpopulation. This subpopulation could not be increased by TSA treatment.

B) AID$^R\psi V^-$ is the precursor cell line of AID$^R\psi V^-$E2A$^{-/-}$. Due to the knockout of the E2A gene, HM in this cell line is strongly impaired. Also in this case, TSA treatment could not increase HM.

C) The cell line AID^{R}Cl.2 diversifies ist rearranged Ig light chain gene by gene conversion (GCV). Due to GCV, the predominantly sIgM(-) cell line becomes sIgM(+). In this case TSA treatment had a strong influence on the diversification processand the sIgM(+) population was greatly enhanced.

9 Acknowledgements

Mein besonderer Dank gilt Herrn Professor Horst Zitzelsberger, der sich ganz besonders für den reibungslosen Ablauf meiner Doktorarbeit eingesetzt hat. Des Weiteren bedanke ich mich besonders bei Herrn apl. Prof. Dr. Jerzy Adamski für die Betreuung dieser Arbeit an der TU München. Bei Professor Michael Atkinson bedanke ich mich sehr für die Korrektur der Doktorarbeit, die mir eine große Hilfe war und Denkanstoss zu weiteren Ideen gegeben hat.

Bei Professor Jean-Marie Buerstedde bedanke ich mich für die Überlassung des Themas und bei Professor Roland Lauster für die Vermittlung.

Hiroshi Arakawa danke ich für fachlichen Rat und das zur Verfügung stellen von Plasmiden und Zelllinien. Randy B. Caldwell danke ich für zahlreiche konstruktive Diskussionen von denen ich hoffentlich auch in Zukunft noch viele erleben darf.

Ganz herzlich und besonders bedanke ich mich bei meiner Kollegin Sabine Schmidl für die fantastische Teamarbeit und unsere vielen spannenden und fruchtbaren Diskussionen, die einen großen Teil zu meiner Freude an der Arbeit beigetragen haben.

Vielen Dank genau so an Arundhathi Sriharshan, Claire Brellinger und Steffen Heuer für die Unterstützung bei der experimentellen Arbeit und ihre Freundschaft, die mich immer wieder motiviert hat.

Mein größter und innigster Dank gilt meinem langjährigen Lebensgefährten Achim Wachsmuth, ohne dessen Zuspruch ich wohl niemals dieses schöne Studium begonnen hätte und der mir während der ganzen Zeit zu Seite stand. Umso trauriger macht es mich, dass er den Abschluss nicht mehr miterleben darf. Seine lange und schwierige Krankheit hat mir einmal mehr gezeigt, dass der Weg den ich einschlage, für mich nicht nur Beruf, sondern Berufung ist.

Ganz besonders möchte ich meiner Familie und vor allem meinen lieben Eltern Helmut und Inge danken, die mir in jeder Lebenslage den Rücken stärkten und mich immer wieder zu neuen Leistungen antrieben. Großer Dank gilt auch allen meinen Freunden, die gerade in dieser schwierigen Zeit fest zu mir hielten und mir gezeigt haben, wie stark unsere Freundschaft ist.

Own publications and research achievements

Journal papers

Schoetz U, Cervelli M, Wang YD, Fiedler P, Buerstedde J-M; E2A expression stimulates Ig hypermutation; Journal of Immunology; 1;177(1):395-400; 2006/July

Batrak V, Blagodatski A, Schmidl S, **Schoetz U**, Caldwell RB, Arakawa H and Buerstedde J-M; A cis-acting diversification activator both necessary and sufficient for AID mediated hypermutation, PLoS Genet. 2009 Jan;5(1):e1000332. Epub 2009 Jan 9.

Schoetz U, **Schmidl S**, Sriharshan A, Arakawa H, and Caldwell RB; A detailed deletion analysis of the chicken Ig light chain enhancer and its role for AID induced hypermutation, projected

Book chapter

Caldwell RB, Fiedler P, **Schoetz U** and Buerstedde J-M; Chapter 11: Gene Function Analysis using the chicken B-cell Line DT40; Gene Function Analysis; Series: Methods in Molecular Biology, Vol. 408; Ochs, Michael (Ed.) 2007, Approx. 399 p. 73 illus., Hardcover, ISBN: 978-1-58829-734-1

Patent

Schoetz U and Buerstedde J-M; Method for Enhancing Somatic Hypermutation, Gene Conversion and Class Switch Recombination. filed on October 27, 2005; application number Nr. 05023503.5; publication number EP 1 780 272.

Invited talk at a Domestic Meeting

Schoetz U, **Schmidl S**, Batrak V, Blagodatskiy A, Arakawa H, Caldwell RB and Buerstedde J-M; The role of transcription factors and Cis-acting regulatory elements for immunoglobulin repertoire development. 5th B cell forum, Bad Bevensen, April 30 – May 2, 2007
http://www.molim.uni-erlangen.de/akb/docs/medingen_program.pdf, p.4

Poster presentation and research report

Poster presentation 5th B cell forum, Bad Bevensen, April 30 – May 2, 2007, http://www.molim.uni-erlangen.de/akb/docs/medingen_abstracts.pdf, p.46

Poster presentation 6th B cell forum, Eibsee, April 10 – 12, 2008

Research report JSPS Summer program 2008: http://www.jsps.go.jp/english/esummer/data/2008/report_all.pdf, pp. 243-245

Lebenslauf

Persönliche Daten

Geburtsdatum	17.11.1976
Geburtsort	Kötzting
Staatsangehörigkeit	deutsch

Schulausbildung

1987 bis 1996	Dominicus-von-Linprun Gymnasium Viechtach Abschluss: Allgemeine Hochschulreife

Hochschulausbildung

04/1997 bis 09/1997	Germanistik, Indogermanistik, Religionswissenschaft Universität Regensburg
10/1997 bis 04/1999	Ökotrophologie TU München
10/1999 bis 08/2005	Medizinische Biotechnologie TU Berlin
03/2005 bis 08/2005	Diplomarbeit Helmholtz Zentrum München Institut für Molekulare Strahlenbiologie Titel: The Role of Bach-2 and E2A for Immunoglobulin Gene Transcription and Repertoire Development Abschluss: Diplom-Ingenieur

Dissertation

Seit 08/2005	Doktorandin am Institut für Molekulare Strahlenbiologie des Helmholtz Zentrum München

Erklärung

Ich erkläre an Eides statt, dass ich die der Fakultät des Wissenschaftszentrums Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München zur Promotionsprüfung vorgelegte Arbeit mit dem Titel:

Diversifikation der Immunglobulingene: Analyse der molekularen Mechanismen in der Hühner B Zellinie DT40

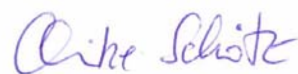
Im
Institut für Molekulare Strahlenbiologie
Helmholtz Zentrum München
Neuherberg

unter Anleitung und Betreuung durch
Prof. Dr. Jerzy Adamski

ohne sonstige Hilfe erstellt und bei der Abfassung nur die gemäß § 6 Abs. 5 angegebenen Hilfsmittel benutzt habe.

Ich habe die Dissertation in dieser oder ähnlicher Form in keinem anderen Prüfungsverfahren als Prüfungsleistung vorgelegt.

Die Promotionsordnung der Technischen Universität München ist mir bekannt.



München, den 12.05.2009 ……………………………………………

Unterschrift