UNIVERSITY OF TOR VERGATA

Faculty of Sciences

PhD school in Cellular and Molecular Biology
XIX cycle

POLYMORPHISMS AND DNA METHYLATION: TWO WAYS FOR FUNCTIONAL DIFFERENCES IN THE 3' REGULATORY REGION OF THE IGH LOCUS

by Vincenzo Giambra

Mentors:
Prof. Domenico Frezza
Prof. Barbara K. Birshtei

ABSTRACT

The IgH locus in mouse and human has a 3' regulatory region (3'RR) with multiple DNaseI hypersensitive sites. In the human, but not in the mouse, the sites (HS3, HS1.2 and HS4) are duplicated. One unit is downstream of the Cα-1 gene and a second unit is downstream of the Cα-2 gene. Human HS1,2 enhancers show polymorphic features.

In the mouse, HS3A, HS1.2, HS3B and HS4 are enhancers involved in the expression and class switching of immunoglobulin heavy chain genes. A recently identified downstream region, which contains hypersensitive sites HS5, HS6 and HS7, has been hypothesized to serve as an insulator of the Igh locus. This downstream region is associated with marks of active chromatin throughout B cell development and contains binding sites for CTCF, a protein associated with mammalian insulators. CTCF binding to many of its cognate DNA sites is prevented by DNA methylation. Previous studies using genomic Southern analysis have shown changes in DNA methylation in the upstream region of the murine 3' RR during B cell development.

In the first part of this work I identified the polymorphic structure of human HS1,2, and its distribution in some populations and in some immunological diseases. The data suggest that the HS1,2 enhancer that lies downstream of the Cα-1 gene has four alleles, one of which, allele *2, is more frequent in some immunological disorders and less frequent in the sub-Saharan region. I have also observed using EMSA that protein binding is different in the four alleles.

Furthermore I have studied changes in DNA methylation in the murine 3'RR during B cell development by digesting genomic DNA with methylation-sensitive restriction enzymes, such as HpaII and MaeII, followed by PCR. The data revealed that the 3'RR is methylated in embryonic stem cells. ES cells derived from histone H1 depleted mice showed a reduction in methylation as compared to their respective wild-type counterparts. I have detected a progressive loss of DNA methylation during B cell development. DNase I HS sites HS4, HS5 and HS7 are the earliest regulated and unmethylated sites in cell lines reflecting early stages of B development, while the HS1.2 and HS3B enhancers are unmethylated

only in plasma cell lines. DNA methylation is also reduced in splenic B cells stimulated with LPS, and LPS plus IL4 to undergo class switch recombination.

These experiments suggest that the DNA methylation pattern of the 3'RR is regulated by the H1 linker histone and related to B cell development and activation.

I have also used EMSA to detect the influence of DNA methylation on protein binding to HS5. Results indicate that USF and CTCF proteins bind HS5 *in vitro* when their specific binding sites are unmethylated, but not when these sites are methylated.

These observations suggest that DNA methylation and polymorphic regions could change the binding of insulator and/or transcription factors to the 3'RR, thereby impacting on the functions of the 3'RR during B cell development.

ABBREVIATIONS

3'RR	3' Regulatory Region
AID	Activation-induced cytidine deaminase
BCR	B cell receptor
C	Constant region
CSR	Class Switch Recombination
GB	Genbank
HS	DNAse I hypersensitive site
I	I-region promoter
Ig	Immunoglobulin
IgH	Immunoglobulin heavy chain locus
LCR	Locus Control Region
RFLP	Restriction fragment length polymorphism
S	Switch region
SHM	Somatic Hypermutation
V	Variable region

TABLE OF CONTENTS

INTRODUCTION

1. Immune system and evolution

The immune system of mammals is composed of a complex constellation of cells, organs and tissues, arranged in an elaborate and dynamic communications network and equipped to optimize the response against invasion by pathogenic organisms. The immune system is, in its simplest form, a cascade of detection and adaptation, culminating in a system that is remarkably effective.

The immune system protects the body from infection by employing three basic strategies:

- Creating and maintaining a barrier that prevents bacteria and viruses from entering the body.
- If a pathogen breaches the barriers, and gets into the body, the innate immune system is equipped with specialized cells that detect, and often eliminate, the invader before it is able to reproduce and cause potentially serious injury to the host.
- If a pathogen is able to successfully evade the innate immune cells, the immune system activates a second, adaptive immune response against the pathogen. It is through the adaptive immune response that the immune system gains the ability to recognize a pathogen, and to mount an even stronger attack each time the pathogen is encountered. The principal elements of human adaptive immune system are well known. These are called T (for thymus-derived) and B (for bursa- or bone-marrow-derived) lymphocytes (Cooper et al., 1965). T and B lymphocytes can specifically recognize and respond to antigenic determinants of potentially hazardous pathogens and toxins.

Throughout evolution, the immune system has used a remarkably extensive variety of solutions to meet fundamentally similar requirements for host protection (**Fig.1**). Adaptive and innate immunity were thought to be temporally separable processes involving different cell types. While innate immunity is generally considered to be more phylogenetically ancient, the adaptive system has diversified and refined its functions over evolutionary time such that it persists in multiple forms and shows lineage-specific diversity (Litman et al., 2005).

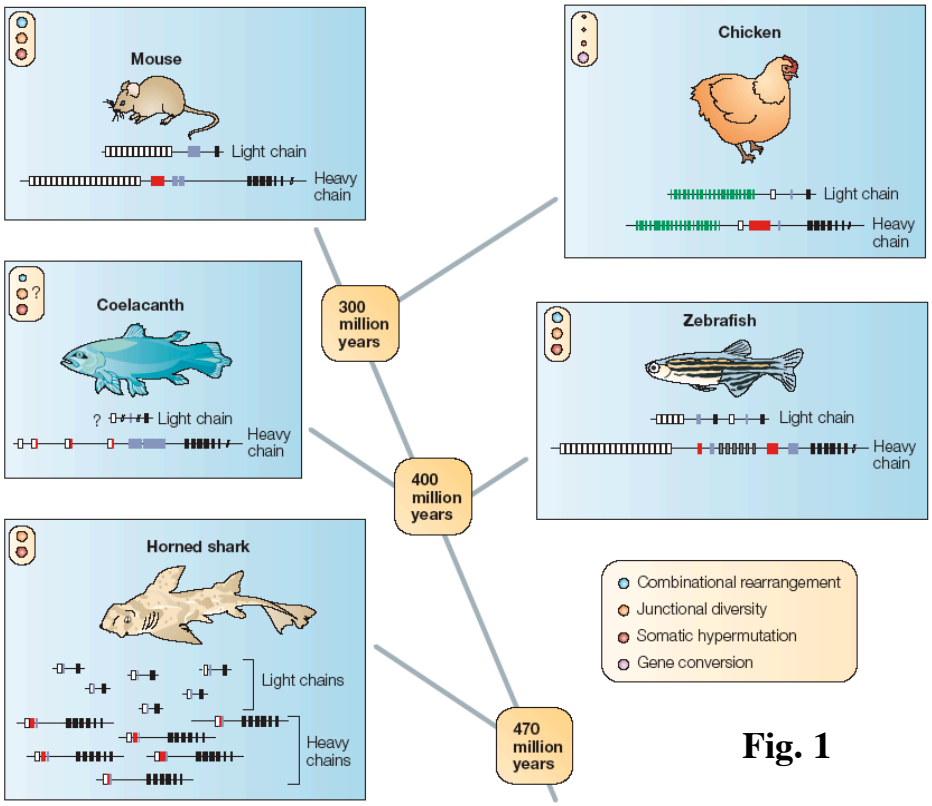


Fig. 1

Fig.1-Immune phylogeny of immunoglobulin heavy and light chain in selected jawed-vertebrate models. Throughout evolution, the immune system has used a remarkably extensive variety of solutions to meet fundamentally similar requirements for host protection, combinatorial rearrangement, junctional diversity, somatic hypermutation and gene conversion (Litman et al., 2006).

Phylogenetic studies of immune molecules clearly indicate that various mechanisms are used to diversify relatively limited amounts of genetic material to create a diverse set of receptor structures, the complexity of which is immense. The events that are likely to have occurred during the evolution of immunoglobulin and other complexes, as well as some of the other diverse immune receptors, show the successful integration of pathways that share important properties with those that are used to achieve genetic variation in viruses, bacteria, fungi and protozoan parasites. There is a co-evolutionary struggle in which selection acts on DNA in the host and in pathogens to mediate sequence variation and diversity (**Fig.2**). In this co-evolutionary struggle, the host and the

pathogen use similar mechanisms to modify selection and outcome. In various pathogens, several fundamental mechanisms of DNA change (as DNA recombination and gene hypermutation) can lead to the evasion of host detection and to the subsequent clonal selection of the pathogenic variant. In the host, analogous processes can increase the capacity of the immune system to recognize and to defeat the pathogens (Litman et al., 2006).

This view is supported by the diversity of immune mechanisms that are now being discovered in a lineage- and taxon-dependent manner. It is likely that understanding the complex mechanisms that affect immune function in multicellular organisms will also provide valuable information about different mechanisms of pathogenic activity.

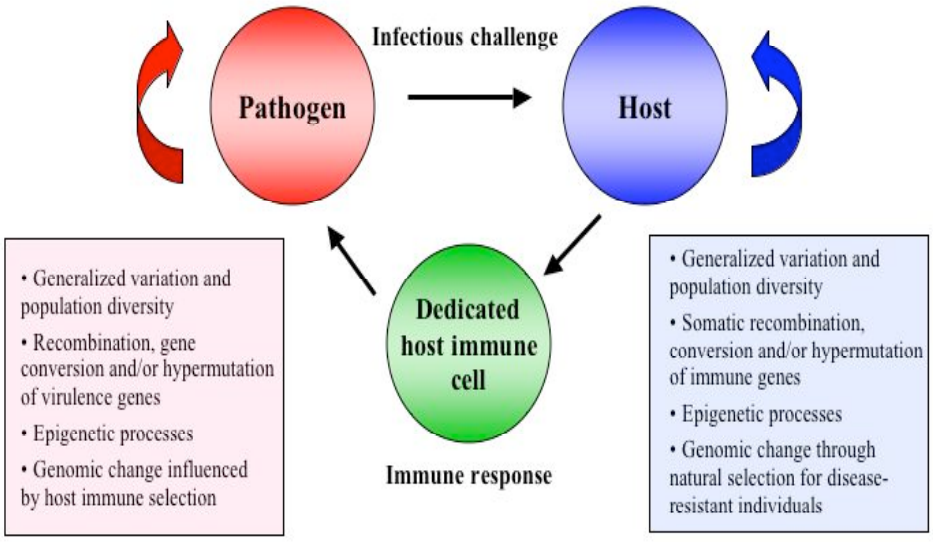


Fig. 2

Fig.2-Mechanistic similarities between pathogenicity and protection. In a co-evolutionary struggle, the host and the pathogen use similar mechanisms to modify selection and outcome. In various pathogens, several fundamental mechanisms of DNA change (such as DNA recombination and gene hypermutation) can lead to the evasion of host detection and to the subsequent clonal selection of the pathogenic variant. In the host, analogous processes can increase the capacity of the immune system to recognize and defeat the pathogens (Litman et al., 2006)

2. Structure of the immunoglobulin

Antibodies (immunoglobulin) are the key effector molecules of the adaptive immune system. Only differentiated B cells, the plasma cells, produce soluble antibodies (they are ~ 20% of total serum protein) by a tightly regulated process.

All antibodies are composed of two identical heavy chains and two identical light chains. Both the light chains and the heavy chains have a variable (V) and a constant (C) region and are held together by disulfide bonds.

The variable regions contain the antigen binding site of the antibody molecule (**Fig.3**). Although the specificity of the antibody depends on both the heavy and light chains, properties such as half life, complement fixation and placental transfer, depend only on the heavy chain. The gene encoding heavy chain can be modified by DNA rearrangements using a process called class switch recombination (CSR).

We can distinguish five different isotypes for the constant region (α, δ, ε, γ and μ); each is encoded from different DNA segments and defines the specific class of the antibodies, i.e. IgA, IgD, IgE, IgG and IgM.

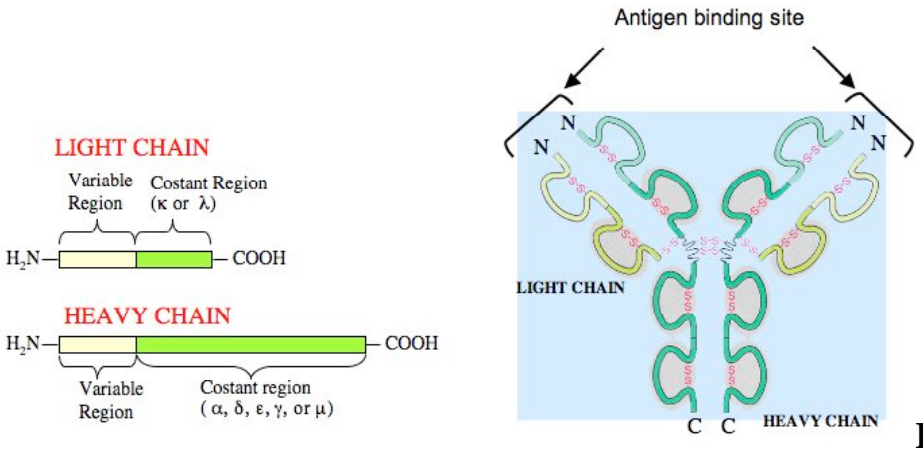


Fig. 3

Fig.3 Immunoglobulin Structure. Schematic representation of an antibody molecule with two identical heavy and light chains. Both chains have an N-terminal variable region and a C-terminal constant region and are held together by disulfide bonds. The variable regions contain the antigen binding site of the antibody molecule.

3. The murine *IgH locus* and 3' Regulatory Region (3'RR)

In the mouse the Immunoglobulin Heavy Chain gene (*IgH*) cluster is localized at Chromosome 12F1 and spans ~ 3Mb (**Fig.4A**). The variable region is close to the telomere and has ~101 VH segments, followed by ~13 DH and 4 JH segments. Downstream of the variable region gene segments, we can distinguish eight heavy chain constant region genes (Cμ, Cδ, Cγ3, Cγ1, Cγ2b, Cγ2a, Cε, and Cα). Regulatory elements are present upstream and downstream of constant region genes. The intronic enhancer, Eμ, is located upstream and a set of four enhancers (hs3A, hs1.2, hs3B and hs4) contained in a 3' Regulatory Region (3'RR) is downstream of the constant region genes (Khamlichi et al., 2000).

The 3'RR can be subdivided into two units (**Fig.5**). The first unit includes an extensive palindrome (~25 kb) containing three of the murine 3' enhancers, hs3A, hs1.2 and hs3B and families of locally repetitive elements (Chauveau and Cogne, 1996; Saleque et al., 1997). Although no biological function has yet been associated to this palindromic structure, it has been reported that under certain circumstances a cruciform structure can form in a DNA sequence with an inverted repeat or palindromic sequence (Howell et al., 1996; Shlyakhtenko et al., 2000). A palindromic structure is also present in the human 3' Regulatory Region and spans ~2.5 kb (Sepulveda et al., 2005). However the implications of this finding are still controversial. The second unit of murine 3' RR contains hs4, the most distal of the 3' enhancers.

A recently identified downstream region, which contains hypersensitive sites HS5, HS6 and HS7, has been hypothesized to serve as an insulator of the IgH locus. This downstream region is associated with marks of active chromatin throughout B cell development and contains binding sites for CTCF, a protein associated with mammalian insulators (Garrett at al., 2005).

Other experiments identified *hole* as the nearest known non-*IgH* gene, followed by *crip1, crip2* and *mta1* (Sepulveda et al., 2005). Another landmark of the region downstream of Cα is an origin of replication (Ori), which is located ~76 kb downstream of Cα in mouse erythroleukemia (MEL) cells. This origin marks the end of ~500 kb replicon that extends 3' to 5' through C, J, and D genes to the most proximal V_H genes (Zhou et al., 2002).

Recently, a cluster of hypersensitive sites was revealed 30 kb upstream of the most 5' VH gene. One of these sites, HS1, is restricted to pro-B cell lines and is accessible to restriction enzyme digestion exclusively in normal pro-B cells, the stage defined by actively rearranging IgH-V loci (Pawlitzky et al., 2006). Additional description of mouse and human 3'enhancers is presented later (See section 6)

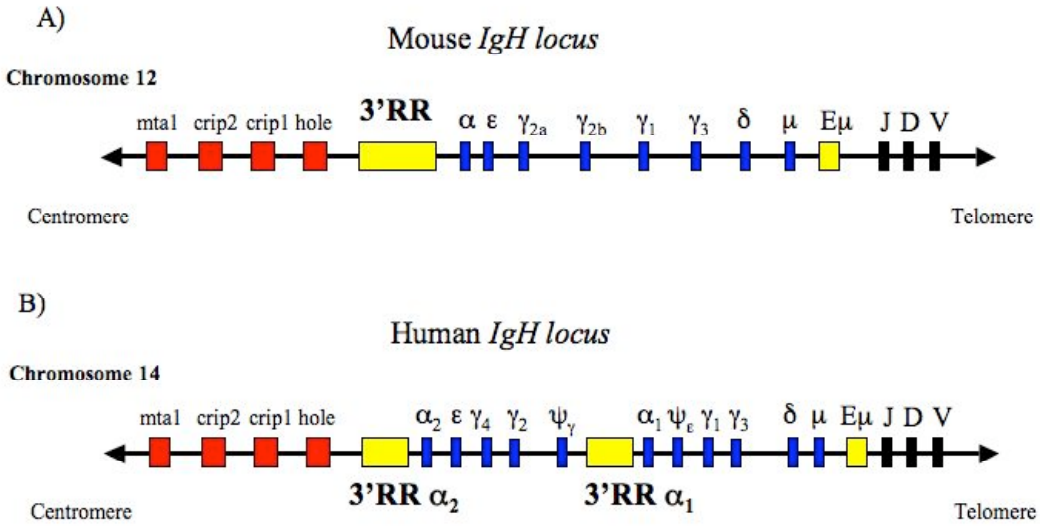


Fig. 4

Fig.4 Schematic map of mouse (A) and human (B) IgH locus. The depicted orientation reflects the order followed by genome projects in which the sequence is presented starting near the centromere and proceeding towards the telomere. Regulatory elements, such as Eμ and the 3'RR, are present near the constant genes. In humans the 3' Regulatory Region (RR) is duplicated, one unit downstream of each of the two alpha genes, α1 and α2.

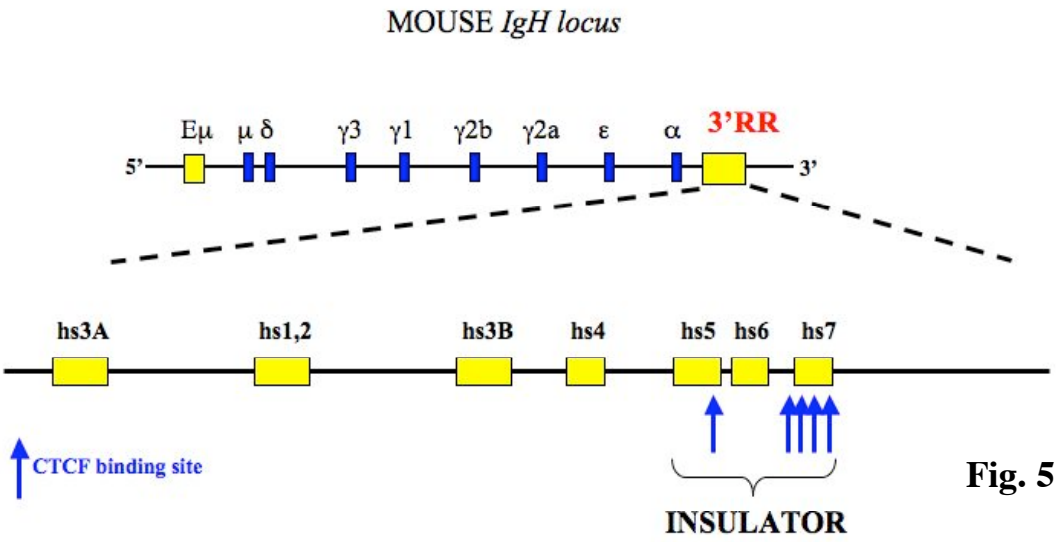


Fig. 5

Fig.5 Schematic map of 3' Regulatory Region (3'RR) in murine IgH locus. In the mouse IgH locus the 3' regulatory region (3'RR) has multiple DNaseI hypersensitive sites. HS3A, HS1.2, HS3B and HS4 are enhancers involved in the expression and class switching of immunoglobulin heavy chain genes. HS5, HS6 and HS7 have been hypothesized to serve as an insulator of the Igh locus. This downstream region contains binding sites for CTCF, a protein associated with mammalian insulator regions.

4. The human *IgH locus* and 3' Regulatory Region (3'RR)

The human *IgH locus* is located at chromosome 14q32.33, spans ~3 Mb and has the same general organization of the murine *IgH* cluster with the variable region close to the telomere (**Fig4B**). There are ~39 functional VH gene segments, followed by 26 DH and 6 JH segments.

In contrast to mouse, there is a partial duplication of constant region genes. Each unit includes four heavy chain genes, an 3' Regulatory Region (3'RR), and a downstream elk-pseudogene (Pinaud et al., 1997; Sadhu et al., 1997, Max et al. 2000).

Thus, the human constant region can be subdivided in 3 clusters (**Fig.6**):

1) Eμ – Cμ – Cδ;
2) Cγ3 – Cγ1 – ψε – Cα1 – 3'RRα1 – ψelk;
3) Cγ2 – Cγ4 – Cε – Cα2 – 3'RRα2 - ψelk.

The first cluster is not involved in the ancestral duplication. The second and the third clusters are separated by the ψγ gene (Flanagan et al., 1984; Lefranc et al., 1982).

In contrast to mouse, each human 3' Regulatory Region contains a set of three enhancers: hs3, hs1.2 and hs4 **(Fig.6)**. Humans lack an orthologue for the mouse hs3B enhancer, and the hs1.2 enhancers are inverted with respect to each other (Chen and Birshtein, 1997; Mills et al., 1997). However, the genomic structure of the two 3'RRs is extensively conserved (Sepulveda et al., 2005). Because the differences between them are very limited, it is technically difficult to identify sequences unique to 3'RR (α1) or 3'RR (α2) (Pinaud et al., 1997).

The 3'RR also shows RFLPs (Frezza et al., 1998) and reflects deletions or duplications, involving large DNA fragments (Bottaro et al., 1991; Rabbani et al., 1996). These rearrangements could change the 3'RR action and so influence the Ig expression (Rabbani et al., 1995).

There is limited sequence similarity between rodent and primates in 3' Igh regulatory regions. Both human and murine regulatory regions contain a palindrome and locally repetitive elements. In primates, repetitive elements are blocks of "switch-like" sequences that differ from the families of inverted and tandem repeats that are present in rodents. Together with enhancers, these "conserved" structural features are predicted to be essential for the activity of the 3' *Igh* regulatory region *in vivo* (**Fig.7**) (Sepulveda et al., 2005).

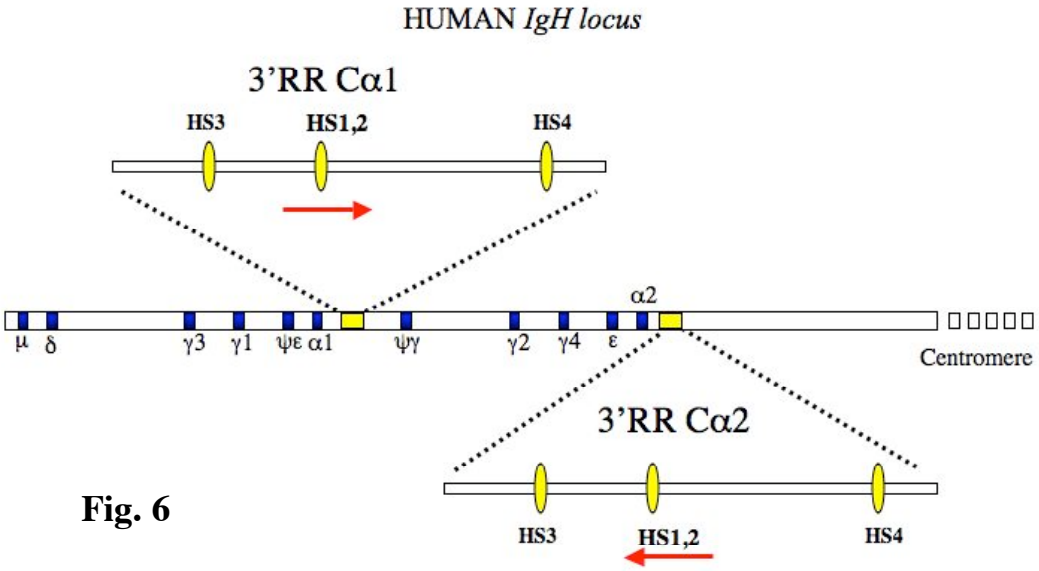


Fig. 6

Fig.6 Schematic map of 3' Regulatory Regions (3'RR) in human IgH locus. In the human IgH locus, the 3' regulatory region (3'RR) is duplicated and each has three DNaseI hypersensitive sites: HS3, HS1.2, and HS4. The genomic structure and sequence of the two 3' RRs are extensively conserved. However the hs1.2 enhancers are inverted with respect to each other

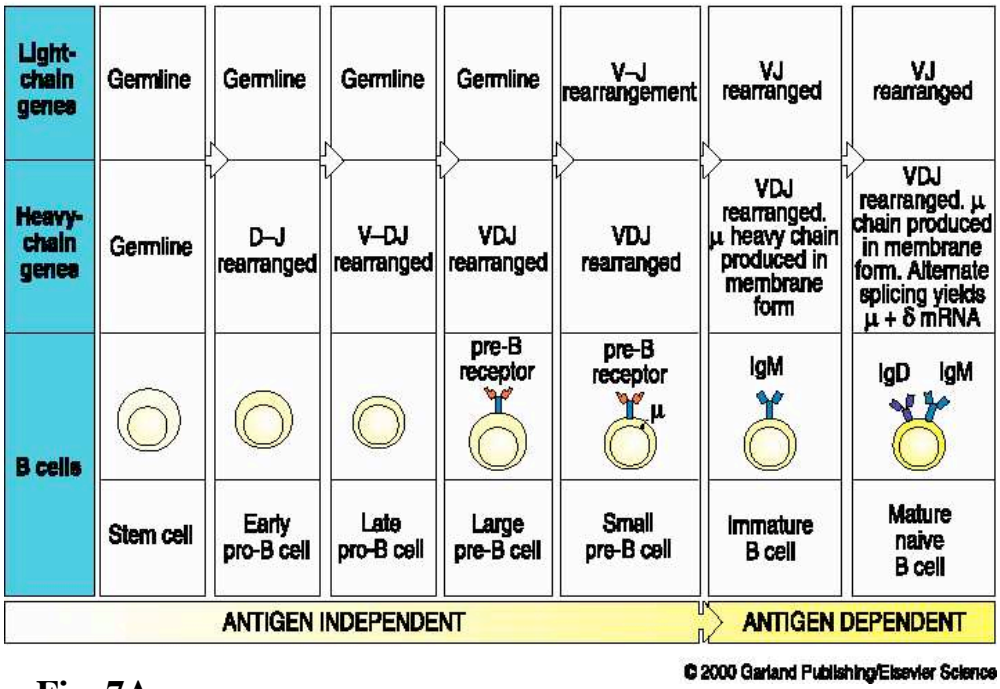


Fig. 7A

Fig.7A Schematic representation of B cell development. Progression from stem cell to mature naïve B cell is associated with V(D)J recombination in the heavy chain locus and VJ rearrangement in the light chain genes.

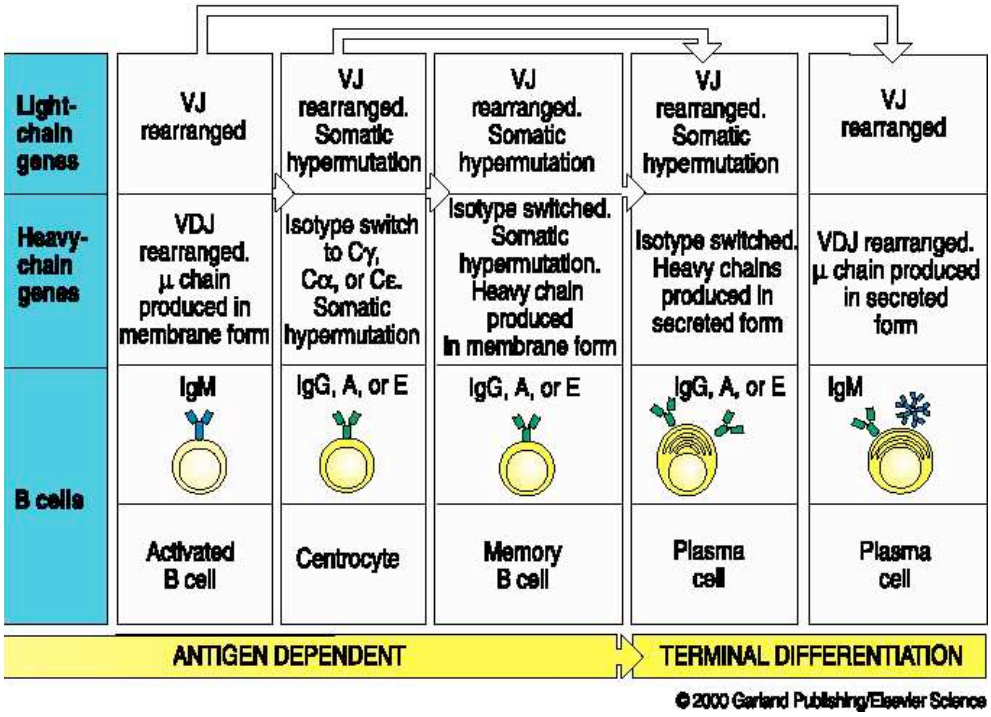


Fig. 7B

Fig.7B Schematic representation of B cell development. After antigen dependent activation, B cells proliferate and continue to differentiate into memory or antibody-producing plasma cells. Activated B cells show class Switch recombination (CSR) and somatic hypermutation.

5. B cell development and immunoglobulin gene rearrangements

B cell development is tightly associated with immunoglobulin gene rearrangements (Manis et al., 1998; Pinaud et al., 2001; Manis et al., 2003.). Three molecular mechanisms contribute to B cell development and to the diversity of the immune repertoire of B cells: V(D)J recombination, class-switch recombination (CSR) and somatic hypermutation (SHM). These three mechanisms involve marked DNA modification and require a fully competent cellular DNA-repair machinery.

5.1 V(D)J recombination and early B cell development

The various lineages of the immune system arise from a common lymphoid progenitor (CLP), which differentiates from a haematopoietic stem cell (HSC) in the bone marrow. T cells further mature in the thymus, whereas B cells develop in the bone marrow. V(D)J recombination is the first rearrangement that characterizes the early B cells and creates productive heavy and light chain alleles. Early in B cell development, at the pro-B cell stage, the D_H and J_H regions of the heavy chain locus are joined together to generate a DJ segment. Then the V_H region is joined to the rearranged DJ region. After a productive VDJ recombination, the rearranged heavy chain allele is transcribed (VDJ-Cm), translated and paired with a surrogate light chain to create a pre B cell (**Fig. 8**).

During the initial phase of V(D)J recombination, the lymphoid-specific recombinase-activating gene (RAG1)/RAG2 factors, together with ubiquitous DNA architectural proteins (high mobility group, HMG, proteins), recognize and bind to recombination signal sequences (RSSs) that flank all V, D, and J gene segments and introduce a DNA double-strand break at the border of the RSS. On the chromosome, coding ends are left as hairpin-sealed structures, whereas signal ends, which are excised from the chromosome, are blunt and 5' phosphorylated. The subsequent steps are executed by the DNA-repair machinery of the non-homologous end-joining (NHEJ) apparatus (Bassing et al., 2002).

After V(D)J recombination in the heavy chain locus and VJ rearrangement of light chain genes, the B cell receptor (BCR) with two identical heavy and two identical light chains is assembled in the membrane. B cells then leave the bone marrow and move to the secondary

lymphoid organs as mature B cells. Mature B cells express IgM and IgD molecules from the same transcript by alternative splicing.

If B cells are not activated by antigen, they will die in few days, but after positive selection by antigen, the B cells will proliferate and continue to differentiate into memory or antibody-producing plasma cells. Some activated B cells undergo class switch recombination (CSR) and somatic hypermutation in germinal centers.

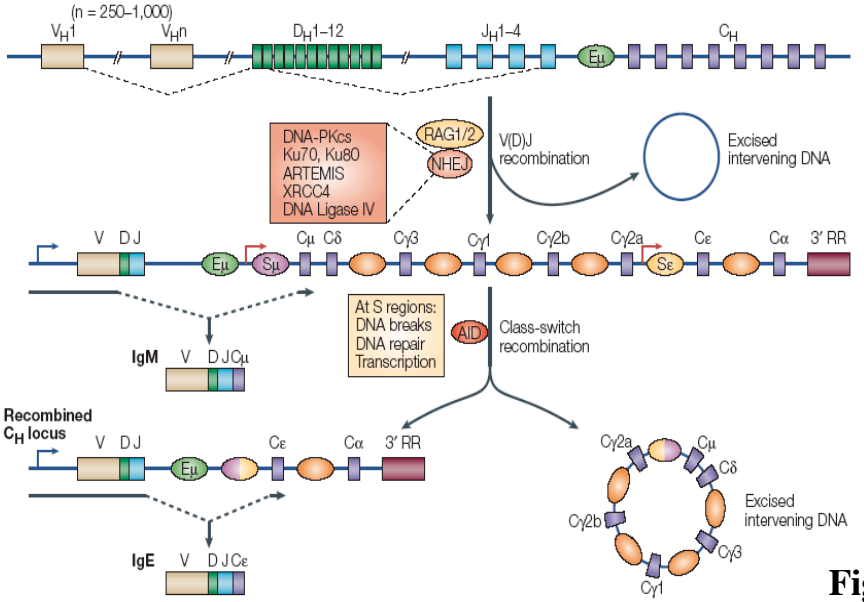


Fig. 8

Fig.8 Rearrangements at the immunoglobulin heavy chain locus. The variable region of the immunoglobulin heavy chain is assembled from component variable (VH), diversity (DH), and joining (JH) gene segments by V(D)J recombination. The process of rearrangement involves cleavage of the recombination signal sequences in the DNA, which flank the rearranging gene segments. Transcription across the locus is driven by a promoter upstream of the rearranged VDJ segment (blue arrow), which facilitates the synthesis of a μ heavy chain. This then associates with a light chain, thereby forming an IgM molecule, which is displayed on the cell-surface of a B cell. Subsequently, secondary isotypes are produced by class-switch recombination (CSR), a process that exchanges the constant region of the heavy chain (CH) with a set of downstream constant-region genes (CSR to IgE is shown). (Chaudhuri J. C. et al. 2004).

5.2 Somatic hypermutation (SHM)

Class switch recombination (CSR) and somatic hypermutation (SHM) are two distinct immunoglobulin gene diversification processes. CSR involves recombination between switch (S) regions to alter the C region of the IgH and therefore the effector function of the immunoglobulin molecule. By contrast, SHM introduces non-templated point mutations in the variable region of rearranged immunoglobulin heavy and light chain genes. SHM underlies the process of affinity maturation, which results in the preferential outgrowth of B cells expressing an immunoglobulin that has high affinity for its cognate antigen.

The mutations introduced by SHM are predominantly point mutations. Transition mutations occur about twice as frequently as transversion mutations and a high proportion of mutations arise in the hotspot motif DGYW (where D denotes adenosine (A), guanosine (G) or thymidine (T); Y denotes cytidine (C) or T; and W denotes A or T) or its reverse complement WRCH (where R denotes A or G; and H denotes T, C or A), showing that SHM is influenced by the primary sequence of the DNA (Odegard et Schatz, 2006).

Several *cis* and *trans* acting elements have been implicated in the regulation of SHM (Li et al., 2004). SHM is regulated by activation-induced cytidine deaminase (AID), Rad51 family (XRCC2 and XRCC3; involved in the regulation of chromosome stability), and several error prone DNA polymerases.

Both SHM and CSR require transcription and AID-mediated deamination of cytidine residues on the non-template DNA strand, which is exposed as single-stranded DNA during the elongation reaction.

Some evidence, based on transgenic mouse experiments, suggest that the 3' enhancers, hs3 and hs4, are involved in regulation of SHM (Terauchi et al., 2001). However more recent data showed that in the hs3-and hs4 knock out mice, SHM was not influenced (Morvan et al., 2003). Therefore, it is possible that hs3 and hs4 can cooperate with other regulatory factors to be involved in this process (Odegard et al., 2006).

5.3 Class switch recombination (CSR)

Much is known about the mechanism and regulation of CSR in mice and humans, although this process is still not understood. CSR occurs by intrachromosomal deletional recombination between switch (S) region sequences located upstream of each of the C_H genes. Before CSR, usually B cells express IgM and the VDJ-Cμ gene is transcriptionally active. If cells are treated to induce CSR, a specific promoter (I) for germline transcription is induced (e.g. Iγ3). Thus the active Sμ and specific S regions (e.g. Sγ3) become accessible to AID, which deaminates dC resulting in dU. The dU residues are substrates for uracil DNA glycosylase (UNG) which excises dU, leaving "abasic" residues. Endonucleases attack the "abasic" residues, creating nicks. Double-strand breaks are created and ligated by unknown mechanisms (**Fig.9**) (Stavnezer et al., 2004)

CSR regulation involves *cis* and *trans* acting elements. Although B cell development continues to progress after selected individual enhancer elements have been knocked out by targeted deletion (e.g. Eμ (Chen et al., 1993), hs3A (Manis et al., 1998), hs1.2 (Manis et al., 1998)), CSR is severely impaired in hs3B-hs4 KO animals (Pinaud et al., 2001) (**Fig.10**). There is a downregulation of the levels of Cm expression in resting, but not in LPS stimulated mature B cells in the hs3B-hs4 KO (Pinaud et al., 2001). *In vitro* stimulation of splenic B cells from hs3B-hs4 KO animals showed reduced production of Cγ3, Cγ2b, Cγ2a, Cα and Cε isotypes, while Cγ1 and Cμ secretion remained normal. This correlated with significant reduction in serum levels of Cγ3, Cγ2b, Cγ2a, Cα and Cε (Pinaud et al., 2001) (Dunnik WA. et al., 2005).

A recent study suggests that in mouse all elements necessary for recruitment of the recombination machinery are present in the transgene containing HS3B and HS4. These enhancers probably provide something more specific than mere increased accessibility of switch regions (Laurencikiene et al., 2006).

Hs3-hs4 KO mice show a decrease of germline transcription class switching. This implies that the 3' enhancers create contact and regulate the activity of I region promoters. However, CSR to γ1 is not impaired in these knockout mice, suggesting that the γ1 heavy chain genes are regulated in a different way.

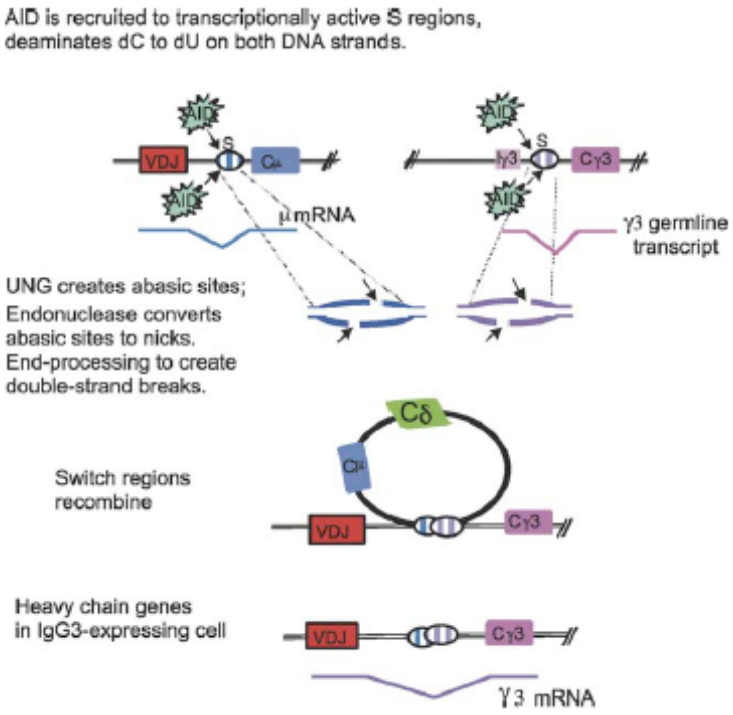


Fig. 9

Fig.9 Model for class switch recombination (CSR). The top line shows a portion of the Ig heavy chain locus in mouse B cells expressing IgM. The VDJ-Cμ gene is transcriptionally active. If cells are treated to induce CSR to IgG3, the promoter (I) for germline transcripts is induced. Thus the active Sμ and Sγ3 regions become accessible to AID, which deaminates dC to dU. The dU residues are substrates for the uracil DNA glycosylase (UNG), which could excise dU, leaving abasic residues. The endonucleases attack the abasic residues, creating nicks. Then double-strand breaks are created and ligated by unknown mechanisms (Stavnezer et al., 2004).

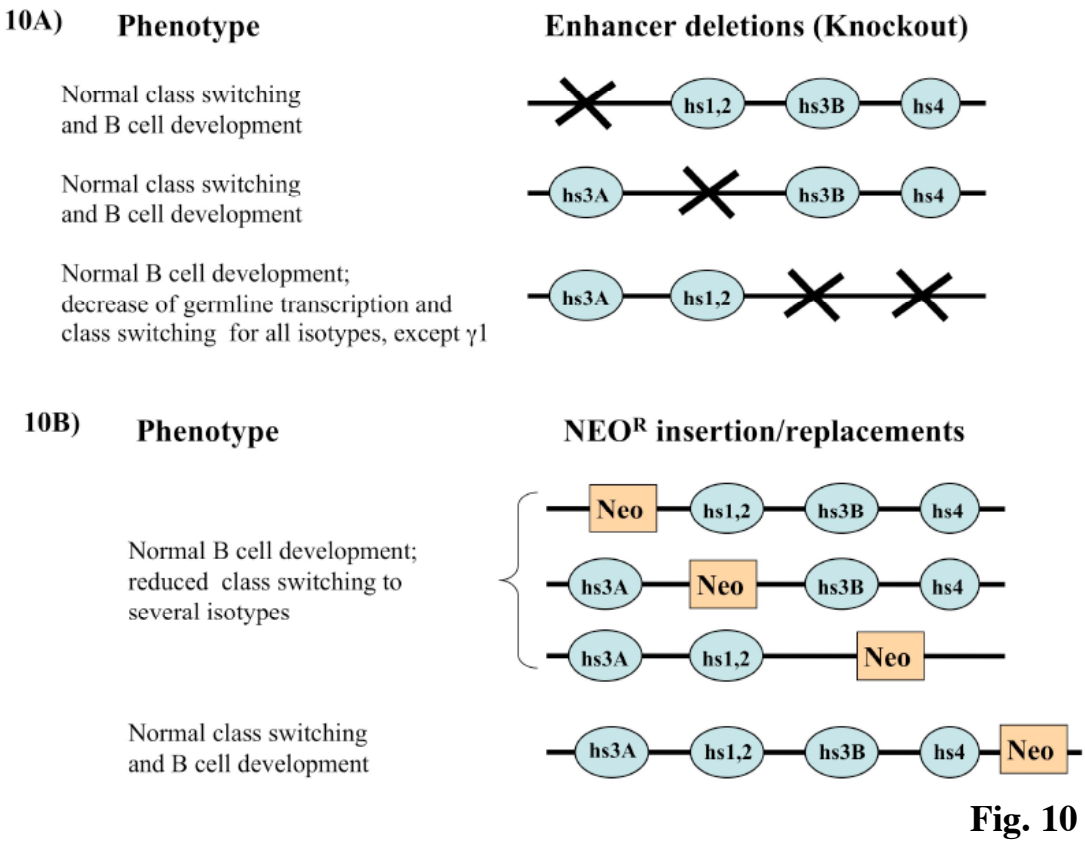


Fig. 10

Fig.10 ***In vivo*** **phenotypes for murine knockout and Neo^R replacement of 3' enhancers. A)** Clean knockout of hs3A and hs1.2 enhancers in the mouse has no effect on B cell development and class switching, suggesting that they are dispensable (Manis et al., 1998) However, the deletion of both hs3B and hs4 impaired class switch recombination to all isotypes, except γ1(Pinaud et al., 2001). **B)** The replacement of the 3' enhancers by Neo^R affects class switching (Cogne et al., 1994; Manis et al., 1998) but the insertion of a Neo-cassette downstream of hs4 does not affect B cell development or class switching (Manis et al., 2003)

6. Regulatory elements of immunoglobulin heavy chain locus

Regulation of transcription is mediated partially by the interaction of *trans*-acting DNA-binding proteins with specific motifs in promoter and enhancer elements. Although enhancers are defined by their ability to augment the level of transcription initiated from promoter elements, dissection of motifs within enhancers has revealed a more complex regulation including both positive and negative components (Ernst et al., 1995). Many of these motifs are present in the regulatory elements of disparate genes and bind to ubiquitous *trans*-acting DNA-binding proteins; yet, transcriptional regulation can be restricted in a tissue-, lineage-, and stage-specific manner. This has led to the notion of "combinatorial regulation" of transcription in which the net regulatory effect is due to a unique combination of relatively ubiquitous *trans*-acting DNA-binding proteins that interact with the regulatory elements of a specific gene (Ernst et al., 1995).

The loss (~90%) of immunoglobulin expression in a plasma cell line by the deletion of the 3' regulatory region (Michaelson et al., 1995) and, as discussed above, the decrease of germline transcription during class switching in hs3-hs4 KO mice (Pinaud et al., 2001) suggest that the 3' enhancers might regulate both V_H and I-region promoters *in vivo*.

6.1 Variable- and I-region promoters

Each V gene segment and each heavy chain (except δ) is associated with a promoter region. B cell specificity depends only on a conserved octamer motif (ATGCAAAT) and a TATA-box (Sleckman et al., 1996). Other motifs, such as E boxes, C/EBP and Ets sites, contribute to Ig promoter activity (Hatada et al., 2000).

I region promoters respond to a variety of stimuli. As discussed above, I region-driven germline transcription is regulated by the 3' enhancers (Pinaud et al., 2001). However the mechanism of these enhancers under different stimulatory conditions is not clear.

In a hypothetical model, the activation of the I-region promoters involves the 3' enhancers. Some observations indicate that the I-region promoters are required for AID to target the non-coding DNA strand (Li et al., 2004). Transcription permits the formation of R-loops between RNA

and DNA. In fact the transcribed DNA hybridizes temporarily with the RNA. It was hypothesized that AID might modify RNA or DNA structures in some way to make them more stable or better substrates for repair endonucleases (Manis et al., 2002).

Finally more recent data suggest that in the absence of UNG and MSH2, AID may occasionally act at the μ switch region in an apparently processive manner, but there is no marked preference for targeting of the transcribed versus nontranscribed strand (even in areas capable of R loop formation) (Xue et al., 2006).

Deamination of cytidine by AID results in a C to U transition, recognized by the enzyme uracil N-glycosylase, which removes the uracil from DNA. Thus the expression of germline transcripts precedes the activity of AID and uracil N-glycosylase. It is possible that the 3' enhancers might cooperate with I-region promoters in recruiting several transcription factors.

6.2 The Eμ enhancer

The murine intronic enhancer, Eμ, was the first mammalian enhancer to be described in the heavy chain locus (**Fig.4**) (Gillies et al., 1983; Banerji et al., 1983). It lies between the J_H gene segments and the *C*μ gene and is closely associated with two matrix attachment regions (MARs) (Cockerill et al., 1987). Motifs within Eμ that potentially bind to *trans*-acting DNA binding proteins have been mapped in detail (Staudt et al., 1991). A 220-bp region of Eμ defines the minimal region required for transcriptional activity. Mutational analysis in transfection assays indicated that most of the E motifs are functionally redundant. However, EMSA experiments determined that the E motifs were bound by different protein complexes, in spite of their sequence similarity (Weinberger et al., 1986). The role of individual binding motifs has been analyzed (Fernex et al., 1995). For instance, the E-box motif, μE3, in conjunction with motifs upstream of μE3, appears to be essential for Eμ -mediated accessibility to the V(D)J recombinase.

Targeted mutation studies have definitively shown that Eμ plays a role in regulating rearrangement of the heavy chain locus. In one study, hit-and-run replacement of Eμ with a short oligonucleotide resulted in slightly diminished D_H-to-J_H rearrangement (70% of normal) but more substantially

inhibited V_H to-DJ_H rearrangement of the targeted allele (Serwe et al., 1993). In a second study, replacement of Eμ with a neomycin resistance (neo^r) gene resulted in a dramatic decrease in the ability of the J_H locus to undergo recombination. Notably, insertion of the neo^r gene upstream of Eμ, without deletion of endogenous sequences, resulted in a similar *cis*-acting inhibition of J_H rearrangement (Chen J. et al., 1993). However, plasma cell lines that show a spontaneous deletion of Eμ are still capable to express high levels of immunoglobulin (Aguilera et al., 1985; Eckhardt et al., 1985). These observations suggest that there might be regulatory elements other than Eμ that are responsible for the upregulation of IgH expression at later stages of B cell development, likely, the 3' enhancers HS4 (Eckhardt et al., 1985; Pettersson et al., 1990). It has been suggested that there is a temporal shift in enhancer function between the Eμ enhancer and the 3' enhancers. In this model, early in B cell development, Eμ is active and as B cell development progresses, the 3' enhancers start to synergize with each other and with Eμ. In mature B cells, in which the 3' enhancers are active, Eμ activity begins to diminish. Finally, at the plasma cell stage, the combined activity of the 3' enhancers would be roughly the same as for Eμ (Ong et al., 1998).

6.3 The enhancers in the murine 3' Regulatory Region

Soon after the identification of Eμ, it became clear that other enhancers might exist outside the J_H-Cμ intron. In fact cell lines were reported that efficiently transcribed their IgH genes despite deletion of Eμ (Aguilera et al., 1985; Eckhardt et al., 1985). Search for additional regulatory elements based on a DNase I hypersensitivity assay led to the discovery of other murine enhancers called: HS3A, HS1.2, HS3B and HS4 (Pettersson et al., 1990; Giannini et al., 1993; Madisen et al., 1994; Michaelson et al., 1995) (**Fig.5**).

Activity of the 3' IgH enhancers was mainly assayed by transient transfection in cell lines thought to represent different stages of B cell differentiation and by transgenic models. These observations showed that the activity of the murine 3' enhancers is B cell specific and developmentally regulated (Khamlichi et al., 2000). In fact while HS1.2, HS3A and HS3B are active at late B cell differentiation stage, HS4 seems to be active throughout B cell development (Madisen et al., 1994; Michaelson et al., 1995). This expression pattern correlates quite well with

the DNA methylation pattern of the 3' region. In fact, the 3' regulatory region is hypermethylated at the pre-B cell stage and becomes demethylated at the plasma cell stage, but it is not clear whether this applies to HS4 (Giannini et al., 1993). One aim of my work has been to define better the DNA methylation pattern of the 3' region in mouse.

An important question concerning the multiple enhancers is their potential synergy in CSR and transcription activation. To understand better the complexity of problem, one has to bear in mind that the 3' regulatory region is ~200kb from Eμ (before class switching) and that the four HSs are kilobases distant from each other within a palindromic structure (except for HS4). It has been proposed that the activity of the 3' enhancers is down-regulated during early stages of B cell development. In vitro, HS3A and HS3B enhancers are regulated negatively by heterodimers composed by Bach2 and a small Maf protein (Muto et al., 1998). Pax5 has been proposed to repress the HS1.2 activity (Singh et al., 1996). Pax5 is expressed from early stages of B cell development but is not expressed at the plasma cell stage (Urbanek et al., 1992). Thus, the HS1.2 enhancer is activated at late B cell stage. Although the deletion of HS3A and HS1.2 has no effect on B cell development and class switching, it is possible that the activities of the 3' enhancers *in vivo* are redundant (Saleque et al., 1999). In fact both GL ε and γ2b promoters synergized strongly with the HS1.2 enhancer in activated primary B cells, a mature B cell line, and a plasma cell line. The principal activity of HS1.2 in activated primary B cells occurs within a 310-bp fragment that includes NF-κB, OCT, and nuclear factors of activated B cells (Ets/AP-1) sites. By mutating the consensus sequences for various transcription factors, it was determined that sites in HS1.2 are important for synergy with the GL ε and γ2b promoters. It suggests that different sites in HS1.2 might selectively interact with the GL e and γ2b promoters (Laurencikiene et al., 2001).

Reporter assays showed that HS4 is active at all stages of B cell development. However, the function of HS4 *in vivo* remains unknown, especially at early stages of B cell development. It is possible that it is involved in early regulation of chromatin changes at the 3' end of IgH locus because it is associated with markers indicative of open chromatin (Garrett et al., 2005). Murine HS4 is regulated by NF-κB, octamer binding proteins, and Pax5. It appears that Oct-1 and NF-κB binding activities positively regulate HS4 activity, whereas Pax5 is a repressor (Michaelson et al.1996). Recently it has been reported that human hs4 is regulated differently. EMSAs and Western analysis of human B cells before and after

stimulation with anti-IgM plus anti-CD40 showed a complex binding pattern formed by NF-κB, Oct-1, and Oct-2, but not by Pax5 (Sepulveda et al., 2004). This opens the possibility that other human 3' enhancers (HS3 and HS1.2) are also regulated differently from the mouse.

It has been suggested that the mouse 3' enhancers have a common regulatory mechanism because they share binding of a group of transcription factors, i.e. Pax5, NF-κB and octamer binding proteins (Michaelson et al., 1996).

Recently it was suggested that a boundary for the CSR was located within close proximity to HS4 involving DNase I hypersensitivity sites HS5, HS6 and HS7 (Garrett et al., 2005). These three segments are positioned between the IgH locus and the next non-IgH genes in a region with potential for insulator function. Moreover, preliminary experiments i.e. testing HS5 and HS6 in reporter constructs, did not detected enhancer activity in a plasma cell line. Insulators are DNA elements that protect genes from inappropriate neighboring regulatory signals by enhancer blocking activity and/or by delineating chromatin domains. All vertebrate insulators tested so far have been associated with CTCF binding sites (West et al., 2002). In this work, it was shown that CTCF binding sites are occupied *in vivo* in HS5, HS6 and HS7 in cell lines representing various stages of B cell development and normal splenic B cells (Garrett et al., 2005). The presence of a constitutively undermethylated region (Giannini et al., 1993), located in the vicinity of the mapped CTCF binding sites, suggests that these sites are accessible to CTCF binding. However the role of individual CTCF sites in the 3' RR remains to be elucidated. One aim of this thesis is to define better the correlation between DNA methylation and binding of specific proteins like CTCF.

6.4 The enhancers in human 3' Regulatory Region

In contrast to murine 3' enhancers, the human 3' enhancers, HS3, HS1.2 and HS4, are duplicated as described above (Chen et al., 1997; Mills et al., 1997; Pinaud et al., 1997) (**Fig.6**).

The human α1 and α2 HS1.2 enhancers both reside near the centers of ~10kb palindromes, with each palindrome closely flanked by a single copy of HS3 immediately adjacent to the 5' end and an HS4 unit located ~4 kb downstream (Pinaud et al., 1997). By comparison, mouse HS1.2 is centrally positioned in a considerable larger (~24 kb) palindrome that

contains a copy of HS3 on each end, with HS4 once again located ~4 kb downstream of the palindrome (Chauveau et al., 1996). Certain functional elements in murine enhancers, including Pax5 sites, do not appear to be conserved in the human HS1.2 or HS4 (Mills et al., 1997).

The 3' enhancers are also responsible for the upregulation of C-myc expression in human Burkitt's lymphoma and in mouse plasmacytomas. In these cells, a reciprocal translocation juxtaposes 3' enhancers with c-myc. It has been observed that the mouse 3' enhancers could increase the expression and the histone acetylation of the translocated c-myc (Madisen et al., 1994).

Similar to the enhancers comprising the murine 3' regulatory region, human HS3, HS1.2 and HS4 interact synergistically with each other in transient transfections and all three enhancer elements may be needed in the activation of Cα genes before switching (Hu et al., 2000). In addition, there is some evidence for specific enhancer – promoter interactions, as demonstrated by differential effects in the upregulation of the Iγ3 and Iα region promoters (Hu et al., 2000). There is a complex polymorphism of human α1 HS1.2 enhancers, in which multiple alleles have been generated through inversions and internal deletions and/or duplication (Denizot et al., 2001). This polymorphism is characterized by the presence of variable numbers of tandem repeats (VNTR) within the core enhancer. In addition the screening of patients with IgA nephropathy toward renal failure shows that one allele is significantly correlated with the disease and the α1 hs1.2 enhancer controls the level of IgA production in patients and the evolution of IgA nephropathy toward renal failure (Aupetit et al., 2000). This suggests that DNA repeated sequences could carry potential transcription factor binding sites that may boost the transcription of the α1 gene. Interestingly, minisatellites located within 3' regulatory elements and binding transcription factors have been reported in a few other cases (Kominato et al., 1997; Maeng et al., 1998), and constitute one aspect through which the occurrence of each repeat helps improve gene expression in the course of evolution. Recently, it was shown that within hs1.2, three regions (1, 2, and 3) are all necessary, but individually not sufficient, for enhancement of transcription. In region 2, a HoxC4 site and a HoxC4/embedded octamer (HoxC4/Oct) site are conserved across human, mouse, rat, and rabbit. These two sites recruit HoxC4 and Oct-1/Oct-2, which act synergistically with the Oca-B coactivator to yield the full hs1.2-enhancing activity (Kim et al., 2004).

The distribution and the influence of this polymorphism in various diseases involving immunoglobulin production deserves to be investigated (Denizot et al., 2001). It is not clear if α2 HS1.2 is polymorphic. Thus one aim of my project has been to study the two HS1.2 enhancers and to see the allelic distribution of a1 HS1.2 in some healthy populations and in patients with various diseases involving immunoglobulin production.

While a complex restriction length polymorphism is documented for the human HS1.2 enhancer, a non polymorphic restriction pattern was found for human HS3 and HS4 showing the stability of regions encompassing these two elements within the human species (Guglielmi et al., 2004).

7. The enhanceosome model

One of the central problems in understanding gene regulation is to explain how specific sets of genes are selected for expression during cell growth, differentiation or in response to environmental cues. In molecular terms, gene activity is specified by *cis*-DNA elements, enhancers and promoters, which provide regulatory infrastructure. In recent years, a model has emerged of how the different transcription factors, general DNA binding proteins (as histones and HMG), co-activators, basal transcription factors and chromatin modifying activities work together to activate transcription. In several cases specificity in gene transcription is achieved by the assembly of higher-order three-dimensional transcription factor/enhancer DNA complexes, termed enhanceosomes. The enhanceosomes activate transcription by recruiting chromatin-modifying activities and basal transcription factors to a specific promoter. In this way, a specific gene is selected for activation only if all the enhanceosome components are present in the same nucleus. Therefore, a specific gene is expressed only if specific signals are sensed and appropriately interpreted by a cell. For this reason, genes responding to a single signal assemble the corresponding enhanceosome only in response to this signal, whereas genes responding to multiple signals could assemble multiple, but specific enhanceosomes. Each set of enhancers might show architectural differences in their organization and also cooperative interactions among the different components of the enhanceosome that are essential for its assembly and activity (Merika et al., 2001).

The IFN-β enhancer is an example. It is activated upon virus infection and contains three positive regulatory binding sites recognized by NF-κB, members of the IRF family, and the ATF-2/c-Jun heterodimer (Merika et al., 2001). After virus infection, enhanceosome formation is activated. The study of the IFN-β enhancer has suggested that there is a step-wise recruitment of transcription factors and chromatin modifiers, which in turn, triggers the opening of the promoter region.

8. Gene regulation at distance and DNA insulators

Eukaryotic genomes necessarily are organized in domains with distinct functions. In fact an active gene might be surrounded by constitutively silenced chromatin structures. To an extent, the identity of these domains is maintained by classical transcriptional regulatory elements, such as enhancers, silencers and upstream activating sequences (UAS). In other cases, however, specific DNA sequences and their associated binding proteins have a role in establishing or maintaining discrete inter-domain boundaries. Such sequence elements have been called insulators.

Insulators are DNA sequence elements that prevent inappropriate interactions between adjacent chromatin domains. We can distinguish two types of insulator activity: enhancer-blocking and barrier activity. The first protects from activation by enhancers. The second protects against heterochromatin-mediated silencing.

Some compound insulators possess both enhancer-blocking and barrier activities, and enhancer-blocking insulators also protect against certain types of transcriptional repressors (Gaszner et al., 2006).

The DNA insulators are involved in the transcriptional control and in the enhancer activity. In fact it is recognized their role in the long-range interactions with the enhancer elements and in the three-dimensional organizations of chromatin within the nucleus. Insulation has emerged as a major mechanism for epigenetic control of gene expression, in particular at imprinted loci (Bell et al., 2000).

Recently some studies suggested that both kinds of insulator elements exploit the function of the other regulatory elements within the nucleus by establishing divisions between various regulative elements (enhancers, silencers, promoters and insulators). In the past it has been difficult to explain the position dependence of enhancer-blocking insulators. Recently two models were created (Gaszner et al., 2006). The first is a direct contact model. In this case, enhancers might function by directly interacting with their designated promoters. The enhancer-blocking insulators might have a steric effect that prevents enhancers from contacting other promoters, either by favouring intra-loop enhancer-promoter interactions or preventing inter-loop contacts (**Fig.11**).

Alternatively, there could be an activating signal that travels processively from enhancer to promoter (the tracking model of enhancer action). This signal could be, for example, a helicase complex that modifies histones or alters nucleosome structure, or it could be RNA polymerase

itself, launched from the enhancer. Then this signal could be blocked by an enhancer-blocking insulator as it tries to traverse the nucleoprotein structure at the base of the loop that the insulator generates.

Therefore, the loop architecture that is connected with insulator action, might be one specialized application of a more general set of regulatory mechanisms that assist in bringing together distant regulatory elements and genes and stabilizing inter-chromosomal interactions. Proteins, such as CTCF and USF, are well suited for these mechanisms, which would be quite different from their modes of action at enhancer-blocking insulators (West et al., 2004).

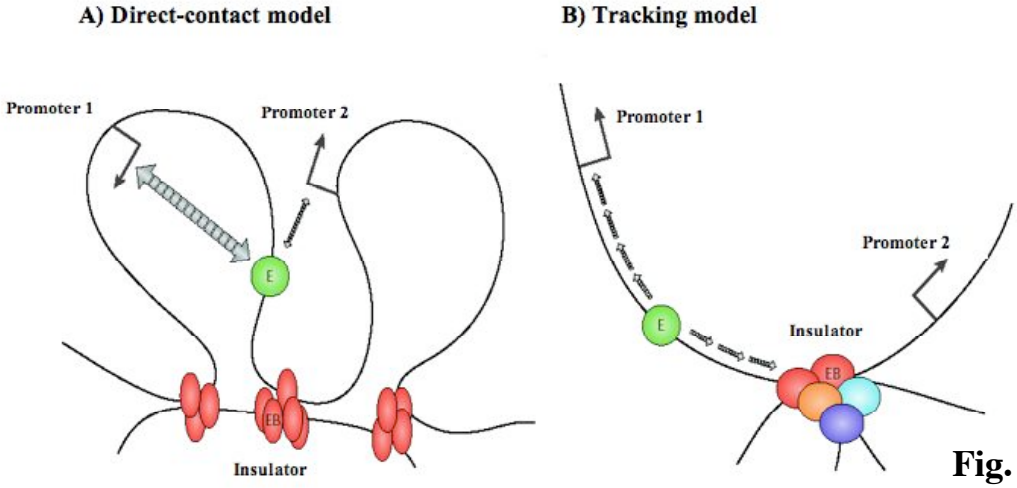


Fig. 11

Fig.11 Models for the enhancer-blocking activity A) The direct-contact model focuses on the formation of topologically closed looped chromatin domains. An assumption of this model is that the frequency of intra-loop enhancer-promoter interactions is higher than that of inter-loop interactions. This can be achieved by the existence of a mechanism that either facilitates intra-loop interactions or inhibits inter-loop interactions. **B)** Inherent to the tracking model is the idea that transcriptional activation involves the processive transfer of a signal from the enhancer to the promoter. Specific interactions between the activation signal and the enhancer-blocking complex disrupt the transfer (Gaszner et al., 2006).

9. Disorders related to immunoglobulin production: the autoimmune diseases

Autoimmune diseases are complex multisystem disorders. In many cases, their etiology is unknown, but genetic, hormonal and environmental factors are important. The population burden of autoimmune disorders is large and underestimated, because the incidence and prevalence of individual autoimmune diseases are not high. Autoimmune diseases are preceded by a long preclinical phase in which patients can be identified by the presence of characteristic autoantibodies.

Autoimmune disease can be divided into either organ-specific illnesses, such as herpetiform dermatitis and Hashimoto disease, or systemic illnesses, such as rheumatoid arthritis, systemic lupus erythematosus, Crohn's disease, psoriasis and celiac disease. Pathogenesis may be mainly mediated by autoimmune T lymphocytes. However virtually all autoimmune diseases are associated with circulating autoantibodies, which bind self-proteins (**Table 1**). Furthermore, for many diseases these autoantibodies are found in serum samples many years before disease onset (Scofield, 2004).

Numerous studies have found that autoimmune diseases have a genetic predisposition (Shepshelovich et al., 2006). The abnormal immune response probably depends upon interactions between susceptibility genes and various environmental factors. Evidence for genetic predisposition includes increased concordance for disease in monozygotic compared with dizygotic twins and an increased frequency of patients with more than one affected family member (Wanstrat et al., 2001). Susceptibility to autoimmune diseases is a multigenic phenotype affected by a variety of genetic and environmental or stochastic factors. The most potent genetic influence on susceptibility to autoimmunity is the major histocompatibility complex (MHC). Different HLA alleles are linked to different autoimmune diseases. However, there are also non-MHC susceptibility alleles related to autoimmune diseases. These are difficult to identify, predominantly due to extensive genetic heterogeneity and possible epistatic interactions among the multiple genes required for diseases development. An intriguing finding indicates that several alleles affect multiple autoimmune diseases, or simply that many immune system genes are clustered together (Wanstrat et al., 2001). It is consistent with the hypothesis that certain immunological pathways are common to multiple autoimmune diseases, whereas other pathophysiological mechanisms are specific to a particular disease. The

effect of genetic factors on the development of autoimmune disorders could be underestimated. Therefore, the presence of a specific autoantibody in an individual's serum combined with disease prone MHC haplotypes or other susceptibility loci increases exponentially the risk for that person to develop an autoimmune disease in the future (Shepshelovich et al., 2006).

It has not been clear if there is a correlation between the 3'RR of immunoglobulin heavy chain locus and autoimmune diseases. The only linkage that was shown was between patients with IgA nephropathy toward renal failure and α1 hs1.2 enhancer (Aupetit et al., 2000). One aim of my project has been to study the distribution of the α1 hs1.2 polymorphic enhancer in different autoimmune disorders in order to understand the functional significance of different alleles of HS1.2.

Autoimmune disorder	*Self antigen*
Addison's disease	21-hydroxylase
Celiac disease	Transglutaminase
Type 1 diabetes	Insulin, GAD-65
Graves disease	TSH receptor
Hashimito's thyroiditis	Thyroid peroxidase, thyroglobulin
Myasthenia gravis	Acetylcholine receptor
Goodpasture's syndrome	Type 4 collagen
Multiple sclerosis	Myelin basic protein, myelin oligodendritic glycoprotein
Systemic lupus erythematosus	Double-stranded DNA
Sjögren's syndrome	Ro/La ribonuclear particles
Rheumatoud arthritis	Citrillunated cyclic peptide, IgM
Dermatomyositis/poltmyositis	t-RNA synthetases
Scleroderma	Topoisomerase
CREST syndrome	Centromere proteins

Table 1 Selected autoimmune disease and characteristic autoantigens (Shepshelovich et al., 2006)

10. Immunoglobulin production in response to neuroleptics in schizophrenic disorder

Schizophrenia affects 1% of the world's population, but its cause remains obscure. Numerous theories have been proposed regarding the cause of schizophrenia, ranging from developmental or neurodegenerative processes or neurotransmitter abnormalities to infectious or autoimmune processes (Jones et al., 2005). However, nowadays, there is not a clear hypothesis about the etiology of this disorder. The immune alterations in schizophrenia have been described for decades. The unspecific, "innate" immune system shows signs of over-activation in non medicated schizophrenic patients. The question of whether some or all cases of schizophrenia have an immune or autoimmune etiology has been asked now for nearly 100 years. Many general immune abnormalities have been reported over this time. These include morphological changes in lymphocytes, altered levels of $CD4^+$ $CD45RA^+$ T cells, $CD8^+$T cells, $CD5^+$B cells and γδ T cells, increased or decreased levels of γ-globulin in serum, increased levels of circulating cytokines, particularly IL-2, IFN-γ and IL-6, and increased levels of antiviral antibodies (Rothermundt et al., 2001). Despite numerous reports of immune abnormalities in schizophrenia, the hypothesis that these abnormalities are related to the pathogenesis of the disease continues to be viewed with much skepticism. However associations of other autoimmune diseases with particular MHC haplotypes, increased serum levels of autoantibodies, and *in vivo* and *in vitro* replication of some of the functional and ultrastructural abnormalities of schizophrenia by transfer of autoantibodies from the sera of patients with schizophrenia suggest that, in some patients at least, autoimmune mechanisms could play a role in the development of disease (Jones et al., 2005).

Pharmacological treatments of schizophrenia alter also the immune system. In fact, during antipsychotic therapy with neuroleptics, the specific TH-1 cells become activated and, in addition, the B cell system and antibody production increase in some patients (Muller et al., 2000). Other evidences suggest that antipsychotic drugs, e.g. chlorpromazine and the atypical compound clozapine, influence the production of cytokines. Cytokines, organized in networks of related peptides with pleiotropic functions, are pivotal humoral mediators of infection and inflammation, and they play an important role in hematopoiesis, autoimmunity and immunoglobulin production (Pollmacher et al., 2000). Therefore, these

treatments can represent a model to study drug dependent immunological alteration.

AIM OF PROJECT

After these considerations, we want to know how HS1.2A and HS1.2B, the two enhancers of the human 3'RRs, could relate to immunoglobulin production, in terms of structure and allelic distribution in different populations. Finally we want also to see how the accessibility of the murine 3'RR changes during CSR and B cell development by studying DNA methylation state of this locus.

Therefore the principal aims of the project are:

1. To study the polymorphic structure of the two 3' enhancers, HS1.2A and HS1.2B, in the human heavy chain locus of immunoglobulin;

2. To study the distribution of alleles of polymorphic HS1.2 in healthy populations and in patients with different autoimmune disorders;

3. To study the pattern of DNA methylation in the murine 3'RR of immunoglobulin heavy chain locus during B development and under specific stimulation for class switching recombination (e.g. LPS or LPS plus IL4).

MATERIAL AND METHODS

1. *In silico* analysis

Contings of the constant heavy chain region were rescued from GenBank with Blast programs of the NCBI and from maps of the chromosome 14q32 telomeric region (www.ncbi.nlm.nih.gov/mapview/maps.cgi).

Comparison of the sequences of the 3'RRα1 and 3'RRα2 was performed with Clustal W and Gene Jockey programs. A new contig at the 3'RR α1 sequence (called CHR77) was assembled from the junction of five genomic contigs (X76785, Y14407, AL928767, AL928765 and U64453), using the overlapping regions (**Fig. 12**).

The sequencing data of the different clones were also compared by Clustal W and Blast programs.

The WebGene program (in Cabibbo et al., 2002) at the web site http://125.itba.mi.cnr.it/~webgene/genebuilder.html was used to find the consensus sites for transcription factors on the polymorphic HS1.2-A and HS1.2-B enhancers.

2. DNA extraction

The DNA of some human donors was extracted from total blood or peripheral blood lymphocytes using standard methods.

The DNA of other human samples was extracted from buccal epithelial cells embedded in a sterile cotton plug (Becton Dickinson, Sparks, MD, USA) The cotton plugs were removed from the support with a sterile lancet and introduced in a sterile vial with 1.0 ml of TE with SDS 1%, RNAse (100 mg/ml), and 20 mg/ml of proteinase K, then incubated o.n. at 37 °C with gentle shaking. The DNA was harvested according to the standard protocol of Microcon (Millipore, Bedford MA, USA).

3. The selective PCRs for the human HS1.2-A and HS1.2-B

The 5.4-kb fragment of 3'RR-1 was selectively amplified by the primers SA2.5 (5'-GGA TCC CTG TTC CTG ATC ACT G-3') and A2R (5'-GCC CTT CCT GCC AAC CTG-3'), respectively located in U2 and U8 sequences (**Fig.13 and Fig.14**) which are differently orientated in the two 3'RRs. Conversely, to amplify the 3'RR-2 specific fragment of 4.4 kb, the primer A2R was selected respectively within the U8r sequence and it was paired to primer A2F (5'-GCA CTG TCG GCT TAC AGA GG-') within UB2, which is unique in the 3'RR-2.

Reaction conditions were: 1.5 units of Taq Polymerase Platinum High Fidelity (Invitrogen, Carlsbad, CA, USA) buffer 1x Platinum High Fidelity (200 mM Tris–HCl, pH 8.4, 500 mM KCl); dNTPs (0.2 mM), $MgCl_2$ (1.5 mM), primers 15 pmol, water for final volume of 50 ml. The reaction was done at 94 °C for 2 min followed by 10 cycles at 94 °C for 30 s, 59 °C per 30 s, 68 °C per 5 min, followed by 20 more cycles at 94 °C for 30 s, 57 °C for 30 s, 68 °C for 5 min and one final extension at 72 °C for 10 min.

A nested PCR was performed for both loci with the same two primers, P3Frw (5'-GAC TCA TTC TGG GCA GAC TTG-3') and D3Rev (5'-GTC CTG GTC CCA AAG ATG G-3'), in order to obtain the polymorphic region of the enhancer HS1,2. The amplification conditions were performed with 1/10 of the volume of the selective PCR to minimize DNA genomic carryover, using 1 unit of Platinum Taq Polymerase (Invitrogen) and 1x buffer platinum, dNTPs (0.2 mM), $MgCl_2$ (1.5 mM), primers (15 pmol) and adding water to a final volume of 50 ml. The temperature steps of the reaction were as follows: 94 °C for 2 min followed by 30 cycles at 94 °C 30 s, 56 °C for 30 s, 72 °C for 1 min followed by a final extension at 72 °C for 5 min.

4. DNA amplification of HS1.2 alleles from monkeys and apes

Amplification of the HS1.2 enhancer from monkeys was performed with the same human primers (P3frw and D3rev), used for the direct nonselective PCR of the short fragment. The ape enhancers were amplified using the same nested PCRs used for humans except that the annealing temperature was 54°C. Southern blot analysis was performed on genomic DNA from five samples of *Macacus fascicularis* digested by XbaI and hybridized with the human enhancer probe with the fluorescence system (Roche Diagnostic, Germany). DNAs were extracted from blood using the standard protocol.

5. Cloning and analysis of the polymorphisms

PCR products were cloned with TOPO XL PCR Cloning kit (Invitrogen) in the plasmid pCR-XL-TOPO and electroporated in E. coli TOP10 (Invitrogen) competent cells. The cells were plated on ampicillin selective agar medium, X-gal and incubated o.n. until white/blue colonies were visible. Minipreps were performed with SIGMA kit (Illinois, USA) and DNA from positive clones extracted, sequenced with M13 forward and reverse universal primers and analyzed by automatic AppliedBiosystem capillary 3700 apparatus at the BMR Sequencing Service CRIBI (Padova, Italy). DNA fragments particularly rich in GC repeats were resolved by use of DMSO in the sequencing reactions. Polymorphism frequency in the different populations was determined by electrophoresis of PCR products on agarose gels (3%) stained with ethidium bromide. The alleles of HS1.2-A and HS1.2-B were analyzed as shown in **Fig.15.**

6. Human and mouse cell lines

Mouse and human cell lines were maintained in complete-RPMI-1640. All cell culture media were supplemented with 10% FBS (Gemini Bio-Products), 1% Penicillin/Streptomycin (P/S) and β-mercapthoethanol. These media will be referred as “complete media” from here on. All cells were grown at log phase at 37°C in a humidified atmosphere with 5% CO_2.

7. Mouse primary B cells

Normal mouse B cells were obtained from spleens of C57B/6 female animals 6-8 weeks of age. Briefly, spleens were surgically removed and placed in PBS buffer (1x) and then homogenized between two glass slides. The cell suspension was filtered through cheesecloth and centrifuged at 1,800 rpm for 5 min at 4-10°C (Beckman). The cells were resuspended in 3 ml of red blood cells (RBC) lysis solution (Puregene), incubated for 3 min at room temperature (RT), and recovered by cetrifugation. T cells were removed by negative selection using anti-CD43 (Ly-48) MACS Microbeads (Miltenyi Biotec, Auburn, CA) as indicated by the manufacturer. B cell purity was 95.3% as evaluated by FACS analysis of CD43(-), B220(+), CD3e(-) cells.

8. LPS or LPS plus IL4 induced class switch recombination in mouse B cells

Mouse primary B cells (10^6 cells/ml in 20ml RPMI) were induced to undergo CSR *in vitro* by addition of LPS (50 μg/ml Calbiochem) or LPS plus IL4 (final concentration of 50 ng/ml- R&D Systems). Non-stimulated B cells and cells stimulated for 48h, 72h and 96h were analyzed for the expression of CTCF, USF1 and USF2. Total RNA prepared from 5 x 10^6 cells per sample was used to monitor for germiline transcripts by RT-PCR using specific primer pairs for mouse γ1, γ2b and γ3 at 0, 48, 72 and 96h. Cell surface expression of the switched isotypes as well as cell cycle progression were monitored by FACS analysis.

9. DNA extraction from cell lines and splenic B cells

10^7 cells were centrifuged at 1,800 rpm for 5 min at 4-10°C (Beckman). The pellet was resuspended in 0.5 ml of lysis buffer (10mM tris HCl, pH8.5, 5mM EDTA, 0.2% SDS, 0.2M NaCl) and 10 µl of 20 mg/ml Proteinase K. The solution was incubated at 65°C for 1h. One volume of chloroform was added and the sample was mixed and centrifuged at full speed for 15 minutes. The upper phase was recovered and the DNA was precipitated by one volume of isopropanol. The DNA was recovered and resuspended in distillated water.

10. Amplification by PCR of HpaII and MaeII sites

The different restriction sites for HpaII and MaeII were amplified in individual PCR. The used primers were reported in the **Table 2**. The reaction conditions were: 1 unit of Z Taq. (generously given by the Scherer Lab in Albert Einstein College of Medicine), buffer 1x Platinum High Fidelity (200 mM Tris–HCl, pH 8.4, 500 mM KCl); dNTPs (0.2 mM), $MgCl_2$ (1.5 mM), primers 15 pmol, water for final volume of 50 µl. The reaction was done at 94 °C for 2 min followed by 27 cycles at 94 °C for 30 s, 60 °C per 30 s, 72 °C per 30 sec and one final extension at 72 °C for 7 min.

PCR products were analyzed on agarose gels (1.2%) stained with ethidium bromide. The intensity of bands was measured by Chemimager V 5.5 software. Each PCR reaction for the different sites was used to amplify the digested and undigested DNA (**Fig.23**). Each signal of PCR products was normalized with another, obtained by amplification of a fragment, called HS4 (see **Table2**), without restriction sites for HpaII or MaeII. The normalized signals of digested and undigested DNA were compared between them. This comparison was expressed as percent.

Primers to amplify the HpaII and MaeII sites

Location Restriction site*	Primer Name	Primer Sequence	Product size
7034 (HpaII)	6911-5'	GCC TCC TGT GCC AGG GAA G	400 bp
	7281-3'	GGT TGG GTG TGG TGG CAC ATG	
7531-7545 (MaeII)	7281-5'	CAT GTG CCA CCA CAC CCA ACC	270 bp
	7608-3'	GCT GTA AGG GGA CAC TGG GGG	
10040 (MaeII)	9953-5'	GCC ATG AAC AGA GAG TGG AGG	391 bp
	10344-3'	GGT CTT GGT CAC ATG ACA TAG GGT	
11439 (MaeII)	11290-5'	AGG CAA GAG TAG CCA GAG GG	270 bp
	11560-3'	GAT CAT GCC CTC AGC CCA G	
13424 (MaeII)	12721-5'	GTG CCC CCT TTC TCT AGG TG	864 bp
	13585-3'	GGG ACA GAA ACA GAA AGA TGG GG	
14256 (HpaII)	14101-5'	CCC CTA AGC TTC TGG CCC	370 bp
	14471-3'	CTG GCT ATG GAG ACA GAC CCT	
14597 (MaeII)	14471-5'	AGG GTC TGT CTC CAT AGC CAG	281 bp
	14711-3'	CAC AGG GCT TGA TGT TGG TGA C	
14860 (MaeII)	14711-5'	GTC ACC AAC ATC AAG CCC TGT G	409 bp
	15120-3'	CCC ATA CCA GCC TTG CCC	
15758 (HpaII)	15707-5'	TGG TGT CCA TCA GGA TCC TGG	379 bp
	16086-3'	GGT GCA GGC TTC AGG TGA G	
17896-17910 (MaeII)	17729-5'	TCT CAA CCC AGG GTA GTA CAC C	330 bp
	18059-3'	CCC TCC CTC CAT TTC CAA GC	
20256 (MaeII)	20022-5'	ATC CTT CAG AGC ACC CTG GAA C	311bp
	20333-3'	TGA GAT GCA GGG ACA GGG AG	
21033 -21047 (MaeII)	20867-5'	CTC CAA ACC CAG GGT AGT ATA CCC	234 bp
	21101-3'	CAC TCT GGC AAA GCC TCC C	
23243 (HpaII)	2291-5'	CCT GCT GTT CAC CAG GTC C	402 bp
	23393-3'	GGA AAC CAG GGG TCA CAG G	
25002 (HpaII)	HS3B-5'	TGG TTT GGG GCA CCT GTG CTG AG	796 bp
	HS3B-3'	TCG GTC CTG GTA ACT GGC TAC TG	
25189 (MaeII)	HS3B-5'	TGG TTT GGG GCA CCT GTG CTG AG	796 bp
	HS3B-3'	TCG GTC CTG GTA ACT GGC TAC TG	
26414 (MaeII)	26217-5'	CGT GTC CCC TCT CCC CAC	623 bp
	26840-3'	TGG GTA GAG GCA TGG CTG G	
27326 (MaeII)	2A	GTA CAG GGC ACA AGG ATG GTG	192 bp
	2B	CTT GAG TGT GAC CCA GGC TTG	
29107 (HpaII)	11A	GTT GAC ACT GGC ACT CCC ACT CCG AG	823 bp
	13B	CAC CAT GGC TGC CCT GGG CCA AG	
29133-29690 (MaeII)	11A	GTT GAC ACT GGC ACT CCC ACT CCG AG	823 bp
	13B	CAC CAT GGC TGC CCT GGG CCA AG	
30475-30684 (HpaII)	15A	GGG GTG TAT CCT CTG GCT ACA GGA G	834 bp
	17B	GTG CGC GCA CAC ACG CAG GGG AAA CCC AG	

* the numbers refer to Bac199 GenBank Accession Number AF450245

Table 2

Location Restriction site*	Primer Name	Primer Sequence	Product size
31326 (HpaII)	18A	GCT TGC CTT CTG GGC TAG AAC ATG G	517 bp
	19B	CGT CTG CTT TCA GGA CAG GAT GGT GGG G	
31417 (HpaII)	19B for	CCC CAC CAT CCT GTC CTG AAA GCA GAC G	136 bp
	19.5 B	CAC CCC TGA TCT CCT AGT TCA TTA CCC	
31423 (MaeII)	19 B for	CCC CAC CAT CCT GTC CTG AAA GCA GAC G	591 bp
	21B	CTG GAA ATG TGG GCC AAC TCT TCA CC	
31492 (HpaII)	19.5 A	GGG TAA TGA ACT AGG AGA TCA GGG GTG	481 bp
	21B	CTG GAA ATG TGG GCC AAC TCT TCA CC	
33355 (HpaII)	27A	GGG CTC CAG AGT GGT GTA ATA CAG GG	594 bp
	28B	GTA CTC CTG GTC GCT TTA GCT CAG TG	
33834 (MaeII)	27A	GGG CTC CAG AGT GGT GTA ATA CAG GG	594 bp
	28B	GTA CTC CTG GTC GCT TTA GCT CAG TG	
34221 (MaeII)	30A	GTC CCA TAT TCC ACA CAG ATC CAT G	360 bp
	30B	GGC TTG GCA CAG CAT GCA GTG GTG	
34942 (HpaII)	33A	TGT GTG GTT TAG ACA CTT ATG TCT C	560 bp
	34B	TGG GCA TTA CCA TGA GGT GG	
36033 (HpaII)	37A	GGT AGA CTC TCA GGC ACA AGT GAC A	741bp
	39B	ATG TAT TTA GTG GGG CAG TC	
36777 (HpaII)	40A	GAA ATA TAC CAC TAG CCT TGG A	589 bp
	41B	GTC AAC CAG AGT GCC ATG GA	
37149 (HpaII)	42A	AGC TTC GAG GAC CCT CAT TC	525 bp
	43B	AAG GAG CTT CAG TTG GGG AA	
HS4	3A	TAT CAC TAG CCC TGA TAT GG	715 bp
	5B	GAC TGG CCA GGT CTG GGC CCC AGA GC	

* the numbers refer to Bac199 GenBank Accession Number AF450245

Table 2

Table 2 The primers used to amplify the different HpaII and MaeII sites in the murine 3'RR.

11. Electrophoretic mobility shift assays (EMSA)

Nuclear extract preparation and EMSA were performed as described previously (Schreiber et al., 1989). Briefly, 50-70x10^6 cells were resuspended in 800 μl of Buffer A (10mM Hepes, pH8.0, 10mM KCl, 0.1 mM EDTA, 0.1 mM EGTA). Then, NP-40 (Sigma) (from a 10% NP-40 stock solution) was added to a final concentration of 0.5%. The cells were vortexed for 15 seconds and the nuclei were recovered by centrifugation at 4,000 rpm for 3 min (Centrifuge 5415D, Eppendorf). The pellet was resuspended in 100 μl of Buffer C (20mM Hepes, pH 8.0, 400 mM NaCl, 1mM EDTA, 1mMEGTA) and incubated for 20 min on ice. Every 4 minutes the contents were mixed. Samples were centrifuged 7 minutes at 14,000 rpm and the supernatant containing the nuclear protein extracts was recovered. The protein content was determined by the standard Bradford technique. Nuclear extracts were stored at -80°C. All buffers were supplemented with a cocktail of protease inhibitors (P-8465, Sigma). EMSA gels and running conditions were described in Sepulveda et al., 2004. Briefly, 5-10 μg of proteins were mixed with 5 μg of Poly-dIdC (27-7880-01, Amersham), the ds DNA competitors or antibodies, and the binding buffer (60% glycerol, 60mM Hepes, pH 8.0, 20mM Tris-Cl, pH 8.0, 250 mM KCl, 5mM EDTA, pH 8.0 5mM DTT) for 20 min on ice. The ^{32}P-labelled probe was then added (30,000-50,000 cpm) and an extra 30 min incubation was performed. Samples were run on 5% PAGE at 15-20mA at room temperature with 0.5x TBE (45 mM Tris base, 45 mM boric acid, 1 mM EDTA pH8.0). Gels were dried and exposed on MS-Kodak films for various amounts of time.

12. ds DNA competitors and Antibodies

The competitor for octamer binding proteins was 5'- CCA TTT GCA TAT TTG CAT ATT TGC ATC C-3' (Michaelson et al., 1996); for Nf-kB, 5'-GAG AGG GGA TTC CCC GAT TAG CTT TCG GGG AAT CCC CTC T-3' (Fujita et al., 1992); and for CTCF (FII), 5'- AGG CGC GCC CCC AGG GAT GTA ATT ACG TCC CTC CCC CGC TAG GGG GCA GCA GGC GCG CCT-3' (Bell et al., 2000). All oligonucleotides were annealed with their respective complementary sequences with the exception of the NF-κB site, which is self-complementary.

The following rabbit antibodies were used: anti-USF-1 (H-86) (sc-8983), anti-USF-2 (C-20) (sc-862) from Santa Cruz Biotechnologies and anti-CTCF

mix of monoclonal antibodies from Lobanenkov's lab.

13. Synthesis and methylation of probes

The probes for the human alleles were obtained by PCR using as templates the vectors where the alleles were cloned and the primers: AL for 5'-CCA GAA ATA GCT TGC ACG ATT CTC C – 3' and AL rev 5' – GTC CTG GGG GAG GGG -3'. The probes of 70 bp or less for the CpG screening were synthesized by Fisheroligo Scientific (Fisher Scientific Company, PA).

The probes were methylated by a specific CpG Methylase, M.Sss I (New England Biolabs, Ipswitch, MA). The conditions of reactions were 100 Units of M.SssI, 1X NE Buffer 2 (50mM NaCl, 10mM Tris-HCl, 10mM $MgCl_2$, 1 mM dithiothreitol), 160 μM S-adenosylmethionine and 1 μg of DNA. Each reaction was incubated over night at 37°C. To check the methylation state of probes, an aliquot of each fragment, treated by M.SssI was digested by methylation sensitive restriction enzymes, such as HpaII or MaeII (**Fig. 28**).

14. ^{32}P labeling of DNA fragments

Double stranded oligonucleotides were labeled with [γ-^{32}P]ATP (5.0 mCi/ml; Perkin Elmer) by using T4 polynucleotide kinase. 500 ng of dsDNA was incubated at 37°C for 1h in the presence of 10U of PNK (49215228, Roche) and 5μl of fresh [γ-^{32}P]ATP, in a total volume of 30 μl. The probe was purified by mini Quick Spin Columns (11814397001, Roche Applied Science) as protocol. Two μl of the solution was used to determine cpm/μl (Tricarb 2900 TR, Liquid Scintillation Analyzer, Packard). Probes were stored at -20°C.

RESULTS

1. *In silico* analysis of contigs from GenBank

The search of GenBank (GB) identified five DNA sequences of the human 3'RR-1, corresponding to the GenBank accession numbers: X76785 (Gualandi et al., 1995), Y14407 (Pinaud et al., 1997), AL928767, AL928765 and U64453 (Sadhu et al., 1997). The analysis of the overlaps allowed the reconstitution of a new assembled contig of 51 kb, called CHR77 (**Fig.12**). The assembled sequence corresponds to the complete 3'RR-1 in agreement with previous reports (Chen et al., 1997; Mills et al., 1997 and Chauveau et al., 1998).

A partial 3'RR-2 sequence (from Mills et al., 1997) allowed us to identify the GenBank accession no. AL928742 contig, which is 176 kb long and covers the four constant genes γ2, γ4 ε and α2 together with the complete 3'RR-2 (**Fig.12**). The 3'RR-2 region is represented in a few independent sequences as compared to 3'RR-1, but their colinearity with the AL928742 contig confirms the identity of the 3'RR-2. The reported contigs represent the most up to date information about this region of the human genome, except for the polymorphisms, described below.

The reconstitution of the 3'RR-1 as a continuous sequence allowed the direct comparison of the two heavy chain constant region. The analysis of the contigs confirmed the previous data obtained by genomic Southern analysis, which had shown the presence in both loci of internal duplications and inversions (**Fig.14**) (Mills et al., 1997) (Chen et Birshtein, 1997). The 3' termini of the duplication of the heavy chain constant region include the LTR but not the unique sequence of the ERV-K10 retrovirus present at the extreme 3' end of the 3'RR-1. The two loci present 16 conserved sequence units (U) and 6 repeated sequences (R), which are numbered starting from the poly-A site of the membrane exon of two constant genes (**Fig.13**). Three inversions are present and identify the two 3'RRs: the "Ua" or "Ub" sequences. In particular an ~4 kb region flanking HS1.2 is inverted in agreement with previous reports (Chen et al., 1997; Mills et al., 1997 and Chauveau et al., 1998). Interestingly rat and mouse HS1.2 sequences are inverted with respect to each other as in the two human *loci*. These specific sequences are present only in one of the two regions and were utilized to select the primers for PCR distinctive analyses (**Fig.14**).

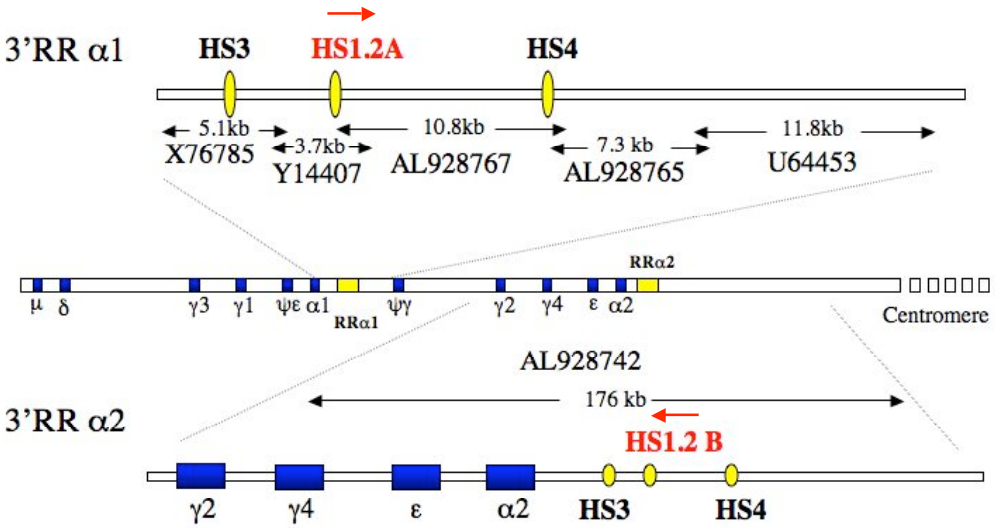


Fig. 12

Fig.12 Schematic map of genomic contigs used to study the human 3' Regulatory Regions. To study the two polymorphic enhancers, HS1.2A and HS1.2B (in red), genomic contigs of the constant heavy chain region were rescued from GenBank. For the 3'RR α1 sequence, a new sequence (called CHR77) was assembled from the junction of five genomic contigs (X76785, Y14407, AL928767, AL928765 and U64453), using the overlapping regions. The contig AL928742 was used the 3'RR α2 sequence.

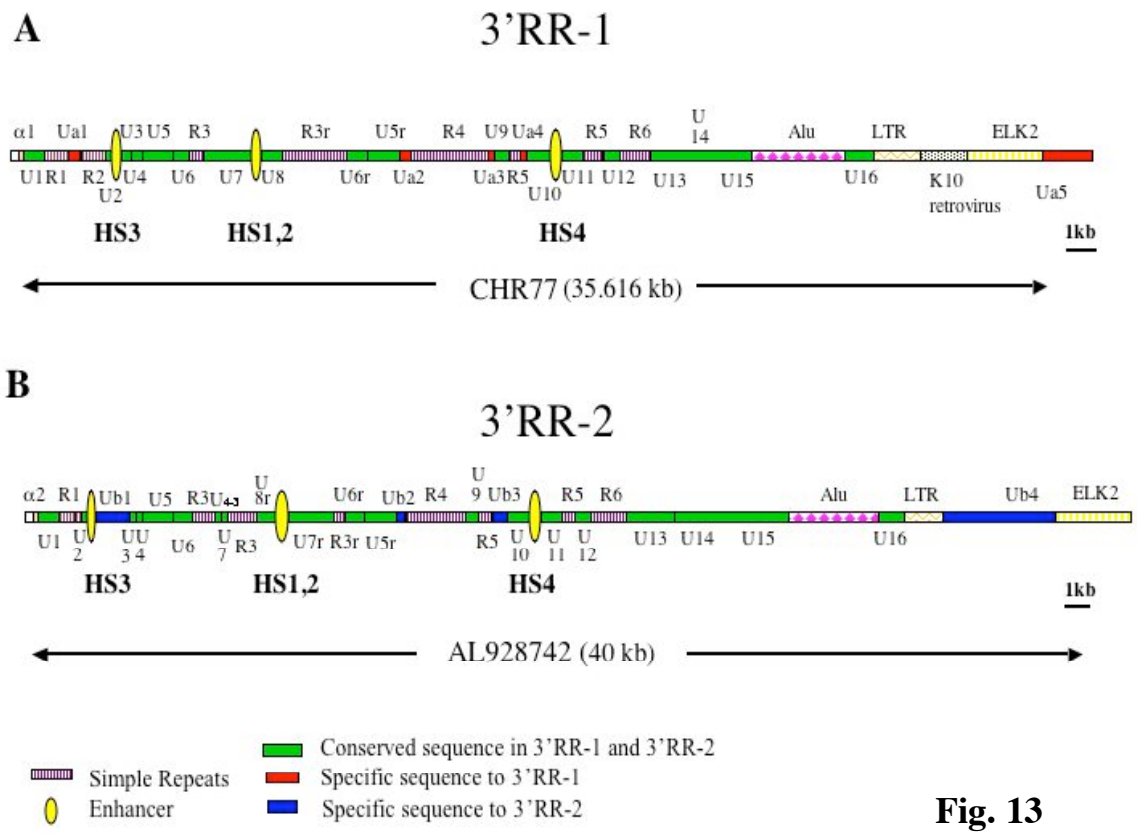


Fig. 13

Fig.13 Genomic organization of the two 3'RRs in the immunoglobulin heavy chain locus of humans: the 3'RR-1 (A) and the 3'RR-2 (B). The conserved sequence units (U), are labelled from left to right with increasing numbers starting from the poly-A site of the membrane Cα exons. Sequences with tandem repeats (R) are also numbered from left to right. The boxes, green if common to 3'RR-1 and 3'RR-2, blue or red with the letter U followed by a or b if unique for one of the 3'RRs. The r that follows the Arabic number indicates that the sequence is duplicated in a reverse form. The border of the duplication is inferred by the homology between the two 3'RRs and the interruption is signed on the right of both A and B enlargements. The yellow elliptic inserts represent the three enhancers of the 3'RRs.

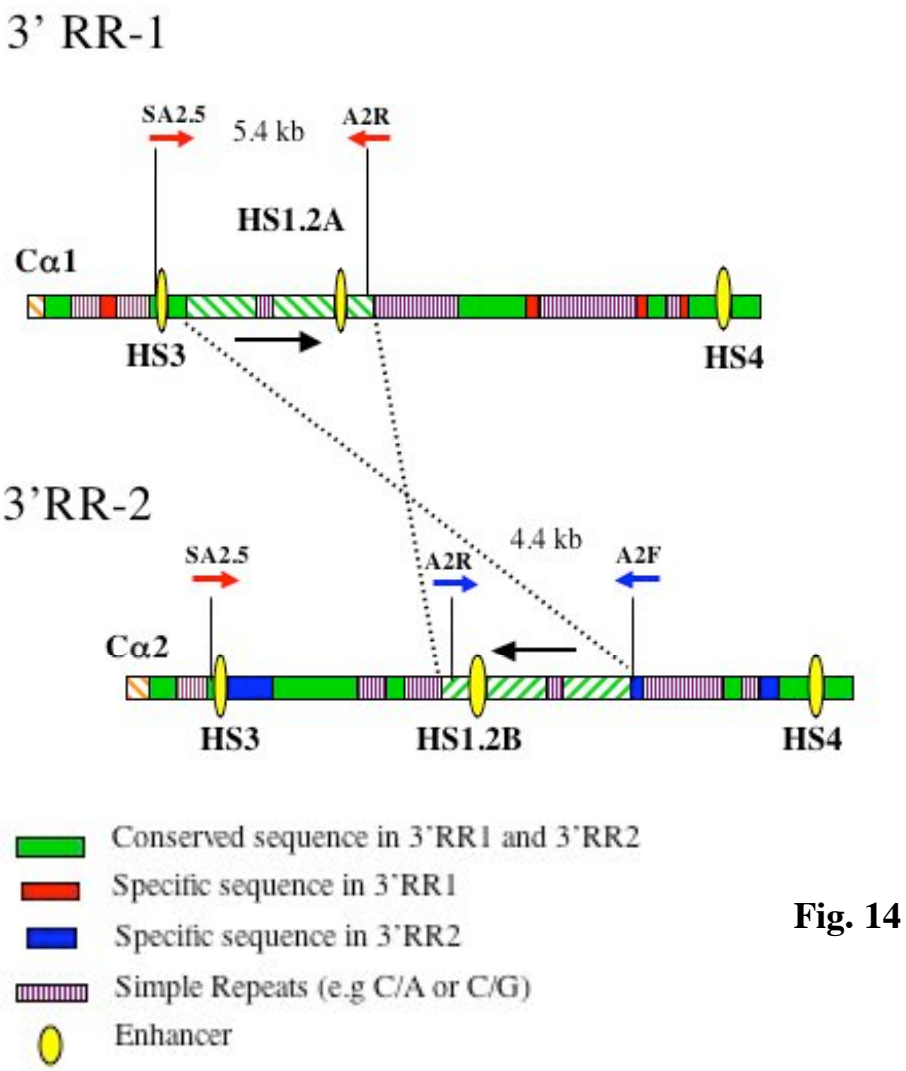


Fig. 14

Fig.14 HS1.2 in the 3'RR-1 is inverted with respect to HS1.2 in 3'RR-2. The inversion and the specific sequences of 3'RR-2 permit amplification of HS1.2-A and HS1.2-B specific fragments.

2. The polymorphisms of the HS1.2-A and HS1.2-B enhancers

In agreement with previous reports (Chen et al., 1997; Mills et al., 1997 and Chauveau et al., 1998), we also observed that the enhancer HS1.2-B (3'α2) is inverted in comparison with the enhancer HS1.2-A (3'α1) (**Fig.14**).

The selective amplification, cloning and sequence determination of HS1.2-A and –B enhancers revealed the occurrence of the alleles reported in **Fig.15** (GenBank accession no. AJ544218, AJ544219, AJ544220, AJ544221, AY530200 and AY530201). Each allele was sequenced twice from at least four clones from five independent genomes, either homozygous or heterozygous. The polymorphic sequences are present immediately downstream of the core of the enhancers. The four alleles of 3'RR-1 and the two alleles of 3'RR-2 have differences in the number of copies (one to four) of a conserved 38-bp element. The repetitions of the 38-bp element are separated by spacers of 20, 16 and 14 bp, respectively. In alleles 1A, 2A and 3A, the polymorphic sequences are bordered by two external 17-bp elements. In alleles 3B,4A and 4B, the external 17-bp element next to the core of enhancer HS1.2 is replaced by a 31-bp element (**Fig.15**). Moreover, the 38-bp element of these alleles (3B, 4A and 4B) presents a transition from T to C at the first position and a transition from C to T at the 15th nucleotide. No other polymorphic sites were detected. The differences do not alter the consensus for the transcription factors considered in this work and described later. Finally allele *4 is identical in both 3'RRs, but inverted.

The presence of different consensus sites for transcription factors in HS1.2 enhancers is related to different organizations of the four alleles. The repeated element of 38 bp is present from one to four unvarying copies in the four alleles, and the copies are separated by different spacers in both A and B loci, as shown in **Fig.16**. The 38-bp element presents one copy of the Nf-kB (P50) site, which consequently is present from one to four times in the four alleles. The spacer of 16 bp contains a consensus for the CMYB factor, which consequently is present in one copy in allele 2A, and in alleles 3 and 4 of both loci. The 31-bp external element specific to alleles 3B, 4A and 4B contains sites for MyoD and μE5. Since the alternative 17-bp external element does not present known *cis* elements, the replacement of this element with that of the 31 bp in allele 3B increases the number of binding sites for transcription factors of this 3B in comparison with allele

3A. This is the main difference between alleles 3A and 3B. For more details, see **Fig. 16**.

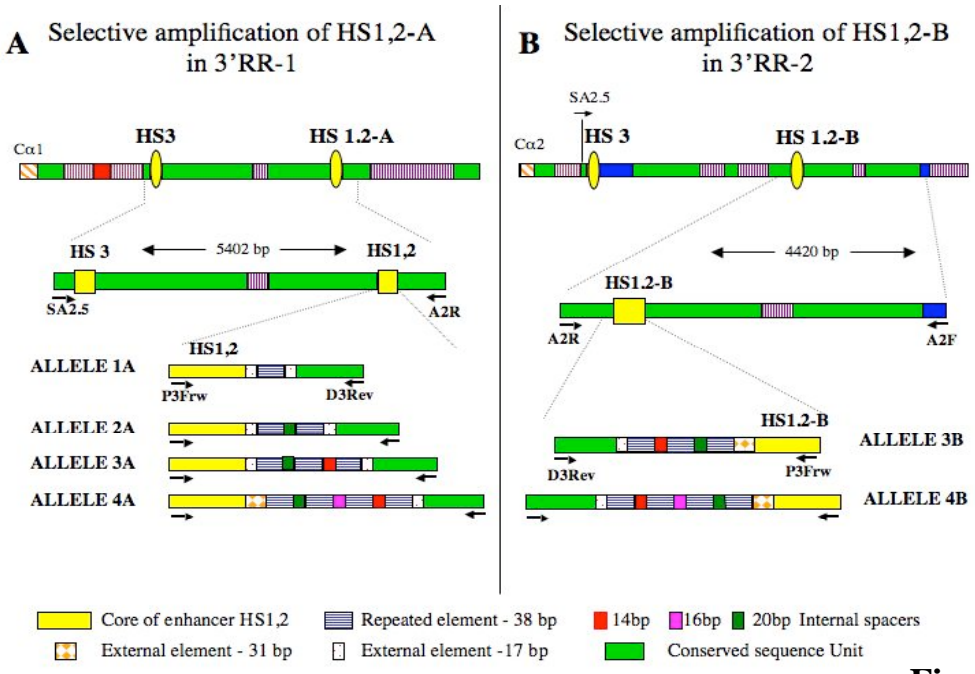
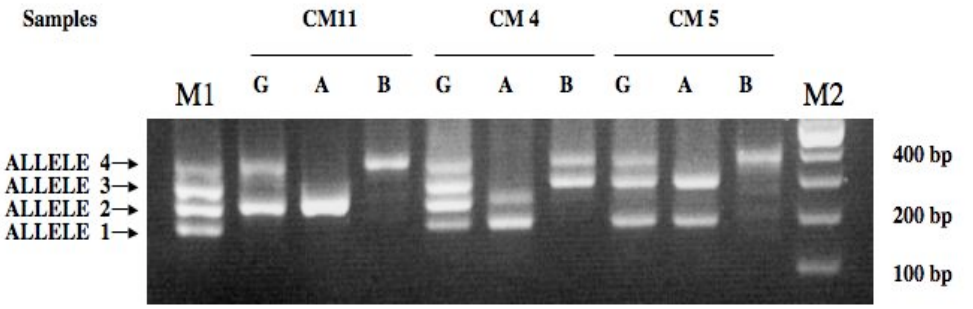


Fig. 15A

Fig.15A Strategy of PCR selective amplifications for the HS1,2-A and HS1,2-B loci and structure of the polymorphisms. The primer positions are indicated by the arrows. The HS1,2-B is inverted compared to HS1,2-A; therefore, the primer A2R is inverted in the two sequences and coupled with primer SA2,5 amplifies only the HS1,2-A. The primer A2F corresponds to the Ub2 unique sequence specific for 3'RR-2 and coupled to A2R primer amplifies only the HS1,2-B enhancer. In the lower part of the figure are represented schematically in part A the four alleles of HS1,2-A and in part B, the two alleles with inverted HS1,2-B. The details of the repeated elements of the different alleles are schematically depicted in the legend and show the different elements separating the 38-bp repeated sequence.



G: No selective amplification;
A: Selective amplification for HS1,2 A;
B: Selective amplification for HS1,2 B;
M1: Allelic Marker;
M2: Marker 100bp;

Fig. 15B

Fig.15B Agarose 3% gel electrophoresis of the different alleles of HS1,2-A and -B from three different genomic samples. Lane 1: M1, the standard size of the four alleles of a double heterozygous sample; lanes 2–9: genomic samples CM11/CM4/CM5, G, non selective amplification; A, selective amplification of HS1,2-A locus and B HS1,2-B locus. In the G lanes, the alleles from both loci are visible. In A or B lane, the alleles of selective amplification are visible. Lane 11: M2 is the size marker for the ladder of 100 bp.

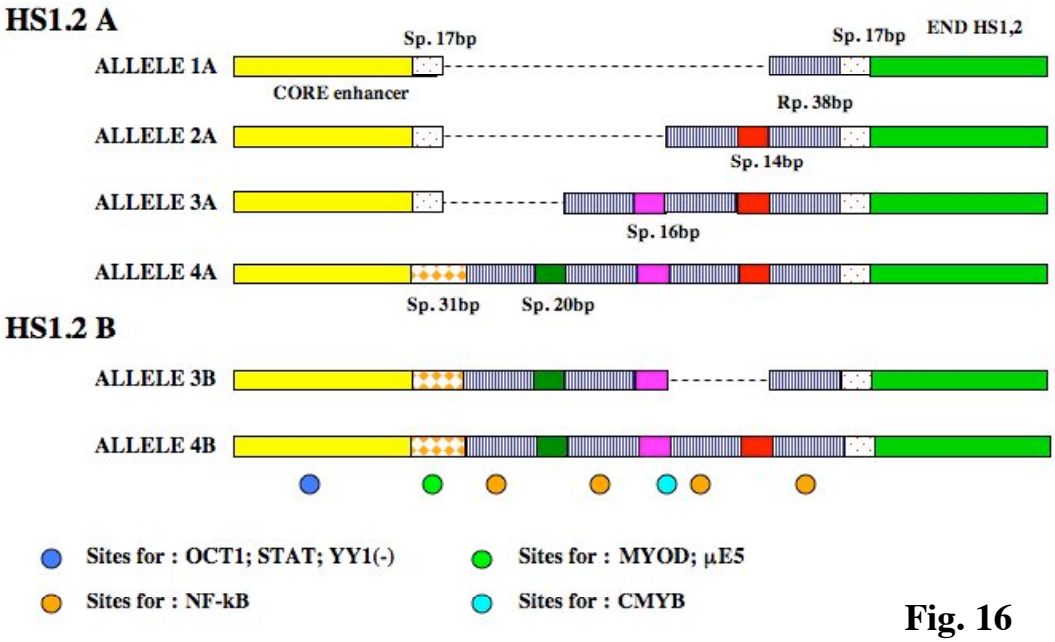


Fig. 16

Fig.16 Schematic representation of the six alleles piled up for the conserved and duplicated elements. The structure of the polymorphisms consists of the elements of 38-bp repeated and separated by spacers. The names of the transcription factors are reported in the legend with the corresponding symbols mapped on the alleles. The sequence of the consensus of the transcription factors reported varies from 9 to 20 bp and the corresponding sequences are fetched with WebGene program.

3. Changes of protein binding in the different HS1.2-A alleles by EMSA

To see if the polymorphism modifies nuclear protein binding of HS1.2-A enhancers, four independent EMSAs were done using the nuclear extracts from three different cell lines, FLEB (Pro B), JJN3 (Plasmacytoma) and SULTAN (Burkitt's lymphoma), and for each a different allele as probe (**Fig.17**). The data show that the protein binding changes. In fact, the binding to *2A, *3A and *4A alleles show one additional band with respect to allele *1A. This suggests that different nuclear proteins can bind to individual alleles. Therefore, the polymorphic region seems to alter the protein binding to HS1.2-A enhancer *in vitro*.

The Webgene predictions suggest that NF-κB is involved in these patterns of protein binding. Therefore, we checked the presence of NF-κB binding in allele *2 by EMSA using a specific DNA competitor (**Fig.17**). The probe was the repeat of allele *2 with predicted consensus for NF-κB. The data show that probably NF-κB binds the allele *2 *in vitro*. It reinforces the hypothesis that the polymorphism can change the binding for NF-κB in HS1.2-A enhancers and so modulate the enhancer activity.

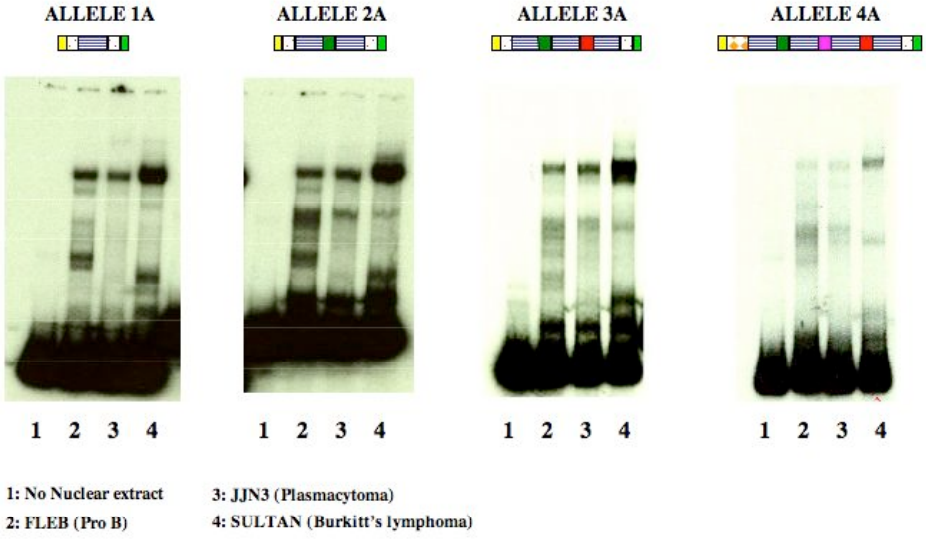


Fig. 17A

Fig. 17B

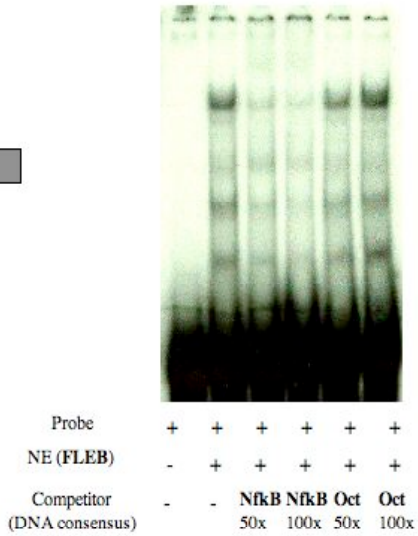


Fig.17-A) EMSA using the four HS1.2A alleles as probes and nuclear extracts of different cell lines. The figure shows four independent EMSA in which the four HS1.2-A alleles were used as probes. This experiment shows that different nuclear proteins bind to individual alleles. Thus the polymorphic region alters the protein binding of HS1.2-A enhancer *in vitro*. **B) *In vitro* NF-κB binds the allele *2**. EMSA using nuclear extract of FLEB (proB cell line), the first repeat of allele *2 as probe, and DNA consensus for NF-κB and Oct as competitors. The experiment shows that NF-κB binds the first repeat of allele *2 *in vitro*. This suggest the possibility that NF-κB can be involved in the activity of allele *2 for the Ig expression.

4. Analyses of HS1.2 in monkeys and apes

Southern blot evidence suggests that both the ψelk insertion and the large locus duplication occurred after the lineage divergence between New and Old World monkeys (Harindranath et al., 1998). Therefore, the nucleotide sequences of the enhancer HS1.2 of *C. moloch, G. gorilla* and *Pan troglodytes* (chimpanzee) which we isolated and sequenced (GenBank accession no. AY649406, AY655739 and AY669117) were compared to rodents (GenBank accession no X96607, AF314408 of mouse and rabbit, respectively) and humans (GenBank accession no. AJ544218). The alignment of **Fig. 18** shows the conservation of the core of the enhancer and the variability of the polymorphic region starting with the 31- or the 17-bp external spacers. Compared to man, only one base is substituted in *Gorilla* inside the core region of the enhancer. In the entire enhancer, *G. gorilla* and *P. troglodytes* have only 4 and 5 nucleotides of difference, respectively, where for *C. moloch* the differences rise to 29 bp. The apes so far analyzed have a conserved copy of the allele 1A, and the HS1.2-B locus has not yet been cloned but it is predicted to be present with the 3'RR-2 duplicated region. Other alleles have not yet been found in monkeys or apes. The 31-bp initial spacer, present in the human alleles of 3'RR-2, has been found only in the monkey but not in the apes which were analyzed. Genomic southern analysis of 5 *M.fascicularis* genomic DNAs, digested with *XbaI,* detected polymorphism of the HS1.2 region (data not shown).

Fig.18 The alignment of the HS1.2 enhancer sequence from different species. The alignment starts with the corresponding number of the sequences reported in the GenBank. The accession numbers in order are as follows: X96607 from nt. 5125 to 5275 (mouse), AF314408 from nt. 729 to 546 (rabbit), AY649406 from nt. 1 to 292 (*C. moloch*), AY655739 from nt. 1 to 290 (*G. gorilla*), AY669117 from nt. 1 to 286 (*P. troglodytes*) and AJ544218 from nt. 1 to 286 (human). The different conserved DNA elements are ahown by coloured boxes. The core of the enhancer (pink) is highly conserved in the six species analysed. The 17-bp element (blue) on both sides of the 38-bp element present in humans and apes is absent from mouse, rabbit and *Callicebus.* The 31-bp element (orange) substitutes, the 5V 17-bp element and is conserved in *Callicebus* and is similar in rabbit. The 38-bp element (yellow) is conserved in mouse and rabbit and highly conserved from *Callicebus* to man and is the structure that gives rise to the duplications of the polymorphisms. The symbol * corresponds to nucleotides conserved in all mammalian species reported.

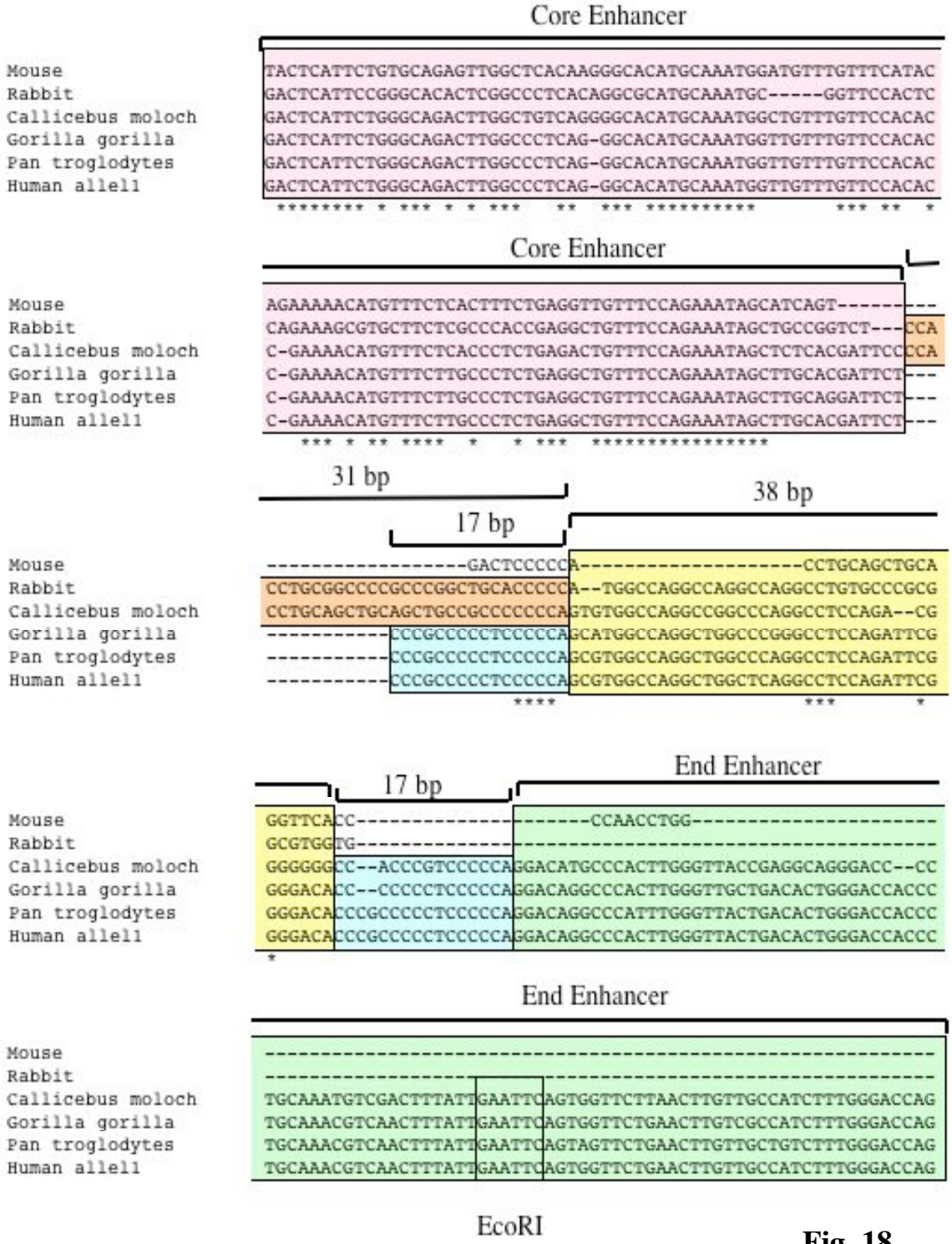

Core Enhancer
Mouse
Rabbit
Callicebus moloch
Gorilla gorilla
Pan troglodytes
Human allel1
TACTCATTCTGTGCAGAGTTGGCTCACAAGGGCACATGCAAATGGATGTTTGTTTCATAC
GACTCATTCCGGGCACACTCGGCCCTCACAGGCGCATGCAAATGC-----GGTTCCACTC
GACTCATTCTGGGCAGACTTGGCTGTCAGGGGCACATGCAAATGGCTGTTTGTTCCACAC
GACTCATTCTGGGCAGACTTGGCCCTCAG-GGCACATGCAAATGGTTGTTTGTTCCACAC
GACTCATTCTGGGCAGACTTGGCCCTCAG-GGCACATGCAAATGGTTGTTTGTTCCACAC
GACTCATTCTGGGCAGACTTGGCCCTCAG-GGCACATGCAAATGGTTGTTTGTTCCACAC
Core Enhancer
Mouse
Rabbit
Callicebus moloch
Gorilla gorilla
Pan troglodytes
Human allel1
AGAAAAACATGTTTCTCACTTTCTGAGGTTGTTTCCAGAAATAGCATCAGT----------
CAGAAAGCGTGCTTCTCGCCCACCGAGGCTGTTTCCAGAAATAGCTGCCGGTCT---CCA
C-GAAAACATGTTTCTCACCCTCTGAGACTGTTTCCAGAAATAGCTCTCACGATTCCCCA
C-GAAAACATGTTTCTTGCCCTCTGAGGCTGTTTCCAGAAATAGCTTGCACGATTCT---
C-GAAAACATGTTTCTTGCCCTCTGAGGCTGTTTCCAGAAATAGCTTGCAGGATTCT---
C-GAAAACATGTTTCTTGCCCTCTGAGGCTGTTTCCAGAAATAGCTTGCACGATTCT---
31 bp
38 bp
17 bp
Mouse
Rabbit
Callicebus moloch
Gorilla gorilla
Pan troglodytes
Human allel1
----------GACTCCCCCA----------CCTGCAGCTGCA
CCTGCGGCCCCGCCCGGCTGCACCCCCA--TGGCCAGGCCAGGCCAGGCCTGTGCCCGCG
CCTGCAGCTGCAGCTGCCGCCCCCCCAGTGTGGCCAGGCCGGCCCAGGCCTCCAGA--CG
----------CCCGCCCCCTCCCCCAGCATGGCCAGGCTGGCCCGGGCCTCCAGATTCG
----------CCCGCCCCCTCCCCCAGCGTGGCCAGGCTGGCCCAGGCCTCCAGATTCG
----------CCCGCCCCCTCCCCCAGCGTGGCCAGGCTGGCTCAGGCCTCCAGATTCG
End Enhancer
17 bp
Mouse
Rabbit
Callicebus moloch
Gorilla gorilla
Pan troglodytes
Human allel1
GGTTCACC----------CCAACCTGG----------
GCGTGGTG----------
GGGGGGCC--ACCCGTCCCCCAGGACATGCCCACTTGGGTTACCGAGGCAGGGACC--CC
GGGACACC--CCCCCTCCCCCAGGACAGGCCCACTTGGGTTGCTGACACTGGGACCACCC
GGGACACCCGCCCCCTCCCCCAGGACAGGCCCATTTGGGTTACTGACACTGGGACCACCC
GGGACACCCGCCCCCTCCCCCAGGACAGGCCCACTTGGGTTACTGACACTGGGACCACCC
End Enhancer
Mouse
Rabbit
Callicebus moloch
Gorilla gorilla
Pan troglodytes
Human allel1
TGCAAATGTCGACTTTATTGAATTCAGTGGTTCTTAACTTGTTGCCATCTTTGGGACCAG
TGCAAACGTCAACTTTATTGAATTCAGTGGTTCTGAACTTGTCGCCATCTTTGGGACCAG
TGCAAACGTCAACTTTATTGAATTCAGTAGTTCTGAACTTGTTGCTGTCTTTGGGACCAG
TGCAAACGTCAACTTTATTGAATTCAGTGGTTCTGAACTTGTTGCCATCTTTGGGACCAG
EcoRI

Fig. 18

5. The frequency of the HS1.2 alleles

The frequency of the different alleles of enhancers HS1.2-A and –B has been determined in two samples of independent healthy donors from Italy and Cameroon. In the Italian group, alleles 1A and 2A are respectively present with a frequency of 0.44 and 0.39. Allele 3A is the most rare, with a frequency of 0.06, whereas allele *4A is present with a frequency of 0.10. In Cameroon, the distribution of four HS1.2-A alleles is different with respect to Italian distribution. In the African population, the allele 2A frequency is significantly decreased, from 0.395 in Italy to 0.087 in Cameroon. For each group, the χ^2 analysis is highly significant for equilibrium test (**Fig.19**).

For HS1.2-B, alleles 4B and 3B are represented with the respective sequences of 0.93 and 0.07 in Italy and 0.96 and 0.03 in Cameroon. The HS1.2-B enhancer has been described in a sample of French population and shown to be present only in the form of allele 4 (Pinaud et al., 1997). Alleles 1 and 2 were not detected for enhancer HS1.2-B but this might depend on the size of the analyzed sample.

The frequency of the different alleles of enhancers HS1.2-A has been determined in other populations of different geographic areas. In **Fig.20,** we report the frequencies of fourteen populations. In all groups, the observed genotypes are in agreement with those expected under Hardy-Weinberg equilibrium. The allele *3A shows the highest frequency among Africans populations. Allele *2A is significantly lower in African in comparison with both European and, to a lesser extent, Asian populations. The allele 4A is also higher in Africans while allele 1A does not seem to differentiate the major human population groups. An analysis of molecular variance (AMOVA) shows that HS1.2-A polymorphism is a useful anthropological marker.

The demonstration that HS1.2-A allele *2 has a role in the progression of IgA nephropathy (Aupetit et al., 2000), suggests that its geographic distribution and its low incidence among African populations could correlate with the fact that European, Asian and Asian derived populations are more prone to Berger disease (Levy et al., 1996;Hall et al., 2004).

A

HS1.2 ALLELES	FREQUENCIES OF HS1.2-A IN ITALY	FREQUENCIES OF HS1.2-B IN ITALY	FREQUENCIES OF HS1.2-A IN CAMEROON	FREQUENCIES OF HS1.2-B IN CAMEROON
1	0.439 ± 0.022	/	0.442 ± 0.038	/
2	0.395± 0.021	/	0.087 ± 0.021	/
3	0.061 ± 0.011	0.070 ± 0.018	0.250 ± 0.033	0.035 ± 0.018
4	0.105 ± 0.014	0.930 ± 0.018	0.221 ± 0.032	0.965 ± 0.018
Total sample	248	100	86	72
χ^2	1.47	0.02	0.65	0.02
P	0.750 < P< 0.500	0.900	0.975	0.900

B

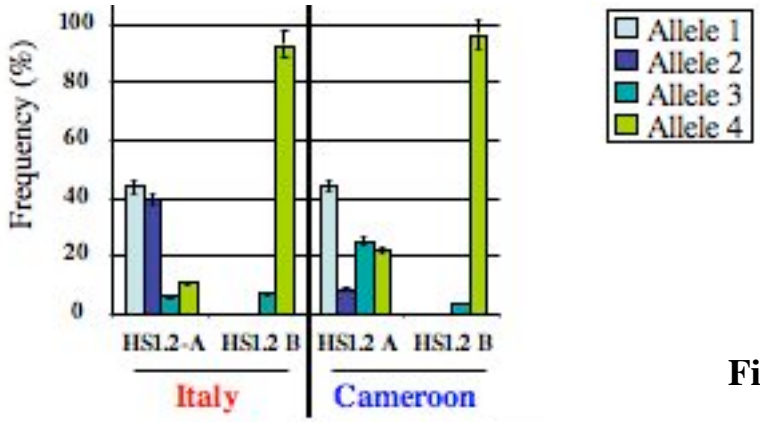


Fig. 19

Fig.19 Allelic distribution of HS1.2-A and HS1.2-B enhancers in two different populations. **(A)** Table of allelic frequencies of the enhancers, HS1.2-A and -B, in Italy and Cameroon. **(B)** Graph of allelic frequencies, expressed in per cent.

Populations	Total	Alleles			
		* 1A	* 2A	* 3A	* 4A
Benin	100	0.590 ± 0.035	0.000 - 0.01	0.245 ± 0.030	0.165 ± 0.026
Cameroon	86	0.442 ± 0.038	0.087 ± 0.021	0.250 ± 0.033	0.221 ± 0.032
Burkina	101	0.322± 0.033	0.049 ± 0.015	0.416± 0.035	0.213 ± 0.029
Tuareg	100	0.535 ± 0.035	0.150 ± 0.025	0.115 ± 0.022	0.200 ± 0.028
Marocco	95	0.531 ± 0.036	0.184 ± 0.028	0.116 ± 0.023	0.168 ± 0.027
Libya	102	0.431 ± 0.035	0.363 ± 0.034	0.073 ± 0.018	0.132 ± 0.024
Mongolia	95	0.658 ± 0.034	0.263 ± 0.032	0.016 ± 0.009	0.063 ± 0.018
Siberia	70	0.650 ± 0.040	0.279 ± 0.038	0.014± 0.010	0.057 ± 0.020
Ladakh	86	0.558 ± 0.038	0.366± 0.037	0.012 ± 0.008	0.064 ± 0.019
Iceland	96	0.536 ± 0.036	0.307 ± 0.033	0.036 ± 0.013	0.120 ± 0.023
Bulgaria	95	0.363 ± 0.035	0.558 ± 0.036	0.016 ± 0.009	0.063 ± 0.018
Turkey	103	0.563 ± 0.035	0.383± 0.034	0.024 ± 0.011	0.029 ± 0.012
Spain	104	0.457± 0.034	0.346 ± 0.033	0.058 ± 0.016	0.139 ± 0.024
Italy	248	0.439 ± 0.022	0.395 ± 0.021	0.061 ± 0.011	0.105 ± 0.014

Fig. 20A

Fig.20 Allelic distribution of HS1.2-A enhancer in some populations of different geographic areas. **(A)** Table of allelic frequencies of HS1.2-A enhancer in fourteen populations.

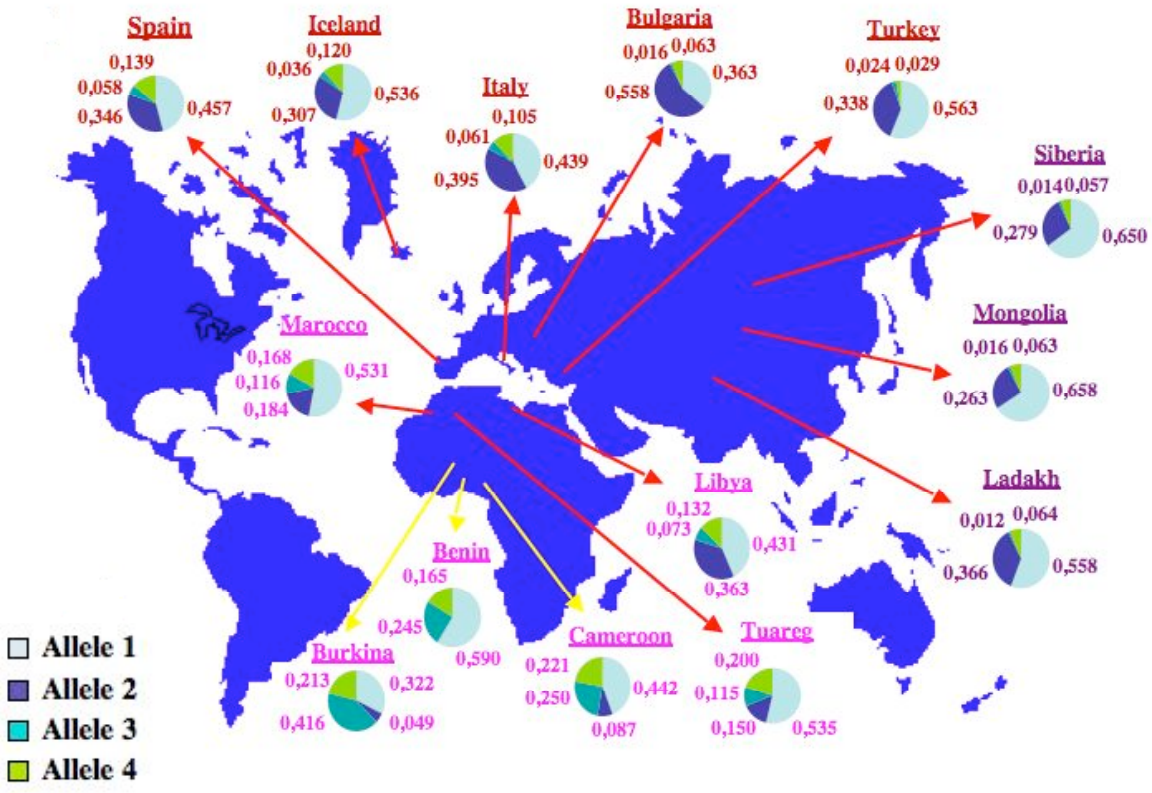


Fig. 22B

Fig.20 Allelic distribution of HS1.2-A enhancer in some populations of different geographic areas. (B) Graph of allelic frequencies.

6. The frequency of the HS1.2-A alleles in some autoimmune diseases

The demonstration that HS1.2-A alleles have a role in the progression of IgA nephropathy (Aupetit et al., 2000), a disease with immunoglobulin deregulation, prompted us to evaluate the influence of polymorphic enhancer HS1.2 in the development of some systemic and organ target autoimmune or diseases.

We have studied two systemic (Lupus erythematosus and systemic sclerosis) and five organ target autoimmune diseases (Crohn's disease, Celiac disease, psoriasis, Herpetiform dermatitis and Hashimoto's disease). As a control, we analyzed a healthy population and patients affected by ulcerative colitis, a non autoimmune disease. We also monitored the allelic distribution of HS1.2-A in a group of patients with high levels of immunoglobulin. This group contained patients with disorders unrelated to an immunological etiology, but, however, at least one class of immunoglobulin out the standard levels. All samples are from the same geographic area from individuals who are not related.

The distribution of allelic frequencies is reported in **Fig.21**. We found that allele *2 of the HS1,2-A enhancer was significantly more frequent in all systemic disorders that we have considered, in herpetiform dermatitis and in patients with high levels of Ig, with respect to healthy controls. The increase was higher in patients with psoriasis where the frequency of allele *2 was 0.663 with respect to 0.395 in healthy controls. Patients with psoriatic arthritis also had a higher frequency of the allele *2 (from 0.59 to 0.75), in the 28 subjects recruited, but the limited sample analyzed did not allow to determine statistical significance. Similarly, herpetiform dermatitis patients in the presence of celiac disease showed the frequency of allele *2 increased from 0.50 to 0.67.

The analysis of genotypes for the four alleles of the HS1,2-A enhancer showed a highly significant increase of frequency of the allele *2 in the systemic sclerosis cohort *vs* controls (57% vs 40% $p<0.0001$). The frequency to find both *2 alleles was increased in ACA+ patients (42%) and anti-Scl70+ patients (31%) compared to the control group (15%) (not shown).

Finally, the frequency of allele *2 increased in patients with high levels of immunoglobulin, suggesting that this allele may improve the production of antibodies.

These data strongly support the idea that allele *2 confers a risk to develop many autoimmune diseases and provides novel evidence of a

genetic predisposition in autoimmune disorders and a possible candidate for a non-HLA locus. This marker may also be relevant in determining some associated conditions.

Samples	Total	Alleles			
		*1A	*2A	*3A	*4A
Control (Italian Healthy)	248	0.439 ± 0.022	0.395 ± 0.021	0.061 ± 0.011	0.105 ± 0.014
Crohn's disease	58	0.353± 0.044	0.595 ± 0.045	0.009 ± 0.009	0.043 ± 0.019
Celiac disease	115	0.335 ± 0.031	0.517 ± 0.033	0.009 ± 0.06	0.139 ± 0.023
Lupus erythematosus	59	0.364 ± 0.44	0.559 ± 0.046	0.034 ± 0.017	0.042 ± 0.018
Systemic sclerosis	155	0.410 ± 0.028	0.558 ± 0.028	0.006 ± 0.004	0.026 ± 0.009
Psoriasis	86	0.279 ± 0.034	0.663 ± 0.036	0.012 ± 0.008	0.046 ± 0.016
Herpetiform dermatitis	38	0.316 ± 0.053	0.618 ± 0.056	0.013 ± 0.013	0.053 ± 0.026
Hashimoto's disease	32	0.531 ± 0.062	0.406 ± 0.061	0.000 ± 0.000	0.062 ± 0.030
Ulcerative colitis	34	0.470 ± 0.060	0.485 ± 0.061	0.029 ± 0.020	0.015 ± 0.015
Patients with high levels of Ig	53	0.311 ± 0.045	0.528 ± 0.048	0.066 ± 0.024	0.094 ± 0.028

Fig. 21A

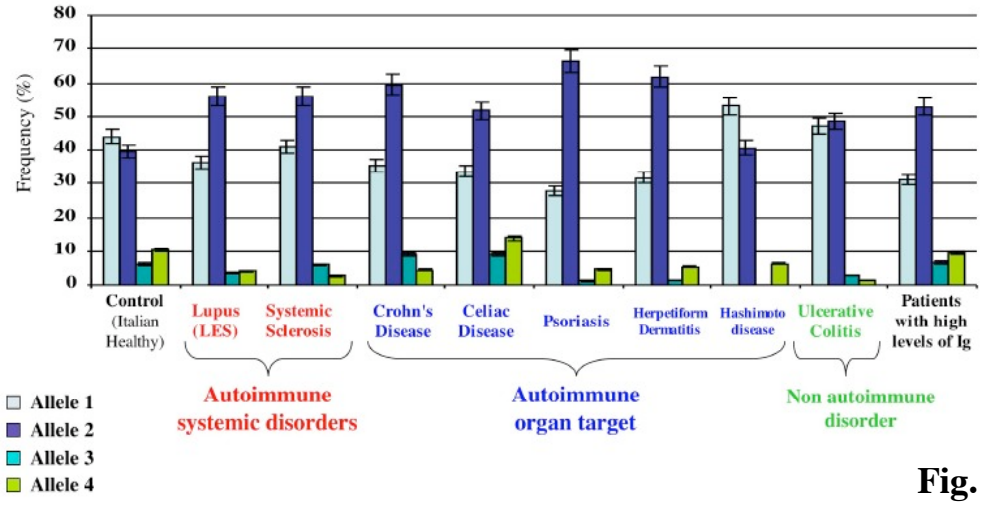


Fig. 21B

Fig.21 Allelic distribution of HS1.2-A enhancer in patients with immunological disorders. **(A)** Table of allelic frequencies of HS1.2-A enhancer . **(B)** Graph of allelic frequencies, expressed in per cent. The distinction in non autoimmune, autoimmune systemic and autoimmune organ target disorders is in agreement with the Encarta classification (http://it.encarta.msn.com/encyclopedia_761563416/Malattie_autoimmuni.html).

7. The frequency of the HS1.2-A alleles and humoral antibodies in response to neuroleptics

During antipsychotic therapy with neuroleptics, the specific TH-1–related immune answer becomes activated, and in addition the B cell system and antibody production increases in some patients (Muller et al., 2000). Thus it is possible that the different responses to neuroleptics can be related to different genetic backgrounds. For this reason, we have studied the distribution of HS1.2-A in patients with a schizophrenic disorder and treated with an antipsychotic therapy.

We divided the schizophrenic patients in four subgroups: 1) patients with standard levels of immunoglobulin; 2) patients with at least one class of Ig, higher with respect to standard levels; 3) patients with at least one class of Ig, lower with respect to standard levels and 4) patients with at least one class of Ig, lower and one higher with respect to standard levels. The distribution of allelic frequencies is reported in **Fig.22**. We found that the allele *1 of the HS1.2-A enhancer was significantly more frequent in patients with lower levels of Ig. In fact the frequency of allele *1 is 0.660 with respect to 0.439 in the healthy control and to 0.438 in all patients. The frequency of allele *2 increases with respect to control in all three subgroups.

These data suggest the idea that different responses to antipsychotic therapy can be related to different genetic backgrounds in schizophrenic patients.

Samples	Total	Alleles			
		*1A	*2A	*3A	*4A
Control (Italian Healthy)	248	0.439 ± 0.022	0.395 ± 0.021	0.061 ± 0.011	0.105 ± 0.014
Total schizophrenics	88	0.438 ± 0.033	0.473 ± 0.033	0.018 ± 0.009	0.071 ± 0.017
Schizophrenics with **normal** levels of Ig	17	0.382 ± 0.083	0.559 ± 0.085	0.029 ± 0.029	0.029 ± 0.029
Schizophrenics with **high** levels of Ig	29	0.345 ± 0.062	0.465 ± 0.065	0.034 ± 0.024	0.155 ± 0.047
Schizophrenics with **low** levels of Ig	25	0.660 ± 0.067	0.300 ± 0.065	0.020 ± 0.020	0.020 ± 0.020
Schizophrenics with **high and low** levels of Ig	17	0.323 ± 0.080	0.559 ± 0.085	0.000 ±	0.118 ± 0.055

Fig. 22

Fig.22 Allelic distribution of HS1.2-A enhancer in patients with schizophrenic disorder and different immunological patterns after treatment with an antipsychotic therapy. **(A)** Table of allelic frequencies of HS1.2-A enhancer.

8. DNA unmethylation patterns of 3'RR in murine cell lines reflecting different stages of B development

To see how the protein binding can be regulated in 3' RR, we have studied the state of DNA methylation in some CpGs of 3'RR in murine cell lines that reflect different stages of B development. DNA methylation can modulate directly or indirectly the activity of important factors for B cell development, such as Pax5 and Ets proteins (Maier et al., 2003), and also plays an important role in the regulation of factors, such as CTCF which is involved in imprinting processes (Bell et al., 2000) and insulator activity (Ishihara et al., 2006).

Therefore, we have studied the DNA methylation state of CpGs present in restriction sites for methylation sensitive endonucleases, such as HpaII and MaeII. The genomic DNA of different cell lines was digested by HpaII or MaeII and then used as a template to amplify the sequences flanking different restriction sites (**Fig23**). The signals of PCR products were quantified by Chemimager V 5.5 software and then compared with that obtained in the same way from undigested DNA. The comparison between digested and undigested DNA was expressed as percent of "unmethylation" where 100% means a fully unmethylated site 0% identifies a fully methylated site. In each graph, the horizontal axis shows the position of the analyzed CpGs and the vertical axis shows the corresponding percents of "unmethylation".

Fig. 24 shows the graph for ES cell WW6. In WW6, the 3'RR is fully methylated. In fact the percent of "unmethylation" does not exceed the 20% for any of the CpGs.

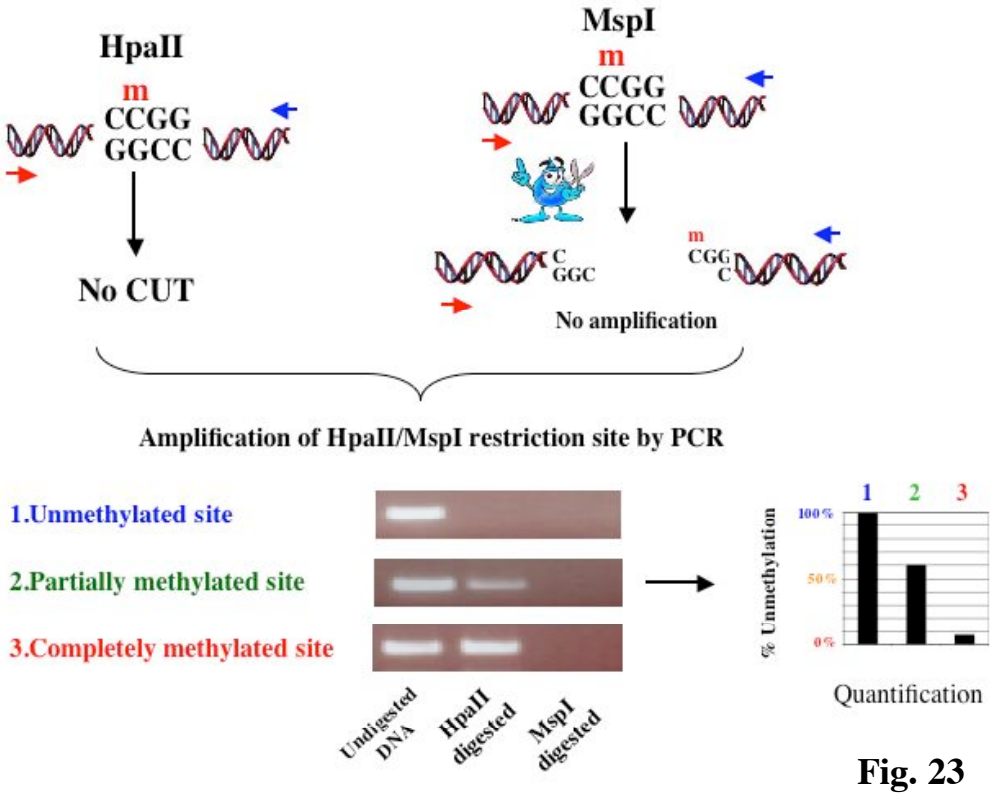


Fig. 23

Fig.23 How we study DNA methylation. DNA methylation was studied by digesting genomic DNA with methylation-sensitive restriction enzymes, HpaII and MaeII (not shown). These enzymes do not cut methylated CpG sites. After digestion, fragments were amplified and the intensity of signals was analyzed using densitometry imaging (Chemimager software). Amplification signals were normalized (see material and methods), compared between digested and undigested DNA and reported as percent of "Unmethylation". MspI was used as a control of HpaII digestion because it is a HpaII isoschizomer and not methylation-sensitive.

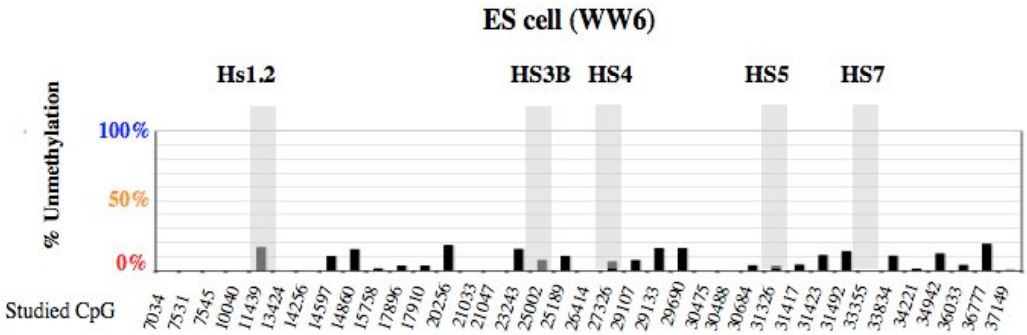


Fig. 24

Fig.24 The murine 3'RR is fully methylated in ES cell line (WW6). The graph shows the comparison between the PCR products obtained from the digested and undigested DNA for each considered CpG. It was expressed as percent of "unmethylation" where 100% means a fully unmethylated site and a fully methylated site has a value of 0%. The horizontal axis shows the different analyzed CpGs (the numbers refer to position in BAC199M11, GenBank Accession Number AF450245) and the vertical axis shows the corresponding percent of "unmethylation". The dark boxes indicate the sites inside the different DNaseI hypersensitve sites of murine 3'RR, except for HS3A. WW6 ES cells present a fully methylated 3'RR.

To see how the DNA methylation state of 3'RR may change during B cell development, we analyzed the genomic DNA of cell lines reflecting different stages of B cells. The **fig.25** shows that CpG sites are selectively undermethylated during B cell development. In particular, the site related to HS4 is unmethylated in all stages of B cell maturation. The sites inside and between HS5 and HS7 became unmethylated after pre B cells. Sites related to HS1.2 and HS3B are unmethylated only in a plasma cell line. It suggests that DNA methylation can be related to B cell development in some CpG sites.

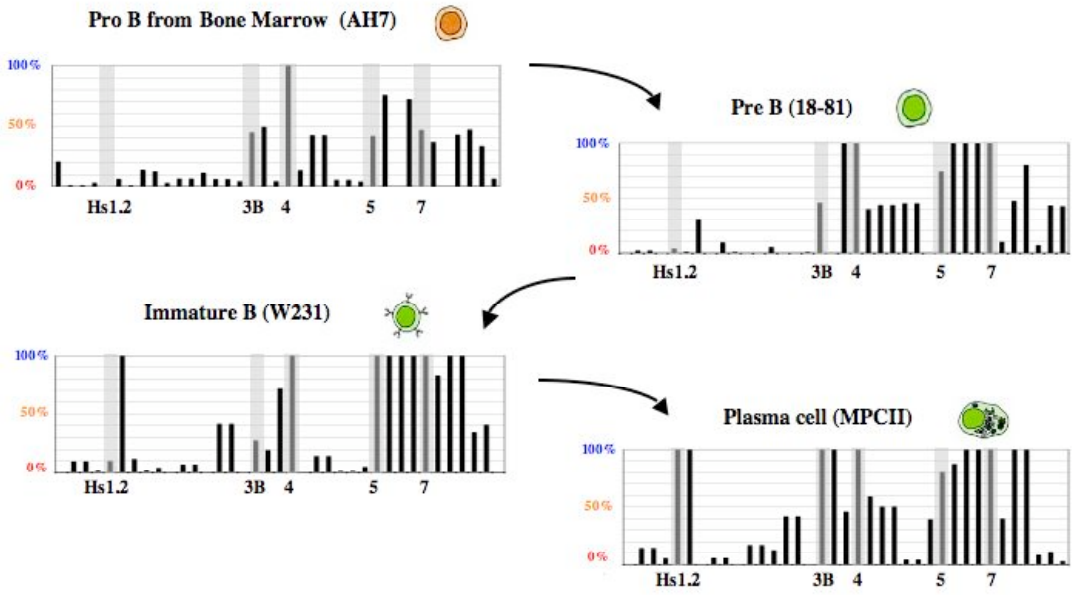


Fig. 25

Fig.25 DNA unmethylation patterns in cell lines reflecting different stages of B development. The site related to HS4 is unmethylated in all stages of B cell maturation. Sites inside and between HS5 and HS7 became unmethylated in pre B cells. Sites related to HS1.2 and HS3B are unmethylated only in a plasma cell line. This suggests that CpG sites are selectively demethylated during B cell development.

9. DNA methylation patterns of 3'RR in unstimulated and stimulated spleenic B cells

The previous data suggest that the DNA methylation is a selective process for the murine 3'RR. Therefore, we asked if the DNA methylation of 3'RR in B cells can be influenced by a specific stimulation for class switching recombination. We, therefore, stimulated splenic B cell by LPS, or LPS plus IL4 to activate the class switching, respectively for IgG3 - IgG2b, or IgG1 - IgE, in a specific way. We have also stimulated splenic T cells by αCD3, as a control, to determine whether changes in the DNA methylation of 3'RR were specific for B cells.

In **Fig.26** the data are reported. They show that the stimulations decrease the DNA methylation levels in different ways. These modifications are specific for B cells. In fact the stimulated T cells do not show the same patterns of stimulated B cells. In unstimulated B cells, only the site in HS4 is partially unmethylated, whereas, except one site between HS5 and HS7, the others are fully methylated. The stimulation by LPS reduces the DNA methylation level in the sites inside all hypersensitive sites, i.e. enhancers and insulators. Other sites downstream and upstream of DNaseI hypersensitive sites also became less methylated. LPS plus IL4 stimulation has more effect in the region downstream of HS4. In particular the sites between HS5 and HS7 became fully unmethylated. These data support the idea that B cell specific stimulation can influence the DNA methylation pattern of 3'RR and create specific epigenetic changes, related to the different stimulations.

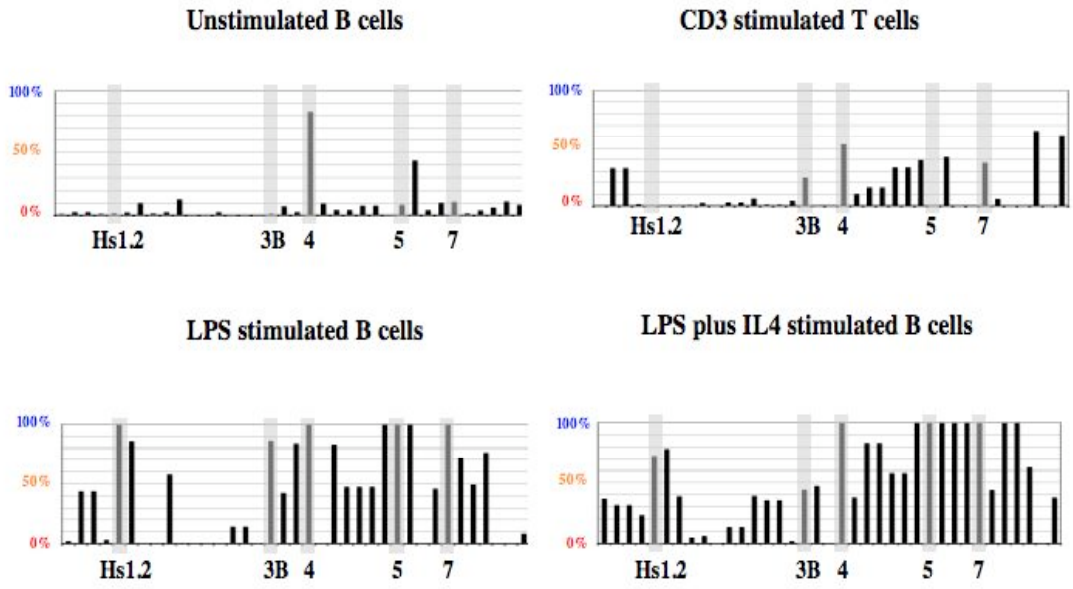


Fig. 26

Fig.26 DNA unmethylation patterns in unstimulated and stimulated spleen B cells and in stimulated resting T cells. The site related to HS4 is unmethylated before and after two independent B cell stimulation. On the contrary, the sites related to HS1.2, HS3B, HS5 and HS7 are less methylated only after the stimulation in B cells, but not in T cells. This suggests that B cell specific stimulation can influence the DNA methylation state of 3'RR.

10. DNA methylation patterns of 3'RR in non B cells

Our data show that CpG sites in the 3'RR are selectively undermethylated during B cell development and in response to the different stimuli. Therefore, to see which sites are specifically unmethylated in B cells, we studied the DNA methylation patterns of 3'RR in different non B cells. We used hematopoietic cells, such as Bac1.2f5 macrophage cell line, MEl erythroid cell line, BW5147 T cell line and peritoneal cells, extracted from mouse, and from liver.

Fig.27 shows that 3'RR has segments of undermethylation in all these cells, screened. This is in contrast to the fully methylated region in ES cells (**Fig24**). Some CpG sites are fully or partially unmethylated. However, DNA methylation patterns are different for all cells. In fact, some sites, located between HS1.2 and HS3B, are fully unmethylated in all hematopoietic cells but not in liver cells or in B cells. The CpGs in the DNaseI hypersensitive sites also show different patterns, except for HS4. In fact, in each non B cell, the site in HS4 does not appear fully unmethylated, such as in B cell lines (**Fig.25**). In HS1.2, instead, we see unmethylation in macrophages, erythroid and peritoneal cells, but not in liver and T cells; in HS5 the CpG site is fully unmethylated only in T cells, and, finally, HS7 is less methylated in erythroid and liver cells.

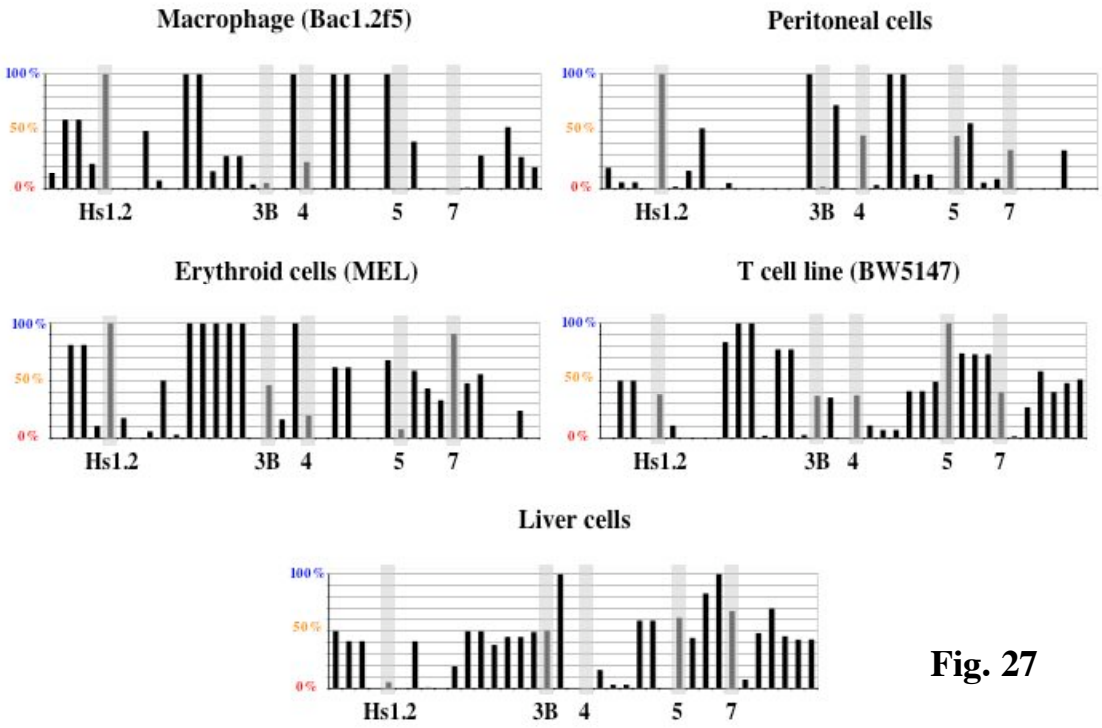


Fig. 27

Fig.27 DNA methylation patterns in non B cells. 3'RR has segments of undermethylation in all non-B cells, screened. This is in contrast to the fully methylated region in ES cells (**Fig24**). Some CpG sites are fully or partially unmethylated in different way for all cells.

11. DNA methylation at specific sites in 3'RR requires histone H1

Linker histone H1 plays an important role in chromatin folding. H1 depletion causes dramatic chromatin structure changes, including decreased global nuclesome spacing, reduced local chromatin compaction, and decreases in certain core histone modification. A recent work has shown that H1 depletion affects the expression of only a small number of genes. Many of the affected genes are imprinted and are, therefore, normally regulated by DNA methylation. These results indicate that histone H1 can participate in epigenetic regulation of gene expression by contributing to the maintenance or establishment of specific DNA methylation patterns (Fan et al., 2005).

For this reason, we have studied the DNA methylation pattern of 3'RR in H1 knock out ES cells (**Fig.28**). The data show that ES cells derived from histone H1 depleted mice showed a reduction in DNA methylation in the 3'RR as compared to their respective wild-type counterparts. CpGs sensitive to H1 depletion are few and localized downstream of HS4. Two additional sites became fully unmethylated in H1 knock out ES cells, one upstream and one downstream of HS7. Finally one site downstream of HS1.2 shows a reduction in DNA methylation. These data suggest that DNA methylation at specific sites in 3'RR requires histone H1.

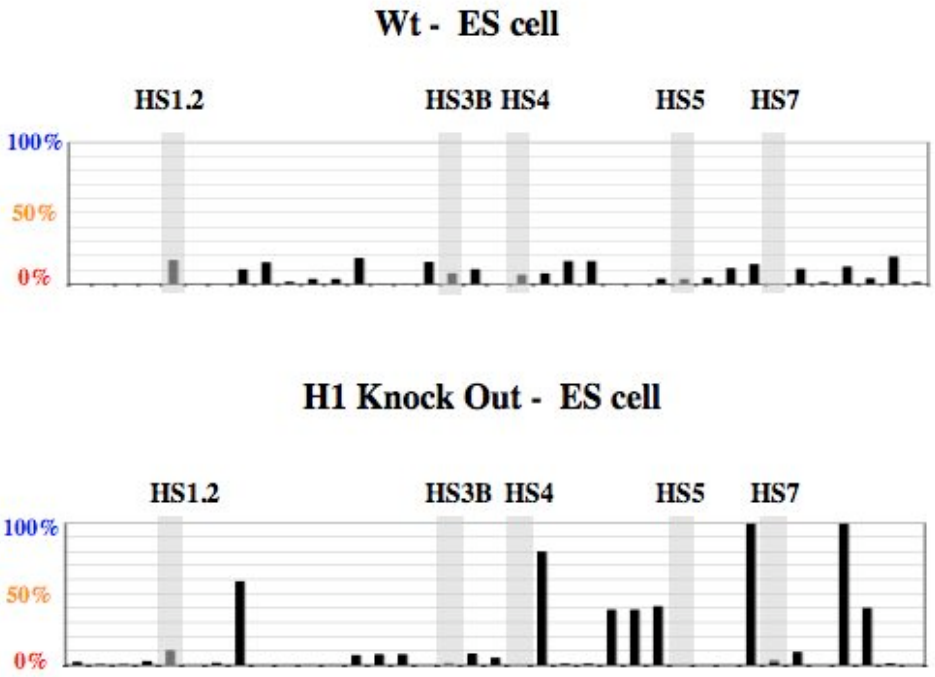


Fig. 28

Fig.28 DNA unmethylation patterns in wt and H1 knock out ES cells. ES cells derived from histone H1 depleted mice showed a reduction in DNA methylation as compared to their respective wild-type counterparts. CpGs sensitive to H1 depletion are few and localized downstream of HS4. Two sites became fully unmethylated in H1 knock out ES cells, one upstream and one downstream HS7.

12. EMSA shows CTCF and USF binding to unmethylated HS5

The HS5 region shows insulator activity, variation in binding for CTCF (Garrett et al., 2005) and also changes in the DNA methylation during B cell development (**Fig.25**). This suggests that the DNA methylation may regulate CTCF binding in HS5, which can be related to insulator activity of HS5. We examined methylated and unmethylated CpGs for protein binding *in vitro* by EMSA. We used probes for each CpG of HS5 and methylated them by a specific CpG Methylase, M.SssI (see material and methods) (**Fig.29**).

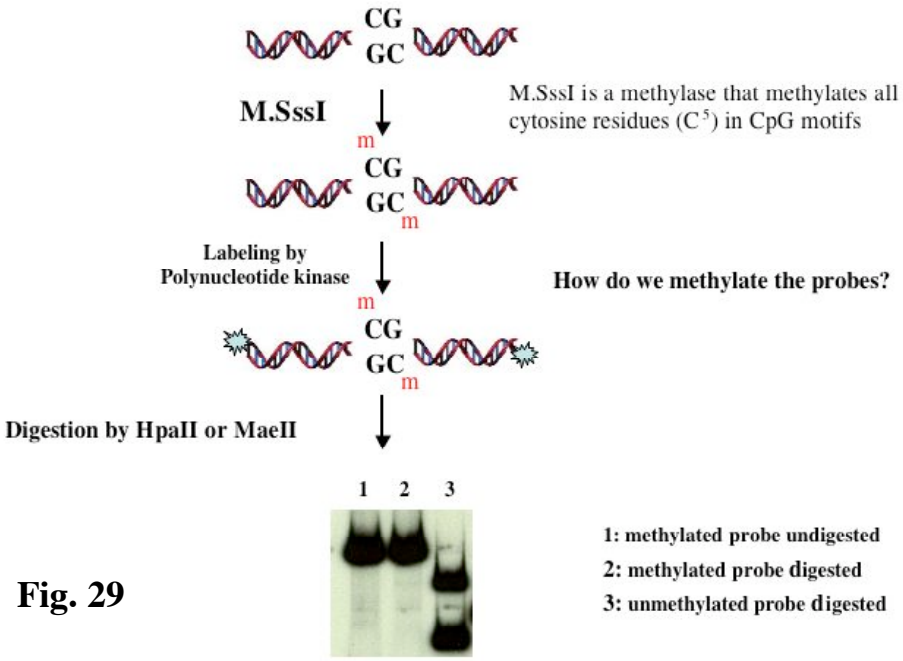


Fig. 29

Fig.29 We study the difference in binding of proteins by EMSA using unmethylated and methylated probes in HS5. Probes were methylated by a specific CpG Methylase, M.SssI (New England Biolabs, Ipswitch, MA). To check the methylation state of probes, an aliquot of each fragment treated by M.SssI was digested by methylation sensitive restriction enzymes, such as HpaII or MaeII. However only some CpGs are in HpaII or MaeII sites and were detected by digestion.

In HS5, we have identified three CpGs that, when methylated, prevent protein binding *in vitro*. **Fig. 30** shows the data for the CpGs in position 31417 and 31427 (the numbers refer to GenBank Accession Number AF450245). When these CpGs were methylated within a fragment extending from 31397 to 31456, no binding was detected in EMSA. The same fragment with unmethylated CpGs showed binding. Here we identify the same region of binding for CTCF that was shown for HS5 in Garrett et al., 2005. To confirm the presence of CTCF binding by EMSA, we use a dsDNA competitor that has its consensus site of imprinting control region (ICR) in Igf2/H19 locus. The presence of competition shows that CTCF is the protein influenced by the methylated probe (**Fig. 30**).

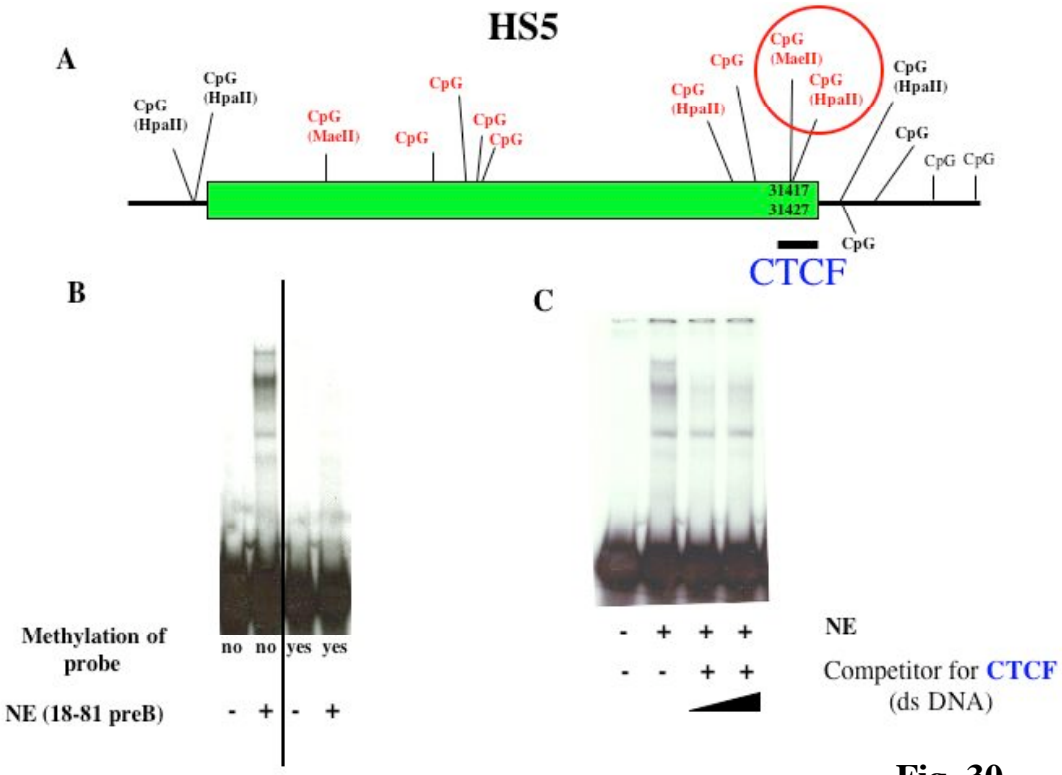


Fig. 30

Fig. 30 EMSA shows CTCF binding to unmethylated HS5. A) Schematic map of HS5 and location of studied CpGs. The brackets enclose the region containing the analyzed sites for the methylation-sensitive restriction enzymes. **B)** EMSA using methylated and unmethylated probes for both 31417 and 31427 CpGs. In this case, the methylated probe in the two CpGs prevents the binding of nuclear proteins of 18-81 pre-B cell lines. **C)** EMSA using the dsDNA competitor for CTCF. This experiment shows that the CTCF binds the probe and that methylation prevents the CTCF binding in HS5.

The CpG in position 30684 is the third CpG that we have identified to change the protein binding. When this CpG is methylated, there is no binding using the nuclear extract of Haftl1, a pro-B cell line (**Fig.31**). In this case, the analysis of the sequence (http://www.cbrc.jp/research/db/TFSEARCH.html) showed that this CpG was inside an E box, a consensus site for proteins as USF, E2A and nMyc.

USF proteins are involved in insulator activity (West et al., 2004) and their binding is also influenced by DNA methylation (Fujii et al., 2006).

Therefore, we investigated the presence of USF binding in this region using specific antibodies for USF1 and USF2 by EMSA. The data demonstrate that these proteins bind HS5 and their binding is influenced by DNA methylation *in vitro* (**Fig. 31**).

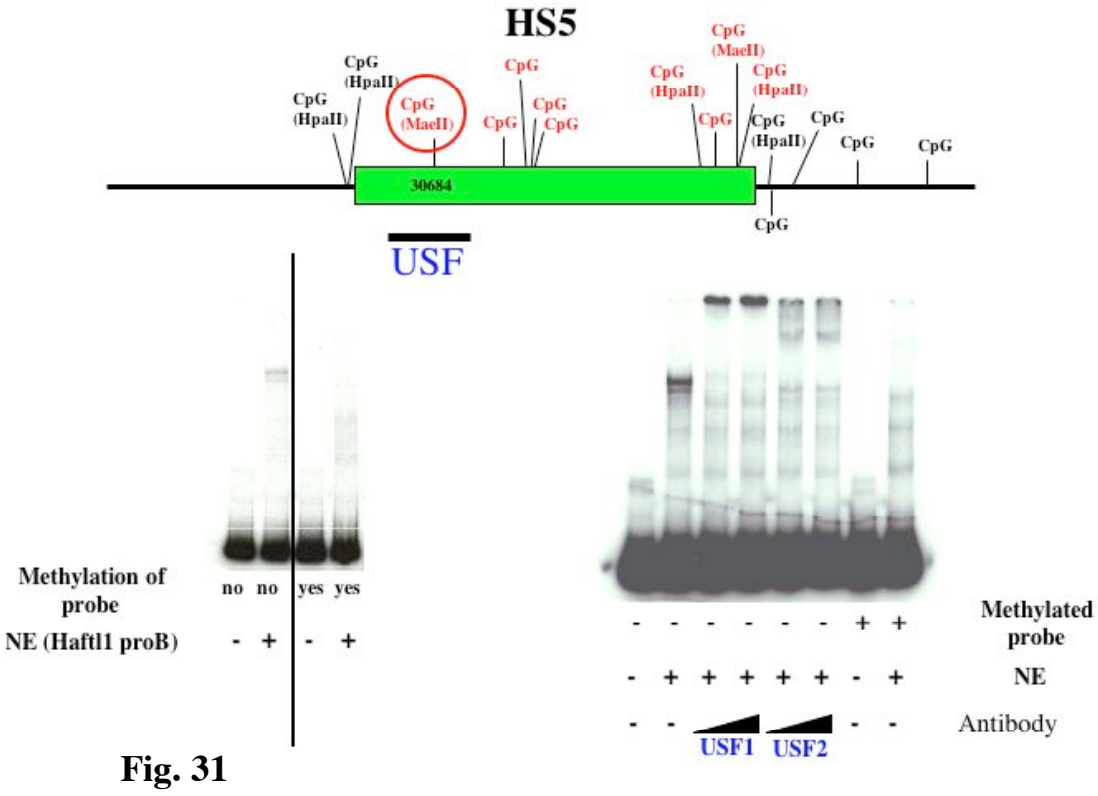


Fig. 31

Fig. 31 EMSA shows USF binding to unmethylated HS5. A) Schematic map of HS5 and location of studied CpGs. The brackets enclose the region containing the analyzed sites for the methylation-sensitive restriction enzymes. **B)** EMSA using methylated and unmethylated probes for 30684 CpGs. The methylated probe in this CpG shows no binding for nuclear proteins of Haftl1 pro-B cell lines. **C)** EMSA using antibodies for USF1 and USF2 at two different concentrations. This experiment shows that the USF1 and USF2 bind the probe. and that methylation prevents the USF binding in HS5.

DISCUSSION

1. The polymorphism of HS1.2 enhancers, in the human 3'RRs

In contrast to murine 3' enhancers, the human 3' enhancers, HS3, HS1.2 and HS4, are duplicated as described above (Chen et al., 1997; Mills et al., 1997; Pinaud et al., 1997) (**Fig.6**). Therefore, genomic contigs spanning the 3' duplicated regions of the human IgH gene locus and present in GenBank since the end of 2002, have allowed us to assemble a continuous nucleotide sequence of the entire region, from the Cα-1 gene through 3'RR-2. This provides new information about the structure of the two related 3'RRs. Based on this information, we have developed a strategy for specific amplification of each duplicated region, utilizing the sequences that diverge in the two 3'RRs. The regions corresponding to enhancers HS1.2 of both 3'RRs were amplified from several independent genomes and then sequenced, confirming the differences observed in GenBank and allowing us to distinguish allelic variants of each locus.

Nonetheless, it is impossible to obtain information about the structure of haplotypes with different combinations of the alleles, due to the difficulty of amplifying in a single reaction a fragment spanning about 180 kb.

The analysis of the human HS1.2 enhancers has provided two main points of information: 1) four alleles are present in the HS1.2-A locus, but only two in the HS1.2-B locus; 2) the polymorphic structure in the two loci seems different. The different alleles are distinguished on the basis of their length. They are characterized by a repeated region of 38 bp and by spacers that change in number (**Fig.16**). Our data also shows that allele *4 is identical, but inverted in both 3'RRs. This suggests that the duplication of the locus took place in a genome harboring that allele. The comparison of the monkey and apes DNA sequences showed that the core enhancer is the more conserved part of the enhancer. However, the repeated region of 38 bp is also present and well conserved in the analyzed species, suggesting that different allelic forms could be already present before the duplication. Other studies of the alleles present in monkeys and apes should help to recognize the polymorphic structure of the ancestral gene.

An obvious consequence of the presence of different HS1.2 alleles resides in the variation of the number of copies of consensus sequences for several transcription factors. In particular, the number of sites for NF-κB

varies from one to four in the alleles. EMSA shows that *in vitro* the polymorphic structure changes the binding pattern in the four alleles and that *in vitro* NF-κB binds the allele*2. This can be of primary importance, as enhancers of the two 3'RRs loci influence the regulation of immunoglobulin production (Denizot et al., 2001) and B lymphocyte maturation (Andersson et al., 2000).

The present work is the first to provide information on the world distribution of the HS1.2-A polymorphism, highlighting the fact that its frequencies are differentially distributed in different geographic groups. The alleles 3A and 4A are at their highest frequencies among Africans. The allele*2 is significantly lower in Africans in comparison with Europeans and, to a lesser extent, Asians. Finally, analysis of molecular variance of the allelic frequencies indicates that the HS1.2-A polymorphism can be considered as a reliable anthropogenetic marker.

Our data provide novel evidence of a possible candidate for a non-HLA locus to genetic predisposition in some systemic autoimmune diseases, such as Lupus erythematosus and systemic sclerosis, and organ target autoimmune diseases, e.g. Crohn's disease, celiac disease, psoriasis, Hashimoto's disease. This marker may also be relevant in determining some associated conditions as celiac disease and herpetiform dermatitis and could be considered as a factor of risk.

This enhances the importance of this polymorphism in epidemiological studies of these immunological disorders, especially in view of the current environmental conditions, such as changes in diet and exposure to pesticides, poisons and particular drugs, to which different populations are exposed. The allelic distribution of HS1.2-A in schizophrenic patients subjected to neuroleptics gives an example that genetic background may affect immunoglobulin patterns that result from antipsychotic therapy.

2. The patterns of DNA methylation in the murine 3'RR

To see how the protein binding can be regulated in 3' RR, we have also studied the state of DNA methylation in some CpGs of the 3'RR in murine cell. Methylation of cytosine within CpG dinucleotides is the major modification of DNA in mammalian genomes. With few exceptions, DNA methylation represses transcription either by blocking binding of transcription factors to their CpG-containing binding sites or by recruiting methyl-binding proteins, which in turn recruit chromatin-remodeling machinery to induce the formation of a repressive chromatin structure (Wade, 2001). Most DNA methylation occurs symmetrically. During the cell cycle, newly replicated DNA incorporates unmethylated cytosine. The DNA methyltransferase 1 protein (Dnmt1) recognizes hemimethylated DNA and methylates the daughter strand, reestablishing a fully methylated state. Another form of DNA methylation is mediated by the *de novo* methyltransferases Dnmt3a and Dnmt3b. Unlike Dnmt1, these two proteins show no preference for hemimethylated substrates, and their expression patterns correlate with de novo methylation during development (Doerfler et al., 2006). In addition to their DNA-methylating capabilities, Dnmt proteins recruit histone deacetylases to DNA, which in turn repress transcription (Burgers et al., 2002). Thus, DNA methylation and histone deacetylation appear to be intimately associated processes leading to transcriptional repression.

In contrast to the detailed understanding of the DNA methylation process, the mechanisms promoting demethylation are unclear. As no DNA demethylase has yet been identified, the mechanism of demethylation likely involves remodeling of the chromatin architecture, allowing sequence-specific transcription factors to access their binding sites, activate transcription, and prevent Dnmt1 from methylating newly synthesized strands during DNA replication. Thus, over several rounds of division, daughter cells effectively become demethylated at recently activated genes (Doerfler et al., 2006).

In the immune system, DNA methylation plays an important role in the transcriptional regulation of several lineage-specific genes. Rearrangement of genes to assemble the mature BCR or T cell receptor (TCR) genes in B or T cells, respectively, generally proceeds in multiple, developmentally regulated steps. The methylation status of gene segments encoding BCR or TCR chains has been correlated with V(D)J rearrangement (Bergman et al., 2004,), although demethylation alone is not sufficient to activate

recombination (Cherry et al., 2000). In B cells, a rare recessive mutation in the *Dnmt3b* gene causes an immunodeficiency disease called ICF (immunodeficiency, centromeric region instability, and facial anomalies). Individuals with ICF have normal numbers of B cells, no or reduced levels of serum immunoglobulin, and altered expression levels of many B-cell-specific genes (Ehrlich et al., 2001), suggesting that the methylation status of these genes contributes to their regulation.

DNA methylation was previously investigated in murine 3'RR with genomic Southern analysis of HpaII digested DNA. Regions in the 3'RR showed cell type specific variations in methylation. In particular progressive changes were observed during B cell development. However only the ends of large fragments (3'α1.2 and PstI 2.3) were monitored, but not specific sites between them. They showed that in both lymphoid and T cells, sequences between Cα and 3'α enhancer were hypermethylated. In no lymphoid cells, the region (PstI2.3) downstream of the 3'α enhancer was considerably undermethylated, whereas in T cells, it was hypermethylated (Giannini et al.,1993). Observations showed the presence of other DNasI hypersensitive sites, called HS5, HS6 and HS7 in the murine 3'RR downstream of HS4, which were associated with histone modifications characteristic of open chromatin environments. CTCF binding sites were detected in the vicinity of HS4 and HS5-HS6 together with a cluster of four very strong sites in HS7. Thus it is possible that the 3' RR extends several kilobases beyond HS4. Although the role of each individual DNaseI hypersensitive site in the 3' RR remains to be elucidated, the association of these sites with changes in histone modification and *in vitro* insulator activity suggests an important role in their overall regulation of the locus (Garrett et al., 2005).

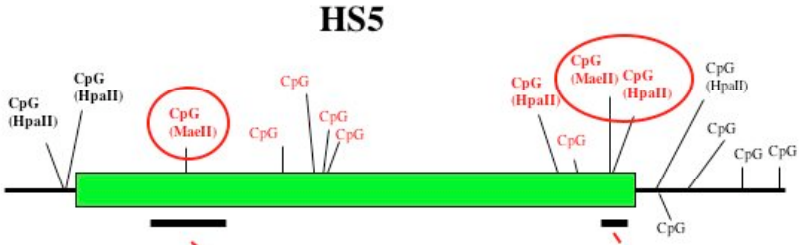
In this thesis, I analyzed the DNA methylation of specific CpGs in the murine 3'RR by a quantitative PCR assay and I have shown differences in DNA methylation patterns of cell lines, reflecting different stages of B development, also between resting splenic B cells and splenic B cells stimulated to undergo class switching with LPS, or with LPS plus IL4. From this analysis, I observed that HS4 is the first DNaseI hypersensitive site to become unmethylated during the B cell development. The region downstream of HS4 at first becomes unmethylated in pre B cell lines. HS1.2 and HS3B become unmethylated only in plasma cell lines.

I have shown also that the CSR stimulation decreases the DNA methylation of 3'RR in a specific way and that different types of stimulation create different patterns of DNA methylation.

In the analyzed non B cells, the 3'RR is not fully methylated such as ES cell line, but only specific CpG sites are methylated in different way. It suggest that, in these cells, there are some open regions which are different with respect to B cells and to each others and could be negatively involved in the 3'RR activity.

These data suggest also that the histone H1 regulates the levels of DNA methylation in 3'RR. In fact the comparison of DNA methylation patterns in wild type and H1 knock out ES cells shows that H1 maintains the DNA methylation level in 3' RR in specific sites. This suggests that H1, directly by physical interaction or indirectly by other factors, could recruit specific DNA methyltransferases in this locus or preclude the demethylation processes. H1 could represent an other pathway of epigenetic regulation of 3'RR.

We have also investigated the influence of DNA methylation on protein binding in HS5. We demonstrated that *in vitro* DNA methylation regulates the binding of USF proteins and CTCF. In fact the DNA methylation prevents the binding these proteins. This is in agreement with the hypothesis that HS5 is an insulator region of 3'RR (**Fig. 32**).



	USF	CTCF
Binding in methylated DNA	NO	NO
Binding in unmethylated DNA	YES	YES

Fig. 32

Fig. 32 Binding site regions influenced by DNA methylation in HS5. *In vitro* DNA methylation prevents the binding of USF proteins and CTCF.

The role of USF and CTCF in an insulator region (HS4) is amply analyzed in the chicken β-globin locus (Gaszner et al., 2006). In this locus, it has been seen that insulators are DNA sequence elements that prevent inappropriate interactions between adjacent chromatin domains and present at least two different functional regions One type of insulator region establishes domains that separate enhancers and promoters to block their interaction, whereas a second type creates a barrier against the spread of heterochromatin. Therefore it has been described a model of insulator

action, involving the clustering or tethering of different insulator elements. This model supposes that the insulator is a complex of many factors with different functions, an enhancer blocking activity, most closely associated proteins as CTCF, and a barrier activity that prevent the spread of heterochromatin, related to the USF proteins (Yusufzai et al., 2004; West et al., 2004). The role of USF is also to mediate the recruitment of specific enzymes for the histone modifications, such as the histone acetyltransferases (HATs) and the histone metyltransferase (HMTs) that are an essential component for the insulator function. However it is important to note that while USF is necessary for the barrier activity in the β-globin locus, it is not sufficient. In fact other protein binding sites are also required (Recillas-Targa et al., 2002).

HS5 in the murine 3'RR shows a structure similar to insulator of the β-globin locus in terms of protein binding. HS5 as insulator was investigated before (Garrett et al., 2005), however other observations will be required to clarify the function of this region and the role of *in vivo*. These data suggest that the binding for different factors can be influenced by DNA methylation in some specific CpG sites. Other experiments will be necessary to explain how these factors and DNA methylation, in general, are involved in immunoglobulin production.

3. Conclusions

Therefore, the picture that emerges from this study is that protein binding in 3'RR can be modulated in two different ways by polymorphic regions and DNA methylation. In this thesis I have studied two examples: one, in the human 3'RRs and one in the murine 3'RR. In the first, the HS1.2 polymorphism defines different genetic backgrounds in the human population and could modulate the capacity of immunoglobulin production by changing the protein binding. In the second, I have studied how the DNA methylation executes functional differences during the B cell development by changing the binding for proteins such as CTCF and USF, related to insulator activity. Both studies show how the processes of immunoglobulin production and diversification are highly regulated, suggesting that there are different pathways of control, some, more static, as the DNA polymorphisms, and others, more dynamic, like DNA methylation.

REFERENCES

- Aguilera RJ, Hope TJ, Sakano H. - Characterization of immunoglobulin enhancer deletions in murine plasmacytomas. - EMBO J. 1985 Dec 30;4(13B):3689-93.
- Andersson, T., Samuelsson, A., Matthias, P., Pettersson, S., - The lymphoid-specific cofactor OBF-1 is essential for the expression of a V(H) promoter/HS1,2 enhancer-linked transgene in late B cell development.- Mol. Immunol. 2000; 37, 889– 899.
- Aupetit C, Drouet M, Pinaud E, Denizot Y, Aldigier JC, Bridoux F, Cogne M. - Alleles of the alpha1 immunoglobulin gene 3' enhancer control evolution of IgA nephropathy toward renal failure.- Kidney Int. 2000 Sep;58(3):966-71.
- Banerji J, Olson L, Schaffner W. - A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes.- Cell. 1983 Jul;33(3):729-40.
- Bassing CH, Swat W, Alt FW.-The mechanism and regulation of chromosomal V(D)J recombination.- Cell. 2002 Apr;109 Suppl:S45-55.
- Bell AC, Felsenfeld G. - Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene.- Nature. 2000 May 25;405(6785):482-5.
- Bergman Y, Cedar H.- A stepwise epigenetic process controls immunoglobulin allelic exclusion.-Nat Rev Immunol. 2004 Oct;4(10):753-61.
- Bottaro A., Cariota U., De Marchi M., Carbonara A.O.,- Pulsed field electrophoresis screening for immunoglobulin heavy chain constant region (IGHC) multigene deletions and duplications. – Am. J. Hum. Genet. 48, 745-756.
- Burgers, W. A., F. Fuks, and T. Kouzarides. DNA methyltransferases get connected to chromatin. Trends Genet. 2002; 18:275-277.
- Cabibbo A., Grant PR., Helmer-Citterich M., - The Internet for Cell and Molecular Biologists, Current Applications and Future Potentials- 2002 Horizon Scientific Press, Norfolk, UK.
- Chaudhuri J, Alt FW. - Class-switch recombination: interplay of transcription, DNA deamination and DNA repair.- Nat Rev Immunol. 2004 Jul;4(7):541-52. Review.

- Chauveau C., Cogne M., - Palindromic structure of the IgH 3' locus control region. – Nat. Genet. 1996 Sep;14(1):15-6.
- Chauveau, C., Pinaud, E., Cogne, M.,. Synergies between regulatory elements of the immunoglobulin heavy chain locus and its palindromic 3' locus control region. Eur. J. Immunol. 1998; 28, 3048–3056.
- Chen C., Birshtein B.K.- Virtually identical enhancers containing a segment of homology to murine 3'IgH-E(hs1,2) lie downstream of human Ig C alpha 1 and C alpha 2 genes. J. Immunol. 1997 Aug 1;159(3):1310-8.
- Chen J., Young F., Bottaro A., Stewart V., Smith RK., Alt FW., - Mutations of the intronic IgH enhancer and its flanking sequences differentially affect accessibility of the J_H locus. EMBO J. 1993, 12:4635-45.
- Cherry, S. R., C. Beard, R. Jaenisch, and D. Baltimore. V(D)J recombination is not activated by demethylation of the kappa locus. Proc. Natl. Acad. Sci. USA 2000; 97:8467-8472
- Cockerill PN., Yuen M-H., Garrard WT., - The enhancer of the immunoglobulin heavy chain locus is flanked by presumptive chromosomal loop anchorage elements.- J. Biol. Chem. 1987, 262:5394-97
- Cogne M., Lansford R., Bottaro A., Zhang J., Gorman J., Young F., Cheng H.L., Alt F.W., - A class switch control region at the 3' end of the immunoglobulin heavy chain locus. – Cell 1994, Jun 3;77(5):737-47.
- Cooper, M.D., Peterson, R.D., and Good, R.A. - Delineation of the thymic and bursal lymphoid systems in the chicken. Nature 1965; 205,143–146.
- Denizot Y, Pinaud E, Aupetit C, Le Morvan C, Magnoux E, Aldigier JC, Cogne M. - Polymorphism of the human alpha1 immunoglobulin gene 3' enhancer hs1,2 and its relation to gene expression.- Immunology. 2001 May;103(1):35-40.
- Doerfler W., Böhm P. – DNA Methylation basic mechanisms- Springer Editor.
- Dunnick WA, Shi J, Graves KA, Collins JT.- The 3' end of the heavy chain constant region locus enhances germline transcription and switch recombination of the four gamma genes.- J Exp Med. 2005 May 2;201(9):1459-66

- Eckhardt LA, Birshtein BK. - Independent immunoglobulin class-switch events occurring in a single myeloma cell line. -Mol Cell Biol. 1985 Apr;5(4):856-68.
- Ehrlich, M., K. L. Buchanan, F. Tsien, G. Jiang, B. Sun, W. Uicker, C. M. Weemaes, D. Smeets, K. Sperling, B. H. Belohradsky, N. Tommerup, D. E. Misek, J. M. Rouillard, R. Kuick, and S. M. Hanash. DNA methyltransferase 3B mutations linked to the ICF syndrome cause dysregulation of lymphogenesis genes. Hum. Mol. Genet. 2001; 10:2917-2931.
- Ernst P, Smale ST. 1995. Combinatorial regulation of transcription I: general aspects of transcriptional control. *Immunity* 2:311–19
- Fan Y., Nikitina T., Zhao J., Fleury T.J., Bhattacharyya R., Bouhassira E.E., Stein A., Woodcock C.L., and Skoultchi A. I.,- Histone H1 Depletion in Mammals Alters Global Chromatin Structure but Causes Specific Changes in Gene Regulation. - Cell 123, 1199–1212, December 29, 2005.
- Fernex C., Capone M., Ferrier P., - The V(D)J recombinational and transcriptional activities of the immunoglobulin heavy-chian intronic enhancer can be mediated through distint protein binding sites in a transgenic substrate. – Mol. Cell. Biol. 1995; 15:3217 – 26.
- Flanagan J.G., Lefranc M.P., Rabbitts T.H. – Mechanisms of divergence and convergence of the human immunoglobulin alpha 1 and alpha 2 constant region gene sequences. – Cell 1984 Mar;36(3):681-8.
- Frezza D., Camacho-Vanegas O., Fruscalzo A., Favaro M., Giorgi S., Scotto d'Abusco A.S., Gualandi G.- The region 3' to Calpha1 gene of human IG heavy chain displays a polymorphic duplicated sequence and encodes an RNA associated with polysomes. Gene.1998 Sep 28;219(1-2):19-24.
- Fujii G, Nakamura Y, Tsukamoto D, Ito M, Shiba T, Takamatsu N.- CpG methylation at the USF-binding site is important for the liver-specific transcription of the chipmunk HP-27 gene.- Biochem J. 2006 Apr 1;395(1):203-9.
- Fujita T, Nolan GP, Ghosh S, Baltimore D.- Independent modes of transcriptional activation by the p50 and p65 subunits of NF-kappa B.- Genes Dev. 1992 May;6(5):775-87.
- Garrett FE, Emelyanov AV, Sepulveda MA, Flanagan P, Volpi

S, Li F, Loukinov D, Eckhardt LA, Lobanenkov VV, Birshtein BK. Chromatin architecture near a potential 3' end of the Igh locus involves modular regulation of histone modifications during B-Cell development and in vivo occupancy at CTCF sites. - Mol Cell Biol. 2005 Feb;25(4):1511-25.

- Gaszner M, Felsenfeld G. - Insulators: exploiting transcriptional and epigenetic mechanisms.- Nat Rev Genet. 2006 Sep;7(9):703-13.
- Giannini SL, Singh M, Calvo CF, Ding G, Birshtein BK.- DNA regions flanking the mouse Ig 3' alpha enhancer are differentially methylated and DNAase I hypersensitive during B cell differentiation.- J Immunol. 1993 Mar 1;150(5):1772-80.
- Gillies SD, Morrison SL, Oi VT, Tonegawa S.- A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene.- Cell. 1983 Jul;33(3):717-28.
- Gualandi,G., Frezza,D., Scotto d'Abusco,A., Bianchi,E., Gargano,S., Giorgi,S., Fruscalzo,A. and Calef,E.- Integration of an Epstein-Barr virus episome 3' into the gene encoding immunoglobulin heavy-chain alpha 1 in a lymphoblastoid cell line.- Gene. 1995 Dec 12;166(2):221-6.
- Guglielmi L, Truffinet V, Magnoux E, Cogne M, Denizot Y. - The polymorphism of the locus control region lying downstream the human IgH locus is restricted to hs1,2 but not to hs3 and hs4 enhancers.- Immunol Lett. 2004 Jun 15;94(1-2):77-81.
- Hall, Y. H., Fuentes, E. F., Chertow, G. M. & Olson, J. L. - Race/ethnicity and disease severit` in IgA nephropathy. *BCM Nephrology* 2004*;* **5**, 10.
- Harindranath N, Mills FC, Mitchell M, Meindl A, Max EE.- The human elk-1 gene family: the functional gene and two processed pseudogenes embedded in the IgH locus.- Gene. 1998 Oct 23;221(2):215-24.
- Hatada EN, Chen-Kiang S, Scheidereit C. - Interaction and functional interference of C/EBPbeta with octamer factors in immunoglobulin gene transcription.- Eur J Immunol. 2000 Jan;30(1):174-84.
- Howell M.L., Schroth G.P., Ho P.S.- Sequence-dependent effects of spermine on the thermodynamics of the B-DNA to Z-DNA transition - Biochemistry. 1996 Dec 3;35(48):15373-82

- Hu Y, Pan Q, Pardali E, Mills FC, Bernstein RM, Max EE, Sideras P, Hammarstrom L. - Regulation of germline promoters by the two human Ig heavy chain 3' alpha enhancers.- J Immunol. 2000 Jun 15;164(12):6380-6.
- Ishihara K, Oshimura M, Nakao M.- CTCF-dependent chromatin insulator is linked to epigenetic remodeling.- Mol Cell. 2006 Sep 1;23(5):733-42
- Jones AL, Mowry BJ, Pender MP, Greer JM. - Immune dysregulation and self-reactivity in schizophrenia: do some cases of schizophrenia have an autoimmune basis?- Immunol Cell Biol. 2005 Feb;83(1):9-17.
- Khamlichi AA., Pinaud E., Decourt C., Chauveau C., Cogne M., - The 3' IgH regulatory region: a complex structure in a search for a function. - Adv. Immunol. 2000; 75:317-45.
- Kim EC, Edmonston CR, Wu X, Schaffer A, Casali P.- The HoxC4 homeodomain protein mediates activation of the immunoglobulin heavy chain 3' hs1,2 enhancer in human B cells. Relevance to class switch DNA recombination.- J Biol Chem. 2004 Oct 1;279(40):42258-69.
- Kominato Y, Tsuchiya T, Hata N, Takizawa H, Yamamoto F.- Transcription of human ABO histo-blood group genes is dependent upon binding of transcription factor CBF/NF-Y to minisatellite sequence.- J Biol Chem. 1997 Oct 10;272(41):25890-8
- Laurencikiene J, Deveikaite V, Severinson E.- HS1,2 enhancer regulation of germline epsilon and gamma2b promoters in murine B lymphocytes: evidence for specific promoter-enhancer interactions.- J Immunol. 2001 Sep 15;167(6):3257-65.
- Laurencikiene J, Tamosiunas V, Severinson E.- Regulation of {epsilon} germline transcription and switch region mutations by IgH locus 3' enhancers in transgenic mice.- Blood. 2006 Sep 12;
- Lefranc M.P., Lefranc G., Rabbitts T.H.- Inherited deletion of immunoglobulin heavy chain constant region genes in normal human individuals. –Nature 1982, Dec 23;300 (5894):760-2.
- Levy, M. & Berger, J. - Worldwide perspective of IgA nephropathy. *Am J Kidney Dis* 1988; **12**, 340–347.
- Li Z, Woo CJ, Iglesias-Ussel MD, Ronai D, Scharff MD.- The generation of antibody diversity through somatic hypermutation and class switch recombination. - Genes Dev. 2004 Jan 1;18(1):1-11.Click here to read

- Litman GW, Cannon JP, Dishaw LJ.-Reconstructing immune phylogeny: new perspectives. - Nat Rev Immunol. 2005 Nov;5(11):866-79.
- Madisen L, Groudine M. - Identification of a locus control region in the immunoglobulin heavy-chain locus that deregulates c-myc expression in plasmacytoma and Burkitt's lymphoma cells.- Genes Dev. 1994 Sep 15;8(18):2212-26.
- Maeng JH, Yoon JB. - The human PTFgamma/SNAP43 gene: structure, chromosomal location, and identification of a VNTR in 5'-UTR.J Biochem (Tokyo). 1998 Jul;124(1):23-7.
- Maier H, Colbert J, Fitzsimmons D, Clark DR, Hagman J.- Activation of the early B-cell-specific mb-1 (Ig-alpha) gene by Pax-5 is dependent on an unmethylated Ets binding site.- Mol Cell Biol. 2003 Mar;23(6):1946-60.
- Manis JP, Michaelson JS, Birshtein BK, Alt FW.- Elucidation of a downstream boundary of the 3' IgH regulatory region. - Mol Immunol. 2003 Jan;39(12):753-60.
- 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. - Class switching in B cells lacking 3' immunoglobulin heavy chain enhancers. - J Exp Med. 1998 Oct 19;188(8):1421-31.
- Max EE, Mitchell M, Mills FC.- Multiple recombination events near the 3' boundary of the human IgH locus duplication.- Mamm Genome. 2000 Oct;11(10):938-40.
- Merika M, Thanos D.- Enhanceosomes.- Curr Opin Genet Dev. 2001 Apr;11(2):205-8.
- Michaelson JS, Giannini SL, Birshtein BK. - Identification of 3' alpha-hs4, a novel Ig heavy chain enhancer element regulated at multiple stages of B cell differentiation. - Nucleic Acids Res. 1995 Mar 25;23(6):975-81.
- Michaelson JS, Singh M, Snapper CM, Sha WC, Baltimore D, Birshtein BK. - Regulation of 3' IgH enhancers by a common set of factors, including kappa B-binding proteins.- J Immunol. 1996 Apr 15;156(8):2828-39.
- Mills F.C., Harindranath N., Mitchell M., Max E.E. – Enhancer complex located downstream of both human immunoglobulin Calpha genes. – J. Exp. Med. 1997, Sep 15;186(6):845-58.

- Morvan CL, Pinaud E, Decourt C, Cuvillier A, Cogne M. - The immunoglobulin heavy-chain locus hs3b and hs4 3' enhancers are dispensable for VDJ assembly and somatic hypermutation. Blood. 2003 Aug 15;102(4):1421-7.
- Muller N, Riedel M, Gruber R, Ackenheil M, Schwarz MJ. - The immune system and schizophrenia. An integrative view.- Ann N Y Acad Sci. 2000;917:456-67.
- Muto A, Hoshino H, Madisen L, Yanai N, Obinata M, Karasuyama H, Hayashi N, Nakauchi H, Yamamoto M, Groudine M, Igarashi K.- Identification of Bach2 as a B-cell-specific partner for small maf proteins that negatively regulate the immunoglobulin heavy chain gene 3' enhancer.- EMBO J. 1998 Oct 1;17(19):5734-43.
- Odegard VH, Schatz DG - Targeting of somatic hypermutation.- Nat Rev Immunol. 2006 Aug;6(8):573-83.
- Ong J, Stevens S, Roeder RG, Eckhardt LA. - 3' IgH enhancer elements shift synergistic interactions during B cell development.- J Immunol. 1998 May 15;160(10):4896-903.
- Pawlitzky I, Angeles CV, Siegel AM, Stanton ML, Riblet R, Brodeur PH. - Identification of a candidate regulatory element within the 5' flanking region of the mouse Igh locus defined by pro-B cell-specific hypersensitivity associated with binding of PU.1, Pax5, and E2A. - J Immunol. 2006 Jun 1;176(11):6839-51.
- Pettersson S, Cook GP, Bruggemann M, Williams GT, Neuberger MS.- A second B cell-specific enhancer 3' of the immunoglobulin heavy-chain locus.- Nature. 1990 Mar 8;344(6262):165-8.
- Pinaud E, Khamlichi AA, Le Morvan C, Drouet M, Nalesso V, Le Bert M, Cogne M.- Localization of the 3' IgH locus elements that effect long-distance regulation of class switch recombination. - Immunity. 2001 Aug;15(2):187-99.
- Pinaud E., Aupetit C., Chauveau C., Cogne M.- Identification of a homolog of the C alpha 3'/hs3 enhancer and of an allelic variant of the 3'IgH/hs1.2 enhancer downstream of the human immunoglobulin alpha 1 gene. – Eur.J. Immunol. 1997 Nov;27(11):2981-5.
- Pollmacher T, Haack M, Schuld A, Kraus T, Hinze-Selch D. - Effects of antipsychotic drugs on cytokine networks.- J Psychiatr Res. 2000 Nov-Dec;34(6):369-82.

- Rabbani H., Kondo N., Smith C.I.E., Hammarström L., - The influence of gene deletions and duplications within the IGHC locus on serum immunoglobulin subclass levels.- Clin. Immnol. Immunopathol. 1995; 76, 214-218.
- Rabbani H., Pan Q., Kondo N., Smith C.I.E., Hammarström L.,- Duplications and deletions of the human IGHC locus: evolutionary implications.- Immunogenetics 1996; 45, 136-141.
- Recillas-Targa, F., Pikaart, M.J., Burgess-Beusse, B., Bell, A.C., Litt, M.D., West, A.G., Gaszner, M., and Felsenfeld, G. Positioneffect protection and enhancer blocking by the chicken beta-globininsulator are separable activities. Proc. Natl. Acad. Sci. USA 2002; *99*,6883–6888.
- Rothermundt M, Arolt V, Bayer TA. Review of immunological and immunopathological findings in schizophrenia. *Brain Behav. Immun*. 2001; **15**: 319–39.
- Sadhu A.I., Shen M.L., Hackbarth M., Hume E., McKeithan T.W.,- CpG-rich sequences close to the site of duplication within the human IGH constant region. – Immunogenetics. 1997;45(6):365-70.
- Saleque S, Singh M, Birshtein BK. - Ig heavy chain expression and class switching in vitro from an allele lacking the 3' enhancers DNase I-hypersensitive hs3A and hs1,2.- J Immunol. 1999 Mar 1;162(5):2791-803.
- Saleque S., Singh M., Little R.D., Giannini S.L., Michaelson J.S., Birshtein B.K. – Dyad symmetry within the mouse 3’ IgH regulatory region includes two virtually identical enahncers (C alpha3’E and hs3). – J. Immunol. 1997 May 15;158(10):4780-7
- Schreiber E, Matthias P, Muller MM, Schaffner W.- Rapid detection of octamer binding proteins with 'mini-extracts', prepared from a small number of cells.- Nucleic Acids Res. 1989 Aug 11;17(15):6419
- Scofield RH. Autoantibodies as predictors of disease. Lancet 2004; 363: 1544.
- Sepulveda MA, Emelyanov AV, Birshtein BK. - NF-kappa B and Oct-2 synergize to activate the human 3' Igh hs4 enhancer in B cells.- J Immunol. 2004 Jan 15;172(2):1054-64.
- Sepulveda MA, Garrett FE, Price-Whelan A, Birshtein BK. - Comparative analysis of human and mouse 3' Igh regulatory regions identifies distinctive structural features. - Mol Immunol. 2005 Mar;42(5):605-15.

- Serwe M., Sablitzky F., V(D)J recombination in B cells is impaired but not blocked by targeted deletion of the immunoglobulin heavy chain intron enhancer. EMBO J. 1993, 12:2321-27
- Shepshelovich D, Shoenfeld Y. - Prediction and prevention of autoimmune diseases: additional aspects of the mosaic of autoimmunity.- Lupus. 2006;15(3):183-90.
- Shlyakhtenko LS, Hsieh P, Grigoriev M, Potaman VN, Sinden RR, Lyubchenko YL. A cruciform structural transition provides a molecular switch for chromosome structure and dynamics. - J Mol Biol. 2000 Mar 10;296(5):1169-73.
- Singh M, Birshtein BK.- Concerted repression of an immunoglobulin heavy-chain enhancer, 3' alpha E(hs1,2).- Proc Natl Acad Sci U S A. 1996 Apr 30;93(9):4392-7.
- Sleckman BP, Gorman JR, Alt FW. - Accessibility control of antigen-receptor variable-region gene assembly: role of cis-acting elements. - Annu Rev Immunol. 1996;14:459-81.
- Staudt LM., Lenardo MJ., - Immunoglobulin gene transcription. Annu. Rev. Immunol. 1991; 9:373-98.
- Stavnezer J, Amemiya CT.- Evolution of isotype switching. - Semin Immunol. 2004 Aug;16(4):257-75.
- 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.Click here to read
- Urbanek P, Wang ZQ, Fetka I, Wagner EF, Busslinger M.- Complete block of early B cell differentiation and altered patterning of the posterior midbrain in mice lacking Pax5/BSAP.- Cell. 1994 Dec 2;79(5):901-12.
- Wade, P. A. Methyl CpG binding proteins: coupling chromatin architecture to gene regulation. Oncogene 2001; 20:3166-3173
- Wanstrat A,Wakeland E. The genetics of complex autoimmune diseases: non-MHC susceptibility genes. *Nature Immunology* 2001; **2**: 802–809.
- Weinberger J, Baltimore D, Sharp PA. - Distinct factors bind to apparently homologous sequences in the immunoglobulin heavy-chain enhancer.- Nature. 1986 Aug 28-Sep 3;322(6082):846-8.

- West AG, Huang S, Gaszner M, Litt MD, Felsenfeld G.- Recruitment of histone modifications by USF proteins at a vertebrate barrier element.- Mol Cell. 2004 Nov 5;16(3):453-63.
- West, A. G., M. Gaszner, and G. Felsenfeld. - Insulators: many functions, many mechanisms. Genes Dev. 2002; 16:271–288.
- Xue K, Rada C, Neuberger MS.- The in vivo pattern of AID targeting to immunoglobulin switch regions deduced from mutation spectra in msh2-/- ung-/- mice.- J Exp Med. 2006 Sep 4;203(9):2085-94.
- Yusufzai, T.M., Tagami, H., Nakatani, Y., and Felsenfeld, G.- CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species. Mol. Cell 2004; *13*, 291–298.
- Zhou J, Ashouian N, Delepine M, Matsuda F, Chevillard C, Riblet R, Schildkraut CL, Birshtein BK. - The origin of a developmentally regulated Igh replicon is located near the border of regulatory domains for Igh replication and expression. - Proc Natl Acad Sci U S A. 2002 Oct 15;99(21):13693-8.