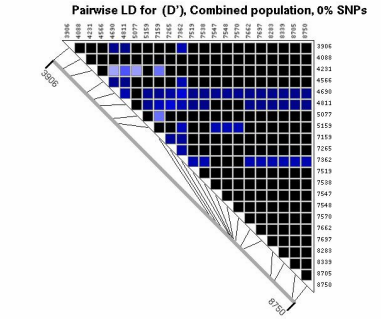
UNIVERSITY OF ROME
"TOR VERGATA"

Faculty of Sciences

PhD school in Cellular and Molecular Biology

XXI cycle

Association of HS1,2A polymorphism with several diseases and analysis of the haplotypes of the region surrounding the enhancer



Claudia Mattioli

A.A. 2008/2009

Mentor: Prof. Domenico Frezza

PhD Coordinator: Prof. Gianni Cesareni

TABLE OF CONTENTS

ABSTRACT

In humans two regulatory regions (3'RR-1 and 3'RR-2) lying downstream of each of the two heavy chain constant α genes are able to enhance immunoglobulin transcription from the heavy chain promoters in a tissue-and-stage-specific manner. Each human regulatory region harbours three enhancer elements: HS3, HS1,2 and HS4. Among the enhancers, only HS1,2 proved to be polymorphic. The selective amplification of HS1,2A revealed the occurrence of four alleles (1; 2; 3; 4) which show variable frequencies throughout the world population. One aim of this thesis is to study the distribution of HS1,2A alleles in the Italian population. The Italian population turns out to be rather homogeneous for this locus. Nevertheless, natural barriers impair genetic flux in Sardinia in contrast to the homogeneity generally found in the rest of Italy. A similar effect is also found in Spain's Sierra de Gredos region.

Previous studies have demonstrated the association of the HS1,2A polymorphism with some immunological diseases. Gel shift experiments have shown that two different protein complexes bind the HS1,2A alleles 1 and 2. These data suggest a role of the HS1,2A polymorphism in the regulation of B cell functions. In this thesis we extended the analysis to selective IgA deficiency and to several autoimmune diseases. We found a significant increase of allele 1 frequency in a group of patients affected by IgA deficiency in respect to the control. On the other hand, allele 2 frequency is significantly increased in individuals affected by different autoimmune diseases in respect to the control. Moreover, allele 2 associates strongly with the acute form of a given autoimmune disease in respect to the milder form. The association of HS1,2A polymorphism with the alteration of immunoglobulin serum levels was also investigated in this thesis. Allele 1 frequency increase associates significantly with immunoglobulin serum levels below the standard values, while allele 2 frequency increase correlates with immunoglobulin serum levels above the standard values. It is known that several host genetic factors influence the progression of viral pathologies. Previous studies have shown the correlation of HS1,2A polymorphism with AIDS progression in a cohort of Libyan patients. Here we found an association of HS1,2A polymorphism with the decrease of CD4+ T cell levels in a group of HIV positive subjects from South Africa. We also showed that the HS1,2A polymorphism could be involved in the individual response to Hepatitis C virus infection.

Subsequently, the analysis was extended to a region lying between the enhancers HS3 and HS1,2A of the 3'RR-1. In total, 22 single nucleotide polymorphisms in strong linkage disequilibrium with each other were identified. We found a strong significant correlation of the haplotypes with the HS1,2A alleles. We hypothesized that the HS1,2A polymorphism and the haplotypes of the 3'RR-1 could affect epigenetic changes or transcription factor binding to the regulative region, so influencing the regulative region activity in immunoglobulin germline transcription, in class switch recombination and in immunoglobulin production.

ABBREVIATIONS

AID	Autoimmune disease
BCR	B cell receptor
CD	Crohn's Disease
C_H	Constant region heavy chain gene
C_L	Constant region light chain gene
CSR	Class Switch Recombination
DH	*Dermatitis herpetiformis*
D_H	Diversity region heavy chain gene
Eμ	Immunoglobulin Heavy Chain Enhancer
ERA	Early Rheumatoid arthritis
GT	Germline transcription
GWA	Genome-wide association studies
HLA	human leukocyte antigen
HS	DNAse I hypersensitive site
I	I-region promoter
Ig	Immunoglobulin
IGAD	IgA deficiency
IgH	Immunoglobulin heavy chain locus
IgHC	Immunoglobulin heavy chain constant region
J_H	Joining region heavy chain gene
LCR	Locus Control Region
LSRA	Long Standing Rheumatoid Arthritis
MHC	Major Histocompatibility Complex
PA	Psoriatic Arthritis
PP	Plaque Psoriasis
PS	Psoriasis
RA	Rheumatoid Arthritis
RCU	Rectocolitis Ulcerosis
RFLP	Restriction fragment length polymorphism
3'RR	3' Regulatory Region
S	Switch region
SHM	Somatic Hypermutation
SLE	Systemic lupus erythematosus
SNP	Single Nucleotide Polymorphism
V_H	Variable region heavy chain gene
V_k	Variable region light chain gene

INTRODUCTION

1. Structure of the immunoglobulin

The soluble immunoglobulins (Ig) represent the 20% of total serum protein and are product by plasma cells by a tightly regulated process. The antibodies (immunoglobulins) are composed by two heavy and two light chains. Both the light chains and the heavy chains have a variable (V) and a constant (C) region which are held together by disulfide bonds (Fig. 1). The variable regions both of the heavy and light chain form the antigen binding site determining the antibodies specificity; whereas other properties of antibodies molecules such as half life, complement fixation and placental transfer, depend only on the constant region of the heavy chain. There are five different isotypes for the constant regions (α, δ, ε, γ,and μ); each isotype is encoded from a different DNA segment and defines the specific class of the antibodies (IgA; IgD; IgE; IgG and IgM).

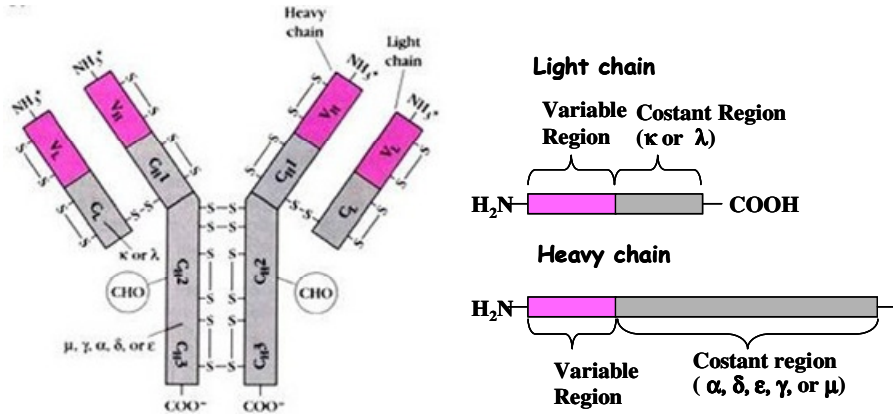


Fig. 1. Immunoglobulin structure. Schematic representation of an antibody molecole with two identical heavy (H) and light (L) chain. Both chain have variable (V) and constant (C) sub-domains which are hold together by disulfide bonds. The light chain have two subdomain (V_L-C_L) while the heavy chain have four subdomain (V_H-C_{H1}-C_{H2}-C_{H3}-C_{H4}).

2. The Immunoglobulin Heavy Chain (IgH) locus in mouse and human

In the mouse the Immunoglobulin heavy chain gene (IgH) cluster is localized on chromosome 12 and spans ~3 megabases. The variable region is close to the telomere and has ~ 101 V_H segments, followed by ~13D_H and 4 J_H segments. The mouse heavy chain constant region gene (IgHC) cluster, lying downstream of the variable region, is composed of eight functional genes, including four Cγ genes and one Cα gene (Cμ-Cδ-Cγ3-Cγ1-Cγ2b-Cγ2a-Cε-Cα). The human IgH locus is located on chromosome 14, spans ~3Mb and has the same general organization of the murine IgH cluster. There are ~39 functional V_H gene segments, followed by 26 D_H and 6 J_H segments. The human IgHC locus contains nine functional genes and two pseudogenes (Cμ-Cδ-Cγ3-Cγ1-Cψε-Cα1-Cψγ-Cγ2-Cγ4-Cε-Cα2), organized into two γ–γ–ε–α blocks. It has evolved through a series of duplications, followed by mutations and specialization of the new genes (Fig. 2). The locus is still evolving and up to 20% of the European population and 44% of the Asian population show duplications of single or multiple IgHC genes. The evolution of the human and mouse IgHC loci after the divergence of the two species has resulted in differences in the gene organization, the number of genes, and the function of selected IgHC genes. Thus, even if both species have the same IgG subclass encoding genes, the product of human and mouse are different in structure and function; CSR to IgG subclass or IgA is also differentially regulated in mouse and human. (Hammarstrom *et al*, 2007)

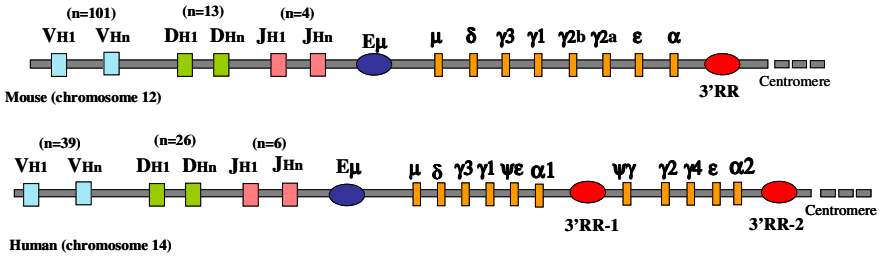


Fig. 2. The Immuoglobulin Heavy Chain (IgH) Locus in Mouse and Human. The map show the variable region (segment V-D-J), the enhancer Eμ, the constant gene and the regulative region at 3' end of the constant region (3'RR). The human IgH constant region is duplicate and two regulatory region (3'RR-1 and 3'RR-2) was identified 3' of each α constant gene.

3. B lymphocyte development

B lymphocyte development begins in the bone marrow and leads to the generation of mature peripheral B cell capable of producing secreted immunoglobulins. Developing B cells undergo an ordered V(D)J recombination process leading to productive assembly of V, D and J genes at the heavy chain locus, and assembly of V and J genes at the light chain locus (Diamant and Melamed, 2004) (Fig. 3). D_H-J_H rearrangements occur at the early pro-B stage, and subsequent V-DJ rearrangement commence at the late pro-B stage. Appropriate signalling through the pre-B cell receptor (BCR) mediates heavy-chain allelic exclusion and induces proliferation of large pre-B cell. Subsequently, at the small pre-B stage, light chain- rearrangement begins (Wang and Clark, 2003). At the immature stage, B cells expressing surface IgM undergo positive and negative selection events to promote the formation of signalling competent non-autoimmune repertoire. Then immature B cells emigrate to the periphery and undergo alternative splicing to produce IgD membrane form becoming mature naïve B cell. The final stage of B-lymphocyte development depends on antigenic stimulation and appropriate T-cell help provided by CD40 ligand. This process take place in the germinal center where activated B cells undergo somatic hyper mutation (SHM) and affinity maturation, and, in most cases, class switch recombination (CSR) to express IgG-IgA and IgE receptors (see below). Some of these cells differentiate into memory B cells and give rise to high affinity antigen-specific repertoire (Diamant and Melamed, 2004). Other cells differentiate into plasma-cell producing the immunoglobulin secret form in response to primary antigen exposure.

3.1 Class Switch Recombination (CSR)

Immunoglobulin (Ig) class switching, or isotype switching, is a process by which B lymphocytes shift from production of IgM to one of the IgG3, IgG1, IgG2b, IgG2a, IgE or IgA classes and subclasses in mouse, or to IgG3, IgG1, IgA1, IgG2, IgG4, IgE, and IgA2 in humans. This process is mediated by the deletional DNA recombination between the switch (S) region of the Ig heavy chain constant region μ gene (Sμ) and one of the downstream S regions located 5' to each C_H (except for the δ gene). CSR creates a novel transcriptional unit encompassing the original variable/diversity/joining (VDJ) fragment, plus the new heavy chain constant (C_H) gene, therefore

generates a new type of Ig molecule with original antigen-binding specificity and novel effector functions associated with the IgH chain. This process provides the basis for the versatile, humoral, immune functions of Ig molecules (Zhang, 2003) (Fig. 3). Transcription of the germline IgH gene, termed Ig germline transcription (GT), is the initial step for CSR. This process is activated and directed by cytokines and synergized by the costimulation of CD40. Ig GT appears to be optimized by the 3'RR via its locus control region (LCR) function providing for efficient GT and CSR. CSR process event suggest clearly that transcription through the S region plays a primary role in targeting CSR his own. Consequently, a mechanism influencing GL promoter's transcription is involved directly in the regulation of CSR (Manis *et al*, 2000).

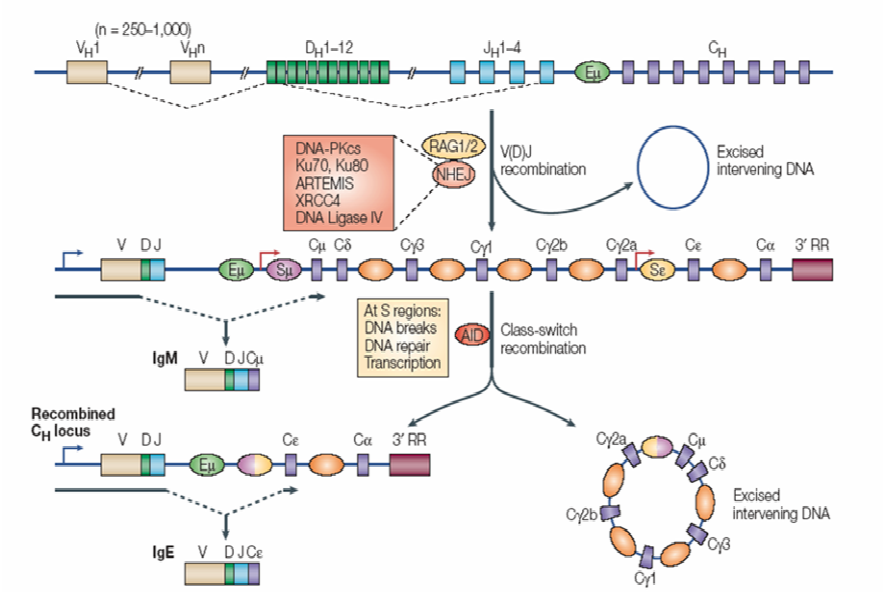


Fig. 3. Rearrangements of the immunoglobulin heavy chain locus. The variable region of the immunoglobulin heavy chain is assembled from the component variable (V_H), diversity (D_H) and joining (J_H) gene segments by V(D)J recombination. Transcription across the locus is driven by a promoter upstream of the rearranged VDJ segment (blu arrow), which allow the synthesis of a μ heavy chain. Then the heavy chain associate with a light chain forming an IgM molecule, which is display on the cell-surface of a B cell. Secondary isotypes are then produced by class-switch recombination mechanism (CSR), a process that exchanges the constant region of the heavy chain (C_H) with a set of downstream constant region gene (CSR to IgE is shown) (Chaudhuri and Alt, 2004)

4. Long-range gene activation

Gene activity is regulated in a tissue- and temporal-specific manner by *cis*-acting DNA elements such as promoters, enhancers and silencers; and *trans*- acting elements such as basal transcription factors, co-activators and chromatin modifying enzymes, which work together to activate transcription upon stimulation by specific signals. Enhancer elements activate high levels of transcription of linked genes from proximal or distal location. Although most enhancers are located tens of Kilobases (kb) away, some have been found at distance of up to a megabase from their gene target. Enhancer elements, therefore, have the potential to activate a number of neighbouring genes over a large chromosomal region. Most enhancers appear to be promiscuous and their action must be restricted in order to prevent the activation of non-target gene. DNA insulators are *cis*-elements founded to possess both enhancer blocking and heterochromatin barrier activities (West and Fraser, 2005). Locus control regions (LCRs) were functionally defined by their ability to enhance the expression of linked genes to physiological levels in a tissue specific and copy number-dependent manner at ectopic chromatin sites. Although their composition and locations relative to their cognate genes are different, LCRs have been described in a large spectrum of mammalian gene system (Li *et al*, 2002). LCRs typically contain several DNase I hypersensitive sites with enhancer activity or insulators elements. The discovery of the LCR in the β-globin locus and the characterization of LCR in many other loci reinforce the idea that the regulation of gene expression is not founded only on proximal elements such as promoters, enhancers and silencers, but also on long-range interactions of various *cis*-regulatory elements and dynamic chromatin alterations. The mechanism of long-range transactivation by the LCRs is poorly understood. Several models have been postulated: looping, tracking, facilitated-tracking and linking model (Li *et al*, 2002) (Fig. 4). All the models implicate the ability of LCRs to alter chromatin conformation to allow transcription factors binding LCR element to interact with transcription factors that bind promoter elements, enhancing gene transcription. Recent studies have shown that upon their transcriptional activation, genes can migrate beyond their chromosomal territories to foci enriched of RNA polymerase II and transcriptional machinery, known as transcriptional factories. In this model, gene promoters have been observed to contact to linked enhancer elements by a chromatin loop. Trans-acting factors recruited to enhancers and promoters determine interaction between looped chromatin fibres (West and Fraser, 2005).

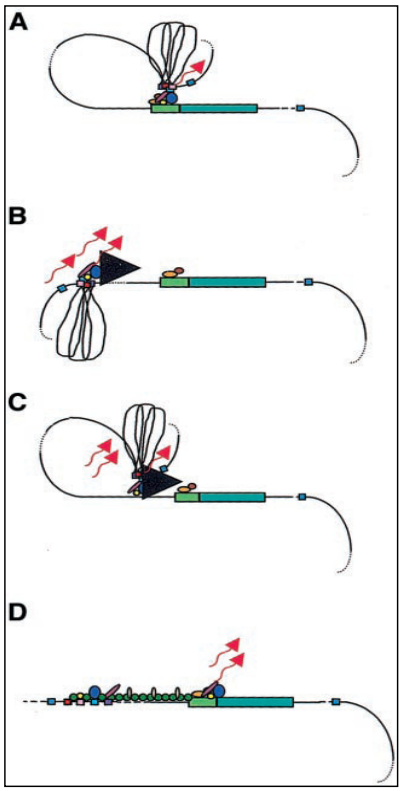


Fig. 4. Models of LCR function. (green rectangular box= putative gene; light green = promoter region; coloured ovals and circles = transcription factors; small coloured box = hypersensitive sites (HSs); wavy arrows = DNA transcripts. The flanking DNA sequences of the HSs are depicted as loops between the HS sites) (A) Looping model. The LCR directly interacts with the gene promoter by looping out the intervening DNA, allow the transcription factor binding the HSs to contact the transcription factor binding the promoter. (B)Tracking model. The transcription factor complex of LCR tracks down the DNA sequence (black arrowhead), until encountering transcription factor bound the gene promoter. (C) Facilitated tracking model. Looping process occurs to deliver the transcriptional factor binding the LCR proximal to gene promoter, then tracking process occur until they encounter the transcription factors bound to the gene promoter. (D) Linking model. Sequential bonding of transcription factor along the DNA directs changes in chromatin conformation and define the transcriptional domain. The transcription factors are linked to one another from the LCR to the gene promoter by non –DNA binding proteins and chromatin modifiers (shown as small green circles) (Li *et al*, 2002).

5. Regulatory elements of the immunoglobulin heavy chain locus

The regulation of human immunoglobulin heavy chain gene expression is incompletely understood, despite clinically significant conditions in which specific isotypes are inappropriately up- or downregulated, e.g., allergies due to inappropriate IgE response, and various forms of immunodeficiency associated with low IgA expression. Clearly, cytokines and interactions between B and T cells play a role in regulate isotype switching, and *cis*-elements in the IgH gene locus which mediate these effects have been described in mouse and human (Mills *et al*, 1997). Key *cis*-regulators of B cell-specific expression of IgH genes are V_H promoters, I region promoters located upstream of each heavy chain constant region gene, and two sets of enhancers: the intronic enhancer (Eμ) and a complex regulatory region that lies 3' of the IgH gene locus (3'RR) (Fig. 2). The immunoglobulin heavy chain enhancer (Eμ), located in the intron separating coding sequences for the heavy chain variable and constant regions, was the first transcriptional enhancer shown to function in a tissue-specific manner. This tissue specificity, along with the fact that Eμ was required for the efficient expression of cloned IgH genes, suggest that it was an essential controlling element in the developmentally-regulated expression of IgH genes (Lieberson *et al*, 1991). However, first observation in mouse myeloma cells supported the notion that additional enhancers might lie within the IgH locus (Zhang, 2003; Lieberson *et al*, 1991). The attention was focused on a region located downstream of the Cα gene. This region includes several elements displaying lymphoid-specific transcriptional enhancer activity in transient or stable transfection assay, and is known collectively as 3' regulatory region (3'RR). Relative roles of the two sets of enhancers are gradually being elucidated. Analysis of B cell lines that lack Eμ have led to the conclusion that Eμ is involved early in B cell development, e.g., in V to D-J joining and in μ heavy chain gene expression, and is dispensable at later stages of B cell differentiation, i.e., in plasma cell lines (Ernst and Smale, 1995); while the 3' enhancers are involved in process that occur late in B cell differentiation, including up-regulation of IgH expression and secretion, and class-switching (Zhang, 2003).

5.1 The murine 3' IgH Regulatory Region (3'RR)

The murine 3'RR spans ~ 28 kb downstream of Cα and harbours four enhancers, each associated with DNase I hypersensitivity : HS3a, HS1,2, HS3b and HS4 (Garrett *et al*, 2005). The first 3'IgH element to be identified, HS1,2, is located 13 kb downstream of the α gene membrane exon in the mouse. Whereas two DNase I-hypersensitive sites (HS1, HS2) were mapped within 3'αE, additional sites (HS3 and HS4) were identified 25 and 29 kb downstream of the α gene (Chauveau *et al*, 1998). The 3' IgH regulatory region consists of a long palindrome, including two copies of the very same enhancer (HS3a and HS3b) duplicated in inverted orientation on both sites of the HS1,2 enhancer (Saleque *et al*, 1997). HS3A maps immediately 3' of the Cα and upstream of HS1,2; while HS3b lies 13 kb downstream of HS1,2, and HS4 lies another 4 kb further 3' (Ong *et al*, 1998). Finally, the 11-kb region immediately downstream of HS4 contained the DNase I hypersensitive sites HS5, HS6 and HS7, that have been hypothesized to serve as an insulator of the IgH locus; and several CTCF target sites, a protein associated with mammalian insulators.(Garrett *et al*, 2005) (Fig 5). Several studies revealed that the potential physiological function of these 3' Igα enhancers includes regulation of the IgH locus rearrangement, transcription and Ig CSR, although additional roles could not be excluded. The individual HS fragments tested are able to synergize, although differentially, with the V_k, V_H, and IgH germline promoters, including the germline promoter for γ2b,γ3,α,and ε, and non-Ig promoters (such as the c-myc promoter) so as to enhance the promoter-directed transcription activity in transient transfection assays while demonstrating cell-type and development stage specificity (Zhang, 2003). In particular HS1,2 has been presented as specifically active in activated B lymphocytes and plasma cells and it is trans-activated by mitogens such as LPS and PMA in resting B cells (Laurencikiene *et al*, 2001; Andersson *et al*, 1998). HS3b and HS4 elements are essential in promoting both CRS and regulation of GT, and pair wise deletion of HS3b and HS4 had a dramatic effect on both the process (Pinaud *et al*, 2001); while deletion of HS4 only has a moderate effect (Zhang *et al*, 2007). HS3a shows only weak enhancer activity. However, both pairs (HS3A-HS1,2 and HS3B-HS4) of enhancers efficiently increase expression of GL ε promoter (Laurencikiene *et al*, 2007). Based on transient transfection experiments, it was concluded that the four enhancers are rather weak when they individually drive transcription of reporter genes. However, their combinations displayed strong transcriptional synergies, especially

when their normal palindromic arrangement was in respected (Ong *et al*, 1998; Chauveau *et al*, 1998). Moreover, in stably transfected plasmacytoma cell lines, a cassette bearing HS1,2, HS3b, and HS4 could confer high-level, tissue–specific expression of a linked c-myc gene in a position-independent and copy-dependent manner, suggesting that the 3' regulatory region could act as a LCR (Madisen and Groudine, 1994).

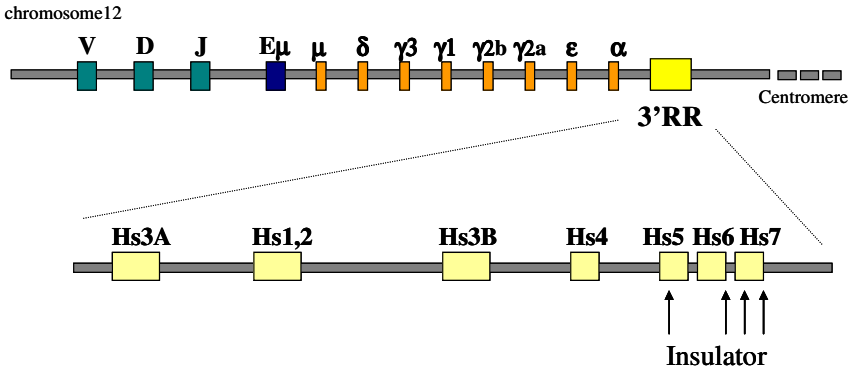


Fig. 5. Murine 3' IgH Regulatory Region (3'RR). In the mouse IgH locus the 3'RR is located downstream of the Cα constant gene and harbours several DNase I hypersensitive sites (HS). HS3a, HS1,2; HS3b, and HS4 are enhancers involved in the expression and class switching of immunoglobulin heavy chain gene, while HS5, HS6 and HS7 have been hypothesized to serve as an insulator of the IgH locus.

5.1.1 Regulation of murine 3'RR activity

During the immune response, B cells switch from expression of IgM to IgG, IgA or IgE. This process involves two events: activation of sterile GL transcripts from specific C_H gene and DNA recombination. Gene targeting experiments indicate that synthesis of GL transcripts is required for switch recombination to occur. External stimuli (cytokines or direct cell-cell contact) direct CSR to a particular C_H gene via their ability to modulate GL transcription. Different promoters can be activated by a different set of stimuli, in this way the specificity of CSR in response to external stimuli is assured. External signals have also varying effects on specific IgH locus

3'enhancers. Many transcriptional-binding sites have been identified in the 3'Igα enhancer regions, some of that confer the inducible activity of the enhancers by the external stimuli such as LPS, IL4, and αCD40, which activate the germline transcription (Laurencikiene *et al*, 2001).
HS1,2 contains several activation and repression elements. Proteins that bind E5; NF-κB; NFE; Oct; NF-αP (ets-like site), NF-AB (Ets/AP-1) motifs positively regulate HS1,2 in activated B and plasma cells (Laurencikiene *et al*, 2001; Michaelson *et al*, 1996). The contribution of E5 site to enhancer activity is inhibited in early stage of development by the dominant negative regulator Id3, which is expressed in early B cells but downregulated in plasma cells (Mills *et al*, 1997; Zhang, 2003). NF-κB is a key regulator of B cell development. NF-κB subunits, especially p50 and c-Rel, have been shown to play critical and different roles in regulating B cell proliferation and CSR. It has been shown that p50 selectively augment HS1,2 activity in LPS activated B cells, whereas c-Rel is required for optimal HS1,2 induction in B cells activated through CD40 (Zelazowski *et al*, 2000). However additional transcription factors are required to allow NF-κB to act as a strong activator. For example, PU.1 interacts with NF-κB (p50/c-Rel) and the two proteins can co-operate to activate enhancer linked reporter genes (Linderson *et al*, 2001). It was also been identified binding sites within the HS4 for NF-κB; Oct-1; Oct-2 and BSAP (Michaelson *et al*, 1996). BSAP has been shown to suppress HS1,2/HS4 activity in B cell lines that correspond to early differentiation stages. In contrast, there is evidence that HS1,2/HS4 activity is not blocked by BSAP in activated mature B cell. (Laurencikiene *et al*, 2001). NF-κB has been demonstrate to be important for the enhancer activity of HS4, as the HS4 fragment with the NF-κB mutation blocks the enhancer activity for c-myc promoter-directed transcription (Michaelson. *et al*, 1996). OCA-B is a coactivator for Oct-1 and Oct-2 also involved in the positive regulation of 3' enhancers activity (Zhang, 2003). A number of HS3-binding common transcription factors such as OCT, NF-κB, and BSAP have been identified. In addition, YY1 is a specific activator for HS3 and Eμ enhancer. Retinoblastoma protein (Rb) inhibited the binding to HS3 in resting cells, while upon stimulation with class-switching activators, Rb becomes hyperphosphorylated, YY1 is released and can bind HS3 (Gordon *et al*, 2003). Therefore specific binding sites of the 3'enhancers may selectively interact with different promoter in a differential stage-specific manner; and specific factors seems to contribute to enhancer activity in different ways, depending on the promoter (Laurencikiene *et al*, 2001). On the other hand murine 3' enhancers have a common regulatory mechanism because they

share binding of a common group of transcription factors (Michaelson *et al*, 1996). Epigenetic changes could also regulate the activity of the 3' RR. Since before the discover of the structure of murine 3'RR became clear that the regulative region was hypermethylated in pre-B and T cell lines and became progressively demethylated as B cell differentiation continued (Giannini *et al*, 1993). Then, by measuring levels of acetylated histones H3 and H4 and of dimethylated H3 (K4) with chromatin immunoprecipitation assay, it was found that, early in B cell development, chromatin encompassing the enhancers of the 3'RR began to attain stepwise modifications typical of an open conformation (Garrett *et al*, 2005). It has been shown DNA methylation and histone modifications of the 3'RR are co-ordinately regulated only in B cells by the B cell-specific transcription factor BSAP and histone H1 (Giambra *et al*, 2008). The HS4 enhancer was associated with active chromatin initially in pro- and pre- B cells and then together with HS3A, HS1,2 and HS3b in B and plasma cells (Garrett *et al*, 2005). In contrast to HS4, the enhancers HS3a, HS1,2 and HS3b are demethylated only late in development, in particular demethlylation of HS1,2 was detect only in plasma cells (Giambra *et al*, 2008). A biphasic model for the activity of this region was proposed: one unit, consisting of the 3'-most enhancer, HS4, is active early and throughout B cell development. The second unit, which comprises HS3a, HS1,2 and HS3b, becomes active later in development, when it contributes to such processes as class switching and increased levels of Ig heavy chain gene transcription in plasma cells (Saleque *et al*, 1997).
Therefore the 3'RR enhancer's activities and the specific stage and temporal activation patterns are governed by epigenetic mechanisms and proteins interacting with multiple binding sites identified within the individual elements (Andersson *et al*, 1998).

5.1.2 Long range activation by murine 3'RR

The 3'RR is a distal regulatory element undergoing the mechanism of long-range activation to stimulate the Ig heavy chain genes transcription. The transcription factor binding the enhancer may cooperate to form a specific activing surface. This surface would also interact with activating proteins in the promoter region and with the basal transcription machinery (Laurencikiene *et al*, 2001). Insertion of a PGK-*neo* cassette into the IgH locus short-circuits the ability of the 3' IgH regulatory region to facilitate

germline transcription of dependent IgH genes upstream but not downstream the insertion. This observation support the existence of a long range 3' IgH regulatory region required for Ig GL transcription and CSR to multiple IgH genes (Samara *et al*, 2006). The model of long range interaction of the LCR with the promoters is not clear. Recent studies show that V_H-3'RR interaction and myc-3'RR interactions involve a physical contact between the intervening DNA element, and a DNA looping process is postulated to occur (Ju *et al*, 2007). On the other hand, a looping model was excluded for the I-3'RR interaction, and a linking-tracking model was hypothesized to be the correct mechanism (Oruc *et al*, 2007). The long range interaction suggests that a chromatin remodelling mechanism and modification of the accessibility of the interacting surface occurs. Epigenetic changes are involved in CSR regulation and Ig GL transcription by *cis*-acting elements. Qin and Tang demonstrate that both LPS and CD40 signalling cause SWI/SNF complex (a complex altering nucleosome conformation in an ATP dependent manner) to dissociate from HS1,2 and associate with their responsive IγG2 GL promoter, suggesting a specific regulatory mode for the ATP-dependent chromatin remodeler during CSR. Increase of overall histone acetylation was parallel to the association of SWI/SNF complex at IγG2b GL promoter. Moreover, up-regulation of transcriptional activity in a linked c-myc gene by the murine 3' RR enhancers is accompanied by a widespread increase in histone acetylation along the linked gene, indicating that the 3' enhancers may regulate gene expression through histone hyperacetylation (Zhang, 2003).

5.2 The human 3' IgH Regulatory Regions

The human and mouse enhancer elements in the 3' regulatory regions share significant homology (Mills *et al*, 1997) and have been proposed to have similar functional properties. There are, however, significant differences (Sepulveda *et al*, 2005). The partial duplication encompassing the four gene γ3, γ1, ψε, α1 in apes did generate the cluster γ2, γ4, ε, α2 at the 3' end of the ψγ gene around 60 million of years ago (Harindranath *et al*, 1998). The duplication of the cluster of the constant heavy genes in the evolution of primates creates new opportunities for the immunoresponse not only with a double repertoire of heavy genes but also with the new 3' enhancer complex (Kawamura *et al,* 1992). There are two 3' enhancer regions in humans: RR-1 at the 3' end of the Cα-1 gene and RR-2, at the 3'

end of the Cα-2 gene (Mills *et al*, 1997) In respect to the mouse, HS3b enhancer element is absent and the 3'enhancers are organized differently (Sepulveda *et al*, 2005). The human α1 and α2 HS1,2 enhancers, known as HS1,2A and HS1,2B respectively, both reside near the centers of ~10 kb palindromes, with each palindrome closely flanked by a single copy of HS3 immediately adjacent to the 5'end and an HS4 unit located ~4kb downstream (HS3-HS1,2-HS4) (Hammarstrom *et al*, 2007) (Fig. 6). The different arrangement of the 3'α enhancers in mice and human (HS3A-HS1,2-HS3B-HS4 *versus* HS3-HS1,2-HS4) may cause some functional differences in these control region. The genomic structure and the sequence of two 3'RRs is highly conserved (Fig. 7). The major structural difference between the RR-1 and RR-2 is that the DNA segment contains HS1,2 is inverted between the two loci (Mills *et al*, 1997). The 3' human enhancers might be regulated differentially from that of the equivalent mouse enhancer. Similarly to the mouse, human HS4 is active from the early stages of the B cell lineage onward, but, unlike the murine model, BSAP binding is not an essential feature for HS4 activation in human pre-B (Mills *et al*, 1997). HS1,2A and HS1,2B fragments show equally strong enhancer activity on the GL α1 e α2 promoters in both orientations when transiently transfected into a mature B cell line. However, there is no activity in human pre-B cell line and T cell line. HS3 shows no enhancer activity by itself in any of the cell lines, whereas a modest effect is noted using HS4 in the three mature B cell lines. However, the combination of the α2HS3-HS1,2-HS4 fragments, which together form a potential locus control region, display a stronger enhancer activity than the individual fragments on α1, α2, and γ3 promoters; with a different effect on the α1 and α2 promoters as compared with the γ3 promoter (Hu *et al*, 2000).

The human HS1,2 elements are the strongest enhancers of the 3'RRs. The transcription factor HoxC4,Oct-1/Oct-2 (homeodomain proteins) and the coactivator Oca-B act synergistically to effect the HS1,2 enhancing activity. The complex recruitment is negligible in pro-B cells, moderate in pre-B cells and maximal in germinal center B cells and plasma cells, where HoxC4, Oct-1/Oct-2/OcaB expression correlates with HS1,2 activation (Kim *et al*, 2004). As HS1,2 carries a dominant role over HS3 and HS4 in the overall activity of the 3'RR, the structure and the function of HS1,2A and HS1,2B enhancers have been further investigated.

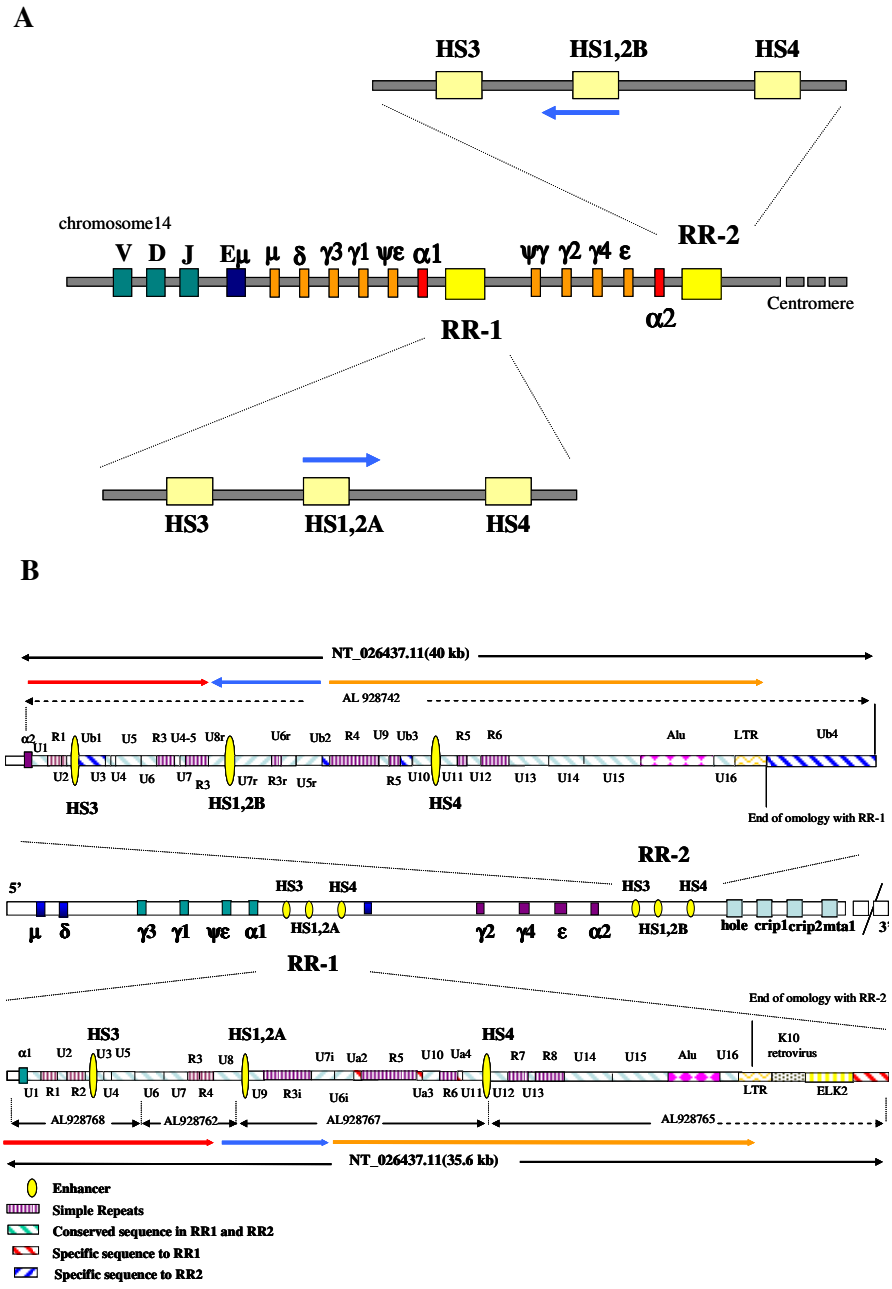

A
HS3
HS1,2B
HS4
chromosome14
V D J Eμ μ δ γ3 γ1 ψε α1
ψγ γ2 γ4 ε
RR-2
Centromere
α2
RR-1
HS3
HS1,2A
HS4
B
NT_026437.11(40 kb)
AL 928742
α2 U1 R1 Ub1 U5 R3 U4-5 U8r U6r Ub2 R4 U9 Ub3 R5 R6 Alu LTR Ub4
U2 U3 U4 U6 U7 R3 U7r R3r U5r R5 U10 U11 U12 U13 U14 U15 U16
HS3
HS1,2B
HS4
End of omology with RR-1
RR-2
5'
HS3 HS4 HS3 HS4
μ δ γ3 γ1 ψε α1 HS1,2A γ2 γ4 ε α2 HS1,2B hole crip1 crip2 mta1 3'
RR-1
End of omology with RR-2
HS3 HS1,2A HS4
α1 U2 U3 U5 R3 U8 U7i Ua2 R5 U10 Ua4 R7 R8 U14 U15 Alu U16 K10 retrovirus
U1 R1 R2 U4 U6 U7 R4 U9 R3i U6i Ua3 R6 U11 U12 U13 LTR ELK2
AL928768 AL928762 AL928767 AL928765
NT_026437.11(35.6 kb)
Enhancer
Simple Repeats
Conserved sequence in RR1 and RR2
Specific sequence to RR1
Specific sequence to RR2

Fig. 6. Human 3' IgH Regulatory Regions (3'RR-1 and 3'RR-2). (A) In the human IgH locus two 3'RR (3'RR-1 and 3'RR-2) are located downstream of each Cα constant gene (Cα1 and Cα2). Each regulative region harbour three enhancer elements: HS1,2; HS3; and HS4. HS1,2A element of RR-1 is inverted in respect HS1,2B element of RR-2. (B) The genomic structure and the sequence of two 3'RRs is highly conserved. The human 3'RRs consist of simple repeated (R); unique sequence conserved (U) and unique sequences specific for RR-1 (Ua) or RR-2 (Ub). The enhancers elements, the coding regions (i.e. ELK2) or other elements (i.e. ALU) are also shown. Four genes (hole, crip1, crip2, mta1) maps at the 3' end of RR-2.

5.2.1 Polymorphic structure of the HS1,2A and HS1,2B enhancers

Among the enhancers, only HS1,2 show polymorphic features, while HS3 and HS4 polymorphism were not described. Perhaps the generation of HS1,2 multiple alleles both in rodents and humans has likely been favoured by its central position within a large palindromic region (Guglielmi *et al*, 2004). The HS1,2A enhancer is located 9kb downstream of the human α1 gene, while HS1,2B is located 11 kb downstream the α2 gene. The selective amplification and cloning of HS1,2A and –B enhancers revealed the occurrence of four alleles of HS1,2A (1A; 2A; 3A; 4A)[1] and two alleles of HS1,2B (3B; 4B). The polymorphic sequences are present immediately at the 3' to the core of the enhancer. The alleles vary on the number of copies (one to four) of a conserved 38-bp element. The repetitions of the 38-bp element are separated by variable spacers of 20 ,16, and 14bp. In all alleles, the polymorphic sequences are bordered by two external 17-bp elements, with exception of alleles 3B, 4A and 4B. In these alleles, the external 17-bp element next to the core of enhancer HS1,2 is replaced by a 31-bp element (Giambra *et al,* 2005) (Fig. 7). The 135-bp HS1,2 core homology sequence contains motif important for B cell-specific enhancer's activity identical to the murine AP1, OCT, E5, ETS, although their function has not yet been demonstrated. Additional binding sites for several transcription factors are situated outside the core, within the polymorphic sequence (Mills *et al*, 1997). An obvious consequence of different HS1,2 alleles resides in the variation of the number of copies of the consensus sequences for several transcription factors. The number of sites for NF-κB/P50 varies from one to four in the six alleles (Giambra *et al,* 2005). This can be of primarily importance because different HS1,2 alleles can influence HS1,2 activities in

[1] *In this thesis we refer to the HS1,2A alleles simply as 1, 2, 3 and 4.*

the regulation of germline transcription and immunoglobulin production. Preliminary data have shown that HS1,2A RFLP polymorphism contribute to the genetic predisposition to several diseases characterized by alteration of B cell activation and Ig production.

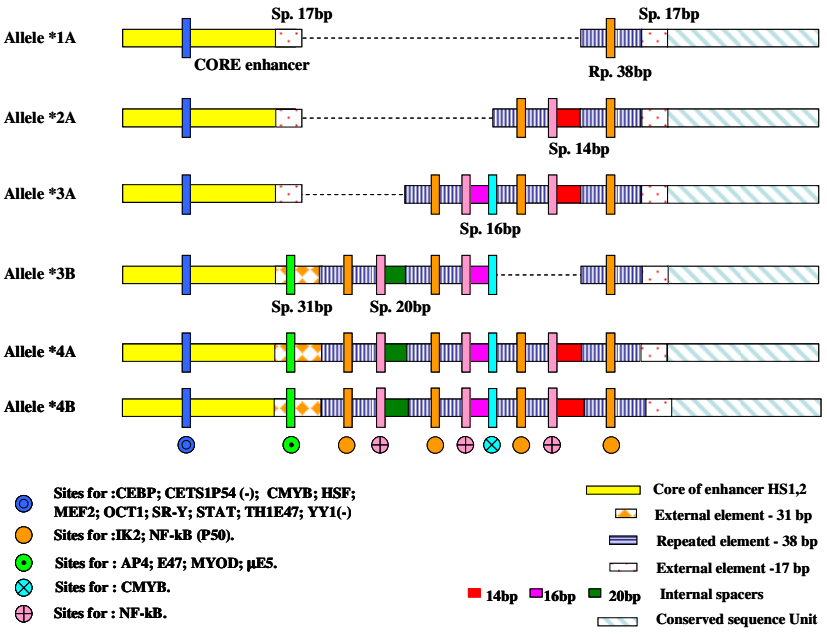


Fig. 7. HS1,2A and HS1,2B alleles. There are four HS1,2A alleles (1A; 2A; 3A; 4A) and two HS1,2B alleles (3B; 4B). The core of enhancer, the repeated element of 38 bp, the external elements of 31 and 17 bp and the internal spacers of 14, 16, and 20 bp are mapped. Binding sites for several transcription factors are shown (Giambra V. *et al*, 2005).

5.2.2 Functional role of the HS1,2A enhancer

Even before the discover of HS1,2A allelic polymorphism, a significant association between several HS1,2A genotype and circulating levels of IgA1 and IgA2 have been reported (Denizot *et al*, 2001). An other study has shown that dysregulation of IgA production in IgA nephropathy patients is

associated with HS1,2A polymorphism, in particular one allele is significantly correlated with a milder form of the disease, whereas another one is associated with severe evolution leading to renal failure (Aupetit. *et al*, 2000). The mechanism leading to IgA dsyregulation is not known, although lymphocyte proliferation and Ig production, or hyper-production, could be related to cytokines inducing NF-κB recruitment. So NF-κB could modulate the function of HS1,2A polymorphism leading to the different clinic features of the disease. Following studies show that homozygosity of allele 2 in Celiac Disease, which is an autoimmune disease characterized by a dysregulation of IgA antiendomysium antibodies (anti-EMA), increase of 30. 8% as compared to the control group (Frezza *et al*, 2004). Moreover, a recent study on systemic sclerosis (SSc), an autoimmune disease characterized by an important B cell activation, shows an increased frequency of the allele 2 in the group of patients high significant in respect to the control groups (Frezza *et al*, 2007) These data suggest an intriguing function of HS1,2A polymorphism in the regulation of B cell activation and Ig production. One aim of this work is to investigate the role of HS1,2A polymorphism as genetic factor increasing the risk to develop several diseases associated with alteration of B cell functions (see below).

5.2.3 World distribution of the HS1,2A allele frequencies

The involvement of HS1,2A in the predisposition to several immune disorders have led to know the distribution of HS1,2A variants in population from different continents, and to determine whether the polymorphism is associated to specific evolutionary factors. HS1,2A alleles 3 and 4 show the highest frequencies among African populations, and allele 2 is significantly lower in Africa in comparison with both European and, to a lesser extend, Asian population. HS1,2A allele 1 does not seem to differentiate the major human groups. Analysis of molecular variance of the allele frequencies indicate that HS1,2A polymorphism can be considered as a reliable anthropogenetic marker (Unpublished data; Giambra *et al*, 2006) (Fig. 8). This enhances the importance of this polymorphism in epidemiological studies related to immune-disorders, especially in view of the current environmental modifications, such as changes in diet, and exposure to pesticides, poisons, and particular drugs, to which these peoples could go towards. On the other hand, HS1,2B allelic frequencies are not variable across the continents.

One aim of this work is to investigate HS1,2A allelic distribution in several Italian regions in order to understand the genetic micro-variation of the whole of Italy. Another aim is to study the distribution of HS1,2A alleles in several European healthy populations living in geographic isolation condition in order to verify if inter-population genetic flux is impaired by natural barriers.

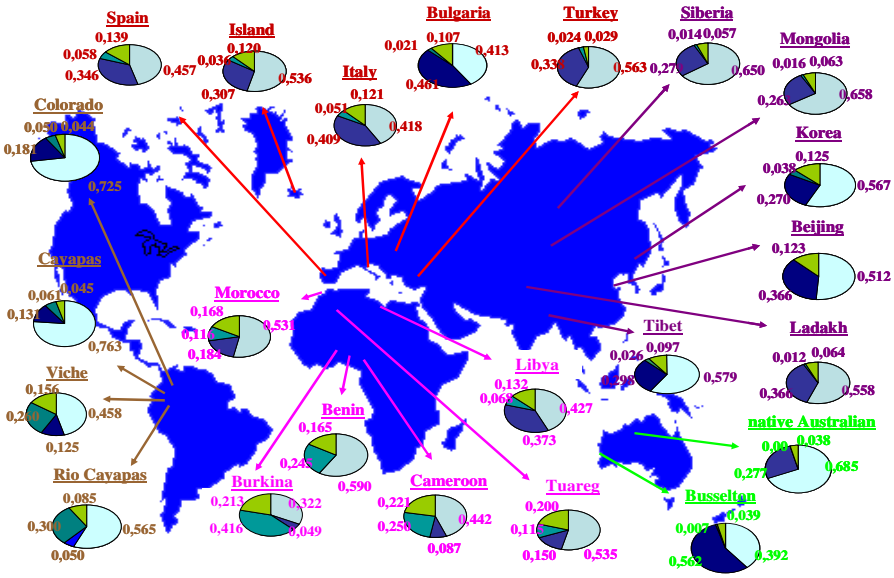


Fig. 8. World distribution of the HS1,2A allele frequencies. The HS1,2A allele frequencies vary through world populations according to the belong to geographical groups.

6. Genetic variants: polymorphisms and haplotypes

Genetic factors are known to have an important role in many common diseases, and the identification of genetic determinant for such disease could provide insight into disease pathogenesis, revealing new therapeutic strategies (Kingsmore *et al*, 2008). More than thousand genes for heritable "mendelian" disorders have been identified, in which variation in a single gene is both necessary and sufficient to cause disease. Common disorders, in contrast, are thought to be determining by the combined effects of many different DNA variants interacting with environmental factors (The International Hap Map Consortium, 2005). The case-control analysis compares the allelic frequency differences of a particular polymorphism between the case and the control groups. There are various type of polymorphisms such as deletions, insertions, restriction fragment length polymorphisms (RFLP) and single nucleotide polymorphism (SNP).

Single nucleotide polymorphisms (SNP) are the most common form of the genetic variation in the human genome. In the recent years, there has been a growing interest in using SNPs to map the susceptibility genes for complex disease (Lee *et al*, 2005). However, usually complex diseases have relatively low genetic components and multiple genes involved in pathogenesis. The human genome has about 10 million of SNPs, so the determination of the SNPs influencing a common disease could appear very difficult. Thankfully, there are often strong statistical associations among polymorphisms in the human genome, such that the presence of a particular variant at one site on a chromosome can predict the presence of a particular variant at another site. This correlation is known as linkage disequilibrium (LD); a particular combination of allele along the chromosome is termed "haplotype". (Goldstein and Cavalleri, 2005). In the recent years the International Hap Map Consortium developed the HapMap Project which has had the purpose to discover the association between these variants. Thanks to HapMap project 3.1 million SNPs have been genotyped from 269 DNA samples. The resulting HapMap has a SNPs density of around one per Kilobase and is estimed to contain approximately 25-35% of all the 9-10 million common SNPs (minor allele frequency (MAF)≥0.05) of the human genome (The International Hap Map Consortium, 2005; The International Hap Map Consortium, 2007).

Genome-wide association studies (GWA) use SNPs data from the HapMap project to investigate on allelic frequency differences between cases and controls (Kruglyak, 2008). A recent work has examined around 14000 cases (2000 individuals for each of 7 common disease) and a shared set of around 3000 controls, observing association at many previously identified loci and discovering several additional susceptibility ones (The Wellcome Trust Control Consortium, 2007).

Even if the Hap Map projects have had been useful for the investigation on common variants, the rare ones will be more difficult to discover. Moreover, the variants within one Kb region will not be detected. These variants can be identified only by genome re-sequencing. In addition, whereas recognition of functional variants in coding regions is easy, detecting functional change in non-coding DNA is more difficult. This is because we do not have a ready connection between nucleotide differences and function for these sequences, and because regulatory regions can be located far from the coding region so are often difficult to identify (Kruglyak, 2008). Therefore, genetic predisposition conferred by non-coding regions such as LCR; enhancer or other regulative regions have been poorly investigated. Nonetheless, it could be very interesting to understand how sequence change can influence regulative region function. Polymorphisms or haplotypes could affect epigenetic change or transcription factor binding to 3'RR, influencing the regulative region activity in Ig germline transcription, CSR stimulation and Ig production. In this light, 3' regulatory region polymorphisms and haplotypes could contribute to predisposition to autoimmune disease or other diseases characterized by a dysregulation of Ig production.

The central aim of my thesis is to investigate the role of HS1,2A polymorphism and other genetic variants of the 3'RR-1 in the immunity-system dysregulation of several diseases. Case-control analysis was performed comparing HS1,2A allelic frequencies of several immunological diseases in respect to the control groups. Moreover, the correlation of HS1,2A allele frequencies with the alteration of serum Ig concentrations in two different groups was investigated. The role of HS1,2A polymorphism in the progression of viral infections was also evaluated.

Finally, genome re-sequencing of the non-coding regions surrounding HS3 and HS1,2A enhancers was performed in order to discover polymorphisms and haplotypes associated with HS1,2A alleles in a control group and in a cohort of psoriatic patients.

7. Immunological diseases

7.1 Autoimmune diseases

Autoimmune diseases (AID) are complex multisystem and multifactorial disorders. While many cell types participate and contribute to the autoimmune disorders, lymphocytes are the key regulators in the initiation and propagation of the autoimmune response. The production of pathogenic autoantibodies during an autoimmune response is an indication that one or more autoreactive B cell clones have circumvented central and/or peripheral tolerance check-points (Diamant and Melamed, 2004). Autoantibodies might not be directly responsible for many of the manifestations of autoimmune disease, but they are markers of possible future disease in presently healthy individuals (Tab. 1). Susceptibility to autoimmune disease is a multigenic phenotype affected by a variety of changeable factors as microbiological factors, smoking, diet, Vitamin D deficiency, Ultraviolet light exposure and psychological stress; and unchangeable factors such as autoantibodies, genetics factors and primary autoimmune disorders (Shepshelovich and Shoenfeld, 2006).

Numerous studies have proved that the genetic predisposition is a main factor in the susceptibility to develop AID. The heritage model of AID is complex and multiple genes are involved in its aetiology.

The most potent genetic influence on susceptibility to autoimmunity is the Major Histocompatibility Complex (MHC). The human MHC is rather small (3.6Mb), but it is densely packed with immunologically important genes in strong linkage disequilibrium with each other. More than 30 autoimmune disease show association with alleles of the ancestral HLA haplotype 8.1 (Lie and Thorsby, 2005). On the other hand, different HLA alleles are often linked to different autoimmune diseases (Tab. 2). The class II MHC allele HLA-DR4 and related alleles are known to be the major genetic risk factors for rheumatoid arthritis (RA). Early studies showed that 70% of patients with classic RA express HLA-DR4 compared with 28% of control individuals (Shepshelovich and Shoenfeld, 2006).

However, genetic predisposition conferred by HLA is not enough to the development of the autoimmune disorders. AIDs are complex diseases that require the confluence of many genes and environmental factors. Linkage studies done both in humans as in rodents have clearly indicated that non-MHC genes also contribute to the susceptibility to develop AID. (Serrano *et*

al, 2006). A new finding indicates that several polymorphisms affect multiple autoimmune diseases. OCTN haplotype is associated both with Crohn's Disease (CD) and psoriatic arthritis (PA), suggesting that these 2 diseases may share some common genetic control in pathways of inflammation (Hon *et al*, 2005). On the other hand several pathophysiological mechanisms are specific to a particular disease. IBD5 risk haplotype within the chromosome 5q31-33 is associated with CD only (Armuzzi *et al*, 2003), while two distinct risk haplotypes for psoriasis on the same chromosomal region have been identified (Li *et al*, 2008).

In the last years GWA studies have proven powerful in identifying SNPs that are significantly associated with multi-factorial diseases. A number of putative susceptibility loci for various autoimmune diseases have been identified, but in many cases the contribution of these areas to disease phenotype are unknown. Obviously, the effect of genetic factors on the development of autoimmune disorders could not be overestimated. The presence of a specific autoantibody in an individuals serum combined with MHC haplotypes or other susceptibility loci increases exponentially the risk for that person to develop an autoimmune disease in the future. The identification of the genes and genetic pathways involved in these diseases will allow elucidating the mechanism triggering the pathogenesis and open the possibility to new therapeutic options.

One aim of this work is to study HS1,2A polymorphism association with several autoimmune disease such as organ-specific illnesses (*dermatitis herpetiformis*) and systemic illnesses (rheumatoid arthritis, systemic lupus erythematosus, crohn's disease and psoriasis).

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

Tab. 1. Selected autoimmune disease and characteristic autoantigens. (Shepshelovich and Shoenfeld, 2006)

Autoimmune disease	*MHC antigen*	*Relative risk*
Ankylosing spondylitis	B27	69.1
Reiter's syndrome	B27	37.0
Rheumatoid arthritis	Dw4/DR4	3.8
Sjögren's syndrome	Dw3	5.7
Systemic lupus erythematosus	DR3 (Caucasian)	2.6
	DR2 (Japanese)	5.3
	DQ3 (Chinese)	11.5
Celiac disease	DR3	11.6
Pernicious anemia	DR5	5.4
Pemphigus vulgaris	DR4 (Jews)	14.6
Type 1 diabetes mellitus	DR4	3.6
	DR3	4.8
	DR2	0.2
	BfF1	15.0
Graves disease	DR3	3.7
Addison's disease	Dw3	10.5
Hashimoto's thyroiditis	DR5	3.2
Multiple sclerosis	DR2	6.0
Goodpasture's syndrome	DR2	15.9

Tab. 2. MHC antigens and related autoimmune disease (Shepshelovich and Shoenfeld, 2006)

7.1.1 Organ-specific illnesses

Dermatitis herpetiformis (DH) is a multi-factorial cutaneous disease like psoriasis and psoriatic arthritis. Patients with DH, which is sometime referred to as the cutaneous phenotype of CD, have asymptomatic or frank CD and a skin rash with anti-transglutaminase cutaneous IgA deposits. CD and DH both show anti-transglutaminase type 2 (TGM2) autoantibodies. So CD and DH are closely related disease which share a genetic background and a common pathogenesis characterized by gluten sensitivity and production of auto-antibodies (Karrell *et al*, 2002; Hervonen *et al*, 2000). HS1,2A polymorphism association with CD has previously been evaluated. In this thesis we investigated the role of the polymorphism as genetic factor increasing the risk to develop DH.

7.1.2 Systemic illnesses

Rheumatoid arthritis is a systemic autoimmune disease characterized by chronic inflammation of senovial tissue that often leads to a progressive joint destruction. RA patients can be divided into two cohorts: patients with early RA (ERA, disease duration<12 months) and patients with long disease duration (more than one years disease duration, long standing RA (LSRA). Studies in twins clearly show a genetic contribution to disease susceptibility (MacGregor *et al*, 2000)), and the most important genetic risk factor is HLA region (MacGregor *et al*, 1995; Huizinga *et al*, 2005), even if other important genetic associations have been evaluated (Begovich *et al*, 2004; Johansson *et al*, 2006). Since CSR could be important in autoimmune chronic inflammatory diseases (ACIDs) we tested the hypothesis that the HS1,2A locus polymorphism could be involved in RA pathogenesis, influencing a small or long disease duration.
Psoriasis (PS), a clinically heterogeneous disease with inherited susceptibility, is considered a T-cell mediated inflammatory skin disease, but humoral immune process has also been evaluated. PS has been variously reported associated with IgA-related disorders, such as CD, DH, IgA nephropathy and others (Lapeyre *et al*, 2006; Kallel Selami *et al*, 2006; Zadrazil *et al*, 2006). Up to 16% of psoriatic patients show presence of IgG and IgA celiac associated antibodies (antiendomysial, antigliadin, antitransglutaminase) (Woo *et al*, 2004). Psoriasis can be divided into plaque psoriasis (PP) and psoriatic arthritis (PA). In comparison to plaque psoriatic

ones, the PA patients seem to have an increase number of intraepithelial lymphocytes of the duodenal mucosa. HS1,2A polymorphism could modulate CSR process and Ig production leading to different clinic feature of the two diseases.
Systemic lupus erythematosus (SLE) is a chronic, inflammatory autoimmune disorder. The exact causes of SLE are not known. The immune system produces especially autoantibodies against nuclear DNA. The autoimmune reaction causes the inflammation that affects specific organs (joints, kidneys, skin, etc) (Zhang *et al*, 2008). Multiple genetic risk factors, combined with random environmental factors, are thought to be responsible for the abnormal immune response. The involvement of HS1,2A polymorphism in developing SLE was evaluated in this work.
Crohn's disease is a chronic inflammatory disease of the intestines. In patients with CD the immune system is abnormally and chronically activated, resulting in chronic inflammation and ulceration. It is unknown if the abnormality in the functioning of the immune system is a cause, or a result, of the disease. It has been shown that the inflammation involves several factors: genetic factors, the immune system itself, and the environment (i.e. foreign substances). Recently a gene called NOD2 has been identified as being associated with CD (Sathiyasekaran and Shivbalan, 2006). One aim of this work is to study the distribution of HS1,2A alleles in a group of CD patients in comparison with a group of patients affected by a non-immunological disease, namely ulcerousis recto colitis (RCU).

7.2 IgA deficiency

Selective immunoglobulin IgA deficiency (IGAD) is the most common primary immunodeficiency, with a prevalence of about 1/600 in European populations. The genetic basis of IGAD is poorly understood and susceptibility loci have been described, such as IGAD 1 located at 6p21 (Vorechovsky *et al*, 1999). Moreover, mutations in several factors involved in class switch recombination control are associated with IgA defect and other Ig defects (Martin and Dixit, 2005; Salzer *et al*, 2005). The correlation between HS1,2A allele frequencies and the risk to develop IGAD was analyzed in this work.

8. Alteration of serum immunoglobulin levels in schizophrenic patients and healthy subjects

Schizophrenia is a severe psychiatric disorder that affects 1% of world's population. The immune abnormalities reported in schizophrenia include morphological change in lymphocytes, altered levels of CD4+ and CD8+ T cells, increased levels both of circulating cytokines and antiviral antibodies. IgG, IgM, IgA, IgE antibodies were altered in schizophrenic patients (Rothermundt *et al*, 2001; Jones *et al*, 2005). Alteration of immunoglobulin concentration could be detected also in healthy subjects, even if it not be considered a pathological condition (data from Massoud). Schizophrenic patients and healthy individuals could be considered two good, independent and very different systems to evaluate the association of HS1,2A allele frequencies with the alteration of serum Ig concentrations.

9. Genetic factors influencing the progression of viral infections

9.1 Genetic factors associated with AIDS progression

In HIV infected subjects, the progression to AIDS has been associated with a number of host genetic factors (Carrington *et al*, 1999; Bugeja *et al*, 2004) and virological variants (Saksena *et al*, 1996; Casartelli *et al*, 2003), suggesting that the individual immune-response to HIV might be influenced at many levels of the virus-host interaction. The principal targets of HIV are the CD4+ T-lymphocytes whose decline causes the impairment of the host immune-defence (Rodriguez *et al*, 2006; Zandman-Goddard and Shoenfeld, 2002). CD4+ T cell levels are an indicator of disease progression: the CD4+ levels decreased with the progression of the diseases (Bourinbaiar *et al*, 2005). However, alteration in the humoral immune-response mediated by B-cells, including hypergammaglobulinemia (De Milito, 2004) has been observed during HIV infection, particularly in the late stage of the disease (Root-Bernstein and Rallo, 2004).

Among the genetic factors which influence the progression to AIDS, the role of HS1,2A polymorphism has been evaluated in a cohort of infected children from a nosocomial outbreak with a monophyletic strain of HIV occurred at the Benghazi Children's Hospital in Libia (unpublished data). According to the different definition from the literature, the HIV patients were divided in four subgroups: Long Term Non-Progressors (any critical events after 6 or more years from infection); Slow Progressors (moderate clinical manifestation and immunosuppression after 6 years from infection); Fast Progressors (patients with severe clinical manifestation and immunosuppression before 6 years from infection); Uncertain (subjects on treatment for which therapy do not allow a univocal classifications; Centers for Disease Control and Prevention, 1994) (Tab. 3). The results show that the homozygous condition for allele 2A was significantly increased in HIV Fast Progressors subjects in respect to Long Term Non-Progressors subjects. Thus, the homozygous condition for allele 2 associate specifically with AIDS progression. Starting from these data, one aim of the project is investigated the role of HS1,2A polymorphism in influencing the individual response to AIDS progression in a cohort of HIV infected African patients.

	n	Sex (% male)	Median age at the first observation (IQR[1])	Median Nadir %CD4+ (IQR[1])	% ARV[2]
Long Term Non Progressor	37	43.2%	5.21 (1.50-10.28)	28.8% (25.0-31.0)	0.0
Slow Progressor	9	88.9%	5.51 (4.51-7.20)	23.0% (19-25)	0.0
Fast Progressor	34	52.9%	3.89 (2.10-6.49)	12% (9-16)	68.2
Uncertain	48	51.1%	1.81 (1.32-28.86)	21% (17-33)	100.0

Tab. 3. Demographic, immunological and therapeutical characteristics of groups classified according to clinical progression (1. IQR Inter-quartile range 2. % of patients which started Antiretroviral Therapy) (Trachtenberg and Erlich, 2001).

Genotypes n(%)	Long Term Non Progressor	Fast Progressor	*P*	O.R.
1/1	7 (19)	4 (12)	0.51	1.75
1/2	17 (46)	11 (31)	0.33	1.77
1/3	2 (5)	2 (6)	1.00	0.91
1/4	4 (11)	4 (12)	1.00	0.90
2/2	2 (5)	7 (21)	0.04	0.22
2/3	1 (3)	1 (3)	1.00	0.91
2/4	3 (8)	5 (15)	0.47	0.51
3/4	1 (3)	0	1.00	n.d.

Tab. 4. HS1,2A genotypes in Long Term Non Progressors and Fast Progressors patients. The number of subjects per genotype and percent (%) have been reported. Fisher's exact test two-sides p-value and odds ratio (O.R.) are shown in the two last columns.

9.2 Genetic factors associated with Hepatitis C progression

Hepatitis C virus (HCV) infection is the main cause of chronic liver disease in the world. The natural history of chronic HCV infection can vary dramatically between individuals. Host factors (like toxins, metabolic factors, liver specific factors, genetic factors, demographics, immunosuppression) and viral factors (like viral protein) affect fibrosis progression in chronic HCV infection (Feld and Liang, 2006). Recent studies show that immunological factor as cytokines and host genetic variations, rather than direct HCV action, seem to play an important role in the pathogenesis of HCV infection.

Spontaneous viral clearance occurs in about 25% of acutely infected individuals. At the genetic level, in human infection there is the linkage between specific human leukocyte antigen (HLA) class I and class II molecules with clearance of virus, indicating that the presentation of specific viral peptides through molecules leads to protective responses.

Around 75% of individuals initially infected with HCV will become chronically infected, usually for decades. Individuals with chronic infection develop progressive hepatic fibrosis leading to an increased risk of cirrhosis, liver failure and hepatocellular carcinoma (Lloyd *et al*, 2007). During chronic infection a strong polyclonal activation of B cells leading to auto-reactive antibodies production, cryoglobulinemia and lymphomas occurs. B cell activation depends on HCV replication and genetic or environmental factors are involvement in the process. Several polymorphism of a variety of candidate gene have been shown to be associated with progressive fibrosis: MHC class II haplotypes; cytokine polymorphism such as CCR5 (chemokine receptor 5), MCP-1 and -2 (Monocyte chemotactic protein-1 and -2), IL-10 (Interleukin-10) and others (Lloyd *et al*, 2007; Saito *et al*, 2004; Feld and Liang, 2006).

The correlation of HS1,2A polymorphism with the progression of Hepatitis C was investigated in this thesis. HCV patients are divided in different groups according the individual response to the HCV infection. HS1,2A allelic frequencies were evaluated for each group.

AIMS OF THE PROJECT

The principal aims of the project are:

1. To study the distribution of HS1,2A alleles in Italian population and in several European healthy populations living in geographic isolation.

2. To study the association of HS1,2A polymorphism with several autoimmune diseases and selective IgA deficiency.

3. To study the correlation of HS1,2A allele frequencies with the alteration of serum Ig concentration.

4. To study the role of HS1,2A polymorphism in the progression of HIV and HCV infections.

5. To study the genetic variants of the regions surrounding the HS1,2A and HS3 enhancers.

MATERIALS AND METHODS

1. Materials

1.1 Solutions

Lysis Buffer (1X)

- 100 mM Tris-HCl pH 8.0
- 100 mM NaCl
- 10 mM EDTA pH 8.0
- 2% SDS

TAE (50X)

- 242 g Tris base
- 57.1 ml Acetic Acid
- 100 ml 0.5 M EDTA (pH 8)

1.2 Enzymes

Platinum Taq DNA Polymerase High Fidelity; Platinum Taq DNA Polymerase (Invitrogen, Carlsbad, CA).

1.3 Primers

1.3.1 PCR primers

Selective PCR

SA2.5 5'- GGA TCC CTG TTC CTG ATC ACT G -3'
A2R 5'- GCC CTT CCT GCC AAC CTG -3'

Nested PCR_HS1,2A

P3 5’- GAC TCA TTC TGG GCA GAC TTG -3’
D3 5’- GTC CTG GTC CCA AAG ATG G -3’

Nested PCR_ from 87168787 nt to 87166773 nt

6648 F 5’- CCT GTC TGA TTT CTG AG -3’
α3R 5’- TCC TCC TCC ATC CCC ATC TG -3’

Nested PCR_ from 87165310 nt to 87163622 nt

γ2F 5’- CAC ATG CCC TCC GCT C -3’
D3 5’- GTC CTG GTC CCA AAG ATG G -3’

1.3.2 Sequencing primers

Primer_ from 87168787 nt to 87166773 nt

6648 F 5’- CCT GTC TGA TTT CTG AG -3’
266F 5’- AGG AGG ACT TGG TGG CGA GA -3’
657F 5’- GGC TTT CCA GGT CCA GGG TT-3’
1076F 5’- AGC TGC AAG CCC CTC TCA CA -3’
8199F 5’- GGA CAG GGA TGG ACG CTG G -3’
559R 5’- GGT CTC GGT TTT GGG GCA T -3’
7545R 5’- CTC TGC ATG TCT GTC TCC -3’
1222R 5’- TGC TCC TGG GCC ATG TGT GT -3’
1617R 5’- ACC CTG CTG GCT CTC CTG TT -3’
α2R 5’- AGG CTG TGC GTG TTG GAA G -3’
8608R 5’- TGG AAC GTG ACC AGT GTG -3’

Primer_ from 87165310 nt to 87163622 nt

γ2F 5’- CAC ATG CCC TCC GCT C -3’
469F 5’- CTT CAT TCC TTG TGC AC -3’
884F 5’- TGA CAC ATA GTG TGC CC -3’
1325F 5’- TTG AGT GAC TCA TTC TGG -3’
485R 5’- GTG CAC AAG GAA TGA AG -3’

773R 5’- TGA GGT AGG AGG ACT GG -3’
1249R 5’- AGA GAT GCC GAA AAC TC -3’
D3R 5’- GTC CTG GTC CCA AAG ATG G -3’

1.4 DNA samples

1.4.1 Control samples

Italy healthy samples for population genetic studies are collected in collaboration with: Dott. Massoud, Tor Vergata University Hospital, Rome, Italy; Prof. Costa, University of Padua, Padua, Italy; Dott.ssa Cianci, Paediatric Clinic, University of Brescia, Brescia, Italy; Prof. Raimondi, University of Valle D’Aosta, Aosta, Italy; Prof. Fani and Prof. Bazzicalupo; Prof. Rocchi and Prof.ssa Archidiacono, University of Bari, Bari, Italy; Prof.ssa Cordeddu, Dept. of Ematology, Oncology and Molecular Medecine, ISS, Rome, Italy; Civil Hospital of Messina; Prof. Novelletto, Tor Vergata University, Rome, Italy; Prof.ssa Rickards, Tor Vergata University, Rome, Italy. Sample collection and extraction of Spanish populations were conducted by Martinez-Labarga, Tor Vergata University, Rome, Italy.

1.4.2 Patients

Clinical diagnosis, sample collection and extraction of Early and Long Standing Rheumatoid Arthritis, Systemic Sclerosis and Lupus Erythematosus Systemicus patients were conducted by Dott.ssa Tolusso and Prof. Ferraccioli, Division of Rheumatology, Catholic University, Rome, Italy. Clinical diagnosis and sample collection of selective IgA deficiency, *Dermatitis Herpetiformis*, Plaque Psoriasis and Psoriatic Arthritis were conducted by Dott.ssa Cianci and Prof. Pandolfi, Paediatric Clinic, University of Brescia, Brescia, Italy, and Institute of Internal Medicine, Catholic University, Rome, Italy; Psoriasis samples was also collected by Dott.ssa Esposito, Dept of Dermatology, Tor Vergata University Hospital, Rome, Italy; and by Dott.ssa Tolusso, Division of Rheumatology, Catholic University, Rome, Italy. Morbo di Chron and Rectocolitis Ulcerosis samples were collected by Dott.ssa Biancone, Dept of Internal Medicine, Tor Vergata University, Rome, Italy. HIV positive and negative samples collection was

conducted by Dott. Buttò and Dott. Chiappi, ISS, Rome, Italy. To clinical diagnosis and sample collection of HCV sample we thanks to Dott. Puoti, chief of the Dept. of Internal Medicine and Liver Unit, Marino General Hospital, Marino, Rome, Italy. Finally, clinical diagnosis of schizophrenic patients was conducted by the supervision of Prof. Rubino, Dept of Neuroscience, Vergata University Hospital, Rome, Italy.

2. Methods

2.1 DNA extraction

DNA was extracted from serum samples or oral steril plug (Becton Dickinson, USA) sampling of epidermal cells. 300 microliters of blood samples or the cotton plugs [2] were introduce in a sterile vial with 300 microliters of Lysis Buffer 1X and 20 μg/ml of proteinase K, recombinat (Roche, Germany), then incubated at 56 °C o.n. or 65 °C for 1-2 h. After the lysis the sample are briefly centrifuged and the cotton plugs are removed. An equal volume of phenol/chloroform/isoamilic alcohol 25:24:1 respectively was added. The vials are gently inverting for 3-4 times and then centrifuged for 3' at 13.000 rpm in an Eppendorf refrigerated minifuge. The supernatant was then re-extracted with chloroform/isoamilic alcohol 24:1, gently inverting for 3-4 times and then centrifuged for 3' at 13.000 rpm. The supernatant was processed following the suggestions of the "Microcon 100" method (Budowle *et al*, 2000).

2.2 HS1,2A genotyping

To estimate the frequencies of the 4 alleles of HS1,2A we carried out a selective PCR, which amplified the HS1,2A region, but not the identical inverted HS1,2B region (Giambra *et al*, 2005). The 5.4Kb fragment of the RR-1 (NT_026437.11 from 87168994 nt to 87163590 nt) was selectively amplified by the primer SA2.5 and A2R. The selective PCR was performed on genomic DNA (≈ 20 ng) extracted as described above. PCRs were carried out in 50 microliters of reaction volume containing: 2 microliters of extracted DNA (50 ng), 1,5 U *Platinum Taq DNA Polymerase High Fidelity* (Invitrogen, Carlsbad, CA), 15 pmoles of each primer, 1.5 mM $MgSO_4$, 50 microM each dNTP and 1X buffer *High Fidelity* (600 mM Tris-SO_4 pH 8.9, 180 mM [$(NH_4)_2SO_4$] (Invitrogen, Carlsbad, CA), by using GeneAmp PCR System 9700 (Applied Biosystems, Foster City, CA).The cycle conditions were: first denaturation step 94°C 2', then 94°C 30'', 59°C 30'', 68°C 5' for 10 cycles and 94°C 30'', 57°C 30'', 68°C 5' for 20 cycles, finally 72°C 10'.

[2] *The cotton plugs were removed from the support with a sterile lancet under laminar flow hood*

A nested PCR was performed using the two primers P3 Frw and D3 Rev in order to obtain the polymorphic region of the enhancer HS1,2A (NT_026437.11 from 87163981 nt to 87163622 nt). It was performed with 1/25 of the volume of the first PCR. This second PCR was carried out with the same volumes and concentrations used in the first PCR reaction, except for the use of 1U of *Platinum Taq DNA polymerase* and 1X buffer *Platinum Taq* (Invitrogen, Carlsbad, CA). Cycle conditions were: first denaturation step, 94°C 2' ; then 94°C 30", 56°C 30", 72°C 30", for 30 cycles; after the last cycle 72°C 5'. PCR products were analyzed on a 2 % agarose gel stained with "Syber Safe DNA gel stain" (Invitrogen, Carlsbad, CA).

2.3 Sequencing

The 5.4Kb fragment of the RR-1 was amplified by the selective PCR (see below). Subsequently, two nested PCRs are performed to amplified two different regions. The first region (NT_026437.11 from 87168787 nt to 87166773 nt) was amplified using the primers 6648 F and α3R; the second region (NT_026437.11 from 87165310 nt to 87163622 nt); was amplified using the primers γ2F and D3. Both the PCRs were carried out with 1/25 of the volume of the selective PCR and they were performed with the same volumes and concentrations used in the selective PCR reaction except for the use of 1U of *Platinum Taq DNA polymerase High Fidelity* (Invitrogen, Carlsbad, CA). Cycle conditions for the first reaction were: first denaturation step, 94°C 2' ; then 94°C 30", 56°C 30", 68°C 2', for 30 cycles; after the last cycle 72°C 10'. Cycle conditions for the second were the same of the first PCR except for the Tm= 60 °C. 1/10 of PCR products were analyzed on a 1 % agarose gel stained with "Syber Safe DNA gel stain" (Invitrogen, Carlsbad, CA). The 9/10 of PCR product were purified following the suggestions of the "Montage PCR Centrifugal Filter Devices" method (Millipore, Bedford MA, USA), or alternatively, of "Sure Clean Plus" method (Bioline, Randolph MA, USA). The labeling reactions were carried out on a MicroAmp- Optical 96- Well Reaction Plate (Applied Biosystem, Forster City, CA, USA) and in 10 microliters of reaction volume. The labeling reaction conditions were: 1 microliter of purified PCR product; 5 pmol of the sequencing primer; 1 microliter of Big Dye Terminator Cycle Sequencing Kit v 1.1 or v 3.1 (Applied Biosystem, Forster City, CA, USA); 1.5 microliters of Big Dye Terminator v1.2, v 3.1 5X Sequencing Buffer

(Applied Biosystem, Forster City, CA, USA). Cycle condition for the labeling reactions are: 96°C 10'', 50°C 5'', 60°C 4' for 30 cycles.
The labeling products were purified following the suggestions of the "Agencourt CleanSEQ" Protocol (Beckman Coulter, Beverly, MA, USA). Finally, 20 microliters of the Labeling DNA purified with 10 microliters of formamide was used for the Sequencing reaction (in collaboration with Dott. G. Vaccari, ISS, Rome).

2.4 Statistical analysis

The HS1,2A allele frequencies were analyzed for the Hardy-Weinberg disequilibrium. *P*-value was set at 0.05. P-value was calculated by Chi square calculator at web site http://www.fourmilab.ch/rpkp/experiments/analysis/chiCalc.html.
Comparisons between frequencies of categorical variables in different groups were performed by Fisher's exact test (recommended) or chi-square test (less exact, but more widely known) depending on sample size restrictions. Yate's continuity correction was used. Two-sided *P*-value was set at 0.05. Odds ratio with 95% Confidence Interval (95% CIs) using the approximation of Woolf were calculated. Graphpad Instat (San Diego, California, USA), Fstat version 2.9.3.2 (Institute of Ecology, UNIL, Lausanne, Switzerland) and Smith's Statistical Package, version 2.80 (Pomona College, Claremont California USA) were used for statistical analysis. Linkage Disequilibrium analysis was carried at web site http://pharmgat.org/Tools/pbtoldplotform. EM algorithm was used to estimate haplotype frequencies by Arlequin ver 3.0 (Excoffier *et al*, 2005). Amova analysis were performed using Arlequin ver 3.0 (Excoffier *et al*, 2005). CpG island analysis was carried by EMBL-EBI database at web site http://www.ebi.ac.uk/emboss/cpgplot/. Transcription factor binding sites analysis was performed by TFSEARCH program at web site http://www.cbrc.jp/research/db/TFSEARCH.html.

RESULTS

1. HS1,2A polymorphism in healthy populations

1.1 HS1,2A allelic frequencies in Italian population

The HS1,2A allelic distribution throughout Italy was investigated. A total of 1122 healthy donors coming from 12 different Italian regions were genotyped (Tab 1A). HS1,2A allelic frequencies, the relative standard error and Hardy Weinberg p-value were reported for each group (Tab. 1B). All the populations proved to be in Hardy Weinberg equilibrium ($p>0.05$) except for Lazio and Molise. In most regions allele 1 and allele 2 show the highest frequencies, allele 4 shows a lower frequency in respect to allele 1 and 2, and allele 3 shows the lowest frequency, as expected from previous studies on Italian population (Mattioli *et al*, 2008; Giambra *et al*, 2005). Concerning the other regions, Sicily shows an increase of allele 1 frequency with a value of 0.5; Sardinia, Umbria and Friuli have an increase of allele 2 which is present at a frequency of 0.595, 0.572 and 0.580 respectively; Lazio and Veneto show an increase of allele 4 with a frequency of 0.165 and 0.176 respectively. To evaluate the genetic homogeneity of the Italian population for HS1,2A polymorphism the "Analysis of Molecular Variance" (AMOVA) was performed by Arlequin ver 3.0 (Excoffier *et al*, 2005). The 12 Italian populations analyzed were divided into three major groups: north, centre and south. The total variability has three genetic components: among groups, among populations within groups and, thirdly, within populations. Results show that the major component of genetic variability for HS1,2A locus is the variance within populations (99.15%), as expected from population genetic studies (Barbujani *et al*, 1996; Barbujani and Goldstein, 2004). The other two components have a marginal effect, with the variance among populations of the same group showing a higher value (1.19%) in respect to the variance among groups (north, centre and south) (-0.34%). The negative variance observed means that, on average, members of different groups do not differ for that marker mare than members of the same group (Barbujani and Goldstein, 2004). From these data we can conclude that the Italian population is rather homogeneous for HS1,2A polymorphism (Mattioli *et al*, 2008).

A

HS1,2A genotypes obs (exp)	Regions					
	Veneto	Lombardy	Valle d'Aosta	Tuscany	Apulia	Sardinia
1/1	8 (9.3)	6 (8.1)	14 (18.3)	8 (7.5)	19 (18.1)	11 (8.3)
2/2	15 (17.0)	18 (20.3)	26 (28.4)	11 (13.2)	28 (30.5)	20 (32.1)
3/3	0 (0.7)	0 (0.01)	1 (0.5)	0 (0.1)	0 (0.01)	0 (0.01)
4/4	1 (2.9)	1 (0.1)	1(0.2)	1(0.8)	1 (0.6)	1 (4.5)
1/2	28 (25.1)	31 (25.7)	53 (45.6)	21(19.9)	49 (47)	38 (32.5)
1/3	9 (5.0)	1 (0.7)	9 (6)	3 (2.1)	1 (0.8)	0 (0.5)
1/4	6 (10.4)	4 (5.4)	2 (4)	2 (4.5)	3 (6.8)	2 (12.3)
2/4	20 (14.1)	9 (8.6)	6 (5)	10 (6.6)	12 (8.8)	42 (24.2)
2/3	2 (6.8)	0 (1.1)	4 (7.5)	3 (2.9)	1 (1.1)	2 (1.1)
3/4	5 (2.8)	1 (2.9)	0 (0.7)	0 (0.7)	0 (0.1)	0 (0.4)
subjects	94	71	116	59	114	116
HS1,2A genotypes obs (exp)	Regions					
	Sicily	Friuli	Calabria	Umbria	Lazio	Molise
1/1	28 (25)	8 (11.6)	13 (12.7)	3 (5.5)	32 (22.4)	16 (11.8)
2/2	14 (13.3)	28 (33.6)	28 (35.1)	15 (20.9)	19 (21.2)	17 (12.8)
3/3	1 (0.5)	0 (0.1)	4 (5.3)	1 (0.02)	0 (0.1)	0 (0.02)
4/4	1 (0.4)	0 (0.2)	2 (0.6)	0 (0.7)	0 (3.6)	0 (0.25)
1/2	33 (36.5)	48 (39.4)	45 (39.1)	30 (21.5)	31 (43.5)	9 (19.9)
1/3	6 (7)	1 (2.4)	0 (4.1)	0 (0.6)	2 (2.8)	0 (0.8)
1/4	5 (6.5)	3 (3.1)	3 (5.5)	1 (3.9)	12 (18.0)	0 (2.8)
2/4	6 (4.7)	6 (5.2)	9 (8.4)	12 (7.5)	32 (17.5)	5 (3.6)
2/3	6 (5.1)	6 (4.1)	4 (6.3)	0 (1.2)	5 (2.8)	2 (1)
3/4	0 (0.9)	0 (0.3)	0 (0.9)	0 (0.2)	0 (1.1)	0 (0.1)
subjects	100	100	108	62	133	49

B

Region	N	Alleles				p
		1	2	3	4	
Veneto	94	0.314 ± 0.034	0.425 ± 0.036	0.085 ± 0.020	0.176 ± 0.028	0.056
Lombardy	71	0.338 ± 0.040	0.535 ± 0.042	0.014 ± 0.010	0.113 ± 0.027	0.313
Valle d'Aosta	116	0.397 ± 0.032	0.495 ± 0.033	0.065 ± 0.016	0.043 ± 0.013	0.133
Tuscany	59	0.356 ± 0.044	0.474 ± 0.046	0.051 ± 0.020	0.119 ± 0.030	0.207
Apulia	114	0.399 ± 0.032	0.517 ± 0.033	0.009 ± 0.006	0.075 ± 0.017	0.098
Sardinia	116	0.267 ± 0.029	0.526 ± 0.033	0.009 ± 0.006	0.198 ± 0.026	0.785
Sicily	100	0.500 ± 0.035	0.365 ± 0.034	0.070 ± 0.018	0.065 ± 0.017	0.785
Friuli	100	0.340 ± 0.033	0.580 ± 0.035	0.035 ± 0.013	0.045 ± 0.015	0.142
Calabria	108	0.343 ± 0.032	0.528 ± 0.034	0.055 ± 0.015	0.074 ± 0.018	0.153
Umbria	62	0.298 ± 0.041	0.572 ± 0.044	0.016 ± 0.011	0.114 ± 0.028	0.090
Lazio	133	0.410 ± 0.030	0.399 ± 0.030	0.026 ± 0.009	0.165 ± 0.023	0.001
Molise	49	0.400 ± 0.050	0.510 ± 0.050	0.020 ± 0.014	0.070 ± 0.030	0.001

Tab. 1. A HS1,2A genotypes of 12 Italian regions. The number of observed genotypes (obs) was reported. In bracket the number of expected genotypes (exp). **B** HS1,2A allelic frequencies ± standard error of 12 Italian regions. (N=number of total subjects; p=Hardy-Weinberg p-value).

1.2 HS1,2A allelic frequencies in populations living in geographic isolation

HS1,2A allelic frequencies were determined in Spain and Italian populations living in a condition of geographic isolation. The Sierra de Gredos population, living in a valley of Spain isolated by mountains, was compared with Madrid who represents the Spanish population. The Sardinian population, isolated by Mediterranean Sea, was compared with a representative Italian sample. Finally we compared Vera and Bejar, which are the two populations of Sierra de Gredos living in an isolated condition in respect of one to the other. HS1,2A genotypes distribution, the allelic frequencies with the standard error and Hardy-Weinberg equilibrium are reported for each group (Tab. 2). All the populations proved to be in Hardy-Weinberg equilibrium ($p>0.05$). An increase of allele 2 and a decrease of

allele 4 and 1 are reported for Sierra de Gredos population in respect to Madrid group, as well as Bejar population in respect to Vera group. The Sardinian population shows an increase of allele 2 and 4 frequencies and a decrease of allele 1 frequency in respect to Italian population. To compare HS1,2A allelic frequencies among the groups χ^2 test was performed. Significant HS1,2A allelic frequency differences between Sierra de Gredos and Madrid groups, as well as Sardinian and Italian groups were found ($p<0.001$). No significant differences between Vera and Bejar populations were found ($p=0.106$). The result suggests that inter-population genetic flux between both Sardinian and Italian populations and Gredos and Madrid populations is impaired for this locus by natural barriers. On the other hand, inter-population genetic flux is not impaired between Vera and Bejar populations, perhaps because the two populations belong to the same geographical area.

A

HS1,2A genotypes obs (exp)	Spanish populations		Italian populations	
	Sierra de Gredos	Madrid	Sardinia	Italy
1/1	17 (16.8)	23 (21.7)	11 (8.3)	69 (48)
2/2	25 (26.5)	12 (12.4)	20 (32.1)	54 (46.8)
3/3	1 (0.04)	1 (0.3)	0 (0.01)	0 (0.6)
4/4	0 (0.3)	3 (2.01)	1 (4.5)	3 (4.8)
1/2	44 (42.2)	33 (32.9)	38 (32.5)	67 (94.8)
1/3	1 (1.6)	6 (5.5)	0 (0.5)	13 (10.7)
1/4	3 (4.5)	10 (13.2)	2 (12.3)	14 (30.4)
2/4	8 (5.7)	12 (10)	42 (24.2)	47 (30)
2/3	1 (2.1)	3 (4.2)	2 (1.1)	7 (10.5)
3/4	0 (0.2)	1 (1.7)	0 (0.4)	6 (3.4)
subjects	100	104	116	280
Alleles Fr ± SE	Spanish populations		Italian populations	
	Sierra de Gredos	Madrid	Sardinia	Italy
1	0.410 ± 0.035	0.457 ± 0.034	0.267 ± 0.029	0.414 ± 0.021
2	0.515 ± 0.035	0.346 ± 0.033	0.526 ± 0.033	0.409 ± 0.021
3	0.020 ± 0.009	0.058 ± 0.016	0.009 ± 0.006	0.046 ± 0.009
4	0.055 ± 0.016	0.139 ± 0.024	0.198 ± 0.026	0.131 ± 0.014
genomes	200	208	232	560
χ^2	0.7	2.4	1.7	8.7
Deg. fr.	1	5	4	5
p	0.403	0.791	0.785	0.121

B

HS1,2A genotypes obs (exp)	Sierra de Gredos populations	
	Bejar	Vera
1/1	8 (7.6)	9 (9.2)
2/2	16 (16.2)	9 (10.6)
3/3	0	1 (0.1)
4/4	0 (0.1)	0 (0.2)
1/2	22 (22.2)	22 (19.8)
1/3	0	1 (1.7)
1/4	1 (1.5)	2 (3.0)
2/4	3 (2.3)	5 (3.2)
2/3	0	1 (1.8)
3/4	0	0 (0.3)
subjects	50	50
Alleles Fr ± SE	Sierra de Gredos populations	
	Bejar	Vera
1	0.390 ± 0.048	0.430 ± 0.049
2	0.570 ± 0.049	0.460 ± 0.049
3	0	0.040 ± 0.019
4	0.040 ± 0.019	0.070 ± 0.025
genomes	100	100
χ^2	0.4	0.8
Deg. fr.	1	1
p	0.527	0.371

Tab. 2. HS1,2A genotypes distribution and allelic frequencies ± standard error (Fr ± SE) in Spanish, Italian (**A**) and Sierra de Gredos (**B**) populations. The number of observed genotypes (obs) was reported. In bracket the number of expected genotypes (exp). Hardy-Weinberg equilibrium was reported for each group (Deg.fr.=degrees of freedom; χ2=chi square; p=p-value).

2. Association of HS1,2A polymorphism with immunological diseases

2.1 Increase of HS1,2A allele 1 enhancer frequency in selective IgA deficiency

The correlation between HS1,2A polymorphism and selective IgA deficiency (IGAD) was investigated in this work. A cohort of 49 IGAD patients and 71 healthy subjects of the same geographical area were genotyped (Tab. 3). The results show a significant difference for the subjects homozygous for the allele 1 that has a frequency of 37% in the IgAD patients and 7% in the control, with a 5.28 folds increase. The heterozygous subjects with alleles 1/2 have a marked shift from 23% in IgAD to 45% in the control subjects. The frequency of the four alleles in IGAD patients and in the control group with standard error and the relative Hardy Weinberg equilibrium are reported in table 4. Unfortunately, both groups do not result in Hardy Weinberg equilibrium. Allele 1 frequency increases in IGAD patients compared to the control group (from 0.331 to 0.520), whereas allele 2 frequency decreases from 0.528 in the control group to 0.367 in IGAD group. A graphic representation of HS1,2A allelic frequencies of IGAD patients and of the control group was reported in figure 1. The histogram shows the inversion of the frequencies among allele 1 and 2 in the two groups. HS1,2A allelic frequency differences between IGAD patients and the control group proved to be significant (χ^2=10.2; Deg.fr.=3; p=0.018). To compare the variation of allele 1 and allele 2 frequencies among the two groups chi-square test[3] was performed. To analyze retrospective case-control studies odds ratio (O.R.) was calculated. We found significant differences of allele 1 frequency compared with non 1 allele frequency between the control and IGAD groups, with a χ^2 value of 7.8 and a relative p-value of 0.005 (O.R.=0.456) . Allele 2 *versus* non 2 comparison shows a χ^2 value of 5.4 with a significant p-value of 0.02 (O.R.=1.928). Finally, the statistical analysis of the variation of allele 1 *versus* allele 2 among the groups shows a χ^2 value of 7.4 with a p-value of 0.006 (O.R.=2.261). The results demonstrate that allele 1 increase and allele 2 decrease of frequencies confer the genetic predisposition to IgA deficiency disease (Mattioli *et al*, 2009).

[3] *with Yate's correction, two-sided p-value*

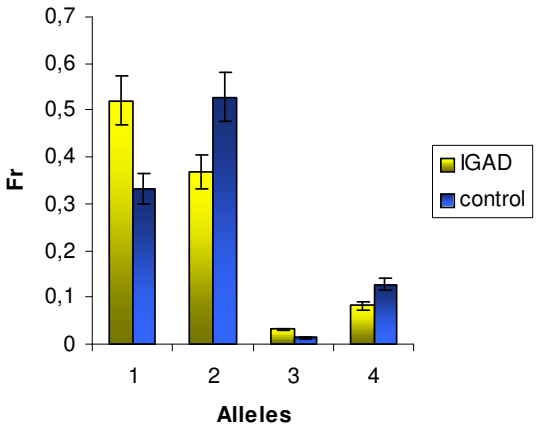
HS1,2A genotypes	IGAD			control		
	obs.	exp.	%	obs.	exp.	%
1/1	18	13.2	37	5	7.8	7
2/2	10	6.5	20	15	19.8	21
1/2	11	18.7	23	32	24.8	45
1/3	2	1.6	4	1	0.6	1.5
1/4	2	4.2	4	4	5.9	6
2/3	0	1.2	-	0	1.0	-
2/4	5	2.9	10	13	9.5	18
3/4	1	0.2	2	1	0.2	1.5
subjects	49			71		

Tab. 3. HS1,2A genotypes of IgA deficiency patients (IGAD) and control group. The number of genotypes both observed (obs) and expected (exp) and percent (%) was reported.

HS1,2A Alleles	IGAD		control	
	Fr ± SE	n	Fr ± SE	n
1	0.520 ± 0.050	51	0.331 ± 0.039	47
2	0.367 ± 0.049	36	0.528 ± 0.042	75
3	0.031 ± 0.017	3	0.014 ± 0.009	2
4	0.082 ± 0.028	8	0.127 ±0.028	18
genomes		98		142
Deg. fr.	2		1	
χ^2	9.6		5.4	
p	0.008		0.020	

Tab. 4. HS1,2A allelic frequencies ± standard error (Fr ± SE) in IgA deficiency (IGAD) patients and control group. The number of observed genomes (n) and Hardy-Weinberg equilibrium were reported for each group (Deg.fr.=degrees of freedom; χ2=chi square; p=p-value).

Fig. 1. Graphic representation of the comparison of HS1,2A allele frequencies in IgA deficiency cohort (IGAD) with the control (χ2=10.2; Deg.fr.=3; *p*=0.018).



2.2 Increased frequency of HS1,2A allele 2 in several autoimmune diseases

2.2.1. Increased frequency of HS1,2A enhancer allele 2 in *Dermatitis Herpetiformis*, Plaque Psoriasis and Psoriatic Arthritis

HS1,2A polymorphism association with multi-factorial cutaneous diseases like *Dermatitis Herpetiformis* (DH), plaque psoriasis (PP) and psoriatic arthritis (PA) was investigated. We analyzed a group of 37 DH patients, 61 PP patients, and 28 patients affected by PA. The control group consisted of a large cohort of 265 healthy individuals from the same geographical area. The genotypes distribution, the allele frequencies ± standard error and the Hardy Weinberg equilibrium are reported in table 5. The observed genotypes matched Hardy-Weinberg expected values for all groups except for *Dermatitis Herpetiformis*, perhaps due to the small number of the sample. The distribution of the four allele frequencies is similar among the three pathological groups but different in respect to the control sample. In the three diseases, the frequencies of 1, 3 and 4 alleles are decreased in respect to the control group. On the other hand, we observe a dramatic increase of allele 2 frequency which shifts from 0.392 in the control group to 0.635, 0.590, and 0.750 in DH, PP and PA respectively. The comparisons of the allelic frequencies between pathologic and healthy groups measured by χ^2 significance test showed very low p-values for DH, PP and PA ($p<8x10^{-4}$, $p<2x10^{-4}$ and $p<10^{-5}$ respectively) emphasizing the strong divergence from the healthy group. The statistical comparison of 2 *versus* non 2 allele frequencies by chi square test shows a significant difference between the diseases and the control group with a p-value of 0.0004, 0.0001 and <0.0001 for DH, PP and PA diseases respectively. We also measured the odds ratios with the corresponding confidence intervals (Tab. 6). These results demonstrate a correlation among the *Dermatitis Herpetiformis*, Plaque Psoriasis and Psoriatic Arthritis diseases and the increase of allele 2 frequency (Fig. 2). A similar trend characterizes the three diseases and suggests that the presence of allele 2 correlates with a clinical risk in developing the immune-disorders (Cianci *et al*, 2008).

HS1,2A genotypes obs (exp)	Patients			Ctr
	DH	PP	PA	
1/1	6 (3.3)	9 (7.2)	1 (1.3)	57 (53.4)
2/2	16 (14.9)	21 (21.2)	15 (15.7)	38 (40.7)
3/3	0 (0.01)	0 (0)	0 (0.01)	2 (0.9)
4/4	0 (0.1)	0 (0.2)	0 (0.01)	4 (2.6)
1/2	10 (13.9)	23 (24.8)	10 (9)	93 (93.3)
1/3	0 (0.3)	0 (0.3)	0 (0.2)	12 (14.0)
1/4	0 (1.2)	1 (2.4)	0 (0.2)	19 (23.8)
2/4	4 (2.5)	6 (4.2)	1 (0.7)	25 (20.8)
2/3	1 (0.7)	1 (0.6)	1 (0.7)	14 (12.2)
3/4	0 (0.1)	0 (0.1)	0 (0.02)	1 (3.1)
subjects	37	61	28	265
Alleles Fr ± SE (n)	Patients			Ctr
	DH	PP	PA	
1	0.297 ± 0.053 (22)	0.344 ± 0.043 (42)	0.214 ± 0.054 (12)	0.449 ± 0.022 (238)
2	0.635 ± 0.070 (47)	0.590 ± 0.044 (72)	0.750 ± 0.057 (42)	0.392 ± 0.021 (208)
3	0.014 ± 0.014 (1)	0.008 ± 0.008 (1)	0.018 ± 0.017 (1)	0.059 ± 0.010 (31)
4	0.054 ± 0.026 (4)	0.058 ± 0.021 (7)	0.018 ± 0.017 (1)	0.100 ± 0.013 (53)
genomes	74	122	56	530
χ^2	5.0	2.1	0.3	3.1
Deg. fr.	1	1	1	5
p	0.025	0.147	0.568	0.676

Tab. 5. HS1,2A genotypes distribution, the allelic frequencies ± standard error (Fr ± SE) in patients with *Dermatitis Herpetiformis* (DH), Plaque Psoriasis (PP), Psoriatic Arthritis (PA) and healthy control (Ctr). The number of observed genotypes (obs) was reported. In bracket the number of expected genotypes (exp) and of observed genomes (n). The Hardy-Weinberg equilibrium were reported for each group (Deg.fr.=degrees of freedom; χ2=chi square; p=p-value).

DH *vs* Ctr				
Deg. fr. = 1	$\chi^2 = 8.291$	p = 0.004	O.R. = 2.17	c.i. = 1.29-3.53
PP *vs* Ctr				
Deg. fr. = 1	$\chi^2 = 15.025$	p = 0.0001	O.R. = 2.23	c.i. = 1.50-3.33
PA *vs* Ctr				
Deg. fr. = 1	$\chi^2 = 17.198$	p < 0.0001	O.R. = 3.68	c.i. = 1.96-6.90

Tab. 6. Statistical analysis of the variation of 2 versus non 2 allele frequencies in DH, PP, PA patients in respect to control subjects (Ctr) (Deg.fr.=degree of freedom, χ^2 =chi-square test with Yate's correction, p=two-sided p-value; O.R.=odds ratio; c.i.=95% confidence interval using approximation of Woolf).

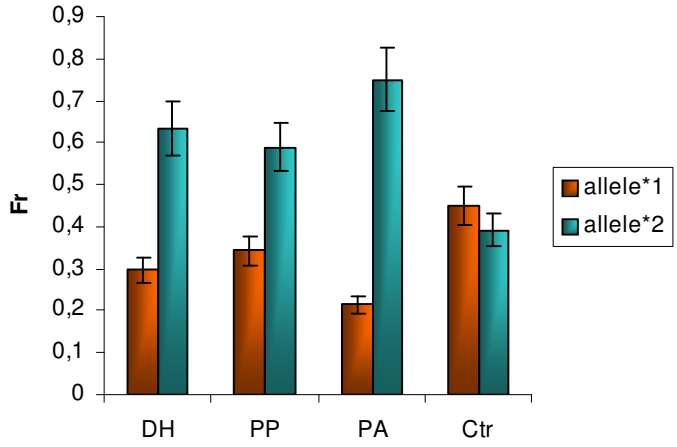


Fig. 2. Graphic representation of 1 and 2 HS1,2A allele frequencies in the diseases (DH; PP and PA) and in the control (Ctr).

2.2.2 Allele 2 of the HS1,2A enhancer associates with Rheumatoid Arthritis

We tested the hypothesis that the HS1,2A locus polymorphism could be involved in Rheumatoid Arthritis (RA) pathogenesis, influencing small or long disease duration. The study cohort consisted of 100 patients with early rheumatoid arthritis (ERA) and of 114 patients with long standing rheumatoid arthritis (LSRA). ERA patients had a mean age of 54.9 ± 13.8 years, 75% were female; the mean disease duration was 0.6 ± 0.3 years. The LSRA patients (n=114) had a mean age of 57.1 ± 13.6 years, 85% were female, the mean disease duration was 13.7 ± 10.0 years (data from Ferraccioli). A first study used the RA sample of 100 Italian individuals with early arthritis and 248 healthy controls from the same ethnic background. The genotype distribution showed significant differences among the two groups (Tab. 7). The percentage of 2/2 genotype was higher in ERA patients (27.0%) than in control group (14.9%; p=0.008, OR=2.11, 95%CIs=1.20-3.70). These results were supported by genotyping a second, completely independent cohort of 114 LSRA patients. In this cohort of patients 34.2% of the individuals carried the 2/2 genotype (p<0.001, OR=2.97, 95%CIs=1.76-5.00) and 62.7% were heterozygous for the allele 2 (p=0.005, OR=2.58, 95%CIs=1.86-3.56). The allele frequencies ± standard error and the Hardy Weinberg equilibrium are reported in table 8. The control group resulted in Hardy Weinberg equilibrium, whereas ERA and LSRA groups showed p-values just under the statistic threshold of 0.05. The allelic frequency distribution of ERA and LSRA groups showed significant differences compared to the control. Allele 2 frequency shift from 0.395 in the control to 0.530 in ERA patients (p=0.001) and to 0.627 in LSRA patients (p<0.0001). Allele 3 frequency decreases in ERA patients (0.002) compared to the control (0.060; p=0.031), while alleles 1 and 4 have the same frequency both in ERA patients and controls. LSRA patients show a significant decrease of allele 1, 3 and 4 frequencies with a value of respectively 0.347; 0.013 and 0.013 compared to the control (allele 1: 0.440; p=0.018; allele 3: 0.395; p=0.003 and allele 4: 0.105; p<0.0001). The allelic frequency differences between the group of patients and the control group were significant (ERA *vs* Ctr: χ^2=13.2; Deg.fr.=3; p=0.004; LSRA *vs* Ctr: χ^2=46.2; Deg.fr.=3; p<0.001). Chi-square test comparing 2 *versus* non 2 allele frequencies between ERA and control; LSRA and control; ERA and LSRA are reported in table 9. The statistical comparison shows a significant difference between ERA and control groups (p=0.001, OR=0.58, 95%CIs=0.41-0.81) and

between LSRA and control groups (p<0.0001, OR=0.39, 95%CIs=0.28-0.54). Allele 2 frequency increase of LSRA *versus* ERA patient groups resulted close to the threshold of significance (p=0.053, OR=1.49, 95%CIs=1.01-2.19). Our results demonstrate a correlation between allele 2 frequency increase and the risk to develop early and long standing rheumatoid arthritis diseases. Moreover, allele 2 frequency increase is stronger in long standing RA patients compared to early RA patients, suggesting a role of HS1,2A polymorphism in influencing disease duration or other specific features of the two diseases (Tolusso *et al*, 2009).

Genotype obs (%)	**Ctr** N=248	**ERA** N=100	**OR (95%C.I.)**	***p***	**LSRA** N=114	**OR (95%C.I.)**	***p***
1/1	52 (21.0)	11 (11.0)	0.47 (0.23-0.94)	0.03	10 (8.8)	0.36 (0.18-0.74)	0.004
2/2	37 (14.9)	27 (27.0)	2.11 (1.20-3.70)	0.01	39 (34.2)	2.96 (1.76-5.00)	<0.001
3/3	2 (0.8)	-	0.49 (0.02-10.32)	1.00	-	0.43 (0.02-9.05)	1.00
4/4	4 (1.6)	5 (5.0)	3.21 (0.84-12-22)	0.13	-	0.24 (0.01-4-45)	0.31
1/2	84 (33.9)	46 (46.0)	1.66 (1.04-2.67)	0.03	59 (51.8)	2.09 (1.33-3.29)	0.001
1/3	12 (4.8)	3 (3.0)	0.61 (0.17-2.20)	0.57	-	0.08 (0.00-1.49)	0.02
1/4	18 (7.3)	2 (2.0)	0.26 (0.06-1.15)	0.07	-	0.05 (0.00-0.91)	0.001
2/3	13 (5.2)	1 (1.0)	0.18 (0.02 -1.42)	0.08	3 (2.6)	0.9 (0.14-1.75)	0.40
2/4	25 (10.1)	5 (5.0)	0.47 (0.17-1.26)	0.14	3 (2.6)	0.24 (0.07-0.82)	0.01
3/4	1 (0.8)	-	0.83 (0.03-20.34)	1.00	-	0.72 (0.03-17.84)	1.00

Tab. 7. Distribution of genotypes of the HS1,2 A polymorphism in ERA patients, LSRA patients and control (Ctr). The number of observed genotypes (obs) and percent (%; in bracket) was reported. P-values are calculated from Fisher's exact test performed in cases (ERA and LSRA) versus controls. (*p*=two sides p-value; O.R.=odds ratio; 95%C.I.=95% confidence interval using approximation of Woolf).

Alleles **Fr ± SE** **(n)**	**Ctr** N=248	**ERA** N=100	***p***	**LSRA** N=114	***p***
Allele 1	0.440 ± 0.022 (218)	0.365 ± 0.034 (73)	0.075	0.347 ± 0.031 (79)	0.018
Allele 2	0.395 ± 0.022 (196)	0.530 ± 0.035 (106)	0.001	0.627 ± 0.032 (143)	<0.0001
Allele 3	0.060 ± 0.011 (30)	0.020 ± 0.010 (4)	0.031	0.013 ± 0.007 (3)	0.003
Allele 4	0.105 ± 0.014 (52)	0.085 ± 0.020 (17)	0.485	0.013 ± 0.007 (3)	<0.0001
genomes	496	200		228	
Deg. fr.	5	3		1	
χ^2	3.5	10.5		4.2	
p	0.618	0.01		0.04	

Tab. 8. HS1,2A allelic frequencies ± standard error (Fr ± SE) in healthy control (Ctr), ERA and LSRA patients. Hardy-Weinberg equilibrium was reported for each group. (Deg.fr.=degrees of freedom; χ2=chi square; p=p-value). In brackets the number of observed genomes (n). Fisher's exact test was performed in cases (ERA and LSRA) versus controls (*p*=two sides p-value).

ERA *vs* Ctrl				
Deg. fr. = 1	χ^2 = 10.008	p = 0.001	OR = 0.58	c.i. = 0.41-0.81
LSRA *vs* Ctr				
Deg. fr. = 1	χ^2 =32.849	p < 0.0001	OR = 0.39	c.i. = 0.28-0.54
ERA *vs* LSRA				
Deg. fr. = 1	χ^2 =3.747	p = 0.053	OR = 1.49	c.i. = 1.01-2.19

Tab. 9. Statistical analysis of the variation of 2 versus non 2 allele frequencies in ERA and LSRA patients compared to the control subjects (Ctr) (Deg.fr.=degrees of freedom, χ^2=chi-square test with Yate's correction, p=two-sided p-value; O.R.=odds ratio; c.i=95% confidence interval using approximation of Woolf).

2.2.3 Increased frequency of HS1,2A enhancer allele 2 in Rheumatoid Arthritis, Systemic Sclerosis and Lupus Erythematosus Systemicus

The involvement of HS1,2A polymorphism in the predisposition to systemic autoimmune disorders, such as Rheumatoid Arthritis (RA), Systemic Sclerosis (SS) and Lupus Erythematosus Systemicus (LES) was investigated. For HS1,2A allele analysis a group of 233 RA, of 135 SS and of 135 LES patients were genotyped. The control group consisted of 248 healthy individuals of the same geographical area. The HS1,2A genotypes, the allelic frequencies ± the standard error and the Hardy Weinberg p-value are collected in table 10. There are some significant differences in genotype and allelic frequency distribution between the patient and the control groups. The genotype distribution is similar in the three pathological groups but different compared to the control sample (Tab. 10A). A significant increase of genotype 2/2 was observed in Rheumatoid Arthritis (30%); Systemic Sclerosis (33.3%) and Lupus Erythematosus Systemicus (31.6%) compared to the control group (14.9%; $p<0.001$). Moreover, genotype 1/2 is significant increased in RA cohort (48.5%) compared to the control group (33.9%; $p=0.001$); whereas genotype 1/1 is significant decreased (9.9%) in RA patients compared to the control group (21.0%; $p<0.001$). Genotype 1/2 is significant increased (48.5%) in SS group compared to the control group (33.9%, $p=0.003$). As shown in table 10B, we found a significant increase of allele 2A frequency in the three pathology groups (57,7%; 57,4% and 54,9% in RA, SS and LES respectively) in respect to the control group (39,5%; $p<0.0001$). Allele 1, 3 and 4 frequencies are significantly decreased in autoimmune pathologies compared to the control group except for allele 4 of SS group (7,9%; $p=0.25$). There are significant allelic frequency differences among RA, SS and LES patients and the control group (RA *vs* Ctr: $\chi^2=40.93$, Deg.fr.=3, $p<0.001$; SS *vs* Ctr: $\chi^2=38.84$, Deg.fr.=3, $p<0.001$; LES *vs* Ctr: $\chi^2=20.01$, Deg.fr.=3, $p<0.001$). The graphic representation comparing 1 and 2 allele frequency differences between the group of patients (RA, SS and LES) and control is shown in figure 3. This results demonstrate clearly a correlation among RA, SS and LES autoimmune systemic diseases and the increase of HS1,2A allele 2 frequency.

A							
genotype obs (%)	Ctr N=248	RA N=233	*p*	SS N=135	*p*	LES N=135	*p*
1/1	52 (21.0)	23 (9.9)	<0.001	20 (14.8)	0.14	19 (14.3)	0.11
2/2	37 (14.9)	70 (30.0)	<0.001	45 (33.3)	<0.001	42 (31.6)	<0.001
3/3	2 (0.8)	-	-	-	-	-	-
4/4	4 (1.6)	5 (2.1)	0.75	1 (0.7)	0.37	4 (3.0)	0.45
1/2	84 (33.9)	113 (48.5)	0.001	53 (39.3)	0.003	50 (37.6)	0.47
1/3	12 (4.8)	3 (1.3)	0.03	2 (1.5)	0.15	2 (1.5)	0.15
1/4	18 (7.3)	3 (1.3)	0.01	2 (1.5)	0.01	4 (3.0)	0.11
2/3	13 (5.2)	5 (2.1)	0.09	5 (3.7)	0.62	3 (2.3)	0.19
2/4	25 (10.1)	11 (4.7)	0.03	7 (5.2)	0.19	9 (6.8)	0.28
3/4	1 (0.8)	-	-	-	-	-	-
B							
Alleles Fr ± SE (n)	Ctr N=248	RA N=233	*p*	SS N=135	*p*	LES N=135	*p*
Allele 1	0.440 ± 0.022 (218)	0.354 ± 0.022 (165)	0.01	0.359 ± 0029 (97)	0.03	0.353 ± 0.029 (95)	0.02
Allele 2	0.395 ± 0.022 (196)	0.577 ± 0.023 (269)	<0.001	0.574 ± 0.030 (155)	<0.001	0.549 ± 0.030 (148)	<0.001
Allele 3	0.060 ± 0.011 (30)	0.017 ± 0.006 (8)	0.001	0.026 ± 0.010 (7)	0.03	0.019 ± 0.008 (5)	0.010
Allele 4	0.105 ± 0.014 (52)	0.052 ± 0.010 (24)	0.002	0.041 ± 0.012 (11)	0.002	0.079 ± 0.016 (21)	0.25
genomes	496	466		270		270	
Deg. fr.	5	3		3		3	
$\chi 2$	3.5	6.4		1.5		1.2	
p	0.618	0.09		0.682		0.753	

Tab. 10. A Distribution of genotypes of the HS1,2A polymorphism in RA; SS; LES and control (Ctr) groups. The number of observed genotypes (obs) and percent (%; in bracket) was reported. **B** HS1,2A allelic frequencies ± standard error in RA; SS; LES and control (Ctr) groups. Hardy-Weinberg equilibrium was reported for each group (Deg.fr.=degrees of freedom; $\chi 2$=chi square; p=p-value). In brackets the number of observed genomes (n). (RA=Rheumatoid Arthritis, SS=Systemic Sclerosis, LES=Lupus Erythematosus Systemicus; *p*=two sides p-value calculated from Fisher's exact test performed RA, SS and LES versus controls).

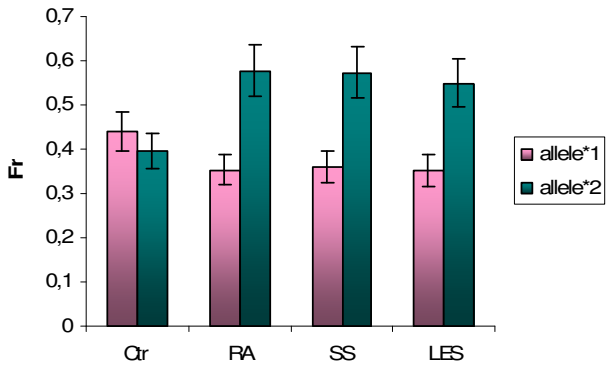


Fig. 3. Graphic representation of 1 and 2 HS1,2A allele frequencies in autoimmune diseases (RA; SS and LES) and in the control group (Ctr).

2.2.4 Allele 2 of the HS1,2A enhancer of the Ig 3' regulatory region associates with Crohn's disease

HS1,2A allelic distribution of a group of patients affected by an autoimmune disease namely Crohn's disease (CD), was compared with the allelic distribution of a patient group affected by a non-immunological disease, namely the ulcerousis-recto-colitis (RCU). The aim is demonstrate that HS1,2A allele 2 frequency specifically increases in immune-related disorders in respect to non-immunological ones. A cohort of 105 CD patients and 64 RCU patients were genotyped. The control group consisted of 265 healthy individuals of the same geographical area. The results are shown in table 11. The control group turned out to be in Hardy Weinberg equilibrium, whereas the p-values of CD and RCU groups proved to be just under the statistic threshold of 0.05. Allele 2 frequency increases and allele 1 frequency decreases in CD (0.500 and 0.362 respectively) in respect to the control group (0.392, p=0.008; 0.449, p=0.032 respectively). On the other hand, RCU patients do not show any significant allelic frequency differences in respect to the control group. The comparisons of HS1,2A allelic frequencies between pathologic and healthy groups measured by χ^2 significance test showed very low significant p-values for the difference of the four allele frequencies between CD and control (χ^2=22.4; Deg.fr=3; p<0.0001) but a high non-significant p-value between RCU and control (χ^2=4.2; Deg.fr.=3; p=0.240). The results demonstrate that HS1,2A polymorphism influence specifically the predisposition to immunological diseases, while it could not be considered a genetic factor conferring the risk to diseases not showing immune system dysregulation.

A

HS1,2A genotypes obs (exp)	Ctr	CD	RCU
1/1	57 (53.4)	15 (13.7)	17 (13.1)
2/2	38 (40.7)	23 (26.2)	12 (10.9)
3/3	2 (0.9)	0 (0.02)	0 (0.01)
4/4	4 (2.6)	1 (1.6)	0 (0.9)
1/2	93 (93.3)	39 (38)	18 (24)
1/3	12 (14.0)	3 (1.1)	1 (0.8)
1/4	19 (23.8)	4 (9.4)	5 (6.8)
2/4	25 (20.8)	20 (13)	10 (6.2)
2/3	14 (12.2)	0 (1.5)	1 (0.8)
3/4	1 (3.1)	0 (0.4)	0 (0.2)
subjects	265	105	64

B

Alleles Fr ± SE (n)	Ctr N=265	CD N=105	*p*	RCU N=64	*p*
1	0.449 ± 0.022 (238)	0.362 ± 0.033 (76)	0.032	0.453 ± 0.044 (58)	1.000
2	0.392 ± 0.021 (208)	0.500 ± 0.035 (105)	0.008	0.414 ± 0.034 (53)	0.688
3	0.059 ± 0.010 (31)	0.014 ± 0.008 (3)	0.010	0.016 ± 0.011 (2)	0.044
4	0.100 ± 0.013 (53)	0.124 ± 0.023 (26)	0.357	0.117 ± 0.028 (15)	0.627
genomes	530	210		128	
Deg. fr.	5	2		1	
χ^2	3.2	8.4		4.1	
p	0.675	0.015		0.043	

Tab. 11. A Distribution of genotypes of the HS1,2A polymorphism in CD; RCU and control (Ctr) groups. The number of observed genotypes (obs) was reported. In bracket the number of expected genotypes (exp). **B** HS1,2A allelic frequencies ± standard error in healthy control, CD and RCU patients. Hardy-Weinberg equilibrium was reported for each group (Deg.fr.=degrees of freedom; $\chi 2$ =chi square; p=p-value). In brackets the number of observed genomes (n). Fisher's exact test was performed in cases (CD and RCU) versus the control (CD=Crohn's disease; RCU=ulcerousis recto colitis; *p*=two sides p-value).

3. Correlation of HS1,2A polymorphism with immunoglobulin serum levels

The role of HS1,2A polymorphism in the alteration of serum Ig concentrations was investigated into two independent cohorts of subjects: a group of 88 schizophrenic patients and a group of 133 healthy individuals namely general population. The concentration of the Ig in the different classes and IgG subclasses detected in blood serum of both normal and schizophrenic individuals indicate the presence of four subgroups: (1) all Ig levels within the normal range, (2) one or more classes or subclasses with too high levels, (3) one or more classes or subclasses with too low levels, and (4) a small group with both too low and too high levels (e.g. IgE above the norm and IgG1 below the norm)[4]. Mean values of the Ig classes and subclasses from general population and schizophrenic patients are reported in table 12. The schizophrenic group presents a lower means of IgG1 (5.60) and a higher means of IgG3 (5.76) compared to the general population (6.26, p=0.025 and 4.27, p=0.015 respectively). However, the weak levels of statistical significant do not allow us to take in serious consideration this finding. Thus no significant overall inter-group difference was register on Ig classes. The HS1,2A genotypes and the allelic frequencies ± the standard error for each group are reported in table 13. The analysis of the four alleles of the enhancer HS1,2A in the different groups with abnormal Ig plasma levels reveals a significant variation of frequency when compared to the total cohort (Tab. 13B). On the whole the cohort of schizophrenic patients is very similar to an Italian control group of the same geographical area. The groups with increased Ig levels have a frequency decrease for the allele 1 of 11% and 10% over the total for the control and schizophrenic subjects respectively. The allele 2 frequency in the same groups has an increase of 14% and 1% respectively. The groups with decreased Ig levels have an increase of the 5% or 22% of the frequency of the allele 1 and a decrease of 6% and 14% for allele 2 in the general population and schizophrenic subjects compared to the total cohorts respectively. The relative variation of the allele frequencies among the increased group compared to the decreased for allele 1 has a Δ of 17% and 33% respectively, whereas for the allele 2 has a Δ of 20% and 15% for the general population and schizophrenic subjects respectively.

---

[4] *- Ig+ or Ig- refer to Ig plasma levels increase or decrease above/below the standard values*

The difference of the four allele frequencies between the two normal groups in the general population and schizophrenic patients is due to the exiguous number in the schizophrenia. The distribution of control and schizophrenia subjects with Ig levels above and below the norm across the four alleles reveals significant differences. As regards the general population, the two Ig groups were unevenly distributed across the alleles (p=0.046) and the same was true, with stronger probability, for the schizophrenia sample (p=0.0008). After having pooled the two cohorts together, the chi-square distribution remained highly significant (p<0.001). When statistical comparisons were restricted to the first two alleles (Tab. 14), the results clearly indicated that in both cohorts there were significantly higher frequencies of allele 2 among subjects with Ig levels above the norm and of allele 1 among subjects with Ig levels below the norm. The chi-square distribution yielded for the two cohorts pooled together a highly significant value (p=0.0003).
These results show that in both samples, the group with Ig levels below the normal range showed an increase of allele 1 frequency and a decrease of the allele 2 frequency; on the other hand the group showing Ig levels above the norm had a decreased frequency of allele 1 and an increased frequency of allele 2. These data suggest that HS1,2A polymorphism could be correlated with the alteration of immunoglobulin serum levels in two independent and very different cohorts (Frezza *et al*, 2009; Mattioli *et al*, 2007).

	GP (N=133)		S (N=88)		*p*
	M	SD	M	SD	
IgA(mg/dl)	184.10	206.75	206.75	83.00	0.099
IgE (UI/ml)	73.76	236.58	236.58	143.61	0.124
IgM (mg/dl)	123.73	110.43	110.43	63.97	0.137
IgG (mg/dl)	1072.08	1031.22	1031.22	308.61	0.342
IgG1 (g/L)	6.26	5.60	5.60	1.96	0.025
IgG2 (g/L)	3.60	3.34	3.34	1.28	0.196
IgG3 (g/L)	4.27	5.76	5.76	2.13	0.015
IgG4 (g/L)	5.43	5.44	5.44	7.93	0.991

Tab. 12. Ig serum concentration in general population (GP) and schizophrenic (S) groups (M=mean; SD=standard deviation; *p*=two-sided p-value calculated from chi-square test)

A

HS1,2A genotypes obs (exp)	GP			
	Normal	Ig+	Ig-	Tot
1/1	16	3 (3.8)	12 (8.8)	32 (22.4)
2/2	2	9 (13.1)	6 (4.8)	19 (21.2)
3/3	0	0 (0)	0 (0)	0 (0.1)
4/4	0	0 (0.8)	0 (1.4)	0 (3.6)
1/2	5	18 (14.2)	7 (13)	31 (43.5)
1/3	2	0 (0.6)	0 (0.4)	2 (2.8)
1/4	3	2 (3.5)	7 (6.9)	12 (18.0)
2/4	11	10 (6.5)	8 (5.1)	32 (17.5)
2/3	2	2 (1.1)	1 (0.3)	5 (2.8)
3/4	0	0 (0.3)	0 (0.2)	0 (1.1)
subjects	41	44	41	133
HS1,2A genotypes obs (exp)	S			
	Normal	Ig+	Ig-	Tot
1/1	2 (2.5)	5 (3.3)	12	21(16.8)
2/2	5 (5.3)	6 (6.1)	2	18 (17.3)
3/3	0 (0.01)	0 (0)	0	0 (0.003)
4/4	0 (0.01)	2 (1.4)	0	2 (1.1)
1/2	8 (7.3)	8 (8.9)	9	30 (34.1)
1/3	0 (0.4)	0 (0)	0	0 (0.5)
1/4	1 (0.4)	2 (4.3)	0	5 (8.8)
2/4	0 (0.5)	7 (5.8)	2	11 (8.9)
2/3	1 (0.5)	0 (0)	0	1 (0.5)
3/4	0 (0.03)	0 (0)	0	0 (0.1)
subjects	17	30	25	88

B

		Alleles (Fr ± SE)				N
		1	2	3	4	
GP	Normal	0.512 ± 0.055	0.268 ± 0.049	0.049 ± 0.024	0.171 ± 0.041	41
	Ig+	0.295 ± 0.048	0.546 ± 0.053	0.023 ± 0.016	0.136 ± 0.036	44
	Ig-	0.463 ± 0.055	0.342 ± 0.052	0.012 ± 0.012	0,183 ± 0.043	41
	Tot	0.410 ± 0.030	0.399 ± 0.030	0.026 ± 0.009	0.165 ± 0.023	133
S	Normal	0.382 ± 0.083	0.560 ± 0.085	0.029 ± 0.029	0.029 ± 0.029	17
	Ig+	0.333 ± 0.061	0.450 ± 0.064	0	0.217 ± 0.053	30
	Ig-	0.660 ± 0,067	0.300 ± 0.065	0	0.040 ± 0.028	25
	Tot	0.437 ± 0.037	0.443 ± 0.037	0.006 ± 0.006	0.114 ± 0.024	88

Tab. 13. A Distribution of genotypes of the HS1,2A polymorphism in GP and S groups. The number of observed genotypes (obs) was reported. In bracket the number of expected genotypes (exp). **B** HS1,2A allelic frequencies ± standard error (Fr ± SE) in GP and S groups. (GP=general population; S=schizophrenia; normal=subjects with normal Ig levels; Ig+=subjects with Ig levels above the norm; Ig-=subjects with Ig levels below the norm; Tot=total group).

	Ig+		Ig-		χ^2	*p*	95% C.I.	O.R.
	1	2	1	2				
GP	26	48	38	28	6.20	0.013	0.20-0.79	0.40
S	20	27	33	15	5.59	0.018	1.28-6.89	2.97
GP+S	46	75	71	43	12.87	0.0003	1.59-4.56	2.69

Tab. 14. Statistical testing of the variation of allele 1/ 2 frequencies among Ig+ and Ig- subjects. (GP=general population; S=schizophrenia; χ^2=Chi-square test with Yate's correction, *p*=two-sided p-value; O.R.=odds ratio; c.i=95% confidence interval using approximation of Woolf).

4. Association of HS1,2A polymorphism with the progression of viral infections

4.1 HS1,2A allele 3 association with AIDS progression in an African cohort of HIV infected subjects

The role of the HS1,2A polymorphism in influencing the individual response to HIV infection was evaluated. Patients sample consisted of a group of 162 HIV positive subjects coming from the region of Swaziland (South Africa). The AIDS patients were divided into two groups according the serum circulating levels of lymphocyte CD4+. It is known that CD4+ serum levels decrease during the progression of HIV infection (CD4+>500 = normal level; CD4+<500= abnormal level) (data from ISS, Rome). A group of 36 AIDS patients with CD4+ serum levels <300 was compared to a group of 57 AIDS patients with CD4+ serum levels >700. The HS1,2A genotypes, the allele frequencies ± standard error and the Hardy Weinberg equilibrium of HIV infected groups are reported (Tab. 15). We found significant allelic frequency differences of CD4<300 group compared to CD4>700 group. Despite this, the genotype distribution do not seemed to vary significantly between the two groups. The HS1,2A allelic distribution in the cohort of HIV patients is very similar to HS1,2A allele frequencies of an African control group of the same geographical area. Allele 1 and 3 show the highest frequencies both in patient and control groups. No significant allele 1 frequency differences are found between the two groups of AIDS patients. On the other hand; allele 3 frequency is higher in CD4+<300 group compared to CD4+>700 group, with a value that shift from 0.344 to 0.210 respectively (p=0.028); whereas allele 4 frequency is lower in CD4+<300 group compared to CD4+>700 group (CD4+<300: 0.188; CD4>700: 0.316; p= 0.026). Chi-square test of the variation of allele 3 *versus* non 3 in CD4+<300 group in respect to CD4+>700 group confirmed the results (χ2=4.35; Deg.fr.=1; p=0.037; OR=2.12; c.i.=1.09-4.09). The statistical comparison of the allelic frequencies showed very low significant p-values among CD4<300 and CD4>700 groups (χ^2=11.35; Deg.fr.=3; p=0.010). From the data analysis result clearly that the increase of allele 3 frequency correlates with the decrease of CD4+ T cells levels. Therefore allele 3 of HS1,2A seems to influence the alteration of the individual immune response during HIV infection leading to AIDS progression.

A

Genotype (n, %)	HIV+		*p*
	CD4+<300	CD4+>700	
1/1	8 (22.2)	16 (28.1)	0.630
2/2	1 (2.8)	0 (0)	0.387
3/3	8 (22.2)	5 (8.8)	0.122
4/4	4 (11.1)	9 (15.8)	0.760
1/2	1 (2.8)	1 (1.7)	1.000
1/3	9 (25)	8 (14.0)	0.270
1/4	4 (11.1)	12 (21.0)	0.268
2/4	1 (2.8)	0 (0)	0.387
2/3	0 (0)	0 (0)	/
3/4	0 (0)	6 (10.5)	0.079
subjects	36	57	

B

Alleles Fr ± SE (n)	HIV+		*p*
	CD4+<300	CD4+>700	
1	0.437 ± 0.058 (30)	0.465 ± 0.047 (53)	0.548
2	0.031 ± 0.020 (4)	0.009 ± 0.009 (1)	0.075
3	0.344 ± 0.056 (26)	0.210 ± 0.038 (24)	0.028
4	0.188 ± 0.046 (12)	0.316 ± 0.043 (36)	0.026
genomes	72	114	
χ^2	4.5	7.3	
Deg. fr.	1	3	
p	0.034	0.063	

Tab. 15. A Distribution of genotypes of the HS1,2 A polymorphism in CD4+<300 and CD4+>700 HIV positive groups. The number of observed genotypes (obs) and percent (%; in bracket) was reported. **B** HS1,2A allelic frequencies ± standard error in CD4+<300 and CD4+>700 HIV positive groups. Hardy-Weinberg equilibrium was reported for each group (Deg.fr.=degrees of freedom; χ2=chi square; p=p-value). In brackets the number of observed genomes (n) (HIV+=HIV positive subjects; CD4+<300=subjects with CD4+ level <300; CD4+>700=subjects with CD4+ level >700; *p*=two sides p-value calculated from Fisher's exact test performed in CD4+<300 versus CD4+>700).

4.2 HS1,2A allele 2 association with progressive HCV infection

HS1,2A polymorphism association with Hepatitis C progression in an Italian cohort of HCV infected subjects was investigated. A cohort of 82 HCV infected patients was analyzed (data not shown). The acutely infected HCV subjects were divided into two groups according the individual response to HCV progression of infection: patients who were chronically infected (I; n=64) and patients showing spontaneous viral clearance (CL; n=18). Then, the group of patients chronically infected was divided into two subgroups: the patients with hepatic cirrhosis (CR; n= 53) and the ones without cirrhosis (NCR; n=29). The HS1,2A genotypes, the allele frequencies ± standard error and the Hardy Weinberg equilibrium for all groups of patients are shown in table 16. All the groups proved to be in Hardy-Weinberg equilibrium except for CL group, perhaps due to the small number of the sample. We found no significant differences in genotype distribution or allelic frequencies between the group of chronically infected patients and the group of patients showing spontaneous viral clearance. We observed a decrease of allele 1 and an increase of allele 2 frequencies in the infected group (0.344 and 0.516 respectively) in respect to CL group (0.389 and 0.444 respectively); even if the differences even if the difference didn't reach the statistical significance (p=0.694 and p=0.572 respectively). Neither did we observe any differences in genotype distribution and allelic frequencies between cirrhotic and non-cirrhotic patients. We conclude that among the host genetic factors, HS1,2A polymorphism could play a role in the pathogenesis of HCV infection. In particular, allele 2 frequency increase and allele 1 frequency decrease seem to be associated with chronic infection, albeit in a not significant way due to the small number of sample. Finally, HS1,2A polymorphism does not associate with progressive hepatic cirrhosis.

A						
Hs1,2A genotype (n, %)	HCV		*p*	I		*p*
	I	CL		NCR	CR	
1/1	5 (7.8)	2 (11.1)	0.645	4 (7.5)	3 (10.3)	0.693
2/2	13 (20.3)	3 (16.8)	1.000	11 (20.8)	5 (17.2)	0.778
3/3	1 (1.6)	0 (0)	1.000	1 (1.9)	0 (0)	1.000
4/4	0 (0)	2 (11.1)	0.046	2 (3.8)	0 (0)	0.537
1/2	29 (45.3)	9 (50)	0.793	25 (47.2)	13 (44.8)	1.000
1/3	0	0	/	0	0	/
1/4	5 (7.8)	1 (5.5)	1.000	4 (7.5)	2 (7)	1.000
2/4	10 (15.6)	1 (5.5)	0.441	5 (9.4)	6 (20.7)	0.184
2/3	1 (1.6)	0	1.000	1 (1.9)	0 (0)	1.000
3/4	0	0	/	0	0	/
subjects	64	18		53	29	
B						
Alleles Fr ± SE (n)	HCV		*p*	I		*p*
	I	CL		NCR	CR	
1	0.344 ± 0.042 (44)	0.389 ± 0.081 (14)	0.694	0.349 ± 0.046 (37)	0.362 ± 0.063 (21)	0.866
2	0.516 ± 0.044 (66)	0.444 ± 0.083 (16)	0.572	0.500 ± 0.048 (53)	0.500 ± 0.065 (29)	1.1296
3	0.023 ± 0.013 (3)	0	1.000	0.028 ± 0.016 (3)	0	0.5529
4	0.117 ± 0.028 (15)	0.167 ± 0.062 (6)	.410	0.122 ± 0.032 (13)	0.138 ± 0.045 (8)	0.8096
genomes	128	36		106	58	
χ^2	3.3	7.8		5.0	2.2	
Deg. Fr.	1	1		2	1	
p	0.070	0.005		0.081	0.138	

Tab. 16. A Distribution of genotypes of the HS1,2A polymorphism in in I, CL, NCR, CR groups of HCV patients. The number of observed genotypes (obs) and percent (%; in bracket) was reported. **B** HS1,2A allelic frequencies ± standard error in I, CL, NCR, CR groups of HCV patients. Hardy-Weinberg equilibrium was reported for each group (Deg.fr.=degrees of freedom; χ2=chi square; p=p-value). In brackets the number of observed genomes (n). (I=infected; CL=viral clearance; NCR=non-cirrhotic; CR=cirrhotic; *p*=two sides p-value calculated from Fisher's exact test performed in cases versus controls).

5. Analysis of the haplothypes of the regions surrounding the HS1,2A and HS3 enhancers

5.1 The SNPs of the regions surrounding the HS1,2A and HS3 enhancers

One aim of this thesis is to discover other genetic variants of the regions surrounding the enhancers HS3 and HS1,2A of the 3'RR-1 associated with both HS1,2A allele 1 and allele 2. Genome resequencing of two regions encompassing the enhancers was performed (Fig.5). The first region spans over 2.2 kb and lies between the repeated sequence R2 and the conserved sequence U6, encompassing the enhancer HS3. The second region spans of 1.7 kb and lies between the conserved sequences U8 and U9; this region includes the enhancer HS1,2A. A 5.4 kb fragment including both regions was amplified by a selective PCR. Subsequently the two regions were specifically selected by two nested PCRs. The NCBI chromosome clone NT_026437.11 (from 87168750 nt to 87163906 nt) was the reference sequence. Figure 5 shows a schematic map of the identified SNPs. The first nucleotide is the "wild type" nucleotide (in reference to NT_026437.11), while the second nucleotide is the "mutated" or the "alternative" nucleotide. Each SNP was tagged with the same numeration of NT_026437.11 without the first four numbers (i.e. the SNP lying at 87168750 position of NT_026437.11 is tagged as 8750). The first region (R2-U6) harbours 14 SNPs : 2 SNPs (8750; 8705) belong to the conserved sequence U3; 1 SNP (8339) belongs to HS3; 1 SNP (8283) belongs to U4 and 10 SNPs belong to U6 (7697; 7662; 7570; 7548; 7547; 7538; 7519; 7362; 7265 and 7159). In the second region (U8-U9) 8 SNPs are identified: 7 SNPs belong to U8 (5159; 5077; 4811; 4690; 4566; 4231; 4088) and 1 SNP (3906) belongs to HS1,2A. In total, 22 SNPs were identified across the two regions. The central region lying between the end of U6 and the beginning of U8 was not analyzed. There are two reasons for this choice. The first reason is that we decided to analyze the regions including HS3 and HS1,2A in order to identify polymorphisms within the two enhancers. The second is that, even if the central region is shorter in respect to the other ones, it is very difficult to amplify because it contains several repeated sequences (R3 and R4). In order to facilitate the analysis, from now on we will consider the two regions as a unique one (Mattioli *et al*, 2007).

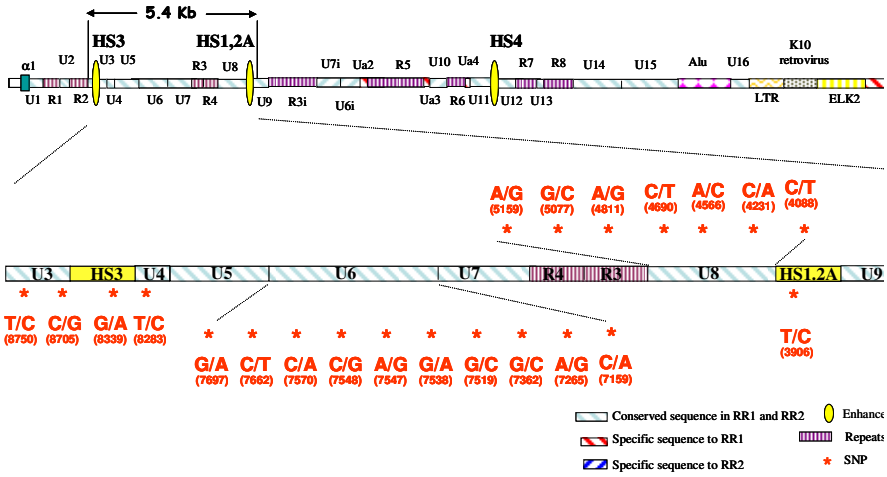


Fig. 5. Schematic map of the 3'RR-1 showing the SNPs identified in the region lying between the enhancers HS,3 and HS1,2A.

5.2 The HS1,2A genotype analysis

An extended cohort of psoriatic arthritis patients and healthy subjects from the same geographical area was genotyped for HS1,2A polymorphism (data not shown). An equal number of individuals carrying allele 1 and 2A in homozygosis were selected in both groups. A few numbers of subjects heterozygous 1/2 were also analyzed. The selected sample consisted of a group of 30 healthy subjects and 29 psoriatic arthritis patients. Among the 30 healthy subjects, 11 are homozygous for the allele 1; 14 are homozygous for the allele 2; and 5 are heterozygous 1/2. Among the 29 psoriatic arthritis patients, 12 are 1/1; 12 are 2/2 and 3 are 1/2. The genotype analysis by genome-resequencing confirmed the giving HS1,2A genotypes of the individuals of the control and patient groups. Therefore control subjects and psoriatic patients carrying 1/1 and 2/2 HS1,2A genotype in the same ratio were analyzed. The first aim is to compare the individuals carrying 1/1 genotype with the individuals carrying 2/2 genotype. Then we compared the control and psoriatic patient groups homozygous for the allele 1 and for the allele 2.

5.3 The Linkage Disequilibrium (LD) analysis

Linkage disequilibrium (LD) analysis among the identified SNPs was performed. D' values calculated for each pair of SNPs are represented by the scale of blue (Fig. 6). D' value shifts from a minimal value of 0 (no linkage; white) and a maximal value of 1 (maximum linkage; black). As shown in the map, the major part of the identified SNPs are in maximum linkage disequilibrium with each other (D'=1; black); while many SNPs are in strong linkage disequilibrium with each other (D'=0.8-0.9; dark blue). Only few SNPs are in weak linkage with each other (D'=0.30-0.40; light blue). In particular, the 4231-4690 and 4231-5077 pairs of SNPs show the lower linkage disequilibrium value (D'=0.30), following by 4231-7159 and 5077-7159 pairs of SNPs (D'=0.40), and by 4231-4811 pair of SNPs (D'=0.50). The results suggest that SNP 4231 and SNP 5077 account for the major part of the D' deviation from its highest value. However, LD plot shows widespread maximum linkage disequilibrium among the identified SNPs. The strong statistical association among the polymorphisms allows the identification of the haplotypes in the sequenced region.

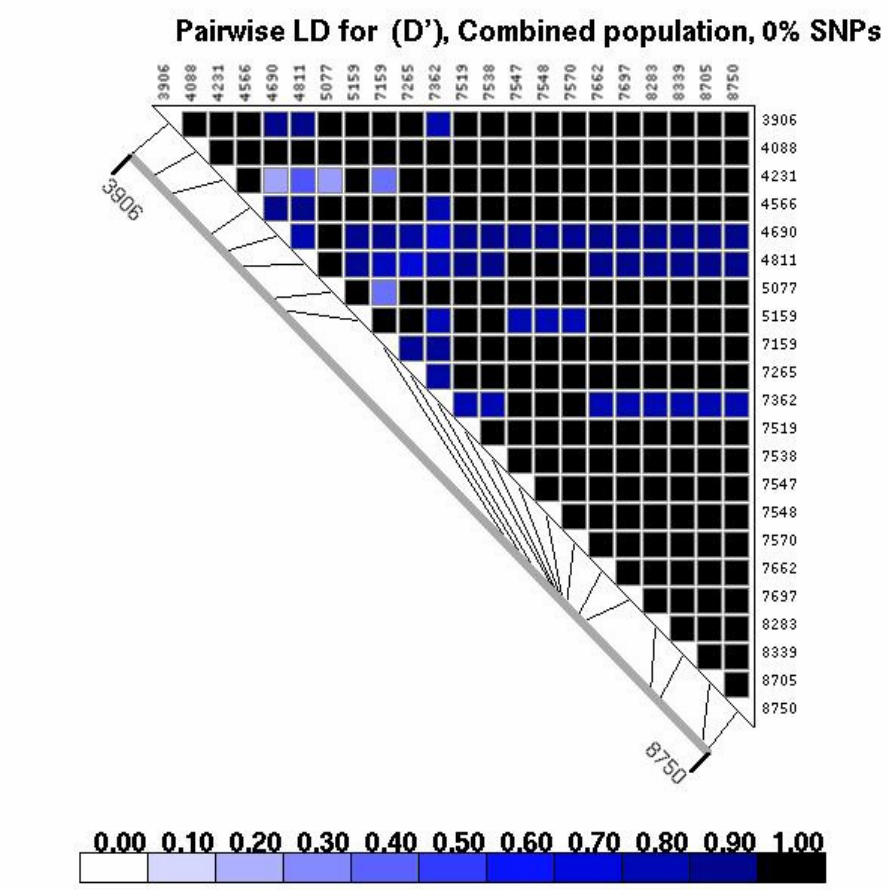


Fig. 6. Linkage Disequilibrium (LD) plot among the identified SNPs. (Black: D'=1, maximum linkage; white: D'=0, minimum linkage).

5.4 The haplotype frequency analysis

5.4.1. The haplotypes of the allele 1 and allele 2 groups

First controls and patients homozygous for the allele 1 were pooled together, as well as control and patients homozygous for the allele 2. We obtained two groups: a group of 23 individuals (including controls and

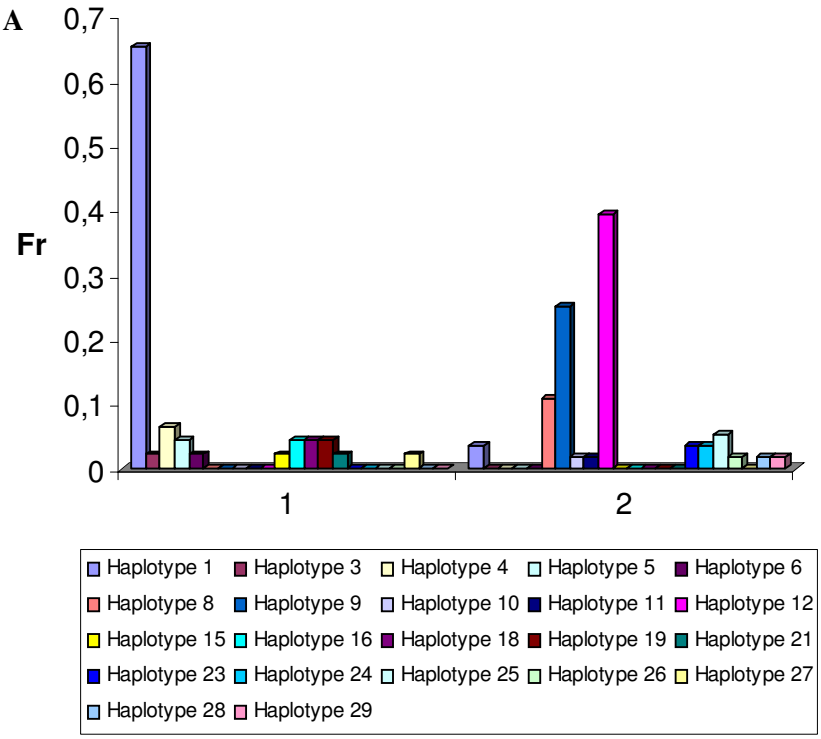
patients) homozygous for the allele 1 (allele 1 group) and a group of 28 individuals (including controls and patients) homozygous for the allele 2 (allele 2 group). The aim is to compare the haplotype frequencies of the allele 1 and allele 2 groups; in order to identify haplotypes associated with one allele or another. As we could not distinguish between the plus and minus strand of the sequences, homozygous 1/1 and 2/2 genotypes were used in order to correlate the haplotypes with allele 1 or allele 2. EM algorithm was used to estimate haplotype frequencies of both groups by Arlequin ver 3.0 (Excoffier *et al*, 2005). Table 17 shows the frequencies ± the standard error of the haplotypes of the allele 1 and allele 2 groups. We found a large number of haplotypes poorly represented in allele 1 group (haplotypes 3, 4, 5, 6, 15, 16, 18, 19, 21, 27) and in allele 2 group (haplotypes 10, 11, 23, 24, 25, 26, 28, 29); a small number of haplotypes highly represented in allele 2 group (haplotypes 9, 12) and in both groups (haplotype 1); and only 1 haplotype medium represented in allele 2 group (haplotype 8). Statistical analysis of the variation of the haplotype frequencies of allele 1 group in respect to allele 2 group was performed by Fisher's exact test. No significant variation of the haplotype frequencies between the two groups was found for the haplotypes poorly represented ($p>0.05$). Beside this, we found significant haplotype frequency differences between the two groups for the four major haplotypes ($p<0.0001$). Haplotype 1 frequency is significantly increased in allele 1 group (0.652) compared to allele 2 group (0.036; $p<0.0001$). Haplotype 8, 9 and 12 frequencies are significantly increased in allele 2 group compared to allele 1 group: the haplotype frequencies shift from 0.000 in allele 1 group to 0.107, 0.250 and 0.393 in allele 2 group respectively ($p=0.026$; $p<0.0001$; $p<0.0001$ respectively). A graphic representation of the haplotype frequency differences between allele 1 and allele 2 groups is reported in figure 7A. The picture shows a strong statistical association of haplotype 1 with allele 1 group and of haplotypes 8, 9 and 12 with allele 2 group. The nucleotide structure of the 29 haplotypes is shown in figure 7B. The nucleotide sequence is numbered decreasing of NT_026437.11 (the first SNP is 8750 and the last 3906). If we look to SNPs differences among the four major haplotypes (1; 8; 9 and 12) we observe a strong nucleotide difference between haplotype 1 and haplotype 8, 9, and 12, which is of 77.3; 81.8 and 86.4% respectively. In particular, haplotype 1 harbours the highest number of wild type nucleotides and only 1 SNP in the alternative form. Haplotypes 8, 9, and 12 have the majority of alternative nucleotides and only 6, 5, and 4 wild type nucleotides respectively. Among the most representative

haplotypes, the haplotype with the highest frequency in allele 2 group is the one with the highest percent of alternative nucleotides (haplotype 12; 0.393; 81.8%) following by haplotype 9 (0.250; 77.3%) and 8 (0.107; 72.7%). The results demonstrate a strong specific association between HS1,2A allele 1 and 2 and the haplotypes identified in the region lying between the enhancers HS3 and HS1,2A. In particular, the haplotype with the highest number of wild type nucleotide (H1) is associated with allele 1; while the haplotypes with the highest number of alternative nucleotides (H 8; 9 and 12) are associated with allele 2. Finally, the importance of the correlation of the haplotypes with allele 2 depends on the structure of the haplotypes themselves. The correlation with allele 2 is more significant as the number of alternative nucleotides of the haplotype is higher. These results demonstrate that other genetic variants over the HS1,2A polymorphism could be involved in the 3'RR-1 functions. The identified haplotypes strongly correlate with HS1,2A alleles and together could influence the activity of the regulatory region.

H	Allele 1(23)		Allele 2(28)		
	Fr ± SE	n	Fr ± SE	n	*p*
1	0.652 ± 0.033	30	0.036 ± 0.013	2	<0.0001
3	0.022 ± 0.010	1	-		0.451
4	0.065 ± 0.017	3	-		0.085
5	0.043 ± 0.014	2	-		0.198
6	0.022 ± 0.010	1	-		0.451
8	-		0.107 ± 0.021	6	0.026
9	-		0.250 ± 0.030	14	<0.0001
10	-		0.018 ± 0.009	1	1.000
11	-		0.018 ± 0.009	1	1.000
12	-		0.393 ± 0.034	22	<0.0001
15	0.022 ± 0.010	1	-		0.451
16	0.043 ± 0.014	2	-		0.198
18	0.043 ± 0.014	2	-		0.198
19	0.043 ± 0.014	2	-		0.198
21	0.022 ± 0.010	1	-		0.451
23	-		0.036 ± 0.013	2	0.495
24	-		0.036 ± 0.013	2	0.495
25	-		0.053 ± 0.013	3	0.242
26	-		0.018 ± 0.013	1	1.000
27	0.022 ± 0.010	1	-		0.451
28	-		0.018 ± 0.013	1	1.000
29	-		0.018 ± 0.013	1	1.000

Tab. 17. Haplotype frequencies ± standard error (Fr ± SE) in allele 1 and allele 2 groups. In brackets the number of subjects. Two sides p-value (*p*) is calculated from Fisher's exact test of allele 1 versus allele 2. (H=haplotype; n=number of observed haplotypes)

A



B

Haplotype 1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Haplotype 3	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	1	0	1	0	0	0
Haplotype 4	0	0	0	0	0	0	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0
Haplotype 5	0	0	0	0	0	0	1	1	1	0	0	1	0	0	0	0	0	1	0	0	0	0
Haplotype 6	0	0	0	0	0	0	1	1	1	0	0	1	0	0	1	0	0	0	0	0	0	0
Haplotype 8	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0	1	1	0	0	0	1
Haplotype 9	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1	1	0	0	0	1
Haplotype 10	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1	1	1	0	1	1
Haplotype 11	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0	0	1
Haplotype 12	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	0	0	1
Haplotype 15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0
Haplotype 16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0
Haplotype 18	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0
Haplotype 19	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	1	0	0
Haplotype 21	0	0	0	0	0	0	1	1	1	0	0	1	0	0	0	0	1	0	1	0	0	0
Haplotype 23	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	0	1	1	0	0	0	1
Haplotype 24	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	0	0	1
Haplotype 25	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	0	1	1
Haplotype 26	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	0	1	0	1
Haplotype 27	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0
Haplotype 28	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	0	1	1	0	0	0	1
Haplotype 29	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	0	1	0	1

Fig. 7. A Graphic representation of haplotype frequencies in allele 1 and allele 2 groups. **B** Nucleotide structure of the haplotypes (0=wild type nucleotide; 1=alternative nucleotide)

5.4.2. The haplotypes of the control and patient allele 1 groups

Healthy and psoriatic individuals homozygous for the allele 1 were compared. The control group consisted of 11 homozygous for the allele 1 (Ctr 1), while psoriatic group consisted of 12 individuals homozygous for the allele 1 (Ps 1). EM algorithm was used to estimate haplotype frequencies of both groups by Arlequin ver 3.0 (Excoffier *et al*, 2005). Table 18 shows the haplotype frequencies ± the standard error of Ctr 1 and Ps 1 groups. Also in this case we found a large number of haplotypes poorly represented both in Ctr 1 group (haplotype 3, 4, 5, 6) and Ps 1 group (4, 5, 15, 16, 18, 19, 21, 27) but only 1 haplotype highly represented in both groups (haplotype 1). There are no significant haplotype frequency variations between the two groups for the haplotypes poorly represented ($p>0.05$). Haplotype 1 frequency is lower in Ps 1 group (0.542) in respect to Ctr 1 group (0.773). Despite this, no significant haplotypes 1 frequency difference between the two groups was found ($p=0.129$). Figure 8 shows a graphic representation of haplotype frequencies in Ctr 1 and Ps 1 groups. From these data result evident that there are no significant haplotype frequency differences between the two groups homozygous for the allele 1.

H	Ctr 1(11)		Ps 1(12)		
	Fr ± SE	n	Fr ± SE	n	*p*
1	0.773 ± 0.029	17	0.542 ± 0.034	13	0.129
3	0.045 ± 0.014	1	-		0.478
4	0.091 ± 0.020	2	0.042 ± 0.014	1	0.590
5	0.045 ± 0.014	1	0.042 ± 0.014	1	1.000
6	0.045 ± 0.014	1	-		0.478
15	-		0.042 ± 0.014	1	1.000
16	-		0.083 ± 0.020	2	0.478
18	-		0.083 ± 0.020	2	0.478
19	-		0.083 ± 0.020	2	0.478
21	-		0.042 ± 0.014	1	1.000
27	-		0.042 ± 0.014	1	1.000

Tab. 18. Haplotype frequencies ± standard error (Fr ± SE) in Ctr 1 and Ps 1 groups. Two sides p-value (*p*) is calculated from Fisher's exact test of Ctr 1 versus Ps 1 groups. In brackets the number of subjects. (Ctr 1= control group 1/1; Ps 1=psoriatic patients 1/1; H= haplotype; n= number of observed haplotypes).

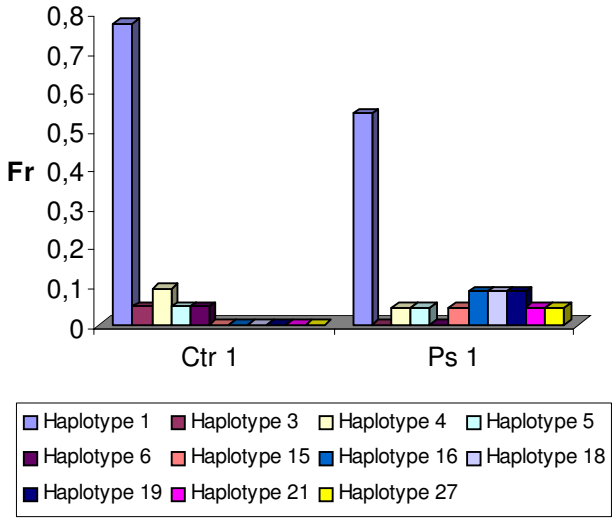


Fig. 8. Graphic representation of haplotype frequencies in Ctr 1 and Ps 1 groups (Ctr 1= control group 1/1; Ps 1= psoriatic patients 1/1).

5.4.3. The haplotypes of the control and patient allele 2 groups

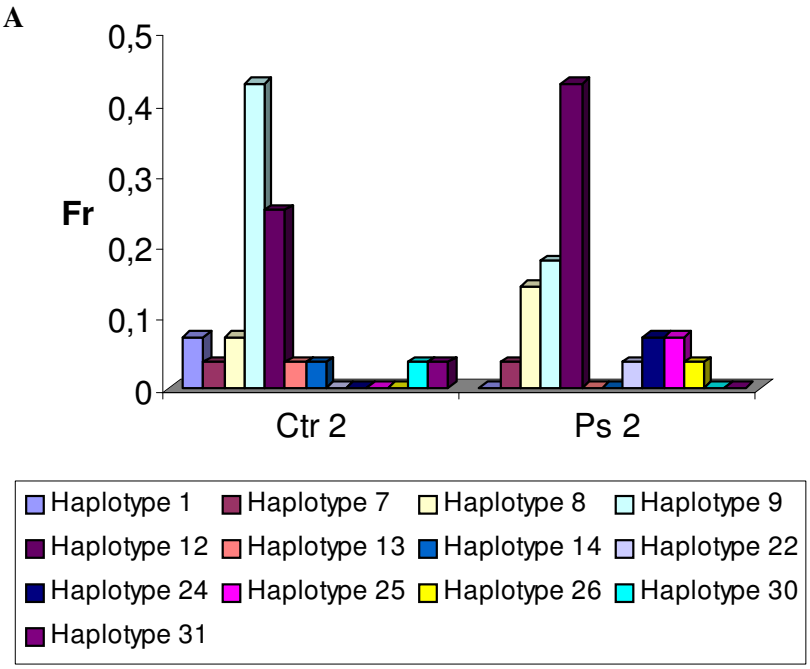
Healthy and psoriatic individuals homozygous for the allele 2 were compared. The control group consisted of 14 individuals homozygous for the allele 2 (Ctr 2), while psoriatic group consisted of 14 individuals homozygous for the allele 2 (Ps 2). EM algorithm was used to estimate haplotype frequencies of both groups by Arlequin ver 3.0 (Excoffier *et al*, 2005). Table 19 shows the haplotype frequencies ± the standard error of Ctr 2 and Ps 2 groups. Several new haplotypes were found in Ctr and Ps allele 2 groups in respect to the total group (Tab. 17), such as haplotype 7, 13, 14, 22, 30 and 31. The haplotypes showing the lowest frequencies are haplotypes 7, 13, 14, 30, 31 in Ctr 2 and 7, 22, 26 in Ps 2 with a frequency of 0.036; haplotypes 1, 8 in Ctr 2 and 24, 25 in Ps 2 with a frequency of 0.071. Haplotype 8 shows a medium frequency in Ps 2 (0.143). The haplotypes showing the highest frequencies are haplotypes 9 and 12, with a frequency of 0.428 and 0.250 in Ctr 2 respectively; of 0.178 and 0.428 in Ps 2 respectively. The comparison of the haplotype frequency of the Ctr 2 and

Ps 2 groups showed no significant differences between the two groups. In spite of that, haplotype 9 frequency decreases in Ps 2 compared to Ctr 2 groups (p=0.079), while haplotype 12 frequency increases in Ps 2 compared to Ctr 2 groups (p=0.259). A graphic representation of haplotype frequencies of Ctr 2 and Ps 2 groups was reported in figure 9. The histogram shows the inversion of the frequencies among haplotypes 9 and 12 in the two groups. We found that only one nucleotide change between the two haplotypes (Fig. 7B). In particular, haplotype 9 presents the wild form of SNP 7265, whereas haplotype 12 shows the alternative form of this SNP.
In conclusion, we found haplotype frequency differences between the two groups homozygous for the allele 2, even if the differences even if the difference didn't reach the statistical significance. To obtain significant p-values may be necessary to increase the number of sample. However, these preliminary data demonstrate that the haplotypes associated with allele 2 in a control group are different in respect to the haplotypes associated with allele 2 in a psoriatic patient group, suggesting that different haplotypes could affect the activity of the 3' regulatory region (Frezza *et al*, 2008).

H	Ctr 2(14)		Ps 2(14)		*p*
	Fr	n	Fr	n	
1	0.071 ± 0.018	2	-		0.481
7	0.036 ± 0.013	1	0.036 ± 0.013	1	1.518
8	0.071 ± 0.018	2	0.143 ± 0.024	4	0.648
9	0.428 ± 0.034	12	0.178 ± 0.026	5	0.079
12	0.250 ± 0.030	7	0.428 ± 0.026	12	0.259
13	0.036 ± 0.013	1	-		1.000
14	0.036 ± 0.013	1	-		1.000
22	-		0.036 ± 0.013	1	1.000
24	-		0.071 ± 0.018	2	0.481
25	-		0.071 ± 0.018	2	0.481
26	-		0.036 ± 0.013	1	1.000
30	0.036 ± 0.013	1	-		1.000
31	0.036 ± 0.013	1	-		1.000

Tab. 19. Haplotype frequencies ± standard error (Fr ± SE) in Ctr 2 and Ps 2 groups. Two sides p-value (*p*) is calculated from Fisher's exact test of Ctr 2 versus Ps 2 groups. In brackets the number of subjects (Ctr 2=control group 2/2; Ps 2=psoriatic patients 2/2; H=haplotype; n=number of observed haplotypes).

A



B

Haplotype 1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Haplotype 7	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	1	1	0	0	0	1
Haplotype 8	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0	1	1	0	0	0	1
Haplotype 9	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1	1	0	0	0	1
Haplotype 12	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	0	0	1
Haplotype 13	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0	1
Haplotype 14	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1
Haplotype 22	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	0	1	1	0	0	0	1
Haplotype 24	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	0	0	1
Haplotype 25	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	0	1	1
Haplotype 26	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	0	1	0	1
Haplotype 30	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1	1	1	0	0	1
Haplotype 31	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0	1	1

Fig. 9. A Graphic representation of haplotype frequencies in Ctr 2 and Ps 2 groups. **B** Nucleotide structure of the haplotypes (0=wild type nucleotide; 1=alternative nucleotide)

5.5 The "analysis of molecular variance" (AMOVA)

To evaluate the degree of genetic heterogeneity of the examined groups the "analysis of molecular variance" (AMOVA) was performed using Arlequin ver 3.0 (Excoffier *et al*, 2005).
The genetic variability of allele 1 and allele 2 groups was investigated. The genetic structure is composed by two groups (allele 1 and allele 2) and two populations within both groups (Ctr 1 and Ps 1; Ctr 2 and Ps 2). Three components account for the total of genetic variability: the variability within populations (Vc); the variability among populations within groups (Vb) and the variability among groups (Va) (Tab. 20). The percentage of variation was reported for each component. The results show that the major component of the variance is the variability within populations (Vc=72.37%; corresponding to the mean of the differences among the individuals of the same population), whereas the other components of variability (Va and Vb) show a lower value, as expected from population genetic studies (Barbujani *et al*, 1996; Barbujani and Goldstein, 2004). Moreover, the variability among groups (Va) presents a higher percent of variation (27.11%) in respect to the variability among populations (Vb=0.52%). Therefore allele 1 and 2 groups differ with each other much more than the populations within each group. The p-value associated with the variability component Vc is extremely significant (0.000). On the other hand, Va and Vb p-values have proved to have no significance, perhaps due to the small number of sample. Nevertheless, we can conclude that AMOVA test confirms the results obtained by the analysis of haplotype frequencies. In fact the differences between allele 1 and allele 2 groups are stronger in respect to the differences between the two populations (control and patients) within each group.
Moreover, AMOVA analysis was carried to estimate the genetic variability of control and psoriatic groups both homozygous for the allele 1. The same analysis was carried to evaluate the genetic differences between the individuals of control and psoriatic groups both homozygous for the allele 2. The genetic structure is composed by only one group (allele 1 or allele 2) and two populations within each group (control and psoriatic individuals homozygous for allele 1 or allele 2). Therefore the genetic variability possess two components: Va, that is the variability between the populations (control and patients) of a same group (the total group with one allele or the other); and Vb, that is the variability within populations. Allele 1 group showed a Va value of -1.75 and a Vb value of 101.75; whereas allele 2 group showed a Va of 3.26 and a Vb of 96.74. Also in this case the major

component of the variance is the variability within the populations (Vb). The Va component is increased in allele 2 group compared to allele 1 group. Therefore, as expected from previous results, the variability between control and psoriatic allele 2 groups is higher in respect to the variability between control and psoriatic allele 1 groups.

Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation
Among groups	1	3.747	0.13229 Va	27.11
Among populations within groups	2	0.770	0.00253 Vb	0.52
Within populations	47	16.600	0.35319 Vc	72.37
Total	50	21.118	0.48801	

Fixation Indices
FSC : 0.00710
FST : 0.27626
FCT : 0.27109

Significance tests (1023 permutations)

Vc and FST : P(rand. value < obs. value) = 0.00000
P(rand. value = obs. value) = 0.00000
P-value = 0.00000+-0.00000

Vb and FSC : P(rand. value > obs. value) = 0.31672
P(rand. value = obs. value) = 0.02151
P-value = 0.33822+-0.01633

Va and FCT : P(rand. value > obs. value) = 0.00000
P(rand. value = obs. value) = 0.31476
P-value = 0.31476+-0.01316

Tab. 20. Analysis of molecular variance of allele 1 group *versus* allele 2 group.

5.6 The CpG plot analysis

Different haplotypes could affect the CpG island structure of the regulatory regions. CpG plot of the 5.4 Kb region was performed with haplotype 1, 8, 9 and 12 by EMBL-EBI database (Fig. 10). The peak represents the observed CpG islands in relation to the expected ones along the giving region. As the picture shows, there are some differences (black small arrow) in CpG distribution between haplotype 1 (Fig. 10A) and haplotypes 12 (Fig. 10B). On the other hand, no differences were found among haplotypes 8, 9, and 12 (data not shown). It could be hypothesized that different CpG structures of the 5.4 Kb region carrying different haplotypes could affect different methylation patterns or epigenetic changes influencing the activity of the 3' RR-1.

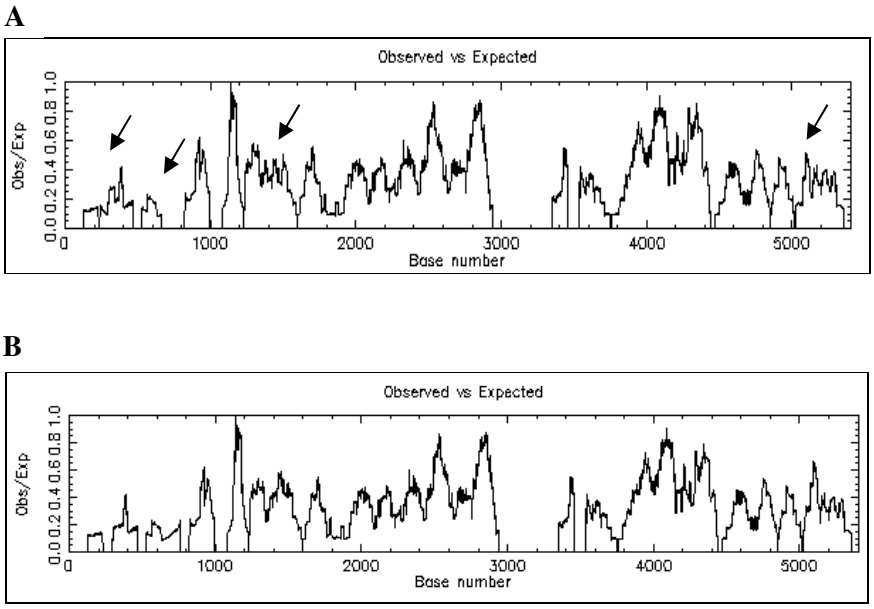


Fig. 10. CpG plot of the 5.4 Kb region carrying haplotype 1 (**A**) and 12 (**B**). (Black small arrow: CpG differences between haplotype 1 and haplotype 12).

5.7 The transcription factor binding sites analysis

Different haplotypes could affect the binding site sequences for several transcription factors involved in the activity of the 3' RR-1. Transcription factor binding site analysis of the 5.4 Kb region was carried with haplotypes 1, 8, 9, and 12 by TFSEARCH program. Hypothetical binding sites for several transcription factors were acquired or lost according to the presence of the wild type or the mutated form of a particular SNP. For example SNP 8750 is present in the wild type form ("T") in haplotype 1 and in the mutated form ("C") in haplotypes 8, 9 and 12 (Fig. 11). Nucleotide "T" determines the binding sites for the transcription factors YY1, HSF and CdxA, while the haplotypes with the mutated nucleotide ("C") lose the binding sites for YY1 and CdX4.

Previously we found that the 7265 is the only nucleotide which changes between haplotypes 9 and 12, and that haplotype 9 correlates with Ctr 2 group while haplotype 12 correlates with Ps 2 group even if in a not significant manner (Fig. 12A). SNP 7265 shows the wild type form ("A") in haplotype 9 and no hypothetical binding sites were detected. On the other hand, SNP 7265 shows the alternative nucleotide ("G") in haplotype 12 and a hypothetical binding site for ADR1 (TGGAGA) was found (Fig. 12B). Therefore different haplotypes could influence the binding of several transcription factors involved in the regulation of the activity of the regulative region, so influencing the function of the region itself.

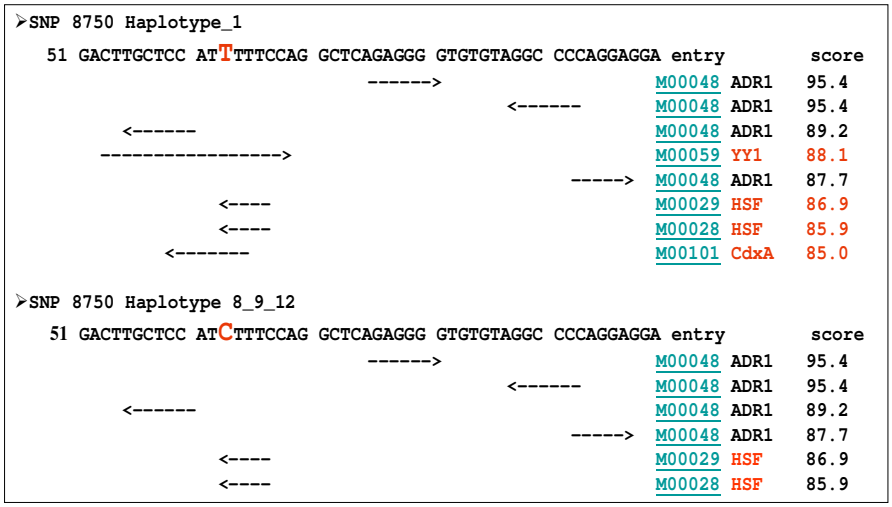


Fig. 11. Transcription factor binding sites analysis of the sequence including the SNP 8750 of haplotype 1 and of the haplotypes 8-9-12 (red: 8750 SNP; transcription factor binding sites for 8750 SNP).

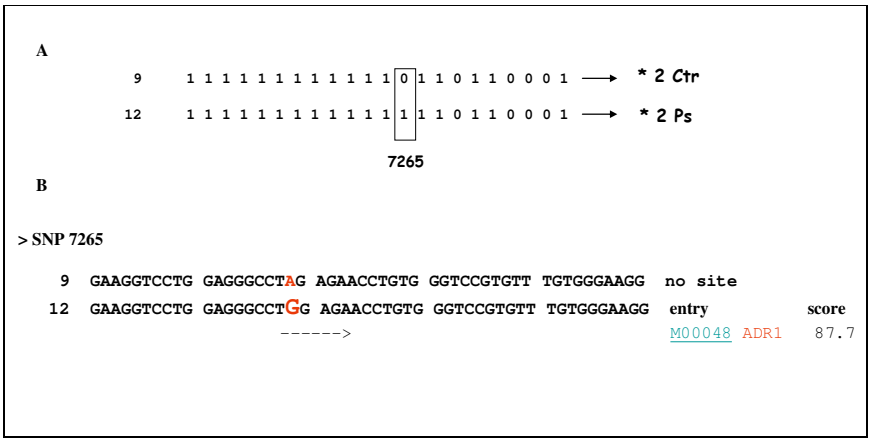


Fig. 12. A SNP difference between haplotype 9 and 12. **B** Transcription factor binding sites analysis of the sequence including the 7265 SNP of the haplotypes 9 and 12 (red: 7265 SNP; small arrow: binding sites for ADR1).

DISCUSSION

1. The HS1,2A polymorphism

The regulation of human immunoglobulin heavy chain gene expression is incompletely understood. Both in human and in mouse, the enhancer Eμ and the regulatory regions (3'RRs) lying downstream of the heavy chain constant α gene are able to enhance immunoglobulin transcription from the heavy chain promoters in a tissue and stage specific manner. Several experiments both in vivo and in vitro have elucidated the role of the enhancers of the 3' RR in mouse, whereas little is known about the mechanism of function of the two human regulatory regions (RR-1 and RR-2). The duplication of the human 3'RRs along with four constant genes make experiments difficult, also because the structure and the sequence of the two 3'RRs are highly conserved. Nevertheless, the polymorphic structure of the two HS1,2 enhancers is well-characterized (Giambra *et al*, 2005).

1.1 Distribution of HS1,2A alleles in the Italian population and in several other healthy European populations living in geographic isolation.

Previous studies have demonstrated that HS1,2A (of 3'RR-1) polymorphism is variable through world population, whereas HS1,2B (of 3'RR-2) alleles do not vary across the continent (Giambra *et al*, 2006). The first aim of the project is to extend the analysis for the HS1,2A locus to the whole of Italy, in order to verify the degree of genetic variability by AMOVA. Italy resulted rather homogeneous for this locus, even if the genotype distribution shows some differences among the populations analyzed (Mattioli *et al*, 2008). In particular, we found an increase of allele 2 frequency in Sardinia in respect to the other regions. We hypothesized that the differences are due to the geographic isolation of the Sardinian population. We also found significant differences in HS1,2A allele frequencies between Italian and Spanish population isolated by natural barriers in contrast with Italian and Spanish populations not subjected to such isolation. So we can conclude that the natural barriers impaired the genetic flux for this locus among the population analyzed.

1.2 Association of HS1,2A polymorphism with several autoimmune diseases and selective IgA deficiency.

In vitro transfection experiments demonstrate clearly both in human and mouse the importance of the enhancers of the 3' RRs in the regulation of immunoglobulin production. However, unlike in the case of mice, in human no experiments could be performed in vivo. To overcome this limitation, preliminary association studies of the HS1,2A polymorphism with several immunological diseases were performed. In contrast to "mendelian disorders", in which variation in a single gene is both necessary and sufficient to cause disease, the "common disorders" such as autoimmune ones are thought to be determined by the combined effects of many different DNA variants interacting with environmental factors. Numerous studies have demonstrated the association of the HLA haplotypes and non-MHC polymorphisms with several immunological disorders. Despite this, only few association studies have investigated the non-coding regions (i.e. the regulative regions) as genetic factors increasing the risk to the predisposition to common diseases. Preliminary studies demonstrate the correlation of HS1,2A allele 2 of the RR-1 with the predisposition to celiac disease (Frezza *et al*, 2004). One aim of this thesis is to extend the analysis to several immunological diseases characterized by alteration of immunoglobulin production and of B cell functions. We found that allele 2 frequency increased in several independent cohorts of patients affected by different autoimmune diseases (*dermatitis herpetiformis*; rheumatoid arthritis, systemic lupus erythematosus, crohn's disease and psoriasis) in respect to the control group (Cianci *et al*, 2008; Tolusso *et al*, 2009; Mattioli *et al*, 2007). A preliminary study showed that the dysregulation of IgA production in IgA nephropathy patients is associated with HS1,2A polymorphism, and in particular one allele is significantly correlated with a milder form of the disease, whereas another allele is associated with severe evolution leading to renal failure (Aupetit *et al*, 2000). Here we found that HS1,2A allele 2 associates strongly with the acute form of a given autoimmune disease in respect to the milder form. For example, in a cohort of LSRA patients allele 2 frequency increased quite significantly in respect to the ERA patients (p=0.053) (Tolusso *et al*, 2009). Similarly, allele 2 was present at a frequency of 0.590 in a cohort of patients affected by plaque psoriasis, whereas the frequency increased to 0.750 in the cohort of patients with psoriatic arthritis, which is the acute form of psoriasis disease (Cianci *et al*, 2008). In both cases allele 2A frequency increased significantly in respect to

the control (0.392; $p<0.001$). However, the role of HS1,2A in the predisposition to immunological diseases is not limited to the autoimmune ones. We found a statistical association of HS1,2A allele frequencies with selective IgA deficiency, which is an immune related disease characterized by the decrease of IgA serum levels. In particular, in the IgA deficiency group allele 1 frequency increases (0.520) and allele 2 decreases (0.367) respectively ($p=0.006$) and also in respect to the control group (allele 1: 0.331; allele 2: 0.528; $p<0.05$) (Mattioli *et al*, 2009). On the other hand, the analysis of HS1,2A allelic frequency distribution in an autoimmune disease (CD) in respect to a disease with a non-immunological aetiology (RCU) demonstrates that the increase of allele 2 frequency is specific for immunological diseases. The association of the HS1,2A polymorphism with independent and different cohorts of patients reinforces this finding. HS1,2A allele 2 could be considered a genetic variant increasing the risk of developing immunological disorders. In autoimmune diseases, the increase of allele 2 frequency seems to be associated with the up-regulation of "self" antibodies' serum levels. On the other hand, in selective IgA deficiency, the increase of allele 1 frequency correlates with the down-regulation of immunoglobulin production.

1.3 Correlation of HS1,2A allele frequencies with the alteration of serum Ig concentration.

Previous studies have shown that the HS1,2A genotype strongly associates with the circulating levels of IgA1 and IgA2 (Denizot *et al*, 2001). The association of HS1,2A polymorphism with the alteration of immunoglobulin serum levels was further investigated in two independent cohorts of subjects. In both samples, the group with Ig levels below the normal range showed an increased frequency of allele 1 and a decreased frequency of allele 2; on the other hand, the group with Ig levels above the norm showed a decreased frequency of allele 1 and an increased frequency of the allele 2 (Frezza *et al*, 2009; Mattioli *et al*, 2007). EMSA experiments have demonstrated that two different protein complexes bind HS1,2A 1 and 2 alleles (Frezza *et al*, 2009). This means that HS1,2A polymorphism could regulate transcription factor binding to the enhancer itself, so influencing the activity of the RR-1 in the regulation of germline transcription and immunoglobulin production.

1.4 Role of HS1,2A polymorphism in the progression of HIV and HCV infections

In viral pathologies such as AIDS and Hepatitis C the host immunity system plays an important role in the progression of the disease. During viral infection, viral factors interact with the host factors. Among the host factors, the genetic component seems to have an important role in the individual response to the virus. In particular, the genetic factors influencing the immunological response to the infection could be essential for the progression of the disease. From literature we know that several genetic factors influence the progression to AIDS (Carrington *et al*, 1999; Bugeja *et al*, 2004). Among them, HS1,2A polymorphism seemed to correlate to AIDS progression in a paediatric cohort infected with a monophyletic HIV-strain (unpublished data). This study showed that the homozygous condition for allele 2 increased in a Fast Progressor cohort in respect to a Long Term Non-Progressor cohort of Libyan infected patients. Taking this as a starting-point, one aim of the project focused on the involvement of HS1,2A polymorphism in influencing AIDS progression in a cohort of African individuals affected by HIV. Our findings show that the increase of allele 3A frequency seems to correlate with the decrease of CD4+ T cell levels. We know that CD4+ levels decreased with the progression of the disease (Bourinbaiar and Abulafia-Lapid, 2005). Therefore the HS1,2A allele 3 seems to be associated with the progression of HIV infection. Previous studies on healthy populations worldwide have shown that allele 3 possess the highest frequency in the African continent, while allele 2 frequency decreases in Africans in respect to Europeans and Asians. Probably allele 3 in Africans carries the same function of the corresponding European allele 2. Finally, HS1,2A polymorphism's association with Hepatitis C progression in an Italian cohort of HCV infected subjects was investigated. In this case, we observe an increase of allele 2A and a decrease of allele 1 in the patients chronically infected in contrast with the ones that undergo spontaneous virus clearance. Therefore HS1,2A polymorphism could be involved in the individual response to the Hepatitis C virus, even if the results even if the difference didn't reach the statistical significance perhaps due to the small number of samples.

2. The genetic variants of the regions surrounding the HS1,2A and HS3 enhancers.

As the human 3'RR-1 spans over 30 kb it is likely that other genetic components beyond the HS1,2A enhancer are involved in the regulation of the function of the region. The final aim of the project focuses on the study of the region lying between the enhancers HS3 and HS1,2A; this region spans 5.4 kb and is amplified specifically by a selective PCR for RR-1. The Hap Map project has found only one SNP within the 5.4 kb region, lying at the end of U6 (87166675 of NT_026437.11) (data from HapMap database). Therefore we investigated to other possible polymorphisms of two regions surrounding the enhancers HS3 and HS1,2A. By using genome resequencing we found in total 22 SNPs which are in strong linkage disequilibrium with each other. A control group and a group of psoriatic patients carrying the 1/1 and 2/2 genotypes were selected. Haplotypes linked specifically with one allele (1) or the other (2) were identified. In particular, the haplotypes (8, 9, 12) showing the highest number of mutated nucleotide are associated with allele 2; whereas the haplotype (1) with the highest number of wild type form (referred to NT_026437.11) is associated with allele 1. Moreover, we found a directly proportional correlation between the number of mutated SNPs and the frequency of the 8, 9, 12 haplotypes. The results suggest that other genetic variants as well as the HS1,2A polymorphism could also contribute to genetic predisposition to immune diseases or other diseases; the same variants could also be involved in the 3'RR-1 function. In particular, haplotype 1 seems to contribute to the predisposition to selective IgA deficiency and could associate with the decrease of antibodies' serum levels. On the other hand, haplotypes 8, 9 and 12 seem to contribute to the predisposition to immunological diseases, to viral infection progression, and could also correlate with the over-expression of serum immunoglobulin levels. Therefore, a specific HS1,2A allele linked to specific haplotypes of the 5.4 kb region could influence the activity of 3'RR-1 in Ig germline transcription, in CSR stimulation and in Ig production (Mattioli *et al*, 2007; Frezza *et al*, 2008). Genetic variants such as polymorphisms and haplotypes could affect epigenetic change or transcription factors binding, so influencing the activity of the 3' regulatory region. The CpG plot analysis of haplotype 1 and the haplotype 12 shows several differences in hypothetical methylation sites. We can hypothesize that the decrease of methylated sites determines an open chromatin conformation leading to an increase of the activity of the regulatory region in the stimulation of germline transcription.

This condition could be associated, for example, with the increased levels of serum antibodies. Therefore, according to this model, haplotype 12 could be associated with a decreased level of the methylated sites. Further experiments may clarify our findings. The analysis of the hypothetical transcription factor binding sites demonstrates that the different haplotypes could bind different transcription factors. For example the wild type form of SNP 8750 present in haplotype 1 shows a hypothetical binding site for YY1, which is lost in the alternative form present in haplotypes 8, 9 and 12. From literature it is known that YY1 is a specific activator for HS3 and Eμ enhancers (Gordon *et al*, 2003). In addition, SNP 8750 lies very near to enhancer HS3. Thus we can hypothesize that YY1 binds haplotypes 8, 9, or 12, leading to an increased activity of the enhancer HS3 of the 3'RR-1. Haplotypes 8, 9, and 12 could associate with the increase of allele 2 because they contribute in different ways to the same goal - in this case, the stimulation of the RR-1 activity. In future, EMSA experiments with different haplotypes could shed light on the mechanism influencing the activity of the regulatory region. Moreover, genome- resequencing of the entire 3'RR-1 could be important for the identification of other genetic variants which participate in the process.

We found that a specific HS1,2A allele is associated with specific haplotypes. On the other hand, different haplotypes could be associated with the same HS1,2A allele. The increase of allele 2 in our group of patients respect to the control could not in itself explain the different activity of the 3' regulatory region. Probably other genetic variants determine the differences between these patients in respect to the healthy subjects. We found that haplotypes 9 is associated with the control group carrying allele 2, whereas haplotype 12 is associated with psoriatic patients carrying allele 2. In this case the difference turned out to be not significant, perhaps due to the small number of sample. Anyway, it could be interesting especially given the structure of the two haplotypes. Haplotypes 9 and 12 can be distinguished only by 1 SNP (87167265 of NT_026437.11) which is present in the wild type form in haplotype 9 and in the mutated form in haplotype 12. SNP 7265 may account for the different role of the two haplotypes (9 and 12) in the regulation of the regulative region functions by means of the differential binding of the transcription factor ADR1. Further studies on the role of ADR1 in B cell functions could explain the significance of this finding. (Frezza *et al*, 2008).

CONCLUSION

In this thesis we demonstrated that:

1. The Italian population is rather homogeneous for the HS1,2A locus. AMOVA analysis confirmed the results. Nevertheless, natural barriers impair genetic flux in Sardinia in contrast to the homogeneity generally found in the rest of Italy. A similar effect is also found in Spain's Sierra de Gredos region.

2. HS1,2A polymorphism associates with the development of several immunological diseases. Allele 2 frequency increased in several independent cohorts of patients affected by different autoimmune diseases (*dermatitis herpetiformis*; rheumatoid arthritis, systemic lupus erythematosus, crohn's disease and psoriasis) in respect to the control group. Moreover, HS1,2A allele 2 associates strongly with the acute form of a given autoimmune disease in respect to the milder form. On the other hand, allele 1 frequency increased in selective IgA deficiency in respect to the control group. Therefore HS1,2A allele 2 could be considered a genetic variant increasing the risk of developing several immunological disorders, whereas allele 1 increases the risk to the predisposition to IgA deficiency disease.

3. HS1,2A polymorphism correlates with the alteration of immunoglobulin serum levels in two independent and very different cohorts. In both samples, the group with immunoglobulin serum levels above the norm showed an increased frequency of the allele 2, whereas the group with immunoglobulin serum levels below the normal range showed an increased frequency of allele 1.

4. HS1,2A polymorphism influence the host's immune response to HIV and HCV infections. HS1,2A polymorphism seems to associate with the progression of HIV infection. Allele 3 frequency increase correlates with the decrease of CD4+ T cell levels, influencing the

alteration of the individual immune response during HIV infection leading to AIDS progression. Moreover, HS1,2A polymorphism could be involved in the individual response to the HCV virus. Allele 2 frequency increase seems to be associated with chronic infection, albeit in a not significant way.

5. Other genetic variants as well as the HS1,2A polymorphism contribute to genetic predisposition to immune diseases or other diseases. We found 22 SNPs in strong linkage disequilibrium with each other in the regions surrounding the enhancers HS3 and HS1,2A. The haplotypes identified associate significantly with HS1,2A alleles. In particular, the haplotypes showing the highest number of alternative nucleotides are associated with allele 2; whereas the haplotype with the highest number of wild type nucleotides associates with allele 1. CpG plot analysis of the haplotypes shows several differences in hypothetical methylation sites. On the other hand, different haplotypes could be associated with the same HS1,2A allele. We found that haplotypes 9 is associated with a control group carrying allele 2, whereas haplotype 12 is associated a group of psoriatic patients carrying allele 2; even if the differences have proved to be not significant.

In conclusion, in this thesis we showed that HS1,2A polymorphism correlates with the development of several immunological diseases, with the host's immune response to viral infection, with antibody serum levels. This suggests that different alleles could influence the activity of the 3'RR-1. EMSA experiments have shown that different alleles bind different transcription factor complexes to the enhancers. This could be a possible mechanism by which HS1,2A polymorphism could influence the activity of the 3' RR-1. Moreover, we discovered the haplotypes of a 5.4 kb region within the 3'RR-1 and we found a correlation of these haplotypes with HS1,2A alleles. Therefore multiple genetic factors could influence the activity of the human 3'RR-1 leading to a differential regulation of GT transcription, of class switch recombination and of B cell functions.

REFERENCES

Andersson, T., Fuebring, C., Borrebaeck, C. A.K and Pettersson, S. (1998). Temporal expression of a VH promoter-Cμ transgene linked to the IgH HS1,2 enhancer. *Mol Immunol* 36:19-29.

Armuzzi A., Ahmad T., Link K-l., De Silva A., Cullen S., van Heel D., Orchard T.R., Welsh K.I., Marshall S.E. and Jewell D. P. (2003). Genotype-phenotype analysis of the crohn's disease susceptibility haplotype on chromosome 5q31. *Gut* 52:1133-1139.

Aupetit, C., Drouet, M., Pinaud, E., Denizot, Y., Aldigier, J. C., Bridoux, F., and Cogne, M. (2000). Alleles of the alpha1 immunoglobulin gene 3' enhancer control evolution of IgA nephropathy toward renal failure. *Kidney Int* 58:966-971.

Barbujani, G. and Goldstein, D.B. (2004). Africans and Asians Abroad: Genetic Diversity in Europe. *Annu Rev Genomics Hum Genet* 5:119-150.

Barbujani, G., Stenico, M., Excoffier, L. and Nigro, L. (1996). Mitochondrial DNA sequence variation across linguistic and geographic boundaries in Italy. *Hum Biol* 68:201-215.

Begovich, A.B., Carlton, V.E., Honigberg, L.A., Schrodi, S.J., Chokkalingam, A.P. and Alexander, H.C. (2004). A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (PTPN22) is associated with rheumatoid arthritis. *Am J Hum Genet* 75:330-337.

Bourinbaiar A.S. and Abulafia-Lapid R. (2005). Autoimmunity, alloimmunization and immunotherapy of AIDS. *Autoimmun Rev* 4:403-409.

Budowle B., Smith J., Moretti T. and Di Zinno J. (2000). DNA Typing Protocols: Molecular Biology and Forensic Analysis. Eaton: Eaton Publishing.

Bugeja, M.J., Booth, D.R., Bennetts, B.H., Guerin, J., Kaldor, J.M., Stewart, G.J. and Australian Long-Term Non-Progressor Study Group (2004).

Analysis of the CCL3-L1 gene for association with HIV-1 susceptibility and disease progression. *AIDS* 18:1069-1071.

Carrington, M., Nelson, G.W., Martin, M.P., Kissner, T., Vlahov, D., Goedert, J.J., Kaslow, R., Buchbinder, S., Hoots, K. and O'Brien, S.J. (1999). HLA and HIV-1: heterozygote advantage and B*35-Cw*04 disadvantage. *Science* 283:1748-1752.

Casartelli, N., Di Matteo, G., Argentini, C., Cancrini, C., Bernardi, S., Castelli, G., Scarlatti, G., Plebani, A., Rossi, P. and Doria, M. (2003). Structural defects and variations in the HIV-1 nef gene from rapid, slow and non-progressor children. *AIDS* 17:1291-1301.

Centers for Disease Control and Prevention. (1994). Revised classification system for human immunodeficiency virus infection in children less than 13 years of age. *MMWR Morb Mortal Wkly Rep* 43.

Chauveau, C., Pinaud, E. and Cogne, M. (1998). Synergies between regulatory elements of the immunoglobulin heavy chain locus and its palindromic 3' locus control region. *Eur J Immunol* 28:3048-3056.

Chaudhuri, J. and Alt, F.W. (2004). Class-switch recombination: interplay of transcription, DNA deamination and DNA repair. *Nat Rev Immunol* 4:541-552.

Cianci, R., Giambra, V., Mattioli, C., Esposito, M., Cammarota, G., Scibilia, G., Magazzù, G., Orlando, A., Sandri, G., Bianchi, L., Gasbarrini, G.B., Pandolci F. and Frezza, D. (2008).Increased frequency of Ig Heavy Chain HS1,2-A Enhancer *2 Allele in Dermatitis Herpetiformis, Plaque Psoriasis and Psoriatic Arthritis. *J Invest Dermatol* 128:1920-1924.

De Milito, A. (2004). B lymphocyte dysfunctions in HIV infection. *Curr HIV Res* 2:11-21.

Denizot, Y., Pinaud, E., Aupetit, C., Le Morvan, C., Magnoux, E., Aldigier, J. C. and Cogne, M. (2001). Polymorphism of the human α1 immunoglobulin gene 3' enhancer HS1,2 and its relation to gene expression. *Immunology* 103:35-40.

Diamant, E. and Melamed, D. (2004). Class switch recombination in B lymphopoiesis: a potential pathway for B cell autoimmunity. *Autoimmunity* 3:464-469.

Ernst, P. and Smale, ST. (1995). Combinatorial regulation of transcription II: The immunoglobulin mu heavy chain gene. *Immunity* 2:427-438.

Excoffier, L., Laval, G. and Shneider, S. (2005). Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evol Bioinform Online* 1:47-50.

Feld, J.J., and Liang, T.J. (2006). Hepatitis C-Identifying Patients with Progressive Liver Injury. *Hepatology* 43:194-206.

Frezza, D., Giambra, V., Cianci, R., Fruscalzo, A., Giufrè, M., Cammarota, G., Martìnez-Labarga, C., Rickards, O., Scibilia, G., Sferlazzas, C., Bartolozzi, F., Starnino, S., Magazzù, G., Gasbarrini, G. B. and Pandolfi, F. (2004). Increased frequency of the immunoglobulin enhancer HS1,2 allele 2 in coeliac disease. *Scand J Gastroenterol* 39:1083-1087.

Frezza, D., Giambra, V., Mattioli, C., Piccoli, K., Massud, R., Siracusano, A., DI Giannantonio M., Birshtein, B.K. and Rubino, I. A. (2009). Allelic frequencies of 3' Ig Heavy Chain Locus enhancer HS1,2-A associated with Ig levels in patients with schizophrenia. *Int J Immun and Farmacol* IN PRESS.

Frezza, D., Esposito, M., Mattioli, C., Vaccari, G., Giambra, V., Cimenti, S. and Bianchi, L. (2008). Haplotypes associated with alleles of the Ig heavy chain enhancer HS1,2 of the 3'Regulatory Region (3'RR) in patients with Psoriatic Arthritis. Psoriasis - From gene to clinic. December 4/6, London. UK.

Frezza, D., Giambra, V., Tolusso, B., De Santis, M., Bosello, S., Vattoni, S., Trioso, G., Valentini, G. and Ferraccioli, G. (2007). Polimorphism of immunoglobulin enhancer element HS1,2A: allele 2 associated with Systemic Sclerosis. Comparision with HLA-DR and DQ alleles. *Ann Rheum Dis* 66:1210-1215.

Garrett, F.E., Emelyanov, A.V., Sepulveda, M.A., Flanagan, P., Volpi, S., Li, F., Loukinov, D., Eckhardt, L.A., Lobanenkov, V.V. and Birshtein, BK. (2005). 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* 25:1511-1525.

Giambra, V., Fruscalzo, A., Giufrè, M., Martinez-Labarga, C., Favaro, M., Rocchi, M. and Frezza, D. (2005). Evolution of human IgH3'EC duplicated structures: both enhancers HS1,2 are polymorphic with the variation of transcription factor's consensus sites. *Gene* 346:105-114.

Giambra, V., Martínez-Labarga, C., Giufré, M., Modiano, D., Simporé, J., Gisladottir, B. K., Francavilla, R., Zhelezova, G., Kilic, S. S., Crawford, M., Biondi, G., Rickards, O. and Frezza, D. (2006). Immunoglobulin enhancer HS1,2 polymorphism: a new powerful anthropogenetic marker. *Ann Hum Genet* 70:946-950.

Giambra, V., Volpi, S., Emelyanov, A.V., Pflugh, D., Bothwell, A.L., Norio, P., Fan, Y., Ju, Z., Skoultchi, AI., Hardy, R.R., Frezza, D. and Birshtein, B.K. (2008). Pax5 and linker histone H1 coordinate DNA methylation and histone modifications in the 3' regulatory region of the immunoglobulin heavy chain locus. *Mol Cell Biol* 28:6123-6133.

Giannini, S. L., Singh, M., Calvo, C. F., Ding, G. and Birshtein, B. K. (1993). DNA regions flanking the mouse Ig 3'alpha enhancer are differentially methylated and DNAse I hypersensitive during B cell differentiation. *J Immunol* 150:1772-1780.

Goldstein, D.B. and Cavalleri, G.L. (2005). Understanding human diversity. *Nature* 437:1241-1242.

Gordon, S. J., Saleque, S. and Birshtein, B.K. (2003). Yin Yang 1 is a lipopolysaccharide-inducible activator of the murine 3' Igh enhancer, HS3. *J Immunol* 170:5549-5557.

Guglielmi, L., Truffinet, V., Magnoux, E., Cognè, M. and Denizot, Y. (2004). 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* 94:77-81.

Hammarstrom, P.Q., Zhao, Y. and Hammarstrom, L. (2007). Class Switch Recombination: a comparison between mouse and human. *Adv Immunol* 93:1-57.

Harindranath, N., Mills, F.C., Mitchell, M., Meindl, A. and Max, E.E. (1998) The human elk-1 gene family: the functional gene and two processed pseudogenes embedded in the IgH locus. *Gene* 221:215-224.

Hervonen, K., Karell, K., Holopainen, P., Collin, P., Partanen, J. and Reunala, T. (2000). Concordance of dermatitis herpetiformis and celiac disease in monozygous twins. *J Invest Dermatol* 115:990-993.

Hon, P., Bruce, I.N., Silman, A., Symmons, D., Newman, B., Young, H., GriffitHS, C.E.M., John, S., Worthington, J. and Barton, A. (2005) Evidence for common genetic control pathways of inflammation for crohn's disease and psoriatic arthritis. *Arthritis Rheum* 52:3596-3602.

Hu, Y., Pan, Q., Pardali, E., Mills, F. C., Bernstein, R. M., Max, E. E., Sideras, P. and Hammarström, L. (2000). Regulation of germline promoters by the two human Ig heavy chain 3' α enhancers. *J Immunol* 164:6380-6386.

Huizinga, T.W., Amos, C.I., van der Helm-van Mil, A.H., Cen, W., van Gaalen, F.A. and Jawaheer, D. (2005) Refining the complex rheumatoid arthritis phenotype based on specificity f the HLA-DRB1 shared epitope for antibodies to citrullinated proteins. *Arthritis Rheum* 52:3433-3438.

Johansson, M., Alestig, L., Hallmans, G. and Rantapaa-Dahlqvist, S. (2006). PTPN22 polymorphism and anti-cyclic citrullinated peptide antibodies in combination strongly predicts future onset of rheumatoid arthritis and has a specificity of 100% for the disease. *Arthritis Res Ther* 8:R19.

Jones, A.L., Mowry, B.J., Pender, M.P. and Greer, J.M. (2005). Immune dysregulation and self-reactivity in schizophrenia: Do some cases of schizophrenia have an autoimmune basis? *Immunol Cell Biol* 83:9-17.

Ju, Z., Volpi, S. A., Hassan, R., Martinez, N., Giannini, S. L., Gold, T. and Birshtein, B. K. (2007). Evidence for physical interaction between the immunoglobulin heavy chain variable region and the 3' regulatory region. *J Biol Chem* 282:35169-35178.

Kallel Sellami, M., Cherif, F., Zitouni, M., Laadhar, L., Mokni, M. and Hajlaoui, K. (2006). Subepidermal bullous dermatosis with IgA antibodies against 200- and 280-kDa epidermal antigens after initiation of ultraviolet B therapy for psoriasis. *Br J Dermatol* 154:193-194.

Karell, K., Korponay-Szabo, I., Szalai, Z., Holopainen, P., Mustalahti, K. and Collin, P. (2002) Genetic dissection between coeliac disease and dermatitis herpetiformis in sib pairs. *Ann Hum Genet* 66:387-392.

Kawamura, S., Saitou, N. and Ueda, S. (1992). Concerted evolution of the primate immunoglobulin alpha-gene through gene conversion. *J Biol Chem* 267:7359-7367.

Kim, E.C., Edmonston, C.R., Wu, X., Schaffer, A. and Casali, P. (2004). The HoxC4 Homeodomain Protein Mediates Activation of the Immunoglobulin Heavy Chain 3' HS1,2 Enhancer in Human B cells. *J Biol Chem* 279:42265-42269.

Kingsmore, S.F., Lindquist, I.E., Mudge, J., Gessler, D. D. and Beavs, W.D. (2008). Genome-wide association studies: progress and potential for drug discovery and development. *Nat Rev Drug Discov* 7:221-230.

Kruglyak L. (2008). The road to genome-wide association studies. *Nat Rev Genet* 9:314-318.

Lapeyre, H., Courville, P., Francois, A., Musette, P., Le Roy, F. and Joly, P. (2006). IgA necrotic renal glomerular vasculitis in pustular psoriasis. *Ann Dermatol Venereol* 133:472-473.

Laurencikiene, J., Deveikaite, V. and Severinson, E. (2001). HS1,2 enhancer regulation of germline ε and γ2b promoters in murine B lymphocytes: Evidence for specific promoter enhancer interactions. *J Immunol* 167:3257-3265.

Laurencikiene, J., Tamosiunas, V. and Severinson, E. (2007). Regulation of ε germline transcription and switch region mutations by IgH locus 3' enhancers in transgenic mice. *Blood* 109:159-167.

Lee, J.E., Choi, J. H., Lee, J.H. and Lee, M.G. (2005). Gene SNPs and mutations in clinical genetic testing: haplotype-based testing and analysis. *Mutat Res* 573:195-204.

Li, Q., Peterson, K.R., Fang, X. and Stamatoyannopoulos, G. (2002) Locus Control Regions. *Blood* 100:3077-3086.

Li Y., Chang M., Schrodi S., Callis-Duffin C.P., Matsunami N., Civello D., Bui N., Catanese J.J., Leppert M.F., Krueger G.G. and Begovich A.B. (2008). The 5q31 variants associated with psoriasis and crohn's disease are distinct. *Hum Mol Genet* 17:2978-2985.

Lie, B.A. and Thorsby, E. (2005). Several genes in the extended human MHC contribute to predisposition to autoimmune diseases. *Curr Opin Immunol* 17:526-531.

Lieberson, R., Giannini, S. L., Birshtein, B. K. and Eckhardt, L. A. (1991). An enhancer at the 3' end of the mouse immmunoglobulin heavy chain locus. *Nucleic Acids Res* 19:933-937.

Linderson, Y., Ebehard, D., Malin, S., Johansson, A., Busslinger, M. and Pettersson, S. (2004). Corecruitment of the Grg4 repressor by PU.1 is critical for Pax5-mediated repression of B-cell-specific genes. *EMBO Rep* 5:291-296.

Linderson, Y., French, N.S., Neurath, M.F. and Pettersson, S. (2001). Context-dependent Pax-5 repression of a PU.1/NF-kB regulated reporter gene in B lineage cells. *Gene* 262:107-114.

Lloyd, A.R., Jagger, E., Post, J.J.,Crooks, Lee-Ann., Rawlinson, W.D., Hahn, Y. S. and Ffench R.A. (2007). Host and viral factors in the immunopathogenesis of primary hepatitis C virus infection. *Immunol Cell Biol* 85:24-32.

Madison, L. and Groudine, M. (1994). 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* 8:2212-2226.

MacGregor, A., Ollier, W., Thomson, W., Jawaheer, D. and Silman, A. (1995). HLA-DRB1*0401/0404 genotype and rheumatoid arthritis: increased association in men, young age at onset, and disease severity. *J Rheumatol* 22:1032-1036.

MacGregor, A.J., Snieder, H., Rugby, A.S., Koskenvuo, M., Kaprio, J. and Aho, K. (2000) Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. *Arthritis Rheum* 43: 30-37.

Manis, J.P., Ming, T. and Alt, F.W. (2000). Mechanism of control of class switch recombination. *Trends Immunol* 23:31-39.

Martin, F. and Dixit, V.M. (2005). Unraveling TACIt functions. *Nat Genet* 37:793-794.

Mattioli, C., Cianci, R., Giambra, V., Tampella, G., Kilic, S.S., Pandolfi, F., Plebani, A. and Frezza, D. (2009). Increase of HS1,2 allele*1 enhancer frequency in selective IgA deficiency. *Ann Hum Genet* SUBMITTED.

Mattioli, C., Giambra, V., Esposito, M., Sali, M., Santangelo, R., Bianchi, L., Pescatori, M. and Frezza, D. (2007). Analysis of the haplotypes of the human 3'Ig Heavy chain Regulatory Region. FISV 2007. September 26/29, Riva del Garda, Italy.

Mattioli, C., Giambra, V., Tomei, S., Vaccari, G. and Frezza, D. (2008).The alleles of the human IgH3' enhancer HS1,2 have different frequencies in world populations: genetic-immunology studies. FISV 2008. September 24/27, Riva del Garda, Italy.

Mattioli, C., Tolusso, B., Giambra, V., De Santis M., Borsello, S., Gremese, E., Peluso, G., Esposito, M., Piccoli, K., Cianci, R., Bianchi, L., Massoud, R., Ferraccioli, G. and Frezza, D. (2007). Ig Heavy Chain 3'HS1,2 – An Enhancer Polymorphism Related to Immunological Diseases and

Dysregulation of Levels of Serum Immunoglobulin. ASEATTA. November 16/19, Freemantle, Australia.

Michaelson, J. S., Singh, M., Snapper, C. M., Sha, W. C., Baltimore, D. and Birshtein, B. K. (1996). Regulation of 3' IgH enhancers by a common set of factors, including kappa B-binding proteins. *J Immunol* 156: 2828-2839.

Mills, F. C., Harindranath, N., Mitchell, M. and Max, E.E. (1997). Enhancer complexes located downstream of both human immunoglobulin Cα genes. *J Exp Med* 186:845-858.

Ong, J., Stevens, S., Roeder, R. G. and Eckhardt, L. A. (1998). 3' IgH enhancer elements shift synergistic interactions during B cell development. *J Immunol* 160:4896-4903.

Oruc, Z., Boumédiène, A., Le Bert, M. and Khamlichi, A.A. (2007). Replacement of Iγ3 germ-line promoter by Iγ1 inhibits class-switch recombination to IgG3. *PNAS* 104:20484-20489.

Pinaud, E., Khamlichi, A. A., Le Morvan, C., Drouet, M., Nalesso, V., Le Bert, M. and Cogne, M. (2001). Localization of the 3' IgH locus elements that effect long-distance regulation of class switch recombination. *Immunity* 15:187-199.

Qin, X. and Tang, H. (2006). Differential regulation of chromatin structure of the murine 3' IgH enhancer and IgG2b germline promoter in response to lipopolysaccharide and CD40 signaling. *Mol Immunol* 43:1211–1220.

Rodriguez, B., Sethi, A.K., Cheruvu, V.K., Mackay, W., Bosch, R.J., Kitahata, M., Boswell, S.L., Mathews, W.C., Bangsberg, D.R., Martin, J., Whalen, C.C., Sieg, S., Yadavalli, S., Deeks, S.G., Lederman, M.M. (2006). Predictive value of plasma HIV RNA level on rate of CD4 T-cell decline in untreated HIV infection. *JAMA* 296:1498-1506.

Root-Bernstein, R. and Rallo, A. (2004). Antigenic complementarity resulting in idiotype-antiidiotype immune complexes: possible contributor to AIDS pathogenesis and autoimmunity. *Autoimmunity* 37:203-210.

Rothermundt, M., Arolt, V. and Bayer, T.A. (2001). Review of immunological and immunopathological findings in schizophrenia. *Brain Behav & Immunity* 15:319-339.

Saito, T., Ji, G., Shinzawa, H., Okumoto, K., Hattori, E., Adachi, T., Takeda, T., Sugahara, K., Ito, J.I., Watanabe, H., Saito, K., Togashi, H., Ishii, K., Matsuura, T., Inageda, K., Muramatsu, M. and Kawata, (2004). Genetic variations in humans associated with differences in the course of hepatitis C. *Biochem Biophys Res* 317:335-341.

Saksena, N.K., Ge, Y.C., Wang, B., Xiang, S.H., Dwyer, D.E., Randle, C., Palasanthiran, P., Ziegler, J. and Cunningham, A.L. (1996). An HIV-1 infected long-term non-progressor (LTNP): molecular analysis of HIV-1 strains in the vpr and nef genes. *Ann Acad Med* 25:848-854.

Saleque, S., Singh, M., Little, R. D., Giannini, S. L.,Michaelson, J. S. and Birshtein, B. K. (1997). Dyad symmetry within the mouse 3' IgH regulatory region includes two virtually identical enhancers (C alpha 3'E and HS3). *J Immunol* 158:4780-4787.

Salzer, U., Chapel, H.M., Webster, A.D., Pan- Hammarstrom, Q., Schmitt-Graeff, A., Schlesier, M., Peter, H.H., Rockstroh, J.K., Schneider, P., Schaffer, A.A., Hammarstrom, L. and Grimbacher, B. (2005). Mutations in TNFRSF13B encoding TACI are associated with common variable immunodeficiency in humans. *Nat Genet* 37:820-828.

Samara, M., Oruc, Z., Dougier, H.L., Essawi, T., Cognè, M. and Khamlichi, A.A. (2006). Germ line transcription of Iγ3 exon in the Ig heavy chain locus. *Int Immunol* 18:581-589.

Sathiyasekaran, M. and Shivbalan, S. (2006). Crohn's disease. *Indian J Pediatr* 73:723-729.

Sepulveda, M.A., Garrett, F.E., Price-Whelan, A. and Brishtein, B.K. (2005). Comparative analysis of human and mouse 3'Igh regulatory regions identifies distinctive structural features. *Mol Immunol* 42:605-615.

Serrano, N.C., Millan, P. and Pàez M.C. (2006). Non-HLA association with autoimmune disease. *Autoimmun Rev* 5:209-214.

Shepshelovich, D. and Shoenfeld, Y. (2006). Prediction and prevention of autoimmune disease: additional aspects of the mosaic of autoimmunity. *Lupus* 15:183-190.

The International Hap Map Consortium (2005). A haplotype map of the human genome. *Nature* 437:1299-1320.

The International Hap Map Consortium (2007). A second generation human haplotype map over 3.1 million SNPs. *Nature* 449:851-862.

The Wellcome Trust Control Consortium (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447:661-677.

Tolusso, B., Frezza, D., Mattioli, C., Fedele, A.L., Bosello, S., Peluso, G., Giambra, V., Pietrapertosa, D., Morelli, A., Gremese, E., De Santis M. and Ferraccioli, G.F. (2009). Allele *2 of the HS1,2A enhancer of the Ig regulatory region associates with Rheumatoid Arthritis. *Ann Rheum Dis* 68:416-419.

Trachtenberg, E.A. and Erlich, H.A. (2001). A review of the role of the human leukocyte antigen (HLA) system as a host immunogenetic factor influencing HIV transmission and course of infection with progression to AIDS. *In: Korber BT, Brander C, Haynes BF, Koup R, Kuiken C, Moore JP, Walker BD, Watkins D, eds. HIV Molecular Immunology Database* 1–60.

Vorechovský, I., Webster, A.D.B., Plebani, A. and Hammastrom, L. (1999). Genetic Linkage of IgA Deficiency to the Major Histocompatibility Complex: Evidence for Allele Segregation Distortion, Parent-of-Origin Penetrance Differences, and the Role of Anti-IgA Antibodies in Disease Predisposition. *Am J Hum Genet* 64:1096–1109.

Wang Leo, D. and Clark Marcus, R. (2003). B-cell antigen-receptor signalling in lymphocyte development. *Immunol* 110:411-420.

West, G.A. and Fraser, P. (2005). Remote control of gene transcription. *Hum Mol Genet* 14:101-111.

Woo, W.K., McMillan, S.A., Watson, R.G., McCluggage, W.G., Sloan, J.M. and McMillan, J.C. (2004). Coeliac disease-associated antibodies correlate with psoriasis activity. *Br J Dermatol* 151:891-894.

Zadrazil, J., Tichy, T., Horak, P., Nikorjakova, I., Zima, P. and Krejci, K. (2006). IgA nephropathy associated with psoriasis vulgaris: a contribution to the entity of psoriatic nephropathy. *J Nephrol* 19(3):382-386.

Zandman-Goddard, G. and Shoenfeld, Y. (2002). HIV and autoimmunity. *Autoimmun Rev* 1:329-337.

Zelazowski, P., Shen, Y. and Snapper, C.M. (2000). NF-κb/p50 and NF-κb/c-Rel differentially regulate the activity of the 3'αE-HS1,2 enhancer in normal murine B cells in an activation-dependent manner. *Int Immunol* 12:1167-1172.

Zhang, B., Alaie-Petrillo, A., Kon, M., Li, F. and Eckhardt, L.A. (2007). Transcription of a productively rearranged Ig VDJC alpha does not require the presence of HS4 in the IgH 3' regulatory region. *J Immunol* 178:6297-6306.

Zhang, J., Jacobi, A.M., Wang, T. and Diamond, B. (2008). Pathogenic autoantibodies in systemic lupus erythematosus are derived from both self-reactive and non-self-reactive B cells. *Mol Med* 14:657-681.

Zhang K. (2003). Accessibility control and machinery of immunoglobulin class switch recombination. *J Leukoc Biol* 73:32-332.