Running titles:

Identification of genetic markers correlated to Psoriasis.

TITLES:

Identification of haplotypes linked to hs1.2 enhancer *2 allele of Immunoglobulin heavy chain locus, and associated with Psoriasis.

Pietro D'Addabbo: pietro.daddabbo@uniba.it

Eliseo Serone: serone@negrisud.it

MDr Maria Esposito: esposito@uniroma2.it

Vincenzo Giambra: vgiambra@bccrc.ca

Dr Gabriele Vaccari: gvaccari@iss.it

Vincenza Colonna: vincenza.colonna@igb.cnr.it

Dr Cesare Gargioli: cesare.gargioli@uniroma2.it

Prof Domenico Frezza: frezza@uniroma2.it

Prof Luca Bianchi: luca.bianchi@uniroma2.it

ABSTRACT

The human Ig heavy-chain genes are organized in two duplicated blocks, with Regulatory Regions at the 3' of both alpha1 and alpha2 constant genes, respectively 3'RR1 and 3'RR2. Enhancers regions were identified in the 3'RRs by DNAse sensitivity assays, phylogenetic analysis and mutational tests. Other evidences suggested the presence of peculiar three-dimensional conformations of the locus with hs1.2, an enhancer including an internal Variable Number Tandem Repeats (VNTR) polymorphism that may shape itself in a G-quadruplex structure. Of note, population studies demonstrated the association of different alleles of the VNTR in the 3'RR1 with susceptibility to a wide range of autoimmune diseases, including Psoriasis.

Here we investigate the distribution of Single Nucleotide Polymorphisms (SNPs) of 3'RR1 in patients with Psoriasis and healthy controls, and their linkage to the two main hs1.2 alleles, looking for markers of specific haplotypes. The locus was sequenced in samples from a selection of hs1.2-allele-homozygous subjects, and 22 SNPs homozygous in at least one sample was identified. All the SNPs involve nucleotides of possible methylation, and we consider these evidences relevant to further investigate the epigenetic effects and modulation of the three-dimensional structures of the alternative allele of the SNPs on this genomic non-coding region with a cis-acting function.

Our study identified two alternative 9-SNPs haplotypes linked to the two main alleles of 3'RR1 hs1.2 enhancer, which would be useful markers at least in the screening of Psoriasis.

INTRODUCTION

Immune diseases originated by complex interactions among multifactorial events, often involving cis-regulative components located in untranscribed genomic regions.

Psoriasis is defined a clinically heterogeneous systemic T-cell-mediated inflammatory skin disease with inherited susceptibility[1,2]. The whole mechanism triggering the symptoms is not fully understood, despite of recent researches progressions[3]. Thus, characterization of genetic contribution to the etiopathology is necessary to provide novel diagnostic and therapeutic approaches.

The 3' Regulatory Region 1 (3'RR1) is a cis-regulative region embedded in the Ig constant-genes cluster on chromosome 14. Previous studies demonstrated that a portion of the 3'RR1 is associated to Psoriasis as well as to other autoimmune diseases[4,5]. Indeed, evaluating alleles frequency of the enhancer hs1.2 in the 3'RR1, it was found unbalanced in Psoriatic patients[6]. The Ig heavy chain locus and the genes at its 3' boundaries are represented in Figure 1 at different scales, utilizing the human genome release hg38.
In non-primates there is a single 3'RR, that is a key-region in B-cell maturation in mouse[7-9]. In primates there are two paralogous regulatory regions, namely 3'RR1 and 3'RR2, derived by a large duplication peculiar of this order[10]. The two human paralogous 3'RRs maintain a high level of similarity to each other, making not easy to discriminate between the two copies of hs1.2, and clarify their specific contribution to the phenotype. Both hs1.2 copies have a polymorphic region, constituted of an invariable core and a 40-bp monomer VNTR. The whole enhancer is embedded in a palindromic region detected in all mammals, despite of the absence of sequence conservation of the palindromic region itself, strongly suggesting a functional activation achieved by a three-dimensional shape[10].
In present work, we deeply analyse a 3'RR1 portion, to define haplotype markers that would allow a most rapid discrimination among hs1.2 alleles and/or paralogous variants (Fig. 1).

Results and discussion

Since it has been reported an association between Psoriasis and the enhancer hs1.2 of 3'RR1, we genotyped a 5kb section of the 3'RR1, on chromosome 14 (Fig. 1C), in European psoriatic patients (P) and healthy controls (H) that were homozygous for the hs1.2 allele (Table 1). We searched for Single Nucleotide Polymorphisms (SNPs) in linkage with the alleles *1 and *2 of hs1.2, i.e. alleles harbouring one or two copies of the 40 bp monomer, respectively.
We identified 22 SNPs, all but one already known and recorded in dbSNP138. As a note, all the found SNPs involve Guanine or Cytosine as an alternative, so being putative sites of methylation, a mechanism that has been studied in mice and showed spot pattern close to hs1.2[11].

The SNP rs373084296 maps in the hs1.2 enhancer, while the others were equally distributed in two clusters (Fig.1 C). We found that the *1 and *2 alleles of hs1.2 are in linkage with 9 SNPs (grey cells in Table 1), defining two genetic haplotypes, respectively E1W and E2A. It is worth to note that the alternative E1W and E2A haplotypes are equally distributed between our P and H cohorts, showing almost the same rate P/H (29/27 and 34/30, respectively; Table 1). This finding at least apparently contrast the known linkage of the *2 allele with autoimmune diseases, as stated in previous analyses[6], but it is to be considered that we performed a selection to isolate subjects homozygous for *1 or *2 allele, introducing a bias that justifies the observation. At same time E2A1 and E2A2, two of the haplotypes of 22 SNPs linked to the *2 allele, showed strikingly divergent P/H rates (1/21 and 18/7, respectively; $p << 0{,}001$; Table 1) despite the bias. This finding seems to assign a critical role to the SNP rs35216181 in the susceptibility to the Psoriasis.

We note that a number of SNPs were assigned to both the two 3'RRs (i.e. multilocus SNPs, Fig. 1A,B), as a consequence of the high similarity level between the duplicated regions including the 3'RRs and the closest Ig genes. SNPs identification and mapping by aligning next-generation-sequencing reads versus a genome reference failed to disclose the authentic position of these SNPs, therefore tagged as multilocus. Traditional Sanger sequencing instead overcome this problem. Because of these mapping issues the hs1.2 *2 allele locus is excluded from most commercial chips used for genomic screenings[5]. We believe this can be the cause of the lack of association of the hs1.2 *2 allele with autoimmune disease in previous whole-genome association studies.

The 3'RR1 is described in details in Fig 1B. We annotated repeats and palindromic region surrounding the hs1.2 enhancer[10]. The hs1.2 could have a pivotal function in the 3'RR1 role, because of its polymorphism and its position within the palindrome. A regulative activity of the chromatin status can be inferred, by influencing the conformational equilibrium modified by the polymorphism of methylation sites.

Also genome-wide analyses of histone proteins methylation were previously performed by ChIP-seq assays[12]. The mono-methylation of lysine 4 of the H3 histone protein is associated with functional enhancers downstream of transcription starts. Variation of frequency of this kind of methylation was observed in the 3'RRs only in a lymphoblastoid

cell line (GM12878) (Fig 1. A and B), indicating functional activation of the region only in Ig actively-expressing cells.

In humans, a *2 allele increased frequency was associated with a wide range of immune diseases, while *1 allele seems protective for long survival elderly[13]. This phenomenon may be balanced by the presence of higher levels of circulating Ig in children with *2 allele in their first five years, giving a possible temporary advantage[14]. Population studies on alleles distribution in Europe, Asia and Africa showed a very low frequency of *2 allele in Africa[15], that is not surprising considering that in this continent the systemic autoimmune diseases are at reduced incidence, when compared to Europe[16].

All these results evidence the presence of specific markers for hs1.2 *2 allele downstream the enhancer itself, that is possible to link to Psoriasis as well as the enhancer itself. It should be a future direction to confirm these SNPs as direct markers of susceptibility, and look for new ones by a wider analysis on the whole 3'RR1. Complete sequencing of the region in more samples could be a source of valuable information about the players of its regulatory function. Moreover, sequencing also the 3'RR2 in the same individuals will be useful either as control, and to check possible association and interaction between the two copies of the 3'RRs.

Methods

We collected blood samples from psoriatic patients (P) and healthy controls (H). Informed consent was obtained in agreement with the Ethical Committee of polyclinic of Tor Vergata. DNA was extracted by standard procedures, and amplified as previously described[17]. *1 and *2 allele hs1.2-homozygous genotypes were discerned by migration on electrophoretic gel, retaining 29 P and 27 H with *1, and 34 P and 31 H with *2 allele. We sequenced 5404 bp, including both the 3'RR1 hs1.2 and hs3 enhancers, (63P and 58H) to discriminate the haplotypes linked to hs1.2 *1 or *2 alleles (Table 1). A nested-PCR was performed on the selected genomes, and sequences analysed for SNPs content (hg38 -chr14:105696209-105701612).

We used the Big Dye Terminator v3.1 chemistry and a four-capillary ABI Prism 3130 Genetic Analyzer (Applied Biosystems).

Acknowledgments

We thank the Policlinic of Tor Vergata. The study was partially possible thanks to Domenico Frezza's personal resources, and for laboratory space, default supplies and salaries thank to the Italian Ministry of Research and University of Tor Vergata.

Conflict of interest

No conflict of interest to declare

References

1. Chandra A, Ray A, Senapati S, Chatterjee R: Genetic and epigenetic basis of psoriasis pathogenesis. *Mol Immunol* 2015; **64:** 313-323.

2. Novelli L, Chimenti MS, Chiricozzi A, Perricone R: The new era for the treatment of psoriasis and psoriatic arthritis: perspectives and validated strategies. *Autoimmun Rev* 2014; **13:** 64-69.

3. Villanova F, Di Meglio P, Nestle FO: Biomarkers in psoriasis and psoriatic arthritis. *Ann Rheum Dis* 2013; **72 Suppl 2:** ii104-110.

4. Aupetit C, Drouet M, Pinaud E *et al*: Alleles of the alpha1 immunoglobulin gene 3' enhancer control evolution of IgA nephropathy toward renal failure. *Kidney Int* 2000; **58:** 966-971.

5. Frezza D, Tolusso B, Giambra V *et al*: Polymorphisms of the IgH enhancer HS1.2 and risk of systemic lupus erythematosus. *Ann Rheum Dis* 2012; **71:** 1309-1315.

6. Cianci R, Giambra V, Mattioli C *et al*: Increased frequency of Ig heavy-chain HS1,2-A enhancer *2 allele in dermatitis herpetiformis, plaque psoriasis, and psoriatic arthritis. *J Invest Dermatol* 2008; **128:** 1920-1924.

7. Birshtein BK: Epigenetic Regulation of Individual Modules of the immunoglobulin heavy chain locus 3' Regulatory Region. *Front Immunol* 2014; **5:** 163.

8. Saintamand A, Rouaud P, Garot A *et al*: The IgH 3' regulatory region governs mu chain transcription in mature B lymphocytes and the B cell fate. *Oncotarget* 2015; **6:** 4845-4852.

9. Wourms MJ, Sulentic CE: The aryl hydrocarbon receptor regulates an essential transcriptional element in the immunoglobulin heavy chain gene. *Cell Immunol* 2015; **295:** 60-66.

10. D'Addabbo P, Scascitelli M, Giambra V, Rocchi M, Frezza D: Position and sequence conservation in Amniota of polymorphic enhancer HS1.2 within the palindrome of IgH 3'Regulatory Region. *BMC evolutionary biology* 2011; **11:** 71.

11. Giambra V, Volpi S, Emelyanov AV *et al*: Pax5 and linker histone H1 coordinate DNA methylation and histone modifications in the 3' regulatory region of the immunoglobulin heavy chain locus. *Molecular and cellular biology* 2008; **28:** 6123-6133.

12. Ernst J, Kheradpour P, Mikkelsen TS *et al*: Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* 2011; **473:** 43-49.

13. Napolioni V, Serone E, Iacoacci V, Carpi FM, Giambra V, Frezza D: The functional VNTR of IGH enhancer HS1.2 associates with human longevity and interacts with TNFA promoter diplotype in a population of Central Italy. *Gene* 2014; **551:** 201-205.

14. Serone E, Daleno C, Principi N *et al*: The change in Ig regulation from children to adults disconnects the correlation with the 3'RR hs1.2 polymorphism. *BMC Immunol* 2014; **15:** 45.

15. Giambra V, Martinez-Labarga C, Giufre M *et al*: Immunoglobulin enhancer HS1,2 polymorphism: a new powerful anthropogenetic marker. *Annals of human genetics* 2006; **70:** 946-950.

16. Chandran V, Raychaudhuri SP: Geoepidemiology and environmental factors of psoriasis and psoriatic arthritis. *J Autoimmun* 2010; **34:** J314-321.

17. Frezza D, Giambra V, Cianci R *et al*: Increased frequency of the immunoglobulin enhancer HS1,2 allele 2 in coeliac disease. *Scand J Gastroenterol* 2004; **39:** 1083-1087.

Table Legend

Table1: Sequenced region and detected SNPs

Code and position (hg38 release) of each detected SNPs are listed. The haplotypes are clustered by presence of *1 or *2 allele of the hs1.2 enhancer and by origin of the sample (Healthy control (H) or Psoriatic patients (P)). The SNPs with grey background generate

two haplotypes that appear differentially linked to the hs1.2 alleles. “+” indicate concordance with the hg38 reference sequence; ** indicate SNP code concordance between dbSNP142 and dbSNP138.

Figure Legend

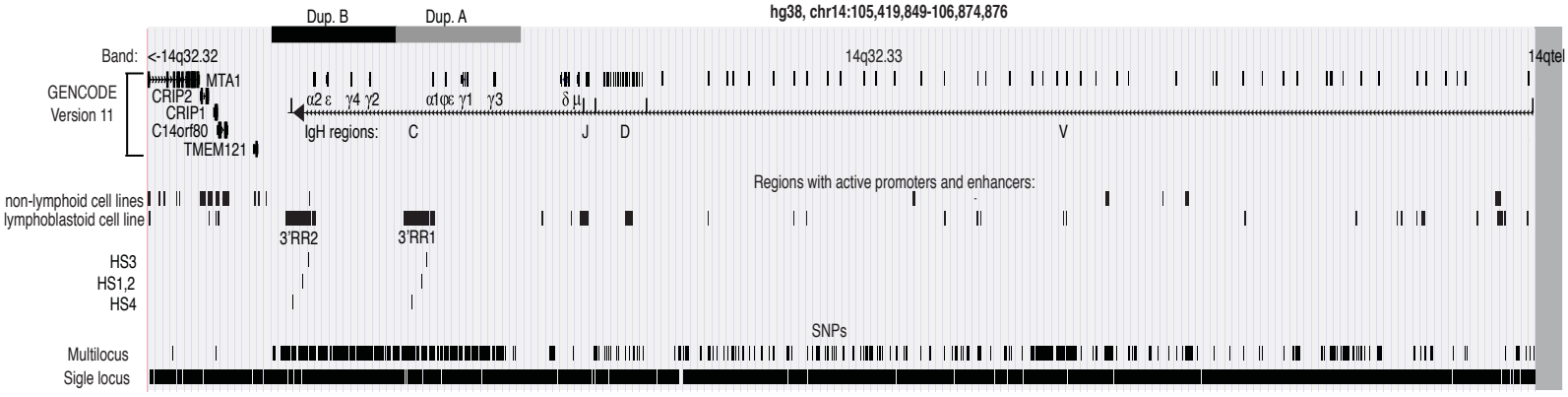
Figure1: analysed region, 3'RR1

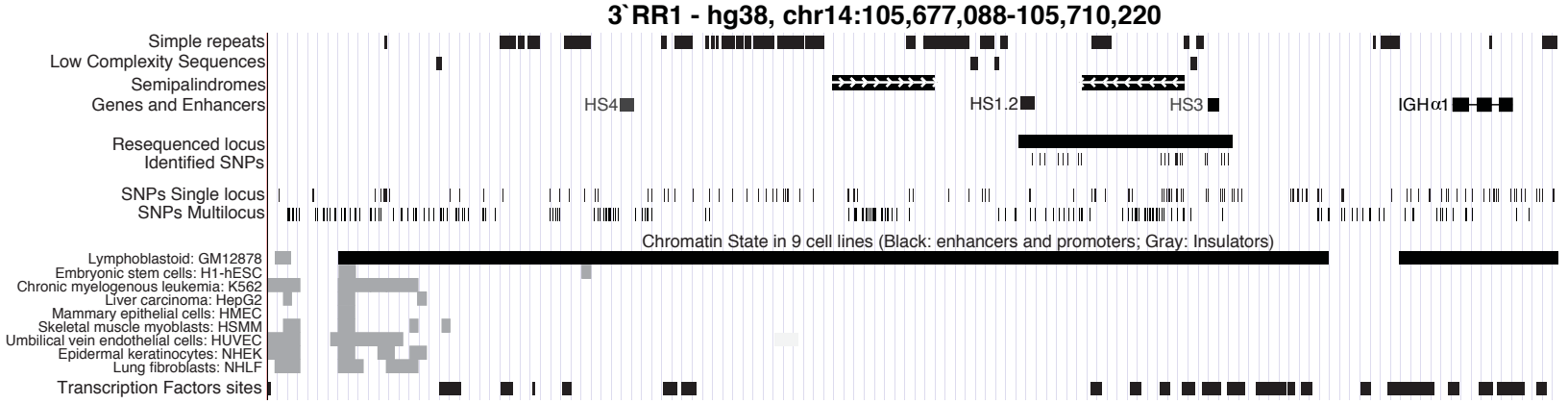
Main elements in the region under investigation and positions of the detected SNPs.

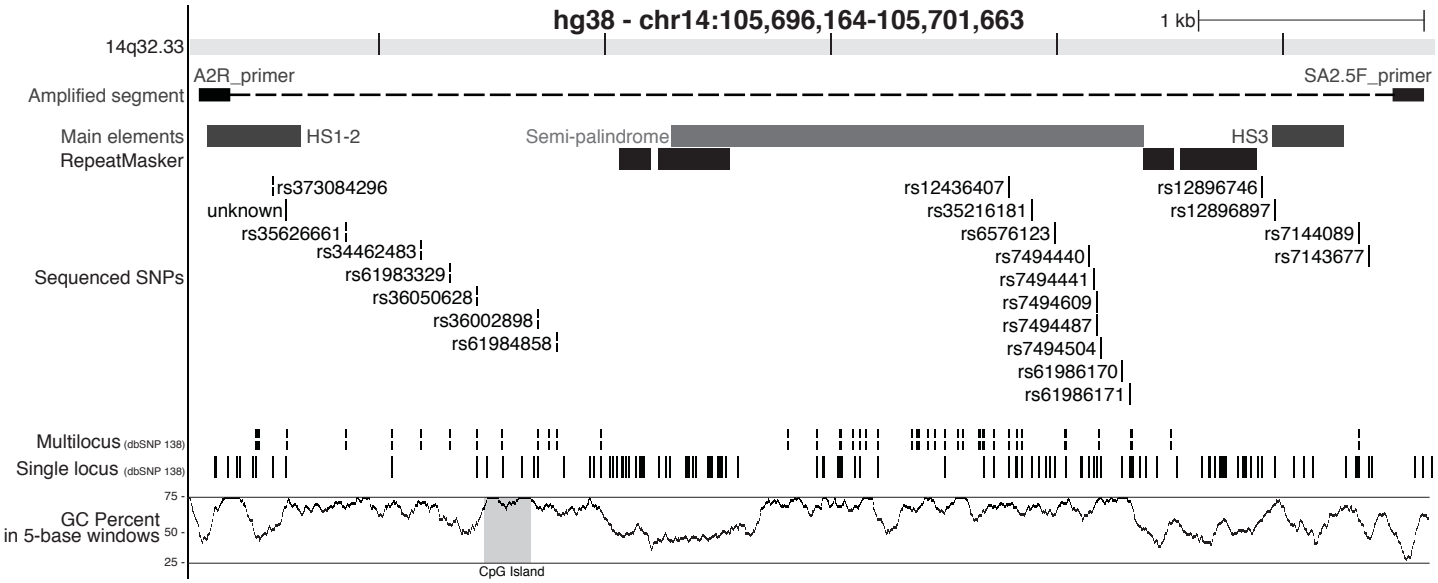
A) The Ig heavy chain region on human chromosome 14, from the variable to the constant region (the transcription direction of the Ig genes is represented by the gapped arrow). The two duplicated regions are marked. The RRs shows activity only in lymphoblastoid cell lines. At the bottom of the schema the positions of known single and multilocus SNPs are shown.

B) Details of the 3’RR1, close to the IgH alpha1 gene. The region includes 2 semipalindromes, matching each other, and 3 enhancers (hs4, hs1.2 and hs3). The amplified and resequenced region is marked. The lymphoblastoid cell line GM12878 shows high transcriptional activity in this region. On the bottom are reported the sites with consensus sequence for transcription-factors binding.

C) The region sequenced, limited by hs3 and hs1.2, and including a semi-palindrome. The SNPs identified are reported (see Table 1). GC percentage and position of CpG islands are also depicted.


hg38, chr14:105,419,849-106,874,876
Dup. B
Dup. A
Band:
<-14q32.32
14q32.33
14qtel
GENCODE
Version 11
MTA1
CRIP2
CRIP1
C14orf80
TMEM121
α2 ε γ4 γ2
α1φε γ1 γ3
δ μ
IgH regions:
C
J
D
V
Regions with active promoters and enhancers:
non-lymphoid cell lines
lymphoblastoid cell line
3'RR2
3'RR1
HS3
HS1,2
HS4
SNPs
Multilocus
Sigle locus


3`RR1 - hg38, chr14:105,677,088-105,710,220
Simple repeats
Low Complexity Sequences
Semipalindromes
Genes and Enhancers
HS4
HS1.2
HS3
IGHα1
Resequenced locus
Identified SNPs
SNPs Single locus
SNPs Multilocus
Chromatin State in 9 cell lines (Black: enhancers and promoters; Gray: Insulators)
Lymphoblastoid: GM12878
Embryonic stem cells: H1-hESC
Chronic myelogenous leukemia: K562
Liver carcinoma: HepG2
Mammary epithelial cells: HMEC
Skeletal muscle myoblasts: HSMM
Umbilical vein endothelial cells: HUVEC
Epidermal keratinocytes: NHEK
Lung fibroblasts: NHLF
Transcription Factors sites


hg38 - chr14:105,696,164-105,701,663
1 kb
14q32.33
A2R_primer
SA2.5F_primer
Amplified segment
Main elements
HS1-2
Semi-palindrome
HS3
RepeatMasker
rs373084296
unknown
rs35626661
rs34462483
rs61983329
rs36050628
rs36002898
rs61984858
rs12436407
rs35216181
rs6576123
rs7494440
rs7494441
rs7494609
rs7494487
rs7494504
rs61986170
rs61986171
rs12896746
rs12896897
rs7144089
rs7143677
Sequenced SNPs
Multilocus (dbSNP 138)
Single locus (dbSNP 138)
GC Percent in 5-base windows
75
50
25
CpG Island

	Non-linked SNP (13)		N2	N3	N4	N5	N6	N7	N8	N9	N10	N11			N14	N15	N16									
	Linked SNP (9)	**L1**											**L12**	**L13**				**L17**	**L18**	**L19**	**L20**	**L21**	**L22**			
	Subregion of 3'RR1	**HS1.2**	**Inner Region**							**Semi-palindrome**										**HS3 Region**						
	Alleles	**T/C**	C/T	C/A	A/C	C/T	A/G	G/C	A/G	C/A	A/G	G/C	**G/C**	**G/A**	A/G	C/G	C/A	**C/T**	**G/A**	**T/C**	**G/A**	**C/G**	**T/C**			
	MultipleAlignments note			Yes	Yes	Yes	Yes	Yes	Yes																	
	SNP code (dbSNP138)	**rs373084296**	n.a.	rs35626661	rs34462483	rs61983329	rs36050628	rs36002898	rs61984858	rs12436407	rs35216181	rs6576123	**rs7494440**	**rs7494441**	rs7494609	rs7494487	rs7494504	**rs61986170**	**rs61986171**	**rs12896746**	**rs12896897**	**rs7144089**	**rs7143677**				
	alternative SNP code (dbSNP142)	**n.a.**	n.a.	n.a.	n.a.	n.a.	**	rs143447395	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**			
	SNP position on chr14 (release hg38)	**105696525**	105696707	105696850	105697185	105697309	105697430	105697696	105697778	105699778	105699884	105699981	**105700138**	**105700157**	105700166	105700167	105700189	**105700281**	**105700316**	**105700902**	**105700958**	**105701324**	**105701369**			
*1	E1W (P=29; H=27)	+	+	+	C	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	E1W1	P	17
		+	+	+	C	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		H	16
		+	+	+	C	+	+	+	+	+	+	+	+	+	G	+	+	+	+	+	+	+	+	E1W2	P	4
		+	+	+	C	+	+	+	+	+	+	+	+	+	G	+	+	+	+	+	+	+	+		H	2
		+	+	+	C	+	+	+	+	+	+	C	+	+	G	G	A	+	+	+	+	+	+	E1W3	P	2
		+	+	+	C	+	+	+	+	+	+	C	+	+	G	G	A	+	+	+	+	+	+		H	4
		+	+	A	C	+	+	+	+	+	+	+	+	+	G	+	+	+	+	+	+	+	+	E1W4	P	2
		+	+	A	C	+	+	+	+	+	+	+	+	+	G	+	+	+	+	+	+	+	+		H	1
		+	+	+	+	+	+	+	+	+	+	C	+	+	+	+	+	+	+	+	+	+	+	OTHERS	P	1
		+	+	+	C	+	+	+	G	+	+	+	+	+	+	+	+	+	+	+	+	+	+		P	3
		+	+	+	+	+	+	+	+	+	+	C	+	+	G	+	A	+	+	+	+	+	+		H	2
		+	+	+	+	+	+	+	+	+	+	C	+	+	G	G	A	+	+	+	+	+	+		H	1
		+	+	+	C	+	+	+	+	+	+	C	+	+	+	+	+	+	+	+	+	+	+		H	1
*2	E2A (P=34; H=31)	**C**	+	+	+	T	G	+	G	A	+	C	**C**	**A**	G	G	A	**T**	**A**	**C**	**A**	**G**	**C**	E2A1	P	1
		C	+	+	+	T	G	+	G	A	+	C	**C**	**A**	G	G	A	**T**	**A**	**C**	**A**	**G**	**C**		H	21
		C	+	+	+	T	G	+	G	A	G	C	**C**	**A**	G	G	A	**T**	**A**	**C**	**A**	**G**	**C**	E2A2	P	18
		C	+	+	+	T	G	+	G	A	G	C	**C**	**A**	G	G	A	**T**	**A**	**C**	**A**	**G**	**C**		H	7
		C	+	+	+	+	G	+	G	+	G	C	**C**	**A**	G	G	A	**T**	**A**	**C**	**A**	**G**	**C**	OTHERS	P	4
		C	+	+	+	T	G	+	G	+	+	+	**C**	**A**	G	G	A	**T**	**A**	**C**	**A**	**G**	**C**		P	1
		C	+	+	+	T	G	+	G	+	+	C	**C**	**A**	G	G	A	**T**	**A**	**C**	**A**	**G**	**C**		P	4
		C	+	+	+	T	G	C	G	A	+	C	**C**	**A**	G	G	A	**T**	**A**	**C**	**A**	**G**	**C**		P	4
		C	+	A	+	T	+	+	G	A	G	C	**C**	**A**	G	G	A	**T**	**A**	**C**	**C**	**G**	**C**		P	1
		C	T	+	+	+	G	+	G	+	+	C	**C**	**A**	G	G	A	**T**	**A**	**C**	**A**	**G**	**C**		P	1
		C	+	+	C	T	G	+	G	A	+	C	**C**	**A**	G	G	A	**T**	**A**	**C**	**A**	**G**	**C**		H	2
		C	T	+	C	T	G	+	G	A	+	C	**C**	**A**	G	G	A	**T**	**A**	**C**	**A**	**G**	**C**		H	1
homozygous hs1.2 allele	Haplotype code (9 Linked SNPs)																							Hap. code (22 SNPs)	Patients/ Healthy controls	#