Author/s:
Rozemuller, E;Eckle, SBG;McLaughlin, I;Penning, M;Mulder, W;de Bruin, H;van Wageningen, S

Title:
MR1 encompasses at least six allele groups with coding region alterations

Date:
2021-12-01

Citation:
Rozemuller, E., Eckle, S. B. G., McLaughlin, I., Penning, M., Mulder, W., de Bruin, H. & van Wageningen, S. (2021). MR1 encompasses at least six allele groups with coding region alterations. Hla, 98 (6), pp.509-516. https://doi.org/10.1111/tan.14390.

Persistent Link:
https://hdl.handle.net/11343/298870

van Wageningen Sake (Orcid ID: 0000-0003-4034-1697)

Full title: MR1 encompasses at least six allele groups with coding region alterations.

Short title: Six MR1 alleles.

Authors
1 Erik Rozemuller[1]
2 Sidonia Barbara Guiomar Eckle[2]
3 Ian McLaughlin[3]
4 Maarten Penning[1]
5 Wietse Mulder[1]
6 Helma de Bruin[1]
7 Sake van Wageningen[1]

[1]GenDx, Yalelaan 48, 3584CM, Utrecht, The Netherlands
[2] Department of Microbiology and Immunology, The University of Melbourne at the Peter Doherty Institute for Infection and Immunity, Melbourne, Australia
[3]Pacific Biosciences, Menlo Park, California, USA

Communicating author; Sake van Wageningen
Email; s.vanwageningen@gendx.com
Address; GenDx, Yalelaan 48, 3584CM, Utrecht, The Netherlands

Data availability statement
Supporting data are available on request from the corresponding author. New alleles were submitted to GenBank (see Figure 1 for accession numbers).

Author contributions
E.R. and S.v.W. designed the study and perfomed the sequence analysis. S.I. performed protein structure analysis. I.M., H.d.B. and S.v.W performed experiments. E.R., M.P., W.M. and S.v.W. analysed data and wrote the manuscript.

Acknowledgements
We thank John Trowsdale and James McCluskey for advice on the manuscript.

Abstract

Unlike classical HLA class I genes, MR1 is assumed to have limited polymorphic positions. We developed a MR1 specific PCR assay and sequenced 56 DNA samples from cells with a diverse set of HLA genotypes. In this relatively small panel we found six allele groups encoding for different MR1 proteins. The two most frequent allele groups found in this panel had a frequency of 71% (MR1*01) and 25% (MR1*02), respectively. Moreover, the panel contained many intronic SNPs and silent variants, with individual samples containing up to 15 heterozygous positions. The data presented here is consistent with marked variation in MR1.

Text

Classical MHC-I proteins, also known as MHC-Ia proteins, are encoded by the most polymorphic genes in the genome and for successful organ and stem cell transplantation MHC compatibility is a prerequisite. Non-classical MHC class I genes, also known as MHC class Ib genes, and MHC-I-like molecules have few alleles and are considered oligo- or even mono-morphic [1, 2, 3]. Classical MHC genes are clustered together on human chromosome 6 and some MHC-Ib genes are interspersed between these loci. MHC-I-like genes genes reside outside the MHC locus. MHC class I-related protein 1 (MR1) is an MHC-I like gene that is on chromosome 1. In spite of this, MR1 shares greater homology with MHC-Ia genes than it does with other MHC-I like genes [4, 5]. The structure of MR1 is that of a typical MHC class I gene, with exon 1 coding for a leader peptide and exons 2-4 coding for extracellular domains, including α1- and α2-domains that form the antigen-binding cleft. Exon 5 and 6 are coding sequences for the transmembrane and cytoplasmic domains, respectively [6]. An unusual feature of MR1 is a large intron 1 of 15kb. MR1 presents microbially derived vitamin B-related metabolites to a set of unconventional T cells called mucosal-associated invariant T cells (MAIT cells) [7, 8, 9]. MAIT cells in mice have been shown to play a role in controlling infections with Legionella longbeachae [10], Francisella tularensis [11], Mycobacterium bovis [12], pathogenic E. coli [13], Clostridium difficile [14] and Klebsiella pneumoniae [15]. Recently, MR1 was highlighted as a monomorphic, pan-cancer, pan-population target for immunotherapies [16]. This study suggested that MR1 is able to present ligands derived from the cancer metabolome, making MR1 expressing cancer cells a target for T cells that become potent killers of human cancer cells. Despite its monomorphic classification, the NCBI variation viewer reveals multiple SNPs in MR1, three of which have been studied more extensively [17, 18, 19]. Stumpf et al.[17] discovered a variant allele of MR1 in a chronic myeloid lymphocyte (CML) patient that was treated with donor lymphocyte infusions (DLI), which acted as an HLA-class II restricted minor histocompatibility antigen, causing mild graft-versus-host disease (GVHD) of the skin. In this study the donor was homozygous and the patient heterozygous for SNP rs2236410. Howson et al.[18] discovered a homozygous mutation in MR1 (R9H) in a patient that had a history of difficult-to-treat viral and bacterial infections. The R9H MR1 protein could no longer bind microbial MR1 ligand (5-OP-RU) and the patient appeared to lack MAIT cells. Seshadri et al.[19] studied intronic SNPs in a cohort of Vietnamese adults and discovered an association of MR1 SNP rs1052632 with MR1 gene expression and susceptibility to tuberculosis. In this paper we report new allele variants for MR1 and map the variant residues to the protein structure.

To study variation of MR1 we developed a PCR assay that amplifies a region spanning intron 1 to the 3' UTR, so that the resulting amplicon of 7kb includes all exons encoding for the MR1 protein. We performed PCR reactions for 56 DNA samples from the genetic testing reference material from the Coriell Cell Repositories (Table 1). We sequenced the samples on an Illumina Miseq and analysed the data using NGSengine software. We analysed the complete amplicon, ignoring only a long A-homopolymer repeat in intron 3 (nucleotide 16972-16985 upstream of the start codon). The sequence data was of high quality with a median read mappability of 90% and a sequence noise percentage in all exon regions below 3%, indicating that the MR1 amplification is very specific.

In this relatively small panel a diverse set of heterozygous genotypes was discovered (Table 1A). In all cases the minor fraction of the heterozygous positions was between 43-50%, indicating a balanced amplification of both alleles in each sample. In this panel we validated three intronic SNPs that were described before using different samples (rs10797666, rs1052632, rs2877545) [19], including the rs1052632 variant that has been associated with susceptibility to tuberculosis. In addition to many intron variants, we found multiple samples with polymorphic nucleotide positions in the exon region compared to the published genomic MR1

sequence (NCBI Ref Seq: NG_042221.1 (Gene ID: 3140)). Almost all variant positions could be linked to a specific SNP in the NCBI variation viewer and many of these SNPs were also found in the 1000 Genomes (1000G) Project (Table 1A). SNPs with a high frequency in the 1000G Project also had a high frequency in the 56 samples panel we studied, indicating these variants are not sequencing artefacts. Although many polymorphisms were found in the introns, none were in the 2 bases that flank any exon and that are important for splicing. Within the exons both silent and non-silent mutations were found. Based solely on sequence data It is difficult to determine if any of these changes also affect splicing. The polymorphisms also included coding region alterations in exons 2 and 3. These exons encode the domains that form the ligand-binding pocket and are the most polymorphic regions in classical MHC-I genes and are the most relevant in the context of transplantation.

To obtain fully phased MR1 allele sequences, MR1 was re-amplified from 5 samples containing alleles encoding for novel MR1 proteins. These amplicons were used to create a library of SMRTbell templates and were sequenced on a PacBio Sequel II System. This method resulted in high quality long reads that covered the complete amplicons and in which phasing of all heterozygous positions could be achieved (Figure 1). Based on this long-read sequencing data new alleles could be described, for which we proposed the following provisional nomenclature; first field differences indicating protein variants, second field reserved for silent mutations and third field used for intron variants (Table 1B). An allele group was found (MR1*02), with a coding region alteration in exon 2 resulting in a His to Arg amino acid change at position 17. Based on the 56 samples in the original panel we calculated an allele frequency of 25% for MR1*02. In an earlier study, this polymorphic position was identified as a HLA-DR-restricted minor histocompatibility antigen in a CML patient after alloSCT [17]. Using a different sample panel Stumpf et al.[17] calculated an identical frequency of 25% for this variant. H17 is not conserved between species (e.g. Pongo Pygmaeus: R, Mouse: P, Bovine: Y, Opossum: Q) [20]. MR1*01 and MR*02 were the most frequent alleles in our panel with respectively 71% and 25%.

In our panel of 56 samples, four more samples were found to have alleles with changes resulting in amino acid variants in the encoded protein (Figure 2). MR1*04 contained 2 amino acid changes. A change at position 17, similar to MR1*02, and additionally an Arg to His change at position 9. Previously a patient homozygous for MR1 R9H has been described. MR1 R9H was unable to present the riboflavin based MAIT cell antigen 5-(2-oxopropylideneamino)-6-D-ribitylaminouracil (5-OP-RU) and the patient lacked circulating MAIT cells [18]. Notably acetyl-6-formylpterin, an MR1 ligand which does not stimulate MAIT cells, could still bind to MR1 [18]. Figure 3 shows a cartoon display of MR1 with the variant residues highlighted in teal in the α1- and α2-domains (Figure 3A) and in the α3-domain and cytoplasmic tail (Figure 3B). None of the polymorphic residues are TCR contacts in crystal structures of MAIT TCRs in complex with MR1-5-OP-RU (Figure 3A). R9 represents a TCR contact in crystal structures of a MAIT TCR complexed with MR1 presenting the drug diclofenac or its metabolite 5-OH-diclofenac [21]. Accordingly R9H might be relevant for TCR recognition of MR1 ligands whose presentation is not disrupted by R9H.

Glutamate is a polar and negatively charged amino acid whilst glycine is non-polar and hydrophobic; histidine is a polar, positively charged and very hydrophilic amino acid whilst glutamine is polar and uncharged. Accordingly, E52G and H90Q represent non-conservative amino acid changes and such changes could impact on the function or folding of MR1 but this warrants testing in functional studies.

I121V represents a conservative change and is located in a region that is unlikely to be of functional importance. I121V is also a polymorphic residue in Opossum.

Isoleucine and valine are both non-polar and hydrophobic so that I244V represents a conservative amino acid change. I244V is furthermore located in a region that is not known and unlikely to be of functional importance. I244V is also a Valine in other species MR1 (e.g. Macaque, Mouse, Bovine and Pig).

Arginine and lysine are both polar, positively charged and very hydrophilic amino acids so that R304K represents a conservative amino acid change in the cytoplasmic tail. Position 304 is also a lysine in MR1 of other species, namely Opossum, Wallaby and Tasmanian devil [20].

Although it is unsure if there is any role for MR1 in histocompatibility, the polymorphic nature of MR1 described here would justify a larger study into the clinical relevance of MR1 variability, both in the context of transplantation and immunotherapy approaches targeting MR1, and MR1 disease-association studies. In the context of transplantation, it is possible that MR1 variants create minor histocompatibility antigens in addition to the ones described [17]. Furthermore, allelic differences in MR1 at the T cell receptor interface could induce alloreactive T cell responses. Other allelic differences in the MR1 protein could impact on antigen presentation which in turn could be associated with diseases, as previously suspcected [18]. Additional variants outside the coding region could also be associated with diseases, affecting MR1 gene regulation, as previously described [19]. Lastly, allelic differences in MR1 at the T cell receptor interface potentially challenges the premise of MR1 representing a pan-population target for immunotherapies.

References

1. Parra-Cuadrado JF, Navarro P, Mirones I, Setien F, Oteo M, Martinez-Naves E. A study on the polymorphism of human MHC class I-related MR1 gene and identification of an MR1-like pseudogene. *Tissue Antigens*. 2000;56(2):170-172. doi:10.1034/j.1399-0039.2000.560211.x
2. J. T, J.C. K. Major histocompatibility complex genomics and human disease. *Annu Rev Genomics Hum Genet*. Published online 2013.
3. Rodgers JR, Cook RG. MHC class IB molecules bridge innate and acquired immunity. *Nat Rev Immunol*. Published online 2005. doi:10.1038/nri1635
4. Krovi SH, Gapin L. Structure and function of the non-classical major histocompatibility complex molecule MR1. *Immunogenetics*. Published online 2016. doi:10.1007/s00251-016-0939-5
5. Hashimoto K, Hirai M, Kurosawa Y. A gene outside the human MHC related to classical HLA class I genes. *Science (80-)*. Published online 1995. doi:10.1126/science.7624800
6. Riegert P, Wanner V, Bahram S. Genomics, isoforms, expression, and phylogeny of the MHC class I-related MR1 gene. *J Immunol*. Published online 1998.
7. Kjer-Nielsen L, Patel O, Corbett AJ, et al. MR1 presents microbial vitamin B metabolites to MAIT cells. *Nature*. Published online 2012. doi:10.1038/nature11605
8. Corbett AJ, Eckle SBG, Birkinshaw RW, et al. T-cell activation by transitory neo-antigens derived from distinct microbial pathways. *Nature*. Published online 2014. doi:10.1038/nature13160
9. Treiner E, Duban L, Bahram S, et al. Selection of evolutionarily conserved mucosal-associated invariant T cells by MR1. *Nature*. Published online 2003. doi:10.1038/nature01433
10. Wang H, D'Souza C, Lim XY, et al. MAIT cells protect against pulmonary Legionella longbeachae infection. *Nat Commun*. Published online 2018. doi:10.1038/s41467-018-05202-8
11. Meierovics A, Yankelevich WJC, Cowley SC. MAIT cells are critical for optimal mucosal immune responses during in vivo pulmonary bacterial infection. *Proc Natl Acad Sci U S A*. Published online 2013. doi:10.1073/pnas.1302799110
12. Chua WJ, Truscott SM, Eickhoff CS, Blazevic A, Hoft DF, Hansen TH. Polyclonal mucosa-associated invariant T cells have unique innate functions in bacterial infection. *Infect Immun*. Published online 2012. doi:10.1128/IAI.00279-12
13. Cui Y, Franciszkiewicz K, Mburu YK, et al. Mucosal-associated invariant T cell-rich congenic mouse strain allows functional evaluation. *J Clin Invest*. Published online 2015. doi:10.1172/JCI82424
14. Smith AD, Foss ED, Zhang I, et al. Microbiota of MR1 deficient mice confer resistance against Clostridium difficile infection. *PLoS One*. Published online 2019. doi:10.1371/journal.pone.0223025
15. Georgel P, Radosavljevic M, Macquin C, Bahram S. The non-conventional MHC class I MR1 molecule controls infection by Klebsiella pneumoniae in mice. *Mol Immunol*. Published online 2011. doi:10.1016/j.molimm.2010.12.002
16. Crowther MD, Dolton G, Legut M, et al. Genome-wide CRISPR–Cas9 screening reveals ubiquitous T cell cancer targeting via the monomorphic MHC class I-related protein MR1. *Nat Immunol*. Published online 2020. doi:10.1038/s41590-019-0578-8
17. Stumpf AN, Van Der Meijden ED, Van Bergen CAM, Willemze R, Falkenburg JHF, Griffioen M. Identification of 4 new HLA-DR-restricted minor histocompatibility antigens as hematopoietic targets in antitumor immunity. *Blood*. Published online 2009. doi:10.1182/blood-2009-03-208017
18. Howson LJ, Awad W, Borstel A Von, et al. Absence of mucosal-associated invariant T cells in a person with a homozygous point mutation in MR1. *Sci Immunol*. Published online 2020. doi:10.1126/SCIIMMUNOL.ABC9492
19. Seshadri C, Thuong NTT, Mai NTH, et al. A polymorphism in human MR1 is associated with mRNA expression and susceptibility to tuberculosis. *Genes Immun*. Published online 2017. doi:10.1038/gene.2016.41
20. Tsukamoto K, Deakin JE, Graves JAM, Hashimoto K. Exceptionally high conservation of the MHC

class I-related gene, MR1, among mammals. *Immunogenetics*. Published online 2013. doi:10.1007/s00251-012-0666-5

21. Keller AN, Eckle SBG, Xu W, et al. Drugs and drug-like molecules can modulate the function of mucosal-associated invariant T cells. *Nat Immunol*. Published online 2017. doi:10.1038/ni.3679
22. Bettinotti MP, Ferriola D, Duke JL, et al. Characterization of 108 Genomic DNA Reference Materials for 11 Human Leukocyte Antigen Loci: A GeT-RM Collaborative Project. *J Mol Diagnostics*. Published online 2018. doi:10.1016/j.jmoldx.2018.05.009
23. Laskowski RA, Hutchinson EG, Michie AD, Wallace AC, Jones ML, Thornton JM. PDBsum: A Web-based database of summaries and analyses of all PDB structures. *Trends Biochem Sci*. Published online 1997. doi:10.1016/S0968-0004(97)01140-7
24. Eckle SBG, Birkinshaw RW, Kostenko L, et al. A molecular basis underpinning the T cell receptor heterogeneity of mucosal-associated invariant T cells. *J Exp Med*. Published online 2014. doi:10.1084/jem.20140484

Table 1 **Variant positions in MR1 in different samples.** Human genomic DNA samples were from the Coriell Institute (New Jersey, USA) and were extracted from ethnically diverse cell lines from the NIGMS Human Genetic Cell Repository [22]. DNA samples were used to amplify MR1 by a Polymerase Chain Reaction (PCR) using Primers (GenDx, Utrecht, NL) and the GenDx-LongMix PCR kit according to the Instructions for Use. The enzyme in this kit is a mixture of a Taq polymerase and a proofreading enzyme with 3'-5'-exonuclease activity for the removal of erroneously built-in, wrong bases. The processivity is primarily driven by the Taq polymerase. Amplicons were processed in the NGSgo workflow for Illumina using NGSgo-LibrX and NGSgo-IndX (GenDx). Libraries were paired-end sequenced (2x150bp) on a MiSeq platform (Illumina). FASTQ files were analyzed in NGSengine HLA typing software (GenDx). A) Samples with 1 or more bases difference compared to MR1*01:01:01 (NCBI Ref Seq: NG_042221.1) are shown. Samples NA17203, NA17204, NA17206, NA17210, NA17226, NA17234, NA17246, NA17274, NA17286, NA17618 were invariant for MR1. Most variant positions could be linked to a specific SNP (Variant ID) in the NCBI variation viewer and many of these SNPs were also found in the 1000 Genomes (1000G) Project. The 1000G MAF is the frequency of the minor allele for the Phase 3 May 2013 call set. Non-silent variants are grouped on top, followed by silent variants and intronic or UTR variants. B) Variant positions in fully phased MR1 alleles.

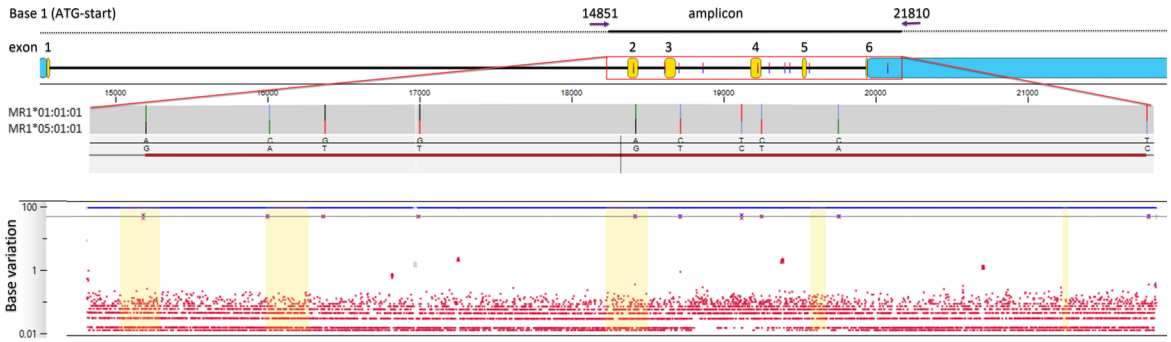
Figure 1. **New MR1 alleles sequenced on a PacBio Sequel II system.** A SMRTbell library was constructed with SMRTbell® Express Template Prep Kit 2.0. Primer annealing and polymerase binding were performed following instructions in the Sample Setup module of SMRT Link analysis software. Resulting fastq files were analyzed using NGSengine software. Shown is a NGSengine overview of sample NA17205. On top the gene overview with exons in yellow and UTR regions in blue. Below the gene overview is the coverage plot of exon 2-exon 6 with heterozygous positions indicated in different colours. The base variation plot shows the heterozygous position around 50% and noise at 1% or lower. Features of the new MR1 alleles are described below the base variation plot. All new alleles for which full phasing information was acquired were submitted to GenBank (MW239102 MR1*01:01:02, MW239103 MR1*02:01:02, MW239104 MR1*01:01:03, MW239105 MR1*06:01:01, MW239106 MR1*05:01:01, MW239107 MR1*02:01:01, MW239108 MR1*03:01:01, MW239109 MR1*01:01:04, MW239110 MR1*04:01:01). First field differences indicate protein variants, second field is reserved for silent mutations and third field is used for intron variants.

Figure 2. **Amino acid alignment of different MR1 alleles.** Shown is the mature protein after removal of the signal peptide.

Figure 3. **A) Differences in MR1 alleles in the α1- and α2-domains.** Cartoon and surface display (light grey) of the MR1 Ag-binding cleft and stick display of 5-OP-RU (black) based on the protein data bank (PDB) deposited crystal structure featuring the human A-F7 MAIT TCR in complex with human MR1-5-OP-RU (PDB ID: 4NQC [8]) . MR1 atoms contacted in hydrogen bonds (there are no residues contacted in salt bridges) based on analysis using PDBsum [23] in any of available ternary crystal structures of human MAIT TCRs in complex with human MR1-5-OP-RU (PDB IDs: 4NQC [8]; PDB IDs: 4PJ7, 4PJ8, 4PJ9, 4PJA, 4PJB, 4PJC, 4PJD [24]) are highlighted in magenta, residue side-chains displayed in stick format and residues labelled in black. All contacts include atoms of side-chains except for Leu65, where the main chain is contacted. The only MR1 residues contacted in all analysed structures are Tyr62 and Tyr152. MR1 residues that differ in the MR1 alleles MR1*02, *03, *04, *05 are highlighted and labelled in teal, and residue side-chains displayed in stick format. **B) Differences in MR1 alleles in the α3-domain and the cytoplasmic tail.** Cartoon display (light grey) of MR1 and β2m (light blue), and stick display of 5-OP-RU (black) based on the PDB deposited crystal structure featuring the human A-F7 MAIT TCR in complex with human MR1-5-OP-RU (PDB ID: 4NQC [8]); schematic of the transmembrane

domain (TM) and cytoplasmic tail (Cyt. tail). The MR1 residues that differ in the MR1 alleles MR1*05 and *06 are highlighted and labelled in teal.

Figure 1



	Frequency	new base	Amino Acid change	AA# (exon#)	Sample	ref	position conserved between species
MR1*01	80/112				NA17205	*Gene ID: 3140*	
MR1*02	28/112	CAT>CGT	His>Arg	17 (exon 2)	NA07439	*Stumpf et al.* [17]	no
MR1*03	1/112	ATC>GTC	Ile>Val	121 (exon 3)	NA17222		no
MR1*04	1/112	CGC>CAC	Arg>His	9 (exon 2)	NA17228	*Howson et al.* [18]	
		CAT>CGT	His>Arg	17 (exon 2)		*Stumpf et al.* [17]	no
MR1*05	1/112	GAG>GGG	Glu>Gly	52 (exon 2)	NA17205		
		CAC>CAA	His>Gln	90 (exon 3)			
		ATT>GTT	Ile>Val	244 (exon 4)			no
MR1*06	1/112	AGG>AAG	Arg>Lys	304 (exon 5)	NA17114		no

	SNP	new base	position		Sample	ref
MR1*02:01:01	rs1052632	G>A	17087	intron 3	NA17222	*Seshadri et al.* [19]
MR1*04:01:01	rs1052632	G>A	17087	intron 3	NA17228	*Seshadri et al.* [19]

Figure 1.tif

Figure 2

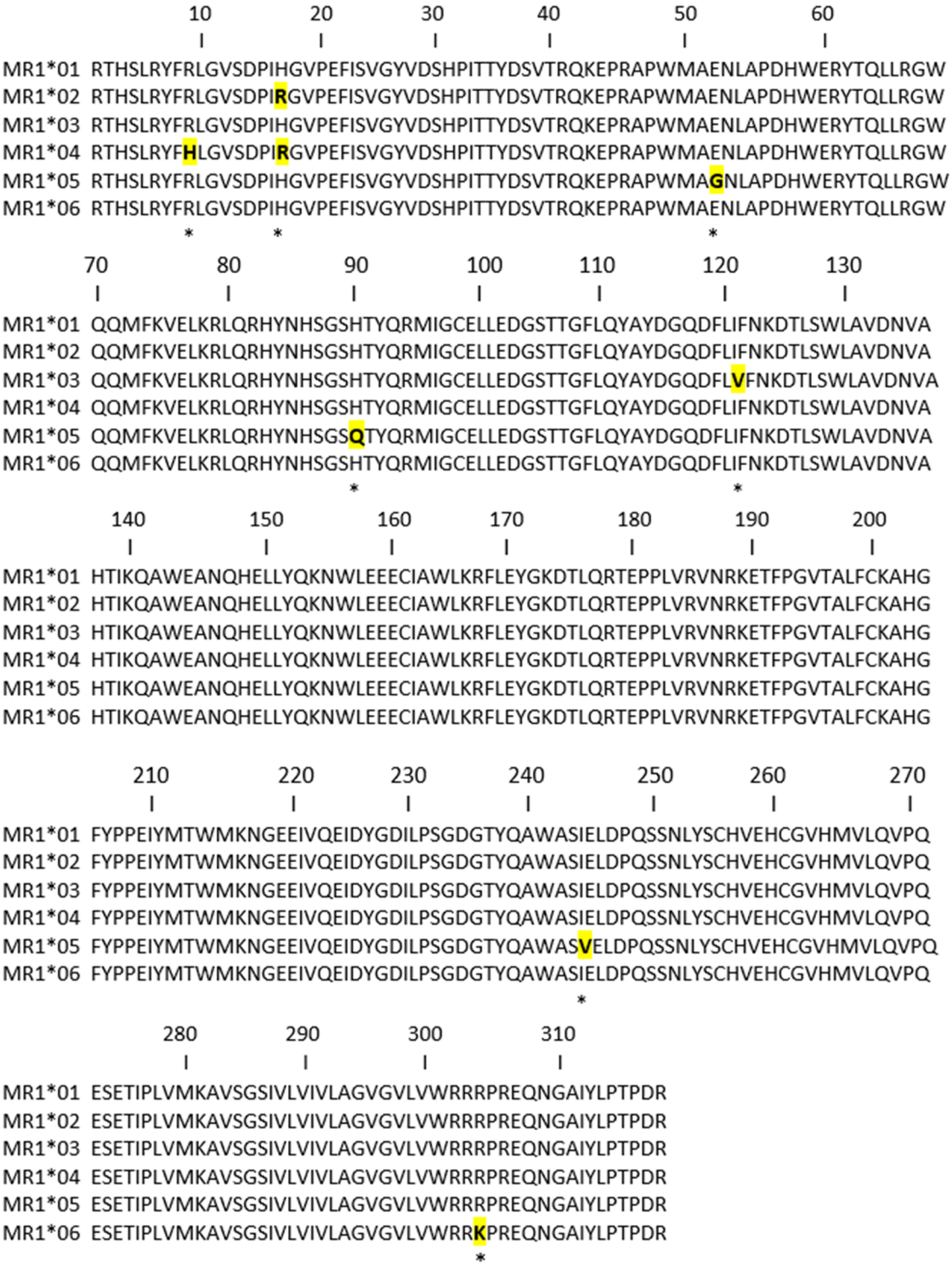
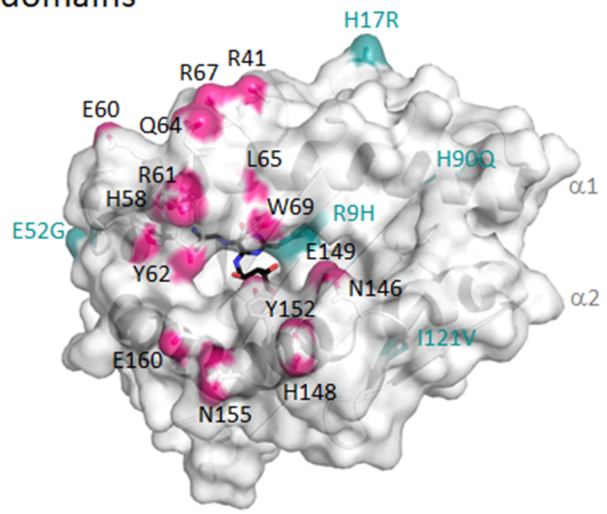


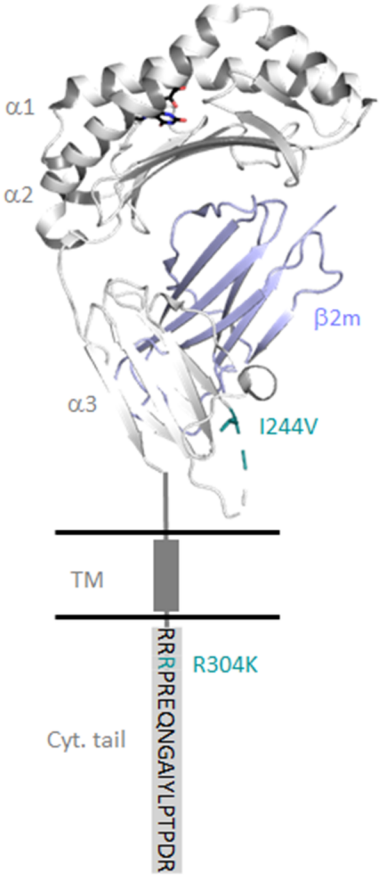
Figure 2.tif

Figure 3

A) Polymorphism in the α1/α2-domains

B) Polymorphism in the α3-domain/cytoplasmic tail





	α1		α2		α3	Cyt tail
MR1*02:	H17R					
MR1*03:			I121V			
MR1*04:	R9H	H17R				
MR1*05:		E52G	H90Q		I244V	
MR1*06:						R304K

Figure 3.tif

Table 1A) Variant positions in MR1 in different samples.

Variant ID	*01:01:01	1000G MAF	NA07439	NA09301	NA10005	NA12244	NA16688	NA16689	NA17019	NA17020	NA17039	NA17057	NA17058	NA17078	NA17084	NA17114	NA17115	NA17119	NA17129	NA17130	NA17201	NA17205	NA17211	NA17212	NA17213	NA17214	NA17219	NA17222	NA17228	NA17229	NA17230	NA17232	NA17237	NA17240	NA17247	NA17256	NA17261	NA17264	NA17267	NA17269	NA17282	NA17285	NA17289	NA17290	NA17295	NA17296	NA17440	NA17466	variant type
rs2236410	A	G=0,1929	G					GG	G			G	G	G	G		G	G	G		G					G		G	G	G	G	GG							G					G		non-silent
rs376289951	A	G=0,0026																			G																												non-silent
rs577996797	C	A=0,0026																			A																												non-silent
rs41268456	G	A=0,0024																											A																			non-silent
rs186300848	A	G=0,0004																			G																												non-silent
rs374009364	A	NA																										G																				non-silent
rs754650505	G	NA														A																																	non-silent
rs61731104	C	T=0,0393															T	T																															silent
rs61731103	T	C=0,0389															C	C																															silent
rs61731101	T	C=0,0226															C	C																															silent
rs3863720	G	A=0,0166				A																A																				A					A	silent
rs35223984	C	T=0,0096																																				T											silent
rs2331994	C	T=0,4978		T			T	TT	T	TT		T	T	T	TT			T	TT	TT	T		T		T		T	T	TT	T	TT			T	T		TT		TT		T	T		TT		TT	intron / UTR
rs3747956	G	G=0,4776		A			A	AA	A	AA		A	A	A	AA			A	AA	AA	A		A		AA		A	A	AA	A	AA	AA		A	A		AA		AA		A	A	AA	A	AA		intron / UTR
rs10797666	A	A=0,4513		C			C	CC	C	CC		C	C	C	CC		C	CC	CC	CC	C		C		CC		C	C	CC	C	CC	CC		C	C		CC		CC		C	C	CC		CC		intron / UTR
rs1052632	G	G=0,4455		A			A	AA	A	AA		A	A	A	AA		A	AA	AA	AA	A		A		AA		A	A	AA	A	AA	AA		A	A		AA		AA		A	A	AA		AA		intron / UTR
rs3845423	C	C=0,4405		T			T	TT	T	TT		T	T	T	TT		T	TT	TT	TT	T	T	T		TT		T	T	TT	T	TT	TT		T	T		TT		TT		T	T	TT		TT		intron / UTR
rs2275469	C	C=0,4403		A			A	CC	A	AA		A	A	A	AA		A	AA	AA	AA	A	A	A		AA		A	A	AA	A	AA	AA		A	A		AA		AA		A	A	AA		AA		intron / UTR
rs6674136	C	C=0,388		T	T		T	TT	T	TT	TT	T	T	T	TT	T	T	TT	TT	TT	T	T	T		TT		T	T	TT	T	TT	TT	T	T	T		TT		TT		T	T	TT		TT		intron / UTR
rs2331993	G	A=0,2173					A			AA					A					A												A											A		A			intron / UTR
rs3789357	C	T=0,2165					T			TT					T					T												T											T		T			intron / UTR
rs3789359	T	C=0,2143					C			CC					C					C												C											C		C			intron / UTR
rs2236411	C	G=0,1058		G													G	G	G	G			G		G				G															G			intron / UTR
rs73050362	A	G=0,0607			G						G																					G															intron / UTR
rs2236412	T	C=0,0519				C					C		C	C								C		C		C				C		C	C			C	C		CC		C					intron / UTR
rs28540353	G	T=0,0389															T	T																														intron / UTR
rs76696261	G	A=0,0284			A																											A															intron / UTR
rs58860674	G	C=0,028									C																																					intron / UTR
rs7545452	T	C=0,025															C	C																														intron / UTR
rs59252681	C	T=0,0248															T	T																														intron / UTR
rs6667601	C	A=0,0236															A	A																														intron / UTR
rs112526759	G	A=0,0104									A																																					intron / UTR
rs59353588	G	C=0,0102	C		C												C																															intron / UTR
rs111999364	G	A=0,0102																															A															intron / UTR
rs75073791	G	A=0,0098																															A															intron / UTR
rs138270245	A	G=0,0052																											G																			intron / UTR
rs544605726	T	G=0,0038																															-															intron / UTR
rs187522902	G	A=0,0034																																										A				intron / UTR
rs150946490	C	T=0,0034												T																																		intron / UTR
rs139847219	C	A=0,0034																																									A				intron / UTR
rs143252572	A	T=0,0034																																									T				intron / UTR
rs571532490	T	C=0,0026																			C																												intron / UTR
rs560526084	G	T=0,0026																			T																												intron / UTR
rs543032425	G	T=0,0026																			T																												intron / UTR
rs531730583	T	C=0,0026																			C																												intron / UTR
rs184678860	G	A=0,0004														A																																	intron / UTR
rs536473637	A	T=0,0002	T																																													intron / UTR
rs548708632	A	G=0,0002														G																																	intron / UTR
rs2877545	C	NA					T			TT					T					T													T											T		T			intron / UTR
rs1318693907	G	NA				A																																											intron / UTR
base 21657	A	NA																																								C							intron / UTR

Table 1B) Variant positions in fully phased MR1 alleles.

Variant ID	*01:01:01	1000G MAF	***01:01:02**	***01:01:03**	***01:01:04**	***02:01:01**	***02:01:02**	***03:01:01**	***04:01:01**	***05:01:01**	***06:01:01**	variant type
rs2236410	A	G=0,1929				G	G		G			non-silent
rs376289951	A	G=0,0026								G		
rs577996797	C	A=0,0026								A		
rs41268456	G	A=0,0024							A			
rs186300848	A	G=0,0004								G		
rs374009364	A	NA						G				
rs754650505	G	NA									A	
rs61731104	C	T=0,0393										silent
rs61731103	T	C=0,0389										
rs61731101	T	C=0,0226										
rs3863720	G	A=0,0166										
rs35223984	C	T=0,0096										
rs2331994	C	T=0,4978				T			T			intron / UTR
rs3747956	G	G=0,4776				A			A			
rs10797666	A	A=0,4513				C			C			
rs1052632	G	G=0,4455				A			A			
rs3845423	C	C=0,4405				T			T	T		
rs2275469	C	C=0,4403				A			A	A		
rs6674136	C	C=0,388				T			T	T		
rs2331993	G	A=0,2173										
rs3789357	C	T=0,2165										
rs3789359	T	C=0,2143										
rs2236411	C	G=0,1058										
rs73050362	A	G=0,0607										
rs2236412	T	C=0,0519										
rs28540353	G	T=0,0389										
rs76696261	G	A=0,0284										
rs58860674	G	C=0,028										
rs7545452	T	C=0,025										
rs59252681	C	T=0,0248										
rs6667601	C	A=0,0236										
rs112526759	G	A=0,0104										
rs59353588	G	C=0,0102					C					
rs111999364	G	A=0,0102										
rs75073791	G	A=0,0098										
rs138270245	A	G=0,0052			G							
rs544605726	T	G=0,0038										
rs187522902	G	A=0,0034										
rs150946490	C	T=0,0034										
rs139847219	C	A=0,0034										
rs143252572	A	T=0,0034										
rs571532490	T	C=0,0026								C		
rs560526084	G	T=0,0026								T		
rs543032425	G	T=0,0026								T		
rs531730583	T	C=0,0026								C		
rs184678860	G	A=0,0004		A								
rs536473637	A	T=0,0002	T									
rs548708632	A	G=0,0002		G								
rs2877545	C	NA										
rs1318693907	G	NA										
base 21657	A	NA										