Journal of Human Genetics (2009) 54, 15–39

www.nature.com/jhg

REVIEW

The HLA genomic loci map: expression, interaction, diversity and disease

Takashi Shiina[1], Kazuyoshi Hosomichi[1], Hidetoshi Inoko[1] and Jerzy K Kulski[1,2]

The human leukocyte antigen (HLA) super-locus is a genomic region in the chromosomal position 6p21 that encodes the six classical transplantation HLA genes and at least 132 protein coding genes that have important roles in the regulation of the immune system as well as some other fundamental molecular and cellular processes. This small segment of the human genome has been associated with more than 100 different diseases, including common diseases, such as diabetes, rheumatoid arthritis, psoriasis, asthma and various other autoimmune disorders. The first complete and continuous HLA 3.6 Mb genomic sequence was reported in 1999 with the annotation of 224 gene loci, including coding and non-coding genes that were reviewed extensively in 2004. In this review, we present (1) an updated list of all the HLA gene symbols, gene names, expression status, Online Mendelian Inheritance in Man (OMIM) numbers, including new genes, and latest changes to gene names and symbols, (2) a regional analysis of the extended class I, class I, class III, class II and extended class II subregions, (3) a summary of the interspersed repeats (retrotransposons and transposons), (4) examples of the sequence diversity between different HLA haplotypes, (5) intra- and extra-HLA gene interactions and (6) some of the HLA gene expression profiles and HLA genes associated with autoimmune and infectious diseases. Overall, the degrees and types of HLA super-locus coordinated gene expression profiles and gene variations have yet to be fully elucidated, integrated and defined for the processes involved with normal cellular and tissue physiology, inflammatory and immune responses, and autoimmune and infectious diseases.
Journal of Human Genetics (2009) **54,** 15–39; doi:10.1038/jhg.2008.5; published online 9 January 2009

Keywords: human leukocyte antigen; major histocompatibility complex; polymorphism; genomic diversity; gene interaction; disease association; gene expression; cancer

INTRODUCTION

It is a decade since the first completely annotated and continuous human major histocompatibility complex (MHC) genomic sequence map was published.[1] The main purpose of the initial genomic sequences was to produce gene and genomic feature maps incorporating known and predicted gene loci. Since then, the MHC genomic sequence template has been used extensively to investigate single nucleotide polymorphism (SNP) and haplotype variation, gene expression, sequence diversity between and within species, and the evolution of the MHC structural organization.[2–8] The continuing strong interest in the MHC genomic sequence stems from its well-established role in regulating inflammation, the complement cascade and the innate and adaptive (acquired) immune responses using the natural killer (NK) and T-cell systems. The MHC locus contributes to restricted cellular interactions and tissue histocompatibility owing to the cellular discrimination of 'self' and 'non-self' that requires an essential knowledge of the effects of MHC-matched and -mismatched donors in transplantation medicine[9] and transfusion therapy.[10] Similarly, a fully annotated MHC genomic and diversity map is useful for understanding autoimmunity[11] and for charting the host response to infectious agents.[12,13] Apart from regulating immunity, the MHC genes may have a role in reproduction and social behavior, such as pregnancy maintenance, mate selection and kin recognition.[14,15] The MHC genomic region also appears to influence central nervous system (CNS) development and plasticity,[16–20] neurological cell interactions,[21,22] synaptic function and behavior,[23,24] cerebral hemispheric specialization,[25] and neurological and psychiatric disorders.[26–30]

The MHC region at ~4 Mb occupies 0.13% of the human genome (3×10^9 bp), but contains ~0.5% (>150) of the ~32 000 known protein coding genes. Many of the MHC gene products are ligands, receptors, interacting proteins, signaling factors and transcription regulators involved in the inflammatory response, antigen processing and presentation as part of the adaptive immune response, and interactions with NK cells and cytokines as part of the innate immune responses. The MHC genomic landscape is composed mainly of genes, retrotransposons, transposons, regulatory elements, pseudogenes and a few remaining undefined sequences. The MHC genomic region is one of the most gene-dense and best-defined regions within the

[1]Division of Basic Medical Science and Molecular Medicine, Department of Molecular Life Science, Tokai University School of Medicine, Isehara, Japan and [2]Centre for Forensic Science, The University of Western Australia, Nedlands, 6009, Western Australia, Australia
Correspondence: Dr T Shiina, Division of Basic Medical Science and Molecular Medicine, Department of Molecular Life Science, Tokai University School of Medicine, 143 Shimokasuya, Isehara 259-1143, Japan.
E-mail: tshiina@is.icc.u-tokai.ac.jp
Received 16 October 2008; revised 10 November 2008; accepted 12 November 2008; published online 9 January 2009

human genome, and the undefined sequences contribute to only a low percentage of the MHC region.

The human leukocyte antigen (HLA) is the name for the human MHC and we will use both names interchangeably in this overview, which outlines the HLA genomic loci, SNP and haplotype diversity, gene interactions and expression, and disease associations. This presentation complements other recent reviews on the human MHC architecture, duplications, diversity, disease and evolution.[5,6,14,31–33]

DEFINITION AND ANNOTATION OF GENE CLASSIFICATIONS

Table 1 is a summary of the latest (16 September 2008) locus information gathered on the genomic sequence of the HLA region providing the official gene and locus symbols, geneIDs, gene type, isoforms, mRNA and protein sequence accession numbers, and Online Mendelian Inheritance in Man (OMIM) identification numbers. The genomic sequence of the HLA region used for the present annotations is the PGF haplotype sequence[34] that was derived from a consanguineous HLA-homozygous cell line carrying the *HLA-A3*, *-B7*, *-Cw7*, *-DR15(DR2)* combination of alleles. This sequence is different from the original HLA virtual genomic sequence that was first reported[1] and reviewed[31] as a continuous, but mixed genomic sequence obtained from different haplotypes. The locus information in Table 1 is divided into five subregions from the telomeric to the centromeric end, the extended class I (*GABBR1* to *ZFP57*), class I (*HLA-F* to *MICB*), class III (*PPIAP9* to *BTNL2*), class II (*HLA-DRA* to *HLA-DPA3*) and the extended class II (*COL11A2* to *KIFC1*) regions. The definition of the extended class I and II regions is ambiguous, and we have included only four well-analyzed loci in the extended class I and 19 in the extended class II regions as shown in Table 1.

Locus information was assembled by using the Entrez Gene database (http://www.ncbi.nlm.nih.gov/sites/entrez) of the National Center for Biotechnology Information (NCBI) and previously published reports and papers.[1,35] The *Homo sapiens* official gene symbols and gene names of the MHC genomic region can be accessed by way of the 'GeneID' using Entrez Gene at NCBI.[36] Of the 224 loci mapped and reported by The MHC Sequencing Consortium in 1999,[1] more than half of them (124 loci per 224 loci) were replaced within 5 years with a new and official gene symbol and name approved by the HUGO Gene Nomenclature Committee (HGNC).[31] Since then, another 21 gene symbols and names have been changed. We have provided only one 'old symbol in 2004 and 2008' in Table 1, but many of the official gene symbols and names have alternate symbols and aliases. For example, the alternative symbols for *HLA-F* (GeneID 143110) are *DADB-68M4.2*, *CDA12*, *HLA-5.4*, *HLA-CDA12* and *HLAF*. There are 11 alternative names for the gene *DDR1* (GeneID 780). The old or alternative gene/locus names and symbols can also be accessed through the GeneID (Table 1) at NCBI.

The assembled loci in Table 1 were classified into four categories of gene status: 'protein coding,' 'gene candidate (candidate),' 'non-coding RNA (NC gene)' and 'pseudogene (pseudo).' The descriptor 'protein coding' means a gene that is transcribed to mRNA and also has a reliable open reading frame (ORF) and/or a known protein product, with the accession numbers for the mRNA and protein sequences provided. The 'gene candidate' is transcribed to mRNA (an mRNA sequence accession number is provided), but has an unknown or uncertain ORF. It may or may not have an accession number for a protein sequence listed. The 'NC gene' is transcribed to mRNA (accession number is provided), but does not have any ORF or a known protein or peptide product. The 'pseudo' is generally not transcribed to mRNA, and it may be a fragmented gene structure or a retrotransposed and unprocessed cDNA structure. Some of the pseudogenes, such as the *P5-1* family in Table 1, are known to be the remnants or hybrids of ancient endoretroviral sequences.[37] Interestingly, SNP variants for one of the members of the *P5-1* family, the gene locus *HCP5* located near *HLA-B*, have been strongly associated with the progression of HIV infection,[13,38] psoriasis vulgaris and psoriatic arthritis.[39]

GENE NUMBERS IN THE HLA REGION

A total of 253 loci have now been identified and/or reclassified in the 3.78 Mb HLA region of the PGF haplotype[34] from *BABBR1* located on the most telomeric side of the extended class I region to *KIFC1* (past name: *HSET*) located on the most centromeric side of the extended class II region (Figure 1 and Table 1). There are an additional 29 loci since the 224 loci were first identified in the HLA region and reported in 1999.[1] The locus numbers of *HLA-DRB* and *RP-C4-CYP21-TNX* subregions generated by gene duplication vary in number and reflect HLA haplotypic differences, as reported earlier.[1] When all the loci of the HLA complex were grouped into four categories of gene status, 133, 19, 22 and 79 loci were classified as protein coding, gene candidates, non-coding RNAs and pseudogenes, respectively. It is clear from Table 1 that the non-HLA genes greatly outnumber the HLA-like genes (HLA-class I, *MIC* and HLA-class II genes). Of the 45 HLA-like genes, 20 were identified as protein coding genes, 4 were NC genes and 21 were pseudogenes. Of the 208 non-HLA genes, 112 were identified as protein coding genes, 20 candidate genes, 18 NC genes and 58 pseudogenes.

Of the total number of 113 non-HLA protein coding genes, 9 (*SFTA2*, *MUC21*, *PSORS1C3*, *MCCD1*, *SLC44A4*, *ZBTB12*, *PRRT1*, *WDR46* and *PFDN6*) were newly identified to be functional loci (Tables 1 and 2). Of them, *PSORS1C3* is one of the associating genes of psoriasis vulgaris.[40] *MCCD1* encodes mitochondrial coiled-coil domain 1 and is highly polymorphic, containing approximately one SNP in every 99 basepairs.[41] *PFDN6* encodes prefoldin subunit 6, and the gene was reported to be overexpressed in certain cancers compared with normal counterparts in a tissue microarray study.[42]

Thirty-three of the non-HLA expressed genes (*GABBR1*, *MOG*, *ZNRD1*, *RNF39*, *TRIM10*, *TRIM39*, *PRR3*, *ABCF1*, *DDR1*, *CCHCR1*, *TCF19*, *POU5F1*, *BAT1*, *ATP6V1G2*, *LTB*, *LST1*, *AIF1*, *BAT3*, *MSH5*, *EHMT2*, *STK19*, *CYP21A2*, *TNXB*, *PPT2*, *AGPAT1*, *AGER*, *TAP2*, *PSMB8*, *PSMB9*, *BRD2*, *COL11A2*, *SLC39A7* and *TAPBP*) and HLA-F appear to express spliced variants with an overall average of 2.6 different kinds of spliced variants per gene. One of the recently identified expressed genes with a relatively large number of spliced variants is *C6orf25* that is located between *LY6G6C* and *DDAH2* within the class III region. This gene has at least seven spliced variants, and it is a member of the immunoglobulin (Ig) superfamily that encodes a glycosylated, plasma membrane-bound cell surface receptor as well as soluble isoforms. Some of the membrane-bound and soluble products encoded by the *C6orf25* splice variants contain two immunoreceptor tyrosine-based inhibitory motifs (ITIMs) that were found to interact by phosphorylation with the SH2-containing protein tyrosine phosphatases SHP-1 and SHP-2.[43]

REGIONAL ANALYSIS OF THE HLA SUPER-LOCUS

The HLA super-locus can be separated into the traditional five HLA regions with 4, 128, 75, 27 and 19 loci within the extended class I, class I, class III, class II and extended class II regions, respectively (Figure 1 and Table 2).

Extended class I region

In this version of the HLA loci, only four genes (*BABBR1*, *SUMO2P*, *MOG* and *ZNP57*) have been included in the extended class I region.

Table 1 Locus information in the HLA region (16 September 2008)

Official symbol	GeneID	Gene type	isoform	mRNA	Protein	function	OMIM	Old symbol*	Old symbol**	Note
EXTENDED CLASS I REGION										
GABBR1	2550	protein coding	a	NM_001470.2	NP_001461.1	o	603540			gamma-aminobutyric acid (GABA) B receptor, 1
			b	NM_021903.2	NP_068703.1					
			c	NM_021904.2	NP_068704.2					
			d	NM_021905.2	NP_068705.2					
SUMO2P	285829	pseudo		X	X	X				SMT3 suppressor of mif two 3 homolog 2 (S. cerevisiae)
MOG	4340	protein coding	a1	NM_206809.2	NP_996532.2	o	159465			myelin oligodendrocyte glycoprotein
			a2	NM_206812.2	NP_996535.2					
			a3	NM_001008228.1	NP_001008229.1					
			a4	NM_206814.3	NP_996537.2					
			b1	NM_002433.3	NP_002424.3					
			b2	NM_001008229.1	NP_001008230.1					
			b3	NM_206811.2	NP_996534.2					
			b4	NM_206813.3	NP_996536.2					
			b5	NM_206810.2	NP_996533.2					
ZFP57	346171	protein coding		NM_001109809.1	NP_001103279.1	o				zinc finger protein 57 homolog
CLASS I REGION										
HCP5P15	353021	pseudo		X	X	X		P5-15	HCP5P15	P5-1 pseudogene 15
HCG4P11	353020	pseudo		X	X	X		HCGIV-11	HCG4P11	HLA complex group 4 pseudogene 11
HLA-F	3134	protein coding	1	NM_001098479.1	NP_001091949.1	o	143110	HLA-F	HLA-F	major histocompatibility complex, class I, F
			2	NM_018950.2	NP_061823.2					
			3	NM_001098478.1	NP_001091948.1					
RPL23AP1	6148	pseudo		X	X	X			RPL23AP1	ribosomal protein L23a pseudogene 1
MICE	4280	pseudo		X	X	X		MICE	MICE	MHC class I polypeptide-related sequence E
HCG9P5	353019	pseudo		X	X	X		HCGIX-5	HCG9P5	HLA complex group 9 pseudogene 5
IFITM4P	340198	NC gene		NR_001590.1	X	X			IFITM4P	interferon induced transmembrane protein 4
3.8-1.5	353010	pseudo		X	X	X		3.8-1.5	3.8-1.5	3.8-1 pseudogene 5
HCP5P14	353018	pseudo		X	X	X		P5-14	HCP5P14	P5-1 pseudogene 14
HCG4P10	353017	pseudo		X	X	X		HCGIV-10	HCG4P10	HLA complex group 4 pseudogene 10
HLA-75	352962	pseudo		X	X	X		HLA-75	HLA-75	major histocompatibility complex, class I, pseudogene 75
HCG4	54435	NC gene		NR_002139.1	X	X		HCGIV-9	HCG4	HLA complex group 4
HCP5P13	353016	pseudo		X	X	X		P5-13	HCP5P13	P5-1 pseudogene 13
HLA-90	352963	pseudo		X	X	X		HLA-90	HLA-90	major histocompatibility complex, class I, pseudogene 90
HCG4P9	353014	pseudo		X	X	X		HCGIV-9	HCG4P9	HLA complex group 4 pseudogene 9
RPL7AP7	353013	pseudo		X	X	X		RPL7B	RPL7AP7	ribosomal protein L7a pseudogene 7
MICG	352967	pseudo		X	X	X			MICG	MHC class I polypeptide-related sequence G pseudogene
HCP2P8	353012	pseudo		X	X	X		HCGII-8	HCP2P8	HLA complex group 2 pseudogene 8
HCP5P12	353011	pseudo		X	X	X		P5-12	HCP5P12	P5-1 pseudogene 12
HCG4P8	353005	pseudo		X	X	X		HCGIV-8	HCG4P8	HLA complex group 4 pseudogene 8
P5-11	352989	pseudo		X	X	X		P5-11	P5-11	P5-1 pseudogene 11
HLA-G	3135	protein coding		NM_002127.4	NP_002118.1	O	142871	HLA-G	HLA-G	major histocompatibility complex, class I, G
LOC100133214	100133214	Candidate		XM_001718031.1	XP_001718083.1	?				similar to PAMP6501
MICF	352957	pseudo		X	X	X		MICF	MICF	MHC class I polypeptide-related sequence F pseudogene
3.8-1.4	353009	pseudo		X	X	X		3.8-1.4	3.8-1.4	3.8-1 pseudogene 4
HCP5P10	352990	pseudo		X	X	X		P5-10	HCP5P10	P5-1 pseudogene 10
HCG4P7	353004	pseudo		X	X	X		HCGIV-7	HCG4P7	HLA complex group 4 pseudogene 7
P5-09	352991	pseudo		X	X	X		P5-09	P5-09	P5-1 pseudogene 9
HLA-H	3136	NC gene		NR_001434.1	X	X		HLA-54	HLA-H	major histocompatibility complex, class I, pseudogene H

Table 1 Continued

P5-07	352992	pseudo		X	X	X		P5-07	P5-07	P5-1 pseudogene 7
HLA-16	352964	pseudo		X	X	X		HLA-16	HLA-16	major histocompatibility complex, class I, pseudogene 16
HCG2P7	80867	NC gene		NR_001318.1	X	X		HCGII-7	HCG2P7	HCGII pseudogene 7
3.8-1.3	353008	pseudo		X	X	X		3.8-1.3	3.8-1.3	3.8-1 pseudogene 3
HCG5P6	352993	pseudo		X	X	X		P5-06	HCG5P6	P5-1 pseudogene 6
HCG4P6	80868	NC gene		NR_001317.1	X	X		HCGIV-6	HCG4P6	HLA complex group 4 pseudogene 6
P5-05	352994	pseudo		X	X	X		P5-05	P5-05	P5-1 pseudogene 5
HLA-K	3138	pseudo		X	X	X		HLA-70	HLA-K	major histocompatibility complex, class I, K
HLA-U	352965	pseudo		X	X	X		HLA-21	HLA-21	major histocompatibility complex, class I, U
HCG4P5	353003	pseudo		X	X	X		HCGIV-5	HCG4P5	HLA complex group 4 pseudogene 5
P5-04	352996	pseudo		X	X	X		P5-04	P5-04	P5-1 pseudogene 4
HLA-A	3105	protein coding		NM_002116.5	NP_002107.3	O	142800	HLA-A	HLA-A	major histocompatibility complex, class I, A
HCP5P3	352997	pseudo		X	X	X		P5-03	HCP5P3	P5-1 pseudogene 3
HCG4P4	353002	pseudo		X	X	X		HCGIV-4	HCG4P4	HLA complex group 4 pseudogene 4
HLA-W	352966	pseudo		X	X	X		HLA-80	HLA-80	major histocompatibility complex, class I, 80
HCG2P6	353006	pseudo		X	X	X		HCGII-6	HCG2P6	HLA complex group 2 pseudogene 6
MICD	4279	pseudo		X	X	X		MICD	MICD	MHC class I polypeptide-related sequence D pseudogene
HCG9	10255	Candidate		NM_005844.2	NP_005835.2	?		HCGIX-4	HCG9	HLA complex group 9
3.8-1.2	353007	pseudo		X	X	X		3.8-1.2	3.8-1.2	3.8-1 pseudogene 3
HCP5P2	352998	pseudo		X	X	X		P5-02	HCP5P2	P5-1 pseudogene 2
HCG4P3	353001	pseudo		X	X	X		HCGIV-3	HCG4P3	HLA complex group 4 pseudogene 3
HLA-J	3137	pseudo		X	X	X		HLA-59	HLA-J	major histocompatibility complex, class I, J
HCG8	80869	NC gene		XR_041146.1	X	X		HCGVIII-1	HCG8	HLA complex group 8 pseudogene 1
ETF1P1	6824	Pseudogene		X	X	X			ETF1P1	eukaryotic translation termination factor 1 pseudogene 1
C6orf12	80862	NC gene		XR_041144.1	X	X		HTEX4	C6orf12	chromosome 6 open reading frame 12
ZNRD1	30834	protein coding	1	NM_014596.4	NP_055411.1	O	607525		ZNRD1	zinc ribbon domain containing 1
			2	NM_170783.2	NP_740753.1					
PPP1R11	6992	protein coding		NM_021959.2	NP_068778.1	O	606670	HCG5	PPP1R11	protein phosphatase 1, regulatory (inhibitor) subunit 11
RNF39	80352	protein coding	1	NM_025236.2	NP_079512.1	O	607524		RNF39	ring finger protein 39
			2	NM_170769.1	NP_739575.1					
TRIM31	11074	protein coding		NM_007028.3	NP_008959.3	O	609316	HCGI	TRIM31	tripartite motif-containing 31
TRIM40	135644	protein coding		NM_138700.3	NP_619645.1	O			TRIM40	tripartite motif-containing 40
TRIM10	10107	protein coding	1	NM_006778.3	NP_006769.2	O	605701	RFB30	TRIM10	tripartite motif-containing 10
			2	NM_052828.2	NP_439893.2					
TRIM15	89870	protein coding		NM_033229.2	NP_150232.2	O		ZNFB7	TRIM15	tripartite motif-containing 15
TRIM26	7726	protein coding		NM_003449.3	NP_003440.1	O	600830	ZNF173	TRIM26	tripartite motif-containing 26
HLA-L	3139	pseudo		X	X	X		HLA-92	HLA-L	major histocompatibility complex, class I, L
FLJ45422	441140	Candidate		NM_001004349.1	NP_001004349.1	?				FLJ45422 protein
LOC100133303	100133303	Candidate		XM_001717921.1	XP_001717973.1	?				hypothetical protein LOC100133303
TRIM39	56658	protein coding	1	NM_021253.2	NP_067076.2	O	605700		TRIM39	tripartite motif-containing 39
			2	NM_172016.1	NP_742013.1					
RPP21	79897	protein coding		NM_024839.1	NP_079115.1	O			RPP21	ribonuclease P/MRP 21kDa subunit
HLA-N	267014	pseudo		X	X	X		HLA-30	HLA-N	major histocompatibility complex, class I, 30
HCG2P5		pseudo		X	X	X		HCGII-5	HCG2P5	HLA complex group 2 pseudogene 5
HCG2P4	387501	pseudo		X	X	X		HCGII-4	HCG2P4	HLA complex group 2 pseudogene 4
MICC	221549	pseudo		X	X	X		MICC	MICC	MHC class I polypeptide-related sequence C
HCG2P3		pseudo		X	X	X		HCGII-3	HCG2P3	HLA complex group 2 pseudogene 3
RANP1	221547	pseudo		X	X	X		TC4	RANP1	RAN, member RAS oncogene family pseudogene 1

Table 1 Continued

HLA-E	3133	protein coding		NM_005516.4	NP_005507.3	O	143010	HLA-E	HLA-E	major histocompatibility complex, class I, E
GNL1	2794	protein coding		NM_005275.2	NP_005266.2	O	143024	HSR1	GNL1	guanine nucleotide binding protein-like 1
PRR3	80742	protein coding	a	NM_025263.2	NP_079539.2	O		CAT56	PRR3	proline-rich polypeptide 3
			b	NM_001077497.1	NP_001070965.1					
ABCF1	23	protein coding	a	NM_001025091.1	NP_001020262.1	O	603429	ABC50	ABCF1	ATP-binding cassette, sub-family F (GCN20), member 1
			b	NM_001090.2	NP_001081.1					
PPP1R10	5514	protein coding		NM_002714.2	NP_002705.2	O	603771	FB19	PPP1R10	protein phosphatase 1, regulatory (inhibitor) subunit 10
MRPS18B	28973	protein coding		NM_014046.3	NP_054765.1	O	611982		MRPS18B	mitochondrial ribosomal protein S18B
PTMAP1	5758	pseudo		X	X	X		PROA-hom	PROAP	prothymosin, alpha pseudogene 1
C6orf134	79969	Candidate	1	NM_145029.2	NP_659466.1	?			C6orf134	chromosome 6 open reading frame 134
			2	NM_024909.2	NP_079185.2					
C6orf136	221545	Candidate	1	NM_001109938.1	NP_001103408.1	?			C6orf136	chromosome 6 open reading frame 136
			2	NM_145029.2	NP_659466.1					
DHX16	8449	protein coding		NM_003587.3	NP_003578.1	O	603405	DBP2	DHX16	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 16
KIAA1949	170954	Candidate		NP_003578.1	NP_597728.1	?	610990		KIAA1949	KIAA1949 protein
NRM	11270	protein coding		NM_007243.1	NP_009174.1	O			NRM	nurim (nuclear envelope membrane protein)
RPL7AP	100133037	pseudo		X	X	X		RPL7A	RPL7AP	ribosomal protein L7 pseudogene
MDC1	9656	protein coding		NM_014641.2	NP_055456.2	O	607593	KIAA0170	MDC1	mediator of DNA damage checkpoint 1
TUBB	203068	protein coding		NM_178014.2	NP_821133.1	O	191130	TUBB	OK/SW-cl.56	tubulin, beta
FLOT1	10211	protein coding		NM_005803.2	NP_005794.1	O	606998	FLOTILLIN	FLOT1	flotillin 1
IER3	8870	protein coding		NM_003897.3	NP_003888.2	O	602996	PRG1	IER3	immediate early response 3
DDR1	780	protein coding	a	NM_013993.2	NP_054699.2	O	600408	DDR	DDR1	discoidin domain receptor tyrosine kinase 1
			b	NM_001954.4	NP_001945.3					
			c	NM_013994.2	NP_054700.2					
GTF2H4	2968	protein coding		NM_001517.4	NP_001508.1	O	601760	TFIIH	GTF2H4	eneral transcription factor IIH, polypeptide 4, 52kDa
VARS2	57176	protein coding		NM_020442.3	NP_065175.3	O			VARS2L	valyl-tRNA synthetase 2, mitochondrial
SFTA2	389376	protein coding		NM_205854.2	NP_995326.1	O			LOC389376	surfactant associated 2
LOC100129065	100129065	Candidate		XM_001723513.1	XP_001723565.1	?				similar to hCG2045728
DPCR1	135656	protein coding		NM_080870.2	NP_543146.1	O	604809		DPCR1	diffuse panbronchiolitis critical region 1
MUC21	135656	protein coding		NM_080870.2	NP_543146.1	O			C6orf205	mucin 21, cell surface associated
LOC729792	729792	Candidate		XM_001131329.1	XP_001131329.1	?				hypothetical LOC729792
HCG22	285834	NC gene		NR_003948.1	X	X				HLA complex group 22
C6orf15	29113	Candidate		NM_014070.2	NP_054789.2	?	611401		C6orf15	chromosome 6 open reading frame 15
CDSN	1041	protein coding		NM_001264.3	NP_001255.3	O	602593	S	CDSN	corneodesmosin
PSORS1C1	170679	protein coding		NM_014068.1	NP_054787.1	O			PSORS1C1	psoriasis susceptibility 1 candidate 1
PSORS1C2	170680	protein coding		NM_014069.2	NP_054788.2	O			PSORS1C2	psoriasis susceptibility 1 candidate 2
LOC100129610	100129610	pseudo		X	X	X				hypothetical LOC100129610
CCHCR1	54535	protein coding	1	NM_001105564.1	NP_001099034.1	O	605310	PG8	C6orf18	coiled-coil alpha-helical rod protein 1
			2	NM_001105563.1	NP_001099033.1					
			3	NM_019052.3	NP_061925.2					
TCF19	6941	protein coding	1	NM_001077511.1	NP_001070979.1	O	600912	SC1	TCF19	transcription factor 19 (SC1)
			2	NM_007109.2	NP_009040.2					
POU5F1	5460	protein coding	1	NM_002701.4	NP_002692.2	O	164177	OTF3	POU5F1	POU domain, class 5, transcription factor 1
			2	NM_203289.3	NP_976034.3					
PSORS1C3	100130889	protein coding		NM_001134284.1	NP_001127756.1	O				psoriasis susceptibility 1 candidate 3
HCG27	253018	Candidate		NM_181717.2	NP_859068.2	?			LOC253018	HLA complex group 27
HCG2P2	387502	pseudo		X	X	X		HCGII-2	HCG2P2	HLA complex group 2 pseudogene 2
HCG9P3	387507	pseudo		X	X	X		HCGIX-3	HCG9P3	HLA complex group 9 pseudogene 3

Table 1 Continued

HLA-C	3107	protein coding		NM_002117.4	NP_002108.4	O	142840	HLA-C	HLA-C	major histocompatibility complex, class I, C
HCG4P2	387504	pseudo		X	X	X		HCGIV-2	HCG4P2	HLA complex group 4 pseudogene 3
KIAA0055P		pseudo		X	X	X		KIAA0055-hom	KIAA0055P	KIAA0055 pseudogene
RPL3P		pseudo		X	X	X		RPL3-hom	RPL3P	ribosomal protein L3 pseudogene
HCG2P1	387484	pseudo		X	X	X		HCGII-1	HCG2P1	HLA complex group 2 pseudogene 1
HLA-B	3106	protein coding		NM_005514.6	NP_005505.2	O	142830	HLA-B	HLA-B	major histocompatibility complex, class I, B
HCG4P1	387503	pseudo		X	X	X		HCGIV-1	HCG4P1	HLA complex group 4 pseudogene 3
DHFRP2	729816	NC gene		XR_042352.1	X	X		DHFRP	DHFRP	dihydrofolate reductase (DHFR) pseudogene
HLA-S	267015	pseudo		X	X	X		HLA-17	HLA-S	major histocompatibility complex, class I, 17
HCP5P8	387508	pseudo		X	X	X		P5-08	HCP5P8	P5-1 pseudogene 8
HCG9P2	387506	pseudo		X	X	X		HCGIX-2	HCG9P2	HLA complex group 9 pseudogene 2
MICA	4276	protein coding		NM_000247.1	NP_000238.1	O	600169	MICA	MICA	MHC class I polypeptide-related sequence A
HLA-X	267016	pseudo		X	X	X		HLA-X	HLA-X	major histocompatibility complex, class I, X
HCP5	10866	protein coding		NM_006674.2	NP_006665.2	?	604676	P5-1	HCP5	HLA complex P5
3.8-1	352961	NC gene		NR_002812.2	X	X		3.8-1	3.8-1	MHC class I mRNA fragment 3.8-1
HCG9P1	387505	pseudo		X	X	X		HCGIX-1	HCG9P1	HLA complex group 9 pseudogene 1
MICB	4277	protein coding		NM_005931.3	NP_005922.2	O	602436	MICB	MICB	MHC class I polypeptide-related sequence B
CLASS III REGION										
PPIAP9	5491	pseudo		X	X	X			PPIP9	peptidylprolyl isomerase A (cyclophilin A) pseudogene 9
LOC100129921	100129921	pseudo		X	X	X				similar to rCG64241
MCCD1	401250	protein coding		NM_001011700.2	NP_001011700.2	O	609624		LOC401250	mitochondrial coiled-coil domain 1
BAT1	7919	protein coding	1	NM_004640.5	NP_004631.1	O	142560	BAT1	BAT1	HLA-B associated transcript 1
			2	NM_080598.4	NP_542165.1					
SNORD117	692233	snoRNA		NR_003140.1	X	X				small nucleolar RNA, C/D box 117
SNORD84	692199	snoRNA		NR_003065.1	X	X				small nucleolar RNA, C/D box 84
ATP6V1G2	534	protein coding	a	NM_130463.2	NP_569730.1	O	606853	ATP6G	ATP6V1G2	ATPase, H+ transporting, lysosomal 13kDa, V1 subunit G isofor
			b	NM_138282.1	NP_612139.1					
NFKBIL1	4795	protein coding		NM_005007.2	NP_004998.2	O	601022	NFKBIL1	NFKBIL1	nuclear factor of kappa light polypeptide gene enhancer
LTA	4049	protein coding		NM_000595.2	NP_000586.2	O	153440	LTA	LTA	lymphotoxin alpha (TNF superfamily, member 1)
TNF	7124	protein coding		NM_000594.2	NP_000585.2	O	191160	TNF-alpha	TNF	tumor necrosis factor (TNF superfamily, member 2)
LTB	4050	protein coding	a	NM_002341.1	NP_002332.1	O	600978	LTB	LTB	lymphotoxin beta (TNF superfamily, member 3)
			b	NM_009588.1	NP_033666.1					
LST1	7940	protein coding	1	NM_007161.2	NP_009092.2	O	109170	LST1	LST1	leukocyte specific transcript 1
			2	NM_205837.1	NP_995309.1					
			3	NM_205838.1	NP_995310.1					
			4	NM_205839.1	NP_995311.1					
			5	NM_205840.1	NP_995312.1					
NCR3	259197	protein coding		NM_147130.1	NP_667341.1	O	611550	1C7	NCR3	natural cytotoxicity triggering receptor 3
LOC100130756	100130756	pseudo		X	X	X				hypothetical LOC100130756
AIF1	199	protein coding	1	NM_001623.3	NP_001614.3	O	601833	AIF1	AIF1	allograft inflammatory factor 1
			3	NM_032955.1	NP_116573.1					
BAT2	7916	protein coding		NM_080686.2	NP_542417.2	O	142580	BAT2	BAT2	HLA-B associated transcript 2
SNORA38	677820	snoRNA		NR_002971.1	X	X				small nucleolar RNA, H/ACA box 38
BAT3	7917	protein coding	1	NM_001098534.1	NP_001092004.1	O	142590	BAT3	BAT3	HLA-B associated transcript 3
			2	NM_004639.3	NP_004630.3					
			3	NM_080702.2	NP_542433.1					
			4	NM_080703.2	NP_542434.1					
APOM	55937	protein coding		NM_019101.2	NP_061974.2	O	606907	ApoM	APOM	apolipoprotein M
C6orf47	57827	Candidate		NM_021184.3	NP_067007.3	?		G4	C6orf47	chromosome 6 open reading frame 47

Table 1 Continued

C6orf47	57827	Candidate		NM_021184.3	NP_067007.3	?		G4	C6orf47	chromosome 6 open reading frame 47
BAT4	7918	protein coding		NM_033177.2	NP_149417.1	O	142610	BAT4	BAT4	HLA-B associated transcript 4
CSNK2B	1460	protein coding		NM_001320.5	NP_001311.3	O	115441	CSK2B	CSNK2B	casein kinase 2, beta polypeptide
LY6G5B	58496	protein coding		NM_021221.2	NP_067044.2	O	610433	G5b	LY6G5B	lymphocyte antigen 6 complex, locus G5B
LY6G5C	80741	protein coding		NM_025262.2	NP_079538.2	O	610434	G5c	LY6G5C	lymphocyte antigen 6 complex, locus G5C
BAT5	7920	protein coding		NM_021160.1	NP_066983.1	O	142620	NG26	BAT5	HLA-B associated transcript 5
LY6G6F	259215	protein coding		NM_001003693.1	NP_001003693.1	O	611404			lymphocyte antigen 6 complex, locus G6F
LY6G6E	79136	NC gene		NR_003673.1	X	X	610437	G6E	LY6G6E	lymphocyte antigen 6 complex, locus G6E
LY6G6D	58530	protein coding		NM_021246.2	NP_067069.2	O	606038	G6D	LY6G6D	lymphocyte antigen 6 complex, locus G6D
LY6G6C	80740	protein coding		NM_025261.1	NP_079537.1	O	610435	G6C	LY6G6C	lymphocyte antigen 6 complex, locus G6C
C6orf25	80739	Candidate	A	NM_025260.2	NP_079536.2	O	606520	G6b	C6orf25	chromosome 6 open reading frame 25
			B	NM_138272.1	NP_612116.1					
			C	NM_138273.1	NP_612117.1					
			D	NM_138274.1	NP_612118.1					
			E	NM_138275.1	NP_612119.1					
			G	NM_138277.1	NP_612121.1					
DDAH2	23564	protein coding		NM_013974.1	NP_039268.1	O	604744	DDAH	DDAH2	dimethylarginine dimethylaminohydrolase 2
CLIC1	1192	protein coding		NM_001288.4	NP_001279.2	O	602872	CLIC1	CLIC1	chloride intracellular channel 1
MSH5	4439	protein coding	1	NM_025259.4	NP_079535.3	O	603382	MSH5	MSH5	mutS homolog 5 (E. coli)
			2	NM_172165.2	NP_751897.1					
			3	NM_002441.3	NP_002432.1					
			4	NM_172166.2	NP_751898.1					
C6orf26	401251	Candidate		NM_001039651.1	NP_001034740.1	?				chromosome 6 open reading frame 26
C6orf27	80737	Candidate		NM_025258.2	NP_079534.2	?	609693	G7c	C6orf27	chromosome 6 open reading frame 27
VARS	7407	protein coding		NM_006295.2	NP_006286.1	O	604137	Val-TRS	VARS2	valyl-tRNA synthetase
LSM2	57819	protein coding		NM_021177.3	NP_067000.1	O	607282	SMRNP	SMRNP	LSM2 homolog, U6 small nuclear RNA associated (S. cerevisiae)
HSPA1L	3305	protein coding		NM_005527.3	NP_005518.3	O	140559	HSPA1L	HSPA1L	heat shock 70kDa protein 1-like
HSPA1A	3303	protein coding		NM_005345.5	NP_005336.3	O	140550	HSPA1A	HSPA1A	heat shock 70kDa protein 1A
HSPA1B	3304	protein coding		NM_005346.4	NP_005337.2	O	603012	HSPA1B	HSPA1B	heat shock 70kDa protein 1B
C6orf48	50854	Candidate	1	NM_001040437.1	NP_001035527.1	?	605447	G8	C6orf48	chromosome 6 open reading frame 48
			2	NM_001040438.1	NP_001035528.1					
SNORD48	26801	snoRNA		NR_002745.1	X	X				small nucleolar RNA, C/D box 48
SNORD52	26797	snoRNA		NR_002742.1	X	X				small nucleolar RNA, C/D box 52
NEU1	4758	protein coding		NM_000434.3	NP_000425.1	O	608272	NEU	NEU1	sialidase 1 (lysosomal sialidase)
SLC44A4	80736	protein coding		NM_025257.2	NP_079533.2	O	606107	NG22	C6orf29	solute carrier family 44, member 4
LOC100128067	100128067	Candidate		XM_001718035.1	XP_001718087.1	?				hypothetical protein LOC100128067
EHMT2	10919	protein coding	a	NM_006709.3	NP_006700.3	O	604599	G9a	BAT8	euchromatic histone-lysine N-methyltransferase 2
			b	NM_025256.5	NP_079532.5					
ZBTB12	221527	protein coding		NM_181842.2	NP_862825.1	O		G10	C6orf46	zinc finger and BTB domain containing 12
C2	717	protein coding		NM_000063.3	NP_000054.2	O	217000	C2	C2	complement component 2
CFB	629	protein coding		NM_001710.5	NP_001701.2	O	138470	BF	BF	complement factor B
RDBP	7936	protein coding		NM_002904.5	NP_002895.3	O	154040	RD	RDBP	RD RNA binding protein
SKIV2L	6499	protein coding		NM_006929.4	NP_008860.4	O	600478	SKI2W	SKIV2L	superkiller viralicidic activity 2-like (S. cerevisiae)
DOM3Z	1797	protein coding		NM_005510.3	NP_005501.2	O	605996	DOM3L	DOM3Z	dom-3 homolog Z (C. elegans)
STK19	8859	protein coding	1	NM_004197.1	NP_004188.1	O	604977	STK19	STK19	serine/threonine kinase 19
			2	NM_032454.1	NP_115830.1					
C4B	721	protein coding		NM_001002029.3	NP_001002029.3	O	120820	C4B	C4B	complement component 4B (Rodgers blood group)
CYP21A1P	1590	pseudo		X	X	X				cytochrome P450, family 21, subfamily A, polypeptide 1 pseudogene
TNXA	7146	NC gene		NR_001284.2	X	X				tenascin XA pseudogene
STK19P	373159	pseudo		X	X	X				serine/threonine kinase 19 pseudogene

Table 1 Continued

C4A	720	protein coding		NM_007293.2	NP_009224.2	O	120810			complement component 4A (Rodgers blood group)
CYP21A2	1589	protein coding	a	NM_000500.5	NP_000491.2	O	201910	P450-C21B	CYP21A2	cytochrome P450, family 21, subfamily A, polypeptide 2
			b	NM_001128590.1	NP_001122062.1					
TNXB	7148	protein coding	1	NM_019105.6	NP_061978.6	O	600985	TNXB	TNXB	tenascin XB
			2	NM_032470.3	NP_115859.2					
CREBL1	1388	protein coding		NM_004381.3	NP_004372.3	O	600984	CREBL1	CREBL1	cAMP responsive element binding protein-like 1
FKBPL	63943	protein coding		NM_022110.3	NP_071393.2	O		NG7	FKBPL	FK506 binding protein like
PRRT1	80863	protein coding		NM_030651.3	NP_085154.3	O		NG5	C6orf31	proline-rich transmembrane protein 1
PPT2	9374	protein coding	a	NM_005155.5	NP_005146.3	O	603298	PPT2	PPT2	palmitoyl-protein thioesterase 2
			b	NM_138717.1	NP_619731.1					
EGFL8	80864	protein coding		NM_030652.2	NP_085155.1	O	609897	NG3	EGFL8	EGF-like-domain, multiple 8
AGPAT1	10554	protein coding	1	NM_006411.2	NP_006402.1	O	603099	LPAAT	AGPAT1	1-acylglycerol-3-phosphate O-acyltransferase 1
			2	NM_032741.3	NP_116130.2					
RNF5	6048	protein coding		NM_006913.3	NP_008844.1	O	602677	G16	RNF5	ring finger protein 5
AGER	177	protein coding	1	NM_001136.3	NP_001127.1	O	600214	RAGE	AGER	advanced glycosylation end product-specific receptor
			2	NM_172197.1	NP_751947.1					
PBX2	5089	protein coding		NM_002586.4	NP_002577.2	O	176311	PBX2	PBX2	pre-B-cell leukemia transcription factor 2
GPSM3	63940	protein coding		NM_022107.1	NP_071390	O		G18	GPSM3	G-protein signaling modulator 3 (AGS3-like, C. elegans)
NOTCH4	4855	protein coding		NM_004557.3	NP_004548.3	O	164951	NOTCH4	NOTCH4	Notch homolog 4 (Drosophila)
C6orf10	10665	Candidate		NM_006781.3	NP_006772.3	?		C6orf10	C6orf10	chromosome 6 open reading frame 10
LOC100131609	100131609	NC gene		XR_039222.1	X	X				similar to rCG55925
BTNL2	56244	protein coding		NM_019602.1	NP_062548.1	O	606000	TSBP	BTNL2	butyrophilin-like 2 (MHC class II associated)
CLASS II REGION										
HLA-DRA	3122	protein coding		NM_019111.3	NP_061984.2	O	142860	HLA-DRA	HLA-DRA	major histocompatibility complex, class II, DR alpha
HLA-DRB9	3132	pseudo		X	X	X		HLA-DRB9	HLA-DRB9	major histocompatibility complex, class II, DR beta 9
HLA-DRB5	3127	protein coding		NM_002125.3	NP_002116.2	O	604776			major histocompatibility complex, class II, DR beta 5
HLA-DRB6	3128	NC gene		NR_001298.1	X	X				major histocompatibility complex, class II, DR beta 6
HLA-DRB1	3123	protein coding		NM_002124.2	NP_002115.2	O	142857	HLA-DRB1	HLA-DRB1	major histocompatibility complex, class II, DR beta 1
HLA-DQA1	3117	protein coding		NM_002122.3	NP_002113.2	O	146880	HLA-DQA1	HLA-DQA1	major histocompatibility complex, class II, DQ alpha 1
HLA-DQB1	3119	protein coding		NM_002123.3	NP_002114.3	O	604305	HLA-DQB1	HLA-DQB1	major histocompatibility complex, class II, DQ beta 1
HLA-DQB3	3121	pseudo		X	X	X		HLA-DQB3	HLA-DQB3	major histocompatibility complex, class II, DQ beta 3
HLA-DQA2	3118	protein coding		NM_020056.2	NP_064440.1	O		HLA-DQA2	HLA-DQA2	major histocompatibility complex, class II, DQ alpha 2
HLA-DQB2	3120	NC gene		NR_003937.1	X	X		HLA-DQB2	HLA-DQB2	major histocompatibility complex, class II, DQ beta 2
HLA-DOB	3112	protein coding		NM_002120.3	NP_002111.1	O	600629	HLA-DOB	HLA-DOB	major histocompatibility complex, class II, DO beta
TAP2	6891	protein coding	1	NM_000544.3	NP_000535.3	O	170261	TAP2	TAP2	transporter 2, ATP-binding cassette, sub-family B (MDR/TAP)
			2	NM_018833.2	NP_061313.2					
PSMB8	5696	protein coding	E1	NM_004159.4	NP_004150.1	O	177046	LMP7	PSMB8	proteasome subunit, beta type, 8 (large multifunctional protease 7)
			E2	NM_148919.3	NP_683720.2					
TAP1	6890	protein coding		NM_000593.5	NP_000584.2	O	170260	TAP1	TAP1	transporter 1, ATP-binding cassette, sub-family B (MDR/TAP)
PSMB9	5698	protein coding	1	NM_002800.4	NP_002791.1	O	177045	LMP2	PSMB9	proteasome subunit, beta type, 9 (large multifunctional protease 2)
			2	NM_148954.2	NP_683756.1					
PPP1R2P1	5505	pseudo		X	X	X		IPP2	PPP1R2P1	protein phosphatase 1, regulatory (inhibitor) subunit 2 pseudogene 1
HLA-Z	267017	pseudo		X	X	X		HLA-Z1	HLA-Z	Class I gene fragment
HLA-DMB	3109	protein coding		NM_002118.3	NP_002109.1	O	142856	HLA-DMB	HLA-DMB	major histocompatibility complex, class II, DM beta
HLA-DMA	3108	protein coding		NM_006120.2	NP_006111.2	O	142855	HLA-DMA	HLA-DMA	major histocompatibility complex, class II, DM alpha
BRD2	6046	protein coding	1	NM_005104.3	NP_005095.1	O	601540	RING3	BRD2	bromodomain containing 2
			2	NM_001113182.1	NP_001106653.1					
HLA-DOA	3111	protein coding		NM_002119.3	NP_002110.1	O	142930	HLA-DOA	HLA-DOA	major histocompatibility complex, class II, DO alpha
HLA-DPA1	3113	protein coding		NM_033554.2	NP_291032.2	O	142880	HLA-DPA1	HLA-DPA1	major histocompatibility complex, class II, DP alpha 1
HLA-DPB1	3115	protein coding		NM_002121.4	NP_002112.3	O	142858	HLA-DPB1	HLA-DPB1	major histocompatibility complex, class II, DP beta 1

Table 1 Continued

RPL32P1	6163	pseudo		X	X	X		RPL32-L	RPL32P1	ribosomal protein L32 pseudogene 1
HLA-DPA2	3114	pseudo		X	X	X		HLA-DPA2	HLA-DPA2	major histocompatibility complex, class II, DP alpha 2
HLA-DPB2	3116	NC gene		NR_001435.1	X	X		HLA-DPB2	HLA-DPB2	major histocompatibility complex, class II, DP beta 2
HLA-DPA3		pseudo		X	X	X		HLA-DPA3	HLA-DPA3	major histocompatibility complex, class II, DP alpha 3
EXTENDED CLASS II REGION										
COL11A2	1302	protein coding	1	NM_080681.2	NP_542412.2	O	120290	COL11A2	COL11A2	collagen, type XI, alpha 2
			2	NM_080680.2	NP_542411.2					
			3	NM_080679.2	NP_542410.2					
RXRB	6257	protein coding		NM_021976.3	NP_068811.1	O	180246	RXRB	RXRB	retinoid X receptor, beta
SLC39A7	7922	protein coding	1	NM_001077516.1	NP_001070984.1	O	601416	RING5	SLC39A7	solute carrier family 39 (zinc transporter), member 7
			2	NM_006979.2	NP_008910.2					
HSD17B8	7923	protein coding		NM_014234.3	NP_055049.1	O	601417	RING2	HSD17B8	hydroxysteroid (17-beta) dehydrogenase 8
RING1	6015	protein coding		NM_002931.3	NP_002922.2	O	602045	RING1	RING1	ring finger protein 1
VPS52	6293	protein coding		NM_022553.4	NP_072047.4	O	603443	HSACM2L	VPS52	vacuolar protein sorting 52 (yeast)
RPS18	6222	protein coding		NM_022551.2	NP_072045.1	O	180473	RPS18	RPS18	ribosomal protein S18
B3GALT4	8705	protein coding		NM_003782.3	NP_003773.1	O	603095	B3GALT4	B3GALT4	UDP-Gal:betaGlcNAc beta 1,3-galactosyltransferase, polypeptide 4
WDR46	9277	protein coding		NM_005452.4	NP_005443.2	O		BING4	C6orf11	WD repeat domain 46
D6S2723E	10470	pseudo		X	X	X		BING5	D6S2723E	unknown
PFDN6	10471	protein coding		NM_014260.2	NP_055075.1	O	605660	HKE2	HKE2	prefoldin subunit 6
RGL2	5863	protein coding		NM_004761.2	NP_004752.1	O	602306	RAB2L	RGL2	RAB2, member RAS oncogene family-like
TAPBP	6892	protein coding	1	NM_003190.3	NP_003181.3	O	601962	TAPBP	TAPBP	TAP binding protein (tapasin)
			2	NM_172208.1	NP_757345.1					
			3	NM_172209.1	NP_757346.1					
ZBTB22	9278	protein coding		NM_005453.3	NP_005444.3	O		BING1	ZNF297	zinc finger and BTB domain containing 22
DAXX	1616	protein coding		NM_001350.3	NP_001341.1	O	603186	DAXX	DAXX	death-domain associated protein
LOC646720	646720	Candidate		XM_933843.2	XP_938936.1	?				similar to mCG56376
MYL8P	442204	pseudo		X	X	X				myosin, light chain 8, pseudogene
LYPLA2P1	285840	NC gene		NR_001444.3	X	X			LYPLA2P1	lysophospholipase II pseudogene 1
KIFC1	3833	protein coding		NM_002263.3	NP_002254.2	O	603763	HSET	KIFC1	kinesin family member C1

Abbreviation: HLA, human leukocyte antigen.
White background and black letters, light gray background and black letters, deep gray background and white letters and black background and white letters indicate 'protein coding gene,' 'gene candidate,' 'non-coding or small RNA (NC RNA or snoRNA)' and 'pseudogene,' respectively. GeneID shows 'NCBI gene ID'. The mRNA and protein columns show the accession numbers of GenBank. In the function column, 'O' is a previously known functional gene, '?' is an unknown or inferred functional gene and 'X' is a non-functional gene. The old symbol columns show the gene symbols reported in [a]1999 and [b]2004.

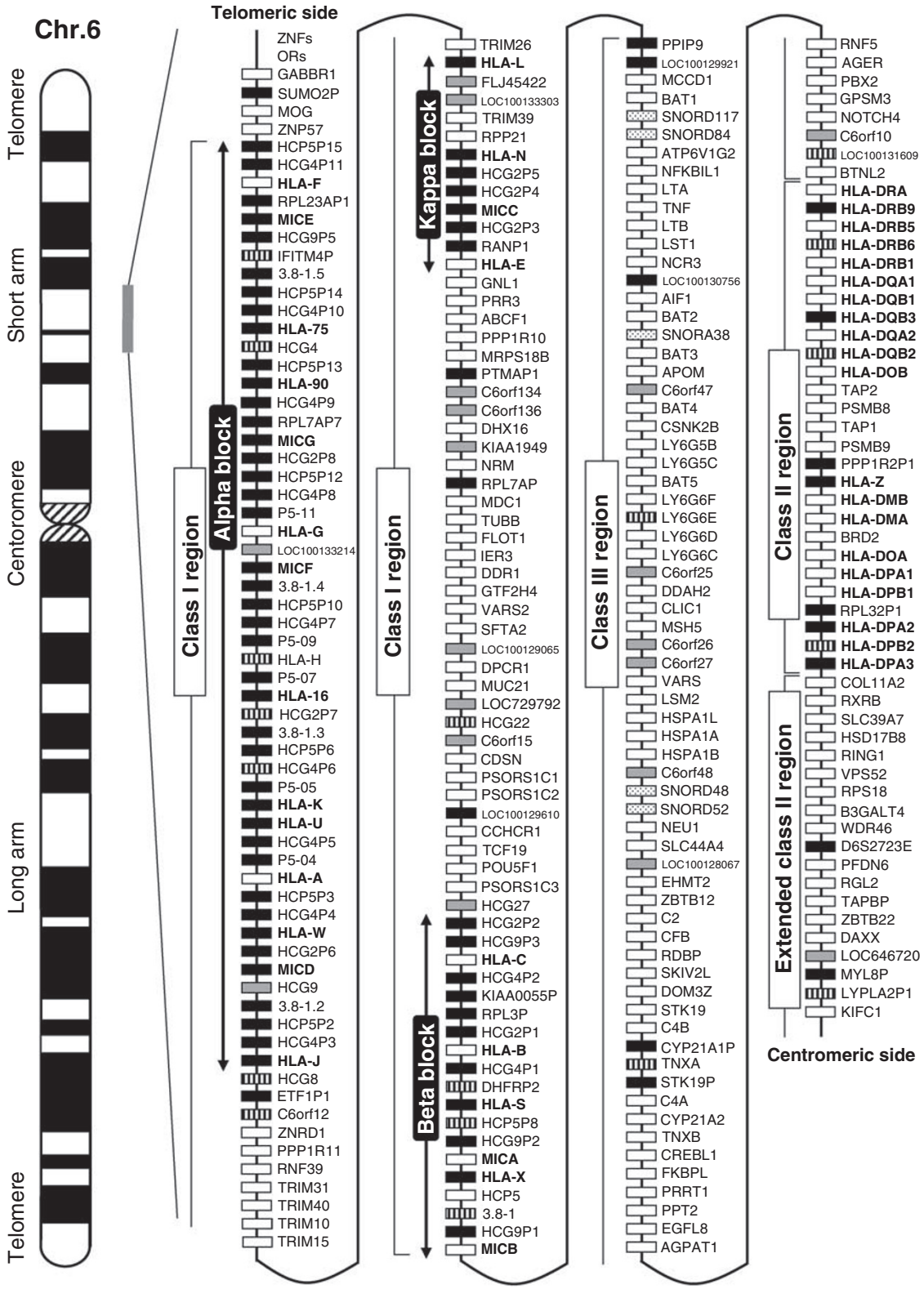

Figure 1 Gene map of the human leukocyte antigen (HLA) region. The major histocompatibility complex (MHC) gene map corresponds to the genomic coordinates of 29 677 984 (*GABBR1*) to 33 485 635 (*KIFC1*) in the human genome build 36.3 of the National Center for Biotechnology Information (NCBI) map viewer. The regions separated by arrows show the HLA subregions such as extended class I, classical class I, class III, classical class II and extended class II regions from telomere (left and top side) to centromere (right and bottom side). White, gray, striped and black boxes show expressed genes, gene candidates, non-coding genes and pseudogenes, respectively. The location of the alpha, beta and kappa blocks containing the cluster of duplicated HLA class I genes in the class I region are indicated.

However, numerous duplicated genes encoding the olfactory receptor, histone, tRNA and zinc-finger protein are located on the telomeric segment of the extended class I region. The hemochromatosis gene (*HFE*) that is similar in structure to an HLA class I gene is located outside the HLA super-locus ~3.6 Mb away on the telomeric side of *HLA-F* and the extended class I region.[44]

Class I region

The class I region contains the six classical and non-classical HLA class I genes. The non-classical HLA class I genes are differentiated from the classical class I genes on the basis that they have limited polymorphism; the tissue distribution of gene expression is restricted and they appear to play a less well-defined role in transplantation medicine.[45] There are 19 HLA class I gene loci, where 3 are classical (*HLA-A*, *-B* and *-C*), 3 non-classical (*HLA-E*, *-F* and *-G*) and 12 non-coding genes or pseudogenes (*HLA-S/17*, *-X*, *-N/30*, *-L/92*, *-J/59*, *-W/80*, *-U/21*, *-K/70*, *-16*, *-H/54*, *-90* and *-75*), clustered within three separate duplication blocks, designated as the alpha, beta and kappa blocks[46] (Figure 1). Of the HLA pseudogenes, *HLA-H/54* appears to encode two mRNA sequences (AK090500 and AK308374), whereas the transcript AK127349 and hypothetical protein FLJ45422 sequence were mapped to a part of overlapping exons of *HLA-L/92*. The *FLJ45422* gene is composed of five exons and contains an Ig domain constant region (IGc) and transmembrane domain, but its polymorphisms and function are unknown.

There are seven *MIC* genes, which are HLA class I-like genes, distributed across the three duplication blocks; two are expressed within the beta block, whereas the remainder are non-expressed pseudogenes within the kappa and alpha blocks.[46–48] These *MIC* genes have been generated with HLA class I genes by several rounds of segmental duplication events.[35] There are 34 non-HLA class I protein coding genes distributed between the duplication blocks that from an evolutionary perspective are termed anchor or framework genes.[48,49]

Overall, there are 128 loci within the 1.8 Mb class I region from *HCP5P15* to *MICB*, with 42 expressed genes, 12 gene candidates, 10 non-coding genes and 64 (50%) pseudogenes (Table 2). Of the 54 protein coding genes and gene candidates, 7 non-HLA genes (*LOC100133214*, *FLJ45422*, *LOC100133303*, *LOC100129065*, *LOC729792*, *HCG22* and *PSORS1C3*) were identified in the region after the previous locus information report.[31] Of the 42 protein coding genes, 4 (*SFTA2*, *MUC21*, *CCHCR1* and *PSORS1C3*) were previously unknown to be functional loci, and *TUBB* received a new official symbol and name (Table 1).

Class III region

The class III region, located between the class I and II regions, contains 75 loci within 0.9 Mb of DNA from *PPIAP9* to *BTNL2* (Table 1), with 55 protein coding genes and 5 (6.7%) pseudogenes (Table 2). Most of the protein coding genes and gene candidates were described earlier in the locus information report of 2004,[31] but three genes (*LY6G6F*, *C6orf26* and *LOC100128067*) were identified more recently. *LY6G6F* belongs to a cluster of *leukocyte antigen-6* (*LY6*) genes in the class III region and it encodes a type I transmembrane protein belonging to the

Table 2 Gene numbers in the HLA region

	Protein coding	*Candidate*	*NC gene*	*Pseudogene*	*Total*
HLA class I genes	6	0	1	12	19
HLA class II genes	12	0	3	4	19
MIC genes	2	0	0	5	7
Total for HLA-like genes	20	0	4	21	45
Non-MHC genes	112	20	18	58	208
Total for all genes	132	20	22	79	253
Extended class I region	3	0	0	1	4
Class I region	42	12	10	64	128
Class III region	55	7	8	5	75
Class II region	17	0	3	7	27
Extended class II region	15	1	1	2	19
Total for all genes	132	20	22	79	253

Abbreviations: HLA, human leukocyte antigen; MHC, major histocompatibility complex.

Table 3 Features of repeat sequences

	Entire region	*Extended class I*	*Class I*	*Class III*	*Class II*	*Extended class II*
Nucleotide length (bp)	3 753 173	164 542	1 781 830	889 503	676 843	240 455
GC (%)	44.7	45.1	45.8	48.9	41.3	49.8
Total repeat sequence (%)	49.5	49.3	53.3	41.6	51.3	46.0
SINEs (%)	17.7	21.8	16.4	22.8	10.1	27.1
Alus (%)	16.0	19.3	14.9	20.9	8.3	25.2
MIRs (%)	1.7	2.4	1.5	1.9	1.9	1.9
LINEs (%)	16.7	8.1	18.5	12.2	23.7	6.4
LINE1 (%)	13.3	6.6	14.8	8.8	20.2	4.6
LINE2 (%)	3.1	1.3	3.3	3.2	3.2	1.7
L3/CR1 (%)	0.3	0.2	0.2	0.2	0.3	0.1
LTR elements (%)	10.7	12.2	14.1	3.0	12.5	7.9
ERVL (%)	3.1	5.2	5.2	0.6	1.2	0.7
ERVL-MaLRs (%)	2.7	2.8	3.7	0.7	3.2	1.4
ERV_classI (%)	3.9	4.2	4.3	1.8	4.8	4.9
ERV_classII (%)	1.0	0.0	0.8	0.0	3.2	0.8
DNA elements (%)	2.9	5.8	2.9	1.8	3.6	3.2

Ig superfamily,[43] which may have a role in signal transduction in response to platelet activation.[50] Of the 55 protein coding genes, 5 (*MCCD1*, *SLC44A4*, *EHMT2*, *ZBTB12* and *PRRT1*) were previously unknown to be functional loci, and three (*VARS*, *LSM2* and *CFB*) had a symbol and name change (Table 1). In addition, five small nuclear RNA sequences (*SNORD84*, *SNORD117*, *SNORA38*, *SNORD48* and *SNORD52*) were identified in the vicinity of the *BAT1*, *BAT2* and *C6orf48* genes, respectively.[51–53] The class III region has no known HLA class I- and class II-like genes, but contains the complement factor genes, *C2*, *C4*, *CFB*, the cytokine genes *TNF*, *LTA* and *LTB*, and many genes with no obvious relationship to immune function or inflammation. The gene combination of *RP-C4-CYP21-TNX* is modular in structure and varies in copy number and has haplotypic variability. Many of the gene products expressed in the class III region have fundamental roles in cellular processes, such as transcription regulation (*BAT1*, *VARS*, *RDBP*, *STK19*, *SKIV2L*, *CREBL1* and *PBX2*), housekeeping (*DOM3Z*, *NEU1*, *AGPAT1*, *CL1C1* and *CSNK2B*), biosynthesis, electron transport and hydrolase activity (*PPT2*, *DDAH2* and *ATP6V1G2*) and protein–protein interactions for either intracellular or intercellular interactions, chaperone function and signaling (*C6orf46, HSPA1A, HSPA1B, BAT3, BAT8, AGAR, RNF5, FKRPL, TNXB, NOTCH4*).

Class II region

The class II region spans 0.7 Mb of DNA and contains the classical class II alpha and beta chain genes, *HLA-DP*, *-DQ* and *-DR* that are expressed on the surface of antigen-presenting cells to present peptides to T-helper cells. There are 27 loci identified within the class II region from *HLA-DRA* to *HLA-DPA3* (Table 1), with 17 protein coding genes, seven gene candidates and five pseudogenes (Table 2). In total, 19 of the loci are HLA class II-like sequences, including the 15 classical HLA class II loci and the four non-classical HLA class II loci (*HLA-DM* and *-DO*). The *HLA-DRB* loci are variable in number and MHC haplotype-dependent. The *HLA-DRB* locus in the PGF haplotype (Table 1) contains four copies of the *HLA-DRB* gene, *HLA-DRB1* (coding), *-DRB5* (coding), *-DRB6* (non-coding) and *-DRB9* (non-coding), whereas the *HLA-DRB* copy numbers vary for other haplotypes.[5] All of the 17 protein coding genes were previously known to be functional genes. Of all the protein coding genes in this region, *BRD2* (alias *RING3*) is the only gene without an established immune function. It is a transcription factor with widespread specificity, possibly remodeling chromatin complexes through interactions with histone acetyltransferase complexes, and its activity is high in myeloid leukemias.[54] Although *BRD2* may have a homologous sequence in yeast and *Drosophila*, it is strongly linked with the MHC of most vertebrates in the evolutionary path from sharks to man.[48]

Extended class II region

The extended class II region spans 0.2 Mb of DNA from *COL11A2* to *KIFC1* (Table 1), with 19 loci; that is, 15 protein coding genes, 1 gene candidate, 1 non-coding gene and 2 pseudogenes (Table 2). There was only one newly identified gene candidate (*LOC646720*) since the locus information report of 2004.[31] However, of the protein coding genes, two (*WDR46* and *PFDN6*) were previously unknown to be functional genes.

Interspersed repeats

Apart from the gene loci, 49.5% of the HLA genomic sequence is composed of interspersed repeat elements, such as SINE (Alu, MIR), LINE (LINE1 and 2, L3/CR1), LTR elements (ERVL, ERV class I and class II) and DNA elements (hAI-Charlie, TeMar-Tigger). Table 3 presents a summary of the repeat elements as detected by RepeatMasker (http://www.repeatmasker.org/). A comparable analysis with slightly different results and annotations (data not shown) was obtained with the repeat analysis program CENSOR.[55]

GENOMIC DIVERSITY

HLA genes

A total of 3201 HLA allele sequences (2215 in class I and 986 in class II) were released by the IMmunoGeneTics HLA (IMGT/HLA) database release 2.22 in July 2008 (http://www.ebi.ac.uk/imgt/hla/). The IMGT/HLA Database is a specialist database for HLA sequences. Ten years ago, the allele numbers were only 964, but since then the numbers have increased by ~200–300 allele sequences each year. Of the 2176 HLA class I alleles, 673, 1077, 360, 9, 21 and 36 alleles were counted in *HLA-A*, *-B*, *-C*, *-E*, *-F* and *-G* genes, respectively (Table 4); 2110 and 66 alleles were counted in the classical and non-classical HLA class I genes, respectively. Of 986 HLA class II alleles, 3, 669, 34, 93, 27, 128, 4, 7, 12 and 9 alleles were counted in *HLA-DRA*, *-DRB*, *-DQA1*, *-DQB1*, *-DPA1*, *-DPB1*, *-DMA*, *-DMB*, *-DOA* and *-DOB* genes, respectively (Table 4), with 954 and 32 alleles in the classical and non-classical HLA class II genes, respectively. In addition, 64 and 30 alleles were detected for the MHC class I-like gene, *MICA* and *MICB*, respectively.

Microsatellites

A total of 1527 microsatellite loci (846 in class I, 295 in class III and 386 in class II) were detected in the COX-MHC sequence (accession number NT_113891) by the Sputnik program (http://espressosoftware.com/pages/sputnik.jsp). Of them, 268 microsatellites (146 in class I, 61 in class III and 61 in the II) were developed as genetic markers.[56] These polymorphic microsatellite markers have been useful for precise map-

Table 4 Number of HLA alleles

Category	*Locus*	*Allele number*	*Protein number*	*Null allele number*
Class I	**HLA-A**	**673**	**527**	**46**
	HLA-B	**1077**	**911**	**38**
	HLA-C	**360**	**283**	**8**
	HLA-E	9	3	0
	HLA-F	21	4	0
	HLA-G	36	14	1
	Pseudogenes	39		
	Total	2215	1742	93
Class II	**HLA-DRA**	**3**	**2**	**0**
	HLA-DRB	**669**	**546**	**8**
	HLA-DQA1	**34**	**25**	**1**
	HLA-DQB1	**93**	**68**	**1**
	HLA-DPA1	**27**	**16**	**0**
	HLA-DPB1	**128**	**114**	**2**
	HLA-DMA	4	4	0
	HLA-DMB	7	7	0
	HLA-DOA	12	3	1
	HLA-DOB	9	4	0
	Total	986	789	13
MHC-like	MICA	64	54	0
	MICB	30	19	2
	Total	94	73	2

Abbreviations: HLA, human leukocyte antigen; MHC, major histocompatibility complex.
This information was obtained from IMGT/HLA Database release 2.22. Bold letters show the HLA genes with classical functions.

ping of disease-related genes within the HLA region in linkage analysis and disease association studies.[57,58] Moreover, they provide a powerful tool to study recombination events in this region, which contributes to haplotypic diversification. Detailed microsatellite marker information is provided by the dbMHC database of the NCBI (http://www.ncbi.nlm.nih.gov/gv/mhc/main.fcgi?cmd=init).

SNPs

A total of 60 928 to 71 569 SNPs were detected in a pairwise analysis of five different genomic sequence assemblies (PGF, Celera, HuRef, C6_COX and C6_QBL), ranging from *GABBR1* to *KIFC1*, by dbSNP (http://www.ncbi.nlm.nih.gov/SNP/). SNP markers are useful for constructing HLA haplotypes and for precise mapping of disease-related genes within the HLA region.[59–62] Figure 2 shows the marked peaks and troughs of the SNP distributions for the pairwise analysis of the five assemblies. The main peak diversities were observed not only in genomic segments harboring the highly polymorphic *HLA-A*, *-B*, *-C*, *-DR*, *-DQ* and *-DP* loci but also within some non-HLA loci such as those telomeric of HLA-C. Therefore, the HLA diversity is not limited to the antigen/T-cell receptor)-interacting sites of the HLA

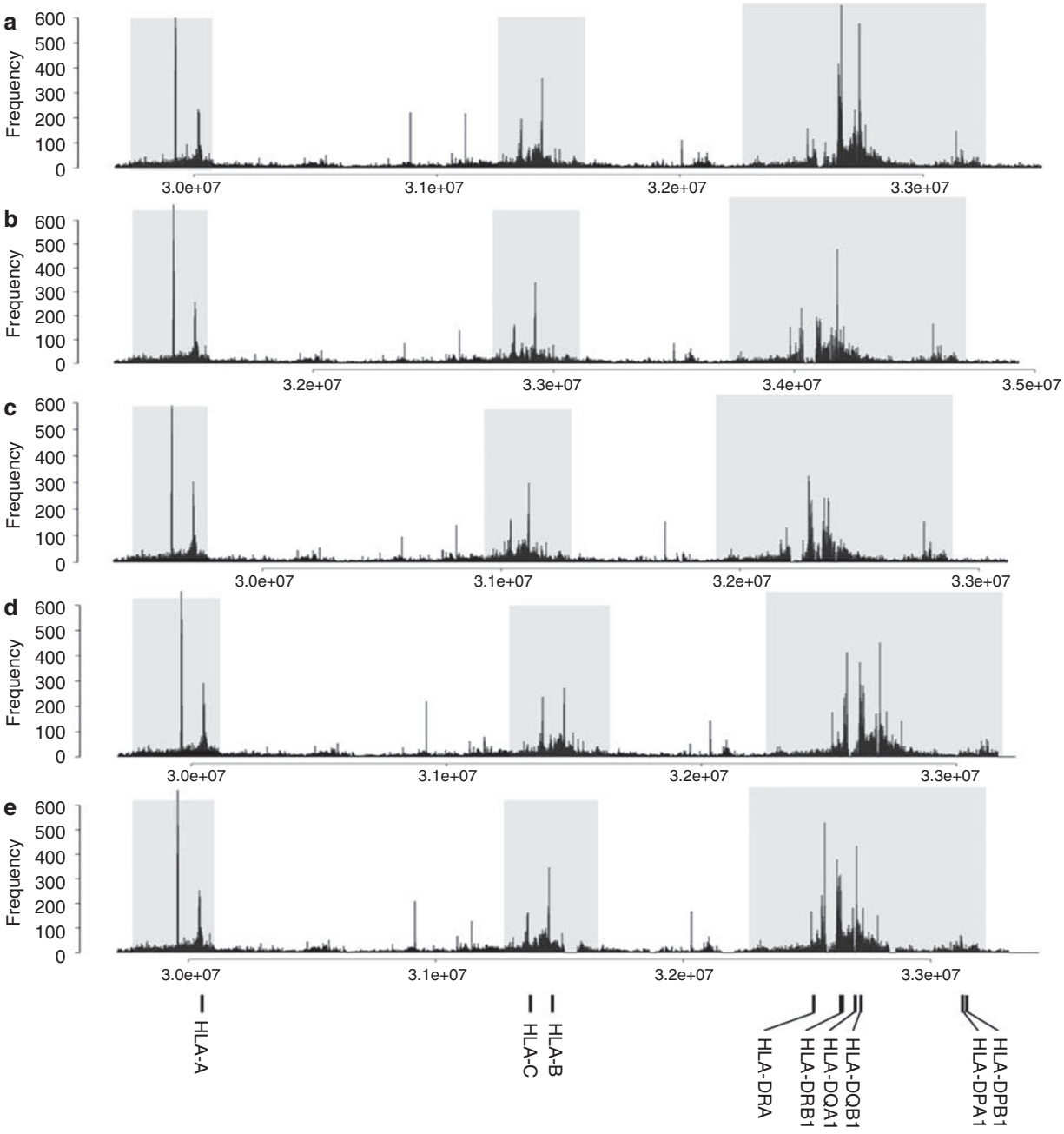

Figure 2 Single nucleotide polymorphism (SNP) distribution within the human leukocyte antigen (HLA) region. Diversity plots (**a**–**e**) drawn by comparing the released SNPs in dbSNP database against the reference assembly sequence determined in 1999[1] (accession no. NT_007592) (**a**), Celera alternate assembly sequence (accession no. NW_923073) (**b**), HuRef alternate assembly sequence based on HuRef SCAF_1103279188254 (accession no. NW_001838980) (**c**), c6_COX sequence (accession no. NT_113891) (**d**) and c6_QBL sequence (accession no. NT_113893 to NT_113897) (**e**). Gray backgrounds show significantly higher SNP regions that may have been generated by hitchhiking diversity.[3]

molecules,[63] but spreads to the surrounding loci as hitchhiking diversity owing to the accumulated effect of overdominant selection acting on HLA loci.[3] Interestingly, several disease-related genes, such as diffuse panbronchiolitis, psoriasis vulgaris, rheumatoid arthritis and sarcoidosis, were identified in the hitchhiking diversity-affected segments.[57,58,64,65] It was hypothesized by Shiina *et al.*[3] that some non-HLA disease alleles co-evolved with the positively selected HLA loci that were in linkage with harmful polymorphisms within the negative or neutrally selected non-HLA loci in response to various selection, population, genetic and environmental factors.

Genomic variation

The HLA genomic variations generated by *HLA-DRB* gene copy number in class II and/or the copy number variations (CNVs) of the *RP-C4-CYP21-TNX* gene combination in class III were previously associated with a number of different autoimmune diseases well before the complete, continuous HLA super-locus sequence was available.[46] The HLA-DR haplotypes consist of a number of copies of coding and non-coding *HLA-DR* genes. The expressed *DRB* sequences have been assigned to four different loci, *DRB1*, *3*, *4* and *5*. The highly polymorphic *DRB1* alleles (Table 4) are present in all haplotypes, whereas *DRB3*, *4* and *5* are present only in some haplotypes, as are the *HLA-DRB2* and *HLA-DRB6* to *-DRB9* pseudogenes. The *HLA-DRB2* pseudogene lacks exon 2 and contains a 20-nt deletion in exon 3, which has interrupted the correct translational reading frame.[66] The common HLA-DR alleles, major allotypes and their association with disease have been reviewed by Marsh.[67] The low and high copy numbers of the *C4* gene in the class III region have been recently associated as risk and protective genes, respectively, for systemic lupus erythematosus (SLE) susceptibility in European Americans.[68]

Genomic variations, such as insertion or deletion (InDel), inversion and other CNV, have been detected in recent genome-wide studies by comparative genomic hybridization (CGH) array mapping, fosmid end mapping, Mendelian inconsistencies, paired-end mapping of 454 sequencing reads, SNP chips and computational mapping of re-sequencing traces.[69–79] From the Database of Genomic Variants (http://projects.tcag.ca/variation/; 26 June 2008), 181 variations (50 InDels, 1 inversion and 130 CNVs) were detected at 49 genomic positions of the HLA region, especially within the HLA class I and II gene regions and a part of the class III region (Table 5). Some InDels are repetitive elements, such as Alu, HERV, L1 and SVA, or were generated by the influence of repetitive elements.[7,34,80–83]

INTRA- AND EXTRA-MHC GENE INTERACTIONS

MHC genes do not function in isolation from other genes in the human genome, but they may interact with other genes inside (local or intra-MHC gene interaction) or outside the MHC region (global or extra-MHC gene interactions). The MHC gene interactions may be viewed as quantitative interactions between alleles at different loci that affect fitness or contribute to complex disease phenotypes (epistasis),[84,85] as simple statistical interactions between alleles at different loci (linkage disequilibrium or LD) as a consequence of functional selection or a hitchhiking effect,[86,87] as functional protein-binding interactions detected by two-hybrid, affinity capture or phage display methods,[88] or as protein–DNA interactions such as those between transcription factors and gene promoter and enhancer regions[89,90] or between replication protein factors and DNA replication sites and elements.[91,92] The study of genetic interactions can reveal gene function, the nature of the mutations, functional redundancy, transcription regulation and protein interactions in normal and disease processes.

Table 6 provides an example of some protein interactions encoded by genes located inside and outside the MHC genomic region. Of the interactions between different genes within the MHC, the most definitively studied examples are those involved in protein dimer formation and peptide presentation in the adaptive immune response. In the former case, the interaction of the HLA class II alpha and beta proteins encoded by the classical class II A and B gene loci, respectively, have long been known to form the alpha and beta heterodimer chains and consequently have received extensive investigations at various levels, including X-ray structural analysis.[93,94] The interaction of proteins involved in antigen presentation, such as HLA class I proteins, TAP1, TAP2, HLA-DM and TAPBP, have also been extensively studied.[95,96] The interactions between the alleles of the HLA-DR haplotypes, which are in strong LD, were found to affect the immune response levels and disease susceptibility. For example, the results obtained for two multiple sclerosis-associated HLA-DR alleles at separate loci of the *HLA-DR2* haplotype in a humanized mice functional assay imply that the LD between these two alleles is due to a functional epistatic interaction.[97] Accordingly, one allele modifies the T-cell response activated by the second allele through activation-induced cell death resulting in a milder form of multiple sclerosis. Other protein interactions encoded by genes within the MHC genomic region include those between RFP5 and BAT5, C4B and C2, CFB and C4B, LTA and LTB, IER3 and BAT3, and between MRPS18B and NFKBIL1.

Examples of protein interactants encoded by genes inside and outside the MHC are more numerous than those encoded by genes within the MHC genomic region. Recent research has focused strongly on the HLA class I interactions with the *killer Ig receptor* (*KIR*) genes and the *leukocyte Ig-like receptor* (*LIR*) gene family encoded in the *leukocyte receptor complex* (*LRC*) on chromosome 19q13.[98,99] Combinations of HLA class I and KIR variants have been associated with autoimmunity, viral infections, pregnancy-related disorders and cancer.[100,101] Similarly, the proteins encoded by the *MICA* and *MICB* genes (Table 6) are known to interact with KLRC4 and KLRK1 that are encoded by the genes on chr 12, to regulate innate immunity by way of the NK cell systems.[47] The proteins encoded by the *C4*, *CFB* and *C2* genes in the HLA class III region are involved in complement activation and consequently interact with proteins encoded by genes from outside the MHC (Table 6). Allelic variations between the MHC complement genes and non-MHC gene sequences have been associated with macular degeneration and SLE.[102] Recently, Lester *et al.*[103] reported finding an epistasis between the MHC *C4* gene region and the *RCAa* block in primary Sjögren syndrome. The *RCAa* block (*regulators of complement activation*, 1q32) contains critical complement regulatory genes such as *CR1* and *MCP*, and the epistasis was attributed to an interaction between *C4* and its receptor, *CR1*, encoded within the *RCAa* block. Furthermore, the *IFN-regulator factor 5* (*IRF5*) gene variants located on chr 7q32 were found to interact with the class I MHC locus in people with psoriasis[104] and possibly other autoimmune diseases.[105]

Most proteins encoded by the 132 protein coding genes within the MHC interact with proteins encoded by genes outside the MHC region. The protein and genetic interactions of the MHC genes listed in Table 1 can be accessed and viewed by way of the GeneID number. For example, the interaction data and online links for the *MDC1* gene (GeneID: 9656), *mediator of DNA damage checkpoint 1*, which is required to activate the intra-S phase and G2/M phase cell cycle checkpoints in response to DNA damage, includes information on the peptide or protein interactants, the interacting genes, the source databases (Human Protein Reference Database (HPRD) or BioGRID)

Table 5 Genomic variations of the HLA region

Region	*Position*[a]	*Variation number*	*Variation type*	*Affected locus*
Class I	chr6:29900413–30083123	26	Copy number	HLA-A, HCG9, HLA-G
	chr6:29923215–29923586	2	InDel	
	chr6:29926851–29926851	1	InDel	
	chr6:30011861–30011861	1	InDel	
	chr6:30106475–30106475	1	InDel	
	chr6:30752672–30924298	3	Copy number	FLOT1, TUBB, NRM, IER3, MDC1, KIAA1949
	chr6:30891138–30891138	1	InDel	
	chr6:30891543–30891703	1	InDel	
	chr6:30894392–30895190	1	InDel	
	chr6:31088899–31088899	1	InDel	
	chr6:31117665–31119504	1	Inversion	
	chr6:31136269–31650287	45	Copy number	C6orf15, CDSN, PSORS1C1, PSORS1C2, CCHCR1, TCF19, POU5F1, HCG27, HLA-C, HLA-B, HCP5, MICB, MCCD1, BAT1, NFKBIL1
	chr6:31379867–31380220	1	InDel	
	chr6:31389749–31390117	11	InDel	
	chr6:31404777–31404777	1	InDel	
	chr6:31430692–31431029	1	InDel	HLA-B
	chr6:31503417–31503528	1	InDel	
	chr6:31504510–31504681	1	InDel	
	chr6:31505358–31505358	1	InDel	
	chr6:31546995–31546995	1	InDel	
Class III	chr6:31803109–31803297	1	InDel	DDAH2
	chr6:31803450–31803929	1	InDel	DDAH2
	chr6:31975718–31978975	1	Copy number	ZBTB12
	chr6:31979491–32317091	17	Copy number	C2, CFB, RDBP, SKIV2L, DOM3Z, STK19, C4B, C4A, CYP21A2, TNXB, CREBL1, FKBPL, PRRT1, PPT2, AGPAT1, RNF5, AGER, PBX2, NOTCH4
	chr6:32343057–32343284	1	InDel	
	chr6:32421761–32422084	1	InDel	C6orf10
Class II	chr6:32467750–32813412	32	Copy number	BTNL2, HLA-DRA, HLA-DRB1, HLA-DRB5, HLA-DQA1, HLA-DQB1,
	chr6:32485882–32486648	1	InDel	
	chr6:32555421–32556081	1	InDel	
	chr6:32579256–32579534	1	InDel	
	chr6:32586215–32586365	1	InDel	
	chr6:32599574–32599574	1	InDel	HLA-DRB5
	chr6:32653188–32653679	1	InDel	
	chr6:32655565–32655887	1	InDel	HLA-DRB1
	chr6:32658616–32658959	1	InDel	HLA-DRB1
	chr6:32662978–32662978	1	InDel	HLA-DRB1
	chr6:32679326–32679613	1	InDel	
	chr6:32734144–32734985	1	InDel	
	chr6:32749252–32749252	1	InDel	
	chr6:32784552–32785009	2	InDel	
	chr6:32816473–32821064	1	Copy number	HLA-DQA2
	chr6:32881415–32881415	1	InDel	
	chr6:32886732–32887798	1	Copy number	
	chr6:32903482–32904136	2	InDel	TAP2
	chr6:32958730–33107258	2	Copy number	HLA-DMB, HLA-DMA, BRD2, HLA-DOA
	chr6:33117527–33117645	1	InDel	
	chr6:33193163–33203995	1	Copy number	
	chr6:33216135–33216241	1	InDel	
	chr6:33252386–33253403	1	Copy number	COL11A2

Abbreviation: HLA, human leukocyte antigen.
This information was sourced from the Database of Genomic Variation (26 June 2008 to present).
[a]The physical position of the variations was taken from the records of the Assembly of the Human Genome (NCBI Build36).

Table 6 Examples of some MHC gene interactions sourced from Entrez gene at NCBI

Gene	*GeneID*	*Interacting gene symbol*	*NCBI GeneID*	*Chromosome*	*Interacting gene product name*
HLA-B	3106				
		B2M	567	15q21–q22.2	Beta-2-microglobulin
		CD8A	925	2p12	CD8 alpha chain of T-cell receptor
		KIR3DL1	3811	19q13.4	Killer cell immunoglobulin-like receptor, three domains, long cytoplasmic tail, 1
		KLRD1	3824	12p13	Killer cell lectin-like receptor subfamily D, member 1
		LILRB1	10859	19q13.4	Leukocyte immunoglobulin-like receptor, subfamily B (with TM and ITIM domains), member 1
		LILRB2	10288	19q13.4	Leukocyte immunoglobulin-like receptor, subfamily B (with TM and ITIM domains), member 2
		TRA@	6955	14q11.2	T-cell receptor alpha locus
		HHV8gp11	935111	Herpesvirus 8	HHV8 group 11 protein
		HIV			
		Peptides	4100	Genome wide	Peptides of various cellular and extracellular gene products
HLA-DRB1	3123				
		HLA-DRA	3122	6p21.3	Major histocompatibility complex, class II, DR alpha
		TRA@	6955	14q11.2	T-cell receptor alpha locus
		HLA-DMA	3108	6p21.3	Major histocompatibility complex, class II, DM alpha
		HLA-DRA	3122	6p21.3	Major histocompatibility complex, class II, DR alpha
		Peptides[a]			Peptides of various cellular and extracellular gene products
HLA-DRA	3122				
		CD63	967	12q12–q13	CD63 molecule
		CD82	3732	11p11.2	CD82 molecule
		HLA-DMA	3108	6p21.3	Major histocompatibility complex, class II, DM alpha
		HLA-DMB	3109	6p21.3	Major histocompatibility complex, class II, DM beta
		HLA-DRB1	3123	6p21.3	Major histocompatibility complex, class II, DR beta 1
		HLA-DRB5	3127	6p21.3	Major histocompatibility complex, class II, DR beta 5
		Peptides[a]	4155	18q23	Peptides of various cellular and extracellular gene products
POU5F1	5460				
		HMGB1	3146	13q12	High-mobility group box 1
		HMGB2	3148	4q31	High-mobility group box 2
		MNAT1	4331	14q23	Menage a trois homolog 1, cyclin H assembly factor
		SOX2	6657	3q26.3–q27	SRY (sex determining region Y)-box 2
TNF	7124				
		KHSRP	8570	19p13.3	KH-type splicing regulatory protein
		BGN	633	Xq28	Biglycan
		CSF1	1435	1p21–p13	Colony-stimulating factor 1
		DCN	1634	12q21.33	Decorin
		PRTN3	5657	19p13.3	Proteinase 3
		TNFRSF1A	7132	12p13.2	Tumor necrosis factor receptor superfamily, member 1A
		TNFRSF1B	7133	1p36.3–p36.2	Tumor necrosis factor receptor superfamily, member 1B
NOTCH4	4855				
		FBXW7	55294	4q31.3	F-box and WD repeat domain containing 7
		MAML1	9794	5q35	Mastermind-like 1
MICA	4276				
		KLRC4	8302	12p13.2–p12.3	Killer cell lectin-like receptor subfamily C, member 4
		KLRK1	22914	12p13.2–p12.3	Killer cell lectin-like receptor subfamily K, member 1
MICB	4277				
		KLRC4	8302	12p13.2–p12.3	Killer cell lectin-like receptor subfamily C, member 4
		KLRK1	22914	12p13.2–p12.3	Killer cell lectin-like receptor subfamily K, member 1
CCHCR1	54535				
		STAR	6770	8p11.2	Steroidogenic acute regulatory protein

Table 6 Continued

Gene	GeneID	Interacting gene symbol	NCBI GeneID	Chromosome	Interacting gene product name
C4A					
		APOA2	336	1q21–q23	Apolipoprotein A-II
		C3AR1	719	12p13.31	Complement component 3a receptor 1
		CR1	1378	1q32	Complement component (3b/4b) receptor 1
		CST3	1471	20p11.21	Cystatin C
		GPR77	27202	19q13.33	G-protein-coupled receptor 77

Abbreviations: MHC, major histocompatibility complex; NCBI, National Center for Biotechnology Information.

and published references (PubMed). The 13 genes found to interact with *MDC1* and listed at Entrez Gene are *ATM*, *BRCA1*, *CHEK2*, *H2AFX*, *NBN*, *SMC1A*, *TP53*, *TP53BP1*, *CENPC1*, *CHEK2*, *GATA4*, *H2AFX* and *HDAC10*. In another example, the protein expressed by the *CCHCR1* gene (ID:54535), which has at least three splice variants, was identified to promote steroidogenesis by interacting with STAR, the steroidogenesis acute regulatory protein[106] encoded by a gene on chr 8p (Table 6), which may be downregulated in psoriatic keratinocytes.[107] A public online service for protein interaction datasets is also provided by BioGRID at http://www.thebiogrid.org/index.php and the HPRD at http://www.hprd.org/index_html. The knowledge extracted from protein interaction databases might assist in a more efficient organization and analysis of genome-wide studies by revealing which gene interactions warrant epistatic investigation.

MHC AND GENOME-WIDE GENE EXPRESSION PROFILING

Most knowledge on MHC gene expression at the transcript and protein levels has depended on individual gene studies (Table 1). However, in recent years, the development of genome-wide gene expression assays, including some or many of the MHC genes, has provided a more global perspective of different expression patterns in immune- and disease-related pathways. Gene expression profiling of normal and diseased cells and/or tissues using oligonucleotides, cDNA or genomic arrays has been a particularly successful by-product of genome sequence research. Global transcriptome studies are performed using various descriptive, experimental and disease conditions, and the data are often deposited into public databases, such as Gene Expression Omnibus (GEO), that can be accessed online for review and/or reanalysis (http://www.ncbi.nlm.nih.gov/geo/).

Genome-wide gene expression data have permitted an examination and comparison of the mRNA profiles expressed by genes both inside and outside the MHC region. For example, in our study of the gene transcription patterns in the skin lesions of four Japanese patients with psoriasis vulgaris and three normal controls, we found that only seven MHC genes (*LY6G6C*, *CDSN*, *TAP1*, *HLA-G*, *HLA-F*, *TUBB* and *CFB*) from a total of approximately 90 MHC protein coding and non-coding genes represented on the HUG95A Affymetrix oligonucleotide array of 12 000 human genes were significantly upregulated in the affected skin compared with normal skin; no significant statistical changes occurred in the expression of the classical HLA class I and II genes.[108] The only MHC gene that was significantly downregulated in the psoriatic lesions was *GABBR1*. Most of the 263 significantly upregulated changes in the psoriatic-affected skin occurred for genes located outside the MHC region that were involved with interferon mediation, inflammation immunity, cell adhesion, cytoskeleton restructuring, protein trafficking and degradation, RNA regulation and degradation, signaling transduction, apoptosis and atypical epidermal cellular proliferation and differentiation. Bioinformatics analysis of the significantly upregulated genes in psoriatic skin compared with normal skin, using a commercially available computer network program (MetaCore) in Figure 3, shows that inflammation and cell cycle regulation were the two most significant molecular pathways involved in psoriasis by way of the *STAT* and *Myc* gene regulatory systems as well as by the MHC genes, *HLA-G* (interacting with *KIR2DL4* and *ILT2* on chr 19), *DDR1* and *TNF* (MetaCore Applications (2007) http://www.genego.com/pdf/PsoriasisCS.pdf). The *HLA-G* locus was recently found to also interact with the IRF5, encoded by gene variants on chr 7q32 in Swedes with psoriasis.[104]

Other investigators have used similar gene microarray assays to identify the patterns of MHC and non-MHC gene transcription in skin lesions of patients with psoriasis,[109,110] atopic dermatitis[111] and porokeratosis, a skin disorder of keratinization.[112] Gene expression profiling of peripheral blood mononuclear leukocytes has been performed on psoriasis patients for disease stage prediction[113,114] and treatments with therapeutic TNF and IFN-gamma antibodies.[115] Leukocytes and/or lymphocytes express more than 75% of the human genome and provide an alternative to tissue biopsies for studies of the association between HLA gene activity and autoimmune diseases, such as psoriasis, asthma, rheumatoid arthritis (RA) and SLE. A number of different MHC-related diseases, including SLE,[114,116] RA[117,118] and OA,[119] have been investigated by gene expression profiling. For example, van der Pouw Kraan *et al.*[120] used cDNA microarray technology to subclassify RA patients and to disclose different disease pathways in rheumatoid synovium. They found that among the 121 genes overexpressed in one of the main tissue groups (RA-I) identified by a hierarchical clustering of gene expression data, 9 genes from the MHC region were indicative of an adaptive immune response, whereas another group (RA-II) expressed genes suggestive of fibroblast dedifferentiation. Microarray analyses of peripheral blood cells from patients with psoriatic arthritis identified downregulation of innate and acquired immune responses as well as the MHC genes from the *PSORS1* and *PSORS2* susceptibility loci.[121]

Peripheral arterial occlusive disease (PAOD: OMIM 606787) is commonly found in elderly patients as a result of atherosclerosis of large and medium peripheral arteries, or aorta, and often coexists with coronary artery disease and cerebrovascular disease. Recently, Fu *et al.*[122] analysed 30 femoral arteries (11 with intermediate and 14 with advanced atherosclerotic lesions and 5 normal femoral arteries) by genome-wide gene expression profiling using the Affymetrix microarray platform and found that most of the MHC class II and complement molecules were significantly upregulated in the intermediate lesions, but not in the advanced lesions. They concluded from the results of their expression study that different immune and inflammatory responses occur at different stages of PAOD and

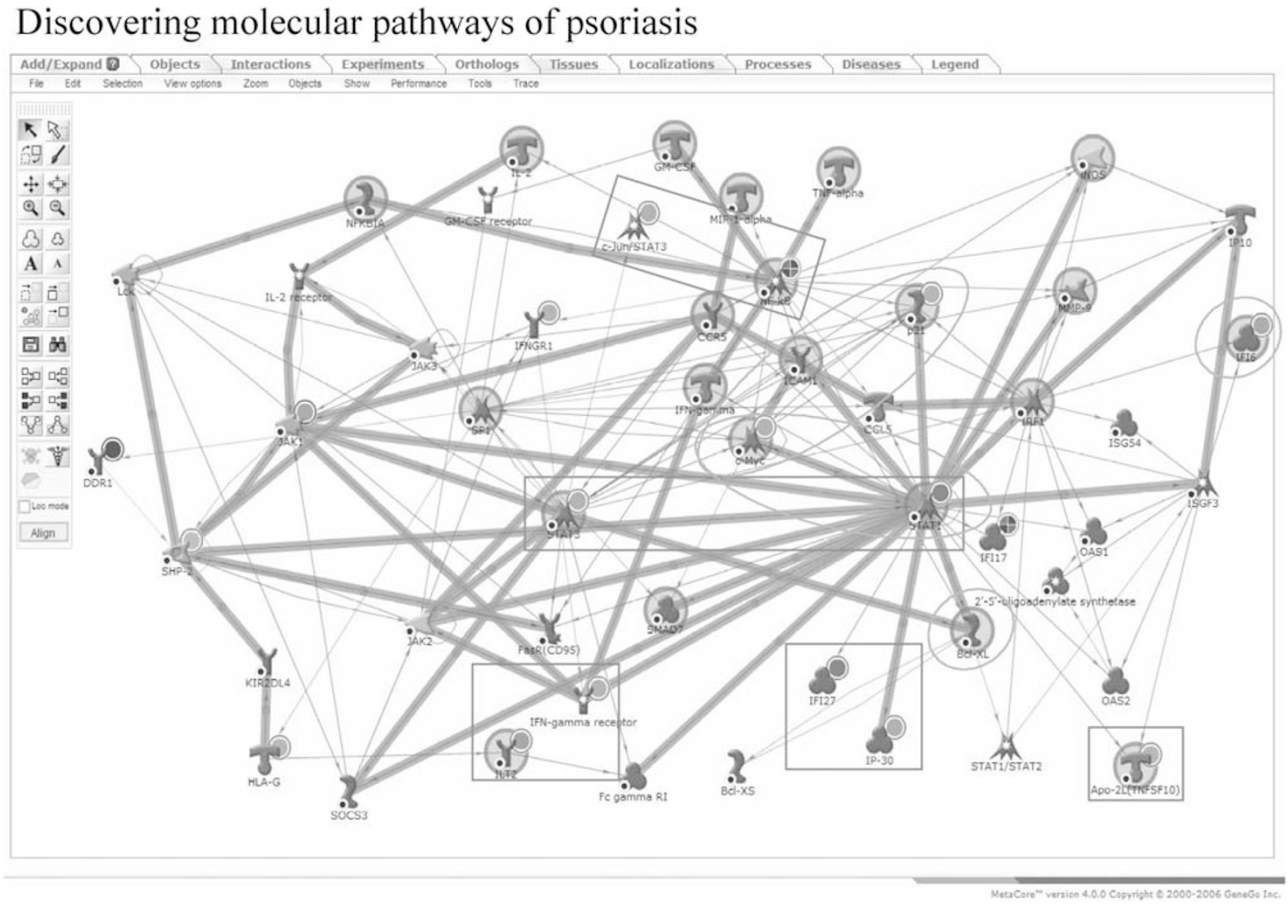

Figure 3 The involvement of major histocompatibility complex (MHC) genes, HLA-G, DDR1 and tumor necrosis factor (TNF)-alpha, in the molecular pathways of psoriasis. The whole-genome microarray data of Kulski *et al*.[108] were evaluated using the MetaCore software package to identify the molecular character and pathways involved in psoriasis. The MHC genes are highlighted by black squares. Red rectangles and orange ovals represent the genes involved in the inflammation and cell cycle regulation pathways (thick blue lines), respectively, and red circles represent overexpressed key transcription regulators. The figure was produced by MetaCore from GeneGo Inc. (St Joseph, MI, USA). The color reproduction of this figure is available on the html full text version of the manuscript.

development of artherosclerotic lesions. The MHC class II and complement gene activity was related in different ways to the Toll-like receptor signaling and NK cell-mediated cytotoxicity enrichment found to take place in the intermediate and advanced atherosclerotic lesions.

HLA-wide gene expression profiling using the Affymetrix microarray platform also allows researchers an opportunity to determine the degree of positive and negative coordination between HLA and non-HLA gene expression in controlled experiments, cell and tissue types, and in population and disease studies. For example, Figure 4 shows the microarray expression profiles for some non-HLA class I genes relative to the expression of the non-classical HLA class I genes, *HLA-E*, *-F* and *-G*, in established cell lines derived from different cancers, with data provided by The Cancer Genome Anatomy Project (http://cgap.nci.nih.gov/Genes). It can be seen in Figure 4 that the *FLOT1* gene was expressed at highest levels in cancer cells derived from the CNS, whereas *DDR1* and *TRIM15* (alias *Hs.591789*) were expressed most strongly in the colonic cancer cell lines. In comparison, the non-classical HLA class I genes were expressed most consistently at moderate to high levels in the cell lines derived from renal carcinomas. The variable expression of *TRIM15* among the different cancer cell types is notable given its possible antiviral role in innate immunity.[123,124]

Although an HLA and global picture of gene expression in tissues and cells can be obtained by using a full set of Affymetrix GeneChips, CGH for SNP analysis in combination with gene expression is still a relatively new and demanding approach for the study of complex diseases. CGH, in an attempt to improve functional genome research and disease associations, is particularly useful for detecting genomic sequence alterations or gene CNVs[125,126] that might be associated with disease. For example, CNVs of defensin genes on chr 8 were found to be strongly associated with Crohn's disease and the skin disease, psoriasis.[127,128] Similar studies on the effects of genomic alteration or CNVs on the expression of MHC genes are still limited, but a few recent reports suggest that this approach might yield important new insights into the interaction between the genes of the MHC and other genomic regions in disease studies. For example, the study by Jiang *et al*.[129] using cDNA microarrays to detect the simultaneous genomic and expression alterations in prostate cancer, has implicated the dysregulation of exogenous antigen presentation through MHC class II and protein ubiquitination during protein-dependent protein catabolism in the tumorigenic process. They found that the expressions of the MHC genes *ABCF1*, *HLA-DRB1* and *HLA-A*, located on the chromosome 6p21, and of the MHC class II chaperone gene, *CD74*, located on 5q32 were both significantly downregulated, probably as a consequence of the *CD74* gene deletion.

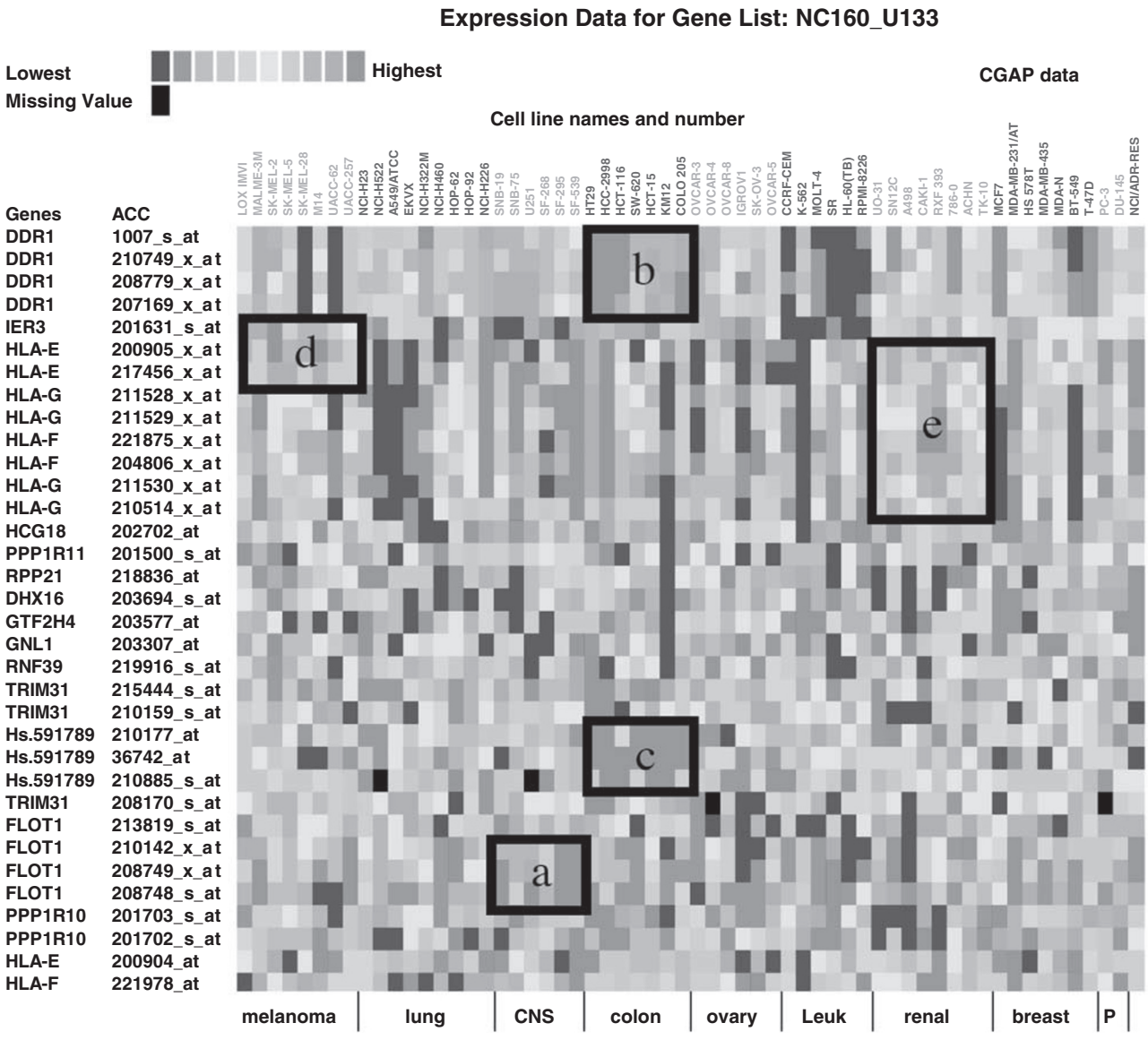

Figure 4 The relative expression of some human major histocompatibility complex (MHC) class I genes in different cancer cell lines. The gene examples from the class I region are non-human leukocyte antigen (HLA) genes (*DDR1*, *IER3*, *HCG18*, *PPP1R11*, *RPP21*, *DHX16*, *GTF2H4*, *GNL1*, *RNF39*, *TRIM31*, *Hs.591789 (TRIM15)*, *FLOT1* and *PP1R10*) and the non-classical HLA class I genes, *HLA-E*, *-F* and *-G*. The data are taken from The Cancer Genome Anatomy Project at the National Cancer Institute (USA) using the batch gene finder to find the expression data for the selected genes of interest (query) in the gene list of NC160_U133 (Affymetrix platform). The present image for the transcriptome analysis was produced online at http://cgap.nci.nih.gov/Genes/BatchGeneFinder using only the selected gene list shown in the image. The level of transcriptional activity in the cells ranged from the highest (red squares) to the lowest (blue squares) according to the color scale indicated at the top left-handed side of the figure. The rectangular blocks labeled (a–e) within the matrix of the figure highlight the detection probes with relatively high expression levels of *FLOT1* in central nervous system (CNS) cancer cells (a), *DDR1* (b) and *Hs.59178* (*TRIM15*) (c) in colon cancer cells, *IER3* and *HLA-E* in melanoma (d) and the non-classical HLA class I genes (e) in the renal cancer cells. Of the list of cancerous tissue at the bottom of the matrix, 'Leuk' is leukemia and 'P' is prostate. The color reproduction of this figure is available on the html full text version of the manuscript.

Genome tiling arrays is another improving methodology that appears useful for future investigations into MHC epigenetics,[130] SNPs,[7] gene–gene interactions[131] and gene expression activity[132] both inside and outside the MHC genomic region by using high-density oligonucleotide arrays with probes chosen uniformly from both strands of the entire genome, including all genic and intergenic regions. Genome-wide protein profiling (proteomics) by using chips, arrays or high-throughput mass spectrometry is a rapidly emerging technology in disease and diversity studies to screen for protein activities such as protein–protein, protein–DNA, protein–drug and protein–peptide interactions; to identify enzyme substrates and to profile immune responses.[133,134] Some of these procedures have been applied specifically to MHC gene functions, particularly to detect and characterize antigen-specific T-cell populations in disease,[135] HLA protein–peptide (antigen) interactions,[136] targeting autoantibody/autoantigen targets[137,138] and to profile other immune responses.[139] Bioinformatic and statistical algorithms are continually being developed to integrate the genomics of DNA variation, transcription and phenotypic data, to provide a system genetics view of disease and to enhance identification of the associations between DNA variation and

diseases as well as to characterize those parts of the molecular networks that drive disease.[140]

MHC AND DISEASE ASSOCIATIONS

The main function of the MHC gene region is to protect itself and its organism against harmful infectious agents (to recognize and deal with foreign organisms and antigens) and to dispense with the damaged, dying or infected cells and tissues. The extremely high levels of polymorphism and heterozygosity within the MHC genomic region provide the immune system with a selective advantage against the diversity and variability of pathogens. However, the high level of polymorphisms and mutations in the MHC has the added risk of generating autoimmune diseases and other genetic disorders. Several hundred autoimmune and infectious diseases have been associated with the MHC since the first report in 1967 that HLA-B antigens were increased in frequency in patients with Hodgkin's lymphoma.[141] At least another 40 different autoimmune diseases were linked to specific HLA types by the end of 1986.[142,143] In an update on the role of the MHC genes in disease, Shiina *et al.*[31] presented an overview of 109 HLA-associated diseases. When PubMed online at NCBI was searched in September 2008 with the keywords 'human MHC (or HLA) gene disease,' 3151 journal publications were listed on the subject of HLA and disease. Using 'HLA' as a keyword to search the Genetic Association Database (GAD) (http://geneticassociationdb.nih.gov/cgi-bin/index.cgi), 500 journal publications were found on HLA gene association and disease between 1999 and 2007. The statistical, biological and medical significance of many of the MHC disease association studies, however, remain unclear or doubtful.

A number of recent reviews are available on HLA and infections,[12,144–146] as well as HLA and autoimmune diseases,[11,31,32,147–150] and will not be considered in any detail here. OMIM is a database of human genes and genetic disorders that provides information and references on the discoverers, chromosomal location, molecular functions, mutations and associations between the genes and disease.[151] There are at least 100 OMIM identifiers concerning the HLA region loci, mostly of expressed genes, that can be accessed through http://www.ncbi.nlm.nih.gov/ or through links from other sites, including Entrez Gene database at NCBI.[36]

The 31 HLA disease associations listed in Table 7 and sourced from the OMIM database[152] are some examples of HLA-associated diseases that have a strong experimental or statistical association with reasonable reproducibility. At least 26 of these diseases have been associated with non-HLA genes encoded within the MHC, with the regulatory cytokines *TNF* and *LTA* contributing to a large number of disease associations by way of mutations or polymorph-

Table 7 MHC monogenic and polygenic disease associations

Disease (symbol)	*MIM no.*[a]	*MHC gene symbol*	*M or P*[b]
Age-related macular degeneration (ARMD1)	603075	CFB	M
Bare lymphocyte syndrome type 1 (BLS1)	604571	TAP1, TAP2, TAPBP	M
C2 deficiency	217000	C2	M
C4 deficiency	120810	C4A, C4B	M
Congenital adrenal hyperplasia (CA21H)	201910	CYP21A2, CYP21, CA21H	M
Ehlers–Danlos syndrome (TNX deficiency))	606408	TNXB	M
Hypotrichosis simplex of the scalp (HTSS)	146520	CDSN	M
Otospondylomegaepiphyseal dysplasia (OSMED)	215150	COL11A2	M
Sialidosis, neuraminidase deficiency	256550	NEU1	M
Stickler syndrome type III (STL3)	184840	COL11A2	M
Ankylosing spondylitis (AS)	106300	HLA-A, HLA-B27	P
Asthma	600807	HLA-G, TNF	P
Autoimmune thyroid disease (AITD)	608173	HLA-DR3	P
Azoospermia, non-obstructive (AZON)	606766	HLA-DRB1, HLA-A, HLA-B	P
Behcet disease (BD)	109650	HLA-B51, MICA	P
Beryllium disease, chronic (CBD)	142858	HLA-DPB1	P
Celiac disease (CD)	212750	HLA-DQA1, CELIAC1	P
Diffuse panbronchiolitis (PBLT)	604809	DPCR1, HLA-B54, HLA-A11	P
Immunoglobulin A deficiency (IGAD)	137100	Unknown	P
Inflammatory bowel disease 1 (IBD1)	266600	TNF	P
Migraine (MGR1)	157300	TNF	P
Multiple sclerosis (MS)	126200	HLA-DRB1, HLA-DQB1	P
Narcolepsy (NL)	161400	HLA-DQB1	P
Psoriasis vulgaris (PV)	177900	HLA-C, PSORS1, other	P
Psoriatic arthritis (PSORAS1)	607507	LTA	P
Rheumatoid arthritis (RA)	180300	HLA-DRB1, NFKBIL1	P
Sarcoidosis	181000	BTNL2	P
Seronegative myasthenia gravis (snMG)	254200	MYAS1, DR1, DR3, DR9	P
Systemic lupus erythematosus (SLE)	152700	TNF, HLA-DR, HLA-B, C4	P
Type I diabetes (T1D)	222100	HLA-DR, HLA-DQ	P
Vitiligo (VIT)	193200	Unknown, D6S265	P

Abbreviation: MHC, major histocompatibility complex.
[a]MIM number provides disease and gene association information and list of references for OMIM at http://www.ncbi.nlm.nih.gov/sites/entrez?db=omim.
[b]M is monogenic disease mutation and P is a suspected polygenic disease.

isms within the gene promoter or coding regions that might affect expression levels.[153–156] Ten of the diseases appear to be monogenic owing to mutations within one of the MHC genes. Adrenal hyperplasia is now well accepted to be the consequence of 21-hydroxylase deficiency and alterations in the *CYP21A2* gene.[157] Some of the *CYP21A2* gene alterations may arise by transference of sequences to *CYP21A2* from the neighboring non-coding *CYP21A1P* pseudogene by gene conversion.[158] It is also generally well accepted that mutations within the *NEU1* gene are responsible for neuraminidase deficiency and sialidosis, which is characterized by the progressive lysosomal storage of sialylated glycopeptides and oligosaccharides,[159] and that *C2* mutations cause C2 deficiency in the process of the complement cascade.[160] Of the 21 multifactorial diseases listed in Table 7, 11 (type I diabetes (T1D), inflammatory bowel disease, multiple sclerosis (MS), AITD, PV, RA, celiac disease (CD), ankylosing spondylitis (AS), SLE, juvenile RA (JRA) and vitiligo (VIT)) were linked most significantly to the HLA region in a recent meta-analysis of 42 independent genome-wide linkage studies.[161] In a recent genome-wide association study of seven common diseases using SNP markers, the MHC associations were strongest for RA, T1D, moderate for CD and weak or absent for bipolar disorder, coronary artery disease, hypertension and type II diabetes.[162] In another recent review and pooled analysis of the MHC in autoimmunity, a number of overlapping HLA class II and TNF alleles and haplotypes were associated with the diseases MS, T1D, SLE, UC, CD and RA.[11]

Most of the 21 multifactorial diseases listed in Table 7 are polygenic with a few specified or unspecified MHC gene alleles possibly interacting in some unspecified way with other genes inside and/or outside the MHC region. The exact MHC genes involved with many of the diseases are still not clearly defined. For example, the association of an HLA genomic region with the onset or maintenance of psoriasis is definite, but which of a number of MHC candidate genes (or combination of genes) ranging between the *MICA* and *CDSN* loci is responsible remains uncertain.[39,58,163–169]

Only a few autoimmune diseases have been related just to the classical class I and II alleles, in spite of the continuing dogma that disease associations are caused by altered or faulty peptide presentation to T cells by polymorphic class I and II gene products. AS is primarily attributed to *HLA-B27*, with minor associations such as *HLA-Cwl* and *-Cw2* or *HLA-DR7* considered secondary because of LD or a hitchhiking effect. Similarly, *HLA-B51* continues to be strongly associated with Behcet syndrome,[170] although other chromosomal regions may be involved.[171] In Caucasian populations of Northern European descent, the DR15 haplotype (*DRB1*1501-DQA1*0102-DQB1*0602*) is hypothesized to be the primary HLA genetic susceptibility factor for MS. Experiments with transgenic mice have confirmed the importance of the *DRB5*0101* and *DRB1*1501* allelic interactions in creating a mild form of MS-like disease,[97] but more severe forms probably depend on other genes[172] such as *T-cell receptor beta*, *CTLA4*, *ICAM1* and *SH2D2A*. Schmidt *et al.*[149] reviewed 72 publications on the HLA association with MS and found that most investigators reported a higher frequency of the DR15 haplotype and/or its component alleles for the MS cases than the controls, but the results may have been biased by poor study designs.

Owing to the difficulty in identifying a single MHC gene that is responsible for disease, some researchers prefer to examine the association between MHC haplotypes and disease susceptibility and resistance.[46] Common Caucasian MHC haplotypes may be accounted for by a limited number of ancestral haplotypes using the alleles of five or more gene loci.[173] The MHC ancestral haplotype (AH) 8.1, characterized by the alleles *HLA-A*01*, *-B*08*, *-DRB1*03*, *-DQB1*02* and *-DQA1*05* has been dubbed the 'autoimmune haplotype' because of its association with numerous autoimmune diseases, including T1D, CD, Graves' disease, SLE and Myasthenia Gravis (MS).[174] The complete MHC genomic sequences for eight haplotypes involved in autoimmune diseases, including the 8.1 AH, have been published.[7] In this regard, Shiina *et al.*[3] proposed, on the basis of comparative genomics between human haplotype sequences and the sequences of chimpanzee and rhesus macaque, that the rapid evolution of the MHC class I genes in primates is likely to have generated new disease alleles in humans through hitchhiking diversity.

The results of MHC disease association studies are complicated by race and population differences, influences of LD, the large polymorphism, copy number and InDel variations between different MHC haplotypes, disease severity and the need for large sample numbers to provide statistical significance. Fernando *et al.*[11] noted in their review of six autoimmune diseases with genetically complex disease traits that nearly all association studies of the MHC in autoimmune and inflammatory disease have been limited to a subset of ~20 genes and performed only in small cohorts of predominantly European origin. As highlighted in a recent review,[5] the MHC association with complex disease phenotypes is dependent on the HLA and non-HLA genes, the genetic code (SNPs, CNV, InDels and inversions), the epigenetic code (DNA methylation and histone modification), biological effects (structural and biochemical changes in gene products and transcriptional regulation) and environmental factors (diet and antigen exposure). Modern HLA and whole genome association studies of SNPs, microsatellites, InDels and CNVs are now broadening toward elucidating gene interactions, epistasis, risk and penetrance of autoimmune diseases,[162] although clear-cut results are often hampered by multiple testing errors and the statistical type I (false positives owing to multiple sample analysis) and statistical type II errors (false negatives owing to insufficient number of samples and other factors). Whole genome gene expression studies in combination with DNA variation and phenotypic data, as a single systematic study, have a greater potential for elucidating disease pathways and dissecting the role of individual genes and genomic loci, similar to the HLA superlocus, that interact in a molecular network. Such studies are still in their infancy, and much experimentation may be needed to overcome the potential data overload as we move rapidly toward a system genetics view of disease.[140]

HLA AND CANCER

The loss of HLA gene expression owing to viral infection, somatic mutations or other causes may have important effects on immune suppression and cancer development.[175] To identify the molecular mechanisms involved in the maintenance of Epstein–Barr virus (EBV)-associated epithelial cancers, Sengupta *et al.*[176] performed genome-wide expression profiling for all human genes and all latent EBV genes in a collection of 31 laser-captured, microdissected nasopharyngeal carcinoma (NPC) tissue samples and 10 normal nasopharyngeal tissues. They determined that all the HLA class I genes, *TAP2* and *HCG9* genes involved in regulating immune response through antigen presentation correlated negatively with increased EBV gene expression in NPC and concluded that antigen display is either directly inhibited by EBV, facilitating immune evasion by tumor cells and/or that tumor cells were selected for their EBV oncogene-mediated tumor-promoting actions. Global gene expression profiling of human papillomavirus (HPV)-positive and -negative head and neck cancers revealed a significant downregulation for two of the MHC genes, *CDSN* and *LY6G6C*, but not other MHC genes in HPV-16-positive head and neck squamous cell carcinomas.[177]

Non-viral tumors frequently lose expression of HLA molecules such as the reduction or total loss in colorectal carcinoma.[178] Cells participating in immune response may fail to exert function without adequate MHC signaling in tumor cells, with the exception of NK cells, which may recognize MHC class I-negative tumor cells. Furthermore, soluble MHC class I-related (MIC) molecules play important roles in tumor immune surveillance through their interaction with the NKG2D receptor on NK, NKT and cytotoxic T cells.[179,180] Interestingly, genome-wide expression profiling has shown that non-steroidal anti-inflammatory drug (NSAID) treatment upregulated HLA class II genes in tumor tissue, but not in normal colon tissue, from the same patient.[181] In total, 23 of the 100 most upregulated genes belonged to MHC class II; *HLA-DM*, *-DO* (peptide loading), *HLA-DP*, *-DQ*, *-DR* (antigen presentation), as did CD4+ T-helper cells, whereas *HLA-A* and *-C* expression were not increased by NSAID treatment.

In breast cancer, metastasis may be suppressed in part by the activity of the *breast cancer metastasis suppressor 1* (*BRMS1*) gene, which can block development of metastasis without preventing tumor growth. In a comparison of gene expression patterns in BRMS1-expressing vs non-expressing human breast carcinoma cells, the BRMS1 expression in 435/BRMS1 cells was strongly correlated with an increased expression of MHC genes, *HLA-DQB1*, *HLA-DRB1*, *HLA-DRB5*, *HLA-DMB*, *HLA-DQA1*, *HLA-DPA1*, *HLA-DRA*, *HLA-DRB4*, *HLA-DMA*, *C1S*, *HLA-B*, *HLA-C* and *HLA-F*.[182] Thus, the induction of MHC class I and II genes may be one mechanism by which 435/BRMS1 cells are kept at low populations, that is, by triggering an immune response that eliminates or reduces their metastasizing potential.

In an interesting paper by Rimsza *et al*.,[183] gene expression profiling data were used to correlate the expression levels of MHCII genes with each other and their transcriptional regulator, *CIITA* (16p13), in 240 cases of diffuse large B-cell lymphoma (240 cases in the LLMPP data set). A correlation map was created for expression of the genes that are telomeric (*HSPA1L*, *HSPA1A*, *BAT8*, *RDBP*, *CREBL1* and *PBX2*), within (*MHCII* genes, *TAP1*, *TAP2*, *PSMB9* and *BRD2*) or centromeric (*RXRB*, *RING1*, *RPS18*, *TAPBP*, *DAXX* and *BAK1*) to the *MHCII* locus. Correlation coefficients among *MHCII* genes were high (0.73–0.92), whereas those between adjacent and intervening genes were low (0.12–0.49). The authors concluded that the loss of *MHCII* expression in non-immune-privileged site diffuse large B-cell lymphoma is highly coordinated and not due to chromosomal deletions or rearrangements. Furthermore, Dave *et al*.[184] showed that gene expression profiling of MHC and non-MHC genes is an accurate, quantitative method for distinguishing Burkitt's lymphoma with the t(8;14) c-*myc* translocation from diffuse large-B-cell lymphoma. Burkitt's lymphoma was readily distinguished from diffuse large-B-cell lymphoma by the high-level expression of c-*myc* target genes and the low-level expression of all the MHC class I genes.

CONCLUSION

The human MHC genomic region is a super-locus composed of at least 250 coding and non-coding genes, the structural organization of which has evolved gradually, involving various mutation, duplication, deletion and genomic rearrangement events over a period of 450–520 Myr, at least from the time of the emergence of sharks (phylum Chordata, subphylum Vertebrata and class Chondrichthyes). A strong and progressive research interest remains toward haplotyping the entire human MHC genomic region by genomic resequencing for SNP, InDel and CNV analysis. The MHC genomic analysis was the prototype for many of the current procedures in genome-wide research, such as haplotyping, SNP and microsatellite analysis, and LD analysis for studies on human population diversity and disease association. The MHC genomic region is now part of the global systems analysis and network programs involved in the storage and dissemination of data on genome-wide gene expression at the level of the proteome, transcriptome, metabolome and phenotome, system and immune pathways, and disease associations using SNP, InDel and microsatellites as genomic markers or haplotype tags for statistical analysis. The degree and type of total MHC coordinated gene expression profiles have yet to be fully defined and understood in the processes of normal physiology, inflammatory and immune responses and autoimmune, chronic and infectious diseases. The field of MHC genomic research will clearly continue to expand into the future with the development of new procedures and studies to gain a better understanding of the intra- and extra-MHC gene interactions and their effects on human diversity and disease.

Website references

http://www.ncbi.nlm.nih.gov/sites/entrez Entrez Gene database
http://www.ncbi.nih.gov/entrez/query.fcgi?db=OMIM. OMIM: Online Mendelian Inheritance in Man
http://www.repeatmasker.org/ RepeatMasker program
http://www.ebi.ac.uk/imgt/hla/IMGT/HLA database: ImMunoGeneTics/HLA Sequence Database
http://espressosoftware.com/pages/sputnik.jsp Sputnik program
http://www.ncbi.nlm.nih.gov/gv/mhc/main.fcgi?cmd=init dbMHC database
http://www.ncbi.nlm.nih.gov/SNP/dbSNP database
http://projects.tcag.ca/variation/ Database of Genomic Variants
http://www.thebiogrid.org/index.php BioGRID: General Repository for Interaction Datasets
http://www.hprd.org/index_html HPRD: Human Protein Reference Database
http://www.ncbi.nlm.nih.gov/geo/ GEO: Gene Expression Omnibus
http://www.genego.com/pdf/PsoriasisCS.pdf MetaCore Applications
http://cgap.nci.nih.gov/Genes The Cancer Genome Anatomy Project
http://geneticassociationdb.nih.gov/cgi-bin/index.cgi GAD: Genetic Association Database

1 The MHC Sequencing Consortium. Complete structure and gene map of a human major histocompatibility complex (MHC). *Nature* **401,** 921–923 (1999).
2 Smith, W. P., Vu, Q., Li, S. S., Hansen, J. A., Zhao, L. P. & Geraghty, D. E. Toward understanding MHC disease associations: partial resequencing of 46 distinct HLA haplotypes. *Genomics* **87,** 561–571 (2006).
3 Shiina, T., Ota, M., Shimizu, S., Katsuyama, Y., Hashimoto, N., Takasu, M. *et al.* Rapid evolution of MHC class I genes in primates generates new disease alleles in man via hitchhiking diversity. *Genetics* **173,** 1555–1570 (2006).
4 Muller-Hilke, B. & Mitchison, N. A. The role of HLA promoters in autoimmunity. *Curr. Pharm. Des.* **12,** 3743–3752 (2006).
5 Traherne, J. A. Human MHC architecture and evolution: implications for disease association studies. *Int. J. Immunogenet.* **35,** 179–192 (2008).
6 Solberg, O. D., Mack, S. J., Lancaster, A. K., Single, R. M., Tsai, Y., Sanchez-Mazas, A. *et al.* Balancing selection and heterogeneity across the classical human leukocyte antigen loci: a meta-analytic review of 497 population studies. *Hum. Immunol.* **69,** 443–464 (2008).
7 Horton, R., Gibson, R., Coggill, P., Miretti, M., Allcock, R. J., Almeida, J. *et al.* Variation analysis and gene annotation of eight MHC haplotypes: the MHC Haplotype Project. *Immunogenetics* **60,** 1–18 (2008).
8 Fu, S., Zhao, H., Shi, J., Abzhanov, A., Crawford, K., Ohno-Machado, L. *et al.* Peripheral arterial occlusive disease: global gene expression analyses suggest a major role for immune and inflammatory responses. *BMC Genomics* **9,** 369 (2008).
9 Claas, F. H. J. & Duquesnoy, R. J. The polymorphic alloimmune response in clinical transplantation. Editorial overview. *Curr. Opin. Immunol.* **20,** 566–567 (2008).
10 Choo, S. Y. The HLA system: genetics, immunology, clinical testing, and clinical implications. *Yonsei Med. J.* **48,** 11–23 (2007).

11 Fernando, M. M., Stevens, C. R., Walsh, E. C., De Jager, P. L., Goyette, P., Plenge, R. M. *et al.* Defining the role of the MHC in autoimmunity: a review and pooled analysis. *PLoS Genet.* **4,** e1000024 (2008).
12 Martin, M. P. & Carrington, M. Immunogenetics of viral infections. *Curr. Opin. Immunol.* **17,** 510–516 (2005).
13 Fellay, J., Shianna, K. V., Ge, D., Colombo, S., Ledergerber, B., Weale, M. *et al.* A whole-genome association study of major determinants for host control of HIV-1. *Science* **317,** 944–947 (2007).
14 Knapp, L. A. The ABCs of MHC. *Evol. Anthropol.* **14,** 28–37 (2005).
15 Ziegler, A., Kentenich, H. & Uchanska-Ziegler, B. Female choice and the MHC. *Trends Immunol.* **26,** 496–502 (2005).
16 Xiao, B. G. & Link, H. Immune regulation within the central nervous system. *J. Neurol. Sci.* **157,** 1–12 (1998).
17 Huh, G. S., Boulanger, L. M., Du, H., Riquelme, P. A., Brotz, T. M. & Shatz, C. J. Functional requirement for class I MHC in CNS development and plasticity. *Science* **290,** 2155–2159 (2000).
18 Boulanger, L. M. & Shatz, C. J. Immune signaling in neural development, synaptic plasticity and disease. *Nat. Rev. Neurosci.* **5,** 521–531 (2004).
19 Cullheim, S. & Thams, S. The microglial networks of the brain and their role in neuronal network plasticity after lesion. *Brain Res. Rev.* **55,** 89–96 (2007).
20 Ohtsuka, M., Inoko, H., Kulski, J. K. & Yoshimura, S. Major histocompatibility complex (*Mhc*) class Ib gene duplications, organization and expression patterns in mouse strain C57BL/6. *BMC Genomics* **9,** 178 (2008).
21 Matsuo, R., Asada, A., Fujitani, K. & Inokuchi, K. LIRF, a gene induced during hippocampal long-term potentiation as an immediate-early gene, encodes a novel RING finger protein. *Biochem. Biophys. Res. Commun.* **289,** 479–484 (2001).
22 Patiño-Lopez, G., Hevezi, P., Lee, J., Willhite, D., Verge, G. M., Lechner, S. M. *et al.* Human class-I restricted T cell associated molecule is highly expressed in the cerebellum and is a marker for activated NKT and CD8+ T lymphocytes. *J. Neuroimmunol.* **171,** 145–155 (2006).
23 Goddard, C. A., Butts, D. A. & Shatz, C. A. Regulation of CNS synapses by neuronal MHC class I. *Proc. Natl Acad. Sci. USA* **104,** 6828–6833 (2007).
24 Tonelli, L. H., Postolache, T. T. & Sternberg, E. M. Inflammatory genes and neural activity: involvement of immune genes in synaptic function and behavior. *Front Biosci.* **10,** 675–680 (2005).
25 Lengen, C., Regard, M., Joller, H., Landis, T. & Lalive, P. Anomalous brain dominance and the immune system: do left-handers have specific immunological patterns? *Brain Cogn.* (2008) (e-pub ahead of print, 30 August 2008).
26 O'Keefe, G. M., Nguyen, V. T. & Benveniste, E. N. Regulation and function of class II major histocompatibility complex CD40, and B7 expression in macrophages and microglia: implications in neurological diseases. *J. Neurovirol.* **8,** 496–512 (2002).
27 Raha-Chowdhury, R., Andrews, S. R. & Gruen, J. R. CAT 53: a protein phosphatase 1 nuclear targeting subunit encoded in the MHC class I region strongly expressed in regions of the brain involved in memory, learning, and Alzheimer's disease. *Brain Res. Mol. Brain Res.* **138,** 70–83 (2005).
28 Cohly, H. H. & Panja, A. Immunological findings in autism. *Int. Rev. Neurobiol.* **71,** 317–341 (2005).
29 Bailey, S. L., Carpentier, P. A., McMahon, E. J., Begolka, W. S. & Miller, S. D. Innate and adaptive immune responses of the central nervous system. *Crit. Rev. Immunol.* **26,** 149–188 (2006).
30 McElroy, J. P. & Oksenberg, J. R. Multiple sclerosis genetics. *Curr. Top Microbiol. Immunol.* **318,** 45–72 (2008).
31 Shiina, T., Inoko, H. & Kulski, J. K. An update of the HLA genomic region, loci information and disease associations: 2004. *Tissue Antigens* **64,** 631–649 (2004).
32 Horton, R., Wilming, L., Rand, V., Lovering, R. C., Bruford, E. A., Khodiyar, V. K. *et al.* Gene map of the extended human MHC. *Nat. Rev. Genet.* **5,** 889–899 (2004).
33 Trowsdale, J. HLA genomics in the third millennium. *Curr. Opin. Immunol.* **17,** 498–504 (2005).
34 Stewart, C. A., Horton, R., Allcock, R. J., Ashurst, J. L., Atrazhev, A. M., Coggill, P. *et al.* Complete MHC haplotype sequencing for common disease gene mapping. *Genome Res.* **14,** 1176–1187 (2004).
35 Wheeler, D. L., Barrett, T., Benson, D. A., Bryant, S. H., Canese, K., Chetvernin, V *et al.* Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **35** (Database issue), D5–D12 (2007).
36 Maglott, D., Ostell, J., Pruitt, K. D. & Tatusova, T. Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res.* **33,** D54–D58 (2005).
37 Kulski, J. K. & Dawkins, R. L. The P5 multicopy gene family in the MHC is related in sequence to human endogenous retroviruses HERV-L and HERV-16. *Immunogenetics* **49,** 404–412 (1999).
38 Colombo, S., Rauch, A., Rotger, M., Fellay, J., Martinez, R., Fux, C. *et al.* Swiss HIV Cohort Study. The HCP5 single-nucleotide polymorphism: a simple screening tool for prediction of hypersensitivity reaction to abacavir. *J. Infect. Dis.* **198,** 864–867 (2008).
39 Liu, Y., Helms, C., Liao, W., Zaba, L. C., Duan, S., Gardner, J. *et al.* A genome-wide association study of psoriasis and psoriatic arthritis identifies new disease loci. *PLoS Genet.* **4,** e1000041 (2008).
40 Chang, Y. T., Chou, C. T., Shiao, Y. M., Lin, M. W., Yu, C. W., Chen, C. C. *et al.* Psoriasis vulgaris in Chinese individuals is associated with PSORS1C3 and CDSN genes. *Br. J. Dermatol.* **155,** 663–669 (2006).
41 Semple, J. I., Ribas, G., Hillyard, G., Brown, S. E., Sanderson, C. M. & Campbell, R. D. A novel gene encoding a coiled-coil mitochondrial protein located at the telomeric end of the human MHC class III region. *Gene* **314,** 41–54 (2003).
42 Ostrov, D. A., Barnes, C. L., Smith, L. E., Binns, S., Brusko, T. M., Brown, A. C. *et al.* Characterization of HKE2: an ancient antigen encoded in the major histocompatibility complex. *Tissue Antigens* **69,** 181–188 (2007).
43 de Vet, E. C., Aguado, B. & Campbell, R. D. G6b, a novel immunoglobulin superfamily member encoded in the human major histocompatibility complex, interacts with SHP-1 and SHP-2. *J. Biol. Chem.* **276,** 42070–42074 (2001).
44 Adams, P. C. & Barton, J. C. Haemochromatosis. *Lancet* **370,** 1855–1860 (2007).
45 Kulski, J. K. & Inoko, H. MHC genes in *The Encyclopedia of the Human Genome* 778–785 (Nature Publishing Group. Macmillan Publishers Ltd, Houndmills, Basingstoke, Hampshire, UK, 2003).
46 Dawkins, R., Leelayuwat, C., Gaudieri, S., Tay, G., Hui, J., Cattley, S. *et al.* Genomics of the major histocompatibility complex: haplotypes, duplication, retroviruses and disease. *Immunol. Rev.* **167,** 275–304 (1999).
47 Bahram, S. MIC genes: from genetics to biology. *Adv. Immunol.* **76,** 1–60 (2000).
48 Kulski, J. K., Shiina, T., Anzai, T., Kohara, S. & Inoko, H. Comparative genomic analysis of the MHC: the evolution of class I duplication blocks, diversity and complexity from shark to man. *Immunol. Rev.* **190,** 95–122 (2002).
49 Shiina, T., Tamiya, G., Oka, A., Takishima, N., Yamagata, T., Kikkawa, E. *et al.* Molecular dynamics of MHC genesis unraveled by sequencing analysis of the 1,796,938 bp HLA class I region. *Proc. Natl Acad. Sci. USA* **96,** 13282–13287 (1999).
50 García, A., Senis, Y. A., Antrobus, R., Hughes, C. E., Dwek, R. A., Watson, S. P. *et al.* A global proteomics approach identifies novel phosphorylated signaling proteins in GPVI-activated platelets: involvement of G6f, a novel platelet Grb2-binding membrane adapter. *Proteomics* **6,** 5332–5343 (2006).
51 Jady, B. E. & Kiss, T. Characterisation of the U83 and U84 small nucleolar RNAs: two novel 2′-O-ribose methylation guide RNAs that lack complementarities to ribosomal RNAs. *Nucleic Acids Res.* **28,** 1348–1354 (2000).
52 Lestrade, L. & Weber, M. J. snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs. *Nucleic Acids Res.* **34,** D158–D162 (2006).
53 Kiss-Laszlo, Z., Henry, Y., Bachellerie, J. P., Caizergues-Ferrer, M. & Kiss, T. Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs. *Cell* **85,** 1077–1088 (1996).
54 Florence, B. & Faller, D. V. You bet-cha: a novel family of transcriptional regulators. *Front. Biosci.* **6,** d1008–d1018 (2001).
55 Kohany, O., Gentles, A. J., Hankus, L. & Jurka, J. Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. *BMC Bioinformatics* **7,** 474 (2006).
56 Gourraud, P. A., Mano, S., Barnetche, T., Carrington, M., Inoko, H. & Cambon-Thomsen, A. Integration of microsatellite characteristics in the MHC region: a literature and sequence based analysis. *Tissue Antigens* **64,** 543–555 (2004).
57 Matsuzaka, Y., Tounai, K., Denda, A., Tomizawa, M., Makino, S., Okamoto, K. *et al.* Identification of novel candidate genes in the diffuse panbronchiolitis critical region of the class I human MHC. *Immunogenetics.* **54,** 301–309 (2002).
58 Oka, A., Tamiya, G., Tomizawa, M., Ota, M., Katsuyama, Y., Makino, S. *et al.* Association analysis using refined microsatellite markers localizes a susceptibility locus for psoriasis vulgaris within a 111 kb segment telomeric to the HLA-C gene. *Hum. Mol. Genet.* **8,** 2165–2170 (1999).
59 Aly, T. A., Eller, E., Ide, A., Gowan, K., Babu, S. R., Erlich, H. A. *et al.* Multi-SNP analysis of MHC region: remarkable conservation of HLA-A1-B8-DR3 haplotype. *Diabetes* **55,** 1265–1269 (2006).
60 de Bakker, P. I., McVean, G., Sabeti, P. C., Miretti, M. M., Green, T., Marchini, J. *et al.* A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. *Nat. Genet.* **38,** 1166–1172 (2006).
61 Smith, W. P., Vu, Q., Li, S. S., Hansen, J. A., Zhao, L. P. & Geraghty, D. E. Toward understanding MHC disease associations: partial resequencing of 46 distinct HLA haplotypes. *Genomics* **87,** 561–571 (2006).
62 Romero, V., Larsen, C. E., Duke-Cohan, J. S., Fox, E. A., Romero, T., Clavijo, O. P. *et al.* Genetic fixity in the human major histocompatibility complex and block size diversity in the class I region including HLA-E. *BMC Genet.* **8,** 14 (2007).
63 Bjorkman, P. J. & Parham, P. Structure, function, and diversity of class I major histocompatibility complex molecules. *Annu. Rev. Biochem.* **59,** 253–288 (1990).
64 Okamoto, K., Makino, S., Yoshikawa, Y., Takaki, A., Nagatsuka, Y., Ota, M. *et al.* Identification of I kappa BL as the second major histocompatibility complex-linked susceptibility locus for rheumatoid arthritis. *Am. J. Hum. Genet.* **72,** 303–312 (2003).
65 Valentonyte, R., Hampe, J., Huse, K., Rosenstiel, P., Albrecht, M., Stenzel, A. *et al.* Sarcoidosis is associated with a truncating splice site mutation in BTNL2. *Nat. Genet.* **37,** 357–364 (2005).
66 Vincek, V., Klein, D., Figueroa, F., Hauptfeld, V., Kasahara, M., O'hUigin, C. *et al.* The evolutionary origin of the HLA-DR3 haplotype. *Immunogenetics* **35,** 263–271 (1992).
67 Marsh, S. G. WHO Nomenclature Committee for Factors of the HLA System. Nomenclature for factors of the HLA system, update July 2000. *Tissue Antigens* **56,** 476–477 (2000).
68 Yang, Y., Chung, E. K., Wu, Y. L., Savelli, S. L., Nagaraja, H. N., Zhou, B. *et al.* Gene copy-number variation and associated polymorphisms of complement component C4 in human systemic lupus erythematosus (SLE): low copy number is a risk factor for and high copy number is a protective factor against SLE susceptibility in European Americans. *Am. J. Hum. Genet.* **80,** 1037–1054 (2007).
69 Tuzun, E., Sharp, A. J., Bailey, J. A., Kaul, R., Morrison, V. A., Pertz, L. M. *et al.* Fine-scale structural variation of the human genome. *Nat. Genet.* **37,** 727–732 (2005).
70 Redon, R., Ishikawa, S., Fitch, K. R., Feuk, L., Perry, G. H., Andrews, T. D. *et al.* Global variation in copy number in the human genome. *Nature* **444,** 444–454 (2006).

71 Conrad, D. F., Andrews, T. D., Carter, N. P., Hurles, M. E. & Pritchard, J. K. A high-resolution survey of deletion polymorphism in the human genome. *Nat. Genet.* **38,** 75–81 (2006).
72 McCarroll, S. A., Hadnott, T. N., Perry, G. H., Sabeti, P. C., Zody, M. C., Barrett, J. C. *et al.* International HapMap Consortium. Common deletion polymorphisms in the human genome. *Nat. Genet.* **38,** 86–92 (2006).
73 Mills, R. E., Luttig, C. T., Larkins, C. E., Beauchamp, A., Tsui, C., Pittard, W. S. *et al.* An initial map of insertion and deletion (INDEL) variation in the human genome. *Genome Res.* **16,** 1182–1190 (2006).
74 Simon-Sanchez, J., Scholz, S., Fung, H. C., Matarin, M., Hernandez, D., Gibbs, J. R. *et al.* Genome-wide SNP assay reveals structural genomic variation, extended homozygosity and cell-line induced alterations in normal individuals. *Hum. Mol. Genet.* **16,** 1–14 (2007).
75 Wong, K. K., deLeeuw, R. J., Dosanjh, N. S., Kimm, L. R., Cheng, Z., Horsman, D. E. *et al.* A comprehensive analysis of common copy-number variations in the human genome. *Am. J. Hum. Genet.* **80,** 91–104 (2007).
76 Korbel, J. O., Urban, A. E., Affourtit, J. P., Godwin, B., Grubert, F., Simons, J. F. *et al.* Paired-end mapping reveals extensive structural variation in the human genome. *Science* **318,** 420–426 2007.
77 Levy, S., Sutton, G., Ng, P. C., Feuk, L., Halpern, A. L., Walenz, B. P. *et al.* The diploid genome sequence of an individual human. *PLoS Biol.* **5,** e254 (2007).
78 Perry, G. H., Ben-Dor, A., Tsalenko, A., Sampas, N., Rodriguez-Revenga, L., Tran, C. W. *et al.* The fine-scale and complex architecture of human copy-number variation. *Am. J. Hum. Genet.* **82,** 685–695 (2008).
79 Kidd, J. M., Cooper, G. M., Donahue, W. F., Hayden, H. S., Sampas, N., Graves, T. *et al.* Mapping and sequencing of structural variation from eight human genomes. *Nature* **453,** 56–64 (2008).
80 Komatsu-Wakui, M., Tokunaga, K., Ishikawa, Y., Kashiwase, K., Moriyama, S., Tsuchiya, N. *et al.* MIC-A polymorphism in Japanese and a MIC-A–MIC-B null haplotype. *Immunogenetics* **49,** 620–628 (1999).
81 Kulski, J. K. & Dunn, D. S. Polymorphic Alu insertions within the major histocompatibility complex class I genomic region: a brief review. *Cytogenet. Genome Res.* **110,** 193–202 (2005).
82 Takasu, M., Hayashi, R., Maruya, E., Ota, M., Imura, K., Kougo, K. *et al.* Deletion of entire HLA-A gene accompanied by an insertion of a retrotransposon. *Tissue Antigens* **70,** 144–150 (2007).
83 Kulski, J. K., Shigenari, A., Shiina, T., Ota, M., Hosomichi, K., James, I. *et al.* Human endogenous retrovirus (HERVK9) structural polymorphism with haplotypic HLA-A allelic associations. *Genetics* **180,** 445–457 (2008).
84 Cordell, H. J. Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans. *Hum. Mol. Genet.* **11,** 2463–2468 (2002).
85 Navarro, A. & Barton, N. H. Effects of multilocus balancing selection on neutral variability. *Genetics* **161,** 849–863 (2002).
86 Miretti, M. M., Walsh, E. C., Ke, X., Delgado, M., Griffiths, M., Hunt, S. *et al.* A high-resolution linkage–disequilibrium map of the human major histocompatibility complex and first generation of tag single-nucleotide polymorphisms. *Am. J. Hum. Genet.* **76,** 634–646 (2005).
87 Blomhoff, A., Olsson, M., Johansson, S., Akselsen, H. E., Pociot, F., Nerup, J. *et al.* Linkage disequilibrium and haplotype blocks in the MHC vary in an HLA haplotype specific manner assessed mainly by DRB1*03 and DRB1*04 haplotypes. *Genes Immun.* **7,** 130–140 (2006).
88 Mathivanan, S., Periaswamy, B., Gandhi, T. K., Kandasamy, K., Suresh, S., Mohmood, R. *et al.* An evaluation of human protein–protein interaction data in the public domain. *BMC Bioinformatics* **7** (Suppl 5), S19 (2006).
89 Gomez, J. A., Majumder, P., Nagarajan, U. M. & Boss, J. M. X box-like sequences in the MHC class II region maintain regulatory function. *J. Immunol.* **175,** 1030–1040 (2005).
90 Müller-Hilke, B. & Mitchison, N. A. The role of HLA promoters in autoimmunity. *Curr. Pharm. Des.* **12,** 3743–3752 (2006).
91 Christova, R., Jones, T., Wu, P. J., Bolzer, A., Costa-Pereira, A. P., Watling, D. *et al.* P-STAT1 mediates higher-order chromatin remodelling of the human MHC in response to IFNgamma. *J. Cell Sci.* **120,** 3262–3270 (2007).
92 Kumar, P. P., Bischof, O., Purbey, P. K., Notani, D., Urlaub, H., Dejean, A. *et al.* Functional interaction between PML and SATB1 regulates chromatin-loop architecture and transcription of the MHC class I locus. *Nat. Cell Biol.* **9,** 45–56 (2007).
93 Brown, J. H., Jardetzky, T. S., Gorga, J. C., Stern, L. J., Urban, R. G., Strominger, J. L. *et al.* Three dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature* **364,** 33–39 (1993).
94 Strominger, J. L. Human histocompatibility proteins. *Immunol. Rev.* **185,** 69–77 (2002).
95 Momburg, F. & Tan, P. Tapasin—the keystone of the loading complex optimizing peptide binding by MHC class I molecules in the endoplasmic reticulum. *Mol. Immunol.* **39,** 217–233 (2002).
96 Sadegh-Nasseri, S., Chen, M., Narayan, K. & Bouvier, M. The convergent roles of tapasin and HLA-DM in antigen presentation. *Trends Immunol.* **29,** 141–147 (2008).
97 Gregersen, J. W., Kranc, K. R., Ke, X., Svendsen, P., Madsen, L. S., Thomsen, A. R. *et al.* Functional epistasis on a common MHC haplotype associated with multiple sclerosis. *Nature* **443,** 574–577 (2006).
98 Martin, A. M., Kulski, J. K., Witt, C., Pontarotti, P. & Christiansen, F. T. Leukocyte Ig-like receptor complex (LRC) in mice and men. *Trends Immunol.* **23,** 81–88 (2002).
99 Thananchai, H., Gillespie, G., Martin, M. P., Bashirova, A., Yawata, N., Yawata, M. *et al.* Cutting edge: allele-specific and peptide-dependent interactions between KIR3DL1 and HLA-A and HLA-B. *J. Immunol.* **178,** 33–37 (2007).
100 Parham, P. Influence of KIR diversity on human immunity. *Adv. Exp. Med. Biol.* **560,** 47–50 (2005).
101 Khakoo, S. I. & Carrington, M. KIR and disease: a model system or system of models? *Immunol. Rev.* **214,** 186–201 (2006).
102 Swaroop, A., Branham, K. E., Chen, W. & Abecasis, G. Genetic susceptibility to age-related macular degeneration: a paradigm for dissecting complex disease traits. *Hum. Mol. Genet.* **16** (spec. no. 2), R174–R182 (2007).
103 Lester, S., McLure, C., Williamson, J., Bardy, P., Rischmueller, M. & Dawkins, R. L. Epistasis between the MHC and the RCA alpha block in primary Sjögren syndrome. *Ann. Rheum. Dis.* **67,** 849–854 (2008).
104 Sánchez, F. O., Linga Reddy, M. V., Sakuraba, K., Ståhle, M. & Alarcón-Riquelme, M. E. IFN-regulatory factor 5 gene variants interact with the class I MHC locus in the Swedish psoriasis population. *J. Invest. Dermatol.* **128,** 1704–1709 (2008).
105 Niewold, T. B., Kelly, J. A., Flesch, M. H., Espinoza, L. R., Harley, J. B. & Crow, M. K. Association of the IRF5 risk haplotype with high serum interferon-alpha activity in systemic lupus erythematosus patients. *Arthritis Rheum.* **58,** 2481–2487 (2008).
106 Sugawara, T., Shimizu, H., Hoshi, N., Nakajima, A. & Fujimoto, S. Steroidogenic acute regulatory protein-binding protein cloned by a yeast two-hybrid system. *J. Biol. Chem.* **278,** 42487–42494 (2003).
107 Tiala, I., Suomela, S., Huuhtanen, J., Wakkinen, J., Hölttä-Vuori, M., Kainu, K. *et al.* The CCHCR1 (HCR) gene is relevant for skin steroidogenesis and downregulated in cultured psoriatic keratinocytes. *J. Mol. Med.* **85,** 589–601 (2007).
108 Kulski, J. K., Kenworthy, W., Bellgard, M., Taplin, R., Okamoto, K., Oka, A. *et al.* Gene expression profiling of Japanese psoriatic skin reveals an increased activity in molecular stress and immune response signals. *J. Mol. Med.* **83,** 964–975 (2005).
109 Zhou, X., Krueger, J. G., Kao, M. C., Lee, E., Du, F., Menter, A. *et al.* Novel mechanisms of T-cell and dendritic cell activation revealed by profiling of psoriasis on the 63 100-element oligonucleotide array. *Physiol. Genomics* **13,** 69–78 (2003).
110 Mee, J. B., Johnson, C. M., Morar, N., Burslem, F. & Groves, R. W. The psoriatic transcriptome closely resembles that induced by interleukin-1 in cultured keratinocytes: dominance of innate immune responses in psoriasis. *Am. J. Pathol.* **171,** 32–42 (2007).
111 Nomura, I., Gao, B., Boguniewicz, M., Darst, M. A., Travers, J. B. & Leung, D. Y. Distinct patterns of gene expression in the skin lesions of atopic dermatitis and psoriasis: a gene microarray analysis. *J. Allergy Clin. Immunol.* **112,** 1195–1202 (2003).
112 Zhang, Z. H., Wang, Z. M., Crosby, M. E., Wang, H. F., Xiang, L. H., Luan, J. *et al.* Gene expression profiling of porokeratosis. *J. Cutan. Pathol.* **35,** 1058–1062 (2008).
113 Koczan, D., Guthke, R., Thiesen, H. J., Ibrahim, S. M., Kundt, G., Krentz, H. *et al.* Gene expression profiling of peripheral blood mononuclear leukocytes from psoriasis patients identifies new immune regulatory molecules. *Eur. J. Dermatol.* **15,** 251–257 (2005).
114 Chaussabel, D., Quinn, C., Shen, J., Patel, P., Glaser, C., Baldwin, N. *et al.* A modular analysis framework for blood genomics studies: application to systemic lupus erythematosus. *Immunity* **29,** 150–164 (2008).
115 Haider, A. S., Lowes, M. A., Suárez-Fariñas, M., Zaba, L. C., Cardinale, I., Khatcherian, A. *et al.* Identification of cellular pathways of 'type 1,' Th17T cells, and TNF- and inducible nitric oxide synthase-producing dendritic cells in autoimmune inflammation through pharmacogenomic study of cyclosporine A in psoriasis. *J. Immunol.* **180,** 1913–1920 (2008).
116 Nzeusseu Toukap, A., Galant, C., Theate, I., Maudoux, A. L., Lories, R. J., Houssiau, F. A. *et al.* Identification of distinct gene expression profiles in the synovium of patients with systemic lupus erythematosus. *Arthritis Rheum.* **56,** 1579–1588 (2007).
117 Bovin, L. F., Rieneck, K., Workman, C., Nielsen, H., Sørensen, S. F., Skjødt, H. *et al.* Blood cell gene expression profiling in rheumatoid arthritis. Discriminative genes and effect of rheumatoid factor. *Immunol. Lett.* **93,** 217–226 (2004).
118 van der Pouw Kraan, T. C., van Baarsen, L. G., Rustenburg, F., Baltus, B., Fero, M. & Verweij, C. L. Gene expression profiling in rheumatology. *Methods Mol. Med.* **136,** 305–327 (2007).
119 Devauchelle, V., Marion, S., Cagnard, N., Mistou, S., Falgarone, G., Breban, M. *et al.* DNA microarray allows molecular profiling of rheumatoid arthritis and identification of pathophysiological targets. *Genes Immun.* **5,** 597–608 (2004).
120 van der Pouw Kraan, T. C., van Gaalen, F. A., Huizinga, T. W., Pieterman, E., Breedveld, F. C. & Verweij, C. L. Discovery of distinctive gene expression profiles in rheumatoid synovium using cDNA microarray technology: evidence for the existence of multiple pathways of tissue destruction and repair. *Genes Immun.* **4,** 187–196 (2003).
121 Batliwalla, F. M., Li, W., Ritchlin, C. T., Xiao, X., Brenner, M., Laragione, T. *et al.* Microarray analyses of peripheral blood cells identifies unique gene expression signature in psoriatic arthritis. *Mol. Med.* **11,** 21–29 (2005).
122 Fu, S., Zhao, H., Shi, J., Abzhanov, A., Crawford, K., Ohno-Machado, L. *et al.* Peripheral arterial occlusive disease: global gene expression analyses suggest a major role for immune and inflammatory responses. *BMC Genomics* **9,** 369 (2008).
123 Nisole, S., Stoye, J. P. & Saïb, A. TRIM family proteins: retroviral restriction and antiviral defence. *Nat. Rev. Microbiol.* **3,** 799–808 (2005).
124 Uchil, P. D., Quinlan, B. D., Chan, W. T., Luna, J. M. & Mothes, W. TRIM E3 ligases interfere with early and late stages of the retroviral life cycle. *PLoS Pathog.* **4,** e16 (2008).
125 Sebat, J., Lakshmi, B., Malhotra, D., Troge, J., Lese-Martin, C., Walsh, T. *et al.* Strong association of *de novo* copy number mutations with autism. *Science* **316,** 445–449 (2007).
126 Iafrate, A. J., Feuk, L., Rivera, M. N., Listewnik, M. L., Donahoe, P. K., Qi, Y. *et al.* Detection of large-scale variation in the human genome. *Nat. Genet.* **36,** 949–951 (2004).

127 Fellermann, K., Stange, D. E., Schaeffeler, E., Schmalzl, H., Wehkamp, J., Bevins, C. L. *et al.* A chromosome 8 gene-cluster polymorphism with low human beta-defensin 2 gene copy number predisposes to Crohn disease of the colon. *Am. J. Hum. Genet.* **79,** 439–448 (2006).
128 Hollox, E. J., Huffmeier, U., Zeeuwen, P. L., Palla, R., Lascorz, J., Rodijk-Olthuis, D. *et al.* Psoriasis is associated with increased beta-defensin genomic copy number. *Nat. Genet.* **40,** 23–25 (2008).
129 Jiang, L., Yu, Z., Du, W., Tang, Z., Jiang, T., Zhang, C. *et al.* Development of a fluorescent and colorimetric detection methods-based protein microarray for serodiagnosis of TORCH infections. *Biosens. Bioelectron* **24,** 376–382 (2008).
130 Tomazou, E. M., Rakyan, V. K., Lefebvre, G., Andrews, R., Ellis, P., Jackson, D. K. *et al.* Generation of a genomic tiling array of the human major histocompatibility complex (MHC) and its application for DNA methylation analysis. *BMC Med. Genomics* **1,** 19 (2008).
131 Serrano, N. C., Millan, P. & Páez, M. C. Non-HLA associations with autoimmune diseases. *Autoimmun. Rev.* **5,** 209–214 (2006).
132 Samanta, M. P., Tongprasit, W. & Stolc, V. In-depth query of large genomes using tiling arrays. *Methods Mol. Biol.* **377,** 163–174 (2007).
133 Borrebaeck, C. A. & Wingren, C. High-throughput proteomics using antibody microarrays: an update. *Expert Rev. Mol. Diagn.* **7,** 673–686 (2007).
134 Tao, S. C., Chen, C. S. & Zhu, H. Applications of protein microarray technology. *Comb. Chem. High Throughput Screen.* **10,** 706–718 (2007).
135 Chen, D. S., Soen, Y., Stuge, T. B., Lee, P. P., Weber, J. S., Brown, P. O. *et al.* Marked differences in human melanoma antigen-specific T cell responsiveness after vaccination using a functional microarray. *PLoS Med.* **2,** e265 (2005).
136 Fortier, M. H., Caron, E., Hardy, M. P., Voisin, G., Lemieux, S., Perreault, C. *et al.* The MHC class I peptide repertoire is molded by the transcriptome. *J. Exp. Med.* **205,** 595–610 (2008).
137 Ho, P. P., Higgins, J. P., Kidd, B. A., Tomooka, B., Digennaro, C., Lee, L. Y. *et al.* Tolerizing DNA vaccines for autoimmune arthritis. *Autoimmunity* **39,** 675–682 (2006).
138 Lueking, A., Huber, O., Wirths, C., Schulte, K., Stieler, K. M., Blume-Peytavi, U. *et al.* Profiling of alopecia areata autoantigens based on protein microarray technology. *Mol. Cell Proteomics* **4,** 1382–1390 (2005).
139 El Essawy, B., Otu, H. H., Choy, B., Zheng, X. X., Libermann, T. A. & Strom, T. B. Proteomic analysis of the allograft response. *Transplantation* **82,** 267–274 (2006).
140 Sieberts, S. K. & Schadt, E. E. Moving toward a system genetics view of disease. *Mamm. Genome* **18,** 389–401 (2007).
141 Amiel, J. L. Study of the leukocyte phenotypes in Hodgkin's disease in *Histocompatibility Testing* 79–81 (eds Curtoni, E.S., Mattiuz, P.L., Tosi, R.M.) (Munksgaard, Copenhagen, 1967).
142 Tiwari, J. L. & Terasaki, P. I. *HLA and Disease Association* (Springer-Verlag, New York, 1985).
143 Naito, S. The association of HLA with diseases in Japanese. *J. Hum. Genet.* **31,** 323–329 (1986).
144 Geluk, A. & Ottenhoff, T. H. HLA and leprosy in the pre and postgenomic eras. *Hum. Immunol.* **67,** 439–445 (2006).
145 Mehra, N. K. & Kaur, G. 14th International HLA and Immunogenetics Workshop: report on joint study on MHC and infection. *Tissue Antigens* **69,** 226–227 (2007).
146 Goulder, P. J. & Watkins, D. I. Impact of MHC class I diversity on immune control of immunodeficiency virus replication. *Nat. Rev. Immunol.* **8,** 619–630 (2008).
147 Lie, B. A. & Thorsby, E. Several genes in the extended human MHC contribute to predisposition to autoimmune diseases. *Curr. Opin. Immunol.* **17,** 526–531 (2005).
148 Jones, E. Y., Fugger, L., Strominger, J. L. & Siebold, C. MHC class II proteins and disease: a structural perspective. *Nat. Rev. Immunol.* **6,** 271–282 (2006).
149 Schmidt, H., Williamson, D. & Ashley-Koch, A. HLA-DR15 haplotype and multiple sclerosis: a HuGE review. *Am. J. Epidemiol.* **165,** 1097–1109 (2007).
150 Jacobson, E. M., Huber, A. & Tomer, Y. The HLA gene complex in thyroid autoimmunity: from epidemiology to etiology. *J. Autoimmun.* **30,** 58–62 (2008).
151 Hamosh, A., Scott, A. F., Amberger, J., Valle, D. & McKusick, V. A. Online Mendaian Inheritance in Man (OMIM). *Hum. Mutat.* **15,** 57–61 (2000).
152 Hamosh, A., Scott, A. F., Amberger, J. S., Bocchini, C. A. & McKusick, V. A. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.* **33,** D514–D517 (2005).
153 Naoum, J. J., Chai, H., Lin, P. H., Lumsden, A. B., Yao, Q. & Chen, C. Lymphotoxin-alpha and cardiovascular disease: clinical association and pathogenic mechanisms. *Med. Sci. Monit.* **12,** RA121–RA124 (2006).
154 London, S. J. Gene-air pollution interactions in asthma. *Proc. Am. Thorac. Soc.* **4,** 217–220 (2007).
155 Sharma, S., Ghosh, B. & Sharma, S. K. Association of TNF polymorphisms with sarcoidosis, its prognosis and tumour necrosis factor (TNF)-alpha levels in Asian Indians. *Clin. Exp. Immunol.* **151,** 251–259 (2008).
156 Vallvé, J. C., Paredes, S., Girona, J., Uliaque, K., Ribalta, J., Hurt-Camejo, E. *et al.* Tumor necrosis factor-alpha −1031 T/C polymorphism is associated with smaller and more proatherogenic low density lipoprotein particles in patients with rheumatoid arthritis. *J. Rheumatol.* **35,** 1697–1703 (2008).
157 Lee, H. H., Lee, Y. J., Wang, Y. M., Chao, H. T., Niu, D. M., Chao, M. C. *et al.* Low frequency of the CYP21A2 deletion in ethnic Chinese (Taiwanese) patients with 21 hydroxylase deficiency. *Mol. Genet. Metab.* **93,** 450–457 (2008).
158 Lee, H. H., Tsai, F. J., Lee, Y. J. & Yang, Y. C. Diversity of the CYP21A2 gene:a 6.2-kb TaqI fragment and a 3.2-kb TaqI fragment mistaken as CYP21A1P. *Mol. Genet. Metab.* **88,** 372–377 (2006).
159 Seyrantepe, V., Poupetova, H., Froissart, R., Zabot, M. T., Maire, I. & Pshezhetsky, A. V. Molecular pathology of NEU1 gene in sialidosis. *Hum. Mutat.* **22,** 343–352 (2003).
160 Sjöholm, A. G., Jönsson, G., Braconier, J. H., Sturfelt, G. & Truedsson, L. Complement deficiency and disease: an update. *Mol. Immunol.* **43,** 78–85 (2006).
161 Forabosco, P., Bouzigon, E., Ng, M. Y., Hermanowski, J., Fishe, R. S. A., Criswell, L. A. *et al.* Meta-analysis of genome-wide linkage studies across autoimmune diseases. *Eur. J. Hum. Genet.* 2008 (e-pub ahead of print, 10 September 2008).
162 Wellcome Trust Case Control Consortium. Genome-wide association study of 14 000 cases of seven common diseases and 3000 shared controls. *Nature* **447,** 661–678 2007.
163 Choi, H. B., Han, H., Youn, J. I., Kim, T. Y. & Kim, T. G. MICA 5.1 allele is a susceptibility marker for psoriasis in the Korean population. *Tissue Antigens* **56,** 548–550 (2000).
164 Asumalahti, K., Veal, C., Laitinen, T., Suomela, S., Allen, M., Elomaa, O. *et al.* Psoriasis Consortium. Coding haplotype analysis supports HCR as the putative susceptibility gene for psoriasis at the MHC PSORS1 locus. *Hum. Mol. Genet.* **11,** 589–597 (2002).
165 Chang, Y. T., Shiao, Y. M., Chin, P. J., Liu, Y. L., Chou, F. C., Wu, S. *et al.* Genetic polymorphisms of the HCR gene and a genomic segment in close proximity to HLA-C are associated with patients with psoriasis in Taiwan. *Br. J. Dermatol.* **150,** 1104–1111 (2004).
166 Ameen, M., Allen, M. H., Fisher, S. A., Lewis, C. M., Cuthbert, A., Kondeatis, E. *et al.* Corneodesmosin (CDSN) gene association with psoriasis vulgaris in Caucasian but not in Japanese populations. *Clin. Exp. Dermatol.* **30,** 414–418 (2005).
167 Helms, C., Saccone, N. L., Cao, L., Daw, J. A., Cao, K., Hsu, T. M. *et al.* Localization of PSORS1 to a haplotype block harboring HLA-C and distinct from corneodesmosin and HCR. *Hum. Genet.* **118,** 466–476 (2005).
168 Martínez-Borra, J., Brautbar, C., González, S., Enk, C. D., López-Vázquez, A. & López-Larrea, C. The region of 150 kb telometic to HLA-C is associated with psoriasis in the Jewish population. *J. Invest. Dermatol.* **125,** 928–932 (2005).
169 Nair, R. P., Stuart, P. E., Nistor, I., Hiremagalore, R., Chia, N. V., Jenisch, S. *et al.* Sequence and haplotype analysis supports HLA-C as the psoriasis susceptibility 1 gene. *Am. J. Hum. Genet.* **78,** 827–851 (2006).
170 Takemoto, T., Naruse, T., Namba, K., Kitaichi, N., Ota, M., Shindo, Y. *et al.* Re-evaluation of heterogeneity in HLA-B*510101 associated with Behçet's disease. *Tissue Antigens* **72,** 347–353 (2008).
171 Karasneh, J., Gül, A., Ollier, W. E., Silman, A. J. & Worthington, J. Whole-genome screening for susceptibility genes in multicase families with Behçet's disease. *Arthritis Rheum.* **52,** 1836–1842 (2005).
172 Oksenberg, J. R., Baranzini, S. E., Sawcer, S. & Hauser, S. L. The genetics of multiple sclerosis: SNPs to pathways to pathogenesis. *Nat. Rev. Genet.* **9,** 516–526 (2008).
173 Degli-Esposti, M. A., Leaver, A. L., Christiansen, F. T., Witt, C. S., Abraham, L. J. & Dawkins, R. L. Ancestral haplotypes: conserved population MHC haplotypes. *Hum. Immunol.* **34,** 242–252 (1992).
174 Price, P., Witt, C., Allcock, R., Sayer, D., Garlepp, M., Kok, C. C. *et al.* The genetic basis for the association of the 8.1 ancestral haplotype (A1, B8, DR3) with multiple immunopathological diseases. *Immunol. Rev.* **167,** 257–274 (1999).
175 Aptsiauri, N., Cabrera, T., Garcia-Lora, A., Lopez-Nevot, M. A., Ruiz-Cabello, F. & Garrido, F. MHC class I antigens and immune surveillance in transformed cells. *Int. Rev. Cytol.* **256,** 139–189 (2007).
176 Sengupta, S., den Boon, J. A., Chen, I. H., Newton, M. A., Dahl, D. B., Chen, M. *et al.* Genome-wide expression profiling reveals EBV-associated inhibition of MHC class I expression in nasopharyngeal carcinoma. *Cancer Res.* **66,** 7999–8006 (2006).
177 Schlecht, N. F., Burk, R. D., Adrien, L., Dunne, A., Kawachi, N., Sarta, C. *et al.* Gene expression profiles in HPV-infected head and neck cancer. *J. Pathol.* **213,** 283–293 (2007).
178 Watson, N. F., Ramage, J. M., Madjd, Z., Spendlove, I., Ellis, I. O., Scholefield, J. H. *et al.* Immunosurveillance is active in colorectal cancer as downregulation but not complete loss of MHC class I expression correlates with a poor prognosis. *Int. J. Cancer* **118,** 6–10 (2006).
179 Wang, H., Yang, D., Xu, W., Wang, Y., Ruan, Z., Zhao, T. *et al.* Tumor-derived soluble MICs impair CD3(+)CD56(+) NKT-like cell cytotoxicity in cancer patients. *Immunol. Lett.* **120,** 65–71 (2008).
180 Groh, V., Wu, J., Yee, C. & Spies, T. Tumour-derived soluble MIC ligands impair expression of NKG2D and T-cell activation. *Nature* **419,** 734–738 (2002).
181 Lönnroth, C., Andersson, M., Arvidsson, A., Nordgren, S., Brevinge, H., Lagerstedt, K. *et al.* Preoperative treatment with a non-steroidal anti-inflammatory drug (NSAID) increases tumor tissue infiltration of seemingly activated immune cells in colorectal cancer. *Cancer Immun.* **8,** 5 (2008).
182 Champine, P. J., Michaelson, J., Weimer, B. C., Welch, D. R. & DeWald, D. B. Microarray analysis reveals potential mechanisms of BRMS1-mediated metastasis suppression. *Clin. Exp. Metastasis* **24,** 551–565 (2007).
183 Rimsza, L. M., Roberts, R. A., Campo, E., Grogan, T. M., Bea, S., Salaverria, I. *et al.* Loss of major histocompatibility class II expression in non–immune-privileged site diffuse large B-cell lymphoma is highly coordinated and not due to chromosomal deletions. *Blood* **107,** 1101–1107 (2006).
184 Dave, S. S., Fu, K., Wright, G. W., Lam, L. T., Kluin, P., Boerma, E. J. *et al.* Lymphoma/Leukemia Molecular Profiling Project. Molecular diagnosis of Burkitt's lymphoma. *N. Engl. J. Med.* **354,** 2431–2442 (2006).