**[Frontiers in Bioscience 6, d914-926, August 1, 2001]**

# GENETIC ORGANIZATION OF THE HUMAN MHC CLASS III REGION

**Caroline M. Milner[1] and R. Duncan Campbell[2]**

*[1]MRC Immunochemistry Unit, University of Oxford, UK, [2]UK HGMP Resource Centre, Hinxton, Cambridge, UK*

**TABLE OF CONTENTS**



## 1. ABSTRACT

The human major histocompatibility complex (MHC), or human leukocyte antigen (HLA) region, encompasses over 4 Mb of DNA on the short arm of chromosome 6 and is traditionally divided into the class I, class II and class III regions. The MHC has now been entirely sequenced and ~220 genes have been defined of which ~62 are in the class III region. It is becoming clear that many of the latter encode proteins that are likely to be involved in the immune and inflammatory responses. The MHC is known to contribute to a large number of immune-related disorders including insulin dependent diabetes mellitus, rheumatoid arthritis, common variable immunodeficiency and IgA deficiency and there is growing evidence that genes within the class III region are important in determining susceptibility to many of these complex conditions. Genes in the class III region have also been implicated in a number of non-immune-related diseases such as congenital adrenal hyperplasia and sialidosis. Now that the full gene content of the class III region is known the stage is set for the identification and characterisation of candidate disease genes, which will allow greater understanding of the causes of many MHC-linked diseases and thus aid the development of improved treatments**.**

## 2. INTRODUCTION

The human major histocompatibility complex (MHC), or human leukocyte antigen (HLA) region, encompasses over 4 Mb of DNA (~0.1% of the genome) on the chromosome band 6p21.3 (1) and was first recognised due to the involvement of the classical HLA class I and class II proteins in graft rejection and antigen presentation to T cells. Extensive characterisation of this region has recently culminated in the determination of the nucleotide sequence of the entire MHC, confirming the presence of ~220 genes (1) (see http://www.sanger.ac.uk/HGP/ Ch6/ MHC .shtml). The MHC is traditionally divided into the class I, class II and class III regions. The class III region, which has been redefined as the ~730 kb of DNA extending from *NOTCH4* to *BAT1*, is now known to contain at least 62 genes (see Figure 1; reviewed in (2)). As putative functions are ascribed to the products of these genes it is becoming increasingly apparent that many of these are involved in the immune and inflammatory responses.

The class III region is extremely gene dense, with one gene every 10 kb on average and many of the genes are very closely packed. For example, *G11* and *C4* are separated by 611 bp (3, 4) while the 3' untranslated region of *PBX2* overlaps the promoter of *RAGE* (5, 6). The 140 kb segment of DNA containing the duplicated *G11*/*C4*/*Z*/*CYP21*/*X*/*Y* gene cluster, which varies in size from 142 to 214 kb, is probably the most complex in the human genome (7, 8). At least 14 distinct transcripts are encoded in this region, which includes overlapping genes and genes within genes (4, 9). In addition, the class III region contains a high density of retroelements such as human endogenous retrovirus (*HERV*) and *Alu* elements (10).

It is well known that the classical HLA class I and class II genes are highly polymorphic and there is evidence for substantial sequence diversity across the whole MHC. In some cases this is associated with non-coding regions, but within the class III region *C4A* and *C4B* are both highly polymorphic while *C2* and *HSP70-HOM* are just two examples of genes where coding polymorphisms occur to a more limited extent (11, 12, 13). An intriguing feature of the MHC is the occurrence of particular combinations of alleles, at loci across this 4 Mb region, more frequently than would be expected based on the frequencies of individual alleles. This non-random association of alleles gives rise to extended, or ancestral, haplotypes (reviewed in (10)). Explanations for the occurrence of ancestral haplotypes include: models of

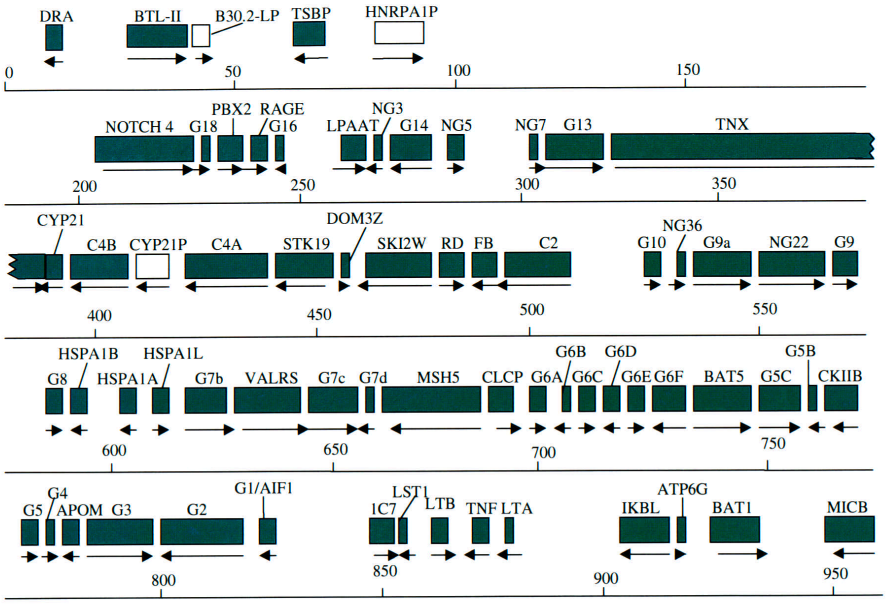


**Figure 1.** A molecular map of the central part of the human MHC. Genes are depicted as boxes with open boxes representing pseudogenes. Arrows indicate the orientation of a gene. The genes *HSPA1A*, *HSPA1B* and *HSPA1L* encode HSP70-1, HSP70-2 and HSP70-HOM, respectively.

selection for apt combinations of alleles at linked loci; population bottle necks, which have not had time to become randomized; and the non-random distribution of recombination sites. Sequence analysis of segments of the MHC from individuals carrying different haplotypes has suggested the presence of blocks of DNA spanning many kilobases that have undergone little or no change due to recombination during human evolution and thus form the basis of ancestral haplotypes (10). Recombination between these 'frozen blocks' has led to the generation of new haplotypes. A second interesting feature of the MHC is the occurrence of extensive duplication, often involving several linked genes or gene fragments. For example, the *G11*/*C4*/*Z*/*CYP21*/*X*/*Y* gene cluster is duplicated to varying degrees on different haplotypes. It has been suggested that these duplication events, together with extensive insertion and deletion, occurred just prior to or at an early stage of human evolution and that this was followed by freezing of the specific blocks of DNA that define human MHC haplotypes (10).

Genetic studies have indicated that genes within the MHC contribute to several hundred diseases, the majority of which are immune-related disorders, e.g. insulin dependent diabetes mellitus (IDDM), rheumatoid arthritis (RA), ankylosing spondylitis (AS), common variable immunodeficiency (CVID) and IgA deficiency (IgAD) (14, 15, 16). Strong associations have been found between many diseases and alleles of classical HLA class II and, to a lesser extent, class I genes. However, in most cases it has not been possible to explain the onset or symptoms of disease in terms of variation in the activity of class I or class II proteins. Furthermore, it is often extended haplotypes, rather than individual alleles, that are associated with disease suggesting that combinations of alleles across the MHC may predispose to, or provide protection against, a disease. Thus, it is very likely that many disease susceptibility genes will be identified within the MHC class III region. For example, a recent study by Schroeder *et al.* (17) has localised the susceptibility locus for IgAD and CVID at the boundary of the class III and class I regions between *G1/AIF1* and *HLA-B*. There is also evidence suggesting that genes in the class III region are involved in susceptibility to AS (18) and IDDM (19). Although most of the MHC-linked diseases are complex, polygenic conditions there are a few instances of diseases that are linked to a single gene in the MHC. For example, congenital adrenal hyperplasia results directly from functional deficiency of the *CYP21B* gene product, due to one of several possible mutations (20), while sialidosis is caused by mutations in the *G9* gene, which encodes a lysosomal sialidase (21). There is evidence that the onset, progression and severity of infectious diseases such as fatal cerebral malaria can also be influenced by genes in the MHC (22). The detailed characterisation of the class III region represents an important step towards determining the contributions of this region to disease.

## 3. CLASS III REGION GENES AND THEIR GENE PRODUCTS

The molecular characterization of the class III region began with the construction of a contig of genomic DNA clones, comprising both cosmids and yeast artificial

chromosomes (23, 24, 25, 26). Physical mapping of the region was achieved by pulsed-field gel electrophoresis, which established the locations of the complement and tumor necrosis factor (*TNF*) genes (27, 28). Subsequent analysis of the cloned genomic DNA using a variety of techniques, such as CpG-island detection, Zoo blot anaysis, Northern blot analysis, cDNA selection, exon trapping and the direct screening of cDNA libraries with cosmid inserts, indicated the presence of at least 50 genes (23, 24, 25, 29, 30, 31). Recently, direct sequence analysis of genomic DNA spanning the entire class III region has been completed (1, 32, 33) (see http://www.sanger.ac.uk /HGP/Ch6/MHC.shtml). This has allowed the precise localizations and genomic structures of the known genes in this region to be determined and, in combination with exon prediction software, has revealed the presence of additional loci bringing the total number of genes in this ~730 kb segment of DNA to ~62 (see Figure 1). The known or putative functions of the products of many of these genes are discussed below.

*NOTCH-4* encodes a member of a family of transmembrane proteins typified by the Notch protein of *Drosophila* (5, 34, 35). Notch and its homologues are widely expressed during embryonic and adult development and participate in cell-cell interactions, mediating a variety of regulatory events such as cell differentiation and development (36). The murine mammary tumour gene, *int-3*, encodes the intracellular domain of Notch4 (37) and there is evidence that Notch4 is important in regulating branching morphogenesis of epithelial cells during the development of the mammary gland (38).

The *PBX2* gene product is a homeodomain-containing protein related to PBX1 (39). The PBX proteins can form heterodimers with homeobox proteins and, thus, may modulate their DNA-binding specificities. Since the homeobox proteins show different expression patterns in various developmental stages and lineages of the hematopoietic system, the PBX proteins might have a role in regulating the expansion of various hematopoietic precursors (40).

The *RAGE* gene encodes a member of the immunoglobulin (Ig) superfamily, which acts as a receptor for advanced glycosylation end products of proteins (RAGE) (5, 41). AGEs accumulate in vascular tissue as a result of ageing, and at an accelerated rate in diabetes, due to the prolonged exposure of proteins to aldoses. The interaction of RAGE with AGEs is involved in mediating effects such as monocyte migration and activation, increased vascular permeability and the activation of NF-κB in endothelial cells. RAGE may contribute to diabetic complications by modulating cellular function following interaction with AGEs and expression of RAGE is upregulated in the vasculature in a range of vasculopathies, in artherosclerotic vascular lesions and in immune inflammation vasculitides (42, 43).

The *hLPAATa* gene encodes a lysophosphatidic acid acetyltransferase (LPAAT) that shares 48% sequence identity with hLPAATβ (which is encoded on chromosome 9) and is localised in the ER (44). Recombinant hLPAATα has been shown to specifically convert LPA to PA, which is the first step in phospholipid synthesis, and to have an affinity for fatty acids with acyl chains of 12-18 carbons in length (44). Of particular interest is the observation that hLPAATα is involved in generating the phospholipid precursor of arachidonic acid, which is essential for the production of most prostaglandins. The latter play important roles in the inflammatory response, the production of pain and fever, the regulation of blood pressure, induction of blood clotting, control of several reproductive functions and regulation of the sleep/wake cycle (45). Although the best known function of phospholipids is the formation of biological membranes, some phospholipids and their metabolites have been implicated in signalling pathways (46). For example, IL-1β and TNFα may activate and signal, at *least* in part, through a common lipid intracellular signalling pathway that requires the activation of LPAAT (47, 48). Furthermore, inhibitors of LPAAT have been shown to block the inflammatory responses mediated by an increase in PA and there is evidence to suggest that LPAAT is activated by IL-1β (47).

The *G14* gene encodes a secreted protein with ~20% identity to human palmitoyl-protein thioesterase 1 (PPT1) (49, 50). This protein was termed PPT2 by Soyombo and Hofmann (49) on the basis that it has palmitoyl-CoA hydrolase activity, although with a substrate specificity different from that of PPT1. Protein modification by the addition of palmitate, or other fatty acids, via labile thioester bonds is a reversible event, which may be important in the regulation of membrane attachment (51). PPT1 is known to be important in the catabolism of lipid-modified proteins by hydrolysing the thioester bonds that link fatty acids to cysteine residues in S-fatty acetylated proteins. Mutations in PPT1 result in the childhood neurogenerative disorder, infantile neuronal ceroid lipofascionosis (INCL), which is characterised by early vision loss and mental deterioration leading to a vegetative state by three years of age (52). This has been ascribed to the failure of PPT1 to localise to the lysosomes resulting in the accumulation of lipid thioesters. The addition of recombinant PPT1 to INCL patient cell lines can reverse this accumulation. The failure of PPT2 to exert a similar effect indicates that these two enzymes have independent functions (49). The structure of PPT2 is distinct from that of PPT1, comprising an N-terminal lipase domain and a C-terminal domain common to the cytokine receptor superfamily. A comparison of the activities of the *G14* protein product and recombinant bovine PPT1 has revealed that, while both enzymes preferentially hydrolyse long acyl chains (>12-14 carbons) the former has a broader range of activity and would be more appropriately termed an S-thioesterase (50). *G14* has been shown to be expressed as a variety of alternatively spliced transcripts and there is evidence, from a monocyte-like cell line, for a secreted form of the protein (50, 53). A number of polymorphisms, two of which result in amino acid changes, have been identified in *G14*. However, none of these show an

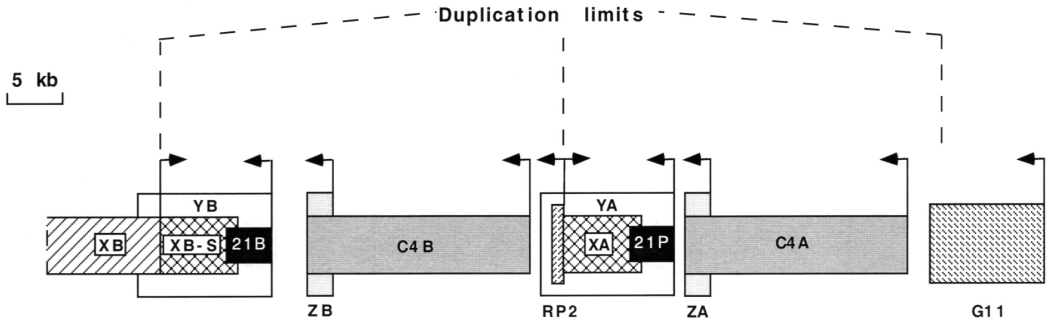


**Figure 2.** A map of the duplicated *G11/C4/Z/CYP21/Y/X* gene cluster. Genes and gene fragments are depicted as boxes, with duplicated loci having the same pattern of shading. Only the 3' end of the ~65 kb XB gene is shown, as indicated by the open-ended box. The 5' to 3' orientation of each locus is indicated by an arrow and the duplication limits are marked by dotted lines.

association with NCL disorders (53). The *G14* gene contains an incomplete CTG repeat, encoding a stretch of Leu residues in the signal peptide of the protein. The expansion of trinucleotide repeats is an important feature in various human genetic disorders, including neurogenic conditions.

The *G13* (*Creb-rp*) protein product contains a leucine zipper consensus sequence preceded by two clusters of basic residues that are separated by an alanine spacer, i.e. a bZip domain (54, 55). This domain is conserved in many transcription factors including the cyclic AMP response element binding protein (CREB) family, which comprises proteins that show sequence-specific binding to DNA (via a basic region) and use dimerization (via a leucine zipper domain) to control activity.

Three components of the complement system, which is the principal effector mechanism of humoral immunity and is important in the clearance of immune complexes, opsonization and cell lysis, are encoded in the class III region (reviewed in (56, 57)). C2 and C4 participate in the classical pathway, which is activated by the binding of C1 to antigen-antibody complexes or following the interaction of mannan binding protein with carbohydrate ligands. Factor B (Bf) is a component of the alternative pathway, which is activated by a diverse set of substances including components of yeast and bacterial cell walls. The *C2* and *Bf* genes are separated by only 421 bp (58) and the fact that they have similar organization and encode functionally related proteins suggests that they arose by duplication of an ancestral locus. This is also thought to be the case for *C4A* and *C4B*, which encode proteins that differ by only four amino acids, but nevertheless have profoundly different covalent binding activities (59, 60, 61). *C4B* can occur as both long and short forms, due to the presence or absence of a *HERV-K* element in intron 9 (62, 63). Both isotypes of C4 are highly polymorphic, whilst C2 and Bf show more limited polymorphism (11, 12).

Most human chromosomes carry two *C4* genes, *C4A* and *C4B*, which form part of a duplicated segment of DNA spanning ~75 kb (64). However, the number of expressed *C4* genes can range from none to four. Complete deficiency of C4, which correlates with severe immune complex diseases, is rare while partial deficiency of C4 is more common and is thought to be associated with an increased susceptibility to systemic lupus erythematosus (SLE), scleroderma and primary biliary cirrhosis (65). Duplications or deletions of *C4* always include the neighbouring *CYP21*, *X*, and *G11* (or *STK19*) loci (see Figure 2). It has been suggested that recombination events involving chromosomes that carry non-functional copies of these genes (i.e. *CYP21P*, *XA* and *RP2*) may result in disease-associated mutations in this segment of the class III region (8, 66).

The *CYP21B* gene and the related pseudogene *CYP21P* lie ~2.5 kb from the 3' ends of *C4B* and *C4A*, respectively, and share 97% overall nucleotide sequence identity (67, 68). *CYP21B* encodes the adrenal enzyme steroid 21-hydroxylase, which is involved in mineralocorticoid and glucocorticoid biosynthesis. *CYP21P* contains deleterious mutations and so cannot encode functional protein. Deficiency of CYP21B is the most frequent cause (95%) of CAH, which is characterised by accumulation of 17-hydroxyprogesterone and production of excess androgen (20). Classical CAH occurs as the milder simple virilising form and the more severe, and life-threatening, salt wasting form. Individuals with the nonclassical late-onset form of CAH remain asymptomatic or develop the symptom of an excess of androgens in childhood or at puberty.

The *TN-X* (*XB*) and *XA* genes overlap the 3' ends of the *CYP21B* and *CYP21P* genes by 481 bp, respectively and are transcribed in the opposite orientation to the *C4* and *CYP21* genes (7, 69) (see Figure 2). Tenascin-X (TN-X) is a member of a family of proteins with structures similar to tenascin (TN-C). The functions of these extracellular matrix proteins are unclear, but TN-C can facilitate cell movement by counteracting the cell adhesion and spreading activity of fibronectin *in vitro* and may also be important in embryogenesis. TN-X has a tissue distribution distinct from that of TN-C, being predominantly expressed in heart and skeletal muscle (70). One of the boundaries of the duplication of the *C4/CYP21/X* region lies within exon 26 of *XB*, such that *XA* lacks the first 26 exons of *XB* (7). *XA*

is expressed as an adrenal-specific transcript, but a premature stop codon suggests that it is probably not expressed at the protein level. The *XB-S* gene, which shares 99% nucleotide sequence identity with *XA*, is transcribed from a promoter within exon 26 of *XB* and is expressed as an adrenal-specific truncated variant of TN-X (71). The *YB* and *YA* genes utilise the *CYP21B/P* promoters and are transcribed in the same orientation as the latter, but use a different array of exons and introns (72). Due to the presence of a duplication boundary within the *Y* genes, they each have unique 3' ends and both *YA* and *YB* are expressed as two alternatively spliced adrenal-specific transcripts (72). Transcripts termed *ZA* and *ZB* have recently been shown to arise from promoters in intron 35 of *C4A* and *C4B*, respectively, and are expressed exclusively in the adrenal gland (9). No protein products have yet been identified for the Y and Z transcripts and their significance remains to be determined.

Genomic sequence analysis of *G11* (or *STK19*) has shown that 913 bp of its 3' end was involved in the duplication event giving rise to the two *C4/Z/CYP21/Y/X* gene clusters (3, 4). *G11* lies ~600 bp upstream of *C4A* and the duplicated fragment, comprising the last 2 and a half exons of *G11*, lies the same distance from *C4B*, i.e. overlapping the 3' end of *YA*. Two alternative products of the G11 gene have been described: G11-Z and G11-Y, which lacks the first 110 amino acids of G11-Z (73). In addition, the use of two differential splice sites at the end of exon 5 results in the presence or absence of four amino acids (Val-Cys-Asp-Cys) in both G11-Y and G11-Z. Recombinant G11-Z protein has been shown to bind ATP and to have $Mn^{2+}$-dependent protein kinase activity that phosphorylates Ser and Thr residues (73). Similar activity was observed for G11-Y, indicating that the catalytic domain of G11 lies between amino acids 111 and 368. Interestingly, the presence or absence of the Val-Cys-Asp-Cys motif within this domain had no obvious effect on the activity of G11-Y. Protein kinases and phosphatases play vital roles in the activation and inactivation of cellular processes that regulate transcription, resulting in specific changes in gene expression. Restricted intracellular distribution may be important in determining the substrate specificity of kinases and the localization of G11-Z to the nuclei of transfected COS7 cells suggests that this protein may be involved in the regulation of transcription in the nucleus (73).

*DOM3Z* encodes a protein product that contains a leucine zipper motif (74). Related proteins (sharing 20-30% sequence identity) have been identified in complex and simple eukaryotes and in a flowering plant. Of these, the *C. elegans* protein, DOM3, has a putative role in the postembryonic development of germ cells.

*SKI2W* encodes a protein that contains an RNA helicase domain and two leucine zipper motifs and shows extensive similarity to the yeast antiviral protein Ski2p (31, 75). Helicases play essential roles in unwinding DNA during replication, repair, recombination and transcription, and RNA during splicing and translation. Ski2p exerts its antiviral action by translational inhibition of viral RNA that lacks a poly(A) tail. Ski2w has been shown to hydrolyse ATP, an important energy source for helicases, and has been localised to the nucleolus and cytoplasm where it is thought to be associated with ribosomes and polyribosomes (75, 76).

G9a contains six contiguous copies of the 33 amino acid ANK repeat (77). This repeat is thought to mediate protein-protein interactions and is found in proteins of diverse function, including *Drosophila* Notch and NF-κB. The C-terminal region of G9a corresponds to a SET domain (~150 amino acids), which is also found at the C-termini of the *Drosophila* trithorax protein and its human homologue ALL-1 (or HRX). The SET domain is important in the action of these proteins as regulators of transcription during development (78).

The *G9* gene encodes a protein with significant sequence identity to bacterial and mammalian sialidases (21, 79, 80). Biochemical analysis and nucleotide sequencing have confirmed the existence of three unique forms of sialidase in mammals, located in the lysosomes, cytosol and plasma. Recombinant G9 protein has sialidase activity with an acidic pH optimum and is localised in lysosomes. The maximal activity of this sialidase and its transport from the ER to lysosomes depends on its association with the lysosomal protein, protective protein/cathepsin A (PPCA) (81). Sialidases are important in modulating many cellular functions by regulating the sialic acid content of glycoproteins and glycolipids. Deficiency in lysosomal sialidase activity (<10% of normal activity), due to a variety of point mutations in the sialidase itself or in PPCA, results in various forms of sialidosis or in galactosialidosis. These lysosomal storage disorders are characterised by developmental and neurological abnormalities (21, 80, 82, 83). Three allelic variants of the murine equivalent of G9 (*Neu1*) have been described, of which $Neu1^a$ confers only ~17% of normal sialidase activity (84). Mice that carry this allele (SM/J and related strains) are small at birth, but do not exhibit the severe symptoms of human sialidosis patients. Comparison of mice expressing the different Neu1 allotypes has implicated the MHC-linked sialidase in several aspects of the immune response. For example, Neu1 has been shown to be required for the regulation of IL4 production during T cell activation, which has important implications for the initiation of Th2- versus Th1-type responses (85).

The *HSP70-1* and *-2* genes (*HSPA1A* and *HSPA1B*) both encode the major heat-inducible 70 kDa heat shock protein (HSP70), while *HSP70-HOM* (*HSPA1L*) encodes a non-heat-inducible protein that is 90% identical to HSP70-1/-2 (86). All hsp70s can bind unfolded proteins and peptides and act as chaperones in the synthesis, folding, assembly, translocation and degradation of proteins during normal cellular processes and following stress. It has been suggested that the cytosolic hsp70s (including HSP70 and HSC70) and the ER-resident chaperone gp96 form a relay line, chaperoning antigenic peptides generated in the cytosol until they bind MHC class I molecules in the ER (87). In support of this, hsp preparations from tumor cells, virally infected cells and normal cells and tissues have been

shown to contain hsp-peptide complexes. Furthermore, HSP70- and gp96-peptide complexes purified from murine methylcholanthrene-induced sarcomas have been found to give rise to tumour-specific immunity mediated by T cells when injected into syngeneic mice, indicating that an immune response is being initiated by tumor-specific peptides carried by hsps (87). In the case of gp96 this has been shown to arise following uptake of the hsp-peptide complexes by a subset of macrophages, which prime antigen-specific cytotoxic T cells (88). Hsp70s have been implicated in many diseases. For example, increased expression of hsp70s has been observed in autoimmune diseases such as scleroderma and SLE, although this may be a consequence of disease rather than a contributory factor (89, 90). However, exposure of keratinocytes to UV light results in accumulation of HSP70 in the nucleoli and a concomitant increase in binding sites for autoantibodies to nuclear antigens which are a feature of erythematosis lesions (91). Furthermore, cytosolic members of the hsp70 family have been detected on the surface of some tumour cells, where they mediate a strong anti-tumour response by rendering cells sensitive to NK cell-mediated lysis (92).

*Hsp70* genes in the freshwater mould, *Achlya klebsiana*, and in *Drosophila auraria* have been shown to form long antiparallel coupled open reading frames (LAC-ORFs), or antisense gene pairs, with NAD-specific glutamate dehydrogenase (*NAD-GDH*)-like genes (93, 94). In *A. klebsiana* the NAD-GDH-like gene is comprised of ten exons and exon 10 is in perfect register, along the complementary strand, with an *hsp70* gene (94). Both genes are inducible by L-glutamine and by abundant organic nitrogen. The *hsp70* gene is also induced by starvation, but not by heat shock, indicating overlapping, but not identical, control of the two genes. The fact that the hydropathy plots for the hsp70 and NAD-GDH-like proteins are almost perfect mirror images suggests that they may interact and it is possible that the hsp70 may assist the folding of the MAD-GDH-like protein. *Hsp70* genes in a number of species have been shown to exist as LAC-ORFs with genes encoding proteins with significant sequence identities to the NAD-GDH-like proteins encoded in *A. klebsiana* and *D. auraria*, suggesting possible evolutionary benefits (94). For example, HSP70-1 is 61% identical to the *D. auraria* hsp70 and the *HSP70-1* gene forms an antisense pair with a gene encoding a protein with 36% identical to the NAD-GDH-like protein of *D. auraria*.

The *VALRS* gene encodes a protein that shares 48.3% identity with the valyl-tRNA synthetase ($Trs^{Val}$) of *Saccharomyces cerevisiae* and functional studies on recombinant G7a have confirmed that it is the human $Trs^{Val}$ (95, 96). G7a has a unique N-terminal region, which may be involved in the interaction of human $Trs^{Val}$ with elongation factor 1H.

*MSH5* encodes a homologue of the bacterial mismatch repair protein, MutS (97, 98, 99). Eukaryotic homologues of MutS fall into two categories: those involved in mismatch repair and those involved in meiotic recombination processes. They typically function as heterodimers and MSH5 has been shown to interact specifically with MSH4. In *S. cerevisiae*, MSH5 is required for meiotic segregation fidelity and crossing-over and may also be important in the response of mitotic cells to DNA-damaging agents (100, 101). Human *MSH5* mRNA is expressed at particularly high levels in testis and protein expression occurs during a defined phase of spermatogenesis suggesting a role in germ cell development. *MSH5* mRNA is also abundant in bone marrow and lymph nodes, so the protein may also be important in the assembly of Ig and/or T cell receptor genes during lymphoid development.

*G6* encodes a regulatory nuclear chloride ion channel protein (hRNCC) (33) that corresponds to the NCCP27 protein described by Valenzuela *et al.* (102). This protein was shown to be associated with the nuclear membrane, but lacks a transmembrane region suggesting that it may be a regulatory subunit of a multi-protein complex. This is supported by the presence of multiple PEST sequences in hRNCC, which mediate rapid degradation of the protein in response to appropriate stimuli. Two CA repeats, one in intron 2 of *G6* and the second in intron 2 of *G6D*, achieved the highest LOD scores of all the markers used in a recent genome-wide screen for susceptibility genes to ankylosing spondylitis (AS) (18).

The *G6A* gene encodes a putative homologue of the enzyme $N^{\omega}$-$N^{\omega}$-dimethylarginine dimethylaminohyrolase (DDAH) (33), which was first described in rat kidney and regulates the production of nitric oxide (NO) by metabolising L-arginine analogues that inhibit NO synthesis (103). G6A is probably not the human equivalent of DDAH, but is likely to have a related function. Interestingly, the L-arginine analogue, $N^{\omega}$-mono-methyl-L-arginine has a variety of anti-inflammatory effects while NO production contributes to the development of IDDM (see references in (33)).

*G6B* encodes a novel member of the Ig superfamily with a single Ig V-like domain (33). Members of this superfamily function mainly in the immune system, cell-cell recognition or structural organisation and regulation of muscle (104). The presence of cysteine residues in G6B in addition to those involved in the conserved disulphide bond of the Ig V domain suggests that it may form hetero- or homo-dimers via disulphide bonding. The cytoplasmic tail of G6B contains a proline-rich domain, characteristic of src homology 3 (SH3) domain-binding proteins, and two putative phosphorylated tyrosine residues. These features are typical of a signal transduction receptor.

*G6C* and *G6D* encode small cysteine-rich proteins that share ~20-27% identity with members of the leukocyte antigen 6 (Ly-6) superfamily (33). Both proteins contain a leader peptide, a Ly-6-like domain, a transmembrane domain and sequences characteristic of GPI-anchored proteins. The Ly-6 proteins are differentially expressed in several hemopoietic lineages, especially T cells, during hemopoiesis and immune responses. They are

thought to be involved in signal transduction and cell activation and may be important in leukocyte development (105, 106). The murine *Ly-6* genes all map to a 630 kb region of chromosome 15 and three *Ly-6* genes have been mapped to human chromosome 8, which is syntenic with murine chromosome 15 (107). G6C and G6D are not direct orthologues of any of the known murine Ly-6 proteins, although they may have related functions. Furthermore, G6C and G6D share only 20% amino acid sequence identity and have distinct expression patterns, with G6C being expressed only in T cells, suggesting that these proteins have distinct roles.

*G6E* is also predicted to encode a Ly-6 family member on the basis of an incomplete sequence. The first two predicted exons encode a leader sequence and the first 24 residues of a Ly-6 domain (33). However, a stop codon at the end of exon 2 and the failure to detect *G6E* transcripts in any of the cell lines tested suggest that this may be a psuedogene.

The *G5a* gene encodes the casein kinase II (ckII) ß subunit (31). CkII is a highly conserved serine protein kinase, consisting of a heterotetramer of α and α' catalytic subunits and two regulatory ß subunits. The ß subunit is required for high activity phosphorylation of the α subunit (108). Many of the substrates of ckII, such as insulin, epidermal growth factor and a variety of cellular and viral oncoproteins, play key roles in cell growth and metabolism (109). CkII activity appears to be upregulated in the neurofibrillary tangles in the brains of Alzheimer's Disease (AD) patients (110).

There is evidence that three alternative transcripts are derived from the differential splicing of nine exons in the *G1*/allograft inflammatory factor 1 (*AIF1*) genomic region (32). *AIF1* (111) is derived from exons 1-6, *G1* (112) contains an alternatively spliced version of exon 4 together with exons 5 and 6, and IFN-γ-responsive transcript 1 (*IRT-1*) (EMBL database accession no. U95213) is derived from exons 4, 5 and 6 together with the intervening intronic sequence. AIF1 is an IFN-γ-inducible molecule expressed in cells of the monocyte lineage (32, 111). It has been associated with the processes involved in chronic allograft rejection and has also been implicated in the development of IDDM (111, 113). G1, AIF1 and IRT-1 all contain a cluster of paired basic residues, which is a feature of hormone precursor proteins. AIF1 also contains a putative EF hand calcium-binding domain. This may not be functional in G1 and IRT-1 due to the truncation of the conserved α-helix preceding the domain in both proteins and the presence of a 66 residue insertion within the conserved loop of the EF hand domain in IRT-1.

*1C7* encodes a putative novel member of the Ig superfamily and differential splicing of this gene has been shown to result in nine alternative transcripts of which six encode putative protein isoforms, 1C7a-f (32). 1C7a, b and c contain sequence encoded by a single exon 2 which corresponds to an Ig V-like domain (104). In 1C7d, e and f alternative splicing results in two exons from the exon 2 region leading to truncation of the Ig domain to form an Ig C2-like domain (114). If these 1C7 isoforms are expressed this will represent the first known example of Ig set switching within a protein. 1C7 is predicted to be a type I integral membrane protein and the absence of cysteines, other than those involved in the conserved disulphide bond of its Ig domain, suggests that it exists as a monomer. The cytoplasmic tails of 1C7a and e contain a number of potential SH3 domain binding motifs and a potential protein kinase C phosphorylation motif. Recently, NKp30, a novel triggering receptor selectively expressed by resting and activated natural killer cells and involved in the recognition and killing of target cells, was shown to correspond to the 1C7c isoform (115).

Leukocyte specific transcript (*LST-1*) is the human equivalent of the murine *B144* gene (116). It is expressed predominantly in monocytes, where it is inducible by IFN-γ, and also at lower levels in some T cells. This has led to the suggestion of a role for LST-1 in the immune response. There is evidence for multiple protein isoforms of LST-1, due to a complex pattern of alternative splicing involving at least nine exons (117, 118). Furthermore, it appears that both membrane-bound and soluble forms of LST-1 are expressed, depending on the usage of two possible open reading frames. Treatment of monocytes with IFN-γ has been shown to result in a shift from expression of both soluble and membrane-bound LST-1 to expression of the soluble from alone (117).

The cytokines TNF, LTα and LTβ belong to the TNF ligand superfamily (119). TNF occurs as membrane bound and secreted homotrimers, both of which are biologically active. LTα also forms secreted homotrimers, but can be retained at the cell surface as heterotrimers with LTβ. TNF is produced by a variety of cells and exhibits numerous inflammatory and immunomodulatory activities as well as being involved in tumor cachexia. LTα is expressed exclusively by lymphocytes. It shows similar activity *in vitro*, but is often less potent than TNF. Recent studies suggest a unique role for secreted LTα in the generation of germinal centres, and membrane-bound LTα has been implicated in the development of the spleen, lymph nodes and Peyer's patches (120, 121, 122). Polymorphisms in the promoter of the TNF gene are thought to play a role in increased susceptibility to fatal cerebral malaria (22).

The IKBL protein contains two copies of the 33 amino acid ANK repeat followed by a third poorly conserved ANK repeat and a short region rich in acidic amino acids, which is typical of the IκB family of transcription factor inhibitors (123). The Rel family of transcription factors, which includes NF-κB, regulates the expression of a wide variety of genes including cytokines, cytokine receptors and stress proteins. In unstimulated cells these transcription factors are maintained in the cytosol due to their association with inhibitors belonging to the IκB family (124). Stimulation of cells by agents such as PMA, IL-1β, TNF and LPS results in the phosphorylation and subsequent degradation of the IκB protein, allowing translocation of the free Rel proteins to the nucleus.

The *ATP6G* gene encodes a homologue of vacuolar ATPase G (V-ATPase G) subunits (32). These $H^+$-translocating ATPases are involved in a broad range of functions, including bone reabsorption, processing of endocytosed receptor-ligand complexes and degradation of cell debris in lysosomes, and may be involved in the immune and inflammatory responses (125). The G subunit is one of two known peripheral components of the $V_1$ catalytic ATPase complex, which are thought to act independently to perform the same or similar functions in different cell types. ATP6G appears to be particularly abundant in monocytes and macrophages (32). On the basis of sequence conservation between species the first 50 amino acids of ATP6G are likely to constitute an important functional domain. A truncated form of ATP6G, which results from alternative splicing and lacks the first 41 amino acids, is therefore unlikely to be functional.

*BAT1* encodes a nuclear protein with significant sequence similarity to the DEAD-box protein family of ATP-dependent RNA helicases (126). These proteins participate in diverse cellular functions such as initiation of translation, RNA splicing, ribosome assembly, spermatogenesis, oogenesis and cell growth and division (127). It is possible that BAT1 may be involved in the complex splicing that has been shown to be a feature of a number of immune-related genes in the MHC, e.g. *LST-1*, *1C7* and *AIF1*.

## 4. CONCLUSION

The genes of the MHC class III region encode proteins of diverse function. However, there is growing evidence that many of these are involved in the immune and inflammatory responses and that genes within the class III region are involved in susceptibility to a number of diseases. Although the gene content of the class III region has now been completely defined it is clear that many of the genes, for example *LST-1*, *1C7* and *AIF1*, are subject to alternative splicing such that the number of transcripts encoded within this regions is far greater than might be expected.

The rapid expansion of the protein and DNA sequence databases over recent years has allowed putative functions to be ascribed to the products of many of the newly identified genes in the class III region and these have been confirmed experimentally in a number of cases. This knowledge, together with the use of the many known polymorphic markers (e.g. dinucleotide repeats and single nucleotide polymorphisms) across the class III region in patient studies, will greatly facilitate the identification of candidate disease genes. As a result of recent improvements in technology the detection of promoter and/or coding polymorphisms in candidate genes and the association of these with diseases are relatively straightforward. However, establishing the effects of polymorphisms on gene expression or protein function *in vitro* and in animal models still represents a major challenge for the future

## 5. ACKNOWLEDGEMENTS

We wish to acknowledge the support of the Medical Research Council and to thank Alison Marsland for secretarial assistance. CMM holds an Arthritis Research Campaign Postdoctoral Fellowship.

## 6. REFERENCES

1. The MHC Sequencing Consortium: Complete sequence and gene map of a human major histocompatibility complex. *Nature* 401, 921-923 (1999)

2. Aguado B, C. M. Milner, & R. D. Campbell: Genes of the MHC class III region and the functions of the proteins they encode. In: HLA and MHC: genes, molecules and function. Eds: Browning M, McMichael A, BIOS Scientific Publishers Ltd, Oxford, 39-75 (1996)

3. Sargent C. A, M. J. Anderson, A.-L. Hsieh, E. Kendall, N. Gomez-Escobar & R. D. Campbell: Characterisation of the novel gene G11 lying adjacent to the complement C4A gene in the human major histocompatibility complex. *Hum Mol Genet* 3, 481-488 (1994)

4. Shen L, L.-C. Wu, S. Sanlioglu, R. Chen, A. R. Mendoza, A. W. Dangel, M. C. Carroll, W. B. Zipf & C. Y. Yu: Structure and genetics of the partially duplicated gene RP located immediately upstream of the complement C4A and the C4B genes in the HLA class III region: molecular cloning, exon intron structure, composite retroposon, and breakpoint of gene duplication. *J Biol Chem* 269, 8466-8476 (1994)

5. Sugaya K, T. Fukagawa, K. Matsumoto, K. Mita, E. Takahashi, A. Ando, H. Inoko & T. Ikemura: Three genes in the human MHC class III region near the junction with the class II: gene for receptor of advanced glycosylation end products, PBX2 homeobox gene and a notch homolog, human counterpart of mouse mammary tumor gene int-3. *Genomics* 23, 408-419 (1994)

6. Li J. & A. M. Schmidt: Characterization and functional analysis of the promoter of RAGE, the receptor for advanced glycation end products. *J Biol Chem* 272, 16498-16506 (1997)

7. Gitelman S. E, J. Bristow & W. L. Miller: Mechanism and consequences of the duplication of the human C4/P450c21/gene X locus. *Mol Cell Biol* 12, 2124-2134 (1992)

8. Yu C. Y.: Molecular genetics of the human MHC complement gene cluster. *Exp Clin Immunogenet* 15, 213-230 (1998)

9. Tee M. K, G. O. Babalola, P. Aza-Blanc, M. Speek, S. E. Gitelman & W. L. Miller: A promoter within intron 35 of the human C4A gene initiates abundant adrenal-specific transcription of a 1 kb RNA: location of a cryptic CYP21

promoter element? *Human Molec Genet* 4, 2109-2116 (1995)

10. Dawkins R, C. Leelayuwat, S. Gaudieri, G. Tay, J. Hui, S. Cattley, P. Martinez & J. Kulski: Genomics of the major histocompatibility complex: haplotypes, duplication, retroviruses and disease. *Immunol Rev* 167, 275-304 (1999)

11. Campbell R. D, I. Dunham, E. Kendall & C. A. Sargent: Polymorphism of the human complement component C4. *Exp Clin Immunogenet* 7, 69-84 (1990)

12. Cross S. J. & W. Thomson: DNA polymorphism of the C2 and factor B gene. *Exp Clin Immunogenet* 7, 53-63 (1990)

13. Milner C. M. & R. D. Campbell: Polymorphic analysis of the three MHC-linked HSP70 genes. *Immunogenetics* 36, 357-362 (1992)

14. JL Tiwari & PI Terasaki: HLA and Disease Associations. Springer-Verlag, New York (1985)

15. Davies J. L, Y. Kawaguchi, S. T. Bennett, J. B. Copeman, H. J. Cordell, L. E. Pritchard, P. W. Reed, S. C. Gough, S. C. Jenkins, S. M. Palmer & .: A genome-wide search for human type 1 diabetes susceptibility genes. *Nature* 371, 130-136 (1994)

16. R Lechler: HLA and Disease. London: Academic Press (1994)

17. Schroeder H. W, Z. B. Zhu, R. E. March, R. D. Campbell, S. M. Berney, S. A. Nedospasov, R. L. Turetskaya, T. P. Atkinson, R. C. Go, M. D. Cooper & J. E. Volanakis: Susceptibility locus for IgA deficiency and common variable immunodeficiency in the HLA-DR3, -B8, -A1 haplotypes. *Mol Med* 4, 72-86 (1998)

18. Brown M. A, K. D. Pile, L. G. Kennedy, D. Campbell, L. Andrew, R. March, J. L. Shatford, D. E. Weeks, A. Calin & B. P. Wordsworth: A genome-wide screen for susceptibility loci in ankylosing spondylitis. *Arthritis Rheum* 41, 588-595 (1998)

19. Lie B. A, J. A. Todd, F. Pociot, J. Nerup, H. E. Akselsen, G. Joner, K. Dahl-Jorgensen, K. S. Ronningen, E. Thorsby & D. E. Undlien: The predisposition to type 1 diabetes linked to the human leukocyte antigen complex includes at least one non-class II gene. *Am J Hum Genet* 64, 793-800 (1999)

20. Morel Y. & W. L. Miller: Clinical and molecular genetics of congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *Adv Hum Genet* 20, 1-68 (1991)

21. Bonten E, S. A. van der, M. Fornerod, G. Grosveld & A. d'Azzo: Characterization of human lysosomal neuraminidase defines the molecular basis of the metabolic storage disorder sialidosis. *Genes Dev* 10, 3156-3169 (1996)

22. McGuire W, A. V. Hill, C. E. Allsopp, B. M. Greenwood & D. Kwiatkowski: Variation in the TNF-alpha promoter region associated with susceptibility to cerebral malaria. *Nature* 371, 508-510 (1994)

23. Sargent C. A, I. Dunham & R. D. Campbell: Identification of multiple HTF-island associated genes in the human major histocompatibility complex class III region. *EMBO J* 8, 2305-2312 (1989)

24. Spies T, M. Bresnahan & J. L. Strominger: Human major histocompatibility complex contains a minimum of 19 genes between the complement cluster and HLA-B. *Proc Natl Acad Sci U S A* 86, 8955-8958 (1989)

25. Kendall E, C. A. Sargent & R. D. Campbell: Human major histocompatibility complex contains a new cluster of genes between the HLA-D and complement C4 loci. *Nucleic Acids Res* 18, 7251-7257 (1990)

26. Abderrahim H, J. L. Sambucy, F. Iris, P. Ougen, A. Billault, I. M. Chumakov, J. Dausset, D. Cohen & D. Le Paslier: Cloning the human major histocompatibility complex in YACs. *Genomics* 23, 520-527 (1994)

27. Carroll M. C, P. Katzman, E. M. Alicot, B. H. Koller, D. E. Geraghty, H. T. Orr, J. L. Strominger & T. Spies: Linkage map of the human major histocompatibility complex including the tumor necrosis factor genes. *Proc Natl Acad Sci U S A* 84, 8535-8539 (1987)

28. Dunham I, C. A. Sargent, J. Trowsdale & R. D. Campbell: Molecular mapping of the human major histocompatibility complex by pulsed-field gel electrophoresis. *Proc Natl Acad Sci U S A* 84, 7237-7241 (1987)

29. Fan W. F, X. Wei, H. Shukla, S. Parimoo, H. Xu, P. Sankhavaram, Z. Li & S. M. Weissman: Application of cDNA selection techniques to regions of the human MHC. *Genomics* 17, 575-581 (1993)

30. Burfoot M. S. & R. D. Campbell: Improved method of gene detection using exon amplification. *Nucleic Acids Res* 22, 5510-5511 (1994)

31. Albertella M. R, H. Jones, W. Thomson, M. G. Olavesen & R. D. Campbell: Localization of eight additional genes in the human major histocompatibility complex, including the gene encoding the casein kinase II beta subunit (CSNK2B). *Genomics* 36, 240-251 (1996)

32. Neville M. J. & R. D. Campbell: A new member of the Ig superfamily and a V-ATPase G subunit are among the predicted products of novel genes close to the TNF locus in the human MHC. *J Immunol* 162, 4745-4754 (1999)

33. Ribas G, M. Neville, J. L. Wixon, J. Cheng & R. D. Campbell: Genes encoding three new members of the leukocyte antigen 6 superfamily and a novel member of Ig superfamily, together with genes encoding the regulatory nuclear chloride ion channel protein (hRNCC) and an N omega-N omega-dimethylarginine dimethylaminohydrolase

homologue, are found in a 30-kb segment of the MHC class III region. *J Immunol* 163, 278-287 (1999)

34. Sugaya K, S. Sasanuma, J. Nohata, T. Kimura, T. Fukagawa, Y. Nakamura, A. Ando, H. Inoko, T. Ikemura & K. Mita: Gene organization of human NOTCH4 and (CTG)n polymorphism in this human counterpart gene of mouse proto-oncogene Int3. *Gene* 189, 235-244 (1997)

35. Li L, G. M. Huang, A. B. Banta, Y. Deng, T. Smith, P. Dong, C. Friedman, L. Chen, B. J. Trask, T. Spies, L. Rowen & L. Hood: Cloning, characterization, and the complete 56.8-kilobase DNA sequence of the human NOTCH4 gene. *Genomics* 51, 45-58 (1998)

36. Fortini M. E. & S. Artavanis-Tsakonas: Notch: neurogenesis is only part of the picture. *Cell* 75, 1245-1247 (1993)

37. Robbins J, B. J. Blondel, D. Gallahan & R. Callahan: Mouse mammary tumor gene int-3: a member of the notch gene family transforms mammary epithelial cells. *J Virol* 66, 2594-2599 (1992)

38. Uyttendaele H, J. V. Soriano, R. Montesano & J. Kitajewski: Notch4 and Wnt-1 proteins function to regulate branching morphogenesis of mammary epithelial cells in an opposing fashion. *Dev Biol* 196, 204-217 (1998)

39. Aguado B. & R. D. Campbell: The novel gene G17, located in the human major histocompatibility complex, encodes PBX2, a homeodomain-containing protein. *Genomics* 25, 650-659 (1995)

40. Lawrence H. J, G. Sauvageau, R. K. Humphries & C. Largman: The role of HOX homeobox genes in normal and leukemic hematopoiesis. *Stem Cells* 14, 281-291 (1996)

41. Neeper M, A. M. Schmidt, J. Brett, S. D. Yan, F. Wang, Y. C. Pan, K. Elliston, D. Stern & A. Shaw: Cloning and expression of a cell surface receptor for advanced glycosylation end products of proteins. *J Biol Chem* 267, 14998-15004 (1992)

42. Schmidt A. M, S. D. Yan & D. M. Stern: The dark side of glucose. *Nat Med* 1, 1002-1004 (1995)

43. Schmidt A. M, S. D. Yan, J. L. Wautier & D. Stern: Activation of receptor for advanced glycation end products: a mechanism for chronic vascular dysfunction in diabetic vasculopathy and atherosclerosis. *Circ Res* 84, 489-497 (1999)

44. Aguado B. & R. D. Campbell: Characterization of a human lysophosphatidic acid acyltransferase that is encoded by a gene located in the class III region of the human major histocompatibility complex. *J Biol Chem* 273, 4096-4105 (1998)

45. Voet D & Voet JG: Biochemistry New York. John Wiley & Sons, (1995)

46. Moolenaar W. H.: Lysophosphatidic acid, a multifunctional phospholipid messenger. *J Biol Chem* 270, 12949-12952 (1995)

47. Bursten S. L, W. E. Harris, K. Bomsztyk & D. Lovett: Interleukin-1 rapidly stimulates lysophosphatidate acyltransferase and phosphatidate phosphohydrolase activities in human mesangial cells. *J Biol Chem* 266, 20732-20743 (1991)

48. Bursten S, R. Weeks, J. West, T. Le, T. Wilson, D. Porubek, J. A. Bianco, J. W. Singer & G. C. Rice: Potential role for phosphatidic acid in mediating the inflammatory responses to TNF alpha and IL-1 beta. *Circ Shock* 44, 14-29 (1994)

49. Soyombo A. A. & S. L. Hofmann: Molecular cloning and expression of palmitoyl-protein thioesterase 2 (PPT2), a homolog of lysosomal palmitoyl-protein thioesterase with a distinct substrate specificity. *J Biol Chem* 272, 27456-27463 (1997)

50. Aguado B. & R. D. Campbell: Characterization of a human MHC class III region gene product with S-thioesterase activity. *Biochem J* 341 ( Pt 3), 679-689 (1999)

51. Bhatnagar R & Gordon J: Understanding covalent modifications of proteins by lipids:where cell biology and biophysics mingle. *Trends Cell Biol* 7, 14-20 (1997)

52. Vesa J, E. Hellsten, L. A. Verkruyse, L. A. Camp, J. Rapola, P. Santavuori, S. L. Hofmann & L. Peltonen: Mutations in the palmitoyl protein thioesterase gene causing infantile neuronal ceroid lipofuscinosis. *Nature* 376, 584-587 (1995)

53. Soyombo A. A, W. Yi & S. L. Hofmann: Structure of the human palmitoyl-protein thioesterase-2 gene (PPT2) in the major histocompatibility complex on chromosome 6p21.3. *Genomics* 56, 208-216 (1999)

54. Min J, H. Shukla, H. Kozono, S. K. Bronson, S. M. Weissman & D. D. Chaplin: A novel Creb family gene telomeric of HLA-DRA in the HLA complex. *Genomics* 30, 149-156 (1995)

55. Khanna A. & R. D. Campbell: The gene G13 in the class III region of the human MHC encodes a potential DNA-binding protein. *Biochem J* 319 ( Pt 1), 81-89 (1996)

56. KBM Reid & Campbell RD: Structure and organization of complement genes. In Complement in Health and Disease. Eds: Whaley K, Loos M. & Weiler J.M., Kluwer Academic Publishers, Dordrecht, 89-125 (1993)

57. Mizuki N, H. Ando, M. Kimura, S. Ohno, S. Miyata, M. Yamazaki, H. Tashiro, K. Watanabe, A. Ono, S. Taguchi, C. Sugawara, Y. Fukuzumi, K. Okumura, K. Goto, M. Ishihara, S. Nakamura, J. Yonemoto, Y. Y. Kikuti, T. Shiina, L. Chen, A. Ando, T. Ikemura & H. Inoko: Nucleotide sequence analysis of the HLA class I region spanning the 237- kb segment around the HLA-B and -C genes. *Genomics* 42, 55-66 (1997)

58. Wu L.-C, B. J. Morley & R. D. Campbell: Cell-specific expression of the human complement protein factor B gene: Evidence for the role of two distinct 5'-flanking elements. *Cell* 48, 331-342 (1987)

59. Belt K. T, M. C. Carroll & R. R. Porter: The structural basis of the multiple forms of human complement component C4. *Cell* 36, 907-914 (1984)

60. Belt K. T, C. Y. Yu, M. C. Carroll & R. R. Porter: Polymorphism of human complement component C4. *Immunogenetics* 21, 173-180 (1985)

61. Yu C. Y, K. T. Belt, C. M. Giles, R. D. Campbell & R. R. Porter: Structural basis of the polymorphism of human complement component C4A and C4B: gene size, reactivity and antigenicity. *EMBO J* 5, 2873-2881 (1986)

62. Dangel A. W, A. R. Mendoza, B. J. Baker, C. M. Daniel, M. C. Carroll, L.-C. Wu & C. Y. Yu: The dichotomous size variation of human complement C4 gene is mediated by a novel family of endogenous retroviruses which also establishes species-specific genomic patterns among Old World primates. *Immunogenetics* 40, 425-436 (1994)

63. Tassabehji M, T. Strachan, M. Anderson, R. D. Campbell, S. Collier & M. Lako: Identification of a novel family of human endogenous retroviruses and characterization of one family member, HERV-K(C4), located in the complement C4 gene cluster. *Nucleic Acids Res* 22, 5211-5217 (1994)

64. Campbell R.D.: The human major histocompatibility complex:a4000kb segment of the human genome replete with genes. In: Cold Spring Harbor Laboratory Press. Eds: Davies K E, Tilghman SM, 1-33 (1993)

65. Hauptmann G, G. Tappeiner, & J. A. Schifferli: Inherited deficiency of the fourty component of human complement. In: Hardwood Academic Publishers. Eds: Rosen F S, Seligmann M, 253-266 (1993)

66. Yang Z, A. R. Mendoza, T. R. Welch, W. B. Zipf & C. Y. Yu: Modular variations of HLA class III genes for serine/threonine kinase RP, complement C4, steroid 21-hydroxylase CYP21 and tenascin TNX (RCCX): a mechanism for gene deletions and disease associations. *J Biol Chem* 274, 12147-12156 (1999)

67. Higashi Y, H. Yoshioka, M. Yamane, O. Gotoh & Y. Fujii-Kuriyama: Complete nucleotide sequence of two steroid 21-hydroxylase genes tandemly arranged in human chromosome: A pseudogene and a genuine gene. *Proc Natl Acad Sci U S A* 83, 2841-2845 (1986)

68. White P. C, M. I. New & B. DuPont: Structure of human steroid 21-hydroxylase genes. *Proc Natl Acad Sci U S A* 83, 5111-5115 (1986)

69. Bristow J, M. K. Tee, S. E. Gitelman, S. H. Mellon & W. L. Miller: Tenascin-X: a novel extracellular matrix protein encoded by the human XB gene overlapping P450c21B. *J Cell Biol* 122, 265-278 (1993)

70. Matsumoto K, Y. Saga, T. Ikemura, T. Sakakura & R. Chiquet-Ehrismann: The distribution of tenascin-X is distinct and often reciprocal to that of tenascin-C. *J Cell Biol* 125, 483-493 (1994)

71. Tee M. K, A. A. Thomson, J. Bristow & W. Miller: Sequences promoting transcription of the human XA gene overlapping P450c21A correctly predicts the presence of a novel, adrenal-specific truncated form of Tenascin-X. *Genomics* 28, 171-178 (1995)

72. Bristow J, S. E. Gitelman, M. K. Tee, B. Staels & W. L. Miller: Abundant adrenal-specific transcription of the human P450c21A "pseudogene". *J Biol Chem* 268, 12919-12924 (1993)

73. Gomez-Escobar N, C. F. Chou, W. W. Lin, S. L. Hsieh & R. D. Campbell: The G11 gene located in the major histocompatibility complex encodes a novel nuclear serine/threonine protein kinase. *J Biol Chem* 273, 30954-30960 (1998)

74. Yang Z, L. Shen, A. W. Dangel, L. C. Wu & C. Y. Yu: Four ubiquitously expressed genes, *RD* (D6S45)*–SKI2W* (SKIV2L) *–DOM3Z–RP1*(D6S60E), are present between complement component genes factor B and C4 in the class III region of the HLA. *Genomics* 53, 338-347 (1998)

75. Dangel A. W, L. Shen, A. R. Mendoza, L.-C. Wu & C. Y. Yu: Human helicase gene SKI2W in the HLA class III region exhibits striking structural similarities to the yeast antiviral gene SKI2 and to the human gene KIAA0052: emergence of a new gene family. *Nucleic Acids Res* 23, 2120-2126 (1995)

76. Qu X, Z. Yang, S. Zhang, L. Shen, A. W. Dangel, J. H. Hughes, K. L. Redman, L. C. Wu & C. Y. Yu: The human DEVH-box protein Ski2w from the HLA is located in nucleoli and ribosomes. *Nucleic Acids Res* 26, 4068-4077 (1998)

77. Milner C. M. & R. D. Campbell: The G9a gene in the human major histocompatibility complex encodes a novel protein containing ankyrin-like repeats. *Biochem J* 290 ( Pt 3), 811-818 (1993)

78. Rozenblatt-Rosen O, T. Rozovskaia, D. Burakov, Y. Sedkov, S. Tillib, J. Blechman, T. Nakamura, C. M. Croce, A. Mazo & E. Canaani: The C-terminal SET domains of ALL-1 and TRITHORAX interact with the INI1 and SNR1 proteins, components of the SWI/SNF complex. *Proc Natl Acad Sci U S A* 95, 4152-4157 (1998)

79. Milner C. M, S. V. Smith, M. B. Carrillo, G. L. Taylor, M. Hollinshead & R. D. Campbell: Identification of a sialidase encoded in the human major histocompatibility complex. *J Biol Chem* 272, 4549-4558 (1997)

80. Pshezhetsky A. V, C. Richard, L. Michaud, S. Igdoura, S. Wang, M. A. Elsliger, J. Qu, D. Leclerc, R. Gravel, L. Dallaire & M. Potier: Cloning, expression and chromosomal mapping of human lysosomal sialidase and characterization of mutations in sialidosis. *Nat Genet* 15, 316-320 (1997)

81. van der Spoel A, E. Bonten & A. d'Azzo: Transport of human lysosomal neuraminidase to mature lysosomes requires protective protein/cathepsin A. *EMBO J* 17, 1588-1597 (1998)

82. Lowden J. A. & J. S. O'Brien: Sialidosis: a review of human neuraminidase deficiency. *Am J Hum Genet* 31, 1-18 (1979)

83. Rudenko G, E. Bonten, W. G. Hol & A. d'Azzo: The atomic model of the human protective protein/cathepsin A suggests a structural basis for galactosialidosis. *Proc Natl Acad Sci U S A* 95, 621-625 (1998)

84. Klein D. & J. Klein: Polymorphism of the Apl (Neu-1) locus in the mouse. *Immunogenetics* 16, 181-184 (1982)

85. Chen X. P, E. Y. Enioutina & R. A. Daynes: The control of IL-4 gene expression in activated murine T lymphocytes: a novel role for neu-1 sialidase. *J Immunol* 158, 3070-3080 (1997)

86. Milner C. M. & R. D. Campbell: Structure and expression of the three MHC-linked HSP70 genes. *Immunogenetics* 32, 242-251 (1990)

87. Srivastava P. K, H. Udono, N. E. Blachere & Z. Li: Heat shock proteins transfer peptides during antigen processing and CTL priming. *Immunogenetics* 39, 93-98 (1994)

88. Suto R. & P. K. Srivastava: A mechanism for the specific immunogenicity of heat shock protein-chaperoned peptides. *Science* 269, 1585-1588 (1995)

89. Deguchi Y. & S. Kishimoto: Elevated expression of heat-shock protein gene in the fibroblasts of patients with scleroderma. *Clin Sci (Colch )* 78, 419-422 (1990)

90. Twomey B. M, V. Amin, D. A. Isenberg & D. S. Latchman: Elevated levels of the 70 kD heat shock protein in patients with systemic lupus erythematosus are not dependent on enhanced transcription of the hsp70 gene. *Lupus* 2, 297-301 (1993)

91. Furukawa F, K. Ikai, N. Matsuyoshi, K. Shimizu & S. Imamura: Relationship between heat shock protein induction and the binding of antibodies to the extractable nuclear antigens on cultured human keratinocytes. *J Invest Dermatol* 101, 191-195 (1993)

92. Multhoff G, C. Botzler, L. Jennen, J. Schmidt, J. Ellwart & R. Issels: Heat shock protein 72 on tumor cells: a recognition structure for natural killer cells. *J Immunol* 158, 4341-4350 (1997)

93. LeJohn H. B, L. E. Cameron, B. Yang & S. L. Rennie: Molecular characterization of an NAD-specific glutamate dehydrogenase gene inducible by L-glutamine. Antisense gene pair arrangement with L-glutamine-inducible heat shock 70-like protein gene. *J Biol Chem* 269, 4523-4531 (1994)

94. Konstantopoulou I, C. A. Ouzounis, E. Drosopoulou, M. Yiangou, P. Sideras, C. Sander & Z. G. Scouras: A *Drosophila* hsp70 gene contains long, antiparallel, coupled open reading frames (LAC ORFs) conserved in homologous loci. *J Mol Evol* 41, 414-420 (1995)

95. Hsieh S. L. & R. D. Campbell: Evidence that gene G7a in the human major histocompatibility complex encodes valyl-tRNA synthetase. *Biochem J* 278, 809-816 (1991)

96. Vilalta A, D. Donovan, L. Wood, G. Vogeli & D. C. Yang: Cloning, sequencing and expression of a cDNA encoding mammalian valyl-tRNA synthetase. *Gene* 123, 181-186 (1993)

97. Her C. & N. A. Doggett: Cloning, structural characterization, and chromosomal localization of the human orthologue of *Saccharomyces cerevisiae* MSH5 gene. *Genomics* 52, 50-61 (1998)

98. Winand N. J, J. A. Panzer & R. D. Kolodner: Cloning and characterization of the human and Caenorhabditis elegans homologs of the *Saccharomyces cerevisiae* MSH5 gene. *Genomics* 53, 69-80 (1998)

99. Bocker T, A. Barusevicius, T. Snowden, D. Rasio, S. Guerrette, D. Robbins, C. Schmidt, J. Burczak, C. M. Croce, T. Copeland, A. J. Kovatich & R. Fishel: hMSH5: a human MutS homologue that forms a novel heterodimer with hMSH4 and is expressed during spermatogenesis. *Cancer Res* 59, 816-822 (1999)

100. Ross-Macdonald P. & G. S. Roeder: Mutation of a meiosis-specific MutS homolog decreases crossing over but not mismatch correction. *Cell* 79, 1069-1080 (1994)

101. Bawa S. & W. Xiao: A mutation in the MSH5 gene results in alkylation tolerance. *Cancer Res* 57, 2715-2720 (1997)

102. Valenzuela S. M, D. K. Martin, S. B. Por, J. M. Robbins, K. Warton, M. R. Bootcov, P. R. Schofield, T. J. Campbell & S. N. Breit: Molecular cloning and expression of a chloride ion channel of cell nuclei. *J Biol Chem* 272, 12575-12582 (1997)

103. Kimoto M, T. Sasakawa, H. Tsuji, S. Miyatake, T. Oka, N. Nio & T. Ogawa: Cloning and sequencing of cDNA encoding NG,NG-dimethylarginine dimethylaminohydrolase from rat kidney. *Biochim Biophys Acta* 1337, 6-10 (1997)

104. Williams A. F. & A. N. Barclay: The immunoglobulin superfamily--domains for cell surface recognition. *Annu Rev Immunol* 6, 381-405 (1988)

105. Rock K. L, H. Reiser, A. Bamezai, J. McGrew & B. Benacerraf: The LY-6 locus: a multigene family encoding

phosphatidylinositol-anchored membrane proteins concerned with T-cell activation. *Immunol Rev* 111, 195-224 (1989)

106. Williams A. F.: Emergence of the Ly-6 superfamily of GPI-anchored molecules. *Cell Biol Int Rep* 15, 769-777 (1991)

107. Kamiura S, C. M. Nolan & D. Meruelo: Long-range physical map of the Ly-6 complex: mapping the Ly-6 multigene family by field-inversion and two-dimensional gel electrophoresis. *Genomics* 12, 89-105 (1992)

108. Pyerin W.: Human casein kinase II: structures, genes, expression and requirement in cell growth stimulation. *Adv Enzyme Regul* 34, 225-246 (1994)

109. Tuazon P. T. & J. A. Traugh: Casein kinase I and II--multipotential serine protein kinases: structure, function, and regulation. *Adv Second Messenger Phosphoprotein Res* 23, 123-164 (1991)

110. Iimoto D. S, E. Masliah, R. DeTeresa, R. D. Terry & T. Saitoh: Aberrant casein kinase II in Alzheimer's disease. *Brain Res* 507, 273-280 (1990)

111. Utans U, W. C. Quist, B. M. McManus, J. E. Wilson, R. J. Arceci, A. F. Wallace & M. E. Russell: Allograft inflammatory factory-1. A cytokine-responsive macrophage molecule expressed in transplanted human hearts. *Transplantation* 61, 1387-1392 (1996)

112. Olavesen M G, W. Thomson, & J. Cheng: Characterisation of a novel gene (G1) in the class III region of the human MHC. In: Proceedings of the 11th International Histocompatibility Workshop.Oxford: Oxford University Press. Eds: Tsuji K, Aizawa M, Sasazuki T, 190-193 (1992)

113. Chen Z. W, B. Ahren, C. G. Ostenson, A. Cintra, T. Bergman, C. Moller, K. Fuxe, V. Mutt, H. Jornvall & S. Efendic: Identification, isolation, and characterization of daintain (allograft inflammatory factor 1), a macrophage polypeptide with effects on insulin secretion and abundantly present in the pancreas of prediabetic BB rats. *Proc Natl Acad Sci U S A* 94, 13879-13884 (1997)

114. Harpaz Y. & C. Chothia: Many of the immunoglobulin superfamily domains in cell adhesion molecules and surface receptors belong to a new structural set which is close to that containing variable domains. *J Mol Biol* 238, 528-539 (1994)

115. Pende D, S. Parolini, A. Pessino, S. Sivori, R. Augugliaro, L. Morelli, E. Marcenaro, L. Accame, A. Malaspina, R. Biassoni, C. Bottino, L. Moretta & A. Moretta: Identification and molecular characterization of NKp30, a novel triggering receptor involved in natural cytotoxicity mediated by human natural killer cells. *J Exp Med* 190, 1505-1516 (1999)

116. Holzinger I, A. de Baey, G. Messer, G. Kick, H. Zwierzina & E. H. Weiss: Cloning and genomic characterization of LST1: a new gene in the human TNF region. *Immunogenetics* 42, 315-322 (1995)

117. de Baey A, B. Fellerhoff, S. Maier, S. Martinozzi, U. Weidle & E. H. Weiss: Complex expression pattern of the TNF region gene LST1 through differential regulation, initiation, and alternative splicing. *Genomics* 45, 591-600 (1997)

118. Neville M. J. & R. D. Campbell: Alternative splicing of the LST-1 gene located in the Major Histocompatibility Complex on human chromosome 6. *DNA Seq* 8, 155-160 (1997)

119. Gruss H. J. & S. K. Dower: Tumor necrosis factor ligand superfamily: involvement in the pathology of malignant lymphomas. *Blood* 85, 3378-3404 (1995)

120. Matsumoto M, S. Mariathasan, M. H. Nahm, F. Baranyay, J. J. Peschon & D. D. Chaplin: Role of lymphotoxin and the type I TNF receptor in the formation of germinal centers. *Science* 271, 1289-1291 (1996)

121. Neumann B, A. Luz, K. Pfeffer & B. Holzmann: Defective Peyer's patch organogenesis in mice lacking the 55-kD receptor for tumor necrosis factor. *J Exp Med* 184, 259-264 (1996)

122. Pasparakis M, L. Alexopoulou, V. Episkopou & G. Kollias: Immune and inflammatory responses in TNF alpha-deficient mice: a critical requirement for TNF alpha in the formation of primary B cell follicles, follicular dendritic cell networks and germinal centers, and in the maturation of the humoral immune response. *J Exp Med* 184, 1397-1411 (1996)

123. Albertella M. R. & R. D. Campbell: Characterization of a novel gene in the human major histocompatibility complex that encodes a potential new member of the I kappa B family of proteins. *Hum Mol Genet* 3, 793-799 (1994)

124. Beg A. A. & A. S. Baldwin: The I kappa B proteins: multifunctional regulators of Rel/NF-kappa B transcription factors. *Genes Dev* 7, 2064-2070 (1993)

125. Crider B. P, P. Andersen, A. E. White, Z. Zhou, X. Li, J. P. Mattsson, L. Lundberg, D. J. Keeling, X. S. Xie, D. K. Stone & S. B. Peng: Subunit G of the vacuolar proton pump. Molecular characterization and functional expression. *J Biol Chem* 272, 10721-10728 (1997)

126. Peelman L. J, P. Chardon, M. Nunes, C. Renard, C. Geffrotin, M. Vaiman, A. Van Zeveren, W. Coppieters, W. A. van de, Y. Bouquet & .: The BAT1 gene in the MHC encodes an evolutionarily conserved putative nuclear RNA helicase of the DEAD family. *Genomics* 26, 210-218 (1995)

127. Schmid S. R. & P. Linder: D-E-A-D protein family of putative RNA helicases. *Mol Microbiol* 6, 283-291 (1992)

**Key Words:** MHC Class III Region, hsp70, Lysophosphatidic Acetyl Transferase, Novel Ly-6 Ssuperfamily Members, Novel IgG Superfamily Members, Candidate Autoimmune Disease Susceptibility Genes, Review

**Send correspondence to**: Dr C. M. Milner, D.Phil., MRC Immunochemistry Unit, Department of Biochemistry, South Parks Road, Oxford OX1 3QU, United Kingdom, Tel: +44-1865-275349, Fax: +44-1865-275729, E-mail: cmilner@molbiol.ox.ac.uk