Human Immunology 80 (2019) 67–78

Contents lists available at ScienceDirect

Human Immunology

journal homepage: www.elsevier.com/locate/humimm

The MHC in the era of next-generation sequencing: Implications for bridging structure with function

Effie W. Petersdorf[a,*], Colm O'hUigin[b]

[a] University of Washington, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, D4-115, Seattle, WA 98109, United States
[b] Frederick National Laboratory for Cancer Research sponsored by the National Cancer Institute, Microbiome and Genetics Core, Building 37, Room 4140B, Bethesda, MD 20852, United States

ARTICLE INFO

Keywords:
Linkage disequilibrium
Phase
Phylogeny
Linage-specific expression
HIV-AIDs
Hepatitis B
Crohn's disease
Unrelated donor hematopoietic cell transplantation
Graft-versus-host disease

ABSTRACT

The MHC continues to have the most disease-associations compared to other regions of the human genome, even in the genome-wide association study (GWAS) and single nucleotide polymorphism (SNP) era. Analysis of non-coding variation and their impact on the level of expression of HLA allotypes has shed new light on the potential mechanisms underlying HLA disease associations and alloreactivity in transplantation. Next-generation sequencing (NGS) technology has the capability of delineating the phase of variants in the HLA antigen-recognition site (ARS) with non-coding regulatory polymorphisms. These relationships are critical for understanding the qualitative and quantitative implications of HLA gene diversity. This article summarizes current understanding of non-coding region variation of HLA loci, the consequences of regulatory variation on HLA expression, the role for evolution in shaping lineage-specific expression, and the impact of HLA expression on disease susceptibility and transplantation outcomes. A role for phased sequencing methods for the MHC, and perspectives for future directions in basic and applied immunogenetic studies of the MHC are presented.

1. Introduction

The MHC has the most associations to human disease than any other region of the human genome [1,2]. Before the availability of SNP typing platforms, studies focused on the evaluation of coding regions of the classical HLA genes (exons 2 and 3 for HLA-A, B and C, and exon 2 for HLA-DRB1, DQB1 and DPB1) to better understand the implications of the ARS on adaptive immunity. Together with crystallographic structures [3,4] DNA typing methods provided the foundation for the identification of many HLA disease associations and for understanding the transplantation barrier [5,6]. More recently, information on non-coding variation within and between HLA genes and their influence on HLA expression has accelerated investigation into the relationship between HLA structure and function (Table 1). The three hallmarks of the MHC, diversity, linkage disequilibrium (LD) and the persistence of alleles far beyond their expected lifetimes (ancient allelic coalescence) are principles that bridge coding and non-coding MHC variation. This perspective will describe the landscape of MHC variation, its impact on HLA gene expression and disease risk, and the role of NGS methods for future investigation into the MHC as the cornerstone of the immune response.

2. MHC diversity, linkage disequilibrium and evolution

Over 18,771 HLA class I and II sequences are currently recognized and the tempo of new sequence discovery is anticipated to increase dramatically with the application of NGS to the HLA system [44, https://www.ebi.ac.uk/ipd/imgt/hla/]. Among jawed vertebrate loci, MHC allelic polymorphism is uniquely functional, well-differentiated and ancient. Generally, extensive polymorphism reflects high mutation rates (such as found at microsatellites), but MHC mutation rates are low [45] and the profound polymorphism is a consequence of balancing selection which reduces allelic turnover rates [46,47]. The selection results in retention of allelic variants that would otherwise be lost by drift. As a consequence, MHC polymorphisms represent an accumulation of millions of years of genetic variation, with divergences far more ancient than found for allelic polymorphism elsewhere in the genome [48,49]. The HLA region encompasses variability exceeding by an order of magnitude the expected accumulation in the absence of selection [50].

Although balancing selection may be limited to MHC genes and their controlling elements, LD across the loci allows polymorphism to accumulate in intergenic regions by hitchhiking [51,52]. The strong maintenance of variation on haplotypes poses challenges in isolating

* Corresponding author.
E-mail addresses: epetersd@fredhutch.org (E.W. Petersdorf), ohuiginc@mail.nih.gov (C. O'hUigin).

https://doi.org/10.1016/j.humimm.2018.10.002
Received 29 August 2018; Received in revised form 24 September 2018; Accepted 1 October 2018
Available online 12 October 2018

Table 1
Sequence features associated with HLA expression and clinical relevance. This table represents a selection of the vast literature on MHC region variation that influences HLA expression, and impact of HLA expression on specific disease or traits. NE, not evaluated; GVHD, graft-versus-host disease; miRNA, microRNA; eQTL, expression quantitative trait locus.

Locus	Sequence Feature/Method	Expression	Clinical Relevance	Citation
Haplotypes	Tiling and splice junction microarray	A1-B8-Cw7-DR2, A3-B7-Cw7-DR15, A26-B18-Cw5-DR3-DQ2 haplotype-specific expression	NE	[7]
	eQTLs	A2-B46-DR9, A33-B58-DR3, A1-B8-DR3 regulatory effects proximate to HLA-A, C, DRB and C4A	NE	[8]
	MicroRNAs	MicroRNAs located within non-coding regions describe local and extended haplotypes	NE	[9]
Class I	Class I regulatory complex ("CRC")	KB1 enhancer conserved in HLA-A and B sequences Locus-specific KB2 divergence Locus-specific IRS	NE	[10]
HLA-A	Methylation DNA sequencing of promoter	Lineage-specific gradient of expression Increased methylation associated with reduced HLA-A expression	NE	[11]
	HLA-A2, B8; C-terminus of α2 and α3 domains	Residue 180 and 239 associated with HLA-A and B expression HLA-A2 Gln^{180} and Gly^{239} associated with higher cell surface expression compared to HLA-B Glu^{180} and Arg^{239}	NE	[12]
	HLA-A11 promoter	Positions -271 to -263 and -242 to -234 involved in binding zinc finger protein ZFX and impact transcriptional activity of promoter.	NE	[13]
HLA-B	Bw4-80I	KIR3DL1/Bw4 combinations have array of binding strengths and correlate with NK cytotoxicity	NE	[14]
	Bw4/Bw6 cell surface expression	Lineage and cell specific expression related to stability and turnover	NE	[15]
HLA-C	Cell surface expression	Continuous allotype-specific HLA-C expression	HIV Viral load Crohn's Disease	[16]
	Inference between HLA-C allele and MFI from ref 16	HLA-C patient/donor mismatches according to level of patient's mismatched allotype	GVHD higher when mismatching against high-expression patient HLA-C-	[17]
	Phylogeny	Ancient escape from miR-148a repression of HLA-C lineages	NE	[18]
	cDNA transfection of HLA-A,B,C	HLA-C mRNA expression lower than HLA-B	NE	[19]
	Rs9264942 ("-35 SNP")	Variant upstream of HLA-C; proxy for miR-148a	HIV viral load	[2,20]
		mRNA levels and cell surface expression for C*06-positive samples from psoriasis cohort	No correlation of HLA-C*06 with -35 SNP in psoriasis	[21]
		Unrelated donor transplants	No correlation of -35 SNP or HLA-C expression levels with unrelated HCT outcomes	[22]
		No correlation between -35 SNP and HLA-C mRNA expression HLA-C expression correlated with HLA-A-B-C-DR and B-C haplotype	NE	[23]
	miR-148a of HLA-C 3′ UTR	Marked by rs926942 Downregulation of HLA-C7 and other alleles with intact microRNA; nucleotide deletion within miR-148a binding site leads to higher HLA-C expression	HIV?	[24]
		Intact miR-148 impacts HIV and Crohn's risks	HIV, Crohn's Disease	[25]
	Rs239541 Within Oct1 binding site in HLA-C promoter	A allele higher affinity than G allele G associated with lower promotor activity	NE	[26]
	Nucleotide sequence of HLA-B, A, C promoter	HLA-B promoter most diverse, but no correlation with mRNA expression No clear correlation between promoter phylogeny and lineage expression	NE	[27]
	Alternative transcript arising from NK-specific upstream promoter	NK intrinsic regulation of HLA-C expression	NE	[28]
	HLA-C*03:03/03:04 mismatch	Low-expression C*03:03/03:04 mismatch is well-tolerated in HCT	HLA-C mismatched unrelated HCT	[29]
Class II	Low-expression HLA-DRB3, DRB4, DRB5 mismatches	Additive effect of patient-donor mismatching for low-expression HLA-DRB3, DRB4, DRB5 on transplant outcomes	Unrelated donor HCT	[30]
	HLA-DR-XL9-DQ haplotypes	IRF4 and CTCF binding site variants modified HLA-DRB1/DQA1/DQB1 transcription and HLA-DR/DQ surface expression	SLE	[31]
	RFX, X2BP, NF-Y, CIITA, Oct-2, Bob-1	CIITA is a molecular switch for class II expression and IFN-γ regulation	NE	[32]
HLA-DRA	Histone modifications and binding to RFX and CIITA	Network of histone modifying proteins together with multi-subunit complex impact class II expression	NE	[33]
HLA-DRB	Promoter sequence	Two promoter point mutations associated with DR8 and DR10 did not affect promoter activity luciferase assay	NE	[34]
	Vitamin D response element in DRB1 promoter on DRB1*15 haplotypes	DR15 expression modulated by stimulation of response element by Vitamin D	Multiple Sclerosis	[35]

(*continued on next page*)

Table 1 (*continued*)

Locus	Sequence Feature/Method	Expression	Clinical Relevance	Citation
	Histone quantitative trait loci	Enrichment of loci on autoimmune disease haplotypes and influence gene expression	Systemic Lupus Erythematosis	[36]
HLA-DQ	HLA-DQB1 transcripts and promoter footprinting	Differential expression of HLA-DQB1 alleles	NE	[37]
	qPCR	Transcription of HLA-DQ associated with polymorphism in promoter of DQB1*03:01 and 03:02.	NE	[38]
		Spacing of W and X1 elements in the promoter		
		Differential expression of DQB1*03:01, 05:01, 06:02	NE	[39]
HLA-DP	Rs9277534 in 3′ UTR	Sequence analysis of rs9277535 and rs927534	Hepatitis B persistence with rs9277534G associated with high transcription levels and cell surface expression of HLA-DP	[40]
		Cell surface expression		
		Validation of r92775534A with lower mRNA expression than rs9277534G	Increased GVHD with mismatching against high-expression HLA-DPB1 mismatch in the transplant recipient	[41]
	Rs9277534A and G phylogenic clades	Patient/donor HLA-DPB1 mismatching for alleles within rs9277534A/G clades	GVHD increased with mismatching for rs9277534G high-expression alleles	[42]
	Rs9277534 and donor-specific antibodies	Rs9277534G high-expression HLA-DP associated with stronger B cell flow crossmatch	Solid organ transplantation	[43]

the true disease susceptibility variant [53–57]. Understanding the nature of LD and MHC haplotypes remains one of the most critical areas of immunogenetics research, and one which will continue to contribute vital information in the NGS era [58].

HLA haplotypes were first described by Ceppellini and coworkers in 1967 [59]. The study of pedigrees was instrumental in the identification of inherited segments characterized by "genetic fixity" of "blocks" of genes within the HLA-B/C and HLA-DR/DQ regions [60–63]. In the SNP era, "SNP blocks" describe the same phenomenon of strong positive LD, albeit among single base changes that travel together [64–66].

The "blockiness" of blocks is a function of recombination events, and their distribution across the MHC resembles that found elsewhere in the human genome, with tight clustering into recombinational hotspots [67,68]. The evolutionary origins of extended haplotypes might be interpreted as a consequence of the localization of recombination to dispersed hotspots within the HLA [69] allied to protracted retention of favorable elements or combinations. Measured HLA recombination rates outside of hotspots vary by haplotypes and location. Furthermore, hotspots themselves can be created or lost over moderate timescales [70,71]. Considerations of functional differentiation and allelic repertoire lead often to consideration of loci as separate entities. However, for tightly linked HLA loci such as HLA-DR and -DQ or HLA-B and -C balancing selection will likely act on haplotypes more often than on the individual loci. The tight LD can lead to shared usage of genomic regulatory elements, as seen of HLA-DR and DQ [31].

LD across the MHC is long-range. In 1999, a contiguous 3.6 Mb sequence of the MHC was reported [72], followed by sequences for 8 HLA haplotypes selected for their relevance in type 1 diabetes and multiple sclerosis [73–75]. These data show conservation of HLA haplotypes at the nucleotide level and blocks of ancestry-specific variation [65,73,76–82]. Extension of these studies to diverse HLA haplotypes using NGS provides an unprecedented visualization of MHC content and organization [83], with immediate benefits to basic and clinical research.

3. Impact of regulatory variation on HLA expression

Early studies of non-coding regions of the MHC used microsatellites to define recombination rates and to identify functional coding and regulatory elements [84,85]. The completion of the sequence of the first human genome paved the way for GWAS [86], and offered new perspectives to the diversity of non-coding regions [76,87].

GWAS have yielded a large number of associations between SNPs and disease phenotypes, the vast majority of which reside in non-coding regions [88]. An impressive number of disease-association SNPs map to the MHC [5,89]. Although the SNPs provide information on regions that might encode a disease gene of interest, they do not describe the biological mechanism(s) underlying the disease or trait. To do so requires a survey for variation that influence gene expression. To this end, the ENCODE (ENCyclopedia Of DNA Elements) consortium was launched to comprehensively annotate DNA elements within promoters, transcriptional regulatory sequences, and targeted regions where chromatin structure and histone modification could alter gene regulation and expression [90–95]. In addition, eQTLs have been instrumental in the development of global map of gene expression variants [96]. In the MHC, expression quantitative locus/loci (eQTLs) may influence gene expression through a single SNP or several haplotype-specific SNPs [7,97–107]. These annotations of human variation provide a rich resource for bridging HLA regulatory polymorphisms with their ARS to understand allotype and haplotype-specific expression and its consequences on disease susceptibility.

3.1. HLA class I expression

It has long been known that classical HLA class I proteins are expressed at different levels [19,108]. In particular, HLA-C appears to be

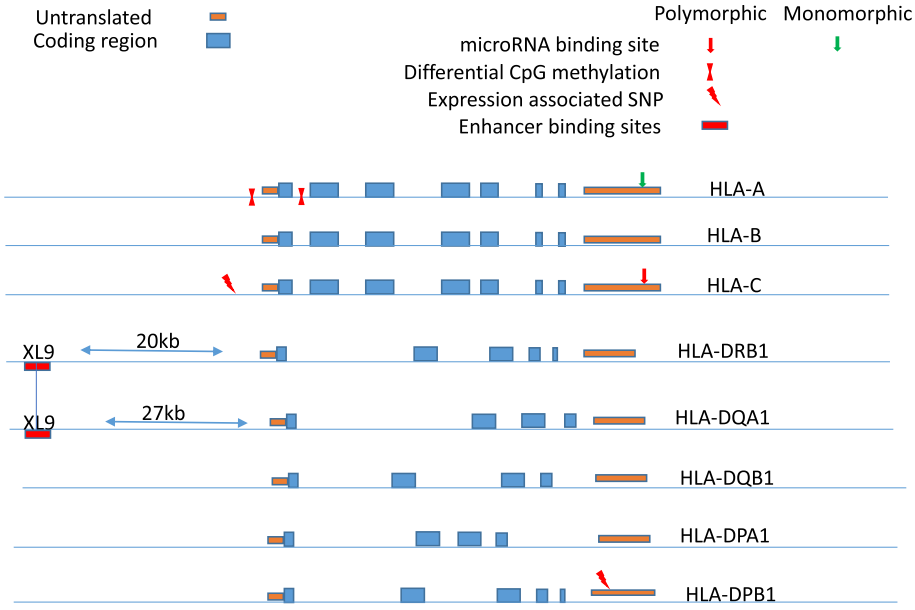

Fig. 1. Simplified representation of HLA genes showing positions of some known polymorphic expression controlling elements or associated SNPs. Genes are oriented left to right in the direction of transcription. Polymorphic expression control can be exerted through differential methylation (HLA-A), post-translational assembly and export (HLA-B), miRNA binding and promoter changes (HLA-C), Enhancer polymorphisms (DRB1, DQB1), or by unknown mechanisms tracked by SNP associations (DPB1). Combinations of the mechanisms are possible and not all expression variation is explained by these known elements.

expressed on the cell surface at levels about one tenth that of HLA-B and HLA-A [109]. Differences in cell surface expression can arise through initial transcription, allele-specific rate limiting formation of peptide-HLA complexes, and differences in cell surface stability. This review is limited to variation that affects HLA cell surface expression (Fig. 1; Table 1). Ultimately, cognate T-cell recognition might limit functional aspects of HLA expression but will not be considered here.

Serologically-related HLA alleles form distinct clades in phylogenetic reconstructions. Usually referred to as lineages, these allelic groupings form a basic distinctive functional unit, related by evolutionary history, serological affinities, binding repertoire and expression levels. For these reasons, much of the analysis of expression differences focuses on patterns found for allelic lineages, rather than patterns specific to unique alleles. Among classical class I loci, 21 lineages are distinguished for HLA-A, 36 for HLA-B and 14 for HLA-C. The prevalence of polymorphism within and between HLA loci mean that elements that control transcription are often polymorphic themselves, resulting in transcriptional differences among lineages.

One of the first health outcomes to implicate HLA expression was HIV viral load (VL) control and HLA-C expression. Early studies identified rs9264942 ("−35" SNP), equidistant from HLA-B and –C, in HIV VL [2,20]. Rs9264942 marked a polymorphic miR-148a binding site in the 3' untranslated region (UTR) of HLA-C. The polymorphism itself was complex, involved multiple significant changes that ablated the miR-148a binding site, and influenced expression levels [24]. The higher expression variant was newer and had likely promoted diversification and expansion of new HLA-C allelic lineages [18]. A second polymorphism of an OCT1 binding site in the promoter region was shown to provide further control of HLA-C expression [26].

These biallelic polymorphisms explain some, but not all of the lineage-associated variation in expression. Lineage-specific antibody reactivity shows that each HLA-C lineage has a characteristic basal level of mRNA and cell surface expression level [16]. When inferred HLA-C lineage expression levels were applied to assess VLs in HIV-infected cases and controls, the association between HLA-C expression and VL was predictive when incorporated alongside other known allelic effects. Furthermore, cytotoxic T-cell restriction of viral peptides correlated well with inferred HLA-C expression levels.

Although average lineage expression patterns can be arranged as a gradient of estimates per lineage, many of the estimates differ only slightly from each other and whether significantly distinct expression categories are functionally relevant remains to be determined. The polymorphic regulators of HLA-C expression that have already been identified are shown in Fig. 1. Although these regulatory polymorphisms can be ascribed to events that occurred on discrete branches of the phylogeny, they can occur singly or in combination on the phylogeny. Consequently, each individual regulatory effect can be mediated by influences from other regulators, so that expression varies unpredictably along branches and high expression is a character that is polyphyletic, occurring independently on several branches (Fig. 2). The lack of concordance of HLA-C lineage expression with known regulators in the phylogeny also suggests the existence of additional undefined modulators of expression.

Although a qPCR gradient of lineage-specific expression levels is observed for HLA-A lineages, like HLA-C no candidate genetic polymorphisms as yet are associated with expression differences [11]. The miR-148a binding site is shared by both HLA-A and HLA-C but is non-polymorphic in HLA-A. Furthermore, available promoter sequences reveal no major differences that can be directly related to changes in expression. The most conspicuous differences in HLA-A lineage expression are found with epigenetic methylation patterns usually associated with downregulation [11]. Despite near identical CpG distribution upstream and in intron 1, representative high-expression HLA-A lineages (HLA-A*24) are less methylated than low-expression lineages (HLA-A*03) when analyzed using bisulfite sequencing. Furthermore, immunoprecipitation of sonicated DNA using DNA-methylation specific antibodies indicate an inverse correlation with expression among lineages. Finally, *in vitro* treatment of cells with the methyltransferases inhibitor 5′-Aza-dCR corroborated results of bisulfite sequencing. In addition to methylation differences, HLA-A lineages also differ in their usage of alternate polyadenylation sites with one of the two variants being associated with posttranscriptional lineage-specific regulation by synaptotagmin binding cytoplasmic RNA interacting protein (Syncrip) [110].

Methylation does not appear to significantly influence lineage-specific expression of HLA-C or HLA-B [11]. qPCR does not reveal any variation in transcription-dependent expression of HLA-B lineages [27]. The binding site for miR-148a is ablated in an identical manner in all lineages, presumably indicating ancestral loss. Neither does high throughput RNA sequencing in two populations reveal differences in transcription levels for HLA-B [15]. On the cell surface however, there is evidence for differential level of HLA-B proteins in a lineage-specific pattern. Antibodies that recognize either universally shared class I epitopes, or specifically the Bw4 or Bw6 epitopes, have enabled

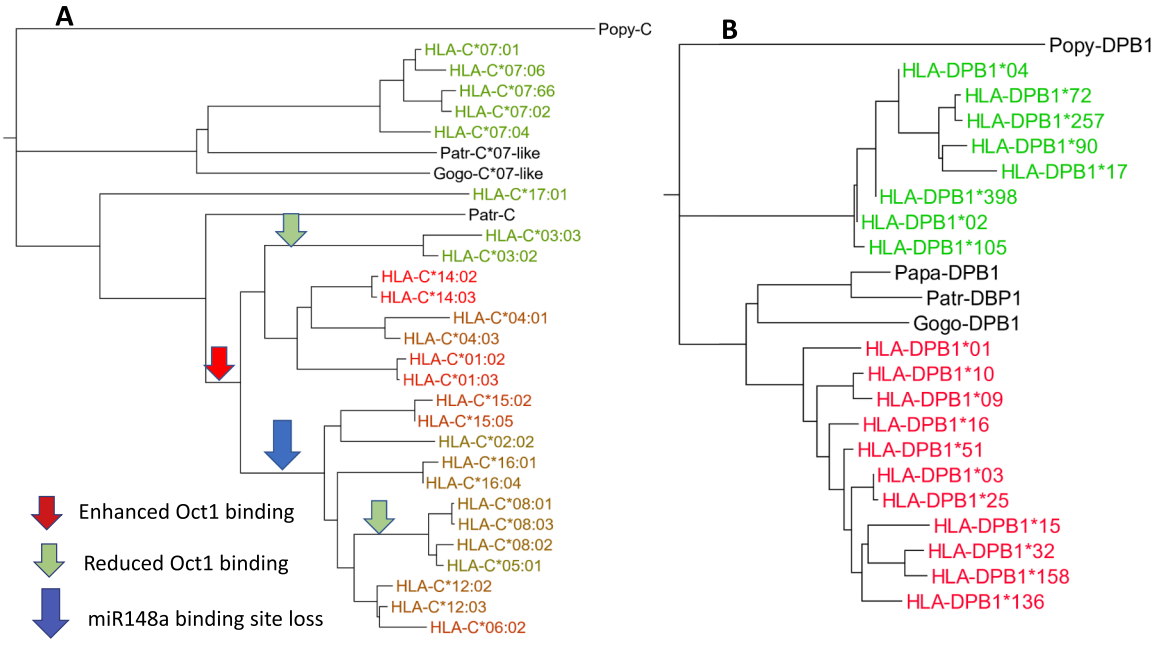

Fig. 2. Phylogeny of expression patterns for HLA-C (A) and inferred patterns for HLA-DPB1 (B). The neighbor-joining tree is based Kimura 2-parameter distances of near full length genomic sequences of HLA-C (4.5 kb) or HLA-DPB1 (11 kb). The orangutan (Popy) is used to root the trees. Available sequences from common (Patr) and pygmy (Papa) chimpanzee as well as gorilla (Gogo) are included and indicate the age of the polymorphisms. Green colors are used to indicate lower expression, red indicate higher expression and black is used where no data are available. Expression levels for HLA alleles have been measured or inferred on an allelic lineage-specific basis. For HLA-C the likely ancestral occurrences of changes in Oct1 binding sites and ablation of miR-148a binding sites are indicated. For HLA-DPB1, the relationship between phylogeny and inferred expression levels may be altered for alleles derived from intragenic recombination events. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

unambiguous inference of lineage specific expression levels on multiple lymphocyte subsets [14,15]. Among Bw6-positive lineages, cell surface HLA-B*08:01 expression is significantly higher than that found for B*07:01 or B*35:01 on lymphocytes. Other Bw6 lineages (B*40, B*15 and B*18) have intermediate levels of expression that are not significantly different from either expression extreme. Using Bw4 antibodies, a gradient of cell surface expression is also found across lymphocyte subsets, with high expression inferred for B*57:01 and B*27:05 contrasting with low expression B*44:02 and B*37:01. Intermediate expression levels are inferred for B*51:01 and B*13:02. The significance of differences in expression levels varies for cell types and allelic lineage combinations. Further complexity is evident with the expression gradient for HLA-B in monocytes differing from that found in lymphocytes.

3.2. HLA class II expression

HLA class II molecules are more functionally specialized than class I in as much as their expression occurs on cellular subsets, primarily B-cells, dendritic cells and macrophages involved in antigen presentation. Class II expression is responsive to interferon-γ signaling and shares some expression induction properties with class I, but it is not clear whether all inducers produce equivalent effects or whether a relationship exists between observed basal and maximal expression for a given locus. For class I, most expression measurements have been made on non-induced PBMC cells where basal expression levels can be compared and suggest a 2-state model (basal and induced). In contrast, it is not apparent that a 2-state model will suffice for class II where expression appears to respond to many inductive or repressive factors.

Only recently have systematic efforts been underway to categorize causes of lineage-specific expression differences in class II molecules. There are 13 HLA-DRB1, 6 HLA-DQA1, 5 HLA-DQB1 and multiple numbered HLA-DPA1 and -DPB1 lineages [https://www.ebi.ac.uk/ipd/imgt/hla/]. Strong LD exists across the HLA-DPB1 locus, and prior GWAS had identified SNPs in LD with HLA-DP as influencing HBV outcomes and vaccination in Asian populations [111,112]. In an independent study of Caucasian- and African-American populations, recovery from HBV infection associated with rs9277534 in the HLA-DPB1 3′ UTR, a marker that correlated with HLA-DP cell surface and transcript expression [40]. HLA-DPB1 expression is high (rs9277534G) or low (rs9277534A), epitomized by DPB1*01:01 and DPB1*04:01 lineages, respectively, and HLA-DPB1 phylogeny segregates into two distinct ancient groupings which seems to predate the origins of the hominines (Fig. 2). In contrast to HLA-C, HLA-DPB phylogeny reflects an evolutionary pattern where changes in expression are rare and the tree is divided into high and low-expressing branches in a monophyletic fashion (Fig. 2). Correlation of the alleles within high and low-expressing branches has recently been made with GVHD in HLA-DPB1-mismatched unrelated donor HCT as described below [41,42], and with postpartum control of hepatitis C virus (HCV) VL [113]. HLA-DPB1 expression-associated rs9277534 does not appear to impact any microRNA binding sites which might be implicated as causative of changes in expression. The long tract of LD ensures that multiple potential regulatory polymorphisms segregate with rs9277534 and any or several can contribute to expression differentiation.

In contrast to the tight LD between HLA-DRB1, DQA1 and DQB1, the monomorphic HLA-DRA locus is separated from HLA-DRB1 by a recombination hotspot. Interferon-γ induced activation of HLA-DRA results in stable modification of chromatin states suggesting involvement of epigenetic regulatory mechanisms in expression induction [33]. However, the lack of polymorphism of HLA-DRA means induction will result in universal HLA-DRA expression. An early study of polymorphisms of the proximal core promoter of HLA-DRB1 [34] found no evidence for substitutions that can markedly alter expression, although later studies revealed a polymorphic vitamin D response element that is lineage-specific and found only in HLA-DRB1*15 [35]. Whereas vitamin D induction did not alter DRB1*07 expression, it doubled DRB1*15 expression, suggesting a lineage-specific effect of vitamin D

response among DRB1*15 alleles.

3.3. Haplotype-specific expression

There is mounting evidence for coordinated expression of HLA loci in a haplotype-specific manner. HLA-DR and DQ are controlled by a shared transcriptional enhancer, XL9, which resides within the 130 kb intergenic region between HLA-DRB1 and DQA1 and is involved in chromatin remodeling [114]. Within XL9, polymorphisms in binding sites for IRF4 and CTCF affect HLA-DR-DQ expression: DR2 and DR3 haplotypes (which are associated with risk of systemic lupus erythematosus, SLE) have a 2.5-fold greater expression (measured by transcript abundance and by cell surface detection) than certain DR4 and DR7 haplotypes [105]. Contrasting expression levels for HLA-DRB1, DQA1 and DQB1 are found for DR3 and DR15 haplotypes [36]. DR3 is associated with significantly higher HLA-DRB1 and lower HLA-DQA1 and DQB1 expression than DR15.

Long-range HLA haplotype-associated gene expression patterns provide new information on functional non-coding intergenic polymorphisms. A locally-designed microarray for the MHC which contained probes specific for sequence variants within and between genes in a phase-specific manner, and for splice junction variants, was applied to construct a high-resolution haplotype-specific transcriptional map for HLA-A1-B8-Cw7-DR3, HLA-A3-B7-Cw7-DR15 and HLA-A26-B18-Cw5-DR3-DQ2 [7]. Cross comparison of expression patterns for these 3 haplotypes uncovered haplotype-specific fingerprints for 96 genes, with the most notable differences attributable to zinc finger proteins, suggesting that transcriptional regulation and DNA methylation figure prominently in haplotype-specific expression. The high usage of alternatively splicing also sets the MHC apart from other regions of the genome. One third of haplotype blocks display intergenic transcriptional activity. The data suggest that several well-known mechanisms for genome-wide gene expression are clustered within the MHC.

Multiple SNPs may act in cis to exert haplotype-specific gene expression [115]. eQTL-associated haplotype-specific differential gene expression of HLA-A2-B46-DR9, HLA-A33-B58-DR3 and HLA-A1-B9-DR3 is concentrated in or near HLA-A, HLA-C, C4A and DRB [8]. These data suggest that the combination of markers on an HLA haplotype have implications beyond the maintenance variation within the ARS, and that regionally clustered haplotype-specific regulatory polymorphisms exert effects on HLA expression.

4. Impact of HLA expression on disease

The theory behind HLA expression in disease relates to the presentation of foreign antigens, which may be more robust when larger numbers of HLA molecules are expressed at the cell surface, resulting in improved immune surveillance. By the same token, high HLA expression may increase presentation of the target needed to elicit a strong attack on host tissues leading to autoimmunity or GVHD in transplantation. Numerous GWAS have identified disease associations with loci that have subsequently been shown to be dependent on HLA expression, such as HBV with HLA-DP or HIV with HLA-C.

4.1. HIV-AIDs

Associations between infections and HLA genotypes were difficult to consistently demonstrate until recent decades. Challenges of small cohorts and diversity in infectious strains or environments led to suggestive rather than strong association discovery. The HIV epidemic represented the emergence of a new infectious disease with strong HIV-HLA associations. Alleles leading to disease susceptibility (B*35) [85] or protection (B*57, B*27) [116,117] were defined, and genotypic associations such as heterozygote advantage characterized [85]. Most of the observed associations could be related to the binding repertoire of the protective alleles [118,119]. Expression differences among alleles had not been extensively described so were generally not taken into consideration. However, once lineage specific expression levels were known, they could be tested and were shown to segregate as an independent and strongly significant variable in models incorporating other known effects [16]. Protection from HIV progression can be measured in multiple ways utilizing VL or CD4-positive cell counts and independent HLA factors might be expected to impact differently on different progression metrics. Higher HLA-C expression itself was shown to be protective - the fifth most significant independent HLA variable distinguishing a cohort of HIV viral load controllers (VL less than 2000) from non-controllers (VL $>$ 10000) in European Americans and the third most among African-Americans. HLA-C expression was also shown to have the most significant effect on progression from seroconversion to low ($<$200) CD4 counts.

HLA-B alleles are known to have some of the strongest effects on HIV outcomes and the strong LD between HLA-B and C makes distinction of effects ascribable solely to HLA-C expression difficult. However, the miR-148a gene on chromosome 7 also shows polymorphism resulting in either high or low expression of miR-148a. Only certain HLA-C alleles and no HLA-B alleles contain the binding site for the microRNA. MicroRNA expression polymorphism influenced association strength (in HIV and in Crohn's disease) only for HLA-C alleles that contain the binding site, demonstrating that expression of HLA-C (and not HLA-B linkage) was the determinant factor [25].

4.2. Crohn's disease

Given the increased immune surveillance for HIV that was associated with increased expression, it was considered likely that an increased risk of autoimmune disease would represent a related manifestation of expression. Inflammatory bowel disease (IBD) represents a dysregulated immune response to gut commensals. Two forms of IBD are common in developed countries, Crohn's disease (CD) and ulcerative colitis (UC). The number of genetic locus associations seen for these diseases now exceeds 200 [120] and includes association signals within the HLA region. The availability of large case-control cohorts as well as HLA-C typing enabled [16] to search for associations with HLA-C expression. A strong signal for increased risk of Crohn's disease related to increased HLA-C expression was seen in case-control cohorts. Interestingly UC did not show a risk association with HLA-C expression suggesting a greater role for Th1 responses in CD.

4.3. Hepatitis B

Infection from HBV can result in transient infection and clearance in about 90–95% of cases or viral persistence and chronic infection in the remainder. Multiple environmental and genetic factors have been identified which associate with HBV outcomes. GWAS of Asian patients pointed to HLA-DP variation as being the most important determinant in chronic HBV infection [111]. Among individuals of European and African descent, rs9277534 located in the 3′ UTR of HLA-DPB1 was found to give stronger associations than prior candidates, including individual HLA-DPB1 alleles themselves [40]. Both the levels of HLA-DPB1 transcription and HLA-DP cell surface expression correlate with rs9277534 identity and the previously identified HLA-DPB1 SNP associations do not segregate independently, but are related through linkage. Although the mechanism of expression differentiation is not yet clearly determined, higher cell surface expression of HLA-DP increases the risk of chronic HBV infection. Whether expression alone or repertoire effects play a role remains to be determined. Both HLA-DR and DQ show generally higher levels of polymorphism and more disease associations than HLA-DP.

4.4. Other autoimmune diseases

SLE is an autoimmune disease which results in a loss of tolerance to

self-antigens and production of autoantibodies. The disease is relatively heterogeneous in its manifestations reflecting underlying complexity. Approximately 50 loci have been associated with risk in GWAS studies, including polymorphisms of the XL9 enhancer which result in a 4-fold increased expression of risk alleles [31]. Autoimmune vitiligo, in which autoreactive T-cells destroy melanocytes, is another complex autoimmune condition with approximately 30 candidate loci identified by GWAS, including HLA-DR and -DQ. Three XL9-region SNPs associated with high HLA-DR and -DQ expression and with increased production of IFN-γ and IL1β, are implicated in the disease, suggesting potential *trans* effects in the development of disease [121].

Any role of HLA expression in infectious and autoimmune disease is determined by simple genotypes of affected individuals. In such cases, both alleles contribute to overall expression levels which can be directly related to disease outcomes. Consideration of dominance relationship between alleles can be included if necessary. At a locus with N alleles, $N*(N+1)/2$ genotypes are considered in relation to outcomes. Transplantation medicine creates an additional degree of complexity for expression studies. Here as before, $N*(N+1)/2$ individual genotypes contribute to overall expression levels. However, both donor and recipient genotypes must be considered and the possible interactions between donors and recipients add an additional consideration to determining outcomes. There will be many more combinations, $(N*(N+1)/2)^2$, that need to be considered. As described in the next section, investigation into the role of HLA expression in transplantation has included analysis of both donor and recipient variation.

4.5. Transplantation

Low-expression mismatches may be better tolerated than high-expression mismatches, as observed with the frequent C*03:03/C*03:04 mismatch [29]. HLA-C allotypes display a continuum of expression [16], hence the risk of GVHD would be predicted to depend on the expression level of the patient's mismatched HLA-C [17]. In fact, not only did the risk of acute GVHD increase in parallel with increasing expression level of the patient's mismatched HLA-C, but the level of HLA-C expression can differentiate high-risk ARS mismatches (high-expression residue 116-mismatch) from low-risk ARS mismatches (low-expression residue-mismatch; high-expression residue-match; low-expression residue-match). Similar findings were observed for C1 and C2 KIR epitopes. These data strongly point to synergistic requirements for ARS (residue 116), HLA ligand/KIR receptor interactions, and expression in determining the immunogenicity of HLA-C mismatches in HCT.

For class II, cumulative mismatching at the low-expression HLA-DRB3, DRB4 and DRB5 genes increases risk [30]. Elucidation of HLA-DP expression in HCT came about through a SNP survey of the MHC in which rs2281389 was validated as a marker for acute GVHD [122]. Rs2281389 is a proxy for HLA-DPB1 exon 2 and rs9277534. To test the hypothesis that the immunogenicity of HLA-DPB1 mismatches depends on the expression level of the recipient's mismatched allotype, the *cis*-linkage of rs9277534A/G with exon 2-defined alleles was determined, and the association of rs9277534A with low-expression and rs9277534G with high-expression was confirmed [41]. Significantly higher risks of GVHD were observed with mismatching against a high-expression patient HLA-DPB1 compared to a low-expression allotype. The role of HLA-DP expression in permissive mismatches was recently validated in Japanese transplants [42]. Phylogenetic segregation of rs9277534A and rs9277534G-linked HLA-DPB1 clades strongly correlated with exon 3-3′ UTR sequences, consistent with findings from registry donors [123]. When HLA-DP expression is evaluated alongside exon 2-defined T-cell epitopes [124], GVHD risk is best defined when both expression and epitopes are taken into consideration [41]. These data further support a model of alloreactivity that requires both ARS variation as well as the quantity of expressed HLA to be immunogenic. Finally, in solid organ transplantation, high-expression rs9277534G HLA-DP antigens are associated with a stronger B-cell flow crossmatch compared to low-expression rs9277534A antigens [43]. These data suggest that HLA-DP expression informs the likelihood of developing a positive crossmatch, and have implications for pre-transplant screening to lower risks of graft rejection.

5. The MHC in the NGS era

NGS of the MHC uncovers substantial novel variation in both coding and non-coding regions, and the potential to phase variants across loci and between genes has utility for understanding the biological consequences of MHC variation [5,125].

5.1. Phased variation within and between loci

NGS of coding and non-coding regions provides information on the nature (biallelic SNP; complex substitution; eQTL; CpG site; microRNA binding site), location (promoter, introns, exons, 3′ UTR), and association of variants, near and far. For example, microRNAs within non-coding regions of the MHC are part-and-parcel of local and extended haplotypes [9]. The long-range nature of HLA haplotypes is ethnogeographically defined [77,78], and sequencing of ethnically diverse populations will provide information on inter- and intra-block content and LD. This information has significant implications for phylogeny and understanding the origins of MHC diversity and implications on expression.

How variation is linked to one another across exons is the basic tenant of HLA typing which requires phase to assign alleles. The potential ability to define phase across long amplicons in single molecules with NGS, is of immense benefit [44,123]. Phase is also necessary to understand *cis* and *trans*-effects of regulatory polymorphisms, where *trans* effects involve target genes within the MHC as well as genes elsewhere in the genome. As many as 48% of *trans*-acting SNPs from GWAS map to the MHC [101], and many of these MHC markers influence the expression of non-MHC genes [87,100,101]. Furthermore, lineage-specific *trans*-associations point to a much broader impact of MHC on regulatory pathways [100]. Hence, NGS of non-coding regions of the MHC may provide dense haplotype maps that will enhance the study of the *cis*- and *trans*-effects, bridge multiple mechanistic pathways that converge on MHC resident variation, and potentially facilitate the identification of genes that cause disease.

The vast majority of HLA alleles have been named for their unique combination of substitutions within the ARS. In 2010 in response to the meteoric increase in the number of novel alleles, the WHO Nomenclature Committee introduced a "future-proof" system with limitless capacity to accommodate novel sequences [https://www.ebi.ac.uk/ipd/imgt/hla/nomenclature/index.html]. A growing number of alleles today have full gene sequence data [https://www.ebi.ac.uk/ipd/imgt/hla/; Prof. Steven Marsh, personal communication]. Based on sequences for 3489 HLA-A, 4356 HLA-B and 3111 HLA-C alleles, most class I substitutions are very rare, being found in one individual or a family, and represent point mutations from common alleles [126]. New alleles with variation in exons 2 and 2 were identified in over 3 million volunteer transplant donors worldwide at a frequency of 1.8010e-4 for HLA-A, 2.18e-4 for HLA-B and 2.13e-4 for HLA-C. We can anticipate that substantial variation will be discovered in both coding and non-coding regions; a future challenge is how much of the phenotype (eg, expression) is part-and-parcel of the name, and how best to capture these features.

Phased haplotypes provide a foundation for improving the genotyping tools used to query, analyze and notate variation. These tools include gene feature enumeration [127]; the design of panels of ancestry-informative markers [128,129]; the design of SNP mapping panels informative for the haplotype and block structure of diverse haplotypes [66,130] and tools with greater capacity for imputation of high-resolution alleles [131]. A complete haplotype map of non-coding and coding regions will greatly enhance SNP array design through the inclusion of haplotype-specific regulatory variants, and may be descriptive of the extraordinary long-range LD that is a hallmark of the MHC. How long is long enough? If one considers that phase across 4 Mb from HLA-A through HLA-DR carries clinical significance in HCT [132], with even longer stretches in selected ancestral haplotypes [77,78], then phased variants across 7 Mb or beyond can be envisioned.

Content of phased HLA haplotypes may help to explain why in many cases, the true causative gene cannot be readily identified from SNP associations [101]. SNPs could be causal themselves, or serve as proxies for other functional variation. Within the MHC, the causal gene may reside at great distances from the original SNP marker due to the long-range LD [133]. Furthermore, ancestry-informative genome-wide SNPs have been used to study population differences [129]; many HLA polymorphisms are also ancestry-informative markers (IMGT/HLA), and complicate the analysis of regulatory variants which may depend on ethnic background of the study population. Alleles with shared ARS variation may have different background haplotypes and evolutionary histories in different ancestries [83,133,134]. Delineation of phased haplotypes of non-coding and coding polymorphisms may help to dissociate the impact of expression from polymorphisms of the ARS or epitopes involved in ligand-KIR receptor binding, in disease risk.

5.2. Correlation of structure with function

HLA expression is a complex phenotype, and continued efforts to comprehensively identify the full spectrum of regulation of HLA expression are needed [11]. Currently identified variants are predicted to account for less than 50% of variation in expression across HLA-C allotypes [26]. For HLA-DR-DQ, additional mechanisms such as *trans*-factors contribute to a phenotype [31,114,135,136]. Furthermore, the influence of the background variation on long [7,8] and short [23,31] haplotypes remains to be defined for diverse blocks. The feasibility of long-range phasing across the MHC therefore will greatly enhance the study of haplotype-associated phenotypic differences.

5.3. Clinically actionable MHC expression variation

To have maximal benefit, use of HLA expression and non-coding variation in clinical practice requires knowledge of what makes a person's MHC personal. An MHC haplotype library inclusive of regulatory expression variants together with the technology for long-range haplotype phasing may provide the necessary resolution to match haplotypes of transplant patients and donors [78,132], identify patients whose HLA-B-C haplotypes predict risky or protective combinations [23], and avoid risky HLA mismatches [17,41]. Current data support the need to define ARS variation (HLA-C residue 116; KIR ligand epitopes; HLA-DP epitopes) along with its linked regulatory polymorphisms to more accurately determine the immunogenicity of HLA mismatches or risk of positive crossmatching [43].

In transplantation, consideration of both the patient's and the donor's haplotypes increases the complexity of matching, but the benefits of matching beyond the ARS are immense and quantifiable. In a landmark study, NGS-matching of transplants beyond exons 2 and 3 of class I provides clinically important information on transplantation survivorship [137]. The potential for strategies to afford this level of matching are available, and holds promise that haplotype-matching may well be within reach.

6. Conclusion

The development of powerful laboratory methods to define the extent of sequence diversity, the linkage of variants over long distances of the MHC, and the implications of structure on function, remain the cornerstone of investigation of the MHC.

Acknowledgements

This project has been funded in whole or in part with federal funds from the Frederick National Laboratory for Cancer Research, under contract no. HHSN261200800001E. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US government. This research was supported in part by the intramural research program of the NIH, Frederick National Laboratory Center for Cancer Research.

Dr. Petersdorf is supported by grants AI069197, CA100019, CA218285 and CA231838 from the National Institutes of Health, USA.

Declaration of interest

The authors declare no conflicts of interest.

Contributors

The authors declare they each materially participated in the preparation of the article, and approve the final article.

References

[1] Wellcome Trust Case Control Consortium, Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls, Nature 447 (2007) 661–678.
[2] J. Fellay, K.V. Shianna, D. Ge, S. Colombo, B. Ledergerber, M. Weale, K. Zhang, C. Gumbs, A. Castagna, A. Cossarizza, A. Cozzi-Lepri, A. De Luca, P. Easterbrook, P. Francioli, S. Mallal, J. Martinez-Picado, J.M. Miro, N. Obel, J.P. Smith, J. Wyniger, P. Descombes, S.E. Antonarakis, N.L. Letvin, A.J. McMichael, B.F. Haynes, A. Telenti, D.B. Goldstein, A whole-genome association study of major determinants for host control of HIV-1, Science 317 (2007) 944–947.
[3] P.J. Bjorkman, M.A. Saper, B. Samraoui, W.S. Bennett, J.L. Strominger, D.C. Wiley, Structure of the human class I histocompatibility antigen, HLA-A2, Nature 329 (1987) 506–512.
[4] J.H. Brown, T.S. Jardetzky, J.C. Gorga, L.J. Stern, R.G. Urban, J.L. Strominger, D.C. Wiley, Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1, Nature 364 (1993) 33–39.
[5] T. Shiina, K. Hosomichi, H. Inoko, J.K. Kulski, The HLA genomic loci map: expression, interaction, diversity and disease, J. Hum. Genet. 54 (2009) 15–39.
[6] C. Anasetti, D. Amos, P.G. Beatty, F.R. Appelbaum, W. Bensinger, C.D. Buckner, R. Clift, K. Doney, P.J. Martin, E. Mickelson, Effect of HLA compatibility on engraftment of bone marrow transplants in patients with leukemia or lymphoma, N. Engl. J. Med. 320 (1989) 197–204.
[7] C. Vandiedonck, M.S. Taylor, H.E. Lockstone, K. Plant, J.M. Taylor, C. Durrant, J. Broxholme, B.P. Fairfax, J.C. Knight, Pervasive haplotypic variation in the spliceo-transcriptome of the human major histocompatibility complex, Genome Res. 21 (2011) 1042–1054.
[8] T.H. Lam, M. Shen, M.Z. Tay, E.C. Ren, Unique Allelic eQTL Clusters in Human MHC Haplotypes, G3 (Bethesda) 7 (2017) 2595–2604.
[9] P.M. Clark, N. Chitnis, M. Shieh, M. Kamoun, F.B. Johnson, D. Monos, Novel and Haplotype Specific MicroRNAs Encoded by the Major Histocompatibility Complex, Sci. Rep. 8 (2018) 3832.
[10] N. Cereb, S.Y. Yang, The regulatory complex of HLA class I promoters exhibits locus-specific conservation with limited allelic variation, J. Immunol. 152 (1994) 3873–3883.
[11] V. Ramsuran, S. Kulkarni, C. O'hUigin, Y. Yuki, D.G. Augusto, X. Gao, M. Carrington, Epigenetic regulation of differential HLA-A allelic expression levels, Hum. Mol. Genet. 24 (2015) 4268–4275.

[12] C. Dellgren, V.A.C. Ekwelum, M. Ormhøj, N. Pallesen, J. Knudsen, J.O. Nehlin, T. Barington, Low Constitutive Cell Surface Expression of HLA-B Is Caused by a Posttranslational Mechanism Involving Glu180 and Arg239, J. Immunol. 197 (2016) 4807–4816.
[13] M. L'Haridon, P. Paul, J.G. Xerri, H. Dastot, C. Dolliger, M. Schmid, N. de Angelis, L. Grollet, F. Sigaux, L. Degos, C. Gazin, Transcriptional regulation of the MHC class I HLA-A11 promoter by the zinc finger protein ZFX, Nucleic Acids Res. 24 (1996) 1928–1935.
[14] J.E. Boudreau, T.J. Mulrooney, J.-B. Le Luduec, E. Barker, K.C. Hsu, KIR3DL1 and HLA-B Density and Binding Calibrate NK Education and Response to HIV, J. Immunol. 196 (2016) 3398–3410.
[15] B. Yarzabek, A.J. Zaitouna, E. Olson, G.N. Silva, J. Geng, A. Geretz, R. Thomas, S. Krishnakumar, D.S. Ramon, M. Raghavan, Variations in HLA-B cell surface expression, half-life and extracellular antigen receptivity, Elife 7 (2018), https://doi.org/10.7554/eLife.34961.
[16] R. Apps, Y. Qi, J.M. Carlson, H. Chen, X. Gao, R. Thomas, Y. Yuki, G.Q. Del Prete, P. Goulder, Z.L. Brumme, C.J. Brumme, M. John, S. Mallal, G. Nelson, R. Bosch, D. Heckerman, J.L. Stein, K.A. Soderberg, M.A. Moody, T.N. Denny, X. Zeng, J. Fang, A. Moffett, J.D. Lifson, J.J. Goedert, S. Buchbinder, G.D. Kirk, J. Fellay, P. McLaren, S.G. Deeks, F. Pereyra, B. Walker, N.L. Michael, A. Weintrob, S. Wolinsky, W. Liao, M. Carrington, Influence of HLA-C expression level on HIV control, Science 340 (2013) 87–91.
[17] E.W. Petersdorf, T.A. Gooley, M. Malkki, A.P. Bacigalupo, A. Cesbron, E. Du Toit, G. Ehninger, T. Egeland, G.F. Fischer, T. Gervais, M.D. Haagenson, M.M. Horowitz, K. Hsu, P. Jindra, A. Madrigal, M. Oudshoorn, O. Ringdén, M.L. Schroeder, S.R. Spellman, J.-M. Tiercy, A. Velardi, C.S. Witt, C. O'hUigin, R. Apps, M. Carrington, International Histocompatibility Working Group in Hematopoietic Cell Transplantation, HLA-C expression levels define permissible mismatches in hematopoietic cell transplantation, Blood 124 (2014) 3996–4003.
[18] C. O'hUigin, S. Kulkarni, Y. Xu, Z. Deng, J. Kidd, K. Kidd, X. Gao, M. Carrington, The molecular origin and consequences of escape from miRNA regulation by HLA-C alleles, Am. J. Hum. Genet. 89 (2011) 424–431.
[19] J.A. McCutcheon, J. Gumperz, K.D. Smith, C.T. Lutz, P. Parham, Low HLA-C expression at cell surfaces correlates with increased turnover of heavy chain mRNA, J. Exp. Med. 181 (1995) 2085–2095.
[20] R. Thomas, R. Apps, Y. Qi, X. Gao, V. Male, C. O'hUigin, G. O'Connor, D. Ge, J. Fellay, J.N. Martin, J. Margolick, J.J. Goedert, S. Buchbinder, G.D. Kirk, M.P. Martin, A. Telenti, S.G. Deeks, B.D. Walker, D. Goldstein, D.W. McVicar, A. Moffett, M. Carrington, HLA-C cell surface expression and control of HIV/AIDS correlate with a variant upstream of *HLA-C*, Nat. Genet. 41 (2009) 1290–1294.
[21] E.H. van den Bogaard, H.J. Tijssen, D. Rodijk-Olthuis, K.P. van Houwelingen, M.J. Coenen, M. Marget, J. Schalkwijk, I. Joosten, Cell Surface Expression of HLA-Cw6 by Human Epidermal Keratinocytes: Positive Regulation by Cytokines, Lack of Correlation to a Variant Upstream of HLA-C, J. Invest. Dermatol. 136 (2016) 1903–1906.
[22] G.A. Hoff, J.C. Fischer, K. Hsu, S. Cooley, J.S. Miller, T. Wang, M. Haagenson, S. Spellman, S.J. Lee, M. Uhrberg, J.M. Venstrom, M.R. Verneris, Recipient HLA-C Haplotypes and microRNA 148a/b Binding Sites Have No Impact on Allogeneic Hematopoietic Cell Transplantation Outcomes, Biol. Blood Marrow Transplant. 23 (2017) 153–160.
[23] F. Bettens, L. Brunet, J.-M. Tiercy, High-allelic variability in HLA-C mRNA expression: association with HLA-extended haplotypes, Genes Immun. 15 (2014) 176–181.
[24] S. Kulkarni, R. Savan, Y. Qi, X. Gao, Y. Yuki, S.E. Bass, M.P. Martin, P. Hunt, S.G. Deeks, A. Telenti, F. Pereyra, D. Goldstein, S. Wolinsky, B. Walker, H.A. Young, M. Carrington, Differential microRNA regulation of HLA-C expression and its association with HIV control, Nature. 472 (2011) 495–498.
[25] S. Kulkarni, Y. Qi, C. O'hUigin, F. Pereyra, V. Ramsuran, P. McLaren, J. Fellay, G. Nelson, H. Chen, W. Liao, S. Bass, R. Apps, X. Gao, Y. Yuki, A. Lied, A. Ganesan, P.W. Hunt, S.G. Deeks, S. Wolinsky, B.D. Walker, M. Carrington, Genetic interplay between HLA-C and MIR148A in HIV control and Crohn disease, Proc. Natl. Acad. Sci. U.S.A. 110 (2013) 20705–20710.
[26] N. Vince, H. Li, V. Ramsuran, V. Naranbhai, F.-M. Duh, B.P. Fairfax, B. Saleh, J.C. Knight, S.K. Anderson, M. Carrington, HLA-C Level Is Regulated by a Polymorphic Oct1 Binding Site in the HLA-C Promoter Region, Am. J. Hum. Genet. 99 (2016) 1353–1358.
[27] V. Ramsuran, P.G. Hernández-Sanchez, C. O'hUigin, G. Sharma, N. Spence, D.G. Augusto, X. Gao, C.A. García-Sepúlveda, G. Kaur, N.K. Mehra, M. Carrington, Sequence and Phylogenetic Analysis of the Untranslated Promoter Regions for HLA Class I Genes, J. Immunol. 198 (2017) 2320–2329.
[28] H. Li, M.A. Ivarsson, V.E. Walker-Sperling, J. Subleski, J.K. Johnson, P.W. Wright, M. Carrington, N.K. Björkström, D.W. McVicar, S.K. Anderson, Identification of an elaborate NK-specific system regulating HLA-C expression, PLoS Genet. 14 (2018) e1007163 .
[29] M.A. Fernandez-Viña, T. Wang, S.J. Lee, M. Haagenson, M. Aljurf, M. Askar, M. Battiwalla, L.-A. Baxter-Lowe, J. Gajewski, A.A. Jakubowski, S. Marino, M. Oudshoorn, S.G.E. Marsh, E.W. Petersdorf, K. Schultz, E.V. Turner, E.K. Waller, A. Woolfrey, J. Umejiego, S.R. Spellman, M. Setterholm, Identification of a permissible HLA mismatch in hematopoietic stem cell transplantation, Blood 123 (2014) 1270–1278.
[30] M.A. Fernández-Viña, J.P. Klein, M. Haagenson, S.R. Spellman, C. Anasetti, H. Noreen, L.A. Baxter-Lowe, P. Cano, N. Flomenberg, D.L. Confer, M.M. Horowitz, M. Oudshoorn, E.W. Petersdorf, M. Setterholm, R. Champlin, S.J. Lee, M. de Lima, Multiple mismatches at the low expression HLA loci DP, DQ, and DRB3/4/5 associate with adverse outcomes in hematopoietic stem cell transplantation, Blood 121 (2013) 4603–4610.
[31] P. Raj, E. Rai, R. Song, S. Khan, B.E. Wakeland, K. Viswanathan, C. Arana, C. Liang, B. Zhang, I. Dozmorov, F. Carr-Johnson, M. Mitrovic, G.B. Wiley, J.A. Kelly, B.R. Lauwerys, N.J. Olsen, C. Cotsapas, C.K. Garcia, C.A. Wise, J.B. Harley, S.K. Nath, J.A. James, C.O. Jacob, B.P. Tsao, C. Pasare, D.R. Karp, Q.Z. Li, P.M. Gaffney, E.K. Wakeland, Regulatory polymorphisms modulate the expression of HLA class II molecules and promote autoimmunity, Elife 5 (2016), https://doi.org/10.7554/eLife.12089.
[32] J.M. Boss, Regulation of transcription of MHC class II genes, Curr. Opin. Immunol. 9 (1997) 107–113.
[33] N.M. Choi, J.M. Boss, Multiple histone methyl and acetyltransferase complex components bind the HLA-DRA gene, PLoS One 7 (2012) e37554 .
[34] A. Kruger, P. Quack, P.M. Schneider, C. Rittner, T. Höhler, Sequence analysis of the DRB1 promoter reveals limited polymorphism with no influence on gene expression, Genes Immun. 2 (2001) 211–215.
[35] S.V. Ramagopalan, N.J. Maugeri, L. Handunnetthi, M.R. Lincoln, S.-M. Orton, D.A. Dyment, G.C. Deluca, B.M. Herrera, M.J. Chao, A.D. Sadovnick, G.C. Ebers, J.C. Knight, Expression of the multiple sclerosis-associated MHC class II Allele HLA-DRB1*1501 is regulated by vitamin D, PLoS Genet. 5 (2009) e1000369 .
[36] R.C. Pelikan, J.A. Kelly, Y. Fu, C.A. Lareau, K.L. Tessneer, G.B. Wiley, M.M. Wiley, S.B. Glenn, J.B. Harley, J.M. Guthridge, J.A. James, M.J. Aryee, C. Montgomery, P.M. Gaffney, Enhancer histone-QTLs are enriched on autoimmune risk haplotypes and influence gene expression within chromatin networks, Nat. Commun. 9 (2018) 2905.
[37] G.T. Nepom, J. Chung, K.A. West, Differential expression of HLA-DQB1 alleles in a heterozygous cell line, Immunogenetics 42 (1995) 143–148.
[38] J.S. Beaty, K.A. West, G.T. Nepom, Functional effects of a natural polymorphism in the transcriptional regulatory sequence of HLA-DQB1, Mol. Cell. Biol. 15 (1995) 4771–4782.
[39] B. Ferstl, T. Zacher, B. Lauer, N. Blagitko-Dorfs, A. Carl, R. Wassmuth, Allele-specific quantification of HLA-DQB1 gene expression by real-time reverse transcriptase-polymerase chain reaction, Genes Immun. 5 (2004) 405–416.
[40] R. Thomas, C.L. Thio, R. Apps, Y. Qi, X. Gao, D. Marti, J.L. Stein, K.A. Soderberg, M.A. Moody, J.J. Goedert, G.D. Kirk, W.K. Hoots, S. Wolinsky, M. Carrington, A novel variant marking HLA-DP expression levels predicts recovery from hepatitis B virus infection, J. Virol. 86 (2012) 6979–6985.
[41] E.W. Petersdorf, M. Malkki, C. O'hUigin, M. Carrington, T. Gooley, M.D. Haagenson, M.M. Horowitz, S.R. Spellman, T. Wang, P. Stevenson, High HLA-DP Expression and Graft-versus-Host Disease, N. Engl. J. Med. 373 (2015) 599–609.
[42] S. Morishima, T. Shiina, S. Suzuki, S. Ogawa, A. Sato-Otsubo, K. Kashiwase, F. Azuma, T. Yabe, M. Satake, S. Kato, Y. Kodera, T. Sasazuki, Y. Morishima, Japan Marrow Donor Program, Evolutionary basis of HLA-DPB1 alleles affects acute GVHD in unrelated donor stem cell transplantation, Blood 131 (2018) 808–817.
[43] N.N. Soe, Y. Yin, N.M. Valenzuela, Y.-P. Jin, M. Rossetti, E.F. Reed, Q.J. Zhang, OR3 HLA-DPB1 single nucleotide polymorphism determines DP molecule expression and B lymphocyte crossmatch results, Human Immunology. 78 (2017) 3.
[44] N.P. Mayor, J. Robinson, A.J.M. McWhinnie, S. Ranade, K. Eng, W. Midwinter, W.P. Bultitude, C.-S. Chin, B. Bowman, P. Marks, H. Braund, J.A. Madrigal, K. Latham, S.G.E. Marsh, HLA Typing for the Next Generation, PLoS One 10 (2015) e0127153 .
[45] Y. Satta, C. O'hUigin, N. Takahata, J. Klein, The synonymous substitution rate of the major histocompatibility complex loci in primates, Proc. Natl. Acad. Sci. U.S.A. 90 (1993) 7480–7484.
[46] A.L. Hughes, M. Nei, Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection, Nature 335 (1988) 167–170.
[47] A.L. Hughes, M. Nei, Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection, Proc. Natl. Acad. Sci. U.S.A. 86 (1989) 958–962.
[48] B. Arden, J. Klein, Biochemical comparison of major histocompatibility complex molecules from different subspecies of Mus musculus: evidence for trans-specific evolution of alleles, Proc. Natl. Acad. Sci. U.S.A. 79 (1982) 2342–2346.
[49] N. Takahata, A simple genealogical structure of strongly balanced allelic lines and trans-species evolution of polymorphism, Proc. Natl. Acad. Sci. U.S.A. 87 (1990) 2419–2423.
[50] N. Takahata, Allelic genealogy and human evolution, Mol. Biol. Evol. 10 (1993) 2–22.
[51] C. O'hUigin, Y. Satta, A. Hausmann, R.L. Dawkins, J. Klein, The implications of intergenic polymorphism for major histocompatibility complex evolution, Genetics. 156 (2000) 867–877.
[52] J.J. Cai, J.M. Macpherson, G. Sella, D.A. Petrov, Pervasive hitchhiking at coding and regulatory sites in humans, PLoS Genet. 5 (2009) e1000336 .
[53] S. Das, P.E. Stuart, J. Ding, T. Tejasvi, Y. Li, L.C. Tsoi, V. Chandran, J. Fischer, C. Helms, K.C. Duffin, J.J. Voorhees, A.M. Bowcock, G.G. Krueger, G.M. Lathrop, R.P. Nair, P. Rahman, G.R. Abecasis, D. Gladman, J.T. Elder, Fine mapping of eight psoriasis susceptibility loci, Eur. J. Hum. Genet. 23 (2015) 844–853.
[54] G. Thomson, L.F. Barcellos, A.M. Valdes, Searching for additional disease loci in a genomic region, Adv. Genet. 60 (2008) 253–292.
[55] A. Brodie, J.R. Azaria, Y. Ofran, How far from the SNP may the causative genes be?

Nucleic Acids Res. 44 (2016) 6046–6054.
[56] S.J. Mack, J. Udell, F. Cohen, K. Osoegawa, S.K. Hawbecker, D.A. Noonan, M.B. Ladner, D. Goodridge, E.A. Trachtenberg, J.R. Oksenberg, H.A. Erlich, Correction: High resolution HLA analysis reveals independent class I haplotypes and amino-acid motifs protective for multiple sclerosis, Genes Immun. (2018), https://doi.org/10.1038/s41435-018-0037-9.
[57] A. Horowitz, Z. Djaoud, N. Nemat-Gorgani, J. Blokhuis, H.G. Hilton, V. Béziat, K.J. Malmberg, P.J. Norman, L.A. Guethlein, P. Parham, Class, Ihaplotypes form two schools that educate NK cells in different ways, Sci Immunol. 1 (3) (2016), https://doi.org/10.1126/sciimmunol.aag1672.
[58] J. Trowsdale, The MHC, disease and selection, Immunol. Lett. 137 (2011) 1–8.
[59] R. Ceppellini, E.S. Curtoni, P.L. Mattiuz, V. Miggiano, G. Scudeller, A. Serra, Histocompatibility Testing, Munksgaard, Copenhagen, 1967, pp. 149–187.
[60] R.L. Dawkins, F.T. Christiansen, P.H. Kay, M. Garlepp, J. McCluskey, P.N. Hollingsworth, P.J. Zilko, Disease associations with complotypes, supratypes and haplotypes, Immunol. Rev. 70 (1983) 5–22.
[61] P. Price, C. Witt, R. Allcock, D. Sayer, M. Garlepp, C.C. Kok, M. French, S. Mallal, F. Christiansen, The genetic basis for the association of the 8.1 ancestral haplotype (A1, B8, DR3) with multiple immunopathological diseases, Immunol. Rev. 167 (1999) 257–274.
[62] E.J. Yunis, C.E. Larsen, M. Fernandez-Viña, Z.L. Awdeh, T. Romero, J.A. Hansen, C.A. Alper, Inheritable variable sizes of DNA stretches in the human MHC: conserved extended haplotypes and their fragments or blocks, Tissue Antigens. 62 (2003) 1–20.
[63] C.A. Alper, C.E. Larsen, D.P. Dubey, Z.L. Awdeh, D.A. Fici, E.J. Yunis, The haplotype structure of the human major histocompatibility complex, Hum. Immunol. 67 (2006) 73–84.
[64] S.B. Gabriel, S.F. Schaffner, H. Nguyen, J.M. Moore, J. Roy, B. Blumenstiel, J. Higgins, M. DeFelice, A. Lochner, M. Faggart, S.N. Liu-Cordero, C. Rotimi, A. Adeyemo, R. Cooper, R. Ward, E.S. Lander, M.J. Daly, D. Altshuler, The structure of haplotype blocks in the human genome, Science. 296 (2002) 2225–2229.
[65] M.M. Miretti, E.C. Walsh, X. Ke, M. Delgado, M. Griffiths, S. Hunt, J. Morrison, P. Whittaker, E.S. Lander, L.R. Cardon, D.R. Bentley, J.D. Rioux, S. Beck, P. Deloukas, 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 (2005) 634–646.
[66] P.I.W. de Bakker, G. McVean, P.C. Sabeti, M.M. Miretti, T. Green, J. Marchini, X. Ke, A.J. Monsuur, P. Whittaker, M. Delgado, J. Morrison, A. Richardson, E.C. Walsh, X. Gao, L. Galver, J. Hart, D.A. Hafler, M. Pericak-Vance, J.A. Todd, M.J. Daly, J. Trowsdale, C. Wijmenga, T.J. Vyse, S. Beck, S.S. Murray, M. Carrington, S. Gregory, P. Deloukas, J.D. Rioux, A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC, Nat. Genet. 38 (2006) 1166–1172.
[67] A.J. Jeffreys, L. Kauppi, R. Neumann, Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex, Nat. Genet. 29 (2001) 217–222.
[68] A.J. Jeffreys, C.A. May, Intense and highly localized gene conversion activity in human meiotic crossover hot spots, Nat. Genet. 36 (2004) 151–156.
[69] L. Kauppi, A. Sajantila, A.J. Jeffreys, Recombination hotspots rather than population history dominate linkage disequilibrium in the MHC class II region, Hum. Mol. Genet. 12 (2003) 33–40.
[70] S.E. Ptak, A.D. Roeder, M. Stephens, Y. Gilad, S. Pääbo, M. Przeworski, Absence of the TAP2 human recombination hotspot in chimpanzees, PLoS Biol. 2 (2004) e155 .
[71] A.J. Jeffreys, R. Neumann, The rise and fall of a human recombination hot spot, Nat. Genet. 41 (2009) 625–629.
[72] Complete sequence and gene map of a human major histocompatibility complex. The MHC sequencing consortium, Nature. 401 (1999) 921–923. https://doi.org/10.1038/44853.
[73] C.A. Stewart, R. Horton, R.J.N. Allcock, J.L. Ashurst, A.M. Atrazhev, P. Coggill, I. Dunham, S. Forbes, K. Halls, J.M.M. Howson, S.J. Humphray, S. Hunt, A.J. Mungall, K. Osoegawa, S. Palmer, A.N. Roberts, J. Rogers, S. Sims, Y. Wang, L.G. Wilming, J.F. Elliott, P.J. de Jong, S. Sawcer, J.A. Todd, J. Trowsdale, S. Beck, Complete MHC haplotype sequencing for common disease gene mapping, Genome Res. 14 (2004) 1176–1187.
[74] R. Horton, R. Gibson, P. Coggill, M. Miretti, R.J. Allcock, J. Almeida, S. Forbes, J.G.R. Gilbert, K. Halls, J.L. Harrow, E. Hart, K. Howe, D.K. Jackson, S. Palmer, A.N. Roberts, S. Sims, C.A. Stewart, J.A. Traherne, S. Trevanion, L. Wilming, J. Rogers, P.J. de Jong, J.F. Elliott, S. Sawcer, J.A. Todd, J. Trowsdale, S. Beck, Variation analysis and gene annotation of eight MHC haplotypes: the MHC Haplotype Project, Immunogenetics 60 (2008) 1–18.
[75] J.A. Traherne, R. Horton, A.N. Roberts, M.M. Miretti, M.E. Hurles, C.A. Stewart, J.L. Ashurst, A.M. Atrazhev, P. Coggill, S. Palmer, J. Almeida, S. Sims, L.G. Wilming, J. Rogers, P.J. de Jong, M. Carrington, J.F. Elliott, S. Sawcer, J.A. Todd, J. Trowsdale, S. Beck, Genetic analysis of completely sequenced disease-associated MHC haplotypes identifies shuffling of segments in recent human history, PLoS Genet. 2 (2006) e9 .
[76] International HapMap Consortium, The International HapMap Project, Nature 426 (2003) 789–796.
[77] W.P. Smith, Q. Vu, S.S. Li, J.A. Hansen, L.P. Zhao, D.E. Geraghty, Toward understanding MHC disease associations: partial resequencing of 46 distinct HLA haplotypes, Genomics 87 (2006) 561–571.
[78] S. Morishima, S. Ogawa, A. Matsubara, T. Kawase, Y. Nannya, K. Kashiwase, M. Satake, H. Saji, H. Inoko, S. Kato, Y. Kodera, T. Sasazuki, Y. Morishima, Japan Marrow Donor Program, Impact of highly conserved HLA haplotype on acute graft-versus-host disease, Blood 115 (2010) 4664–4670.
[79] R.J.N. Allcock, A.M. Atrazhev, S. Beck, P.J. de Jong, J.F. Elliott, S. Forbes, K. Halls, R. Horton, K. Osoegawa, J. Rogers, S. Sawcer, J.A. Todd, J. Trowsdale, Y. Wang, S. Williams, The MHC haplotype project: a resource for HLA-linked association studies, Tissue Antigens. 59 (2002) 520–521.
[80] C. Gabriel, M. Danzer, C. Hackl, G. Kopal, P. Hufnagl, K. Hofer, H. Polin, S. Stabentheiner, J. Pröll, Rapid high-throughput human leukocyte antigen typing by massively parallel pyrosequencing for high-resolution allele identification, Hum. Immunol. 70 (2009) 960–964.
[81] E.E. Baschal, T.A. Aly, J.M. Jasinski, A.K. Steck, J.A. Noble, H.A. Erlich, G.S. Eisenbarth, Type 1 Diabetes Genetics Consortium, Defining multiple common "completely" conserved major histocompatibility complex SNP haplotypes, Clin. Immunol. 132 (2009) 203–214.
[82] E.C. Walsh, K.A. Mather, S.F. Schaffner, L. Farwell, M.J. Daly, N. Patterson, M. Cullen, M. Carrington, T.L. Bugawan, H. Erlich, J. Campbell, J. Barrett, K. Miller, G. Thomson, E.S. Lander, J.D. Rioux, An integrated haplotype map of the human major histocompatibility complex, Am. J. Hum. Genet. 73 (2003) 580–590.
[83] P.J. Norman, S.J. Norberg, L.A. Guethlein, N. Nemat-Gorgani, T. Royce, E.E. Wroblewski, T. Dunn, T. Mann, C. Alicata, J.A. Hollenbach, W. Chang, M. Shults Won, K.L. Gunderson, L. Abi-Rached, M. Ronaghi, P. Parham, Sequences of 95 human MHC haplotypes reveal extreme coding variation in genes other than highly polymorphic HLA class I and II, Genome Res. 27 (2017) 813–823.
[84] J.C. Murray, K.H. Buetow, J.L. Weber, S. Ludwigsen, T. Scherpbier-Heddema, F. Manion, J. Quillen, V.C. Sheffield, S. Sunden, G.M. Duyk, A comprehensive human linkage map with centimorgan density. Cooperative Human Linkage Center (CHLC), Science 265 (1994) 2049–2054.
[85] M. Carrington, Recombination within the human MHC, Immunol. Rev. 167 (1999) 245–256.
[86] B.E. Stranger, E.A. Stahl, T. Raj, Progress and promise of genome-wide association studies for human complex trait genetics, Genetics 187 (2011) 367–383.
[87] M.T. Maurano, R. Humbert, E. Rynes, R.E. Thurman, E. Haugen, H. Wang, A.P. Reynolds, R. Sandstrom, H. Qu, J. Brody, A. Shafer, F. Neri, K. Lee, T. Kutyavin, S. Stehling-Sun, A.K. Johnson, T.K. Canfield, E. Giste, M. Diegel, D. Bates, R.S. Hansen, S. Neph, P.J. Sabo, S. Heimfeld, A. Raubitschek, S. Ziegler, C. Cotsapas, N. Sotoodehnia, I. Glass, S.R. Sunyaev, R. Kaul, J.A. Stamatoyannopoulos, Systematic localization of common disease-associated variation in regulatory DNA, Science 337 (2012) 1190–1195.
[88] L.A. Hindorff, P. Sethupathy, H.A. Junkins, E.M. Ramos, J.P. Mehta, F.S. Collins, T.A. Manolio, Potential etiologic and functional implications of genome-wide association loci for human diseases and traits, Proc. Natl. Acad. Sci. U.S.A. 106 (2009) 9362–9367.
[89] J. Trowsdale, J.C. Knight, Major histocompatibility complex genomics and human disease, Annu Rev Genomics Hum Genet. 14 (2013) 301–323.
[90] ENCODE Project Consortium, E. Birney, J.A. Stamatoyannopoulos, A. Dutta, R. Guigó, T.R. Gingeras, E.H. Margulies, Z. Weng, M. Snyder, E.T. Dermitzakis, R.E. Thurman, M.S. Kuehn, C.M. Taylor, S. Neph, C.M. Koch, S. Asthana, A. Malhotra, I. Adzhubei, J.A. Greenbaum, R.M. Andrews, P. Flicek, P.J. Boyle, H. Cao, N.P. Carter, G.K. Clelland, S. Davis, N. Day, P. Dhami, S.C. Dillon, M.O. Dorschner, H. Fiegler, P.G. Giresi, J. Goldy, M. Hawrylycz, A. Haydock, R. Humbert, K.D. James, B.E. Johnson, E.M. Johnston, T.T. Frum, E.R. Rosenzweig, N. Karnani, K. Lee, G.C. Lefebvre, P.A. Navas, F. Neri, S.C.J. Parker, P.J. Sabo, R. Sandstrom, A. Shafer, D. Vetrie, M. Weaver, S. Wilcox, M. Yu, F.S. Collins, J. Dekker, J.D. Lieb, T.D. Tullius, G.E. Crawford, S. Sunyaev, W.S. Noble, I. Dunham, F. Denoeud, A. Reymond, P. Kapranov, J. Rozowsky, D. Zheng, R. Castelo, A. Frankish, J. Harrow, S. Ghosh, A. Sandelin, I.L. Hofacker, R. Baertsch, D. Keefe, S. Dike, J. Cheng, H.A. Hirsch, E.A. Sekinger, J. Lagarde, J.F. Abril, A. Shahab, C. Flamm, C. Fried, J. Hackermüller, J. Hertel, M. Lindemeyer, K. Missal, A. Tanzer, S. Washietl, J. Korbel, O. Emanuelsson, J.S. Pedersen, N. Holroyd, R. Taylor, D. Swarbreck, N. Matthews, M.C. Dickson, D.J. Thomas, M.T. Weirauch, J. Gilbert, J. Drenkow, I. Bell, X. Zhao, K.G. Srinivasan, W.-K. Sung, H.S. Ooi, K.P. Chiu, S. Foissac, T. Alioto, M. Brent, L. Pachter, M.L. Tress, A. Valencia, S.W. Choo, C.Y. Choo, C. Ucla, C. Manzano, C. Wyss, E. Cheung, T.G. Clark, J.B. Brown, M. Ganesh, S. Patel, H. Tammana, J. Chrast, C.N. Henrichsen, C. Kai, J. Kawai, U. Nagalakshmi, J. Wu, Z. Lian, J. Lian, P. Newburger, X. Zhang, P. Bickel, J.S. Mattick, P. Carninci, Y. Hayashizaki, S. Weissman, T. Hubbard, R.M. Myers, J. Rogers, P.F. Stadler, T.M. Lowe, C.-L. Wei, Y. Ruan, K. Struhl, M. Gerstein, S.E. Antonarakis, Y. Fu, E.D. Green, U. Karaöz, A. Siepel, J. Taylor, L.A. Liefer, K.A. Wetterstrand, P.J. Good, E.A. Feingold, M.S. Guyer, G.M. Cooper, G. Asimenos, C.N. Dewey, M. Hou, S. Nikolaev, J.I. Montoya-Burgos, A. Löytynoja, S. Whelan, F. Pardi, T. Massingham, H. Huang, N.R. Zhang, I. Holmes, J.C. Mullikin, A. Ureta-Vidal, B. Paten, M. Seringhaus, D. Church, K. Rosenbloom, W.J. Kent, E.A. Stone, NISC Comparative Sequencing Program, Baylor College of Medicine Human Genome Sequencing Center, Washington University Genome Sequencing Center, Broad Institute, Children's Hospital Oakland Research Institute, S. Batzoglou, N. Goldman, R.C. Hardison, D. Haussler, W. Miller, A. Sidow, N.D. Trinklein, Z.D. Zhang, L. Barrera, R. Stuart, D.C. King, A. Ameur, S. Enroth, M.C. Bieda, J. Kim, A.A. Bhinge, N. Jiang, J. Liu, F. Yao, V.B. Vega, C.W.H. Lee, P. Ng, A. Shahab, A. Yang, Z. Moqtaderi, Z. Zhu, X. Xu, S. Squazzo, M.J. Oberley, D. Inman, M.A. Singer, T.A. Richmond, K.J. Munn, A. Rada-Iglesias, O. Wallerman, J. Komorowski, J.C. Fowler, P. Couttet, A.W. Bruce, O.M. Dovey, P.D. Ellis, C.F. Langford, D.A. Nix, G. Euskirchen, S. Hartman, A.E. Urban, P. Kraus, S. Van Calcar, N. Heintzman, T.H. Kim, K. Wang, C. Qu, G. Hon, R. Luna, C.K. Glass, M.G. Rosenfeld, S.F. Aldred, S.J. Cooper, A. Halees, J.M. Lin, H.P. Shulha, X. Zhang, M. Xu, J.N.S. Haidar, Y. Yu, Y. Ruan, V.R. Iyer, R.D. Green, C. Wadelius, P.J. Farnham, B. Ren, R.A. Harte, A.S. Hinrichs, H. Trumbower, H. Clawson,

J. Hillman-Jackson, A.S. Zweig, K. Smith, A. Thakkapallayil, G. Barber, R.M. Kuhn, D. Karolchik, L. Armengol, C.P. Bird, P.I.W. de Bakker, A.D. Kern, N. Lopez-Bigas, J.D. Martin, B.E. Stranger, A. Woodroffe, E. Davydov, A. Dimas, E. Eyras, I.B. Hallgrímsdóttir, J. Huppert, M.C. Zody, G.R. Abecasis, X. Estivill, G.G. Bouffard, X. Guan, N.F. Hansen, J.R. Idol, V.V.B. Maduro, B. Maskeri, J.C. McDowell, M. Park, P.J. Thomas, A.C. Young, R.W. Blakesley, D.M. Muzny, E. Sodergren, D.A. Wheeler, K.C. Worley, H. Jiang, G.M. Weinstock, R.A. Gibbs, T. Graves, R. Fulton, E.R. Mardis, R.K. Wilson, M. Clamp, J. Cuff, S. Gnerre, D.B. Jaffe, J.L. Chang, K. Lindblad-Toh, E.S. Lander, M. Koriabine, M. Nefedov, K. Osoegawa, Y. Yoshinaga, B. Zhu, P.J. de Jong, Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project, Nature 447 (2007) 799–816.

[91] B. Maher, ENCODE: The human encyclopaedia, Nature 489 (2012) 46–48.

[92] E. Pennisi, Genomics. ENCODE project writes eulogy for junk DNA, Science 337 (2012) 1159–1161.

[93] J.A. Stamatoyannopoulos, What does our genome encode? Genome Res. 22 (2012) 1602–1611.

[94] M. Yavartanoo, J.K. Choi, ENCODE: A Sourcebook of Epigenomes and Chromatin Language, Genomics Inform. 11 (2013) 2–6.

[95] H. Qu, X. Fang, A brief review on the Human Encyclopedia of DNA Elements (ENCODE) project, Genomics Proteomics Bioinf. 11 (2013) 135–141.

[96] A.L. Dixon, L. Liang, M.F. Moffatt, W. Chen, S. Heath, K.C.C. Wong, J. Taylor, E. Burnett, I. Gut, M. Farrall, G.M. Lathrop, G.R. Abecasis, W.O.C. Cookson, A genome-wide association study of global gene expression, Nat. Genet. 39 (2007) 1202–1207.

[97] N.C. Sheffield, R.E. Thurman, L. Song, A. Safi, J.A. Stamatoyannopoulos, B. Lenhard, G.E. Crawford, T.S. Furey, Patterns of regulatory activity across diverse human cell types predict tissue identity, transcription factor binding, and long-range interactions, Genome Res. 23 (2013) 777–788.

[98] B. Vernot, A.B. Stergachis, M.T. Maurano, J. Vierstra, S. Neph, R.E. Thurman, J.A. Stamatoyannopoulos, J.M. Akey, Personal and population genomics of human regulatory variation, Genome Res. 22 (2012) 1689–1697.

[99] W. Cookson, L. Liang, G. Abecasis, M. Moffatt, M. Lathrop, Mapping complex disease traits with global gene expression, Nat. Rev. Genet. 10 (2009) 184–194.

[100] B.P. Fairfax, S. Makino, J. Radhakrishnan, K. Plant, S. Leslie, A. Dilthey, P. Ellis, C. Langford, F.O. Vannberg, J.C. Knight, Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles, Nat. Genet. 44 (2012) 502–510.

[101] R.S.N. Fehrmann, R.C. Jansen, J.H. Veldink, H.-J. Westra, D. Arends, M.J. Bonder, J. Fu, P. Deelen, H.J.M. Groen, A. Smolonska, R.K. Weersma, R.M.W. Hofstra, W.A. Buurman, S. Rensen, M.G.M. Wolfs, M. Platteel, A. Zhernakova, C.C. Elbers, E.M. Festen, G. Trynka, M.H. Hofker, C.G.J. Saris, R.A. Ophoff, L.H. van den Berg, D.A. van Heel, C. Wijmenga, G.J. Te Meerman, L. Franke, Trans-eQTLs reveal that independent genetic variants associated with a complex phenotype converge on intermediate genes, with a major role for the HLA, PLoS Genet. 7 (2011) e1002197 .

[102] F.W. Albert, L. Kruglyak, The role of regulatory variation in complex traits and disease, Nat. Rev. Genet. 16 (2015) 197–212.

[103] Q. Li, A. Stram, C. Chen, S. Kar, S. Gayther, P. Pharoah, C. Haiman, B. Stranger, P. Kraft, M.L. Freedman, Expression QTL-based analyses reveal candidate causal genes and loci across five tumor types, Hum. Mol. Genet. 23 (2014) 5294–5302.

[104] H. Ongen, C.L. Andersen, J.B. Bramsen, B. Oster, M.H. Rasmussen, P.G. Ferreira, J. Sandoval, E. Vidal, N. Whiffin, A. Planchon, I. Padioleau, D. Bielser, L. Romano, I. Tomlinson, R.S. Houlston, M. Esteller, T.F. Orntoft, E.T. Dermitzakis, Putative cis-regulatory drivers in colorectal cancer, Nature 512 (2014) 87–90.

[105] T. Raj, K. Rothamel, S. Mostafavi, C. Ye, M.N. Lee, J.M. Replogle, T. Feng, M. Lee, N. Asinovski, I. Frohlich, S. Imboywa, A. Von Korff, Y. Okada, N.A. Patsopoulos, S. Davis, C. McCabe, H. Paik, G.P. Srivastava, S. Raychaudhuri, D.A. Hafler, D. Koller, A. Regev, N. Hacohen, D. Mathis, C. Benoist, B.E. Stranger, P.L. De Jager, Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes, Science 344 (2014) 519–523.

[106] A. Ramasamy, D. Trabzuni, S. Guelfi, V. Varghese, C. Smith, R. Walker, T. De, UK Brain Expression Consortium, North American Brain Expression Consortium, L. Coin, R. de Silva, M.R. Cookson, A.B. Singleton, J. Hardy, M. Ryten, M.E. Weale, Genetic variability in the regulation of gene expression in ten regions of the human brain, Nat. Neurosci. 17 (2014) 1418–1428.

[107] S. de Jong, K.R. van Eijk, D.W.L.H. Zeegers, E. Strengman, E. Janson, J.H. Veldink, L.H. van den Berg, W. Cahn, R.S. Kahn, M.P.M. Boks, R.A. Ophoff, PGC Schizophrenia (GWAS) Consortium, Expression QTL analysis of top loci from GWAS meta-analysis highlights additional schizophrenia candidate genes, Eur. J. Hum. Genet. 20 (2012) 1004–1008.

[108] J.J. Neefjes, H.L. Ploegh, Allele and locus-specific differences in cell surface expression and the association of HLA class I heavy chain with beta 2-microglobulin: differential effects of inhibition of glycosylation on class I subunit association, Eur. J. Immunol. 18 (1988) 801–810.

[109] R. Apps, Z. Meng, G.Q. Del Prete, J.D. Lifson, M. Zhou, M. Carrington, Relative expression levels of the HLA class-I proteins in normal and HIV-infected cells, J. Immunol. 194 (2015) 3594–3600.

[110] S. Kulkarni, V. Ramsuran, M. Rucevic, S. Singh, A. Lied, V. Kulkarni, C. O'hUigin, S. Le Gall, M. Carrington, Posttranscriptional Regulation of HLA-A Protein Expression by Alternative Polyadenylation Signals Involving the RNA-Binding Protein Syncrip, J. Immunol. 199 (2017) 3892–3899.

[111] Y. Kamatani, S. Wattanapokayakit, H. Ochi, T. Kawaguchi, A. Takahashi, N. Hosono, M. Kubo, T. Tsunoda, N. Kamatani, H. Kumada, A. Puseenam, T. Sura, Y. Daigo, K. Chayama, W. Chantratita, Y. Nakamura, K. Matsuda, A genome-wide association study identifies variants in the HLA-DP locus associated with chronic hepatitis B in Asians, Nat. Genet. 41 (2009) 591–595.

[112] E. Png, A. Thalamuthu, R.T.H. Ong, H. Snippe, G.J. Boland, M. Seielstad, A genome-wide association study of hepatitis B vaccine response in an Indonesian population reveals multiple independent risk variants in the HLA region, Hum. Mol. Genet. 20 (2011) 3893–3898.

[113] J.R. Honegger, D. Tedesco, J.A. Kohout, M.R. Prasad, A.A. Price, T. Lindquist, S. Ohmer, M. Moore-Clingenpeel, A. Grakoui, C.M. Walker, Influence of IFNL3 and HLA-DPB1 genotype on postpartum control of hepatitis C virus replication and T-cell recovery, Proc. Natl. Acad. Sci. U.S.A. 113 (2016) 10684–10689.

[114] P. Majumder, J.A. Gomez, J.M. Boss, The human major histocompatibility complex class II HLA-DRB1 and HLA-DQA1 genes are separated by a CTCF-binding enhancer-blocking element, J. Biol. Chem. 281 (2006) 18435–18443.

[115] S. Garnier, V. Truong, J. Brocheton, T. Zeller, M. Rovital, P.S. Wild, A. Ziegler, Cardiogenics Consortium, T. Munzel, L. Tiret, S. Blankenberg, P. Deloukas, J. Erdmann, C. Hengstenberg, N.J. Samani, H. Schunkert, W.H. Ouwehand, A.H. Goodall, F. Cambien, D.-A. Trégouët, Genome-wide haplotype analysis of cis expression quantitative trait loci in monocytes, PLoS Genet. 9 (2013), https://doi.org/10.1371/journal.pgen.1003240.

[116] S.A. Migueles, M.S. Sabbaghian, W.L. Shupert, M.P. Bettinotti, F.M. Marincola, L. Martino, C.W. Hallahan, S.M. Selig, D. Schwartz, J. Sullivan, M. Connors, HLA B*5701 is highly associated with restriction of virus replication in a subgroup of HIV-infected long term nonprogressors, Proc. Natl. Acad. Sci. U.S.A. 97 (2000) 2709–2714.

[117] R.A. Kaslow, M. Carrington, R. Apple, L. Park, A. Muñoz, A.J. Saah, J.J. Goedert, C. Winkler, S.J. O'Brien, C. Rinaldo, R. Detels, W. Blattner, J. Phair, H. Erlich, D.L. Mann, Influence of combinations of human major histocompatibility complex genes on the course of HIV-1 infection, Nat. Med. 2 (1996) 405–411.

[118] J.M. Carlson, A.Q. Le, A. Shahid, Z.L. Brumme, HIV-1 adaptation to HLA: a window into virus-host immune interactions, Trends Microbiol. 23 (2015) 212–224.

[119] M.-E. Blais, Y. Zhang, T. Rostron, H. Griffin, S. Taylor, K. Xu, H. Yan, H. Wu, I. James, M. John, T. Dong, S.L. Rowland-Jones, High frequency of HIV mutations associated with HLA-C suggests enhanced HLA-C-restricted CTL selective pressure associated with an AIDS-protective polymorphism, J. Immunol. 188 (2012) 4663–4670.

[120] Y. Momozawa, J. Dmitrieva, E. Théâtre, V. Deffontaine, S. Rahmouni, B. Charloteaux, F. Crins, E. Docampo, M. Elansary, A.-S. Gori, C. Lecut, R. Mariman, M. Mni, C. Oury, I. Altukhov, D. Alexeev, Y. Aulchenko, L. Amininejad, G. Bouma, F. Hoentjen, M. Löwenberg, B. Oldenburg, M.J. Pierik, A.E. Vander Meulen-de Jong, C. Janneke van der Woude, M.C. Visschedijk, International IBD Genetics Consortium, M. Lathrop, J.-P. Hugot, R.K. Weersma, M. De Vos, D. Franchimont, S. Vermeire, M. Kubo, E. Louis, M. Georges, IBD risk loci are enriched in multigenic regulatory modules encompassing putative causative genes, Nat Commun. 9 (2018) 2427.

[121] G. Cavalli, M. Hayashi, Y. Jin, D. Yorgov, S.A. Santorico, C. Holcomb, M. Rastrou, H. Erlich, I.W. Tengesdal, L. Dagna, C.P. Neff, B.E. Palmer, R.A. Spritz, C.A. Dinarello, MHC class II super-enhancer increases surface expression of HLA-DR and HLA-DQ and affects cytokine production in autoimmune vitiligo, Proc. Natl. Acad. Sci. U.S.A. 113 (2016) 1363–1368.

[122] E.W. Petersdorf, M. Malkki, T.A. Gooley, S.R. Spellman, M.D. Haagenson, M.M. Horowitz, T. Wang, MHC-resident variation affects risks after unrelated donor hematopoietic cell transplantation, Sci Transl Med. 4 (2012) 144ra101.

[123] B. Schöne, S. Bergmann, K. Lang, I. Wagner, A.H. Schmidt, E.W. Petersdorf, V. Lange, Predicting an HLA-DPB1 expression marker based on standard DPB1 genotyping: Linkage analysis of over 32,000 samples, Hum. Immunol. 79 (2018) 20–27.

[124] K. Fleischhauer, B.E. Shaw, T. Gooley, M. Malkki, P. Bardy, J.-D. Bignon, V. Dubois, M.M. Horowitz, J.A. Madrigal, Y. Morishima, M. Oudshoorn, O. Ringden, S. Spellman, A. Velardi, E. Zino, E.W. Petersdorf, International Histocompatibility Working Group in Hematopoietic Cell Transplantation, Effect of T-cell-epitope matching at HLA-DPB1 in recipients of unrelated-donor haemopoietic-cell transplantation: a retrospective study, Lancet Oncol. 13 (2012) 366–374.

[125] T. Sasazuki, H. Inoko, S. Morishima, Y. Morishima, Gene Map of the HLA Region, Graves' Disease and Hashimoto Thyroiditis, and Hematopoietic Stem Cell Transplantation, Adv. Immunol. 129 (2016) 175–249.

[126] J. Robinson, L.A. Guethlein, N. Cereb, S.Y. Yang, P.J. Norman, S.G.E. Marsh, P. Parham, Distinguishing functional polymorphism from random variation in the sequences of > 10,000 HLA-A, -B and -C alleles, PLoS Genet. 13 (2017) e1006862 .

[127] S.J. Mack, A gene feature enumeration approach for describing HLA allele polymorphism, Hum. Immunol. 76 (2015) 975–981.

[128] R. Kosoy, R. Nassir, C. Tian, P.A. White, L.M. Butler, G. Silva, R. Kittles, M.E. Alarcon-Riquelme, P.K. Gregersen, J.W. Belmont, F.M. De La Vega, M.F. Seldin, Ancestry informative marker sets for determining continental origin and admixture proportions in common populations in America, Hum. Mutat. 30 (2009) 69–78.

[129] J.R. Kidd, F.R. Friedlaender, W.C. Speed, A.J. Pakstis, F.M. De La Vega, K.K. Kidd, Analyses of a set of 128 ancestry informative single-nucleotide polymorphisms in a global set of 119 population samples, Investig Genet. 2 (2011) 1.

[130] L.C. Schalkwyk, E.L. Meaburn, R. Smith, E.L. Dempster, A.R. Jeffries, M.N. Davies, R. Plomin, J. Mill, Allelic skewing of DNA methylation is widespread across the genome, Am. J. Hum. Genet. 86 (2010) 196–212.

[131] A.T. Dilthey, L. Moutsianas, S. Leslie, G. McVean, HLA*IMP–an integrated framework for imputing classical HLA alleles from SNP genotypes, Bioinformatics. 27 (2011) 968–972.

[132] E.W. Petersdorf, M. Malkki, T.A. Gooley, P.J. Martin, Z. Guo, MHC haplotype matching for unrelated hematopoietic cell transplantation, PLoS Med. 4 (2007) e8 .
[133] E.W. Petersdorf, P. Stevenson, M. Malkki, R.K. Strong, S.R. Spellman, M.D. Haagenson, M.M. Horowitz, T. Gooley, T. Wang, Patient HLA Germline Variation and Transplant Survivorship, J. Clin. Oncol. 36 (2018) 2524–2531.
[134] K. Hosomichi, T. Shiina, A. Tajima, I. Inoue, The impact of next-generation sequencing technologies on HLA research, J. Hum. Genet. 60 (2015) 665–673.
[135] P.J. van den Elsen, Expression regulation of major histocompatibility complex class I and class II encoding genes, Front Immunol. 2 (2011) 48.
[136] P. Majumder, J.A. Gomez, B.P. Chadwick, J.M. Boss, The insulator factor CTCF controls MHC class II gene expression and is required for the formation of long-distance chromatin interactions, J. Exp. Med. 205 (2008) 785–798.
[137] N.P. Mayor, J.D. Hayhurst, T.R. Turner, R.M. Szydlo, B.E. Shaw, W.P. Bultitude, S. Jex-Ray, F. Tavarozzi, K. Latham, C. Anthias, J. Robinson, H. Braud, R. Danby, J. Perry, M.C. Wilson, A.J. Bloor, A. Clark, S. McKinnon, D.L. Marks, A. Pagliuca, M.N. Potter, V.T. Potter, N.H. Russell, K.J. Thomson, J.A. Madrigal, S.G.E. Marsh, Better HLA matching as revealed by ultra-high resolution HLA typing results in superior overall survival, HLA 91 (2018) 337–338.