Immunogenomics: Molecular hide and seek

Marcos M. Miretti and Stephan Beck*

Immunogenomics Laboratory, The Wellcome Trust Sanger Institute, Genome Campus, Hinxton, Cambridge, CB10 1SA, UK
*Correspondence to: Tel: +44 (0)1223 494952; Fax: +44 (0)1223 494919; E-mail: beck@sanger.ac.uk

Abstract

Similar to other classical science disciplines, immunology has been embracing novel technologies and approaches giving rise to specialised sub-disciplines such as immunogenetics and, more recently, immunogenomics, which, in many ways, is the genome-wide application of immunogenetic approaches. Here, recent progress in the understanding of the immune sub-genome will be reviewed, and the ways in which immunogenomic datasets consisting of genetic and epigenetic variation, linkage disequilibrium and recombination can be harnessed for disease association and evolutionary studies will be discussed. The discussion will focus on data available for the major histocompatibility complex and the leukocyte receptor complex, the two most polymorphic regions of the human immune sub-genome.

Introduction

The ongoing 'hide and seek' between pathogens and their host immune systems has led to a molecular arms race which makes immunology one of the most interesting, but also one of the most complex, areas of research. By contrast with the harmless children's game, the outcome of this molecular race determines health versus disease and, in many cases, survival versus death. A prerequisite for immunogenomic research is the availability of comprehensive and fully informative sequence variation and gene maps. A high-quality sequence map of the human genome and successive generations of annotation have been available since 2001.[1,2] Using this sequence as a reference, global and population-specific variation maps of increasing resolution were subsequently generated by the International SNP Consortium and HapMap Projects.[3,4] Because of the extreme polymorphism encountered in some immune regions, separate efforts were focused on regions such as the major histocompatibility complex (MHC; reviewed by Allcock *et al.*[5]), the leukocyte receptor complex (LRC; reviewed by Yawata *et al.*[6]) and others, to complement the global variation map. In addition, several specialised databases, including the Immuno Polymorphism Database[7] and the International Immunogenetics Information System,[8] catalogue many of the classical immune genes and their allelic variants. Using immune ontology definitions,[9] the first comprehensive gene map and database of the human immune sub-genome was recently reported to consist of 1,562 genes (about 7 per cent of human genes) that are distributed at varying densities across all chromosomes except the Y chromosome.[10]

In the following review, some of the benefits and caveats of using a holistic approach to immunogenomic analysis will be discussed with respect to genetic and epigenetic variation and linkage disequilibrium (LD).

Genetic variation

Based on both experimental data[11-15] and inferences from population genetics data,[16-18] there is evidence that recombination is not uniformly distributed across the human genome. In addition to the heterogeneity of recombination rates, it was suggested that the human genome could be subdivided into relatively short fragments with little or no evidence of historical recombination.[19] Under this premise, a genome-wide LD map is being constructed by the International HapMap Project, which aims to generate a resource of DNA variation in human populations with different ancestry.[3] In its first phase, which is close to completion, genotyping data at one single nucleotide polymorphism (SNP)/five kilobase (kb) density will provide enough resolution to portray the LD landscape of the genome, including regions of the immune sub-genome.

LD maps of complete chromosomes are already available;[20-22] however, detailed LD patterns at high SNP density covering large chromosomal regions encoding immune-related gene clusters have long been awaited. LD data have so far been reported for two such clusters: the LRC and the MHC. Within the LRC, the haplotypic nature of the killer immunoglobulin-like receptor (*KIR*) gene cluster causes a

DNA size variation of 120−200 kb as a consequence of the presence/absence of genes in some haplotypes.[23−26] This feature, in addition to intron similarity owing to recent duplications, restricts the availability of informative SNPs for LD studies[27] and poses particular restrictions on large-scale genotyping protocols. Consequently, a detailed LD map for the LRC is not available. For the MHC, a comprehensive LD map is available[16] as a direct consequence of existing SNP resources.[28]

Historically, considerable attention has been paid to the MHC region, resulting in a wealth of valuable information in a variety of areas — that is, origin and maintenance of nucleotide diversity in classical human leukocyte antigen (*HLA*) genes; gene conversion, average recombination rates, presence and distribution of the recombination hot spots; and extended LD in long-range haplotypes.[11,14,15,29−31] The high-resolution LD map available for the MHC permits the detailed study of the allelic correlation, which can be particularly useful for the selection of tagSNP sets for disease association studies. Since the history of recombination between two SNPs can be estimated using D', DNA segments showing no evidence of historical recombination have been defined as LD-blocks.[19] Consequently, haplotypes within such LD-blocks are likely to share a common ancestral haplotype, and the genealogy of an LD-block might be different from the genealogy of neighbouring LD-blocks. Phylogenetic analysis of haplotypes observed in a ∼110 kb LD-block (rs3115569−rs2022533) at the MHC class II−III boundary, the region of lowest recombination rate in the MHC (0.0093 centimorgan [cM]/ megabase [Mb]), surrounded by two recombination hot spots,[16] showed that the pattern of nucleotide differences between haplotypes within this LD-block is difficult to account for by mutation only, but could readily be explained by recombination events (Miretti *et al.*, unpublished). Following detailed inspection of the haplotype nucleotide sequence alignment at variable positions (Figure 1B), putative recombination sites can be identified. The subsequent splitting of the LD-block at a recombination site (Figure 1C) resulted in haplotypic networks presenting distinct clustering in both fragments (Figure 1D), which were also dissimilar from that of the complete LD-block. While it needs to be examined in a larger sample of LD-blocks from diverse populations, the presence of recombination within LD-blocks suggests that the occurrence of LD between two SNPs might not be sufficient to unambiguously detect historical recombination beyond the history of the sample population where LD is being measured. Recombination hot spots within LD-blocks — that is, hot spots that have not left an imprint on LD — have recently been postulated to be either old hot spots en route to extinction or too young to leave a mark on haplotype diversity in Europeans.[18] Additional experimental data based on sperm cell recombination analysis would help to confirm the signature of recombination identified here and to determine if it is an example of a potentially evolutionarily very young hot spot.[18]

Overall recombination rates within the MHC are known to be lower than across chromosome 6 and the genome average,[34,35] and regions of relatively high recombination rates (hot spots) and low recombination rates (cold spots) have been described in males based on microsatellite typing.[15] The fine-scale experimental recombination analysis in a 200 kb segment within the MHC class II region revealed that a few hot spots shape the distribution of LD throughout the segment, that hot spots tend to be delimited to short DNA segments (usually < 5 kb) and that a portion of the recombination is contributed by gene conversion.[11,31,36,37] The question therefore arises as to whether the recombination pattern observed in this segment can be extended to the complete MHC and, ultimately, if it represents a general characteristic of the genome. The heterogeneous model of recombination seems to be a general feature of the genome,[17] where the landscape is moulded by the local distribution and intensity of recombination. According to the recombination rates inferred from population genetic data,[16] the experimentally verified recombination hot spots in the 200 kb segment of the MHC class II region[11] might not represent a paradigm for the entire MHC, as their intensities are ∼2− 100 times higher than hot spots in the rest of the MHC. Figure 2 shows the uneven distribution of hot spots across the complete MHC — integrated with gene annotation, LD-blocks and tagSNP distribution, as represented by the GLOVAR genome browser (www.glovar.org) — and also that the hot spot intensity is far from being uniform. Additional large-scale experimental evidence of hot spot intensity and distribution is required to more accurately assess the concordance with predictions based on population genetic data. Importantly, hot spot features will significantly influence the selection of tagSNPs for disease association studies involving any particular region of the immune sub-genome.

From the detailed distribution of recombination and LD across the MHC, it would be possible to exploit the common variation observed in approximately 80 per cent of the MHC sequence, which is in high LD, to conduct association studies. Even with the current SNP density (1 SNP/1.9 kb), however, between 10−27 per cent of the DNA sequence in the MHC sub-regions is contained in regions of low LD, where tagSNP efficiency — the number of genotyped markers divided by the number of tagging SNPs[38] — would not entail any benefit, meaning that almost every SNP would need to be typed. Figure 2 illustrates the dependence of the tagSNP distribution on the SNP density and LD pattern, which is particularly variable across the MHC. Long regions of strong LD are common in the extended class I sub-region, where 75 tagSNPs are sufficient to recover haplotypic information provided by 408 SNPs. Conversely, 204 tagSNPs are necessary to recover the equivalent information from 466 SNPs in the class II sub-region, where the LD pattern is more interrupted and the decay of LD is more
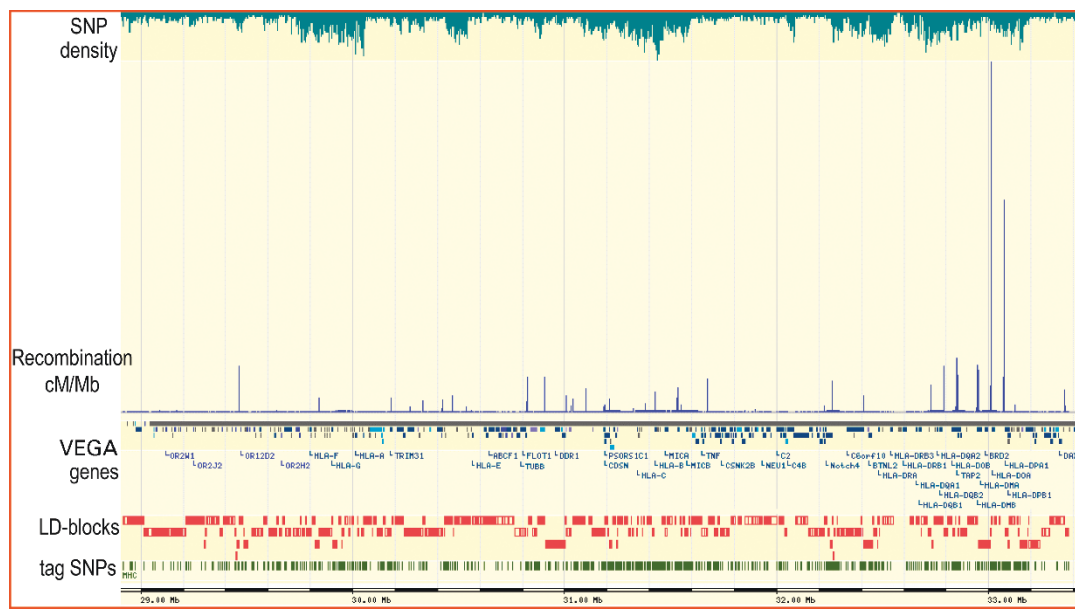
Figure 1. Potential recombination sites within a linkage disequilibrium (LD) block. Patterns of LD for the complete major histocompatibility complex were obtained for a panel of 180 Centre d'Etude du Polymorphisme Humain founder chromosomes[16] and LD-blocks defined according to Gabriel *et al.* criteria[19] employing Haploview.[32] (A) The region selected for the analysis constitutes an LD-block extending over a ∼110 kilobase (kb) DNA segment (National Center for Biotechnology Information built 35 coordinates: 32321766–32431723) containing most of the *C6ORF10* gene, with an average space between single nucleotide polymorphisms (SNPs) of 1.6 kb. This region also represents a recombination cold spot — with recombination rates = 0.00931 centimorgans/megabase, ten times lower relative to that on the neighbouring area — surrounded by two recombination hot spots identified in *NOTCH4* and *C6ORF10* genes.[16] (B) Ten haplotypes (rows) observed in this sample, which were recognised based on alleles at 68 variable positions (SNPs, in columns) within the selected LD-block. Haplotype names and frequencies (per cent) are given on the left. (C) A break in the haplotype alignment suggesting the presence of ancestral recombination within the LD-block. Haplotypes were then split at a potential recombination site into two sub-regions (with 41 and 27 SNPs, respectively) aiming to construct a phylogenetic network in order to check for consistency. (D) The distribution of variation among haplotypes as represented by median-joining trees obtained for both sub-regions employing the NETWORK program.[33] Haplotypes F and D are grouped closely related within each network but belong to different clusters when comparing networks from the two sub-regions. The number of substitutions involved in these branching differences could preferably be explained by recombination events rather than owing to mutations occurring between haplotypes. The distribution pattern of these differences adds consistency to this view and strongly suggests the presence of recombination. The circle size is proportional to the haplotype frequency, and the number of substitutions between nodes is represented by the number of lines between them. The site where the haplotypes were split can be changed to few positions on either side of the current site without significantly modifying the topology of the network.

Figure 2. Data integration in the major histocompatibility complex (MHC) region as represented by the GLOVAR genome browser (www.glovar.org) showing the distribution and intensity of recombination hot spots relative to gene annotation, linkage disequilibrium blocks and tag single nucleotide polymorphism distribution. Note that recombination hot spots within the MHC class II region are comparatively more intense than in the rest of the MHC. Most gene names were excluded owing to space constraints.

pronounced.[16] Also, the tagSNP sets are derived from studies which mostly dismiss loci of minor allele frequency (MAF; <5 per cent), which seem to be present at increased frequency in low LD regions,[39] thus excluding rare variants from the analyses. This raises the question of whether failure to detect disease variants simply results from the exclusion of SNPs with rare alleles from the analysis (discussed below). Furthermore, tagging effectiveness — the proportion of 'hidden' SNPs being detected in LD with SNPs from the tagging set — can be substantially increased by incorporating SNPs with rare alleles (MAF <5 per cent); this effect was even more prominent in low LD regions.[39]

Epigenetic variation

Much less is known about epigenetic variation, both in terms of causality and function. The most frequent and stable form of epigenetic variation is differential DNA methylation. DNA methylation, which was first discovered in 1948, occurs naturally at the carbon-5 position of cytosine (5-methyl cytosine) at CpG dinucleotides.[40] In subsequent years, it was proposed that DNA methylation played an important role in the regulation of gene expression[41,42] and disease aetiology, particularly cancer.[43] The discovery of CpG islands (sequences enriched in cytosine–guanosine dinucleotides) suggested candidate regions in the genome for epigenetic modulation.[44] DNA methylation occurs predominantly at CpG sites[45] and, in addition to gene regulation, is also involved in phenomena such as X-chromosome

inactivation in female mammals, parent-of-origin-specific, mono-allelic gene expression (imprinting) and epigenetic reprogramming in mammalian development.[46] The key technology for detecting 5-methyl cytosine by DNA sequencing was developed in 1992, and is known as bisulphite sequencing.[47]

The need to consider epigenetic variation alongside genetic variation in the context of disease has been highlighted by the finding of high discordance rates in monozygotic twin studies and numerous other studies, confirming that epigenetic factors play a decisive role in the aetiology of virtually all human pathologies (reviewed by Robertson and Wolffe).[48] Therefore, when the SNP Consortium announced in 1999 that it would generate a first-generation genetic variation map of the human genome,[49] the time and opportunity seemed right to generate a first-generation epigenetic (methylation) variation map alongside — which is what the Human Epigenome Consortium announced that it would do.[50] One aspect of the Human Epigenome Project (HEP) that is of particular interest in the context of this review, is the aim and ability to identify methylation variable positions (MVPs), which, together with SNPs, promise to significantly advance our ability to understand and diagnose human disease. MVPs are defined as differentially methylated CpG sites that have the statistical power to discriminate, for example, between biological states such as active versus inactive or healthy versus diseased.

As a result of the HEP and other studies, a detailed genomic methylation map is already available for the MHC[51] and

models for the epigenetic control of the *KIR* expression repertoire have been proposed (reviewed by Uhrberg[52]). For the MHC, methylation profiles have been generated essentially for all expressed genes, demonstrating tissue-specificity (eg for the *C2* locus) and inter-individual heterogeneity (eg for the *TNF* locus). Previously, it has been shown that transcription profiles can be associated with specific haplotypes,[53] with epigenetic states[54] or even with specific epi-alleles in the absence of DNA variation.[55] Overall, the methylation profile of the MHC appears to be strongly bimodal, with good correlation between hypo- and hypermethylation within the upstream regions of genes and their transcriptional activity (Figure 3). For the LRC, fewer data are available and are mostly restricted to the *KIR* genes. *KIR* genes exhibit a high degree of polymorphism and are expressed in a clonally restricted fashion.[56] Each cell expresses a mono-allelic repertoire which is highly variable with respect to number and combination of

receptors. *KIR* genes are regulated by three types of promoters, corresponding to the three different modes of expression: one for *KIR2DL4*, which is constitutively expressed; one for *KIR3DL3*, which is expressed at a very low level; and one which is common to all clonally distributed *KIR* genes.[57] Given the absence of promoter variability among the clonally distributed *KIR* genes, it is, in fact, likely that the variegated expression of these genes is mainly regulated epigenetically, and a model has been proposed to this effect.[52] The model proposes four stages of sequential DNA and histone modifications leading to the observed mosaicism of clonally restricted *KIR* expression but requires some still unproven processes, including locus- and allele-specific DNA demethylation.

Compared with genetic data, the ability to analyse and interpret epigenetic data is still fairly simplistic, although enormous progress has been made, particularly in the past few years. What is becoming increasingly clear, however, is that
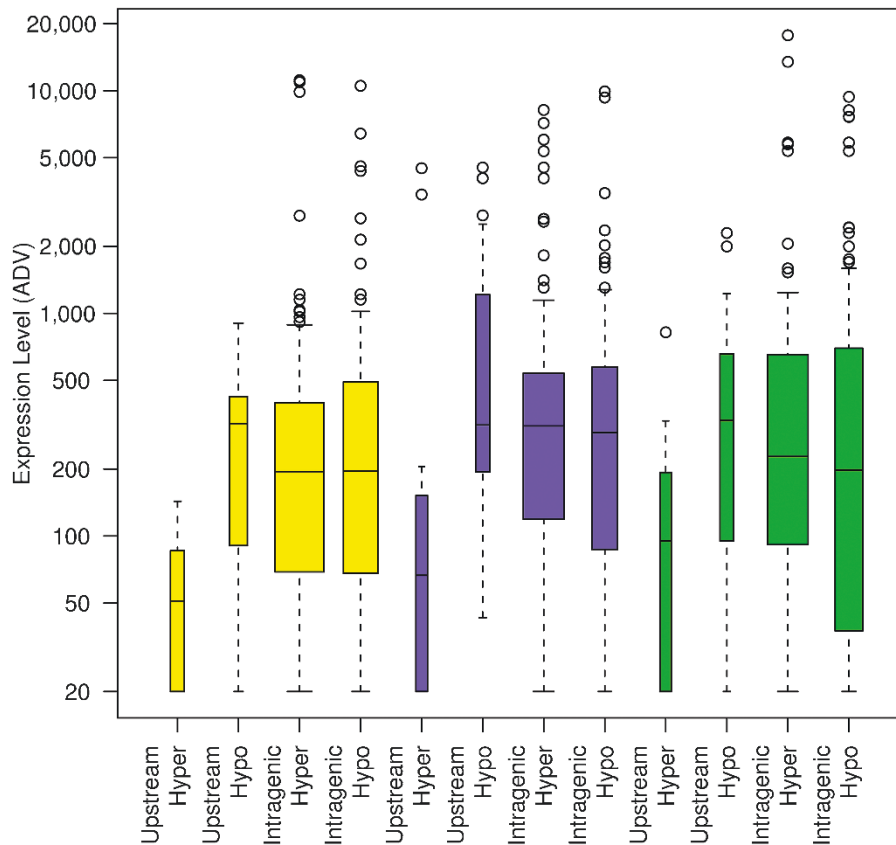

Figure 3. Correlation between DNA methylation and gene expression. DNA methylation data derived from the Human Epigenome Pilot Project were grouped according to 5'-UTR being methylated or unmethylated and the body of the gene being methylated or unmethylated. Each group was then plotted against the corresponding Genomics Institute of the Novartis Research Foundation (GNF) (http://symatlas.gnf.org/SymAtlas) expression level, which is set to anything below 200 as not being expressed and anything above 200 as being expressed. This cluster analysis shows that genes with methylated 5'-UTRs are silent and genes with unmethylated 5'-UTRs are expressed. In this study, no correlation was found between expression and the methylation level within the body of the gene. Reproduced from Rakyan *et al.*[51]

the two are inextricably intertwined, and efforts towards an integrated (epi)genetic approach to common disease are well underway.[58,59]

Application of immunogenomic data for disease association studies

Major progress in genotyping technology has improved our ability to harvest genotypic information from a comprehensive number of loci and has laid down the foundation for the genome-wide common variation survey,[3] enhancing our understanding of human genetic variation. Current genotyping costs, however, make genome-wide association studies still challenging for the large sample sizes that will be required for adequately powered studies. The inter-dependencies of marker numbers, allele frequency, LD, cohort size and power in disease association studies have recently been reviewed[60−62] and will therefore not be discussed in depth here. Essentially, genome-wide studies are still restricted by the allelic spectra underlying complex diseases. The frequency distribution of the disease variants and the proportion of the trait variance for which they are responsible determine the potential power of genetic association studies and, therefore, the feasibility of the study. Disease susceptibility variants associated with Mendelian disorders tend to be modelled by purifying selection and present low population frequencies (<1 per cent).[63] Thus, rare alleles — also including low-frequency, mildly deleterious variants together with those alleles in high LD — might be under-represented in genomic LD surveys and tagSNP sets excluding loci with MAF < 5 per cent. It has been proposed, however, that genetic risk for common diseases and most of the clinically important traits are determined by the joint contribution of disease-predisposing high-frequency alleles that are shared between unrelated affected individuals in the population: the common-disease/common-variant (CD/CV) model.[60,64,65] A less extreme model based on the divergence of the allelic spectra of disease susceptibility variants relative to that of all variants has been proposed for evaluating the impact of allele architecture on common diseases.[61] Ultimately, the applicability of the CD/CV model in disease association studies will eventually rely on the selected markers and on the relationship between the four parameters that affect the apparent size at the marker locus, namely the odds ratio of the disease allele, the disease allele frequency, the marker allele frequency and the LD between the marker and the disease locus (reviewed by Zondervan and Cardon[62]).

Diseases related to defective immune response, such as autoimmune disorders, might comply with the CD/CV model of allelic frequency distribution for susceptibility loci, showing increased allele frequency after being under positive selection for infectious disease resistance or heterozygote advantage.[61,66] Direct extrapolations to the immune sub-genome and autoimmune diseases could be elusive, however; that is, it would be difficult to prove that alleles having an effect on the function of an immune gene — and also contribute to the risk of autoimmune diseases — are beneficial on a population level because they increase the diversity in the immune response.[67] In fact, the overall level of polymorphism in immune genes is, in general, similar to that of non-immune genes,[10] and only a few loci among the genes constituting the immune sub-genome largely contribute to the diversity of the immune response — disregarding loci with somatic recombination. In addition, modes of selection maintaining extreme allelic diversity levels — heterozygote advantage, balanced selection — have been proposed for these few loci, including the classical *HLA* class I and II and *KIR* genes.[29,30,66,68] Consequently, it is probably premature to generalise any disease model for immune genes. A recent study shows, for instance, that multiple disease loci with predominant disease variants are not necessarily required to develop a complex immune disorder.[69] It has been shown that sarcoidosis — a multi-systemic immune disorder initially associated with several classical HLA markers — is associated with a single disease locus independently responsible for most of the predisposing influence.[69] The authors of this study demonstrated that a single transition in exon 5 of the *BTNL2* gene, which alters a splice site and causes a premature stop codon, has a profound effect on the mature protein structure and function.

A number of autoimmune disease-susceptibility loci have been mapped to the MHC and LRC regions.[70−72] Autoimmunity and cancer are associated with the progressive decline in immune functions, such as decline in T cell function, dysregulation of T cell apoptosis and immune senescence during ageing.[73,74] Tolerance induction of T cells is mediated by cytotoxic T lymphocyte antigen 4. Polymorphism in this inhibitory receptor has also been associated with autoimmune diseases.[75,76]

Outlook

Since the term haplotype was coined in 1967,[77] a vast amount of literature has been published employing haplotypes to construct genetic and LD maps, which have enabled the emerging field of immunogenetics to make major contri-butions to immunology, population genetics and medicine.[78] For the MHC, a haplotype hierarchy has been defined based on pedigree and population genetic data, where physically linked classical *HLA* genes constitute 'blocks' of relatively short DNA segments containing alleles in LD at different loci, namely *HLA-B/C*, *TNF*, complotype and *HLA-DQB/DRB*. Combinations of these four basic blocks are frequently found in LD, constituting 'conserved extended haplotypes' (CEH)[79,80] or 'ancestral haplotypes'.[81] More recently, local fine-scale LD blocks based on high-density SNP typing have

been described and meiotic recombination hot spots inferred from population genetic data. The integration of long-range LD in CEH and local fine-scale LD data will provide new and exciting opportunities to explore the evolution and disease associations of the immune sub-genome. While the jury is still out on the validity of the CD/CV model, the recently announced Wellcome Trust Case Control Consortium (WTCCC) is likely to provide the answer to this ongoing debate. The WTCCC proposes to analyse 19,000 samples for common genetic variants in eight common diseases using data generated by the HapMap Consortium. Integration of the genetic and epigenetic datasets discussed here is a necessary next step towards a more holistic approach to immunological research, which is one of the aims of the recently formed International Immunomics Society, IMMIS (http://research. i2r.a-star.edu.sg/IIMMS/). The IMMIS was formed with the main objective of promoting the science of immunomics, which is the interdisciplinary field spanning immunology, immunoinformatics, genomics, proteomics, bioinformatics and related scientific fields.

Acknowledgments

References

1. International Human Genome Sequencing Consortium (2004), 'Finishing the euchromatic sequence of the human genome', *Nature* Vol. 431, pp. 931–945.

2. International Human Genome Sequencing Consortium (2001), 'Initial sequencing and analysis of the human genome', *Nature* Vol. 409, pp. 860–921.

3. The International HapMap Consortium (2003), 'The International HapMap Project', *Nature* Vol. 426, pp. 789–796.

4. Sachidanandam, R., Weissman, D., Schmidt, S.C. *et al.* (2001), 'A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms', *Nature* Vol. 409, pp. 928–933.

5. Allcock, R.J.N., Atrazhev, A.M., Beck, S. *et al.* (2002), 'The MHC haplotype project: A resource for HLA-linked association studies', *Tissue Antigens* Vol. 59, pp. 520–521.

6. Yawata, M., Yawata, N., McQueen, K. *et al.* (2002), 'Predominance of group A KIR haplotypes in Japanese associated with diverse NK cell repertoires of KIR expression', *Immunogenetics* Vol. 54, pp. 543–550.

7. Robinson, J., Waller, M.J., Stoehr, P. and Marsh, S.G.E. (2005), 'IPD — The Immuno Polymorphism Database', *Nucl. Acids Res.* Vol. 33(Suppl. 1), pp. D523–D526.

8. Lefranc, M.-P., Giudicelli, V., Kaas, Q. *et al.* (2005), 'IMGT, the international ImMunoGeneTics information system(R)', *Nucl. Acids Res.* Vol. 33(Suppl. 1), pp. D593–D597.

9. Trowsdale, J. and Parham, P. (2004), 'Defense strategies and immunity-related genes', *Eur. J. Immunol.* Vol. 34, pp. 7–17.

10. Kelley, J., de Bono, B. and Trowsdale, J. (2005), 'IRIS: A database surveying known human immune system genes', *Genomics* Vol. 85, pp. 503–511.

11. Jeffreys, A.J., Kauppi, L. and Neumann, R. (2001), 'Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex', *Nat. Genet.* Vol. 29, pp. 217–222.

12. Jeffreys, A.J., Ritchie, A. and Neumann, R. (2000), 'High resolution analysis of haplotype diversity and meiotic crossover in the human TAP2 recombination hotspot', *Hum. Mol. Genet.* Vol. 9, pp. 725–733.

13. Cullen, M., Erlich, H., Klitz, W. and Carrington, M. (1995), 'Molecular mapping of a recombination hotspot located in the 2nd intron of the human Tap2 locus', *Am. J. Hum. Genet.* Vol. 56, pp. 1350–1358.

14. Cullen, M., Noble, J., Erlich, H. *et al.* (1997), 'Characterization of recombination in the HLA class II region', *Am. J. Hum. Genet.* Vol. 60, pp. 397–407.

15. Cullen, M., Perfetto, S.P., Klitz, W. *et al.* (2002), 'High-resolution patterns of meiotic recombination across the human major histocompatibility complex', *Am. J. Hum. Genet.* Vol. 71, pp. 759–776.

16. Miretti, M.M., Walsh, E.C., Ke, X.Y. *et al.* (2005), 'A high-resolution linkage-disequilibrium map of the human major histocompatibility complex and first generation of tag single-nucleotide polymorphisms', *Am. J. Hum. Genet.* Vol. 76, pp. 634–646.

17. McVean, G.A.T., Myers, S.R., Hunt, S. *et al.* (2004), 'The fine-scale structure of recombination rate variation in the human genome', *Science* Vol. 304, pp. 581–584.

18. Jeffreys, A.J., Neumann, R., Panayi, M. *et al.* (2005) 'Human recombination hot spots hidden in regions of strong marker association', *Nature Genetics* Vol. 37, pp. 601–606.

19. Gabriel, S.B., Schaffner, S.F., Nguyen, H. *et al.* (2002), 'The structure of haplotype blocks in the human genome', *Science* Vol. 296, pp. 2225–2229.

20. Dawson, E., Abecasis, G.R., Bumpstead, S. *et al.* (2002), 'A first-generation linkage disequilibrium map of human chromosome 22', *Nature* Vol. 418, pp. 544–548.

21. Phillips, M.S., Lawrence, R., Sachidanandam, R. *et al.* (2003), 'Chromosome-wide distribution of haplotype blocks and the role of recombination hot spots', *Nat. Genet.* Vol. 33, pp. 382–387.

22. De la Vega, F.M., Isaac, H., Collins, A. *et al.* (2005), 'The linkage disequilibrium maps of three human chromosomes across four populations reflect their demographic history and a common underlying recombination pattern', *Genome Res.* Vol. 15, pp. 454–462.

23. Uhrberg, M., Valiante, N.M., Shum, B.P. *et al.* (1997), 'Human diversity in killer cell inhibitory receptor genes', *Immunity* Vol. 7, pp. 753–763.

24. Uhrberg, M., Parham, P. and Wernet, P. (2002), 'Definition of gene content for nine common group B haplotypes of the Caucasoid population: KIR haplotypes contain between seven and eleven KIR genes', *Immunogenetics* Vol. 54, pp. 221–229.

25. Hsu, K.C., Chida, S., Dupont, B. and Geraghty, D.E. (2002), 'The killer cell immunoglobulin-like receptor (KIR) genomic region: Gene-order, haplotypes and allelic polymorphism', *Immunol. Rev.* Vol. 190, pp. 40–52.

26. Parham, P. (2005), 'MHC class I molecules and kirs in human history, health and survival', *Nat. Rev. Immunol.* Vol. 5, pp. 201–214.

27. Norman, P.J., Cook, M.A., Carey, B.S. *et al.* (2004), 'SNP haplotypes and allele frequencies show evidence for disruptive and balancing selection in the human leukocyte receptor complex', *Immunogenetics* Vol. 56, pp. 225–237.

28. Stewart, C.A., Horton, R., Allcock, R.J.N. *et al.* (2004), 'Complete MHC haplotype sequencing for common disease gene mapping', *Genome Res.* Vol. 14, pp. 1176–1187.

29. Hughes, A. and Nei, M. (1988), 'Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection', *Nature* Vol. 336, pp. 167–170.

30. Hughes, A.L. and Nei, M. (1989), 'Nucleotide substitution at major histocompatibility complex class II loci: Evidence for overdominant selection', *Proc. Natl. Acad. Sci. USA* Vol. 86, pp. 958–962.

31. Jeffreys, A.J. and May, C.A. (2004), 'Intense and highly localized gene conversion activity in human meiotic crossover hot spots', *Nat. Genet.* Vol. 36, pp. 151–156.

32. Barrett, J.C., Fry, B., Maller, J. and Daly, M.J. (2005), 'Haploview: Analysis and visualization of LD and haplotype maps', *Bioinformatics* Vol. 21, pp. 263–265.

33. Bandelt, H., Forster, P. and Rohl, A. (1999), 'Median-joining networks for inferring intraspecific phylogenies', *Mol. Biol. Evol.* Vol. 16, pp. 37–48.

34. Kong, A., Gudbjartsson, D.F., Sainz, J. *et al.* (2002), 'A high-resolution recombination map of the human genome', *Nat. Genet.* Vol. 31, pp. 241–247.

35. Mungall, A.J., Palmer, S.A., Sims, S.K. *et al.* (2003), 'The DNA sequence and analysis of human chromosome 6', *Nature* Vol. 425, pp. 805–811.

36. Kauppi, L., Stumpf, M.P.H. and Jeffreys, A.J. (2006), 'Localized breakdown in linkage disequilibrium does not always predict sperm crossover hot spots in the human MHC class II region', *Genomics* In Press.

37. Kauppi, L., Sajantila, A. and Jeffreys, A.J. (2003), 'Recombination hotspots rather than population history dominate linkage disequilibrium in the MHC class II region', *Hum. Mol. Genet.* Vol. 12, pp. 33–40.

38. Ke, X., Durrant, C., Morris, A.P. *et al.* (2004), 'Efficiency and consistency of haplotype tagging of dense SNP maps in multiple samples', *Hum. Mol. Genet.* Vol. 13, pp. 2557–2565.

39. Ke, X., Miretti, M.M., Broxholme, J. *et al.* (2005), 'A comparison of tagging methods and their tagging space', *Hum. Mol. Genet.* Vol. 14, pp. 2757–2767.

40. Hotchkiss, R. (1948), 'The quantitative separation of purines, pyrimidines and nucleosides by paper chromatography', *J. Biol. Chem.* Vol. 168, pp. 315–332.

41. Riggs, A. (1975), 'X inactivation, differentiation, and DNA methylation', *Cytogenetics and cell genetics* Vol. 14, pp. 9–25.

42. Holliday, R. and Pugh, J. (1975), 'DNA modification mechanisms and gene activity during development', *Science* Vol. 187, pp. 226–232.

43. Feinberg, A. and Vogelstein, B. (1983), 'Hypomethylation distinguishes genes of some human cancers from their normal counterparts', *Nature* Vol. 301, pp. 89–92.

44. Bird, A. (1986), 'CpG-rich islands and the function of DNA methylation', *Nature* Vol. 321, pp. 209–213.

45. Clark, S.J., Harrison, J. and Frommer, M. (1995), 'CpNpG methylation in mammalian cells' Vol. 10, pp. 20–27.

46. Reik, W., Dean, W. and Walter, J. (2001), 'Epigenetic reprogramming in mammalian development', *Science* Vol. 293, pp. 1089–1093.

47. Frommer, M., McDonald, L., Millar, D. *et al.* (1992), 'A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands', *Proc. Natl. Acad. Sci. USA* Vol. 89, pp. 1827–1831.

48. Robertson, K. and Wolffe, A. (2000), 'DNA methylation in health and disease', *Nat. Rev. Genet.* Vol. 1, pp. 11–19.

49. Masood, E. (1999), 'As consortium plans free SNP map of human genome', *Nature* Vol. 398, pp. 545–546.

50. Beck, S., Olek, A. and Walter, J. (1999), 'From genomics to epigenomics: A loftier view of life', *Nat. Biotechnol.* Vol. 17, p. 1144.

51. Rakyan, V.K., Hildmann, T., Novik, K.L. *et al.* (2004), 'DNA methylation profiling of the human major histocompatibility complex: A pilot study for the Human Epigenome Project', *PLoS Biol.* Vol. 2, p. e405.

52. Uhrberg, M. (2005), 'Shaping the human NK cell repertoire: An epigenetic glance at KIR gene regulation', *Mol. Immunol.* Vol. 42, pp. 471–475.

53. Knight, J.C., Keating, B.J., Rockett, K.A. and Kwiatkowski, D.P. (2003), '*In vivo* characterization of regulatory polymorphisms by allele-specific quantification of RNA polymerase loading', *Nature Genetics* Vol. 33, pp. 469–475.

54. Murrell, A., Heeson, S., Cooper, W.N. *et al.* (2004), 'An association between variants in the IGF2 gene and Beckwith-Wiedemann syndrome: Interaction between genotype and epigenotype', *Hum. Mol. Genet.* Vol. 13, pp. 247–255.

55. Rakyan, V.K., Blewitt, M.E., Druker, R. *et al.* (2002), 'Metastable epialleles in mammals', *Trends Genet.* Vol. 18, pp. 348–351.

56. Santourlidis, S., Trompeter, H.-I., Weinhold, S. *et al.* (2002), 'Crucial role of DNA methylation in determination of clonally distributed killer cell Ig-like receptor expression patterns in NK cells', *J. Immunol.* Vol. 169, pp. 4253–4261.

57. Trompeter, H.-I., Gomez-Lozano, N., Santourlidis, S. *et al.* (2005), 'Three structurally and functionally divergent kinds of promoters regulate expression of clonally distributed killer cell Ig-like receptors (KIR), of KIR2DL4, and of KIR3DL3', *J. Immunol.* Vol. 174, pp. 4135–4143.

58. Bjornsson, H.T., Fallin, M.D. and Feinberg, A.P. (2004), 'An integrated epigenetic and genetic approach to common human disease', *Trends Genet.* Vol. 20, pp. 350–358.

59. Murrell, A., Rakyan, V.K. and Beck, S. (2005), 'From genome to epigenome', *Hum. Mol. Genet.* Vol. 14(Suppl. 1), pp. R3–R10.

60. Hirschhorn, J.N. and Daly, M.J. (2005), 'Genome-wide association studies for common diseases and complex traits', *Nat. Rev. Genet.* Vol. 6, pp. 95–108.

61. Wang, W.Y.S., Barratt, B.J., Clayton, D.G. and Todd, J.A. (2005), 'Genome-wide association studies: Theoretical and practical concerns', *Nat. Rev. Genet.* Vol. 6, pp. 109–118.

62. Zondervan, K.T. and Cardon, L.R. (2004), 'The complex interplay among factors that influence allelic association', *Nat. Rev. Genet.* Vol. 5, pp. 89–100.

63. Pritchard, J.K. and Cox, N.J. (2002), 'The allelic architecture of human disease genes: Common disease – common variant... or not?', *Hum. Mol. Genet.* Vol. 11, pp. 2417–2423.

64. Reich, D.E. and Lander, E.S. (2001), 'On the allelic spectrum of human disease', *Trends Genet.* Vol. 17, pp. 502–510.

65. Blangero, J. (2004), 'Localization and identification of human quantitative trait loci: King Harvest has surely come', *Curr. Opin. Genet. Dev.* Vol. 14, pp. 233–240.

66. Nelson, G.W., Martin, M.P., Gladman, D. *et al.* (2004), 'Cutting edge: Heterozygote advantage in autoimmune disease: Hierarchy of protection/susceptibility conferred by HLA and killer Ig-like receptor combinations in psoriatic arthritis', *J. Immunol.* Vol. 173, pp. 4273–4276.

67. Hafler, D.A. and Jager, P.L.D. (2005), 'Applying a new generation of genetic maps to understand human inflammatory disease', *Nat. Rev. Immunol.* Vol. 5, pp. 83–91.

68. Hughes, A.L. (2002), 'Evolution of the human killer cell inhibitory receptor family', *Mol. Phylogenet. Evol.* Vol. 25, pp. 330–340.

69. Valentonyte, R., Hampe, J., Huse, K. *et al.* (2005), 'Sarcoidosis is associated with a truncating splice site mutation in BTNL2', *Nat. Genet.* Vol. 37, pp. 357–364.

70. Barcellos, L.F., Oksenberg, J.R., Begovich, A.B. *et al.* (2003), 'HLA-DR2 dose effect on susceptibility to multiple sclerosis and influence on disease course', *Am. J. Hum. Genet.* Vol. 72, pp. 710–716.

71. Orru, S., Giuressi, E., Carcassi, C. *et al.* (2005), 'Mapping of the major psoriasis-susceptibility locus (PSORS1) in a 70-kb interval around the corneodesmosin gene (CDSN)', *Am. J. Hum. Genet.* Vol. 76, pp. 164–171.

72. Horton, R., Wilming, L., Rand, V. *et al.* (2004), 'Gene map of the extended human MHC', *Nat. Rev. Genet.* Vol. 5, pp. 889–899.

73. Gupta, S. (2005), 'Molecular mechanisms of apoptosis in the cells of the immune system in human aging', *Immunol. Rev.* Vol. 205, pp. 114–129.

74. Hsu, H.-C., Scott, D.K. and Mountz, J.D. (2005), 'Impaired apoptosis and immune senescence — Cause or effect?', *Immunol. Rev.* Vol. 205, pp. 130–146.

75. Gough, S.C.L., Walker, L.S.K. and Sansom, D.M. (2005), 'CTLA4 gene polymorphism and autoimmunity', *Immunol. Rev.* Vol. 204, pp. 102–115.

76. Ueda, H., Howson, J.M.M., Esposito, L. *et al.* (2003), 'Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease', *Nature* Vol. 423, pp. 506–511.

77. Cepellini, R., Curtoni, E. and Mattiuz, P. *et al.* (1967), Genetics of Leukocyte Antigens: A Family study of Segregation and Linkage, Munksgaard, Copenhagen, Denmark.

78. Klein, J. (1986), 'Seeds of time: Fifty years ago Peter A. Gorer discovered the H-2 complex', *Immunogenetics* Vol. 24, pp. 331–338.

79. Awdeh, Z.L., Raum, D., Yunis, E.J. and Alper, C.A. (1983), 'Extended HLA/complement allele haplotypes: Evidence for T/t-like complex in man', *Proc. Natl. Acad. Sci. USA* Vol. 80, pp. 259–263.

80. Yunis, E.J., Larsen, C.E., Fernandez-Vina, M. *et al.* (2003), 'Inheritable variable sizes of DNA stretches in the human MHC: Conserved extended haplotypes and their fragments or blocks', *Tissue Antigens* Vol. 62, pp. 1–20.

81. Dawkins, R., Leelayuwat, C., Gaudieri, S. *et al.* (1999), 'Genomics of the major histocompatibility complex: Haplotypes, duplication, retroviruses and disease', *Immunol. Rev.* Vol. 167, pp. 275–304.