ANALYSIS

DISEASE MECHANISMS

Genetic insights into common pathways and complex relationships among immune-mediated diseases

Miles Parkes[1], Adrian Cortes[2]*, David A. van Heel[3] and Matthew A. Brown[2]*

Abstract | Shared aetiopathogenic factors among immune-mediated diseases have long been suggested by their co-familiality and co-occurrence, and molecular support has been provided by analysis of human leukocyte antigen (HLA) haplotypes and genome-wide association studies. The interrelationships can now be better appreciated following the genotyping of large immune disease sample sets on a shared SNP array: the 'Immunochip'. Here, we systematically analyse loci shared among major immune-mediated diseases. This reveals that several diseases share multiple susceptibility loci, but there are many nuances. The most associated variant at a given locus frequently differs and, even when shared, the same allele often has opposite associations. Interestingly, risk alleles conferring the largest effect sizes are usually disease-specific. These factors help to explain why early evidence of extensive 'sharing' is not always reflected in epidemiological overlap.

Immune-mediated diseases (IMDs) comprise a clinically heterogeneous group of disorders affecting ~3–5% of individuals of European origin. Knowledge of their pathogenic mechanisms remains limited, although genome-wide association studies (GWASs) have produced powerful new insights over the past 7 years. GWASs provide a hypothesis-free survey of the human genome for common variants associated with disease susceptibility[1,2]. The most recent genetic studies of IMDs have used GWAS meta-analysis allied to the Immunochip, which is a single-nucleotide polymorphism (SNP) microarray designed for deep replication and fine mapping[3] (BOX 1). Hundreds of IMD susceptibility loci have now been identified convincingly. Some are disease-specific and these often exert relatively large effects, but it is clear that many others affect the risk of several IMDs.

Overlapping aetiological factors among IMDs have long been suspected owing to shared clinical and immunological characteristics. In addition, individuals with one IMD have an increased risk of developing another, and familial aggregation of multiple IMDs is well established. The patterns are themselves instructive. Broadly, two clusters are identified: seropositive autoimmune diseases, such as coeliac disease, primary biliary cirrhosis (PBC), type 1 diabetes (T1D), autoimmune thyroid disease (AITD) and rheumatoid arthritis[4–6]; and seronegative auto-inflammatory conditions, including Crohn's disease, psoriasis and ankylosing spondylitis, in which autoantibodies are not a common or characteristic feature[7]. Despite the distinct pathogenic mechanisms implied by these subdivisions, shared susceptibility loci have been found across all of these phenotypes since the earliest GWASs.

In recognizing the overlap, the scientific literature has risked over-emphasizing the commonality of pathogenic mechanisms among IMDs. One question that has arisen is why, given that even across the serological divide there are many shared genetic risk loci, the co-occurrence of, for example, inflammatory bowel disease (IBD) and T1D in individuals and families is not more common. In this Analysis, we examine the output of several recent IMD genetic studies carried out using the Immunochip microarray, systematically identify 'pleiotropic' loci shared across IMDs, and discuss the complexities that are evident. We highlight loci that are phenotype-specific, and in particular flag shared loci where distinct variants are associated with different diseases or where the susceptibility allele is shared but in opposite directions. As well as highlighting key immune-regulatory genes, these loci may signpost pathways that are critical for organ selectivity.

In addition, we focus on major areas of interest as the field moves from GWASs to the post-GWAS era,

[1]*Inflammatory Bowel Disease Research Group, Department of Medicine, Addenbrooke's Hospital, University of Cambridge, Cambridge CB2 0QQ, UK.*
[2]*The University of Queensland Diamantina Institute, Translational Research Institute, Princess Alexandra Hospital, 37 Kent Road, Woolloongabba, Brisbane QLD 4102, Australia.*
[3]*Blizard Institute, Barts and The London School of Medicine, Queen Mary University of London, London E1 2AT, UK.*
**These authors contributed equally to this work.*
Correspondence to M.P.
e-mail: mp372@cam.ac.uk
doi:10.1038/nrg3502
Published online 6 August 2013

Deep replication
Attempted replication of a long list of single-nucleotide polymorphisms (SNPs) for which there is some evidence of association in index genome-wide association studies (GWASs) or GWAS meta-analyses. Most studies attempt to replicate a modest number of SNPs (5 to 100); Immunochip studies attempted this for the top 2,000 independent association signals for each phenotype.

Seropositive
The presence of antibodies that are directed against one or more of an individual's own proteins (known as autoantibodies). For example, there are anti-tissue transglutaminase antibodies in coeliac disease, and anticitrullinated peptide antibodies in rheumatoid arthritis.

where the goals move to defining causal variants and understanding their impact on transcriptional regulation and function in the relevant cell types. Important clues relating to environmental triggers and their interaction with genetically mediated mechanisms are described. We also discuss the increasing understanding of the allelic spectrum of IMD susceptibility and how these patterns overlay the complex relationship between different IMDs. The genetic clustering of clinically distinct phenotypes is quantified and discussed, and we reflect on the challenges that remain.

Association mapping and allelic architecture

The main goals of the Immunochip studies were the deep replication of GWAS data and fine mapping of confirmed susceptibility loci (BOX 1). A major incidental benefit has been the study of tens of thousands of patients with distinct IMDs using a common genotyping array. Loci included on an array owing to association with one IMD have been genotyped in many other IMDs, aiding comparisons and driving new gene discovery for many phenotypes. An unresolved question is what statistical threshold should be used to identify new associations, given the high prior probability of association for Immunochip SNPs compared to variants selected randomly from anywhere in the genome (for which $P < 5 \times 10^{-8}$ is conventionally used). For clarity in this Analysis we continue to use the GWAS significance threshold of $P < 5 \times 10^{-8}$, but this is clearly too conservative for Immunochip studies.

Notably, there is a strong correlation between sample size and the number of loci detected in both GWASs[8] and Immunochip studies (FIG. 1). The number of new associations identified through Immunochip for each disease has, unsurprisingly, correlated with the numbers of new samples genotyped (that is, those not previously assessed by GWASs). The differing data-set sizes among IMDs have produced variation in statistical power for drawing firm conclusions regarding associations, particularly with low-frequency risk alleles of modest effect. It can therefore be difficult to say with certainty that a particular phenotype is not associated with such loci, particularly in diseases in which only modest-size sample sets are available. Nevertheless, Immunochip has undoubtedly facilitated comparison of the molecular genetic architecture between IMDs, and has allowed a more thorough exploration of the extent and nature of the overlap, as discussed below.

Fine mapping. Relating to its 'fine mapping' goal, the use of Immunochip has narrowed the association signals to different extents across the various IMDs. Improved resolution has resulted from the use of tens of thousands of samples to define recombination rates and linkage disequilibrium (LD) blocks compared to the few hundred used in HapMap analyses, and from the genotyping of all identified variants within a region (rather than the condensed set present on GWAS chips). However, there are limitations to the Immunochip design, as indicated in BOX 1. These may affect some IMDs more than others, depending on the genetic architecture of each. For some loci the causal variant may not have been genotyped (particularly in cases in which it is a structural variant), and for others it will be impossible to identify using SNP genotyping and association mapping alone (for example, where up to 100 SNPs in tight LD show equivalent association signals). Resolution of causal variants in such instances is likely to require integration of expression and functional data derived from relevant cell types; for example, identifying where SNPs implicated by association analysis overlap regulatory motifs such as transcription factor binding sites in regions of open chromatin.

In coeliac disease, Immunochip has enabled approximately half of known signals to be narrowed to individual genes, and in some cases to subregions of genes. At almost all loci, a 'credible' set of polymorphisms that are likely to contain the causal variant (or variants) has been defined[24]. Notably, the risk attributable to these is not substantially greater than the original markers defined by GWASs. Investigators of rheumatoid arthritis have refined associations to 1 gene or subregion in 19 of 39 densely genotyped loci and, with additional bioinformatics analysis, provided suggestive evidence

Box 1 | An overview of Immunochip

Immunochip is an Illumina Infinium single-nucleotide polymorphism (SNP) microarray. The probes on this array interrogate 195,806 SNPs and 718 small insertion–deletions. The chip was designed in 2009 by investigators of 11 distinct autoimmune and inflammatory diseases[24,95].
There were two major goals:

- Deep replication of meta-genome-wide association studies (GWASs). The top 2,000 independent association signals for each disease were included.
- Fine mapping of GWAS loci. The content included all SNPs for which probes could be designed from the 1000 Genomes Project and resequencing data from various sources within all confirmed GWAS association intervals for each of the 11 immune-mediated diseases (IMDs). This was done without the filtering of SNPs on spacing or linkage disequilibrium, as had been used in earlier GWAS arrays. The chip also has dense coverage of the major histocompatibility complex (MHC) and killer immunoglobulin-like receptor (KIR) loci. Overall, 186 loci are densely covered with SNPs.

The cost of the chip was approximately 20% of that of contemporary GWAS chips, thus enabling far more subjects to be studied. As the number of loci identified strongly correlates with sample size[8], this has substantially increased the number of new IMD genes identified. In inflammatory bowel disease, for example, the total number of confirmed susceptibility loci has increased from 100 to 163, and in ankylosing spondylitis from 13 to 31. Genotyping using a uniform array has also enabled direct comparison across phenotypes much more readily than is possible using data from different GWAS arrays.

Limitations of Immunochip

- It was designed using early 1000 Genomes Pilot data (February 2010 release) and it has incomplete coverage, particularly of lower frequency variation. Some of this variation can be filled in using imputation of genotypes; in some cases such imputed SNP data will show a stronger association signal than that seen for genotyped SNPs.
- It is focused on SNP content rather than structural variants. The most-associated SNP may thus tag the causal variant rather than itself as being causal.
- Approximately 10% of SNPs failed assay design and could not be included on the array.
- The design was based on SNPs identified in individuals of European origin; variation present in non-European-origin populations is under-represented.
- Newly identified loci are not densely covered.

Development of a cost-effective, updated Immunochip is the subject of discussion within the IMD genetics community and will depend on demand.

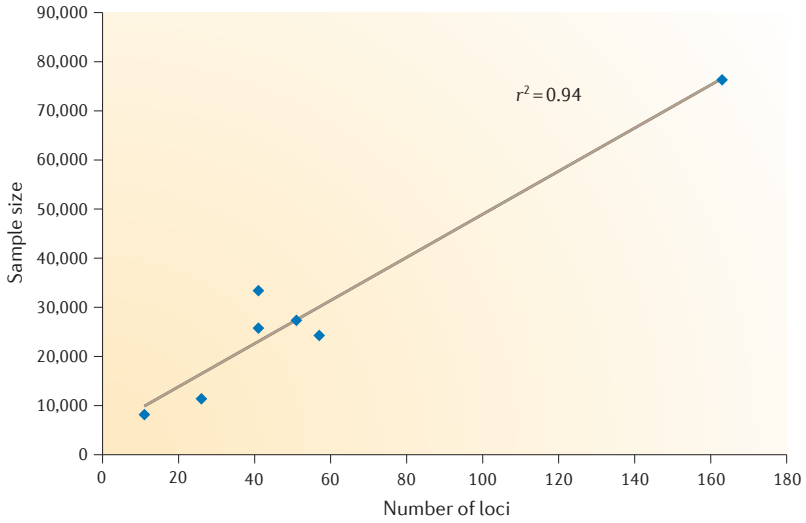


Figure 1 | **Relationship between sample size and the number of associated loci with genome-wide significance.** A plot of sample size (total number of cases plus controls) against the number of associated loci detected that reach the genome-wide significance threshold of $P < 5 \times 10^{-8}$, for Immunochip and genome-wide association studies (GWASs)[3,9,23–26]. Strong correlation is shown: $r^2 = 0.94$ when including a large inflammatory bowel disease study reported previously[25]; even if that large study is excluded the correlation remains strong, with $r^2 = 0.86$.

Seronegative
The absence of antibodies that are directed against one or more of an individual's own proteins (known as autoantibodies).

Auto-inflammatory
A disease, usually of mostly unknown aetiology, in which the immune response seems to be directed against 'self' antigens. The term may be used to include antibody-mediated disease as well as non-autoantibody-mediated pathology.

Pleiotropic
The phenomenon in which variants in a single gene are responsible for several distinct and seemingly unrelated phenotypic effects.

Linkage disequilibrium
(LD). The non-random association of two or more alleles. Alleles that are in LD are found together on the same haplotype more often than would be expected under a random combination of alleles. The pattern of LD in a given genomic region reflects the history of natural selection, mutation, recombination, genetic drift and other demographic and evolutionary forces.

for a single causal variant at 7 of these[9]. In ankylosing spondylitis, less than 10% narrowing of the region was observed for 4 of 8 loci; the resolution here being limited by extensive LD. Fine mapping remains an area of methodological development, and results for other IMDs are awaited[10,11].

Genetic architecture. Although notably few rare (minor allele frequency (MAF) of less than 0.5%) coding mutations causing IMDs have been identified so far, low-frequency variants (MAF of 0.5 to 5%) do contribute to the allelic spectrum of IMDs. Low-frequency and rare variants are generally not tagged by GWAS arrays but they often confer relatively large effect sizes and can both pinpoint likely causal genes among many potential candidates and help to provide functional information on pathogenic mechanisms. For example, early GWASs identified an association of Crohn's disease and ankylosing spondylitis with a locus containing caspase recruitment domain family, member 9 (*CARD9*)[12–14]. After resequencing, an independent deleterious low-frequency splice-site variant in *CARD9* was identified[15]. This variant is protective against Crohn's disease, ulcerative colitis and ankylosing spondylitis, whereas the common variant that was detected in GWASs (and that is not in LD with the low-frequency variant) increases risk and correlates with increased *CARD9* transcription in expression quantitative trait locus (eQTL) analyses[15]. Similar approaches have been successful in resolving interferon induced with helicase C domain 1 (*IFIH1*) as a causal gene in T1D[16], and tumour necrosis factor alpha-induced protein 3 (*TNFAIP3*) in systemic lupus erythematosus (SLE)[17].

In rare cases, highly penetrant mutations within loci identified by GWASs cause a familial or Mendelian disease, which phenotypically overlaps an IMD, further illustrating the spectrum of immune diseases. For example, mutations in the three prime repair exonuclease 1 (*TREX1*) gene cause Aicardi–Goutieres syndrome, which resembles cerebral lupus[18]; insulin (*INS*) mutations lead to neonatal diabetes mellitus[19]; nucleotide-binding oligomerization domain-containing 2 (*NOD2*) mutations distinct from those associated with Crohn's disease cause a systemic granulomatous inflammation termed Blau syndrome[20]; and interleukin-10 receptor beta (*IL10RB*) mutations lead to severe intestinal inflammation in infancy[21], echoing the confirmed association of Crohn's disease and ulcerative colitis to the *IL10* gene locus[22].

Distinct haplotypes within IMD susceptibility loci may confer independent effects. For some loci, the risk that can be attributed to them has increased markedly with the identification of such signals, as seen in the regression-based analysis of the association of *IL12A* with PBC[23]. Further Immunochip analyses have identified multiple associated haplotypes in roughly one-third of GWAS loci. In rheumatoid arthritis, 6 of 45 tested loci have multiple independent effects[9], and this is also true for 14 of 40 loci (including human leukocyte antigen (HLA) loci) in coeliac disease (5 of these have 3 or more risk haplotypes)[24]. Interestingly, where there was more than one association signal at a shared locus, in no instance was one signal concordant and the other discordant in our analysis: they were always either in the same direction (both concordant or both discordant) or one was correlated and the other not.

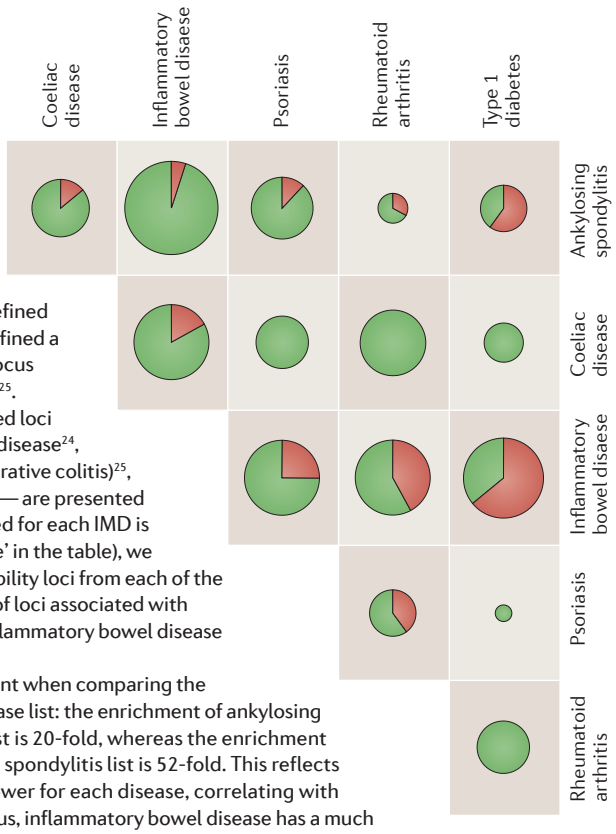
Loci that overlap between diseases

Although pleiotropy has long been recognized, Immunochip studies have highlighted subtleties that were not immediately evident from comparing GWAS results alone. Here we illustrate these subtleties by considering in detail Immunochip results for six key IMDs with large sample sets analysed so far: namely, ankylosing spondylitis[3], coeliac disease[24], IBD[25], psoriasis[26], rheumatoid arthritis[9] and T1D (J. Todd and S. Rich, personal communication) (see Supplementary information S1 (figure) for a summary of the results, Supplementary information S2 (box) for a description of our analytical methods, and Supplementary information S3 (table) for full results). Although sharing is common — 71 loci show associations at $P < 5 \times 10^{-8}$ with two or more of these diseases — it is also complex. Thus, overlap can comprise a shared locus for which the same SNP (or same haplotype) confers increased risk for more than one disease ('correlated and concordant'); a shared locus for which the same haplotype increases risk for one disease but is protective for another ('correlated but discordant'); or a shared locus for which different haplotypes are implicated ('non-correlated').

Many examples were found in each category. Pairwise, there were 416 instances of overlap: 45% were concordant, 14% discordant and 42% were not correlated. The extent of sharing and the degree of concordance depends on which diseases are compared

Box 2 | **Extent and direction of sharing of loci between six immune-mediated diseases**

Reference disease	Number of confirmed loci	Fold enrichment (number of shared loci)					
		Ankylosing spondylitis	Coeliac disease	Inflammatory bowel disease	Psoriasis	Rheumatoid arthritis	Type 1 diabetes
Ankylosing spondylitis	31	-	65 (9)	52 (20)	66 (9)	14 (5)	21 (5)
Coeliac disease	40	60 (9)	-	33 (20)	59 (10)	43 (10)	42 (9)
Inflammatory bowel disease	163	20 (20)	17 (20)	-	17 (21)	13 (22)	12 (19)
Psoriasis	36	63 (9)	61 (10)	39 (21)	-	14 (5)	8 (4)
Rheumatoid arthritis	48	12 (5)	41 (10)	23 (22)	13 (5)	-	22 (8)
Type 1 diabetes	43	20 (5)	41 (9)	19 (19)	6 (4)	23 (8)	-



The extent of sharing of loci between immune mediated diseases (IMDs) can be quantified by the number of overlapping loci and by an enrichment score. The enrichment score quantifies an enrichment odds ratio of one set of genomic intervals in a second set of genomic intervals relative to the genome as a whole. A genomic interval was defined as 250 kilobases (kb) on either side of each lead single-nucleotide polymorphism (SNP) (as only data for lead SNPs and not summary data for all SNPs in the loci were available to us, the boundaries of the loci had to be defined by distance), and overlapping intervals in two IMDs defined a shared locus. Enrichment scores and definition of a locus were calculated using a method described previously[25].

Pairwise enrichment scores and the number of shared loci between six IMDs — ankylosing spondylitis [3], coeliac disease[24], inflammatory bowel disease (Crohn's disease and ulcerative colitis)[25], psoriasis[26], rheumatoid arthritis [9] and type 1 diabetes — are presented above (see the table). The total number of loci reported for each IMD is also given. For each IMD (shown as 'Reference disease' in the table), we computed the fold enrichment observed for susceptibility loci from each of the other five IMDs (columns). Thus, for example, the list of loci associated with ankylosing spondylitis has a 52-fold enrichment of inflammatory bowel disease loci and a 21-fold enrichment of type 1 diabetes loci.

The enrichment score is not symmetric. This is evident when comparing the enrichment scores from the inflammatory bowel disease list: the enrichment of ankylosing spondylitis loci in the inflammatory bowel disease list is 20-fold, whereas the enrichment of inflammatory bowel disease loci in the ankylosing spondylitis list is 52-fold. This reflects the substantially different sample sizes and study power for each disease, correlating with different numbers of confirmed susceptibility loci. Thus, inflammatory bowel disease has a much larger number of confirmed loci, producing a large denominator for all comparisons with this disease; hence, enrichment scores in the inflammatory bowel disease row seem generally low.

A second approach to quantify the extent of sharing of loci between two IMDs is to interrogate whether the reported index SNPs are correlated and, if so, to determine whether effects are concordant or discordant between the IMDs. The figure contains pairwise comparisons between the six IMDs considered, represented as pie charts (see the figure). Each pie chart is sized in proportion to the number of shared loci with correlated signals: green sectors represent concordant associations and red sectors represent discordant associations. See Supplementary information S1 (figure) for graphical representations of disease pairwise comparisons with annotations.

Minor allele frequency
In a population, the frequency of the less common allele of a genetic variant. It has a value of between 0 and 0.5, and may vary among populations.

(see Supplementary information S1,S3 (figure,table)). For example, between IBD and ankylosing spondylitis, the lead SNP was correlated at 19 out of 20 shared loci, whereas for IBD and T1D, 19 loci were shared but the signal was not correlated for 6 of these; and for the 13 loci that shared the same lead SNP the signal was discordant at 8 of them.

To aid interpretation, we have quantified the degree to which one IMD is enriched for loci shared by another (BOX 2). Although scores for IBD using this method are

depressed by the large number of loci identified for this disease, the scores do highlight the close relationship between the seronegative diseases Crohn's disease, ulcerative colitis, ankylosing spondylitis and psoriasis. In addition, they show enrichment between coeliac disease and these conditions, perhaps hinting that immunological interactions in the gut contribute to them all. There is a surprising lack of correlation between the two rheumatological conditions ankylosing spondylitis and rheumatoid arthritis; these diseases only share five loci, three with the same SNP implicated but correlated at just one of these. The enrichment score between them is among the lowest for any pairs, which suggests that distinct pathogenic pathways lead to joint inflammation in each case. So far, most drugs for ankylosing spondylitis are selected for trial owing to previously demonstrated efficacy in the treatment of rheumatoid arthritis; these genetic findings underline the importance of disease-specific drug development based on a true understanding of pathogenic mechanisms.

Although future cross-phenotype studies using genotype-level data will facilitate further understanding of the overlap, the analysis we have performed provides interesting insights. Overall, a surprisingly limited range of genes and pathways are found to contribute substantially to IMD risk. We briefly discuss a few of these below and summarize more of the pleiotropic genes and pathways and their disease associations in TABLE 1. The examples discussed have been selected for their importance, their mechanistic interest, their contemporary relevance derived from recent Immunochip studies and the increasingly evident complexity of their effects in different IMDs.

Major histocompatibility complex. The major histocompatibility complex (MHC) is an extremely gene-dense region with long-range LD and hundreds of immunologically active genes. MHC loci were the first reported genetic associations with IMDs and, for most of the diseases, remain the strongest. Seropositive diseases are typically associated with HLA class II alleles: several classical autoimmune phenotypes (including coeliac disease, T1D, AITD, SLE and others)[27] are associated with the HLA-DR3-DQ2 haplotype; rheumatoid arthritis is associated with different alleles of HLA-DRB1 (REF. 28). By contrast, the seronegative diseases are generally associated with HLA class I alleles, and these tend to be disease-specific. For example, ankylosing spondylitis is associated with HLA-B27 (REF. 29–31), psoriasis with HLA-Cw6 (REF. 32), and Behçet's disease with HLA-B51 (REF. 33). A notable exception is Crohn's disease; although it is associated with SNPs in the MHC class I region, it is not associated with any classical HLA allele[25].

Association of individual diseases with multiple MHC genes has long been suspected. However, the strength of the MHC associations and the extensive LD that they demonstrate has frustrated the dissection of these associations. Using large data sets, dense SNP typing and improved analytical methods — including imputation of HLA alleles — multiple independent MHC signals have recently been reported in multiple sclerosis [34], T1D[35], rheumatoid arthritis[36], psoriasis[37] and ankylosing spondylitis[3]. For example, in addition to a very strong association between HLA-B27 and ankylosing spondylitis, HLA imputation studies using Immunochip data identified a further independent signal at HLA-A*0201 (REF. 3). Similarly, alongside the strong HLA class II association with T1D, an independent association of HLA-B*39 with this disease was found[35]. Recently, by imputing classical HLA alleles and amino-acid polymorphisms, the so-called 'shared epitope' HLA-DRB1 alleles associated with rheumatoid arthritis were refined to three amino acids within the peptide-binding groove, and further associations were identified at HLA-B and HLA-DPB1 (REF. 36). Many more such independent MHC associations are likely to emerge soon.

Although most associations are with classical HLA alleles, it is likely that non-MHC genes in the MHC region also contribute to IMD heritability. Rare variant studies may help to distinguish the contributions of these polymorphisms from those of common haplotypes, as has been achieved for associations between complement genes and SLE[38].

IL-23 and T_H1 pathways. IL-23 and T helper 1 (T_H1) pathways have a key role in regulating the immune response to exogenous antigens, and the contribution of the IL-23 pathway to IMD pathogenesis has been particularly highlighted by genetics studies (FIG. 2; TABLE 1). In health, the IL-23 pathway mediates defence against extracellular bacteria and antimicrobial responses at epithelial surfaces, such as the skin and gut. Early GWASs demonstrated an association between the IL-23 receptor (*IL23R*) coding variant Arg381Gln (rs11209026) and several seronegative IMDs[2,39,40]. Many other variants in this gene were later found to have independent IMD associations.

Although T_H17 lymphocytes (which are $CD4^+$ and produce IL-17) were initially assumed to be the main effectors, several distinct IL-23-responsive IL-17-producing cell types are now implicated. More than one cell type may contribute in different contexts[41]. For many diseases, association with *IL23R* was the first indication that IL-23 and the downstream cytokines IL-17 and IL-22 were involved in their pathogenesis. This finding has led to the development of targeted cytokine inhibitor therapies that are now in widespread clinical use — an early success for the clinical translation of genetic findings.

Many other genes in the IL-23 response pathway have been implicated in IMDs (FIG. 2). Perhaps unsurprisingly, pleiotropic effects are widespread. For example, common variants at the tyrosine kinase 2 (*TYK2*) locus are associated with ankylosing spondylitis, psoriasis, Crohn's disease, ulcerative colitis, T1D, multiple sclerosis, rheumatoid arthritis and PBC — interestingly, with little overlap in the specific allele implicated. Rare *TYK2* polymorphisms add to the complexity: a variant predicted to influence RNA splicing is associated with ankylosing spondylitis with a high odds ratio of 7.7, and a distinct non-synonymous *TYK2* variant predicted to affect phosphorylation activity is associated with multiple sclerosis[42,43].

HLA class II
(Human leukocyte antigen class II). Part of the human chromosome 6 major histocompatibility complex region that encodes HLA-DP, -DQ and -DR alleles. These are expressed on professional antigen-presenting cells and present antigens from extracellular proteins such as those derived from pathogens to $CD4^+$ T cells.

HLA class I
(Human leukocyte antigen class I). Part of the human chromosome 6 major histocompatibility complex region that encodes HLA-A, -B and -C. These molecules can present antigen from inside the cell (including virally encoded proteins) to other immune cells, and are present on most cell types.

Shared epitope
The hypothesis that a subregion of the human leukocyte antigen DR (HLA-DR) molecule involved in peptide presentation is important in rheumatoid arthritis pathogenesis.

T_H17 lymphocytes
(T helper 17 lymphocytes). A subset of $CD4^+$ T helper cells that produce interleukin-17 (IL-17) and that are thought to be important in inflammatory and autoimmune diseases.

Table 1 | **Examples of genes and pathways that are associated with two or more immune-mediated diseases***

Pathway or genes	Positional candidate genes shared by ≥2 diseases (cytogenetic position)‡	Diseases associated with this pathway or ≥1 of these loci (genes)	Diseases for which the main signal is discordant or not correlated with the others (gene)
IL-23 and T_H1	*IL23R* (1p31), *IL12B* (5q33), *IL12A* (3q25), *TYK2* (19p13), *JAK2* (9p24), *STAT3* (17q21), *STAT4* (2q32), *IL27* (16p11) and *CCR6* (6q27)	Ankylosing spondylitis (*IL23R*, *IL12B*, *TYK2*, *JAK2*, and *IL27*); IBD (*IL23R*, *IL12B*, *TYK2*, *JAK2*, *STAT3*, *STAT4* and *IL27*); psoriasis (*IL23R*, *IL12B*, *TYK2* and *STAT3*); coeliac disease (*IL12A* and *STAT4*); rheumatoid arthritis (*TYK2*, *STAT4* and *CCR6*); T1D (*TYK2*); SLE (*TYK2*, *STAT4* and *IL27*); and multiple sclerosis (*IL12B*, *IL12A*, *TYK2* and *STAT3*)	Psoriasis (*IL12B*); psoriasis and rheumatoid arthritis (*TYK2*); multiple sclerosis (*STAT3* and TYK2); IBD and rheumatoid arthritis (*STAT4*)
NF-κB	*REL* (2p16), *TNFAIP3* (6q23), *NFKB1* (4q24) and *TNIP1* (5q32)	IBD (*REL*, *TNFAIP3* and *NFKB1*); psoriasis (*REL*, *TNFAIP3*, *NFKB1* and *TNIP1*); coeliac disease (*REL* and *TNFAIP3*); rheumatoid arthritis (*REL* and *TNFAIP3*); T1D (*TNFAIP3*); SLE (*TNFAIP3* and *TNIP1*); and multiple sclerosis (*NFKB1*)	Psoriasis (*REL* and *TNFAIP3*); rheumatoid arthritis (*REL*) and SLE (*TNFAIP3*)
Aminopeptidase	*ERAP1* (5q15) and *ERAP2* (5q15)	Ankylosing spondylitis (*ERAP1* and *ERAP2*); IBD (*ERAP2*); and psoriasis (*ERAP1* and *ERAP2*)	-
IL-2 and IL-21	*IL2*, *IL21*§ (4q26), *IL2RA* (10p15) and *IL2RB* (22q13)	IBD (*IL2*, *IL21* and *IL2RA*); coeliac disease (*IL2*, *IL21*); rheumatoid arthritis (*IL2*, *IL21*, *IL2RA* and *IL2RB*); T1D (*IL2*, *IL21*, *IL2RA* and *IL2RB*); and multiple sclerosis (*IL2RA*)	IBD (*IL2*, *IL21*) and coeliac disease (*IL2RA*)
IRF family	*IRF4* (6p25), *IRF5* (7q32), *IRF7* (11p15) and *IRF8* (16q24)	IBD (*IRF5* and *IRF8*); psoriasis (*IRF4*); coeliac disease (*IRF4*); rheumatoid arthritis (*IRF5* and *IRF8*); SLE (*IRF5*, *IRF7* and *IRF8*); and multiple sclerosis (*IRF8*)	Psoriasis (*IRF4*) and rheumatoid arthritis (*IRF8*)
T-cell co-stimulation	*CD40* (20q12), *CD28*, *CTLA4*, *ICOS*§ (2q33) and *ICOSLG* (21q22)	Ankylosing spondylitis (*ICOSLG*); IBD (*ICOSLG*); coeliac disease (*CD28*, *CTLA4* and *ICOSLG*); rheumatoid arthritis (*CD40*, *CD28* and *CTLA4*); T1D (*CD28*, *CTLA4*); and multiple sclerosis (*CD40*)	Coeliac disease (*ICOSLG*) and IBD (*CD40*)
PTPN2, *PTPN22*	*PTPN2* (18p11) and *PTPN22* (1p13)	IBD (*PTPN2* and *PTPN22*); coeliac disease (*PTPN2*); rheumatoid arthritis (*PTPN22*); T1D (*PTPN2* and *PTPN22*); and SLE (*PTPN22*)	Crohn's disease (*PTPN22*)
Ubiquitylation	*UBE2L3* (22q11)	Ankylosing spondylitis, IBD, psoriasis, coeliac disease, rheumatoid arthritis, SLE and multiple sclerosis	-
Viral response	*IFIH1* (2q24)	IBD, psoriasis, T1D and SLE	IBD
Other	*IL10* (1q32)	IBD, T1D and SLE	T1D
	IL18RAP (2q12),	IBD, coeliac disease and T1D	-
	FCGR2A (1q23),	Ankylosing spondylitis, IBD (ulcerative colitis), rheumatoid arthritis, T1D, SLE and multiple sclerosis	Ulcerative colitis
	PTGER4 (5p13)	Ankylosing spondylitis, IBD and multiple sclerosis	-
	BACH2 (6q15),	Ankylosing spondylitis, IBD, coeliac disease, T1D and multiple sclerosis	IBD
	CARD9 (9q34),	Ankylosing spondylitis and IBD	-
	ZMIZ1 (10q22),	IBD, psoriasis, coeliac disease and multiple sclerosis	-
	YDJC (22q11),	IBD, psoriasis, coeliac disease, rheumatoid arthritis and SLE	-
	TAGAP (6q25)	IBD, psoriasis, coeliac disease, rheumatoid arthritis, T1D and multiple sclerosis	Rheumatoid arthritis, T1D
	PRDM1 (6q21)	IBD, rheumatoid arthritis and SLE	Rheumatoid arthritis, SLE

BACH2, BTB and CNC homology 1, basic leucine zipper transcription factor 2; *CARD9*, caspase recruitment domain family, member 9; *CCR6*, CC chemokine receptor 6; *CTLA4*, cytotoxic T lymphocyte-associated protein 4; *ERAP1*, endoplasmic reticulum aminopeptidase 1; *FCGR2A*, Fc fragment of IgG, low affinity IIa, receptor (CD32); IBD, inflammatory bowel disease; *ICOS*, inducible T-cell co-stimulator; *ICOSLG*, ICOS ligand; *IFIH1*, interferon induced with helicase C domain 1; *IL18RAP*, interleukin-18 receptor accessory protein; *IL23R*, interleukin-23 receptor; IRF, interferon regulatory factor; *JAK2*, Janus kinase 2; *NFKB1*, nuclear factor of kappa light polypeptide gene enhancer in B cells 1; *PRDM1*, PR domain containing 1, with ZNF domain; *PTGER4*, prostaglandin E receptor 4 (subtype EP4); *PTPN2*, protein tyrosine phosphatase, non-receptor type 2; SLE, systemic lupus erythematosus; *STAT3*, signal transducer and activator of transcription 3; T1D, type 1 diabetes; *TAGAP*, T cell activation RhoGTPase activating protein; T_H1, T helper 1; *TNIP1*, TNFAIP3 interacting protein 1; *TNFAIP3*, tumour necrosis factor alpha-induced protein 3; *TYK2*, tyrosine kinase 2; *UBE2L3*, ubiquitin-conjugating enzyme E2L 3; *ZMIZ1*, zinc finger, MIZ-type containing 1.*For more details and references of the individual studies see the ImmunoBase website. ‡The lead positional candidate gene for the respective association interval (identified by genome-wide association studies and interrogated on Immunochip) is shown. Note that in some cases these genes have not yet been confirmed as causal and that many loci contain several candidate genes. §Two of the loci highlighted here contain more than one strong candidate gene: *IL2* and *IL21*; and *CD28*, *CTLA4* and *ICOS*.

IL-12 and IL-23 cytokines and receptors, which crucially influence T_H1 and T_H17 cell differentiation, respectively, are closely related and their pathways share many components (FIG. 2). In addition, the *IL23R* and *IL12RB2* genes are adjacent, and eQTL data suggest that IMD-associated variants at this locus can affect the expression of both genes. Furthermore, *IL12B,* which is associated with several seronegative IMDs, including IBD, psoriasis, multiple sclerosis and ankylosing spondylitis, encodes the p40 component that is shared between IL-12 and IL-23 heterodimers. Both signal transducer and activator of transcription 3 (*STAT3*)

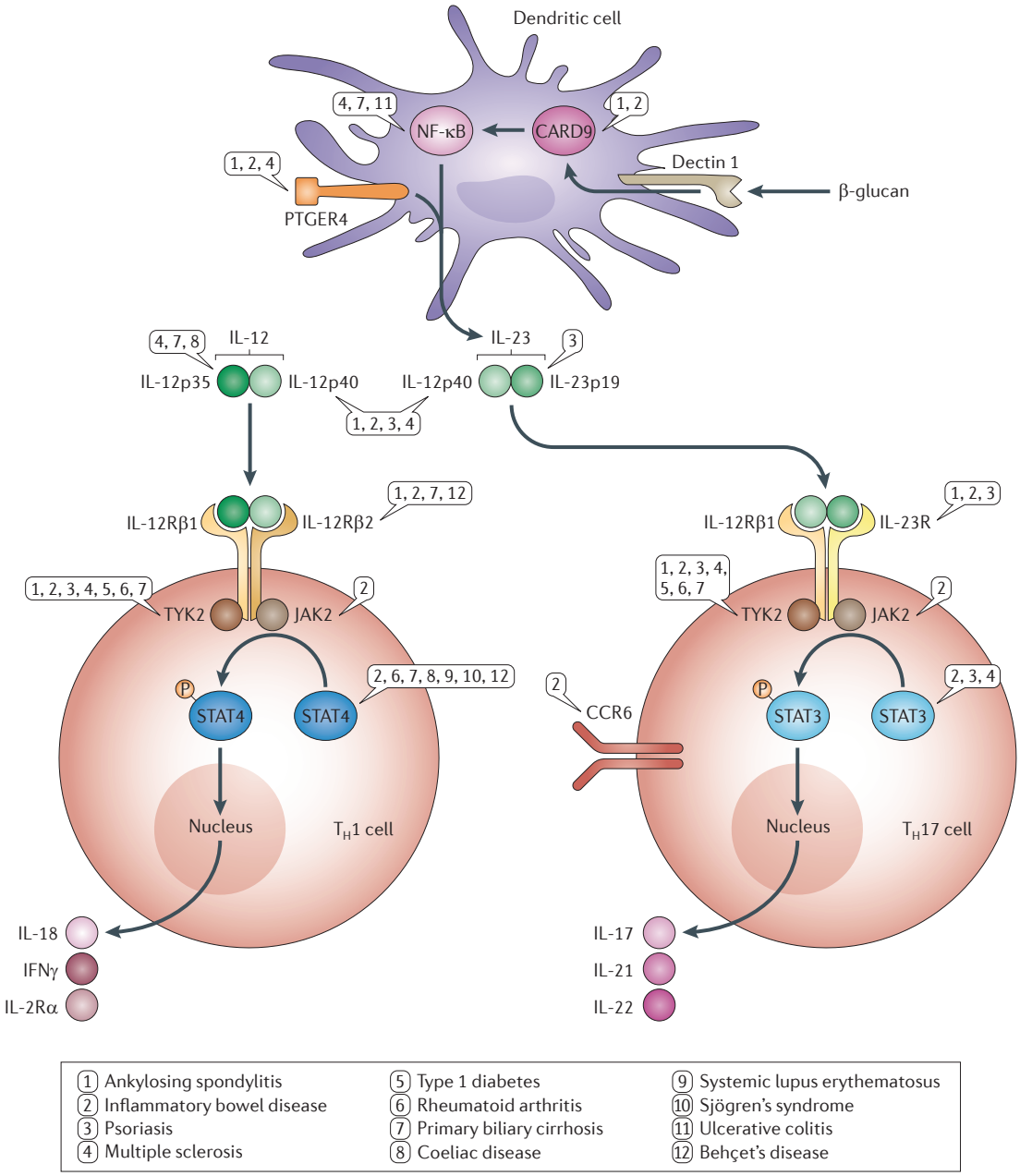


Figure 2 | **Associations of components of the IL-23R response pathway with immune-mediated diseases.** Multiple components of the interleukin-23 receptor (IL-23R) response pathway show genome-wide significant associations with a broad range of immune-mediated diseases (IMDs); that is, these components map within genomic intervals that are associated with the respective disease in genome-wide association studies or Immunochip analyses with $P < 5 \times 10^{-8}$. Although this figure shows these components in the conventional context of T helper 1 (T_H1) and T_H17 lymphocytes, it is now recognized that they are widely expressed in innate lymphoid cells and also atypical lymphocytes. All of these cell types are candidates for their site of action in IMD predisposition, and the specific cell type may vary from phenotype to phenotype. CARD9, caspase recruitment domain-containing protein 9; CCR6, CC chemokine receptor 6; IFNγ, interferon-γ; JAK2, Janus kinase 2; NF-κB, nuclear factor-κB; PTGER4, prostaglandin E2 receptor EP4 subtype; STAT3, signal transducer and activator of transcription 3; TYK2, tyrosine kinase 2.

and *STAT4*, which are expressed in T_H17 and T_H1 lymphocytes, respectively, show notable and overlapping pleiotropic effects (FIG. 2). Interestingly, there are three independent *STAT4* signals for coeliac disease, and a further independent signal at this locus is shared by Crohn's disease and rheumatoid arthritis, but it has discordant effects.

Nuclear factor-κB and its regulators. Nuclear factor-κB (NF-κB) is a highly conserved and ubiquitously expressed master transcriptional regulator of inflammation. It is constitutively present in the cytoplasm and its canonical pathway is activated rapidly by a variety of danger signals, including cytokines (TNF, IL-1β) and microbial motifs (through Toll-like receptors (TLRs) and NOD-like receptors). GWAS meta-analyses and Immunochip studies have identified associations between several NF-κB components and multiple IMDs. Although individual effect sizes are modest, NF-κB components have been highlighted by the enhanced power of bioinformatic techniques to detect patterns in association data and of network analyses to identify causal genes from among many positional candidates, when using very large Immunochip data sets[25,44].

Some susceptibility genes cluster with seropositive or seronegative IMDs, but susceptibility genes involved in the NF-κB pathway are truly pleiotropic. Of the five proteins in the human NF-κB family, loci encompassing the *REL* and *NFKB1* genes have shown the most IMD associations so far: *REL* with rheumatoid arthritis, coeliac disease, psoriasis, Crohn's disease and ulcerative colitis; and *NFKB1* with multiple sclerosis, psoriasis and ulcerative colitis. Genes that are directly involved in NF-κB regulation, for example, by ubiquitylation of NF-κB, have also been widely implicated. For example, ubiquitin-conjugating enzyme E2L 3 (*UBE2L3*), which encodes a protein that ubiquitylates the NF-κB1 precursor p105 to target it for degradation, is associated with rheumatoid arthritis, coeliac disease, SLE, multiple sclerosis, ankylosing spondylitis and Crohn's disease; and the *TNFAIP3* locus is associated with rheumatoid arthritis, coeliac disease, T1D, SLE, scleroderma, psoriasis, Crohn's disease and ulcerative colitis.

TNFAIP3 (also known as A20) is a ubiquitin-editing enzyme with potent anti-inflammatory properties mediated by NF-κB inhibition[45]. A20 expression is induced rapidly by TNF, microbial TLR ligands, NOD-like receptors and other NF-κB-dependent signals to provide important negative feedback. Although associated with many IMDs, *TNFAIP3* shows substantial allelic heterogeneity. The same non-coding variant seems to contribute to rheumatoid arthritis, coeliac disease, T1D and ulcerative colitis, but this is distinct from the lead psoriasis SNP, and SLE and scleroderma are associated with another independent signal, a non-synonymous coding variant known to impair A20 function. In addition, *TNIP1*, which encodes a TNFAIP3-interacting protein, is associated with psoriasis and SLE.

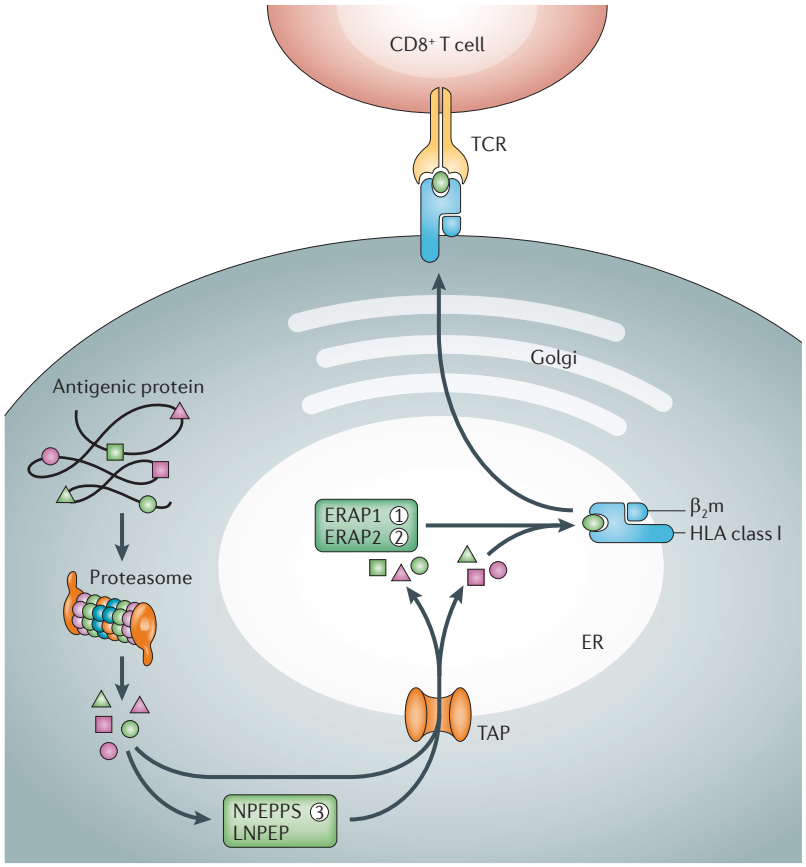
Functional data relating to *TNFAIP3* are interesting: various IMD phenotypes result from conditional knockouts in mice. Animals lacking A20 in myeloid lineages develop an erosive polyarthritis similar to rheumatoid arthritis[46], whereas deletion of A20 in dendritic cells produces T-cell-dependent colitis, ankylosing spondylitis and enthesitis[47]. Mice with A20-deficient B cells spontaneously develop an autoimmune condition analogous to SLE, with increased plasma cells, autoantibodies and renal immunoglobulin deposition[48]. A20 thus seems to have a central role in immune homeostasis, and its cell-specific deletion triggers inflammatory phenotypes closely matching those seen in at least some of the diseases genetically associated with *TNFAIP3*.

Interferon regulatory factor proteins. Another group of proteins whose contribution to IMD pathogenesis has been highlighted by Immunochip studies is the interferon regulatory factor (IRF) family. These transcription factors have diverse roles, including virus-mediated activation of interferon, immunoregulation and modulation of cell growth, differentiation and apoptosis. The locus encoding IRF4, which is widely expressed in the cells of the immune system and regulates both IL-17A and IL-21 production, has been associated with psoriasis and coeliac disease, whereas distinct loci for *IRF5* and *IRF8* have been associated with PBC, rheumatoid arthritis, scleroderma, multiple sclerosis, Crohn's disease and ulcerative colitis. IRF5 polarizes macrophages towards the M1 microbicidal phenotype. It also directly activates transcription of *IL12A*, *IL12B* and *IL23A*, and represses *IL-10*, hence priming a potent T_H1–T_H17 response[49]. *IRF8* is critical for the development of monocytes and dendritic cells, and mutations in this gene lead to human primary immunodeficiencies and loss of antimycobacterial immunity[50]. These associations have raised interest in the overlap among loci for IMD susceptibility, immunodeficiency and risk of infection, as discussed below. In addition, *IRF1* (among other strong positional candidates) has been associated with psoriasis, ulcerative colitis and Crohn's disease.

Aminopeptidases. Aminopeptidases catalyse the cleavage of amino-terminal amino acids from peptides and, among other functions, trim peptides to the optimal length for HLA class I presentation (FIG. 3). Variants in several aminopeptidase genes — endoplasmic reticulum aminopeptidase 1 (*ERAP1*), *ERAP2*, leucyl/cystinyl aminopeptidase (*LNPEP*) and aminopeptidase puromycin sensitive (*NPEPPS*) — are associated with ankylosing spondylitis[2], and there are concordant associations of *ERAP1* with psoriasis[51], and of *ERAP2* with IBD[52].

The *ERAP1* variants show protective association with ankylosing spondylitis and psoriasis. They reduce cleavage of peptides *in vitro* and reduce HLA class I expression. The *ERAP2* allele is also protective for ankylosing spondylitis and Crohn's disease; it causes splice site variation, decay of *ERAP2* mRNA and the absence of translated protein[53]. Together, these findings suggest that reduced peptide presentation by HLA class I molecules is protective for seronegative diseases. *ERAP2* and, to a lesser extent, *ERAP1* are subject to balancing selection[53,54] (as are the HLA loci). This

T_H1 lymphocytes
(T helper 1 lymphocytes). A subset of T helper cells that produce interferon-γ (and other cytokines) and that activate macrophages.



① Ankylosing spondylitis, psoriasis and Behçet's disease
② Ankylosing spondylitis, inflammatory bowel disease and psoriasis
③ Ankylosing spondylitis

Figure 3 | **Associations of components of the major histocompatibility complex class I antigen processing pathway with immune-mediated diseases.** Antigenic proteins are degraded in the proteasome and trimmed by cytoplasmic aminopeptidases (namely, aminopeptidase puromycin sensitive (NPEPPS) and leucyl/cystinyl aminopeptidase (LNPEP)) before delivery by 'transporter associated with antigen processing' (TAP) into the endoplasmic reticulum (ER). Here they are trimmed further by endoplasmic reticulum aminopeptidase 1 (ERAP1) and ERAP2 to achieve the optimal length for loading into human leukocyte antigen (HLA) class I molecules, before presentation on the cell surface. Single-nucleotide polymorphisms within or adjacent to the specific aminopeptidase genes are associated at $P < 5 \times 10^{-8}$ with susceptibility to seronegative immune-mediated diseases ankylosing spondylitis[1,3], psoriasis[51], Behçet's disease[96] and inflammatory bowel disease[52], with strong evidence of interaction between the locus encoding ERAP1 and respective HLA susceptibility alleles for ankylosing spondylitis[14] and psoriasis[51]. β2m, β2-microglobulin; TCR, T cell receptor.

probably reflects the need to preserve genetic diversity in key immune-response genes.

An interesting mechanistic observation reported in the index GWASs was evidence of gene–gene interaction between *ERAP1* and *HLA-B27* in ankylosing spondylitis[14], and between *ERAP1* and *HLA-Cw6* in psoriasis [51]. These represent the most robust examples of epistasis reported in common human diseases so far. The finding suggests that the mechanisms by which *HLA-B27* induces ankylosing spondylitis, and *HLA-Cw6* induces psoriasis, are closely related.

The IL-2 and IL-21 T cell activation pathway. The cytokines IL-2 and IL-21 regulate the growth, proliferation and differentiation of T cells and also regulate thymic T cell maturation. The approximately 400-kilobase (kb) region on chromosome 4q27 encoding *IL2* and *IL21* has been associated with coeliac disease, T1D, ulcerative colitis and rheumatoid arthritis. It has unusually strong LD and, even though tens of thousands of individuals have been studied, is has not been possible to narrow the association signal to either gene[24].

Expression of IL-2R is a defining feature of $CD4^+$ regulatory T cells, but IL-2R is also expressed on $CD4^+$ memory T cells and can be upregulated following stimulation of several innate immune cells. Several haplotypes at the *IL2RA* locus, which encodes one IL-2R subunit, are associated with IBD, T1D, multiple sclerosis, rheumatoid arthritis and vitiligo. *IL2RB*, which encodes the other subunit, is associated with rheumatoid arthritis and T1D. The functional role of *IL2RA* variants has been dissected in detail in human leukocyte samples[55]. Correlation between distinct T1D-associated haplotypes and surface expression of IL-2RA is complex and cell-type specific — issues that may be mirrored for other IMD-associated genes.

Discordant associations

A notable feature that is apparent from the Immunochip data is the number of loci shared between IMDs for which the same SNP shows the strongest association, but in opposite directions (BOX 2; TABLE 1; see Supplementary information S1,S3 (figure,table)). This discordancy is particularly common between seronegative and seropositive diseases, and may help to explain the lack of obvious familial clustering across this divide. Thus, variants at each of *IL27*, *IL10*, *STAT3*, *CD40* and *FCGR2A* (Fc fragment of IgG, low affinity IIa, receptor (CD32)) are associated with increased risk of Crohn's disease, ulcerative colitis, Behçet's disease and ankylosing spondylitis, but are protective against some or all of T1D, rheumatoid arthritis, SLE and multiple sclerosis. For example, at *FCGR2A*, which encodes a cell-surface receptor on phagocytes, His131 is a risk allele for ulcerative colitis but is protective for SLE, multiple sclerosis and T1D. The 'opposite' Arg131 variant, which has lower affinity for immunoglobulin G (IgG) immune complexes, is the risk allele for SLE, multiple sclerosis and T1D.

The non-synonymous variant, Arg620Trp, in the gene encoding protein tyrosine phosphatase 22 (*PTPN22*) confers a sizable odds ratio of 1.78 for rheumatoid arthritis[9] and was identified before the GWAS era[56–58]. The same allele, which is almost certainly causal, was later found to increase the risk of T1D, SLE, vitiligo, AITD and ulcerative colitis. Although ulcerative colitis and Crohn's disease are closely related phenotypes, this variant is actually protective against Crohn's disease. *PTPN22* encodes a lymphoid tyrosine phosphatase that modulates T cell receptor signalling. The Arg620Trp variant facilitates cleavage *in vitro* by calpain 1 and is more readily ubiquitylated, leading to its degradation by the proteasome. The lower levels of PTPN22 result in

impaired dephosphorylation of substrates, thus increasing T cell receptor signalling and activation[59]. The correlation of the Crohn's-disease-associated variant with T cell suppression mirrors the paradoxical association between Crohn's disease and *NOD2* variants that inhibit pro-inflammatory signalling in innate immune cells. These observations add to the evidence that Crohn's disease may represent an atypical immunodeficiency state[25,60].

As discussed above, multiple IL-23 pathway associations are shared across several seronegative IMDs, but the direction of effect is inconsistent for some. For example, the *STAT3* risk allele for Crohn's disease is protective against multiple sclerosis; and the minor allele at rs6887695 that lies adjacent to the *IL12B* gene, which encodes the p40 subunit shared by IL-12 and IL-23, is protective for psoriasis but is a risk allele for Crohn's disease and ankylosing spondylitis.

Consistent with discordant genetic associations, treatments targeting one IMD may inadvertently provoke or exacerbate another. For example, the rs1800696 variant of TNF receptor 1 (*TNFR1*; also known as *TNFRSF1A*), is protective for ankylosing spondylitis but increases the risk of multiple sclerosis. This SNP affects *TNFR1* splicing, leading to loss of the transmembrane domain and an increase in the level of soluble TNFR1 (REF. 61). Notably, anti-TNF therapy is highly effective for ankylosing spondylitis, but can induce and exacerbate multiple sclerosis. In another example, *IL6R* polymorphisms show concordant association with ankylosing spondylitis and rheumatoid arthritis, but discordant association with coronary heart disease (which is increasingly recognized as an inflammatory phenotype). IL-6R inhibition is therapeutic for rheumatoid arthritis[62] but can increase cholesterol and low-density lipoprotein levels, raising concerns about the long-term effects of IL-6R inhibition on vascular health[63]. Also echoing a discordant genetic association, in the context of the IL-23 and T_H1 pathway (discussed above), safety concerns have been raised regarding IL-17-specific antibody therapy, which is effective for psoriasis but exacerbates Crohn's disease[64]. A better understanding of the genetic and immunological signals might have anticipated these problems: Crohn's disease is associated with excess T_H17 activation, whereas deficient T_H17 activation owing to mutations in *STAT3* leads to hyper-IgE syndrome, which is associated with recurrent extracellular bacterial and fungal infections[65]. Understandably, in this context IL-17-specific antibody therapy trialled in Crohn's disease caused mucocutaneous candidiasis in some patients, emphasizing the delicate regulation of gut mucosal immunity by IL-17.

The presence of discordantly associated variants emphasizes the functional importance of the implicated genes, but in most cases their precise biological impact and how this translates to phenotypic differences between IMDs awaits future study.

Disease-specific susceptibility loci

Although many IMD susceptibility loci are pleiotropic, some are convincingly phenotype-specific. Interestingly, these often confer large effect sizes. Consequently, they were among the first complex disease loci to be identified, using modestly powered sample sets for candidate gene analysis, association mapping within linkage intervals or early GWASs.

Phenotype-specific HLA associations have been discussed above. These are a feature of seronegative diseases, whereas seropositive phenotypes often share HLA haplotypes. Non-HLA examples in seronegative diseases include *NOD2* (per-allele odds ratio = 3) and autophagy related 16-like 1 (*ATG16L1*) (per-allele odds ratio = 1.23) in Crohn's disease; and hepatocyte nuclear factor 4 alpha (*HNF4A*) (per-allele odds ratio = 1.23) in ulcerative colitis. The associations at *NOD2* and autophagy genes *ATG16L1* and *IRGM* (immunity-related GTPase family M) first highlighted defective handling of intracellular bacteria as important in Crohn's disease pathogenesis[66–70]; and the *HNF4A* association has implicated epithelial barrier defects as an ulcerative colitis-specific mechanism[71].

Phenotype-specific associations in seropositive IMDs particularly implicate genes that encode relevant autoantigens. Examples of this include *INS* (encoding insulin) in T1D, *TSHR* (encoding thyroid stimulating hormone receptor) in Graves' disease[72], and *PRTN3* (encoding proteinase 3) in anti-neutrophil cytoplasmic antibody (ANCA)-associated vasculitis[73]. Genes that are involved in producing the auto-antigen are also implicated, such as *PADI4* (encoding peptidyl arginine deiminase, type IV) in rheumatoid arthritis[74]; *PADI4* citrullinates peptides targeted by anticitrullinated peptide antibodies, which are highly specific for rheumatoid arthritis.

Given the phenotype specificity of these loci, their generally large effect sizes and their links to specific pathogenic pathways, it is likely that they have key phenotype-determining roles for each IMD. Their penetrance is likely to depend on their co-segregation with a 'permissive' number of pleiotropic risk alleles. Other factors, including discordant loci and 'chance effects' such as the nature or route and timing of environmental exposures (with regard to the developing immune system and recognition of self versus non-self antigens), clearly also have a major role in determining specific IMD phenotypes. Although identifying pleiotropic loci may point to common pathways against which new therapies can be directed, targeting non-pleiotropic pathways may produce effective, disease-specific therapies with less systemic toxicity than drugs that disrupt central immune mechanisms.

Genetic clues for environmental triggers

An emerging theme is the overlap between the genetic loci that confer susceptibility to seronegative IMDs and those that alter the risk of infection. This is particularly the case for Crohn's disease and mycobacteria, for which seven out of eight risk loci for *Mycobacterium leprae* infection are confirmed Crohn's disease susceptibility loci (in six of these cases the same SNP is implicated). Marked overlap is also seen between Crohn's disease and Mendelian susceptibility to mycobacterial disease. The relationships are complex: some associations are concordant and others discordant[25]. It is currently

unclear whether these signals reflect an evolutionary signature of resistance to mycobacterial disease that confers increased risk of Crohn's disease or a direct role of mycobacteria or other microorganisms in Crohn's disease pathogenesis.

Few formal statistical tests for gene–environment interactions in IMDs have been performed so far. There is evidence that smoking interacts with risk alleles at HLA DRB1*16 and HLA A*02 in multiple sclerosis[75]. In rheumatoid arthritis there is an interaction among smoking, anticitrullinated protein antibodies (a hallmark of seropositive rheumatoid arthritis), the 620Trp variant in *PTPN22* and 'shared epitope' HLA-DRB1 alleles[76–78]. Adding these genetic risk factors to smoking alone increases the risk ratio from approximately 1.5 to greater than 20. Different lines of evidence suggest that smoking induces citrullinated and perhaps other autoantigens in the lung, and implicate periodontitis induced by *Porphyromonas gingivalis* in the potential breakdown of self-tolerance to these peptides[79]; rheumatoid arthritis may then occur in genetically susceptible individuals.

Several IMD-associated genetic loci affect host responses to viral infection. For example, the *IFIH1* locus contains a common variant signal, as well as multiple low-frequency coding variants, associated with T1D[16,80]. These variants lead either to lower IFIH1 expression or reduced protein function[81]. The IFIH1 protein senses double-stranded RNA during picornavirus replication. One such picornavirus, enterovirus, has been implicated in T1D aetiology by epidemiological studies. The *IFIH1* locus has also been associated with IBD[25], psoriasis[51], SLE[82] and vitiligo[83]. In addition, the region encoding TLR7 and TLR8 is associated with T1D[84] and coeliac disease[85], and these proteins recognize viral endosomal single-stranded RNA. Furthermore, the IRF family proteins (discussed above) mediate virally induced interferon-α (IFNα) and IFNβ production.

Discussion and future directions

Possible evolutionary origins. The key question that arises is how best to explain the plethora of shared variants that increase predisposition to IMD concordantly or discordantly. Clearly, immune-regulatory mechanisms operate within a range of tolerances. They vary among individuals and among cell types, reflecting genetic and epigenetic diversity and contributing in part to IMD susceptibility.

Many pleiotropic loci encompass genes that affect proximal events in T cell receptor signalling and co-stimulation, hence influencing adaptive immunity; examples include *PTPN22*, *CD45*, cytotoxic T lymphocyte-associated protein 4 (*CTLA4*) and inducible T cell co-stimulator ligand (*ICOSL*). It is intuitive that in environments in which pathogens are encountered frequently, a selective advantage would be conferred by a finely set immune-response trigger that also efficiently primes immune memory to foreign antigens. However, in a modern and clean environment that has altered 'training' of the developing immune system by reduced exposure to microorganisms in infancy, increased exposure to as-yet-unidentified environmental triggers and a contribution from phenotype-specific genes, this finely set trigger may predispose to IMD. Such an argument may, for example, pertain to the common *IL23R* coding variant Arg381Gln that is present in approximately 90% of healthy controls and 95% of patients with Crohn's disease, ankylosing spondylitis and psoriasis; the low-frequency, protective arginine allele attenuates IL-23-induced T_H17 cell effector function[86], whereas the common glutamine allele enhances this. The glutamine allele presumably conferred selective advantage through infection-prone evolutionary history but is now associated with IMD predisposition.

Missing heritability. For all IMDs studied, a substantial proportion of heritability (estimated from epidemiological and twin data) is yet to be accounted for[87]. Part of this heritability probably lies in common variant loci that have effect sizes that are too small to detect using a conventional $P < 5 \times 10^{-8}$ significance threshold, at least using tractable sample sizes. Recent analyses show that many hundreds (perhaps thousands) of such loci contribute to the genetic variance of complex traits, including IMDs[88–91]. The extent to which rare variants contribute to IMD susceptibility is currently unknown, but should become clearer with the widespread application of whole-genome sequencing and the development of new statistical approaches to analyse these data. Although rarer variants associated with IMDs have so far had large effect sizes, this probably represents a biased view as these studies have lacked statistical power to detect smaller effects.

Cross-disease and trans-ethnic studies. Analysing multiple disease data sets simultaneously — for example, grouping seropositive and seronegative data sets and deploying multinomial analyses — is likely to extend knowledge regarding pleiotropy and identify more IMD genes. There is also interest in analysing genotype data in light of more detailed clinical information — for example, using disease subphenotypes and serological biomarkers. For some phenotypes the loss of power incurred by analysing subsets within each disease will be more than offset by the reduced heterogeneity and increased signal-to-noise ratios. Trans-ethnic studies are also underway, and will leverage different patterns of LD to help pinpoint causal variants.

Defining causal genes and the cells in which they act. A key challenge is to identify which gene among the many positional candidates in each association interval is causal, and in which cell type it exerts an effect; resolving this will have a critical bearing on downstream functional analyses. A variety of informatics approaches have been deployed to identify causal genes, including those used in the recent IBD Immunochip analysis[25]. With regard to cell of origin, statistical methods based on transcriptional profiles have been used to identify cell types enriched for the expression of susceptibility genes for specific IMDs[92]. For example, this approach has implicated dendritic cells in Crohn's disease pathogenesis[25].

Importantly, expression analyses are only as good as the data sets interrogated. Current limitations include a prevalence of data from mice rather than humans, the use of transformed cell lines rather than primary cells, expression profiling of unseparated tissue rather than specific cell types, under-representation of several important tissues or cell types and use of unstimulated rather than stimulated cells. Progress in characterizing the complexity of eQTL effects in different cell types is being made[93], and data from ENCODE are helping the interpretation of GWAS and Immunochip results in the context of functional annotation of the genome[94]. As data sets become richer, and particularly when they are correlated with germline genetic and epigenetic data, the utility of pathway and network analysis will increase.

Functional and clinical translation of genetic findings. Functional and clinical translation are among the biggest remaining challenges. The robustly identified loci can be used to underpin hypothesis-driven research into pathogenic mechanisms and explored as potential therapeutic targets. For many diseases, the potential for genetic screening in high-risk populations is approaching feasibility, and requires development and testing. Large genotyped biobanks linked with clinical information, inception cohorts of high-risk individuals and deeply characterized prospective population cohorts would be useful resources across diseases.

There is increasing clinical interest in determinants of prognosis. However, few GWASs have attempted to identify distinct determinants of disease course. GWASs that have 'within-cases' designs, in which cases from opposite ends of the severity spectrum are selected, could explore determinants of prognosis across IMDs; such studies, using adequately powered cohorts, are awaited.

Gene discovery programmes have greatly benefited from the open exchange of genetic data, and this ethos will remain critical to future success. Interaction between the public research community and the pharmaceutical industry will also become increasingly important. The challenge now is to turn the increasing knowledge of pleiotropic pathways to clinical advantage, improving therapies and preventive strategies across the spectrum of IMDs while minimizing the risk of collateral toxicity.

1. Wellcome Trust Case Control Consortium. Genomewide association study of 14,000 cases of seven common diseases and 3,000 controls. *Nature* **447**, 661–683 (2007).
This is a proof-of-principle paper that initiated the GWAS era. It developed many methods and set standards for future GWAS performance and interpretation. It also set an important precedent by making data publicly available; the control genotype set alone from this study has been used in hundreds of GWASs.
2. Wellcome Trust Case Control Consortium & Australo-Anglo-American Spondyloarthritis Consortium (TASC). Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. *Nature Genet.* **39**, 1329–1337 (2007).
3. International Genetics of Ankylosing Spondylitis Consortium. Identification of multiple risk variants for ankylosing spondylitis through high-density genotyping of immune-related loci. *Nature Genet.* **45**, 730–738 (2013).
4. Torfs, C. P., King, M. C., Huey, B., Malmgren, J. & Grumet, F. C. Genetic interrelationship between insulin-dependent diabetes mellitus, the autoimmune thyroid diseases, and rheumatoid arthritis. *Am. J. Hum. Genet.* **38**, 170–187 (1986).
5. Lin, J. P. *et al.* Familial clustering of rheumatoid arthritis with other autoimmune diseases. *Hum. Genet.* **103**, 475–482 (1998).
6. Hemminki, K., Li, X., Sundquist, K. & Sundquist, J. Shared familial aggregation of susceptibility to autoimmune diseases. *Arthritis Rheum.* **60**, 2845–2847 (2009).
7. Thjodleifsson, B., Geirsson, A. J., Bjornsson, S. & Bjarnason, I. A common genetic background for inflammatory bowel disease and ankylosing spondylitis: a genealogic study in Iceland. *Arthritis Rheum.* **56**, 2633–2639 (2007).
8. Visscher, P. M., Brown, M. A., McCarthy, M. I. & Yang, J. Five years of GWAS discovery. *Am. J. Hum. Genet.* **90**, 7–24 (2012).
9. Eyre, S. *et al.* High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. *Nature Genet.* **44**, 1336–1340 (2012).
10. The Wellcome Trust Case Control Consortium *et al.* Bayesian refinement of association signals for 14 loci in 3 common diseases. *Nature Genet.* **44**, 1294–1301 (2012).
11. Boyle, A. P. *et al.* Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* **22**, 1790–1797 (2012).
12. Zhernakova, A. *et al.* Genetic analysis of innate immunity in Crohn's disease and ulcerative colitis identifies two susceptibility loci harboring *CARD9* and *IL18RAP*. *Am. J. Hum. Genet.* **82**, 1202–1210 (2008).
13. Pointon, J. J. *et al.* Elucidating the chromosome 9 association with AS; CARD9 is a candidate gene. *Genes Immun.* **11**, 490–496 (2010).
14. Evans, D. M. *et al.* Interaction between *ERAP1* and *HLA-B27* in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility. *Nature Genet.* **43**, 761–767 (2011).
This study provided the first strongly replicated example of a gene–gene interaction in any common IMD.
15. Rivas, M. A. *et al.* Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. *Nature Genet.* **43**, 1066–1073 (2011).
This study provided early suggestive evidence that low-frequency or rare variants that are not well captured by GWAS chips may be important in susceptibility to IMD.
16. Nejentsev, S., Walker, N., Riches, D., Egholm, M. & Todd, J. A. Rare variants of *IFIH1*, a gene implicated in antiviral responses, protect against type 1 diabetes. *Science* **324**, 387–389 (2009).
17. Adrianto, I. *et al.* Association of a functional variant downstream of *TNFAIP3* with systemic lupus erythematosus. *Nature Genet.* **43**, 253–258 (2011).
18. Crow, Y. J. *et al.* Mutations in the gene encoding the 3′–5′ DNA exonuclease TREX1 cause Aicardi–Goutieres syndrome at the *AGS1* locus. *Nature Genet.* **38**, 917–920 (2006).
19. Stoy, J. *et al.* Insulin gene mutations as a cause of permanent neonatal diabetes. *Proc. Natl Acad. Sci. USA* **104**, 15040–15044 (2007).
20. Miceli-Richard, C. *et al. CARD15* mutations in Blau syndrome. *Nature Genet.* **29**, 19–20 (2001).
21. Glocker, E. O. *et al.* Inflammatory bowel disease and mutations affecting the interleukin-10 receptor. *N. Engl. J. Med.* **361**, 2033–2045 (2009).
22. Franke, A. *et al.* Sequence variants in *IL10*, *ARPC2* and multiple other loci contribute to ulcerative colitis susceptibility. *Nature Genet.* **40**, 1319–1323 (2008).
23. Liu, J. Z. *et al.* Dense fine-mapping study identifies new susceptibility loci for primary biliary cirrhosis. *Nature Genet.* **44**, 1137–1141 (2012).
24. Trynka, G. *et al.* Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. *Nature Genet.* **43**, 1193–1201 (2011).
25. Jostins, L. *et al.* Host–microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* **491**, 119–124 (2012).
This paper demonstrated that the small-effect-size loci identified in large-scale genetic association studies improve our understanding of disease pathogenesis, particularly by enhancing bioinformatic analyses to pinpoint causal genes and pathways.
26. Tsoi, L. C. *et al.* Identification of 15 new psoriasis susceptibility loci highlights the role of innate immunity. *Nature Genet.* **44**, 1341–1348 (2012).
27. Price, P. *et al.* The genetic basis for the association of the 8.1 ancestral haplotype (A1, B8, DR3) with multiple immunopathological diseases. *Immunol. Rev.* **167**, 257–274 (1999).
28. Gregersen, P. K., Silver, J. & Winchester, R. J. The shared epitope hypothesis. An approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis. *Arthritis Rheum.* **30**, 1205–1213 (1987).
29. Brewerton, D. A. *et al.* Ankylosing spondylitis and HL-A 27. *Lancet* **1**, 904–907 (1973).
30. Caffrey, M. F. & James, D. C. Human lymphocyte antigen association with ankylosing spondylitis. *Nature* **9**, 121 (1973).
31. Schlosstein, L., Terasaki, P. I., Bluestone, R. & Pearson, C. M. High association of an HL-A antigen, W27, with ankylosing spondylitis. *N. Engl. J. Med.* **288**, 704–706 (1973).
32. Karvonen, J. HL-A antigens in psoriasis with special reference to the clinical type, age of onset, exacerbations after respiratory infections and occurrence of arthritis. *Ann. Clin. Res.* **7**, 301–311 (1975).
33. Takano, M., Miyajima, T., Kiuchi, M., Ohmori, K. & Amemiya, H. Behcet disease and the HLA system. *Tissue Antigens* **8**, 95–99 (1976).
34. Field, J. *et al.* A polymorphism in the *HLA-DPB1* gene is associated with susceptibility to multiple sclerosis. *PLoS ONE* **5**, e13454 (2010).
35. Nejentsev, S. *et al.* Localization of type 1 diabetes susceptibility to the MHC class I genes *HLA-B* and *HLA-A*. *Nature* **450**, 887–892 (2007).
36. Raychaudhuri, S. *et al.* Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. *Nature Genet.* **44**, 291–296 (2012).
37. Knight, J. *et al.* Conditional analysis identifies three novel major histocompatibility complex loci associated with psoriasis. *Hum. Mol. Genet.* **21**, 5185–5192 (2012).

38. Yang, Y. *et al.* The intricate role of complement component C4 in human systemic lupus erythematosus. *Curr. Dir. Autoimmun.* **7**, 98–132 (2004).
39. Duerr, R. H. *et al.* A genome-wide association study identifies *IL23R* as an inflammatory bowel disease gene. *Science* **314**, 1461–1463 (2006).
40. Cargill, M. *et al.* A large-scale genetic association study confirms *IL12B* and leads to the identification of *IL23R* as psoriasis-risk genes. *Am. J. Hum. Genet.* **80**, 273–290 (2007).
41. Kenna, T. J. & Brown, M. A. The role of IL-17-secreting mast cells in inflammatory joint disease. *Nature Rev. Rheumatol.* **9**, 375–379 (2012).
42. Mero, I. L. *et al.* A rare variant of the *TYK2* gene is confirmed to be associated with multiple sclerosis. *Eur. J. Hum. Genet.* **18**, 502–504 (2010).
43. Ban, M. *et al.* Replication analysis identifies *TYK2* as a multiple sclerosis susceptibility factor. *Eur. J. Hum. Genet.* **17**, 1309–1313 (2009).
44. Raychaudhuri, S. *et al.* Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. *PLoS Genet.* **5**, e1000534 (2009).
45. Ma, A. & Malynn, B. A. A20: linking a complex regulator of ubiquitylation to immunity and human disease. *Nature Rev. Immunol.* **12**, 774–785 (2012).
46. Matmati, M. *et al.* A20 (TNFAIP3) deficiency in myeloid cells triggers erosive polyarthritis resembling rheumatoid arthritis. *Nature Genet.* **43**, 908–912 (2011).
47. Hammer, G. E. *et al.* Expression of A20 by dendritic cells preserves immune homeostasis and prevents colitis and spondyloarthritis. *Nature Immunol.* **12**, 1184–1193 (2011).
48. Chu, Y. *et al.* B cells lacking the tumor suppressor TNFAIP3/A20 display impaired differentiation and hyperactivation and cause inflammation and autoimmunity in aged mice. *Blood* **117**, 2227–2236 (2011).
49. Krausgruber, T. *et al.* IRF5 promotes inflammatory macrophage polarization and T_H1-T_H17 responses. *Nature Immunol.* **12**, 231–238 (2011).
50. Hambleton, S. *et al. IRF8* mutations and human dendritic-cell immunodeficiency. *N. Engl. J. Med.* **365**, 127–138 (2011).
51. Strange, A. *et al.* A genome-wide association study identifies new psoriasis susceptibility loci and an interaction between *HLA-C* and *ERAP1*. *Nature Genet.* **42**, 985–990 (2010).
52. Franke, A. *et al.* Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nature Genet.* **42**, 1118–1125 (2010).
53. Andres, A. M. *et al.* Balancing selection maintains a form of *ERAP2* that undergoes nonsense-mediated decay and affects antigen presentation. *PLoS Genet.* **6**, e1001157 (2010).
54. Cagliani, R. *et al.* Genetic diversity at endoplasmic reticulum aminopeptidases is maintained by balancing selection and is associated with natural resistance to HIV-1 infection. *Hum. Mol. Genet.* **19**, 4705–4714 (2010).
55. Dendrou, C. A. *et al.* Cell-specific protein phenotypes for the autoimmune locus *IL2RA* using a genotype-selectable human bioresource. *Nature Genet.* **41**, 1011–1015 (2009).
56. Begovich, A. B. *et al.* A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (*PTPN22*) is associated with rheumatoid arthritis. *Am. J. Hum. Genet.* **75**, 330–337 (2004).
57. Kyogoku, C. *et al.* Genetic association of the R620W polymorphism of protein tyrosine phosphatase PTPN22 with human SLE. *Am. J. Hum. Genet.* **75**, 504–507 (2004).
58. Bottini, N. *et al.* A functional variant of lymphoid tyrosine phosphatase is associated with type I diabetes. *Nature Genet.* **36**, 337–338 (2004).
59. Zhang, J. *et al.* The autoimmune disease-associated *PTPN22* variant promotes calpain-mediated Lyp/Pep degradation associated with lymphocyte and dendritic cell hyperresponsiveness. *Nature Genet.* **43**, 902–907 (2011).
60. Marks, D. J. *et al.* Defective acute inflammation in Crohn's disease: a clinical investigation. *Lancet* **367**, 668–678 (2006).
61. Gregory, A. P. *et al.* TNF receptor 1 genetic risk mirrors outcome of anti-TNF therapy in multiple sclerosis. *Nature* **488**, 508–511 (2012).
62. Choy, E. H. *et al.* Therapeutic benefit of blocking interleukin-6 activity with an anti-interleukin-6 receptor monoclonal antibody in rheumatoid arthritis: a randomized, double-blind, placebo-controlled, dose-escalation trial. *Arthritis Rheum.* **46**, 3143–3150 (2002).
63. Maini, R. N. *et al.* Double-blind randomized controlled clinical trial of the interleukin-6 receptor antagonist, tocilizumab, in European patients with rheumatoid arthritis who had an incomplete response to methotrexate. *Arthritis Rheum.* **54**, 2817–2829 (2006).
64. Hueber, W. *et al.* Secukinumab, a human anti-IL-17A monoclonal antibody, for moderate to severe Crohn's disease: unexpected results of a randomised, double-blind placebo-controlled trial. *Gut* **61**, 1693–1700 (2012).
65. Minegishi, Y. *et al.* Dominant-negative mutations in the DNA-binding domain of STAT3 cause hyper-IgE syndrome. *Nature* **448**, 1058–1062 (2007).
66. Hampe, J. *et al.* A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in *ATG16L1*. *Nature Genet.* **39**, 207–211 (2007).
67. Ogura, Y. *et al.* A frameshift mutation in *NOD2* associated with susceptibility to Crohn's disease. *Nature* **411**, 603–606 (2001).
68. Hugot, J. P. *et al.* Association of *NOD2* leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* **411**, 599–603 (2001).
69. Rioux, J. D. *et al.* Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. *Nature Genet.* **39**, 596–604 (2007).
70. Parkes, M. *et al.* Sequence variants in the autophagy gene *IRGM* and multiple other replicating loci contribute to Crohn's disease susceptibility. *Nature Genet.* **39**, 830–832 (2007).
71. Barrett, J. C. *et al.* Genome-wide association study of ulcerative colitis identifies three new susceptibility loci, including the *HNF4A* region. *Nature Genet.* **41**, 1330–1334 (2009).
72. Dechairo, B. M. *et al.* Association of the TSHR gene with Graves' disease: the first disease specific locus. *Eur. J. Hum. Genet.* **13**, 1223–1230 (2005).
73. Lyons, P. A. *et al.* Genetically distinct subsets within ANCA-associated vasculitis. *N. Engl. J. Med.* **367**, 214–223 (2012).
74. Suzuki, A. *et al.* Functional haplotypes of *PADI4*, encoding citrullinating enzyme peptidylarginine deiminase 4, are associated with rheumatoid arthritis. *Nature Genet.* **34**, 395–402 (2003).
75. Hedstrom, A. K. *et al.* Smoking and two human leukocyte antigen genes interact to increase the risk for multiple sclerosis. *Brain* **134**, 653–664 (2011).
76. Kallberg, H. *et al.* Smoking is a major preventable risk factor for rheumatoid arthritis: estimations of risks after various exposures to cigarette smoke. *Ann. Rheum. Dis.* **70**, 508–511 (2011).
77. Padyukov, L., Silva, C., Stolt, P., Alfredsson, L. & Klareskog, L. A gene–environment interaction between smoking and shared epitope genes in HLA-DR provides a high risk of seropositive rheumatoid arthritis. *Arthritis Rheum.* **50**, 3085–3092 (2004).
This is a very good example of a gene–environment interaction that influences risk of a major common IMD: rheumatoid arthritis.
78. Mahdi, H. *et al.* Specific interaction between genotype, smoking and autoimmunity to citrullinated α-enolase in the etiology of rheumatoid arthritis. *Nature Genet.* **41**, 1319–1324 (2009).
79. Wegner, N. *et al.* Autoimmunity to specific citrullinated proteins gives the first clues to the etiology of rheumatoid arthritis. *Immunol. Rev.* **233**, 34–54 (2010).
80. Smyth, D. J. *et al.* A genome-wide association study of nonsynonymous SNPs identifies a type 1 diabetes locus in the interferon-induced helicase (*IFIH1*) region. *Nature Genet.* **38**, 617–619 (2006).
81. Downes, K. *et al.* Reduced expression of *IFIH1* is protective for type 1 diabetes. *PLoS ONE* **5**, e12646 (2010).
82. Molineros, J. E. *et al.* Admixture mapping in lupus identifies multiple functional variants within *IFIH1* associated with apoptosis, inflammation, and autoantibody production. *PLoS Genet.* **9**, e1003222 (2013).
83. Jin, Y. *et al.* Genome-wide association analyses identify 13 new susceptibility loci for generalized vitiligo. *Nature Genet.* **44**, 676–680 (2012).
84. Cooper, J. D. *et al.* Follow-up of 1715 SNPs from the Wellcome Trust Case Control Consortium genome-wide association study in type I diabetes families. *Genes Immun.* **10**, S85–S94 (2009).
85. Dubois, P. C. *et al.* Multiple common variants for celiac disease influencing immune gene expression. *Nature Genet.* **42**, 295–302 (2010).
86. Di Meglio, P. *et al.* The IL23R R381Q gene variant protects against immune-mediated diseases by impairing IL-23-induced Th17 effector response in humans. *PLoS ONE* **6**, e17160 (2011).
87. Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).
88. Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nature Genet.* **42**, 565–569 (2010).
This study developed a methodology that enables estimation of the contribution of common variants to common diseases using GWAS data from unrelated individuals. This showed that much of the 'missing heritability' in common diseases is still to be found among common variants.
89. Park, J. H. *et al.* Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. *Nature Genet.* **42**, 570–575 (2010).
90. Stahl, E. A. *et al.* Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. *Nature Genet.* **44**, 483–489 (2012).
91. Bloom, J. S., Ehrenreich, I. M., Loo, W. T., Lite, T. L. & Kruglyak, L. Finding the sources of missing heritability in a yeast cross. *Nature* **494**, 234–237 (2013).
This study used yeast cross-experiments to demonstrate that gene–gene interactions contributed 0–54% of broad-sense heritability, but even in this controlled setting the responsible loci were largely elusive. It identified variants responsible for 72–100% of narrow-sense heritability, indicating that multiple variants of small individual effect are largely responsible for the missing narrow-sense heritability.
92. Trynka, G. *et al.* Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nature Genet.* **45**, 124–130 (2013).
93. Fairfax, B. P. *et al.* Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. *Nature Genet.* **44**, 502–510 (2012).
94. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
This is a summary of an important series of papers that shed light on the role of non-coding variation in the human genome in the control of its function.
95. Cortes, A. & Brown, M. A. Promise and pitfalls of the Immunochip. *Arthritis Res. Ther.* **13**, 101 (2011).
96. Kirino, Y. *et al.* Genome-wide association analysis identifies new susceptibility loci for Behcet's disease and epistasis between *HLA-B*51* and *ERAP1*. *Nature Genet.* **45**, 202–207 (2013).

Acknowledgements
M.P. is supported by the UK National Institute for Health Research (NIHR) Biomedical Research Centre award to Addenbrooke's Hospital and University of Cambridge School of Clinical Medicine. D.A.v.H. is funded by the UK Medical Research Council (G1001158). M.A.B. is supported by a National Health and Medical Research Council (Australia) Senior Principal Research Fellowship.

Competing interests statement
The authors declare no competing financial interests.

FURTHER INFORMATION
ENCODE: http://www.genome.gov/10005107
ImmunoBase: http://www.immunobase.org
Ricopili: http://www.broadinstitute.org/mpg/ricopili/
US National Human Genome Research Institute (NHGRI) catalogue of published genome-wide association studies: http://www.genome.gov/gwastudies
Wellcome Trust Case Control Consortium: http://www.wtccc.org.uk/

SUPPLEMENTARY INFORMATION
See online article: S1 (figure) | S2 (box) | S3 (table)
ALL LINKS ARE ACTIVE IN THE ONLINE PDF