Published in final edited form as:

Nat Rev Genet. 2016 March ; 17(3): 160–174. doi:10.1038/nrg.2015.33.

Autoimmune diseases — connecting risk alleles with molecular traits of the immune system

Maria Gutierrez-Arcelus[1,2,3], **Stephen S. Rich**[4], and **Soumya Raychaudhuri**[1,2,3,5,6]

[1]Division of Genetics, and Division of Rheumatology, Immunology and Allergy, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA

[2]Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts 02142, USA

[3]Partners Center for Personalized Genetic Medicine, Boston, Massachusetts 02115, USA

[4]Center for Public Health Genomics, University of Virginia, Charlottesville, Virginia 22908, USA

[5]Faculty of Medical and Human Sciences, University of Manchester, Manchester M13 9PL, UK

[6]Department of Medicine, Karolinska Institutet and Karolinska University Hospital Solna, Stockholm SE-171 77, Sweden

Abstract

Genome-wide strategies have driven the discovery of more than 300 susceptibility loci for autoimmune diseases. However, for almost all loci, understanding of the mechanisms leading to autoimmunity remains limited, and most variants that are likely to be causal are in non-coding regions of the genome. A critical next step will be to identify the *in vivo* and *ex vivo* immunophenotypes that are affected by risk variants. To do this, key cell types and cell states that are implicated in autoimmune diseases will need to be defined. Functional genomic annotations from these cell types and states can then be used to resolve candidate genes and causal variants. Together with longitudinal studies, this approach may yield pivotal insights into how autoimmunity is triggered.

Critical to the success of the adaptive immune system is the ability to distinguish pathogens from self-antigens (BOX 1). Autoimmunity occurs when a failure in this recognition process leads to erroneous immune responses that damage healthy tissues. The first case of autoimmunity was recognized in 1904, through the observation of autoreactive antibodies in patients, which reacted to self-blood cells[1]. To date, more than 80 diseases have been found to have an autoimmune pathogenesis, with half of these considered to be rare[2]. Worldwide, autoimmune diseases, such as type 1 diabetes mellitus (T1DM), inflammatory bowel disease (IBD) or rheumatoid arthritis (RA), are now estimated to affect 7.6–9.4% of the population[3]. The prevalence of autoimmune diseases is typically higher in women than in men; systemic

Correspondence to S.R. ; Email: soumya@broadinstitute.org

Competing interests statement The authors declare no competing interests.

lupus erythematosus (SLE) is an extreme example, with a 10:1 female to male ratio. In the USA and the UK, auto immune diseases are within the top 10 leading causes of death for women aged up to 65 years and up to 75 years, respectively[4,5]. Moreover, autoimmune disease prevalence can vary by ethnicity (TABLE 1). For example, multiple sclerosis (MS) is ten times more common in North American cohorts than in those from South American countries[3], and SLE is more frequent in individuals of African ancestry than in those of European ancestry[6].

Despite knowledge of the epidemiology of autoimmune diseases, much remains to be understood in how self-tolerance is broken down and how autoimmunity is triggered. Although experiments in mouse models have established the foundation for our understanding of basic immunology, findings have overall not been translated successfully to human disease[7]. Currently, only a handful of alleles exist for which the mechanisms triggering autoimmune disease are defined to some extent.

With the development of high-throughput sequencing technologies, genome-wide association studies (GWAS) have uncovered hundreds of risk loci for autoimmune diseases (see Immunobase), many of which overlap across different disorders[8]. However, the genomic regions implicated by these risk loci are large, with up to a dozen or more potential candidate genes within each locus, and many contain polymorphisms that often have small effect sizes. Furthermore, most putative causative variants fall in non-coding regions of the genome and are enriched in distant regulatory elements. Therefore, besides continuing efforts to fine-map the causative variants and defining the genes involved, new approaches are needed to understand how risk variants affect gene regulation and immune function.

A key next step will be to define *in vivo* and *ex vivo* cellular and molecular immune traits that are influenced by genetic susceptibility factors and that are implicated in the development of autoimmunity. Elucidating how genetic risk variants alter immune traits within the human immune system will help us understand the impact they have on autoimmune disease risk (FIG. 1). This line of research will include the longitudinal measurement of a wide range of immunophenotypes, such as signalling responses, immune cell abundances and serum cytokine levels, in thousands of individuals, and in the context of the individuals' local environmental conditions.

Here, we review recent advances in gene mapping and fine-mapping of autoimmune disease-causing variants. We illustrate the genetic basis behind autoimmune disease by focusing on 12 common autoimmune diseases for which GWAS have been reported (FIG. 1). We then discuss recent functional genomics approaches that have the potential to help define key immune molecular traits, cell types and cell states. Finally, we highlight the necessity of quantifying immune traits to better understand the mechanisms of autoimmunity.

Familial clustering of autoimmune diseases

Autoimmune diseases cluster in families, indicating a substantial genetic component (TABLE 1) as well as a shared (and often unique) environmental component. Many studies suggest that disease concordance in monozygotic twins (that is, genetically identical

individuals, who share the same alleles) is significantly higher than that observed for dizygotic twins (who share one-half of their alleles)[9]. For example, monozygotic twins exhibit 25% concordance for MS, whereas dizygotic twins have 5.4% concordance[10,11]. Additionally, the risk of autoimmune disease in siblings of an affected individual is significantly higher than that of the general population, as measured by the sibling recurrence risk (λ_S; with a high λ_S value (for example, greater than ~5) indicating a high rate of recurrence). For example, psoriasis has a λ_S of ~6 (REF. 12), and the λ_S value for Crohn's disease is ~20 (REFS 13,14). However, λ_S and estimates of disease concordance can be unreliable, given their dependence on disease prevalence, shared environment among siblings, sample size, and the sex and age of ascertained patients[15].

Interestingly, autoimmune diseases co-occur in families more often than expected by their individual population prevalence[16]. For instance, Crohn's disease and ulcerative colitis have very high co-occurrence (pairwise odds ratio, estimated from logistic regression predicting one disease from another, ~67.4)[17] and a significantly high genetic correlation based on findings from GWAS[18]. Even diseases affecting different organ systems can have high co-occurrence. For example, coeliac disease, which affects the small intestine, and T1DM, which affects the pancreas, have high co-occurrence (pairwise odds ratio ~4.2)[17]. These findings suggest that there are common genetic factors across multiple autoimmune diseases.

Genetic factors associated with autoimmunity

Several approaches have been used to map the genetic variants contributing to autoimmune diseases. The first approaches, before the genomics era, were mainly based on families and captured a few of the loci with larger effect sizes. Later, high-throughput genome-wide technologies led to the identification of hundreds of common variants with small to moderate effect sizes. Larger cohorts and standardized technologies targeted for autoimmunity, such as the ImmunoChip, have further advanced the discovery and fine-mapping of disease loci. Finally, studying rare variants has yielded mechanistic insights into autoimmunity.

Discovery of susceptibility loci with large effect sizes

Early linkage analysis in pedigrees with patients enabled the identification of susceptibility loci with large effect size for autoimmune diseases, such as the major histocompatibility complex (MHC) locus for T1DM[19] and SLE[20], nucleotide-binding oligomerization domain-containing 2 (*NOD2*) for Crohn's disease[21] and *IBD5* for IBD[22]. The MHC locus contributes to autoimmune disease risk more significantly than do any other known loci. In T1DM, 30% of disease liability is attributed to the MHC locus, compared with 9% for other loci discovered across the rest of the genome with GWAS[23]. Although the MHC locus is a 3.6-Mb region comprising >250 genes[24], most associations are mediated by a handful of human leukocyte antigen (HLA) genes (FIG. 2a), which encode the receptors that are expressed by antigen-presenting cells to trigger the immune response.

In psoriasis, a series of linkage studies over the course of 9 years led to the identification of *CARD14* as a susceptibility gene, starting with mapping to the long arm of chromosome 17 using polymorphic microsatellite markers in 8 families with multiple affected members[25],

and ending with the identification of *CARD14* mutations that altered splicing in 2 families through targeted genomic capture and sequencing[26]. *CARD14* encodes a caspase recruitment domain-containing protein, and the risk alleles for psoriasis yield an increased activation of nuclear factor-κB (NF-κB) in keratinocytes, which could initiate the recruitment of inflammatory cells[26].

Subsequent candidate gene studies in autoimmune diseases — although generally unsuccessful at identifying reproducible results[27] — yielded several key discoveries. Most notably, a non-synonymous variant in *PTPN22* was shown to be associated with many autoimmune diseases, including T1DM, RA, SLE and Graves disease[28–31]. This gene encodes the tyrosine phosphatase lymphoid phosphatase (LYP), which is involved in signalling pathways during T cell and B cell receptor response. The risk variant affects the binding of LYP to the signalling suppressor SRC kinase[31]. However, the actual functional mechanisms leading to auto immunity are still an active area of investigation, more than 10 years after this discovery. The variant has been shown to both increase and decrease T cell receptor (TCR) activation in T cells, to disrupt B cell tolerance checkpoints and to alter Toll-like receptor (TLR) signalling and the production of type 1 interferon (IFN) in myeloid cells[32].

Another autoimmune disease-relevant gene, linked initially to T1DM through candidate gene studies, is *CTLA4*[33]. This gene encodes an immunoglobulin-like protein expressed on the surface of T helper (T_H) cells that functions as an inhibitor of activation. *CTLA4* is also associated with patients with autoantibody-positive RA[27]. Association of *CTLA4* with other autoimmune diseases, including alopecia areata (a condition in which the body attacks hair follicles, resulting in hair loss)[34], were subsequently established by GWAS[35], which have been successful for the identification of numerous common variants across multiple autoimmune diseases.

Detection of common risk variants through GWAS

Over 100 GWAS have been conducted to identify common variants associated with autoimmune diseases. For most diseases, dozens of susceptibility loci have been discovered, with more than 100 loci identified for RA[36] and IBD[37] (TABLE 1). Similarly to other complex diseases, most single-nucleotide polymorphisms (SNPs) associated with autoimmune disease have small to moderate effect sizes. For example, odd ratios of loci outside the MHC associated with psoriasis range from ~1.1 to 1.6, and those of loci associated with autoimmune thyroid disease range from 1.2 to 1.6 (as listed on Immunobase).

Multiple risk loci are shared between autoimmune diseases, which is consistent with them having common genetic aetiology[8,38]. For example, nine diseases show an association with the *STAT4* locus; notably, however, different SNPs in this locus may drive susceptibility for different diseases. The STAT4 protein plays a major part in cytokine signalling pathways in specific T_H cell populations. In addition, the same allele can increase risk for one disease but be protective for another, as has been shown for eight loci in a study analysing ten autoimmune diseases with paediatric age of onset[39]. Other susceptibility loci are specific for each disease and reflect their uniqueness in their pathology. For example, the insulin gene

(*INS*; whose variable number of tandem repeats, tagged by SNP rs689, is associated with risk) is associated with T1DM but no other autoimmune disease, consistent with an aetiology defined by the destruction of insulin-secreting β-cells[40,41].

Overall, GWAS have widely expanded the number of loci associated with autoimmune diseases, which has enabled researchers to jointly analyse loci and look for common pathways. This approach has led to the observation that autoimmune disease risk genes often cluster in key immunological pathways[42]. Intriguingly, in some of these pathways there is evidence of natural selection. An examination of 40 autoimmune disease risk loci suggested selective pressure driven by pathogens (such as *Plasmodium falciparum*, which causes malaria)[43]. Thus, some risk alleles for autoimmune diseases may have increased in frequency in the population because they have been favourable for fighting infectious disease[44]. For example, *IL23R*, which is associated with six autoimmune diseases, and *TYK2*, which is associated with seven autoimmune diseases, are part of the IL-23R response pathway and present evidence for selection implicating Protozoan pathogens[42,43,45].

Fine-mapping of disease-causing risk variants

Associations between HLA genes and autoimmune diseases have been described since the 1970s[46]; however, pinpointing the alleles driving the HLA associations has been challenging owing to the highly polymorphic nature of these genes and the long-range linkage disequilibrium (LD) across the MHC region. Before 2012, associations of RA with the MHC were explained by the `shared epitope hypothesis', which asserted that RA risk is driven by a common consensus sequence in the protein encoded by *HLA-DRB1*, encompassing amino acids 70–74 along the rim of the antigen-binding groove[47]. These amino acid residues were thought to mediate T cell activation. However, a fine-mapping study in 2012 indicated that ~90% of the MHC risk in RA is attributable to a specific amino acid residue in position 13 at the bottom of the DRβ1 antigen-binding groove, and that amino acids 71 and 74 (whose side chains point into the antigen-binding groove) independently modulate RA susceptibility[48] (FIG. 2b). Independent genetic association effects at *HLA-B* and *HLA-DPB1* are explained by a single amino acid site at the bottom of the binding grooves of the protein products. Thus, amino acid sites that modulate binding to specific antigens mediate RA risk. Similar results have now been confirmed in Asian and African populations that mainly involve the same amino acids but sometimes present differences in effect sizes, potentially driven by differences in minor allele frequency between populations[49,50]. These and other studies underscore the utility of *trans*-ancestral cohorts for fine-mapping genetic associations[51]. Similarly, for T1DM, a secondary association to DRβ1 at positions 13 and 71 explains much of the class II HLA association with T1DM, in addition to the well-known *HLA-DQB1* position 57 association[23]. Certain regions in HLA proteins seem to recur in the context of HLA–disease associations; for example, the DRβ1 pocket 4 includes positions 13, 71 and 74, and has been implicated in antibody-negative RA and follicular lymphoma[52,53] in addition to T1DM and antibody-positive RA[23,48]. These recurrent regions may be critical to autoimmunity and could also be important to induce self-tolerance.

Notably, these and other studies have been enabled by observations made by multiple research groups that intragenic MHC SNPs can be used to infer HLA genotypes[54–57].

Approaches using imputation and large population-specific reference panels have enabled the re-investigation of the MHC locus for a wide range of autoimmune and non-autoimmune diseases, starting with re-mapping associations to HIV controller status, that is, the ability of individuals infected with HIV to exert control over the virus without the need for medicine[57]. As with all imputation approaches, the availability of large population-specific reference panels can be a limiting factor; HLA imputation requires particularly large panels, owing to the need to impute classical alleles, some of which can be rare. As new reference panels have emerged, this approach has been applied to European, Asian and African populations[36,58].

Next-generation sequencing strategies are enabling the effective interrogation of the HLA genes and the MHC region[59–62], although these protocols have yet to be widely adopted in clinical practice or research. Recently, Dilthey *et al.*[63] implemented a population reference graph to infer MHC sequence, taking into account the extended MHC haplotypes from the MHC haplotype project[64] and the HLA alleles reported in the IMGT/HLA database[65]. As sequence reads increase in length, many of the challenges with targeting HLA genes and read-mapping assembly are becoming more tractable[66].

Fine-mapping using the ImmunoChip

In an effort to leverage the common features of autoimmune diseases to discover novel associations and fine-map existing ones, a genotyping chip was designed with dense common and rare variant coverage in susceptibility loci, and in loci with immune-related genes[67,68]. The ImmunoChip encompasses ~180,000 SNPs in 186 loci[68].

One of the first studies using the ImmunoChip was reported for coeliac disease. Thirteen new risk loci were discovered to be associated with coeliac disease, and fine-mapping enabled the more-accurate delimiting of previously discovered non-HLA susceptibility regions[68]. In T1DM, the ImmunoChip enabled replication of known variants, discovery of new loci and fine-mapping of previously found loci[41]. For example, for the *IL2RA* gene, which encodes a subunit of the IL-2 cytokine receptor, researchers found a new variant that is partially linked to previously reported variants, as well as two additional independently associated variants[41]. More recently, researchers fine-mapped 18 IBD risk loci, including previously reported coding variants as well as additional protein-coding, intronic and intergenic variants, to a single likely causal variant[69]. Specifically, the SNP rs6062496, in the intron of the *TNFRSF6B* gene, overlaps an open chromatin region and is predicted to alter a transcription factor binding site for early B-cell factor 1 (EBF1), which is implicated in B cell identity[69]. The ImmunoChip has also been useful in the discovery of novel associations and fine-mapping in other auto immune diseases, including psoriasis, juvenile idiopathic arthritis, RA, ankylosing spondylitis and autoimmune thyroid disease[41,70–74]. ImmunoChip genotyping on samples also includes dense coverage of the MHC locus, and has thus improved the accuracy of HLA imputation[56]. It has also enabled assessment of shared genetic contribution across autoimmune diseases[35].

The regionally dense variant catalogue of ImmunoChip and its large-scale usage in tens of thousands of individuals have significantly improved risk assessment for Crohn's disease and ulcerative colitis, reaching areas under the curve (AUCs) of 0.83–0.86 with machine-

learning algorithms[75]. AUC can be interpreted as the probability that the classifier will assign a higher risk score to a randomly chosen individual positive for the disease than to a randomly chosen individual that does not have the disease. Another method for estimating polygenic risk scores by modelling LD has proven useful for RA, T1DM and coeliac disease[76]. Overall, the power and utility of the ImmunoChip will be expanded as more data are aggregated.

Rare protein-coding variants and autoimmune disease risk

Despite the success of GWAS in finding common SNPs associated with disease, common variants explain only a small percentage of the familial aggregation of common complex diseases. For example, in IBD, the 200 currently known common susceptibility loci explain <15% of the disease variance[37,51]. By contrast, rare protein-coding variants are more likely to have loss-of-function effects and thus are, in theory, more likely to have larger effect sizes in disease[77]. As healthy individuals can carry dozens of loss-of-function variants[78], the presence of rare variants that alter protein function does not necessarily imply that the gene has a role in disease. Statistical reproducibility and functional validation are needed to be confident about the role of rare variant associations with autoimmune disease[79].

Although several groups have ascertained rare variants with the objective of identifying missing heritability[80,81], rare variants have proven to be more useful for dissecting candidate genes within risk loci and gaining insights into the mechanisms of disease than for the discovery of new variants that contribute substantially to disease heritability. Exon sequencing of candidate autoimmunity disease genes has yielded interesting findings for IBD, even if rare variants added less than 0.5% to the explained variance[82]. Five rare variants have been identified in *NOD2*, independent of known common disease variants in the same gene[82]. *NOD2* encodes a pattern-recognition receptor that recognizes bacterial molecules, including muramyl dipeptide (MDP). Two of the rare *NOD2* variants affect the translocation of NOD2 to the membrane or impair NF-κB response mediated by MDP stimulation[82].

Screening for rare variants has led to the identification of protective variants against disease, as in the case for *TYK2* in RA and SLE[83], and *IFIH1* in T1DM[84]. *IFIH1* encodes a cytoplasmic helicase that, on detection of picornavirus RNA, triggers antiviral IFNβ response[85]. The rare protective variants for T1DM cause a loss of function of the protein[84]. However, gain-of-function variants in the same gene have been associated with Aicardi–Gouitères syndrome, an inflammatory disorder with severe neurodevelopmental impairment[86].

In general, it does not seem that rare variants explain common variant associations from GWAS in autoimmune disease[87_89]. This could be because only exons from a subset of genes, and no regulatory regions, have been interrogated so far. In addition, extremely large samples sizes are required to detect rare variant associations[90]. In rare and severe cases of autoimmunity, such as very-early age-of-onset IBD, exome sequencing has identified the causal gene, leading to successful medical treatments based on the detection of protein-coding variants[91]. Thus, studying paediatric age-of-onset autoimmune disease at the genomic level can uncover novel genes and variants with large effect as well as pathways

involved in autoimmunity, some of which may involve already available therapeutic targets for other clinical conditions[39].

Autoimmunity-relevant cell types and cell states

Despite several successes in defining the mechanisms that can lead to autoimmunity described above, an understanding of how the small effects in each locus contribute to triggering autoreactivity and inflammation is lacking. Although risk alleles seem to influence gene regulation, efforts to investigate the function of these alleles are hindered by the complexity of the human immune system. The human immune system is composed of hundreds of different cell types, cellular subsets and cell states (BOX 1). Differences in cell types and cell state can lead to dramatic differences in intracellular gene regulation and functional phenotypes. It is thus critical for the functional follow-up of individual disease alleles to have an *ex vivo* cellular system that appropriately reflects a cell type and cell state in which genetic mechanisms mediate disease risk. When such a system has been defined, it may become possible to define mechanisms of the individual alleles on gene expression and other cellular phenotypes.

Given the polygenic nature of autoimmune diseases, investigators have hypothesized that many genetic risk factors exert their effect in a small number of cell types. For many autoimmune diseases, the specific cell subtypes that mediate disease risk are unclear, that is, the literature implicates different cell types and cellular subsets, often on the basis of studies in mouse models or human observational studies. For example, for RA, synovial fibroblasts, mast cells, B cells and T cells have all been implicated[92–95]. In addition, a cell type can comprise different cellular subsets. For instance, T cells can be subdivided into cytotoxic and T_H cells, and the latter can be further sub divided into various cellular subsets, such as T_H1, T_H2, T_H9, T_H17, regulatory T (T_{Reg}) cells and follicular T_H cells[96]. In MS, it was originally thought that T_H1 cells were involved in disease development, and subsequent findings have pointed to T_H17 cells having an important role; however, the specific roles of T_H1 and T_H17 cells in MS remain to be elucidated[97]. Furthermore, each cell subset population can take on a range of different cellular states in response to external stimuli and environment. Thus, overall it is not trivial to pinpoint pathological drivers. Most strategies developed to pinpoint relevant cell types for disease seek to identify enrichment of cell-type-specific cellular phenotypes (such as gene expression or epigenetic marks) among disease loci.

Cell-type-specific gene expression

Using gene expression levels of a wide compendium of mouse immune tissues[98], Hu *et al.* quantified the tissue specificity of gene expression[99]. For each disease, they assessed which tissue was most enriched in tissue-specific gene expression of genes within disease-associated risk loci. This method found significant enrichment for tissue-specific expression of splenic transitional B cells for genes in SLE risk loci. Similarly, for RA, $CD4^+$ effector memory T cells presented the highest enrichment. For Crohn's disease, epithelial-associated stimulated $CD103^+$ dendritic cells were the most-strongly implicated cells[99]. A subsequent meta-analysis of ulcerative colitis and Crohn's disease studies confirmed the importance of

dendritic cells[37]. In addition, this study found stronger enrichment for activated dendritic cells, highlighting the importance of ascertaining different cellular states[37].

Cell type-specific epigenomic profiles

Researchers have also used histone marks or open chromatin regions as a proxy for active regulatory elements (such as enhancers and promoters). Resources of epigenomic profiles in dozens of human cell types have now been generated and continue to be expanded[100_102]. These resources can be used to identify cell types with enrichment of cell-type-specific chromatin marks in disease susceptibility loci. For example, Trynka *et al.* analysed trimethylation on Lys4 of histone H3 (H3K4me3) marks (a histone modification associated with promoters) in many cell-types. They discovered that T_{Reg} cells have an enrichment of T_{Reg} cell-specific H3K4me3 peaks within susceptibility loci for RA[103]. Enrichment of enhancers and super enhancer marks in disease loci for cell types relevant for auto immune diseases have been discovered[38,41,104,105]. H3K4me1 (a histone mark associated with enhancers) profiles demonstrated that the most enriched cell type for RA, Crohn's disease and ulcerative colitis is stimulated T_H17 subset[106]. This suggests that many risk variants for autoimmune disease may be exerting their effects in the activated state of immune cell types. Overall, these studies highlight the utility of using cell-type-specific molecular phenotypes, such as gene expression or epigenetic marks, to define relevant cell types and cell states for autoimmune disease.

Additionally, knowledge of disease-relevant cell types and cell states combined with cell-specific regulatory annotations can help prioritize candidate causative variants[38,103]. Even in fine-mapped loci, it is often the case that several SNPs in very high LD are associated with the disease. In this case, having information on the regulatory elements of the locus in the cell type relevant for the disease can point to the most pertinent variants to test for functional follow-up (FIG. 2c). This approach proved successful for revealing the regulatory role of a risk variant for SLE[107]. Methods that leverage epigenomic annotations to prioritize non-coding variants associated with disease have been developed[108_113]. Although these methods are being applied to a broad range of phenotypes, they are limited by the narrow compendium of publicly available epigenomic profiles. Hence, as cellular traits are measured in more cell types and states, functional consequence prediction of non-coding variants will improve.

Quantifying immune-related phenotypes

As discussed above, many autoimmune disease variants probably influence disease through alteration of gene regulation in a cell-type-specific manner[103,114] (FIG. 3). Notably, ~90% of candidate causative variants for autoimmune diseases are estimated to be non-coding[38]. It is estimated that causal variants are abundant in enhancers, which tend to be context-specific in their effects. For instance, Farh *et al.* argued that causal autoimmune disease variants were more enriched in T cell enhancers than in T cell promoters[38]. In T1DM, the set of credible susceptibility variants is enriched in enhancer marks found in the thymus and other immune cell types[41]. In RA, a significant enrichment of risk alleles is found in super enhancer regions in $CD4^+$ T cells compared with in typical enhancers[115]. To really define the

mechanisms of these non-coding autoimmune disease variants, we need to understand how they affect not only gene regulation but also function at the cellular level and at the level of the entire immune system, that is, signalling response, cytokine production, cytokine response, cell type counts and antigenic response. We collectively refer to these cellular and systemic immune traits as immunophenotypes.

Expression quantitative trait loci

Most genes have variants correlated to gene expression (expression quantitative trait loci (eQTLs)) in different cell types[116–118]. These variants can be local eQTLs[119,120] (*cis*; FIG. 3b); often a gene has multiple *cis* eQTLs[121]. eQTLs can be cell-type-specific (that is, active in one cell type but not in another), and their effect sizes may vary across cell types[122–127]. Reports have estimated that 11–30% of autoimmune risk loci involve *cis* eQTLs in blood-derived cells or $CD4^+$ T cells[38,51,118,128], and that trait-associated *cis* eQTLs have a higher degree of tissue specificity than expected[118]. In *trans* eQTLs, which can involve an intermediary gene, the variant is distant from the gene (typically >5 Mb away). This type of eQTL has proven more difficult to detect, in part due to their smaller effect size compared with *cis* eQTLs[129]. A proportion of *trans* eQTLs have been shown to be associated with complex traits[130], including a SNP associated with SLE that affects the expression of multiple IFNγ response genes[130].

Early studies investigating the effects of variants on gene expression examined (B cell-derived) lymphoblastoid cell lines[131,132], but recent studies have emphasized primary cells, such as monocytes, B cells, T cells, dendritic cells and neutrophils[117,118,122,125,128,133–136] to capture regulatory variation active in cell types that are highly relevant for autoimmunity. A critical issue is that many cell types of interest (for example, T cells) constitute a relatively small component of peripheral blood. Hence, further dissection of immune cell types (by cell sorting with flow cytometry (BOX 2)), is often limited by the number of cells available.

A compelling example highlighting the utility of eQTL studies for autoimmune disease aetiology is the recent investigation of the *UBE2L3* gene and its association with SLE[137–140]. Using genome-wide eQTL studies, Lewis *et al.* found that the risk haplotype of this locus is associated with increased *UBE2L3* expression in B cells and monocytes, leading to higher protein levels in B cells[141]. Whereas the eQTL SNP (rs140490) is active in B cells and monocytes, it has a negligible effect in $CD4^+$ T cells. *UBE2L3* encodes an E2 ubiquitin-conjugating enzyme, which together with linear ubiquitin chain assembly complex (LUBAC) is necessary for degradation of NF-κB inhibitor-α (IκBα). In healthy individuals, higher expression of *UBE2L3* in B cells and monocytes, driven by the risk haplotype, leads consistently to higher activation of the transcription factor NF-κB, a major regulator of B cell development and survival[115]. In patients with SLE, the susceptibility risk allele is also associated with higher proliferation of peripheral blood plasmablasts, a differentiated form of B cells that produces greater amounts of antibodies[115]. Although details of this disease association remain to be elucidated, these results suggest *UBE2L3* as a potential putative drug target for SLE and demonstrate the utility of intermediate immunophenotypes for the discovery of disease mechanisms[142].

The genetic effects on gene regulation are now being examined in a range of physiological states in monocytes, dendritic cells, $CD4^+$ T cells and endothelial cells[128,133,134,143–145]. A greater proportion of eQTLs are found in these cell types exclusively in stimulated states, often dependent on the stimulus or time after stimulus (FIG. 3a). For example, a *cis* eQTL for *IFNB1* (REF. 133), which encodes the cytokine IFNβ, is active after 2 hours of lipopolysaccharide (LPS) stimulation in monocytes but not in the naive state nor after 24 hours of LPS stimulation. The same SNP is associated in *trans* to 17 genes, all of which are part of the IFNβ signalling cascade, after 24 hours of stimulation. This widespread effect could be mediated by the transcription factor interferon regulatory factor 7 (IRF7), which acts just downstream of *IFNB1* and upstream of most of the other genes affected in *trans*. These studies highlight the importance of ascertaining genetic regulatory variation in different cellular states, as eQTLs for autoimmunity may be missed if only baseline effects are assayed in a single source of cells.

Together with the evidence of variants associated with autoimmune diseases being strongly enriched in immune cell type enhancer sequences, the above findings suggest that disease variants alter gene regulation in a very cell type- and cell state-dependent manner. Furthermore, it is possible that a single variant in a single enhancer has a very small effect on transcription, dependent on the target gene and cell type, so that many variants in several enhancers may be needed to detect a signal in a low sample size[146]. Alternatively, a SNP that affects an enhancer whose target is a transcription factor may alone regulate hundreds of genes, should the correct cell type be implicated.

Epigenetic phenotypes

Epigenetic marks, such as DNA methylation, change among cell types and conditions, marking active or repressed regions in the genome. DNA methylation might influence disease as a mediator of genetic risk or be a signature of environmental exposure that triggers disease, or it may be just a consequence of a disease[147]. Initial studies in discordant monozygotic twins (that is, where one twin has the disease and the other twin does not) in T1DM and in SLE have found differentially methylated regions, often pre-dating the disease diagnosis in T1DM[148] or enriched in immuno logical genes in SLE[149]. Hence, emerging questions are how and to what extent genetic variation affects epigenetic traits, and how changes in epigenetic marks associate with gene expression.

Groups assessing the effect of genetic variants on DNA methylation and histone modifications have found a widespread signal. Their results suggest that one of the major mechanisms by which these variants act is by altering the binding of transcription factors, which in turn could affect the local epigenetic landscape and the transcriptional output of their target genes[150–153]. Another study quantified DNA methylation levels in lymphocytes of several hundred individuals with a chip assaying 450,000 CpG sites. This approach revealed both local (*cis*) and distant (*trans*) methylation QTLs that were in high LD to autoimmune disease variants[154]. This colocalization does not necessarily mean that the actual disease variant is affecting DNA methylation. As higher-resolution technologies for DNA methylation, such as whole-genome bisulphite sequencing, and for ascertaining DNA-sequence variation, such as genome sequencing, become more economical, applied to larger

cohorts and specific cell types, we will be able to dissect the consequences of disease variants on molecular immune-related traits. Mendelian randomization methods[155] have been useful for testing causal relationships in this type of functional genomics studies[150], as well as in disease studies, and have led to the identification of candidate methylation sites that could be mediators of genetic risk for RA[156].

Transcriptional regulation also occurs through chroma tin interactions, that is, when enhancers and promoters come into close proximity via DNA looping[157]. Thus, assessing how genetic variation can affect the three-dimensional structure of the genome will also be useful for assessing the regulatory effect of autoimmune disease risk variants.

Immunophenotypes

Many immunophenotypes exist that can be measured efficiently, including cell type counts, cell proliferation, serum protein levels, surface protein expression levels, and signalling response levels (BOX 2; FIG. 4). Initial studies showed that immunophenotypes such as the ratio of different T cell subsets are heritable in mice and humans[158,159]. Since then, multiple studies have continued to examine the genetic control of lymphocyte abundance in blood, typically involving a few measured phenotypes per study[160–165]. More recently, improvement in high-throughput cellular phenotyping technologies, such as flow and mass cytometry[166,167] (BOX 2), has enabled studies to characterize thousands of immunophenotypes in hundreds to thousands of individuals[168–170], providing a full range of genetic contribution to variation depending on the trait and the study. Overall, for the majority of the immunophenotypes measured, phenotypic variation can be explained in a larger proportion by environment than by genetics[168], which is consistent with the immune system's role in responding to the environment. However, multiple immunophenotypes present predominant genetic contributions to variation, including serum cytokine and chemokine levels, cell population frequencies and signalling response phenotypes[168]. Particularly, serum levels of IL-6 and IL-12p40 are highly heritable[168], and IL-12p40 has been associated with psoriasis and asthma[171,172].

Genetic variants can be associated with variation in immunophenotypes (as quantitative traits). A total of 23 variants have been associated with 132 cell frequency traits in a cohort of Sardinians[169]. In several cases, the SNP associated with the abundance of a cell type mapped to the gene coding for one of the markers defining the cell type. The strongest association involved $CD39^{+}CD4^{+}$ T cells and a SNP in the intron of *ENTPD1*, the gene that encodes CD39. The same SNP was associated with surface protein levels of CD39 (REF. 170); thus, the cell type counts were influenced by the variation of expression of one of the markers. It is critical to consider these possible confounding factors in analyses of cell population frequencies. Orrù *et al.* found a modest overlap of cell abundance associations with disease loci, with only 3 out of the 23 variants being in high LD with reported autoimmune disease variants[169]. Although larger sample sizes may be needed to find more loci with significant associations (and subsequently possibly more overlap with disease loci), these results are consistent with the hypothesis that genetic variation influencing immune cell frequencies does not confer susceptibility to autoimmune disease[128].

Instead, other genetically controlled immunophenotypes might be more enriched in disease loci, particularly phenotypes representing functional responses of specific cellular subsets. For example, MS risk loci are enriched for binding sites of the transcription factor NF-κB[38], which has an important role in immune response signalling. One MS risk variant in the NF-κB locus[173] is associated with higher NF-κB expression and increased signalling response after stimulating naive $CD4^+$ T cells with TNF[174]. These results highlight the relevance of studying the genetic basis of immunophenotypes to dissect the mechanisms by which autoimmune diseases develop. It will be interesting to determine the impact of shared versus opposing (or specific) genetic effects between autoimmune diseases on immunophenotypes as well as cell-type specificity on molecular traits.

Conclusions and perspectives

Although manipulation of mouse models has been a powerful tool for understanding immunology, it is crucial to understand autoimmune disease in the context of human immunophenotypes[7]. Recent studies are focusing on understanding the mechanisms by which genetic risk variants confer susceptibility to disease. Researchers are achieving this by quantifying the natural variation of cellular and molecular phenotypes in cell types and cell states relevant for autoimmune diseases, as well as by measuring a wide range of immuno pheno types. The modest overlap between susceptibility loci and genetic variation affecting immune cell abundances suggests that some immunophenotypes will be more pertinent to ascertain for autoimmune disease than others. Future immunoprofiling studies with larger sample sizes will probably be able to pinpoint the most important read-outs to measure. Within the context of an appropriate *ex vivo* system, emerging genetic engineering approaches, such as CRISPR–Cas9-based genome editing[175–177], could provide insights on altered cellular phenotypes.

Recent genetic studies have demonstrated that many genomic loci of small to moderate effect contribute to autoimmune disease risk; of these, most do not lie in protein-coding regions but cluster in non-coding regions of the genome. Given the regulatory nature of these risk variants, it is essential to identify the causative polymorphisms in enhancers and their target genes. Novel technologies applied to DNA–DNA interactions, such as chromatin interaction analysis by paired-end tag sequencing (ChIA-PET) and Hi-C[157,178], combined with the insights of immunological eQTL studies, will contribute to resolving this issue.

The next wave of eQTL and gene regulation studies will use specific cell types and cell states, focusing on the most-relevant ones for each disease. Such studies may enable the detection of context-specific effects, which can be crucial to disease. These studies are challenging to conduct, as they require an infrastructure to obtain, sort and manipulate cells in addition to the requisite sample size. Moreover, although progress has been made in developing methods to identify the most relevant cell types and cell states based on integration of risk loci and functional genomic annotations, studies are limited by the narrow compendium of human immune epigenomic and transcriptomic profiles. Allele-specific expression is mainly driven by genetic regulatory effects in *cis*[119,179]. Hence, the allelic effects observed in heterozygous sites of a few individuals (FIG. 3c) could yield valuable information on the particular cellular subsets and conditions in which most disease genes are

genetically deregulated. Another technique that could aid in this objective is single-cell RNA sequencing (RNA-seq), as it can reveal undiscovered heterogeneity of cellular populations and states[180]. Approaches that exploit the allelic information in RNA-seq data may also prove useful to distinguish cell types and states that should be ascertained in a large-scale manner.

Technological advancements are enabling the ascertainment of proteomic phenotypes in autoimmunity. Genetic variants can affect transcript levels or mRNA stability[181], which in turn could alter protein levels[182–186]. Alternatively, missense variants can change a protein's stability and influence protein-binding partners[187]. Susceptibility genes for RA, Crohn's disease and MS are enriched for protein–protein interaction networks that are more connected than expected by chance, and these genes are expressed in similar disease-relevant tissues[188,189]. Once technological advances facilitate dynamic assessment of these protein-interaction pheno types at the level of human common variation, directly in the cell types of interest, we will learn a lot more about the functional effect of complex disease susceptibility variants (for example, how a variant that influences splicing may alter the binding partners of the encoded protein, which may affect downstream biological processes).

As we gain a more comprehensive picture of genetic variants associated with the many layers of intermediate phenotypes and disease, it is becoming challenging to disentangle the causal relationships among them (FIG. 1). For example, a SNP may be associated with an immunophenotype and a disease. This could reflect the more intuitive scenario in which the SNP affects the immunophenotype, which in turn influences the disease. However, this situation could also be consistent with the SNP increasing susceptibility to the disease and the manifestation of disease causing a change in the immunophenotype. Furthermore, the SNP may be associated independently with the disease and with the immunophenotype, but these two are not cause of, nor consequence from, each other. Computational methods relying on Mendelian randomization are being developed to address this problem[155], although these require huge sample sizes. Human studies of autoimmunity are often limited by comparison of prevalent cases (versus controls); however, disease-progression studies based on longitudinal observation will ultimately reveal the causal relationships of intermediate phenotypes and disease. Efforts to enhance personalized medicine[190] and biobanks can be instrumental for this purpose[191,192]. For example, blood samples from healthy individuals stored in biobanks could be used as reference samples when a subset of these individuals becomes ill: both healthy individuals and individuals with a disease could be tracked and compared over time to dissect the gradual manifestations pre-disease and how genetic variation may have partially triggered them. The information gained from studies aiming at finding the relevant immunophenotypes for a particular autoimmune disease will aid in making this process more efficient. Overall, the field of autoimmune disease genetics faces difficult challenges but, at the same time, exciting and translational discoveries are now being delivered.

Acknowledgements

This work is supported in part by funding from the Swiss National Science Foundation (M.G.-A.), the US National Institutes of Health (1R01AR063759, 5U01GM092691-05, 1UH2AR067677-01 and U19 AI111224-01 (S.R.)) and the Doris Duke Charitable Foundation Grant #2013097 (S.R.). The authors thank the Raychaudhuri laboratory for

their feedback, in particular H.-J. Westra and C. M. Hong for their careful reading of the manuscript and C. Y. Fonseka for figure support.

Glossary

Adaptive immune system	The part of the immune system that can react to specific pathogens and remember them to elicit a faster immune response if they are re-encountered.
Autoimmunity	When the immune system elicits an immune response against its own body (or non-pathogenic antigens that are commonly found in the body, such as commensal bacteria).
Type 1 diabetes mellitus (T1DM)	A disease characterized by autoreactivity to insulin-secreting β-cells found in the pancreas, resulting in high glucose levels in blood and urine.
Inflammatory bowel disease (IBD)	A term that groups together several conditions, such as Crohn's disease and ulcerative colitis, that are characterized by inflammation of the small intestine and colon. The inflammation is thought to be caused by immune reaction to commensal bacteria.
Rheumatoid arthritis (RA)	Autoimmune disease causing joint inflammation and destruction of cartilage and bone. Patients with RA can present autoantibodies (seropositive) or not (seronegative).
Systemic lupus erythematosus (SLE)	A disease characterized by autoreactivity to ubiquitous molecules, such as DNA or chromatin proteins, which can cause immune reactions against many parts of the body, some of the most common being joints, skin, blood vessels, kidneys and nervous system.
Multiple sclerosis (MS)	A condition of autoimmunity against the myelin sheath that protects the central nervous system.
Self-tolerance	The mechanism by which the immune system does not attack its own body.
Genome-wide association studies (GWAS)	Studies in which genetic variants are ascertained across the whole genome in thousands of patients and thousands of controls, and statistical genetic associations (but not causality) with disease are tested.
Effect sizes	Quantitative measures of the strength of events. In genetic association studies, the odds ratio is often used as an effect size measure, as it reflects how many times

	more likely it is to find the susceptibility variant in cases versus controls.
Immunophenotypes	In this article, we define immunophenotypes as quantitative traits that reflect immune function. They include cell-type abundances, cell proliferation, signalling response, and cytokine and chemokine serum protein levels.
Cytokine	A type of small cell-signalling protein that is secreted by some cells and affects the behaviour of others; cytokines are crucial participants in the immune response.
Psoriasis	An autoimmune disease affecting the skin.
Crohn's disease	Inflammatory bowel disease presenting autoreactivity in any part of the gastrointestinal tract.
Ulcerative colitis	Inflammatory bowel disease mainly affecting the colon.
Coeliac disease	An autoimmune disease of the small intestine caused by an immune reaction to wheat.
Linkage analysis	A genetic association method that uses linkage disequilibrium and recombination principles in families to find markers linked to a disease locus (markers co-segregating with the disease phenotype).
Human leukocyte antigen (HLA)	A type of surface protein that presents antigens. HLA class I proteins present antigens from inside the cell and are expressed by most cells, whereas HLA class II proteins are expressed only by antigen-presenting cells and present extracellular antigens to T cells.
Nuclear factor-κB (NF-κB)	An important transcription factor involved in the immune response and B cell development.
Candidate gene studies	A genetic association method that looks for alleles of a candidate gene (for example, based on a priori functional knowledge) that are associated with a disease, typically using a case-control study design.
Linkage disequilibrium (LD)	A phenomenon describing the nonrandom association (or co-segregation) between two alleles of different loci.
***Trans*-ancestral cohorts**	Groups of individuals pertaining to different ancestry populations.
Imputation	Inference of alleles of non-genotyped variants based on nearby observed genotypes and using a reference panel

	from which the correlation structure between variants (or haplotypes) can be learnt.
Missing heritability	A concept referring to the heritability that has not been explained by known disease susceptibility loci (despite so many having been discovered with genome-wide association studies).
Exome sequencing	A technique that captures the protein-coding sequences of the genome (exons) and sequences them.
Enhancers	Distant transcriptional regulatory elements, often regulating expression in a cell-type- and cell-state-specific manner in *cis*, and often marked by specific chromatin modifications.
Super enhancer	Several enhancer marks clustered together; super enhancers tend to be more cell-type specific than typical enhancer marks.
Expression quantitative trait loci (eQTLs)	Regulatory variant whose genotypes correlate with gene expression levels.
QTLs (Quantitative trait loci)	Loci in the genome that are associated with a quantitative trait. For example, a single-nucleotide polymorphism associated with gene expression levels can be called an expression QTL (eQTL).
Mendelian randomization	A method for inferring causality of a trait (or modifiable exposure) on disease, taking advantage of the randomized trial nature of genetic variation (assuming a random mating pattern).
Allele-specific expression	A phenomenon that occurs when one of the two alleles of a gene in an individual is expressed more than the other.

References

1. Mackay IR. Travels and travails of autoimmunity: a historical journey from discovery to rediscovery. Autoimmun. Rev. 2010; 9:A251–A258. [PubMed: 19883799]
2. Hayter SM, Cook MC. Updated assessment of the prevalence, spectrum and case definition of autoimmune disease. Autoimmun. Rev. 2012; 11:754–765. [PubMed: 22387972]
3. Cooper GS, Bynum MLK, Somers EC. Recent insights in the epidemiology of autoimmune diseases: improved prevalence estimates and understanding of clustering of diseases. J. Autoimmun. 2009; 33:197–207. [PubMed: 19819109] This is a good epidemiology review with prevalence estimates for autoimmune diseases.
4. Walsh SJ, Rau LM. Autoimmune diseases: a leading cause of death among young and middle-aged women in the United States. Am. J. Public Health. 2000; 90:1463–1466. [PubMed: 10983209]

5. Thomas SL, Griffiths C, Smeeth L, Rooney C, Hall AJ. Burden of mortality associated with autoimmune diseases among females in the United Kingdom. Am. J. Public Health. 2010; 100:2279–2287. [PubMed: 20864721]
6. Gilkeson G, et al. The United States to Africa lupus prevalence gradient revisited. Lupus. 2011; 20:1095–1103. [PubMed: 21917913]
7. Davis MM. A prescription for human immunology. Immunity. 2008; 29:835–838. [PubMed: 19100694]
8. Cotsapas C, et al. Pervasive sharing of genetic effects in autoimmune disease. PLoS Genet. 2011; 7:e1002254. [PubMed: 21852963] This study uses GWAS data to show extensive sharing of genetic risk for seven autoimmune diseases beyond the MHC and *PTPN22* loci.
9. Bogdanos DP, et al. Twin studies in autoimmune disease: genetics, gender and environment. J. Autoimmun. 2012; 38:J156–J169. [PubMed: 22177232]
10. Willer CJ, et al. Twin concordance and sibling recurrence rates in multiple sclerosis. Proc. Natl Acad. Sci. USA. 2003; 100:12877–12882. [PubMed: 14569025]
11. Hawkes CH, Macgregor AJ. Twin studies and the heritability of MS: a conclusion. Mult. Scler. 2009; 15:661–667. [PubMed: 19482860]
12. Elder JT, et al. The genetics of psoriasis. Arch. Dermatol. 1994; 130:216–224. [PubMed: 8304761]
13. Satsangi J, et al. Contribution of genes of the major histocompatibility complex to susceptibility and disease phenotype in inflammatory bowel disease. Lancet. 1996; 347:1212–1217. [PubMed: 8622450]
14. Vyse TJ, Todd JA. Genetic analysis of autoimmune disease. Cell. 1996; 85:311–318. [PubMed: 8616887]
15. Raychaudhuri S. Recent advances in the genetics of rheumatoid arthritis. Curr. Opin. Rheumatol. 2010; 22:109–118. [PubMed: 20075733]
16. Criswell LA, et al. Analysis of families in the multiple autoimmune disease genetics consortium (MADGC) collection: the *PTPN22* 620W allele associates with multiple autoimmune phenotypes. Am. J. Hum. Genet. 2005; 76:561–571. [PubMed: 15719322]
17. Eaton WW, Rose NR, Kalaydjian A, Pedersen MG, Mortensen PB. Epidemiology of autoimmune diseases in Denmark. J. Autoimmun. 2007; 29:1–9. [PubMed: 17582741]
18. Bulik-Sullivan B, et al. An atlas of genetic correlations across human diseases and traits. Nat. Genet. 2015; 47:1236–1241. [PubMed: 26414676]
19. Rich SS, Weitkamp LR, Barbosa J. Genetic heterogeneity of insulin-dependent (type I) diabetes mellitus: evidence from a study of extended haplotypes. Am. J. Hum. Genet. 1984; 36:1015–1023. [PubMed: 6594040]
20. Gaffney PM, et al. A genome-wide search for susceptibility genes in human systemic lupus erythematosus sib-pair families. Proc. Natl Acad. Sci. USA. 1998; 95:14875–14879. [PubMed: 9843983]
21. Hugot JP, et al. Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's's disease. Nature. 2001; 411:599–603. [PubMed: 11385576]
22. Rioux JD, et al. Genomewide search in Canadian families with inflammatory bowel disease reveals two novel susceptibility loci. Am. J. Hum. Genet. 2000; 66:1863–1870. [PubMed: 10777714]
23. Hu X, et al. Additive and interaction effects at three amino acid positions in HLA-DQ and HLA-DR molecules drive type 1 diabetes risk. Nat. Genet. 2015; 47:898–905. [PubMed: 26168013]
24. The MHC Sequencing Consortium. Complete sequence and gene map of a human major histocompatibility complex. Nature. 1999; 401:921–923. [PubMed: 10553908]
25. Tomfohrde J, et al. Gene for familial psoriasis susceptibility mapped to the distal end of human chromosome 17q. Science. 1994; 264:1141–1145. [PubMed: 8178173]
26. Jordan CT, et al. PSORS2 is due to mutations in *CARD14*. Am. J. Hum. Genet. 2012; 90:784–795. [PubMed: 22521418]
27. Plenge RM, et al. Replication of putative candidate-gene associations with rheumatoid arthritis in>4,000 samples from North America and Sweden: association of susceptibility with *PTPN22*, *CTLA4*, and *PADI4*. Am. J. Hum. Genet. 2005; 77:1044–1060. [PubMed: 16380915]

28. Begovich AB, 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. 2004; 75:330–337. [PubMed: 15208781]
29. Velaga MR, et al. The codon 620 tryptophan allele of the lymphoid tyrosine phosphatase (LYP) gene is a major determinant of Graves' disease. J. Clin. Endocrinol. Metab. 2004; 89:5862–5865. [PubMed: 15531553]
30. Kyogoku C, et al. Genetic association of the R620W polymorphism of protein tyrosine phosphatase PTPN22 with human SLE. Am. J. Hum. Genet. 2004; 75:504–507. [PubMed: 15273934]
31. Bottini N, et al. A functional variant of lymphoid tyrosine phosphatase is associated with type I diabetes. Nat. Genet. 2004; 36:337–338. [PubMed: 15004560]
32. Rawlings DJ, Dai X, Buckner JH. The role of *PTPN22* risk variant in the development of autoimmunity: finding common ground between mouse and human. J. Immunol. 2015; 194:2977–2984. [PubMed: 25795788]
33. Nisticò L, Buzzetti R, Pritchard LA, et al. The CTLA-4 gene region of chromosome 2q33 is linked to, and associated with type 1 diabetes. Hum. Molec. Genet. 1996; 5:1075–1080. [PubMed: 8817351]
34. Petukhova L, et al. Genome-wide association study in alopecia areata implicates both innate and adaptive immunity. Nature. 2010; 466:113–117. [PubMed: 20596022]
35. Fortune MD, et al. Statistical colocalization of genetic risk variants for related autoimmune diseases in the context of common controls. Nat. Genet. 2015; 47:839–846. [PubMed: 26053495]
36. Okada Y, et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature. 2014; 506:376–381. [PubMed: 24390342]
37. Jostins L, et al. Host–microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature. 2012; 491:119–124. [PubMed: 23128233] This large meta-analysis of GWAS reveals the greatest number of susceptibility loci for IBD.
38. Farh KK-H, et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. Nature. 2015; 518:337–343. [PubMed: 25363779] This study fine-maps autoimmune disease variants and integrates them with immune enhancer and promoter marks, as well as with blood eQTLs.
39. Li YR, et al. Meta-analysis of shared genetic architecture across ten pediatric autoimmune diseases. Nat. Med. 2015; 21:1018–1027. [PubMed: 26301688]
40. Barratt BJ, et al. Remapping the insulin gene/*IDDM2* locus in type 1 diabetes. Diabetes. 2004; 53:1884–1889. [PubMed: 15220214]
41. Onengut-Gumuscu S, et al. Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. Nat. Genet. 2015; 47:381–386. [PubMed: 25751624] This paper increases the resolution of the SNPs most associated with T1DM, highlights the genetic relationships between T1DM and other autoimmune diseases, and uses functional annotation to implicate specific cell types and regulatory DNA (enhancers).
42. Parkes M, Cortes A, van Heel DA, Brown MA. Genetic insights into common pathways and complex relationships among immune-mediated diseases. Nat. Rev. Genet. 2013; 14:661–673. [PubMed: 23917628]
43. Ramos PS, Shedlock AM, Langefeld CD. Genetics of autoimmune diseases: insights from population genetics. J. Hum. Genet. 2015; 60:657–664. [PubMed: 26223182]
44. Raj T, et al. Common risk alleles for inflammatory diseases are targets of recent positive selection. Am. J. Hum. Genet. 2013; 92:517–529. [PubMed: 23522783]
45. Cagliani R, et al. Crohn's's disease loci are common targets of protozoa-driven selection. Mol. Biol. Evol. 2013; 30:1077–1087. [PubMed: 23389767]
46. Stastny P. Association of the B-cell alloantigen DRw4 with rheumatoid arthritis. N. Engl. J. Med. 1978; 298:869–871. [PubMed: 147420]
47. Gregersen PK, Silver J, Winchester RJ. The shared epitope hypothesis. An approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis. Arthritis Rheum. 1987; 30:1205–1213. [PubMed: 2446635] This study describes the shared epitope hypothesis as a common thread for MHC effects.

48. Raychaudhuri S, et al. Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. Nat. Genet. 2012; 44:291–296. [PubMed: 22286218] This study shows that the MHC effects for RA resolve into specific amino acids in the peptide-binding grooves of MHC molecules.
49. Okada Y, et al. Risk for ACPA-positive rheumatoid arthritis is driven by shared HLA amino acid polymorphisms in Asian and European populations. Hum. Mol. Genet. 2014; 23:6916–6926. [PubMed: 25070946]
50. Reynolds RJ, et al. HLA-DRB1-associated rheumatoid arthritis risk at multiple levels in African Americans: hierarchical classification systems, amino acid positions, and residues. Arthritis Rheumatol. 2014; 66:3274–3282. [PubMed: 25524867]
51. Liu JZ, et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. Nat. Genet. 2015; 47:979–986. [PubMed: 26192919]
52. Han B, et al. Fine mapping seronegative and seropositive rheumatoid arthritis to shared and distinct HLA alleles by adjusting for the effects of heterogeneity. Am. J. Hum. Genet. 2014; 94:522–532. [PubMed: 24656864]
53. Foo JN, et al. Coding variants at hexa-allelic amino acid 13 of HLA-DRB1 explain independent SNP associations with follicular lymphoma risk. Am. J. Hum. Genet. 2013; 93:167–172. [PubMed: 23791106]
54. Zheng X, et al. HIBAG—HLA genotype imputation with attribute bagging. Pharmacogenomics J. 2014; 14:192–200. [PubMed: 23712092]
55. Leslie S, Donnelly P, McVean G. A statistical method for predicting classical HLA alleles from SNP data. Am. J. Hum. Genet. 2008; 82:48–56. [PubMed: 18179884]
56. Jia X, et al. Imputing amino acid polymorphisms in human leukocyte antigens. PLoS ONE. 2013; 8:e64683. [PubMed: 23762245]
57. International HIV Controllers Study. The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. Science. 2010; 330:1551–1557. [PubMed: 21051598]
58. Okada Y, et al. Construction of a population-specific HLA imputation reference panel and its application to Graves' disease risk in Japanese. Nat. Genet. 2015; 47:798–802. [PubMed: 26029868]
59. Warren RL, et al. Derivation of HLA types from shotgun sequence datasets. Genome Med. 2012; 4:95. [PubMed: 23228053]
60. Liu C, et al. ATHLATES: accurate typing of human leukocyte antigen through exome sequencing. Nucleic Acids Res. 2013; 41:e142. [PubMed: 23748956]
61. Bai Y, Ni M, Cooper B, Wei Y, Fury W. Inference of high resolution HLA types using genome-wide RNA or DNA sequencing reads. BMC Genomics. 2014; 15:325. [PubMed: 24884790]
62. Huang Y, et al. HLA reporter: a tool for HLA typing from next generation sequencing data. Genome Med. 2015; 7:25. [PubMed: 25908942]
63. Dilthey A, Cox C, Iqbal Z, Nelson MR, McVean G. Improved genome inference in the MHC using a population reference graph. Nat. Genet. 2015; 47:682–688. [PubMed: 25915597]
64. Horton R, et al. Variation analysis and gene annotation of eight MHC haplotypes: the MHC Haplotype Project. Immunogenetics. 2008; 60:1–18. [PubMed: 18193213]
65. Holdsworth R, et al. The HLA dictionary 2008: a summary of HLA-A, -B, -C, -DRB1/3/4/5, and -DQB1 alleles and their association with serologically defined HLA-A, -B, -C, -DR, and -DQ antigens. Tissue Antigens. 2009; 73:95–170. [PubMed: 19140825]
66. Cereb N, Kim HR, Ryu J, Yang SY. Advances in DNA sequencing technologies for high resolution HLA typing. Hum. Immunol. 2015; 76:923–927. [PubMed: 26423536]
67. Cortes A, Brown MA. Promise and pitfalls of the Immunochip. Arthritis Res. Ther. 2011; 13:101. [PubMed: 21345260]
68. Trynka G, et al. Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. Nat. Genet. 2011; 43:1193–1201. [PubMed: 22057235]
69. Huang, H., et al. Association mapping of inflammatory bowel disease loci to single variant resolution. bioRxiv. 2015. http://dx.doi.org/10.1101/028688

70. Cooper JD, et al. Seven newly identified loci for autoimmune thyroid disease. Hum. Mol. Genet. 2012; 21:5202–5208. [PubMed: 22922229]
71. International Genetics of Ankylosing Spondylitis Consortium (IGAS). Identification of multiple risk variants for ankylosing spondylitis through high-density genotyping of immune-related loci. Nat. Genet. 2013; 45:730–738. [PubMed: 23749187]
72. Hinks A, et al. Dense genotyping of immune-related disease regions identifies 14 new susceptibility loci for juvenile idiopathic arthritis. Nat. Genet. 2013; 45:664–669. [PubMed: 23603761]
73. Tsoi LC, et al. Identification of 15 new psoriasis susceptibility loci highlights the role of innate immunity. Nat. Genet. 2012; 44:1341–1348. [PubMed: 23143594]
74. Eyre S, et al. High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. Nat. Genet. 2012; 44:1336–1340. [PubMed: 23143596]
75. Wei Z, et al. Large sample size, wide variant spectrum, and advanced machine-learning technique boost risk prediction for inflammatory bowel disease. Am. J. Hum. Genet. 2013; 92:1008–1012. [PubMed: 23731541]
76. Vilhjálmsson BJ, et al. Modeling linkage disequilibrium increases accuracy of polygenic risk scores. Am. J. Hum. Genet. 2015; 97:576–592. [PubMed: 26430803]
77. Peloso, GM., et al. Phenotypic extremes in rare variant study designs. Eur. J. Hum. Genet. 2015. http://dx.doi.org/10.1038/ejhg.2015.197
78. MacArthur DG, et al. A systematic survey of loss-of-function variants in human protein-coding genes. Science. 2012; 335:823–828. [PubMed: 22344438]
79. Hunt KA, et al. Rare and functional *SIAE* variants are not associated with autoimmune disease risk in up to 66,924 individuals of European ancestry. Nat. Genet. 2012; 44:3–5. [PubMed: 22200769]
80. Dickson SP, Wang K, Krantz I, Hakonarson H, Goldstein DB. Rare variants create synthetic genome-wide associations. PLoS Biol. 2010; 8:e1000294. [PubMed: 20126254]
81. Manolio TA, et al. Finding the missing heritability of complex diseases. Nature. 2009; 461:747–753. [PubMed: 19812666]
82. Rivas MA, et al. Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. Nat. Genet. 2011; 43:1066–1073. [PubMed: 21983784] One of the first deep re-sequencing studies to find rare variants involved in IBD, which also assesses their functional consequences.
83. Diogo D, et al. *TYK2* protein-coding variants protect against rheumatoid arthritis and autoimmunity, with no evidence of major pleiotropic effects on non-autoimmune complex traits. PLoS ONE. 2015; 10:e0122271. [PubMed: 25849893]
84. Nejentsev S, Walker N, Riches D, Egholm M, Todd JA. Rare variants of *IFIH1*, a gene implicated in antiviral responses, protect against type 1 diabetes. Science. 2009; 324:387–389. [PubMed: 19264985]
85. Meylan E, Tschopp J, Karin M. Intracellular pattern recognition receptors in the host response. Nature. 2006; 442:39–44. [PubMed: 16823444]
86. Rice GI, et al. Gain-of-function mutations in *IFIH1* cause a spectrum of human disease phenotypes associated with upregulated type I interferon signaling. Nat. Genet. 2014; 46:503–509. [PubMed: 24686847]
87. Beaudoin M, et al. Deep resequencing of GWAS loci identifies rare variants in *CARD9*, *IL23R* and *RNF186* that are associated with ulcerative colitis. PLoS Genet. 2013; 9:e1003723. [PubMed: 24068945]
88. Bang S-Y, et al. Targeted exon sequencing fails to identify rare coding variants with large effect in rheumatoid arthritis. Arthritis Res. Ther. 2014; 16:447. [PubMed: 25267259]
89. Hunt KA, et al. Negligible impact of rare autoimmune-locus coding-region variants on missing heritability. Nature. 2013; 498:232–235. [PubMed: 23698362]
90. Moutsianas L, et al. The power of gene-based rare variant methods to detect disease-associated variation and test hypotheses about complex disease. PLoS Genet. 2015; 11:e1005165. [PubMed: 25906071]
91. Cardinale CJ, Kelsen JR, Baldassano RN, Hakonarson H. Impact of exome sequencing in inflammatory bowel disease. World J. Gastroenterol. 2013; 19:6721–6729. [PubMed: 24187447]

92. Firestein GS. Evolving concepts of rheumatoid arthritis. Nature. 2003; 423:356–361. [PubMed: 12748655]

93. Lee DM, et al. Mast cells: a cellular link between autoantibodies and inflammatory arthritis. Science. 2002; 297:1689–1692. [PubMed: 12215644]

94. Pap T, Müller-Ladner U, Gay RE, Gay S. Fibroblast biology. Role of synovial fibroblasts in the pathogenesis of rheumatoid arthritis. Arthritis Res. 2000; 2:361–367. [PubMed: 11094449]

95. Lefèvre S, et al. Synovial fibroblasts spread rheumatoid arthritis to unaffected joints. Nat. Med. 2009; 15:1414–1420. [PubMed: 19898488]

96. Li P, Spolski R, Liao W, Leonard WJ. Complex interactions of transcription factors in mediating cytokine biology in T cells. Immunol. Rev. 2014; 261:141–156. [PubMed: 25123282]

97. Lovett-Racke AE, Yang Y, Racke MK. Th1 versus Th17: are T cell cytokines relevant in multiple sclerosis? Biochim. Biophys. Acta. 2011; 1812:246–251. [PubMed: 20600875]

98. Heng TSP, et al. The Immunological Genome Project: networks of gene expression in immune cells. Nat. Immunol. 2008; 9:1091–1094. [PubMed: 18800157]

99. Hu X, et al. Integrating autoimmune risk loci with gene-expression data identifies specific pathogenic immune cell subsets. Am. J. Hum. Genet. 2011; 89:682.This is the first study integrating GWAS loci with cell-type-specific gene expression to find pathogenic cell types of autoimmune diseases.

100. Roadmap Epigenomics Consortium. et al. Integrative analysis of 111 reference human epigenomes. Nature. 2015; 518:317–330. [PubMed: 25693563]

101. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012; 489:57–74. [PubMed: 22955616]

102. Andersson R, et al. An atlas of active enhancers across human cell types and tissues. Nature. 2014; 507:455–461. [PubMed: 24670763]

103. Trynka G, et al. Chromatin marks identify critical cell types for fine mapping complex trait variants. Nat. Genet. 2013; 45:124–130. [PubMed: 23263488]

104. Hnisz D, et al. Super-enhancers in the control of cell identity and disease. Cell. 2013; 155:934–947. [PubMed: 24119843]

105. Parker SCJ, et al. Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. Proc. Natl Acad. Sci. USA. 2013; 110:17921–17926. [PubMed: 24127591]

106. Finucane HK, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat. Genet. 2015; 47:1228–1235. [PubMed: 26414678]

107. Adrianto I, et al. Association of a functional variant downstream of *TNFAIP3* with systemic lupus erythematosus. Nat. Genet. 2011; 43:253–258. [PubMed: 21336280]

108. Ritchie GRS, Dunham I, Zeggini E, Flicek P. Functional annotation of noncoding sequence variants. Nat. Methods. 2014; 11:294–296. [PubMed: 24487584]

109. Kircher M, et al. A general framework for estimating the relative pathogenicity of human genetic variants. Nat. Genet. 2014; 46:310–315. [PubMed: 24487276]

110. Gagliano SA, Barnes MR, Weale ME, Knight JA. Bayesian method to incorporate hundreds of functional characteristics with association evidence to improve variant prioritization. PLoS ONE. 2014; 9:e98122. [PubMed: 24844982]

111. Shihab HA, et al. An integrative approach to predicting the functional effects of non-coding and coding sequence variation. Bioinformatics. 2015; 31:1536–1543. [PubMed: 25583119]

112. Trynka G, et al. Disentangling the effects of colocalizing genomic annotations to functionally prioritize non-coding variants within complex-trait loci. Am. J. Hum. Genet. 2015; 97:139–152. [PubMed: 26140449]

113. Kichaev G, et al. Integrating functional data to prioritize causal variants in statistical fine-mapping studies. PLoS Genet. 2014; 10:e1004722. [PubMed: 25357204]

114. Gusev A, et al. Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. Am. J. Hum. Genet. 2014; 95:535–552. [PubMed: 25439723]

115. Vahedi G, et al. Super-enhancers delineate disease-associated regulatory nodes in T cells. Nature. 2015; 520:558–562. [PubMed: 25686607] This paper describes superenhancers in T cells and reveals that superenhancers associated with RA genes are altered with the drug tofacitinib.
116. Stranger BE, et al. Population genomics of human gene expression. Nat. Genet. 2007; 39:1217–1224. [PubMed: 17873874]
117. Fairfax BP, et al. Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. Nat. Genet. 2012; 44:502–510. [PubMed: 22446964]
118. Raj T, et al. Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes. Science. 2014; 344:519–523. [PubMed: 24786080]
119. Lappalainen T, et al. Transcriptome and genome sequencing uncovers functional variation in humans. Nature. 2013; 501:506–511. [PubMed: 24037378]
120. Liang L, et al. A cross-platform analysis of 14,177 expression quantitative trait loci derived from lymphoblastoid cell lines. Genome Res. 2013; 23:716–726. [PubMed: 23345460]
121. Wen X, Luca F, Pique-Regi R. Cross-population joint analysis of eQTLs: fine mapping and functional annotation. PLoS Genet. 2015; 11:e1005176. [PubMed: 25906321]
122. Dimas AS, et al. Common regulatory variation impacts gene expression in a cell type-dependent manner. Science. 2009; 325:1246–1250. [PubMed: 19644074]
123. Grundberg E, et al. Mapping *cis*- and *trans*-regulatory effects across multiple tissues in twins. Nat. Genet. 2012; 44:1084–1089. [PubMed: 22941192]
124. Fu J, et al. Unraveling the regulatory mechanisms underlying tissue-dependent genetic variation of gene expression. PLoS Genet. 2012; 8:e1002431. [PubMed: 22275870]
125. Gutierrez-Arcelus M, et al. Tissue-specific effects of genetic and epigenetic variation on gene regulation and splicing. PLoS Genet. 2015; 11:e1004958. [PubMed: 25634236]
126. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science. 2015; 348:648–660. [PubMed: 25954001]
127. Nica AC, et al. The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. PLoS Genet. 2011; 7:e1002003. [PubMed: 21304890]
128. Hu X, et al. Regulation of gene expression in autoimmune disease loci and the genetic basis of proliferation in $CD4^+$ effector memory T cells. PLoS Genet. 2014; 10:e1004404. [PubMed: 24968232]
129. Petretto E, et al. Heritability and tissue specificity of expression quantitative trait loci. PLoS Genet. 2006; 2:e172. [PubMed: 17054398]
130. Westra H-J, et al. Systematic identification of *trans* eQTLs as putative drivers of known disease associations. Nat. Genet. 2013; 45:1238–1243. [PubMed: 24013639] This study uses two large cohorts to identify and replicate *trans* eQTLs in blood, highlighting an interesting example implicating SLE.
131. Montgomery SB, et al. Transcriptome genetics using second generation sequencing in a Caucasian population. Nature. 2010; 464:773–777. [PubMed: 20220756]
132. Pickrell JK, et al. Understanding mechanisms underlying human gene expression variation with RNA sequencing. Nature. 2010; 464:768–772. [PubMed: 20220758]
133. Fairfax BP, et al. Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. Science. 2014; 343:1246949. [PubMed: 24604202]
134. Lee MN, et al. Common genetic variants modulate pathogen-sensing responses in human dendritic cells. Science. 2014; 343:1246980. [PubMed: 24604203]
135. Ferraro A, et al. Interindividual variation in human T regulatory cells. Proc. Natl Acad. Sci. USA. 2014; 111:E1111–E1120. [PubMed: 24610777]
136. Naranbhai V, et al. Genomic modulators of gene expression in human neutrophils. Nat. Commun. 2015; 6:7545. [PubMed: 26151758]
137. Han J-W, et al. Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus. Nat. Genet. 2009; 41:1234–1237. [PubMed: 19838193]

138. Gateva V, et al. A large-scale replication study identifies *TNIP1*, *PRDM1*, *JAZF1*, *UHRF1BP1* and *IL10* as risk loci for systemic lupus erythematosus. Nat. Genet. 2009; 41:1228–1233. [PubMed: 19838195]
139. Wang S, et al. A functional haplotype of *UBE2L3* confers risk for systemic lupus erythematosus. Genes Immun. 2012; 13:380–387. [PubMed: 22476155]
140. International Consortium for Systemic Lupus Erythematosus Genetics (SLEGEN). et al. Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in *ITGAM*, *PXK*, *KIAA1542* and other loci. Nat. Genet. 2008; 40:204–210. [PubMed: 18204446]
141. Lewis MJ, et al. *UBE2L3* polymorphism amplifies NF-κB activation and promotes plasma cell development, linking linear ubiquitination to multiple autoimmune diseases. Am. J. Hum. Genet. 2015; 96:221–234. [PubMed: 25640675] This study describes an example of a risk locus for SLE that affects cellular phenotypes and subsequent immunophenotypes, one of which is particularly altered in patients.
142. Dermitzakis ET. Cellular genomics for complex traits. Nat. Rev. Genet. 2012; 13:215–220. [PubMed: 22330769]
143. Barreiro LB, et al. Deciphering the genetic architecture of variation in the immune response to *Mycobacterium tuberculosis* infection. Proc. Natl Acad. Sci. USA. 2012; 109:1204–1209. [PubMed: 22233810]
144. Ye CJ, et al. Intersection of population variation and autoimmunity genetics in human T cell activation. Science. 2014; 345:1254665. [PubMed: 25214635] This paper, which is part of the ImmVar project, studies cell-state-specific eQTLs by subjecting $CD4^+$ T cells to various stimuli in individuals of three different ancestral origins.
145. Romanoski CE, et al. Systems genetics analysis of gene-by-environment interactions in human cells. Am. J. Hum. Genet. 2010; 86:399–410. [PubMed: 20170901]
146. Corradin O, et al. Combinatorial effects of multiple enhancer variants in linkage disequilibrium dictate levels of gene expression to confer susceptibility to common traits. Genome Res. 2014; 24:1–13. [PubMed: 24196873]
147. Rakyan VK, Down TA, Balding DJ, Beck S. Epigenome-wide association studies for common human diseases. Nat. Rev. Genet. 2011; 12:529–541. [PubMed: 21747404]
148. Rakyan VK, et al. Identification of type 1 diabetes-associated DNA methylation variable positions that precede disease diagnosis. PLoS Genet. 2011; 7:e1002300. [PubMed: 21980303]
149. Javierre BM, et al. Changes in the pattern of DNA methylation associate with twin discordance in systemic lupus erythematosus. Genome Res. 2010; 20:170–179. [PubMed: 20028698]
150. Gutierrez-Arcelus M, et al. Passive and active DNA methylation and the interplay with genetic variation in gene regulation. eLife. 2013; 2:e00523. [PubMed: 23755361]
151. Kilpinen H, et al. Coordinated effects of sequence variation on DNA binding, chromatin structure, and transcription. Science. 2013; 342:744–747. [PubMed: 24136355]
152. McVicker G, et al. Identification of genetic variants that affect histone modifications in human cells. Science. 2013; 342:747–749. [PubMed: 24136359]
153. Banovich NE, et al. Methylation QTLs are associated with coordinated changes in transcription factor binding, histone modifications, and gene expression levels. PLoS Genet. 2014; 10:e1004663. [PubMed: 25233095]
154. Lemire M, et al. Long-range epigenetic regulation is conferred by genetic variation located at thousands of independent loci. Nat. Commun. 2015; 6:6326. [PubMed: 25716334]
155. Millstein J, Zhang B, Zhu J, Schadt EE. Disentangling molecular relationships with a causal inference test. BMC Genet. 2009; 10:23. [PubMed: 19473544]
156. Liu Y, et al. Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. Nat. Biotechnol. 2013; 31:142–147. [PubMed: 23334450]
157. Fullwood MJ, et al. An oestrogen-receptor-α-bound human chromatin interactome. Nature. 2009; 462:58–64. [PubMed: 19890323]
158. Kraal G, Weissman IL, Butcher EC. Genetic control of T-cell subset representation in inbred mice. Immunogenetics. 1983; 18:585–592. [PubMed: 6606619]

159. Amadori A, et al. Genetic control of the CD4/CD8 T-cell ratio in humans. Nat. Med. 1995; 1:1279–1283. [PubMed: 7489409]

160. Clementi M, et al. CD4 and CD8 T lymphocyte inheritance. Evidence for major autosomal recessive genes. Hum. Genet. 1999; 105:337–342. [PubMed: 10543402]

161. Evans DM, Frazer IH, Martin NG. Genetic and environmental causes of variation in basal levels of blood cells. Twin Res. 1999; 2:250–257. [PubMed: 10723803]

162. Ferreira MAR, et al. Quantitative trait loci for CD4:CD8 lymphocyte ratio are associated with risk of type 1 diabetes and HIV-1 immune control. Am. J. Hum. Genet. 2010; 86:88–92. [PubMed: 20045101]

163. Hall MA, et al. Genetic influence on peripheral blood T lymphocyte levels. Genes Immun. 2000; 1:423–427. [PubMed: 11196672]

164. Nalls MA, et al. Multiple loci are associated with white blood cell phenotypes. PLoS Genet. 2011; 7:e1002113. [PubMed: 21738480]

165. Okada Y, et al. Identification of nine novel loci associated with white blood cell subtypes in a Japanese population. PLoS Genet. 2011; 7:e1002067. [PubMed: 21738478]

166. Ermann J, Rao DA, Teslovich NC, Brenner MB, Raychaudhuri S. Immune cell profiling to guide therapeutic decisions in rheumatic diseases. Nat. Rev. Rheumatol. 2015; 11:541–551. [PubMed: 26034835]

167. Maecker HT, McCoy JP, Nussenblatt R. Standardizing immunophenotyping for the Human Immunology Project. Nat. Rev. Immunol. 2012; 12:191–200. [PubMed: 22343568]

168. Brodin P, et al. Variation in the human immune system is largely driven by non-heritable influences. Cell. 2015; 160:37–47. [PubMed: 25594173] This study measures a wide range of immunophenotypes in >200 twins, estimating the genetic and environmental contributions to each.

169. Orrù V, et al. Genetic variants regulating immune cell levels in health and disease. Cell. 2013; 155:242–256. [PubMed: 24074872] This study reports genetic associations with immune cell type abundances in >1,000 Sardinian individuals.

170. Roederer M, et al. The genetic architecture of the human immune system: a bioresource for autoimmunity and disease pathogenesis. Cell. 2015; 161:387–403. [PubMed: 25772697]

171. Nair RP, et al. Genome-wide scan reveals association of psoriasis with IL-23 and NF-κB pathways. Nat. Genet. 2009; 41:199–204. [PubMed: 19169254]

172. Morahan G, et al. Association of *IL12B* promoter polymorphism with severity of atopic and non-atopic asthma in children. Lancet. 2002; 360:455–459. [PubMed: 12241719]

173. International Multiple Sclerosis Genetics Consortium. et al. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. Nature. 2011; 476:214–219. [PubMed: 21833088]

174. Housley WJ, et al. Genetic variants associated with autoimmunity drive NFκB signaling and responses to inflammatory stimuli. Sci. Transl Med. 2015; 7:291ra93.This paper implicates NF-κB and its pathway in MS by showing alterations of various immunophenotypes driven by risk alleles.

175. Mali P, et al. RNA-guided human genome engineering via Cas9. Science. 2013; 339:823–826. [PubMed: 23287722]

176. Cong L, et al. Multiplex genome engineering using CRISPR/Cas systems. Science. 2013; 339:819–823. [PubMed: 23287718]

177. Jinek M, et al. RNA-programmed genome editing in human cells. eLife. 2013; 2:e00471. [PubMed: 23386978]

178. Belton J-M, et al. Hi-C: a comprehensive technique to capture the conformation of genomes. Methods. 2012; 58:268–276. [PubMed: 22652625]

179. Buil A, et al. Gene–gene and gene–environment interactions detected by transcriptome sequence analysis in twins. Nat. Genet. 2015; 47:88–91. [PubMed: 25436857]

180. Sandberg R. Entering the era of single-cell transcriptomics in biology and medicine. Nat. Methods. 2014; 11:22–24. [PubMed: 24524133]

181. Pai AA, et al. The contribution of RNA decay quantitative trait loci to inter-individual variation in steady-state gene expression levels. PLoS Genet. 2012; 8:e1003000. [PubMed: 23071454]
182. Hause RJ, et al. Identification and validation of genetic variants that influence transcription factor and cell signaling protein levels. Am. J. Hum. Genet. 2014; 95:194–208. [PubMed: 25087611]
183. Battle A, et al. Impact of regulatory variation from RNA to protein. Science. 2015; 347:664–667. [PubMed: 25657249]
184. Wu L, et al. Variation and genetic control of protein abundance in humans. Nature. 2013; 499:79–82. [PubMed: 23676674]
185. Liu Y, et al. Quantitative variability of 342 plasma proteins in a human twin population. Mol. Syst. Biol. 2015; 11:786. [PubMed: 25652787]
186. Garge N, et al. Identification of quantitative trait loci underlying proteome variation in human lymphoblastoid cells. Mol. Cell. Proteomics. 2010; 9:1383–1399. [PubMed: 20179311]
187. Sahni N, et al. Widespread macromolecular interaction perturbations in human genetic disorders. Cell. 2015; 161:647–660. [PubMed: 25910212]
188. Rossin EJ, et al. Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology. PLoS Genet. 2011; 7:e1001273. [PubMed: 21249183]
189. International Multiple Sclerosis Genetics Consortium. Network-based multiple sclerosis pathway analysis with GWAS data from 15,000 cases and 30,000 controls. Am. J. Hum. Genet. 2013; 92:854–865. [PubMed: 23731539]
190. Topol EJ. Individualized medicine from prewomb to tomb. Cell. 2014; 157:241–253. [PubMed: 24679539]
191. Dendrou CA, et al. Cell-specific protein phenotypes for the autoimmune locus *IL2RA* using a genotype-selectable human bioresource. Nat. Genet. 2009; 41:1011–1015. [PubMed: 19701192]
192. De Jager PL, et al. ImmVar project: Insights and design considerations for future studies of `healthy' immune variation. Semin. Immunol. 2015; 27:51–57. [PubMed: 25819567]
193. Klareskog L, Rönnelid J, Lundberg K, Padyukov L, Alfredsson L. Immunity to citrullinated proteins in rheumatoid arthritis. Annu. Rev. Immunol. 2008; 26:651–675. [PubMed: 18173373]
194. Chattopadhyay PK, et al. Quantum dot semiconductor nanocrystals for immunophenotyping by polychromatic flow cytometry. Nat. Med. 2006; 12:972–977. [PubMed: 16862156]
195. Bendall SC, Nolan GP, Roederer M, Chattopadhyay PK. A deep profiler's guide to cytometry. Trends Immunol. 2012; 33:323–332. [PubMed: 22476049]
196. Purohit S, Sharma A, She JX. Luminex and other multiplex high throughput technologies for the identification of, and host response to, environmental triggers of type 1 diabetes. Biomed. Res. Int. 2015; 2015:326918. [PubMed: 25883955]
197. Kondrat RW, McClusky GA, Cooks RG. Multiple reaction monitoring in mass spectrometry/mass spectrometry for direct analysis of complex mixtures. Anal. Chem. 1978; 50:2017–2021.
198. Gillet LC, et al. Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. Mol. Cell. Proteomics. 2012; 11:O111.016717. [PubMed: 22261725]
199. Bowcock AM, Krueger JG. Getting under the skin: the immunogenetics of psoriasis. Nat. Rev. Immunol. 2005; 5:699–711. [PubMed: 16138103]
200. Selmi C, Lu Q, Humble MC. Heritability versus the role of the environment in autoimmunity. J. Autoimmun. 2012; 39:249–252. [PubMed: 22980030]
201. Akkoc N, Khan MA. Overestimation of the prevalence of ankylosing spondylitis in the Berlin study: comment on the article by Braun *et al.* Arthritis Rheum. 2005; 52:4048–4049. [PubMed: 16320356]
202. Ng SC, et al. Epidemiology of spondyloarthritis in the People's Republic of China: review of the literature and commentary. Semin. Arthritis Rheum. 2007; 37:39–47. [PubMed: 17350674]
203. Liu JZ, et al. Dense fine-mapping study identifies new susceptibility loci for primary biliary cirrhosis. Nat. Genet. 2012; 44:1137–1141. [PubMed: 22961000]
204. Villanueva R, Greenberg DA, Davies TF, Tomer Y. Sibling recurrence risk in autoimmune thyroid disease. Thyroid. 2003; 13:761–764. [PubMed: 14558919]

205. Brix TH, Kyvik KO, Christensen K, Hegedüs L. Evidence for a major role of heredity in Graves' disease: a population-based study of two Danish twin cohorts. J. Clin. Endocrinol. Metab. 2001; 86:930–934. [PubMed: 11158069]

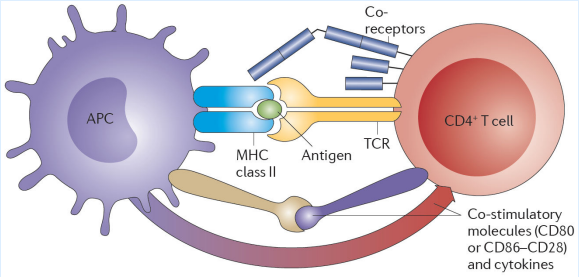
206. Prahalad S, et al. Quantification of the familial contribution to juvenile idiopathic arthritis. Arthritis Rheum. 2010; 62:2525–2529. [PubMed: 20506132]

207. Sadovnick AD, et al. A population-based study of multiple sclerosis in twins: update. Ann. Neurol. 1993; 33:281–285. [PubMed: 8498811]

208. Chen G-B, et al. Estimation and partitioning of (co)heritability of inflammatory bowel disease from GWAS and immunochip data. Hum. Mol. Genet. 2014; 23:4710–4720. [PubMed: 24728037]

209. Jones DE, Watt FE, Metcalf JV, Bassendine MF, James OF. Familial primary biliary cirrhosis reassessed: a geographically-based population study. J. Hepatol. 1999; 30:402–407. [PubMed: 10190721]

Box 1 | Basic principles of autoimmunity



The two major cellular components of the adaptive immune system are B cells and T cells. B cells are antibody-producing cells that mature in the bone marrow. During their development, genomic rearrangements occur to produce a wide range of antibodies that can recognize a diverse antigen repertoire. T cells develop in the thymus and mediate the adaptive immune response by interacting with antigen-presenting cells (such as B cells, or innate immune cells, including dendritic cells and macrophages). Most T cells undergo genomic rearrangements in their α and β T cell receptor (TCR) chains. The TCR recognizes antigens in conjunction with class II major histocompatibility complex (MHC) molecules, which are expressed by antigen presenting cells (APCs; see the figure).TCR recognition of antigens is assisted by co-receptors, such as CD4 and CD3. $CD4^+$ T cell activation also requires additional co-stimulatory molecules, such as T cell expression of CD28, which is the receptor for CD80 or CD86 expressed by APCs. Cytokines are detected by T cells as additional `danger' signals, triggering signalling cascades that regulate immunological response genes and produce more cytokines.

There are several self-tolerance immune mechanisms to protect against B and T cell response to self-antigens. Central tolerance occurs during development, when B or T cells that react strongly to self-antigens are eliminated. However, the human body changes with time, and not all possible self-antigens can be presented in the thymus and bone marrow. Additional peripheral self-tolerance processes exist for when B cells and T cells migrate from their developing organs. For example, if a T cell reacts to a self-antigen but there are no additional `danger' signals, such as cytokines produced by the innate immune system, the cell will be inactivated. Similarly, strong, constant signals are an indication that the antigen presented is a self-antigen (as opposed to a pathogenic antigen, which would usually rapidly increase in concentration) and will not produce an immune reaction.

When self-tolerance mechanisms fail, autoimmunity can emerge (FIG. 1). For example, in type 1 diabetes mellitus (T1DM), the immune system reacts to pancreatic β-cells. In systemic lupus erythematosus (SLE), autoreactivity to DNA and chromatin proteins can occur in a wide range of tissues, including the skin, heart, lungs and blood vessels. Autoimmunity can develop against commensal bacteria in the gut, resulting in inflammatory bowel diseases (IBD). Diseases vary in their autoantibodies based on organ specificity or aetiological mechanism. For example, patients with rheumatoid arthritis (RA) often have anti-citrullinated protein antibodies (ACPAs), which are antibodies against proteins with a post-translational modification that often occurs during

inflammation[193]. However, multiple pathways can lead to autoimmune disease, as RA can develop without the presence of ACPAs.

Box 2 | Immunophenotyping technologies

Flow cytometry

Flow cytometry is a single-cell technique that is used for a wide range of applications owing to its ability to count and sort cells on the basis of a set of markers. Briefly, cells are marked with antibodies conjugated to fluorescent dyes that bind to specific cell-surface or intracellular proteins. The cells are then lined-up in fluid inside a machine, where they pass one-by-one through a laser, which excites the dye molecules. The emission spectra of the dye molecules are then recorded to identify the markers present in each cell. This technique can be used to quantify specific cell-type abundances, to sort specific cellular subsets for subsequent experiments or to measure signalling response if antibodies against the phosphorylated state of a protein are used (a variation known as Phospho-flow). The main limitation of this technique is that even with the most powerful flow cytometry machines, which have four lasers, a maximum of ~20 markers can be used at one time[194]. An extension of this approach can be used to measure cell proliferation; carboxyfluorescein succinimidyl ester (CFSE) is a dye that is easily incorporated and retained inside cells, and its quantity is halved for each cell division. Flow cytometry is then used to quantify this dye in cells. This technology was originally used to detect cell migration in mouse models, but it is now also widely used to measure cell proliferation[128].

Mass cytometry

Mass cytometry also counts cells based on intracellular or cell-surface markers, with the advantage that it can detect over >40 markers simultaneously. It relies on antibodies conjugated to rare earth metals not present in biological samples, which are then detected by a mass spectrometer[195]. Although in this technique the cells die in the process and cannot be used for subsequent experiments, it is instrumental in immunology because it allows quantification of many cell populations in one run[166].

Luminex

Luminex technology has many applications, one of which is the quantification of serum protein levels, such as cytokines and chemokines[168], in a bead-based, multiplex manner. Each bead has a unique internal dye that can be detected by flow cytometry. Additionally, each bead type is coated with an antibody to detect a specific marker (for example, a cytokine). When the samples are run through the beads and the markers of interest are bound to them, the markers are coated with a general dye. The beads are then analysed in a machine that uses lasers to detect the dye of each bead, and the quantity of marker bound to each. The advantage of this technique over others used for protein quantifications, such as enzyme-linked immunosorbent assay (ELISA), is that it can quantify up to 500 markers simultaneously (although it is generally used for about 50) and can be done in a large-scale manner[196].

Mass spectrometry

Mass spectrometry, a technology for peptide detection based on mass-to-charge ratio, has been used to quantify protein levels from plasma, detecting 1,904 peptides pertaining to

342 unique proteins[185]. Accurate quantification is challenging in mass spectrometry, but groups have solved this using relative quantification through labelling peptides with a different isotope per sample (the stable isotope labelling with amino acids in cell culture (SILAC) technique) and by developing methods for direct quantification that involve computational algorithms and machine calibration[197,198].

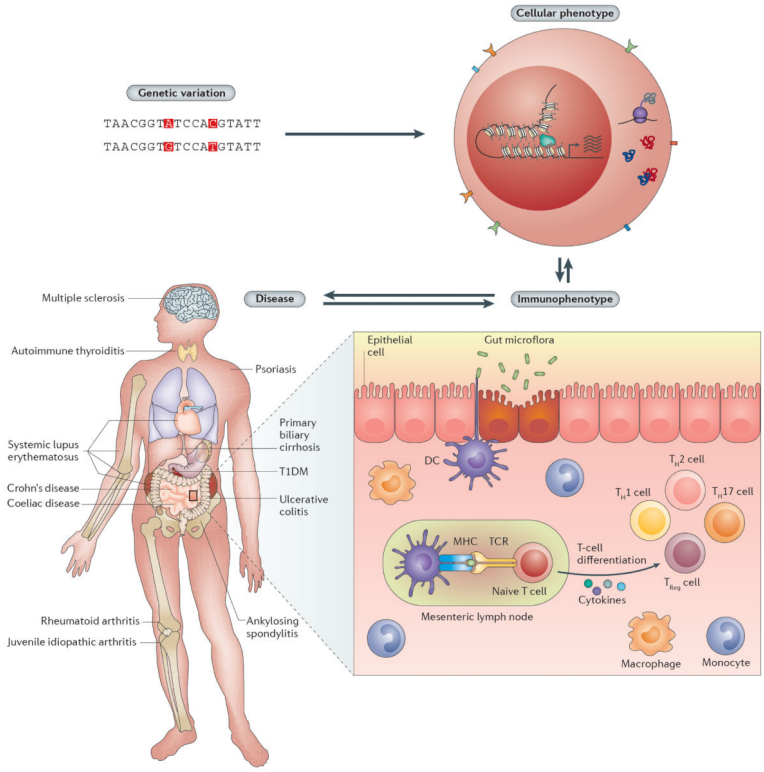


Figure 1. Genetic variation, intermediate immunological phenotypes and disease

Genetic variation (top left) may influence molecular phenotypes, including gene transcription, DNA–DNA interactions, transcription factor binding, histone modifications, DNA methylation, mRNA stability and translation, protein levels, and protein–protein interactions (top right). These cellular processes may affect or be affected by immunophenotypes, such as signalling response, cell-type abundances and cytokine production (bottom right). Immunophenotypes in turn can influence or be influenced by the manifestation of autoimmune diseases and affect different parts of the body (bottom left). DC, dendritic cell; MHC, major histocompatibility complex; TCR, T cell receptor; T_H cell, T helper cell; T_{Reg}, regulatory T cell.

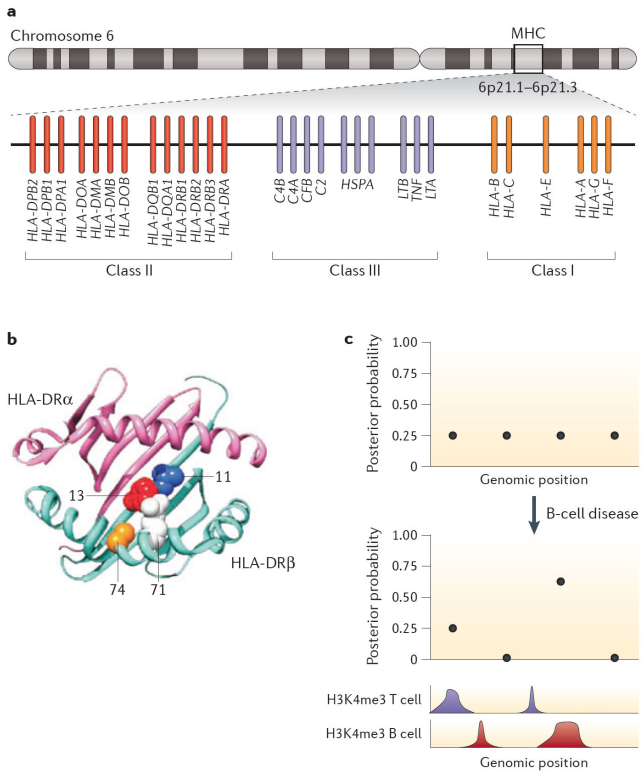


Figure 2. Candidate variant fine-mapping based on functional annotations

Different types of functional annotations, such as missense variants (**a**,**b**) or regulatory marks (**c**), can lead to prioritization of candidate risk variants. **a** | Human leukocyte antigen (HLA) locus in chromosome 6, where genes pertaining to major histocompatibility complex (MHC) class I, II and III are found. **b** | By testing for associations between amino acid residues and rheumatoid arthritis (RA), investigators were able to fine-map independent risk variants that cause changes in amino acids found in the binding pocket of the MHC class II molecule DRβ1. Specifically, ~90% of the MHC risk in RA is attributable to a specific amino acid residue in position 13 at the bottom of the DRβ1 antigen-binding groove, and amino acids 71 and 74 (whose side chains point into the antigen-binding groove) independently modulate RA susceptibility **c** | Other functional annotations, such as histone modifications, can be used to prioritize non-coding candidate risk variants[38,103]. In the hypothetical example shown, four non-coding variants in linkage disequilibrium in a disease susceptibility locus have equal posterior probability of being causative for the disease. However, if one uses information on cell-type-specific regulatory annotation (in this case histone H3 trimethylation on Lys4 (H3K4me3)), and knowledge of the most relevant cell type as the genetic mediator of the disease in question (in this case, B cells), one can assign a higher posterior probability to a variant overlapping a B cell H3K4me3 peak. Part **a** adapted from REF. 199, Nature Publishing Group. Part **b** adapted from REF. 48, Nature Publishing Group.

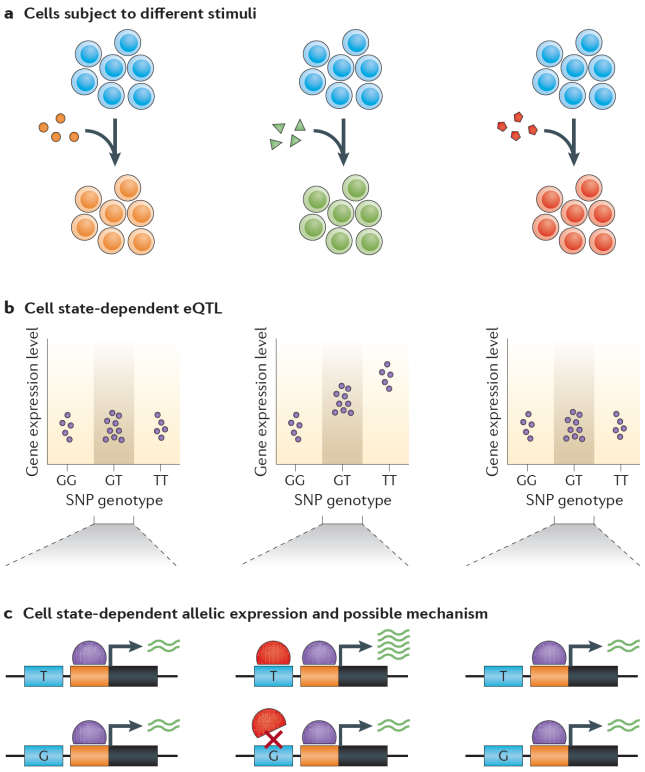


Figure 3. Cell state-dependent eQTLs

a | An immune cell type can be treated with different types of stimuli (such as different cytokines, antigens or non-antigen T cell receptor (TCR) stimulation). **b** | If this is done in many genotyped individuals from a certain population, genetic variants influencing gene expression levels can be found. In this hypothetical example, a single-nucleotide polymorphism (SNP) affects the expression of a gene in the second stimulation condition from part **a** (middle) and not the others (for similar studies, see REFS 117,128,134,143–145). **c** | In heterozygous individuals, allele-specific expression for the affected gene can be observed (see REFS 119,179). A mechanism by which the state-dependent regulatory effect may be acting is by the presence of a transcription factor (red symbol) in the second condition whose regulatory element has a variant that prevents its binding. eQTLs, expression quantitative trait loci.

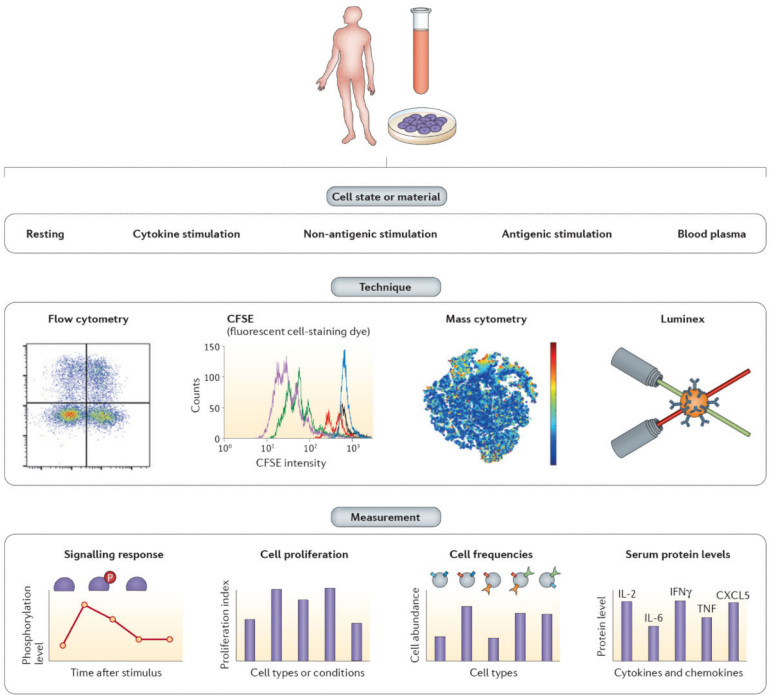


Figure 4. Immunophenotypes

By drawing blood from a single individual, many different immunophenotypes can be measured. These can be measured directly from blood plasma or from cells that can be cultured and subject to different states: resting or under cytokine, non-antigenic or antigenic stimulations (top panel). The petri dish represents stimulation of cells in culture to measure additional response phenotypes. Investigators use different techniques depending on the phenotypes to be measured (middle panel), including flow cytometry, carboxyfluorescein succinimidyl ester (CFSE), mass cytometry, Luminex or mass spectrometry (not shown) (BOX 2). Measurements using these techniques (bottom panel) include signalling response, cell proliferation, cell frequencies and serum protein levels (see REFS 128,168–170 for applications). CXCL5, CXC chemokine ligand 5; IFNγ, interferon-γ; IL, interleukin; TNF, tumour necrosis factor.

Table 1

Prevalence and genetic contribution in autoimmune diseases

Autoimmune disease	No. of loci	Sibling recurrence[14]				Prevalence[3], rate per 100,000 people			Genetic heritability[200] (CIs)	MZ and DZ pairwise concordance[9]	
		Population frequency	Sibling risk	λ_s	MHC λ_s	Denmark[17]	Europe, North America Australia and New Zealand	Asia, Middle East and Latin America		MZ range	DZ range
Ankylosing spondylitis	24	0.13	7	54	4.2	225 (REF 17)	550 (REF 201)	230 (REF 202)	0.97 (0.92–0.99)	50–75	20–27
Crohn's disease	120	0.06	1.2	20	1.3	225	28–53	6.53	1.00 (0.34–1.00), 0.55 *	20–50	3.8–6.5
Primary biliary cirrhosis	20	0.008	0.8	100‡	—	12	4.40	4.18	0.39 (REF 203)	77	0
Systematic lupus erythematosus	20	0.1	2	20	—	32	10–150	19–93§	0.66 *	11.1–40.0	0–4
Autoimmune thyroid disease	9	0.008 (REF. 204)	0.135 (REF 204)	16.9 (REF 204)	—	62//	300–2,980	20–350	0.79 (REF 205)	17.0–22.2	0.0–1.9
Juvenile idiopathic arthritis	23	0.0022 (REF. 206)	0.0257 (REF 206)	11.6 (REF 206)	—	—	120 (REF 206)	—	—	—	—
Psoriasis	35	2.8	17	6	—	197	696–1,527	—	0.66 (0.52–0.77)	35–64	10.14
Type 1 diabetes mellitus	59	0.4	6	15	2.4	946	70–570	31–270	0.88 (0.78–0.94), 0.80 *	13.0–47.4	5.0–11.6
Coeliac disease	41	0.05	3	60	5.2	50	180–1,900	140–900	0.57 (0.32–0.93)¶, 0.87 (0.49–1.00)#	60–75	9.1–11.0
Multiple sclerosis	105	0.1	2	20	2.4	182	50–358	11–101	0.25 (0.00–0.88), 0.76 (0.33–0.88)	5.8–30.8 (REFS 10,11, 207)	2.7–5.4 (REFS 10,11, 207)
Rheumatoid arthritis	81	1	8	8	1.6	381	310–810	120–550	0.68 (0.55–0.79), 0.66 (0.21–0.82) **	0–21	0.0–8.8
Ulcerative colitis	102	0.1	1.2	12	8.3	378	143–294	6–102	0.67 (±0.13)[208]	6.3–18.8	0.0–4.5

Most of the data were taken from the reviews cited at the top of each column section. Data missing from these reviews were added from separate references (cited where appropriate). Numbers are often calculated in different ways and with different assumptions or disease definitions, so the reader should refer to the specific references for details. λ_S, sibling recurrence risk; DZ, dizygotic; MHC, major histocompatibility complex; MZ, monozygotic.

* Heritability rates obtained from known variants analysis.

‡ λ_S is 10.5 in a 1999 study[209].

§ Australians and Aboriginal Australians included in these numbers.

// In Eaton *et al.*[17], autoimmune thyroid disease; in Cooper *et al.* (citing Eaton)[3], hyperthyroidism has 62 and hypothyroidism has 629.

¶ Prevalence 1 in 1,000.

Prevalence 1 in 91.

** First value is for anti-citrullinated protein antibody (ACPA)-positive disease, second value is for ACPA-negative disease.