The Immunogenetic Architecture of Autoimmune Disease

An Goris and Adrian Liston

1Division of Experimental Neurology, University of Leuven, Leuven 3000, Belgium
2VIB and Department of Experimental Medicine, University of Leuven, Leuven 3000, Belgium

Correspondence: an.goris@med.kuleuven.be

The development of most autoimmune diseases includes a strong heritable component. This genetic contribution to disease ranges from simple Mendelian inheritance of causative alleles to the complex interactions of multiple weak loci influencing risk. The genetic variants responsible for disease are being discovered through a range of strategies from linkage studies to genome-wide association studies. Despite the rapid advances in genetic analysis, substantial components of the heritable risk remain unexplained, either owing to the contribution of an as-yet unidentified, “hidden,” component of risk, or through the underappreciated effects of known risk loci. Surprisingly, despite the variation in genetic control, a great deal of conservation appears in the biological processes influenced by risk alleles, with several key immunological pathways being modified in autoimmune diseases covering a broad spectrum of clinical manifestations. The primary translational potential of this knowledge is in the rational design of new therapeutics to exploit the role of these key pathways in influencing disease. With significant further advances in understanding the genetic risk factors and their biological mechanisms, the possibility of genetically tailored (or “personalized”) therapy may be realized.

Autoimmune diseases affect a significant proportion of the population, with >4% of the European population suffering from one or more of these disorders (Vyse and Todd 1996; Cooper et al. 2009; Eaton et al. 2010). Although all autoimmune diseases share similarities in the basic immunological mechanisms, in other aspects, such as clinical manifestation and age of onset, individual diseases vary widely. A few rare autoimmune diseases with Mendelian inheritance patterns within families occur including APS-1 (autoimmune polyendocrine syndrome type 1), IPEX (immunodysregulation, polyendocrinopathy, and enteropathy X-linked) syndrome, and ALPS (autoimmune lymphoproliferative syndrome). Most autoimmune diseases are, however, multifactorial in nature, with susceptibility controlled by multiple genetic and environmental factors.

The genetic component of more common autoimmune diseases can be calculated in several different manners, including the sibling recurrence risk ($\lambda_s$) and the twin concordance rate. The sibling recurrence risk is the ratio of the lifetime risk in siblings of patients to the lifetime population risk, whereas the twin concor-
dance rate measures the proportion of the siblings of affected twins that are also affected. Most common autoimmune diseases, such as multiple sclerosis (MS), type 1 diabetes (T1D), rheumatoid arthritis (RA), and inflammatory bowel disease (IBD) are characterized by a sibling recurrence risk between 6 and 20 (Vyse and Todd 1996), and concordance rates of 25%–50% in monozygotic twins and 2%–12% in dizygotic twins (Cooper et al. 1999). A substantial proportion of relatives may also have subclinical evidence of autoimmunity without developing clinically overt disease. For example, 19% of healthy siblings of MS patients show antibody production in the cerebrospinal fluid, compared to 4% of unrelated healthy controls (Haghghi et al. 2000), whereas 4% of healthy first-degree relatives display lesions that are indistinguishable from those seen in patients and are not seen in unrelated healthy controls (De Stefano et al. 2006). Furthermore, comorbidity with the development of several autoimmune diseases in the same patient and clustering of several autoimmune diseases within families above what is expected by chance appear common (Cooper et al. 2009; Zhernakova et al. 2009). Together these data show a strong genetic component to autoimmune disease development.

STRATEGIES TO IDENTIFY GENETIC FACTORS UNDERLYING AUTOIMMUNE DISEASE

The human genome consists of 3 billion base pairs, and any two genomes typically differ by 0.1% (Kruglyak and Nickerson 2001). Common variants or polymorphisms, where both alleles have a frequency of >1% in the population, account for most (90%) of these differences (Kruglyak and Nickerson 2001; The International HapMap Consortium 2003). Rare variants, each found in <1% of the population, make up the remaining 10% of genetic polymorphism, but are far more numerous than common variants in total counts, given the large size of the human population. In all, 10–15 million common variants and billions of rare sequence variants constitute the human genetic diversity, making the identification of those variants relevant to particular autoimmune diseases challenging (Sawcer 2008).

Different strategies are tailored to each category of variants. Linkage studies search for segregation or coinheritance of a genomic region with disease in families and have been very successful in Mendelian diseases. This strategy has identified mutations in the autoimmune regulator gene (AIRE), the forkhead box P3 gene (FOXP3), and the Fas gene (TNFRSF6) causing APS-1 (The Finnish-German APECED Consortium 1997; Nagamine et al. 1997), IPEX (Bennett et al. 2001; Wildin et al. 2001), and ALPS (Fisher et al. 1995) syndromes, respectively. In most common autoimmune diseases, in contrast, few extended families with Mendelian inheritance patterns exist. Linkage studies have summed signals over multiple smaller families, including affected sibling pairs, instead. These have shown an overall increased allele sharing between affected siblings, confirming the shared genetic risk observed in epidemiological studies, but have only been able to pinpoint precisely those regions exerting large effects on disease risk, such as mutations in NOD2 in Crohn’s disease (CD) (Hugot et al. 2001; Ogura et al. 2001). Indeed, unrealistically high numbers of affected sibling pairs are required to detect the modest effect sizes that we now know are typical for multifactorial disorders by the linkage approach (The International Multiple Sclerosis Genetics Consortium 2005).

Association testing is more suited to detect the contribution of genetic loci to common autoimmune diseases. An association study compares a group of patients and healthy controls and searches for genetic variants that differ in frequency between both groups. Variations in the human leukocyte antigen (HLA) region were the first polymorphisms investigated in association studies (in the 1970s) and turned out to play a major role in most autoimmune diseases, even though precise understanding of the effects is still under investigation owing to the highly polymorphic nature, exceptional linkage disequilibrium, and high gene density of this region (Fernando et al. 2008). Through the investigation of candidate genes, a handful of non-HLA genes influencing risk of autoimmune
disease were identified, e.g., the CTLA4 and PTPN22 genes in T1D (Ueda et al. 2003; Bottini et al. 2004). A major breakthrough came, however, around 2007 with the advent of high-throughput genotyping technology (microarrays), the information from the mapping of the human genome and interindividual genetic variation (International HapMap project), and the realization that the collection of large study populations and the establishment of collaborations is necessary. With these tools, it became possible to undertake genome-wide association studies (GWAS). GWAS are based on the “common disease–common variant” hypothesis, and screen the vast majority (>80%) of common genetic variation for its contribution to disease susceptibility (Hirschhorn and Gajdos 2011). Because of a low prior probability of association amongst the multiple tests performed (typically a few hundred thousand to a million variants screened capture >80% of common variation in the genome), a conservative significance threshold needs to be applied to avoid false positives. It has been shown that results reaching P values <5 × 10⁻⁸ (genome-wide significance level) in a well-powered study of at least 2000 cases and 2000 controls have a high likelihood of being true positive and of replicating in later studies (The Wellcome Trust Case Control Consortium 2007; Sawcer 2008). To date >1200 genomic regions influencing risk for 210 multifactorial diseases have been identified (http://www.genome.gov/gwastudies/ up to 12/2010) and autoimmune diseases have been exceptionally successful, with >30 risk regions each identified in T1D (Barrett et al. 2009), CD (Franke et al. 2010), ulcerative colitis (UC) (Anderson et al. 2011), RA (Stahl et al. 2010), and MS (The International Multiple Sclerosis Genetics Consortium 2011). This suggests a substantial role of common genetic variation in susceptibility to autoimmune diseases, and contrasts with other multifactorial diseases, such as neurodegenerative disorders. The important role of common variation in autoimmune diseases is thought to reflect a history of adaption and selection for variability in regions controlling the immune system, as a defense against intrapopulation pathogen spread (Corona et al. 2010; Casto and Feldman 2011).

**THE EMERGING GENETIC ARCHITECTURE OF COMMON AUTOIMMUNE DISEASES**

GWAS have revealed important conclusions on the genetic architecture of autoimmune diseases. The genetic architecture refers to the number of risk variants, their frequencies in the population (risk allele frequency), and the risks on disease they confer (odds ratio). The odds ratio estimates the fold change in risk for carriers versus noncarriers of a risk allele. Most variants identified follow a multiplicative model, i.e., carrying two copies of the risk allele amounts to the square of the risk of carrying one copy. The genetic architecture of multifactorial autoimmune diseases as emerging from GWAS appears to follow the L shape predicted on the basis of theoretical modeling and animal data (Barton and Keightley 2002). Few if any regions exert a large effect on the risk of disease (odds ratio >3), and a handful have modest effect sizes (odds ratios >2) (Hindorff et al. 2009). In most autoimmune diseases these important loci include the HLA region (Fernando et al. 2008); other examples are NOD2 and IL23R in CD (Hugot et al. 2001; Ogura et al. 2001; Duerr et al. 2006), and PTPN22 in T1D and RA (Begovich et al. 2004; Bottini et al. 2004). Subsequently, there is a long tail of loci with odds ratios of around 1.2 or smaller (Hindorff et al. 2009). Based on currently identified loci, theoretical models predict that many more risk variants are hidden below the genome-wide significance threshold in current GWAS and that a large number of susceptibility loci, possibly thousands, with very small effect sizes may contribute to multifactorial disease (Park et al. 2010). Indeed, the presence of such a polygenic or “en masse” effect has already been shown in MS (The International Multiple Sclerosis Genetics Consortium et al. 2010).

Most of the already identified risk variants are common in the general population (Hindorff et al. 2009), although examples of both common and less common variants in the same gene influencing risk have been reported, e.g., for IL23R in CD (Momozawa et al. 2011) and IFIH1 in T1D (Nejentsev et al. 2009). For many variants, the risk allele is even more common
than the other, protective, allele. This appears counterintuitive at first instance but underlines once again the multifactorial nature of most autoimmune diseases. Immune and infectious agents have been recognized as amongst the strongest selection pressures in human evolution and several autoimmune-associated genes show signs of positive selection, favoring either the protective or risk allele, depending on the case (Hindorff et al. 2009; Corona et al. 2010; Casto and Feldman 2011). Whereas Mendelian disease mutations are controlled by negative selection, risk variants in complex diseases may at least in part be tied to evolutionary adaptations (Blekhman et al. 2008). Pleiotropic effects of these variants may contribute to adaptation and selection, or environmental changes in modern societies may expose the disease risk associated with the variants (Manolio et al. 2009). As an example, whereas the majority of the individuals carry the T1D risk alleles at IFIH1, the rare protective alleles are functionally deleterious and may reduce the ability to mediate an immune response against enterovirus infection (Nejentsev et al. 2009).

Under a polygenic model with many variants independently contributing to disease, the logarithm of the risk is normally distributed in the population (Fig. 1) (Pharoah et al. 2002; Clayton 2009; Sawcer et al. 2010). The genetic risk or burden can be thought of as the number of risk alleles weighted by their effect on disease risk. Most individuals carry a mix of protective and risk alleles and will have an average genetic burden. However, a small subset of individuals differs substantially in genetic risk, e.g., individuals in the 99th percentile versus the first percentile. The distributions of the logarithm of the risk in the general population and in individuals who will effectively go on to develop the disease have the same shape with the average genetic burden being higher in patients than

---

**Figure 1.** Genetic risk distribution in cases and controls. (A) The effectiveness of a hypothetical “diagnostic chip” measuring all genetic variants (including currently unknown variants) can be assessed by considering the distribution of genetic risk at birth among the general population (blue) and affected individuals (red) under a multiplicative model with many risk variants involved (Pharoah et al. 2002; Clayton 2009; Sawcer et al. 2010). The probability distribution of risk is calculated for a population with a prevalence ($\frac{1}{1000}$) and a $\lambda_s$ of 10. The variance in risk observed ($\frac{1}{2} \ln \lambda_s$) is the same for the general population and for individuals who go on to develop the disease. The distribution in cases is, however, shifted toward a higher risk, with a magnitude equal to the variance. (B) The number of false-negative and false-positive diagnostic calls is dictated by the absolute number of healthy (blue) and affected (red) individuals at any given level of risk. Under this scenario a diagnostic chip with a false-negative rate of 50% (i.e., capable of detecting 50% of affected patients) would have only a 3% true-positive rate, as it would produce a positive result for all individuals in the highest 1.6% risk bracket, of which 97% would still come from the healthy population (false positives).
in the general population. However, the distributions overlap to an extent that is dependent on the heritability of the disease (Sawcer et al. 2010). For a typical autoimmune disease with a sibling recurrence risk ($\lambda_s$) of 10, approximately two-thirds of the patients have a genetic risk that is higher than the 95th percentile of the genetic risk in healthy controls (Sawcer et al. 2010). Families with multiple affected relatives appear to share common risk alleles with sporadic patients, but may have a higher genetic load (D’Netto et al. 2009; Gourraud et al. 2011).

A consequence of the polygenic model for complex diseases is that patients are inevitably highly heterogeneous in terms of the particular set of risk alleles they carry (Janssens and van Duijn 2008; Sawcer et al. 2010). It has been suggested that this may translate in different genetically determined disease mechanisms in subgroups of patients (heterogeneity) or a common disease mechanism that is complemented by additional pathways that are more or less predominant in different subgroups (complexity). This concept of heterogeneity or complexity is an important emerging question. A first example of such complexity is found in RA, where the strong association with the HLA-DRB1 risk alleles sharing a specific epitope at positions 70–74 (shared epitope alleles) is specific to the subset of patients producing antibodies to citrullinated peptide antigens (ACPAs) (van der Helm-van Mil et al. 2007; Padyukov et al. 2011). The two forms of the disease are clinically indistinguishable at onset but ACPA positivity is the factor most predictive of future prognosis (van der Helm-van Mil et al. 2007).

In summary, in contrast to rare Mendelian mutations causing childhood-onset autoimmune diseases, the genetic architecture of multifactorial autoimmune diseases is highly complex. Susceptibility rather than causality results from the combined effects of many variants, most of them common in the general population, that each exert a small effect on risk. Many different combinations of risk alleles are able to independently generate a high level of disease risk, without individual loci being necessary or sufficient for the development of disease.

FROM GENETICS TO IMMUNOLOGY

Despite the enormous diversity of genetic variants contributing to susceptibility from individual to individual with the same autoimmune disease, a high degree of similarity is observed between autoimmune diseases in the pool of genetic variants associated with disease. These genetic associations can be dissected into three biologically distinct classes. First, for most autoimmune diseases, the strongest genetic associations are with the HLA locus. Although some HLA haplotype linkages are shared between autoimmune diseases, most HLA associations seem to be specific for a disease (Trowsdale 2011). Second, many of the strong loci that have been associated with one autoimmune disease are also found to be involved in multiple other autoimmune diseases (Makaroff et al. 2008). Typically, the allele of these shared variants that causes risk in one autoimmune disease also causes risk in other autoimmune diseases, indicating that the alleles participate in a shared immunological process common to the development of multiple, clinically distinct, autoimmune diseases (Makaroff et al. 2008). There is a degree of complexity, however, in which precise mechanisms may differ between subgroups of autoimmune diseases. For example, variants in the PTPN22 gene have been associated, with various degrees of risk, with T1D (Smyth et al. 2004), autoimmune thyroid disease (AITD) (Smyth et al. 2004), and RA (The Wellcome Trust Case Control Consortium 2007), among other diseases but are protective against CD (Barrett et al. 2009). Third, a smaller number of associations are observed that are specific to a single autoimmune disease, with no measured impact on other autoimmune diseases, suggesting that they drive a target organ-specific pathway toward disease. These three classes will be discussed in turn.

HLA Associations in Autoimmune Disease

Most, if not all, autoimmune conditions appear associated with the HLA region, if sufficiently large study populations are investigated, and this region is the strongest genetic component
in most of these diseases. The HLA region spans approximately 4 Mb, and is the most gene-dense region in the human genome, containing around 250 genes of which 40% are predicted to exert a function in the immune system (The MHC Sequencing Consortium 1999). The region is characterized by an exceptionally high degree of polymorphism, with >1000 alleles known for HLA-A and -B, for example, and the most extensive degree of linkage disequilibrium within the genome. These features have hampered fine mapping of the functional variants in this region, even decades after the first reports of association. The classic HLA class I (HLA-A, -B, and -C) and class II (HLA-DR, -DQ, and -DP) gene clusters involved in antigen processing and presentation are best characterized and show the strongest association.

Unlike many other autoimmune risk variants (see below), most HLA variants have associations with specific autoimmune diseases, with different variants providing susceptibility to different diseases (Trowsdale 2011). For example, the HLA-DR2 haplotype is protective against T1D, yet predisposes toward MS and SLE (Brand et al. 2005). The unique amino acid sequences of HLA class I and class II alleles determine to a large extent the ability to respond to an antigen, whether foreign or self, and the nature of that response. It is thought that this binding of specific peptides or peptide modifications provides an explanation for their contribution to autoimmune disease. Alterations in binding efficiency may change autoimmune risk in several different manners. By increasing the binding efficiency of otherwise sequestered peptide, HLA risk variants may allow the presentation of potentially immunogenic peptides. Alternatively, by reducing the binding efficiency of particular peptides, HLA risk variants may allow autoreactive T-cell clones to escape tolerance mechanisms. A likely example of the first is the T1D risk allele of HLA-DQ8, which has enhanced binding of insulin peptides, resulting in peripheral activation of autoreactive T cells (Faas and Trucco 1994; Wen et al. 2002). An example of the latter is the MS risk allele of HLA-DR2, which shows poor binding of myelin basic protein, likely causing a failure in thymic tolerance mechanisms and activation of weakly binding autoreactive T cells in the periphery (Li et al. 2005; Maynard et al. 2005). Although HLA has the strongest linkage to particular autoimmune diseases, it is likely that most of the general autoimmune genetic susceptibility component is mediated by non-HLA risk alleles (see below), with HLA risk variants altering the target of autoimmunity. This phenomenon is observed in the nonobese diabetic mouse model, where the replacement of the T1D MHC risk variant, H2IR, with H2IR reroutes the autoimmune response from the pancreas to the thyroid (Braley-Mullen et al. 1999).

Common Autoimmune Risk Variants

Many of the gene variants associated with risk in one autoimmune disease have been shown to also have an association with the development of additional autoimmune diseases. This shared association, and the observation that multiple autoimmune diseases track in the same patients and same families, has led to the model of a cluster of gene variants that decrease immunological tolerance. A large number of genes fall into the category of general autoimmune risk variants, however, it is likely that only a few key immunological pathways are affected.

One immunological phenomenon that is commonly associated with autoimmunity is that of partial T-cell immunodeficiency (Liston et al. 2008). Many of the rare Mendelian autoimmune diseases have a dual association of partial T-cell immunodeficiency and autoimmunity or immune dysregulation, of which Omenn syndrome can be considered a classical example. Partial loss-of-function mutations in RAG1, RAG2, DCLRE1C, IL7RA, and RMRP all produce lymphopenia and result in excessive T-cell activation (Liston et al. 2008). Although these are all examples of single gene variants being sufficient to produce Mendelian autoimmunity, milder gene variants are likely to produce a partial T-cell immunodeficiency, which, in combination with additional genetic insults, may predispose to autoimmunity. The best example of this phenomenon is the association of IL7RA polymorphisms with several common
autoimmune diseases, including MS and T1D (Gregory et al. 2007; Todd et al. 2007). The variants that result in Omenn syndrome show severe loss of function, whereas the variant associated with common autoimmune disease alters splicing efficiency, creating a soluble competitive antagonist form of the receptor (Gregory et al. 2007). This reduction in available IL-7 in turn reduces thymic output of T cells (Broux et al. 2010), replicating on a less severe form the immunological defect present in Omenn syndrome. Similarly, the gain-of-function PTPN22 allele associated with multiple autoimmune diseases results in a reduction in T-cell receptor (TCR) signaling (Smerdel et al. 2004; Vang et al. 2005), and may therefore also contribute to partial immunodeficiency.

Another Mendelian autoimmune disease, IPEX, shows a separate mechanism by which risk variants may increase susceptibility to autoimmunity. IPEX is caused by mutants in FOXP3 (Bennett et al. 2001; Wildin et al. 2001), which in turn eliminate the production of regulatory T cells (Fontenot et al. 2003; Hori et al. 2003; Khattri et al. 2003). The absence of regulatory T cells results in autoreactive T cells undergoing uncontrolled activation, driving multiorgan autoimmunity (Kim et al. 2007). Again, this Mendelian disease finds parallels in common autoimmunity. IL-2 is important for the survival of regulatory T cells (Sadlack et al. 1993; Kramer et al. 1995; Suzuki et al. 1995, 1999; Willerford et al. 1995; Almeida et al. 2002), and variants in IL2RA are associated with T1D and MS. It has been shown that the T1D risk variant of IL2RA decreases receptor expression, which may result in defective regulatory T-cell biology (Wing et al. 2008), so it is notable that a splice variant of CTLA4 associated with autoimmune diseases has reduced production of a tolerogenic soluble isoform (Ueda et al. 2003).

A third cluster of risk variants is observed in cytokine and cytokine signaling genes, including STAT4, TNFA, IL12B, IL23R, and IL10 (Makaroff et al. 2008). Because many autoimmune diseases are strongly associated with a particular effector T-cell lineage (e.g., Th1, Th2, or Th17), it is likely that these variants alter the ease with which naive T cells enter a pathogenic lineage. For example, IBD and psoriasis are both associated with Th17 infiltrate and both show associations with a disease risk allele in IL23R. As IL-23 is critical for the differentiation of Th17 cells (McGeachy et al. 2007), it is likely that this risk variant increases the generation of Th17 cells, and hence the development of IBD and psoriasis. In addition to the three pathways discussed above, there are multiple other cases where the known biology of associated risk variants suggests a common function in impeding tolerance. However, the examples of immunodeficiency, regulatory T cells, and effector T-cell differentiation are sufficient to show the general rule.

Disease-Specific Risk Variants

In addition to HLA and the shared autoimmune variants, additional risk variants are associated with single autoimmune diseases. Unlike the shared autoimmune variants, which are largely in genes expressed within leukocytes, the disease-specific risk variants are often expressed by the target organ. Thus, for example, polymorphisms in INS are associated with T1D, ATG16L1 variants with IBD, and KAZALD1 variants with RA. There are two main biological mechanisms by which these variants may drive disease, and often it is not clear for a given variant which mechanism is active (Makaroff et al. 2008).

The first mechanism is that of a target antigen gene variant which results in poorer tolerance induction. Studies of the rare Mendelian autoimmune disease APS-1 have shown that thymic expression of target organ antigens is essential to prevent autoimmunity (Liston et al. 2003; Liston et al. 2004a; DeVoss et al. 2006). In the case of APS-1, mutations in the gene AIRE destroy entire networks of thymic antigen expression, driving disease against multiple targets (The Finnish-German APECED Consortium 1997; Nagamine et al. 1997). More pertinent to the common autoimmune diseases is when expression of single targets is reduced in the thymus. The best example of this phenomenon is
the T1D INS risk variant, which has a two- to threefold decrease in thymic expression of insulin (Vafadis et al. 1997; Taubert et al. 2007).

The second mechanism by which disease-specific variants can drive autoimmune is through the modification of the target organ. For example, ATG16L1 is involved in the barrier function of the gut, and so the association of variants with IBD risk may represent a failure to prevent inflammatory bacterial infections (Rioux et al. 2007). In a similar vein, KAZALD1 is involved in joint regeneration, so the observed association of variants with RA risk (The Wellcome Trust Case Control Consortium 2007) may represent reduced ability of the joints to heal after autoimmune attack. Until further biological analysis of the mechanism of disease-specific variants is completed, explanations of the pathway to risk will largely remain “Just So” stories, with plausible models that invoke differentiation tolerance induction or differences in target organ function.

FROM IMMUNOLOGY TO GENETICS

Just as genetic data have been used to further understand the biology of autoimmunity, immunological data can be used to further understand the genetics of autoimmunity. Although most GWAS have used clinical presentation as the phenotype, it is becoming increasingly recognized that GWAS can equally use intermediate biological phenotypes as the mapping subject, such as immunological parameters. These “endophenotypes” are quantitative measures that have a biologically plausible mechanism leading to disease. Endophenotypes are thought to be more closely related to gene function than clinical end point phenotypes. As subtraits, endophenotypes are expected to have a simpler genetic architecture than the final clinical manifestation, with fewer influencing variants each of larger effect, making them easier to dissect (Weidinger et al. 2010). Pilot studies in mice suggest that this is indeed the case (Liston et al. 2004b). Furthermore, endophenotypes can be measured within the healthy population, providing more scope for understanding the genetics of the less common autoimmune diseases and eliminating the confounding secondary effects of autoimmune disease on biological processes. GWAS of immunological traits underlying asthma, i.e., IgE production and eosinophil numbers, have been successful in identifying association with single-nucleotide polymorphisms (SNPs) involved in asthma risk, providing critical immunological information of the mechanism of these SNPs (Weidinger et al. 2008; Gudbjartsson et al. 2009). Similar studies of immunological parameters associated with autoimmune diseases are under way, and not only have the potential to provide mechanistic information on already known SNPs, but also the capacity to identify new SNPs with greater power.

MISSING OR HIDDEN HERITABILITY?

Despite enormous progress over recent years, only a fraction of the variance in risk between individuals that has been estimated from epidemiological studies can be explained by known genetic risk factors. Typically, several tens of variants explain at most 25% of the variance (Park et al. 2010). The remainder is described as the “missing heritability” and represents a current challenge in complex genetics (Manolio et al. 2009). Several explanations can be offered. First, additional genetic risk factors may remain to be discovered, having been hidden from previous studies by technical constraints. Alternatively, the known genetic component may account for far more of the variance in susceptibility than currently recognized.

Several different sources can be postulated to make up the hidden genetic risk:

1. The identification of weak effect common variants is limited by study design. Although the variants with the strongest effect on disease susceptibility have been identified, larger studies may identify more variants with a more modest effect (Park et al. 2010; Yang et al. 2010). The effect of these variants can be predicted even without identification, but is unlikely to increase the proportion of variance explained above 50% (Park et al. 2010; The International Multiple Sclerosis Genetics Consortium et al. 2010; Yang et al. 2010).
2. The effect of less common (1%–5%) and rare (<1%) variants is yet to be fully calculated. If “common” diseases actually constitute a heterogeneous set of clinically similar diseases, multiple distinct rare variants may be having a large effect on susceptibility, while simultaneously evading detection by linkage and GWAS approaches (Smith and Lusis 2002). For example, rare mutations in the exonuclease gene TREX1 are found in 0.5% of SLE patients (Namjou et al. 2011), and mutations leading to defective expression of the sialic acid acetylhydrolase (SIAE) are found in 2.4% of patients with autoimmune disease (Surolia et al. 2010). However, the linkage studies in affected sibling pairs provide upper estimates for the effect size that can be expected from a yet unknown rare variant (Sawcer 2008; Forabosco et al. 2009). A rare variant with a frequency of 0.2% and an odds ratio of 20 would result in >60% allele sharing in siblings and should have been detected in linkage studies using a few hundred affected sib pairs (Risch and Merikangas 1996). Simulation studies suggest that there are fewer rare variants with large effect sizes (odds ratios >3) in autoimmune diseases than in other multifactorial disorders (Nejentsev et al. 2009; Mozawa et al. 2011). However, rare variants with weak to moderate effects may have a substantial cumulative effect and are unlikely to be detected using any association approach.

3. The contribution of other types of human genetic variation than sequence variation is yet to be thoroughly investigated. Variation such as copy number variation (CNV), inversions, and complex rearrangements could have an impact on disease susceptibility that has not been picked up in association studies that focus on SNPs. Early data suggest that the net contribution of CNV will be less than that observed with SNPs (The Wellcome Trust Case Control Consortium 2010). The effect of other non-SNP genetic variation will be harder to assess, especially inherited epigenetic variation, which could conceivably have a large impact on disease susceptibility (Meda et al. 2011).

Alternatively, there are multiple explanations by which the effect of known genetic variation on susceptibility may be systematically underestimated:

1. Measures of familial clustering such as the sibling recurrence risk (λs) used for estimation of the total variance tend to be biased upward by an overestimated lifetime risk in siblings in the numerator and an underestimated lifetime risk in the general population in the denominator (Guo 1998; Sawcer et al. 2010).

2. Gene–gene or gene–environment interactions may contribute to part of the missing heritability. The observation of monozygotic concordance rates of less than 100% (typically 25%–50% [Cooper et al. 1999]) shows the influence of either environmental or stochastic events (such as TCR rearrangement) in determining disease development. As environmental variation between dizygotic twins is lower than environmental variation of the population at large, the genetic contribution of risk loci influenced by environmental effects may be reduced in GWAS studies, compared to the total genetic risk calculated from twin concordance rates. Known examples of gene–gene interaction include the risk alleles at HLA-C and ERAP1 in psoriasis (Strange et al. 2010), whereas gene–environment interactions have been observed in RA (Mahdi et al. 2009). Many additional gene–environment interactions are predicted, with variants in key infection-sensing genes, such as NOD2 (Franke et al. 2010), IFIH1 (Smyth et al. 2006; Nejentsev et al. 2009), and FUT2 (McGovern et al. 2010), being associated with autoimmunity.

3. Limitations inherent in current GWAS may underreport the association effect. For example, the SNPs found to be associated with disease may not be the functional SNPs, and thus the effect will be diluted by imperfect correlation (Yang et al. 2010). Similarly, the most associated SNP may not capture the
entire contribution from a locus if there are several risk alleles clustered, as observed frequently in both mice (Fraser et al. 2010) and humans (Ioannidis et al. 2009; Zhernakova et al. 2009). In particular, risk variants at the same locus that are in negative linkage disequilibrium with each other may mask each other’s effect, again as observed in mice (McDuffie 2000). Extensive fine-mapping studies investigating regions of association, such as those using a custom genotyping array covering autoimmune-associated genes developed by the Immunochip consortium of autoimmune disease research groups (Cortes and Brown 2011), are ongoing and will elucidate more precisely the contribution of associated loci.

CONCLUDING REMARKS

Enormous advances have been made in dissecting the genetics of autoimmunity over the past decade, both in identifying genetic risk factors and in understanding the cellular basis for their effect. The ultimate aim of these studies was, of course, to mitigate disease in the clinic. There are three main ways in which our increased understanding of the genetics of autoimmunity has potential for translation—through the identification of drug targets, through genetic identification of at-risk individuals, and through personalized medicine.

With a shown impact on disease development, genes identified to contain disease-associated variation constitute high-quality candidates for therapeutic drug design. Examples of drug design following genetic association are highly limited, however, the reverse shows that the principle of associated genes being high-quality candidates is valid. Several examples of the latter exist, such as the identification of two genes that are targets for current treatments in recent genome-wide association screens for MS (The International Multiple Sclerosis Genetics Consortium 2007; The International Multiple Sclerosis Genetics Consortium and The Wellcome Trust Case Control Consortium 2 2011). With genetic data maturing, the forward drug-design strategy is anticipated to be productive in the near future.

The prospect of translating genetic knowledge into diagnostic identification of at-risk individuals is less rosy. The large component of “missing” genetic risk obviously limits diagnostic accuracy (Kraft et al. 2009). However, even if all genetic risk variants were known the potential for diagnostics is limited. A hypothetical “diagnostic chip” containing all genetic risk factors may give a reasonable sensitivity (proportion of people who are correctly classified as high risk among those individuals who actually develop the disease) and specificity (proportion of people who are correctly classified as low risk amongst people who do not develop the disease) (Clayton 2009). However, positive and negative predictive values (proportion of people classified as high risk who develop the disease and the proportion of people classified as low risk who do not develop the disease) are limited by the low prevalence of most autoimmune diseases, such that even the hypothetical perfect chip would have low diagnostic value (see Fig. 1) (Kraft and Hunter 2009; Sawcer et al. 2010). These predictive values are not expected to be much better than those based on traditional risk factors, such as a family history of the disease (De Jager et al. 2009; Sawcer et al. 2010), except within specific cohorts (Ziegler and Neapom 2010). The value of a diagnostic chip may be increased if gene–environment interactions are understood and both genetic and environmental data was available for individual risk assessment, which is one possible avenue to broad diagnostic utility.

Last, our understanding of the genetics of autoimmunity may translate to personalized medicine, i.e., providing each patient with the right treatment at the right time. Studies correlating genetic factors with treatment response or side effects are in their childhood as yet, and the collection of large study populations and detailed clinical follow-up is a challenge. One of the first examples suggests that genetically determined increased expression levels of interleukin-21 contribute to the development of secondary autoimmune diseases, an important side effect on treatment of MS patients with alemtuzumab (Jones et al. 2009). If further correlations like this are elucidated, the genetic profile of individual
autoimmune patients may be used to recommend or contraindicate particular treatment regimes.

ACKNOWLEDGMENTS

A.G. is supported by the Belgian Neurological Society, the Belgian Charcot Foundation, Wetenschappelijk Onderzoek Multiple Sclerose (WOMS) — Vlaanderen, and Onderzoeksfonds K.U. Leuven/Research Fund K.U. Leuven (OT/11/087). A.L. is supported by the Juvenile Diabetes Research Foundation, European Research Council, FWO, and VIB.

REFERENCES


Genetic Factors Controlling Autoimmunity


A. Goris and A. Liston


The Wellcome Trust Case Control Consortium 2. 2010. Genomewide association study of CNVs in 16,000 cases of eight common diseases and 3000 shared controls. *Nature* **466**: 713–720.


