Abstract

In addition to the HLA locus, over 30 genetic loci have been convincingly associated with risk for rheumatoid arthritis, and the majority of these associations have been identified in the last four years. Although this is a remarkable accomplishment, the majority of the genetic risk for RA still remains to be identified. Some of this “missing heritability” will likely be due to rare genetic variation, and will require extensive resequencing of the genomes of patients with RA. In addition, with few exceptions, the function and role in disease pathogenesis of the newly defined risk genes is unknown. Thus, the initial harvest of RA loci will catalyze new lines of hypothesis driven research to determine their role in disease pathogenesis. In addition, the rapidly advancing genetic technologies should lead to a more complete definition of the genetic underpinnings of RA in the next few years.

It is now obvious to even the general public that the genomic era has arrived. However, some commentators have raised the possibility that, at least in terms of mapping disease genes, the genomic era has already come and gone without much of an impact on medicine.1 This pessimistic view is derived, in part, from the relatively modest number and modest influence of risk genes that have been defined thus far in some disease areas, such as psychiatric disorders and type II diabetes. However, autoimmune disorders, including rheumatoid arthritis (RA), are at the other end of the spectrum of success.2 More than 150 loci have confirmed associations with various autoimmune diseases, and we are still in the early stages of digesting the substantial harvest of new genetic findings for these diseases. It is also clear that there is much more to be done in terms of defining the genetic underpinnings of autoimmunity, and the next stage of the genomic era is likely to yield additional important insights into the genetic architecture of RA. In this brief review, we will consider the current data on the genetics of RA in the context of the rapidly changing landscape of genetics research.

RA Genetics Timeline: A Long Way from HLA

There has been an exponential increase in the number of genes associated with RA in the last several years, as shown in Figure 1. The initial identification of HLA-DR4 led to a flurry of activity defining the molecular basis of this association in the 1980s—the so-called “shared epitope”—an association that is still not fully understood mechanistically, but that appears to be primarily related to the production of an anti-CCP antibody response.3,4 It required another 20 years before evidence of genetic associations outside of the major histocompatibility complex (MHC) were convincingly demonstrated, with the discovery of associations with PTPN22 in Caucasian populations5 and PADI4 in Asian RA populations.6 It is only in the last few years that statistically well-powered case control studies have been performed using DNA chip technologies, available at a reasonable cost, to carry out a genome wide association scan (GWAS). The results of this have been remarkable, with over 35 genetic regions now definitively associated with RA.7,11 It is important to emphasize that the recent GWAS experiments are primarily designed to detect genetic variation that is relatively common, e.g., genetic variants that have frequencies of 5% or more in the general population. The genetic variants used for these mapping studies typically take the form of single nucleotide polymorphisms (SNPs). SNPs are particularly useful because each individual in the population is uniquely characterized by a set of at least
3 million common SNPs distributed across the genome; a set of around 500,000 SNP markers is enough to get a fairly accurate picture of the common variation in a person’s genome. Because of the large number of tests being done, many thousands of cases and controls are required to carry out a GWAS. In addition, because SNP markers can vary dramatically among different ethnic groups, and even among people with different European ancestry, it is critical to make sure that the geographic and ethnic origin of case and controls samples are matched. Statistical methods have been developed now to make this matching possible, using the genetic data alone, and this has dramatically improved our confidence in the results of these studies. A recent meta-analysis of over 5000 RA cases and 20,000 controls provides the best current picture of the patterns of association of RA, with common variants in the human genome in the Caucasian population.11

A Long Way Still to Go

While Figure 1 captures the top regions of association that are statistically robust, there are several things to keep in mind. First, we generally name these regions using the known genes that are in the region of interest. Sometimes there is only one gene nearby that clearly has an immune function; in this case, the assumption that the finding is functionally related to this gene is reasonable. On the other hand, there may be more than one gene in the region or the genetic association may be due to some other genetic feature, such as a micro-RNA or a regulatory element that is acting at a distance. The truth is that the actual causative variation is not known for almost all of these new associations. The only really clear exception to this is PTPN22, where a combination of dense genetic mapping and convincing functional data produce the amino acid change from arginine to tryptophan at amino acid 620, almost certainly a direct cause of disease risk. Even for PTPN22, the way in which this amino acid change actually leads to autoimmunity is still unclear, as discussed below. For each of these associations, there is an enormous amount of work required to identify the likely causative genetic changes in the region, establish some functional effect for these changes, and then integrate them into a scheme of pathogenesis.

In addition to these issues, it is also clear that we have not yet explained most of the genetic risk for RA. Even including the MHC, it is likely that more that 50% of the genetic risk for RA remains to be identified—the so called “missing heritability.”12 One possibility is that multiple rare variants meaning variants with allele frequencies much less than 5%, or even less than 1% may contribute to disease risk in a significant subset of the RA population. A seminal example of rare variant associations for a common phenotype is seen in the risk for low high-density lipoprotein (HDL) levels,13 and rare genetic variation also has been implicated in risk for both lupus14 and type 1 diabetes.15 More recently, Shiv Pillai and colleagues have discovered that rare variants in sialic acid acetylesterase (SIAE) are strongly associated with risk for RA.16 SIAE is an enzyme involved in the negative regulation of B cells and is required for immune tolerance in mice.17 Multiple rare amino acid changes in SIAE have been identified, many of which destroy enzymatic activity or transport of this molecule. In the aggregate, functionally deleterious variants in SIAE confer a relative risk of ~8 for RA. Even though these changes are found in only about 2% of RA patients, they clearly have a dramatic effect on disease risk when they are present. It seems likely that rare variants in other genes that are involved in immune tolerance will be forthcoming.

The example of rare variants that carry relatively high risk leads to a concept of disease genetics, in which a substantial proportion of individuals may have rather “private” genetic reasons for disease susceptibility, analogous to Leo Tolstoy’s famous beginning to Anna Karenina that “Happy families are all alike; each unhappy family is unhappy in its own way.” The discovery of additional rare genetic changes like those in SIAE will require deep re-sequencing of target genes, whole exomes or the entire genome, or both, from multiple individuals. At the moment, the cost of doing this on a genome-wide scale is prohibitive, but costs are likely to come down very quickly. Certainly, within the next 3 to 5 years, we will have a better sense of how much of the miss-
ing heritability for RA is due to rare variation in the human genome.

A recent deep analysis of GWAS data for RA also has revealed that there is a large component of very weak risk contained in the genomes of RA patient populations.10 This could be due to the widespread presence of many different rare risk variants, or it may indicate a background of very low risk from multiple common variants. This raises the question of whether particular combinations of variants, common or rare, are especially potent risk factors. So far, very few examples of this kind of “epistatic” interaction have emerged from the GWAS data on autoimmunity, and a comprehensive analysis is extremely costly in terms of statistical correction for multiple testing. These issues can only be adequately addressed when tens of thousands of samples are available for analysis. A current effort to do this is in progress by an international consortium of investigators for many autoimmune disorders, including RA. The availability of a low cost “ImmunoChip” to type ~200,000 SNPs for less than $50 a person is making this study possible.

Linking Genotype to Phenotype

Despite incomplete information on the genes involved in RA, we are still in the fortunate position of having many “definite” genetic associations to investigate in terms of function and disease pathogenesis, as shown in Figure 1. It has become obvious that gene action cannot be properly understood without examining a relevant phenotype. This implies a need for more precise definition of the clinical disease phenotype, as well as research to investigate specific immunological or other biologic features that characterize individuals as subphenotypes, persons with disease or at risk for disease. Subphenotypes are phenotypes found in many subjects with disease but may also exist independently of disease. For example, the major human leukocyte antigen (HLA) associations with RA are only present in subjects who are positive for anti-CCP antibodies. This strongly suggests that loss of tolerance to particular citrullinated antigens is the relevant functional abnormality to examine with respect to HLA risk alleles in RA, not the disease per se. The presence of anti-CCP antibodies is a phenotype that is associated with RA, but can exist independent of RA, and HLA is one of the genetic variants that contribute to this phenotype. It is currently not clear whether HLA contributes to other phenotypic features of RA, independent of the effect on anti-CCP antibody production.

There are several examples of risk genes for RA that have been associated with specific immunological phenotypes. For example, PTPN22 is an intracellular phosphatase that is known to regulate the threshold for signaling through both T- and B-cell receptors.11 One functional effect of the PTPN22 risk allele (620W) is to raise the threshold for signaling through these receptors. Thus, at least some subsets of B and T cells are more difficult to activate in the presence of the 620W risk allele. At first glance, this seems paradoxical for an allele that increases the risk for autoimmunity and breaking of tolerance. However, it raises several hypotheses. One is that there may be a relatively diminished activation of regulatory T cells (Treg). Alternatively, it may be that reduced receptor signaling results in a relative deficiency of negative selection against self reactive clones, either in the thymus for T cells, or during B-cell development in the bone marrow. Recent data from Eric Meffre and coworkers suggest that alterations in early B-cell selection events may indeed lead to a more autoreactive antibody repertoire, even in normal subjects who carry PTPN22 620W (E. Meffre, unpublished data). Data on Treg function or T-cell selection in PTPN22 620W carriers has not yet been reported. It is only by studying specific immunological functions that one can develop an understanding of how a genetic variant such as PTPN22 620W may predispose to RA. A somewhat analogous example involves the expression of IL2 receptor (CD25), which has recently been shown to be regulated by allelic variants that are involved in both T1D and RA.20

It is likely that many of the genes involved in autoimmunity are actually associated with immune phenotypes that are fairly common in the population, such as the PTPN22 and CD25 associated phenotypes discussed above. We are still in the early stages of understanding the full range of normal variation in human immune response, and this has led to calls for the development of more organized approaches in defining the phenotypic features of the human immune system21 and their relation to genetic variation. In order to make these connections between genotype and phenotype, it will be important to apply new tools for immunophenotyping to populations of normal subjects, thereby avoiding the confounding effects of active disease and drug therapy. Collections of genetically well-characterized normal subjects for such studies are being established and will facilitate targeted functional studies of the many new genes that have been and will continue to be identified for RA.22 This will be essential in order to integrate the new genetic discoveries into a comprehensive understanding of disease pathogenesis, with attendant benefits for diagnosis and new, more targeted, therapeutics.

Disclosure Statement

The author has no financial or proprietary interest in the subject matter or materials discussed, including, but not limited to, employment, consultancies, stock ownership, honoraria, and paid expert testimony.

References


