Seminars in Immunology 23 (2011) 67–83

Contents lists available at ScienceDirect

Seminars in Immunology

journal homepage: www.elsevier.com/locate/ysmim

Review

Genetic predisposition to autoimmunity – What have we learned?

Ekta Rai, Edward K. Wakeland*

Department of Immunology, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd., Dallas, TX 75229, United States

ARTICLE INFO

Keywords:
Autoimmune diseases
GWAS
Association
Variations
Susceptibility

ABSTRACT

Rapid advances in genetic technologies have led to the identification of more than 85 loci that contribute to susceptibility to autoimmune diseases. These susceptibility genes are distributed throughout the innate and adaptive immune systems, indicating that dysregulations in both immune systems participate in the development of autoimmunity. A significant subset of these susceptibility genes are shared between multiple autoimmune diseases. However, the dysregulation of specific pathways, such as the pathogen recognition receptors of the innate immune system and the TNF supergene family, are significantly involved in some autoimmune diseases. Although these findings dramatically increase the details available concerning the nature of genetic predisposition to autoimmunity, a mechanistic understanding of the processes involved has not been achieved. Future studies must focus on correlating phenotypes with specific genotypes to improve our understanding of the immune processes that are dysregulated during the development of autoimmunity.

1. Genetics of autoimmunity

Autoimmune diseases (AID) arise due to a failure of the immune self-tolerance mechanism, where tolerance to 'self' antigens is lost by the immune system, resulting in abnormal destruction of self tissue. AID can be categorized into two broadly defined types of disorders: (1) systemic autoimmune disorders such as rheumatoid arthritis (RA) and systemic lupus erythematosus (SLE), in which the loss of immune tolerance is directed towards systemic antigens and disease manifestations can occur at a variety of different sites in the body; and (2) organ-specific autoimmune disorders in which autoimmunity is predominantly or exclusively directed towards tissue-specific elements. Classic examples of organ-specific autoimmune disorders would include type 1 diabetes (T1D), psoriasis (PS), and multiple sclerosis (MS), in which pathology is very focused on a specific cell or tissue. Other AID, such as inflammatory bowel disease (IBD) [which includes Crohn's disease (CrD) and ulcerative colitis (UC)], and celiac disease (CD) are also specific to a single organ, although some characteristics of the autoimmunity are more similar to systemic autoimmunity or possibly autoaggressive responses to normal bacterial flora or food components. A variety of pathogenic mechanisms are ultimately triggered during the progression of autoimmune disease, and dysregulations involving major cell signaling pathways and the inflammatory responses are consistent features in most AID [1,2].

Susceptibility to AID is complex and involves interactions between both environmental and genetic factors. Environmental factors such as diet, environmental toxins, and/or infections with pathogenic organisms are thought to play a key role in the initiation of AID in genetically susceptible individuals. Genetically susceptible individuals are thought to have genomes that contain an appropriate constellation of susceptibility alleles that modulate the immune system in a fashion leading to a predisposition to AID when exposed to the appropriate environmental trigger. The importance of genetic predisposition to AID has been roughly quantified by calculating a genetic risk ratio, termed λ_S. The λ_S ratio is calculated from the disease prevalence in first degree relatives of an affected proband in comparison to the disease prevalence in the general population. A λ_S risk ratio of more than 5 is generally interpreted as indicative of a significant genetic contribution to the pathogenesis of a disease [3]. For AID, the reported λ_S values often vary between independent studies, but consistently indicate a strong genetic element. The calculated λ_S values reported for the AID discussed in this review are: (1) RA, 2–17 [4]; (2) SLE, 8–29 [5]; (3) MS, 20–40 [6]; (4) UC, 6–9; (5) CrD, 15–35 [7]; and (6) TID, ~15 [8].

The identification of AID susceptibility loci has been pursued using three basic strategies: (1) linkage analyses in pools of affected families, (2) candidate gene association studies using collections of cases and controls; and (3) genome wide association studies (GWAS), generally performed by consortiums of investigators on collections of 1000s of patients and controls. Many early studies of

Abbreviations: AID, autoimmune diseases; GWAS, genome wide association studies; T1D, type 1diabetes; PS, psoriasis; MS, multiple sclerosis; IBD, inflammatory bowel disease; CrD, Crohn's disease; UC, ulcerative colitis; CD, celiac disease; SNP, single nucleotide polymorphism; NGS, next-generation sequencing; PAMPs, pathogen associated molecular patterns; CNV, copy number variations; LPS, lipopolysaccharides.

* Corresponding author. Tel.: +1 214 648 7330.
E-mail address: Edward.wakeland@utsouthwestern.edu (E.K. Wakeland).

1044-5323/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.smim.2011.01.015

AID susceptibility genetics were performed using linkage analysis, which exploits the co-segregation of chromosomal regions with disease in affected families to identify the locations of causative genes. These studies identified many loci, however, with the exception of the HLA region, most susceptibility loci were detected with low statistical significance and findings were often poorly replicated in independent studies. However, these early linkage studies did identify some important non-HLA loci, including NOD2 (nucleotide binding and oligomerization domain) in Crohn's disease [9,10], Fc receptor complex in SLE, and STAT4 (signal transducer and activator of transcription) in RA and SLE [11]. A variety of candidate gene association studies were being performed in parallel with these linkage studies during the earliest phase of genetic analysis for human AID. Candidate gene analysis has remained a favorite strategy of many investigators and has the strength of exploiting prior knowledge of the basic biology of the disease to focus on a subset of genes whose function make them relevant to disease pathogenesis. The experimental approach is to statistically compare the allele and genotype frequencies of genetic markers in unrelated patient (cases) and normal (control) populations. Candidate gene association studies with single nucleotide polymorphisms (SNPs) were found to be superior to linkage analyses for the detection of common alleles that contributed to susceptibility to disease in a population [12]. However, these studies commonly suffered from statistical power issues due to insufficient sample collections of patients and controls and poorly defined population stratification issues. Similar to linkage studies, many candidate gene analyses replicated poorly between independent investigations [13].

During the last three years, genome wide association studies (GWAS), which are in essence an extension of the statistical methods used for candidate gene associations to the analysis of markers spaced throughout the entire genome, have emerged as the method of choice for genetic analyses of susceptibility to complex diseases. GWAS typically are conducted with SNP typing platforms that can genotype >350,000 SNPs (recent versions will now type >2,500,000 SNPs/assay) distributed throughout the genome in a single procedure. In general, GWAS are performed by large consortiums of investigators and involve the analysis of 1000s of patients and controls. Many GWAS utilize a collection of "tagging" SNPs, which are selected based on their capacity to capture information for 65–70% of the common variations throughout the human genome [4,14]. GWAS datasets are viewed as providing an unbiased, genome-wide search for risk loci that contribute to susceptibility to the disease of interest. These studies have several analytical advantages: (1) the dense datasets produced allow the identification of duplicate samples or samples from genetically related individuals in separate studies; (2) the datasets are sufficient to allow the application of strategies to correct for case–control stratification; and (3) the datasets are sufficient to impute SNP genotypes for untyped sites when performing genome-wide meta-analyses across multiple studies. Currently, >640 GWAS have been performed by the global genetics community, and these studies have led to the identification of ~3144 SNPs associations with ~393 diseases/traits. Among these, 55 published GWAS have identified 375 SNPs associations with 8 AID (see GWAS catalog http://www.genome.gov/26525384).

Although GWAS are a potent new strategy for the detection of novel and unbiased genetic associations with susceptibility to complex diseases, the interpretation of GWAS is subject to many important limitations. First, GWAS focus on common variants in the population and these SNPs generally do not occur in genomic regions anticipated to impact gene function. Theoretically, they are selected to represent the presence of a specific constellation or "haplotype" of SNP variations, thus allowing data obtained with a relatively small number of SNPs to reveal overall patterns of variation associated with disease. However, the extent and nature of any functional variations that may underlie differences in "SNP haplotypes" is poorly defined, and as a result, the functional and biological significance of the association of a specific SNP haplotype with disease susceptibility is unknown. Consequently, it is usually impossible to predict the nature of the functional changes that may underlie the association of a particular GWAS-defined candidate gene with disease. Secondly, current guidelines have established extremely stringent thresholds for defining a significant association with disease susceptibility (i.e. $p < 10^{-8}$) and, as a result, many genome scans provide a plethora of suggestive associations (SNPs having p value < 0.05 and $>10^{-8}$) while detecting few or no associations that reach genome-wide "significant" threshold levels [15]. As a result, different groups often discuss different sets of associated alleles for the same disease, which can lead to confusion in the literature and in the interpretation of genetic findings. Thirdly, the common SNPs that are assessed in GWAS represent only a subset of all the variability that is associated with complex diseases. That is, genetic predisposition to a common disease is likely to involve complex interactions between genetic variations exhibiting a broad spectrum of population frequencies, including a subset of "rare" mutations (i.e. <1%) that have a major effect on disease susceptibility in some individuals [13,16]. As a result, GWAS generally can only account for a fraction (i.e. 5–10%) of the total genetic heritability associated with a complex disease. Finally, the functional variation causing the association of a specific SNP with disease susceptibility may or may not be associated with a gene in close proximity to the tagging SNP. Although the causative lesion should be located within the LD block that is tagged by the associated SNP, the functional variation could potentially impact the expression or functional properties of a gene or genes that are more distal, or possibly interacting epistatically with the associated allele to cause the disease [4]. Thus, the precise location and identity of the specific genes mediating the functional variations associated with a specific SNP association can be difficult to predict.

The role of common variants versus rare variants in mediating susceptibility to common diseases has recently re-emerged as a controversial topic [17,18]. The controversy arises predominantly from the fact that the sum of the risk attributable to all of the loci identified in GWAS can only account for a small fraction of the genetic heritability exhibited by many common diseases, thus opening the door to the possibility that rare variants with strong functional effects actually contribute significantly to the overall heritability of common diseases. In this regard, rare variants are not assayed in a GWAS and several studies using sequencing techniques have identified rare mutations or copy number variations that contribute significantly to the pathogenesis of autoimmune diseases [19–25]. Recently developed high-throughput sequencing technologies make it feasible to initiate a comprehensive analysis of the impact of such rare variants on common diseases. The application of next generation sequencing to the analysis of orphan Mendelian genetic diseases has demonstrated the efficacy of this technology for identifying previously unknown rare defective alleles that cause human diseases [26–28]. A comprehensive cataloguing of the bulk of genetic sequence variations occurring within human populations is currently underway and it is reasonable to predict that a detailed database of human genome variability is in the process of being assembled. The most striking outcome of the initial phase of this massive undertaking is the extensive diversity detected within modern human populations. Thus, identifying the causative lesions that underlie both common and rare alleles in mediating susceptibility to AID may be more complex than originally anticipated. Nonetheless, it is clear that both types of alleles are involved.

This review will focus on the nature of the susceptibility loci that underlie the genetic predisposition associated with eight common autoimmune diseases (IBD: CrD, UC; T1D; SLE; RA; PS; MS and CD).