# Fishing for Genes in Autoimmunity

Miriam Regev MD PhD and Elon Pras MD

Gertner Institute of Human Genetics, Sheba Medical Center, Tel Hashomer affiliated with Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel

**ABSTRACT:** Autoimmune diseases are classic examples of multifactorial disorders in which a large number of genes interact with environmental factors to form the final phenotype. Identification of the genes involved in these diseases is a daunting challenge. Initially the search involved the candidate approach where polymorphisms in suspected genes were tested for association in large cohorts of patients and controls. Today, the most widely used method is genome-wide association studies (GWAS), a method based on screening large panels of patients and controls with hundreds of thousands of single nucleotide polymorphisms (SNPs), using microarray-based technology. Unique families in which autoimmune diseases are caused by single genes are another alternative. The identification of candidate genes is often followed by studies that provide biologic plausibility for the findings. The widely expanding list of genes involved in autoimmune conditions show that the same genes frequently underlie the pathogenesis of different autoimmune diseases. Despite all available resources, the main void of heritability in autoimmune conditions is yet to be discovered. Identification of these genes will help define new biological pathways and identify novel targets for the development of new therapeutic drugs.

*IMAJ* 2016; 18: 209–211

**KEY WORDS:** autoimmunity, genome-wide association studies (GWAS), exome sequencing, genetic advantage, HLA

Autoimmune diseases are the prototype of multifactorial inheritance, also called complex inheritance. In contrast to single gene inheritance, this form of inheritance is governed by an additive effect of many genes in combination with environmental factors, both of which may vary from person to person for a given disease. The heterogeneity of these diseases is so large that sometimes single gene disorders cause phenotypes almost indistinguishable from these diseases [1]. In this review we shall focus on the main methods used for the identification of genes involved in the pathogenesis of these disorders.

Many individuals carry genetic variants predisposing to autoimmune diseases but never develop such diseases; others are exposed to associated environmental factors but do not develop disease due to the lack of the genetic components. How and why did genes that predispose for autoimmunity evolve, and why do some individuals who carry these variants develop disease while others do not? It is feasible that millions of years ago subtle changes occurred in genes related to our immune system. Some of these changes conferred increased resistance to infections and therefore were associated with a genetic advantage which over time increased their frequency in the population [2]. Concomitantly however, these changes in certain situations caused a "hyperimmune" state and predisposition to autoimmune diseases. Since most autoimmune disorders appear later in life, these changes probably did not interfere with the genetic advantage conferred by the increased resistance to infections. An example for this hypothesis was recently proffered by Dunstan et al. [3], who showed that a specific subtype of HLA DRB1 confers protection from enteric fever. The same HLA subtype has been associated with increased risk for the development of Crohn's disease and rheumatoid arthritis (RA) [4,5].

Several lines of evidence point to the involvement of genetic factors in autoimmune diseases: first, for many of the autoimmune disorders relatives of the index case are at increased risk for developing the same disorder; for example, the risk of sibs or offspring of an individual with multiple sclerosis (MS) to develop the disease is 2%–5%, while the risk in the general population is around 0.1%, i.e., 20 times higher [6]. Second, the risk in monozygotic twins is higher compared to dizygotic twins: for example, if one of monozygotic twins develops systemic lupus erythematosus (SLE) the risk of his twin brother to develop the disease is 24% compared to 2% for dizygotic twins [7,8]. Adoption studies using twins who have been separated at birth, as well as population migration studies, help us distinguish genetic from environmental factors. It is possible to estimate what proportion of the etiology can be related to genetic factors as compared to environmental factors, a term referred to as heritability ($h^2$). There are a number of methods to calculate $h^2$, the most common one is to use variance information derived from monozygotic and dizygotic twins with the formula:

$$h^2 = \frac{\text{variance in dizygotic twins} - \text{variance in monozygotic twins}}{\text{variance in dizygotic twins}}$$

For a Mendelian disease $h^2$ is 1.0; for a disease that is completely environmental $h^2$ is zero. Heritability for autoimmune diseases varies between the two values. For example, the estimated $h^2$ for ankylosing spondylitis (AS) is 0.7, and for RA 0.55; thus the genetic component in AS is larger than in RA.

Over the years a number of methods have evolved enabling the detection of genes involved in multifactorial diseases.

These include the candidate approach, linkage analysis, and sib pair analysis. However, the great breakthrough came with the development of genome-wide association studies (GWAS) and lately, whole exome and whole genome sequencing.

The main method in use for the detection of genes predisposing to autoimmune diseases is the GWAS [9]. In order to understand the theoretical basis behind this method it is essential to fully understand the basis of association studies in candidate genes. Candidate gene association studies are based on the assumption that a susceptibility allele for a complex disease would be over-represented in patients compared to controls, probably due to a founder effect, a common ancient ancestor. This over-representation is also called linkage disequilibrium (LD). Most often not only is the susceptibility allele in LD, but also polymorphisms in close proximity to it. This is the result of lack of recombination events between such polymorphisms and the susceptibility allele which are very close one to the other. In fact, recombination events occur at specific points along the chromosomes, thus dividing the genome into LD blocks. Polymorphisms that are located in the same block would be in LD one with the other and, therefore, a candidate gene does not necessarily have to be detected by the actual disease susceptibility polymorphism but could be found through polymorphisms in its vicinity. Recent technological advancements allow us to screen the whole genome for LD between a disease and such polymorphisms [10]. This is done by testing a large cohort of patients with a specific disease, as well as controls, using an extremely large set of hundreds of thousands of single nucleotide polymorphisms (SNPs) evenly dispatched throughout the genome. This daunting task is performed on microarray platforms which can simultaneously carry out an enormous number of reactions in a short time and at reasonable cost. The data obtained are subjected to rigorous statistical analysis and replicated several times on separate, independent cohorts in order to avoid false positive results. Identifying a candidate gene is followed by a phase of multiple functional experiments to understand the nature of the gene, how it acts, which pathways it involves, and what role it plays in the pathogenesis of the disease [11]. To date, hundreds of GWAS studies have been performed for a wide variety of autoimmune disorders and these have resulted in the identification of hundreds of genes [12,13].

The function of these genes is highly variable and includes chemotaxis, lymphocyte activation [14], innate defense, adaptive immunity regulation, inflammatory mediators, autophagy, barrier function, NF-kB activation or inhibition, solute carriers, antimicrobial peptides, phagocytosis, and many more. However, above all, these studies emphasize the striking role of the HLA system in the pathogenesis of these disorders. The HLA system is responsible for the presentation of foreign antigens to the body's immune components. An "over-efficient" system may result in the presentation of self antigens as foreign and thus provoke autoimmunity. GWAS studies have reaffirmed what we already know from association studies, but they emphasize the overwhelming importance of the HLA system to the development of autoimmunity in comparison to the rest of the genome.

A surprising candidate that has emerged through multiple GWAS studies and is considered second to the HLA system is the thyrosine phosphatase PTPN22. Initially reported in type I diabetes mellitus (DM) [15], PTPN22 has also been shown to be involved in RA, Addison's disease, alopecia areata, Hashimoto's disease, Grave's disease, systemic sclerosis and other autoimmune conditions [16,17]. Despite hundreds of studies the mechanism of action of PTPN22 has not been fully resolved. Knockout mice models for PTPN22 show enhanced T cell activation and an increase in antibody production; however, in humans the risk allele has a lower degree of T cell activation [18,19]. It was therefore suggested that reduced rather than elevated T cell triggering may be part of the phenotypic predisposition to autoimmunity. In any case PTPN22 is an excellent example of how genes discovered by GWAS studies are hypothesis-generating, in contrast to those proposed by the candidate gene approach which are hypothesis-driven.

Another thyrosine kinase, PTPN2 has been associated with Crohn's disease, DM, RA and Grave's disease [20]. Knockout mice for this gene exhibit a fatal inflammatory wasting syndrome and abnormalities in a number of cell types. PTPN2 has a negative regulatory role on ILR2 signaling in T cells, and studies have shown that Jak1 and Jak2 are among its substrates [21]. These findings led investigators to believe that variants in PTPN2 alter the thresholds for these genes along with other immune system-related signaling pathways.

Other studies have implicated the tumor necrosis factor (TNF) receptor and the NF-kB signaling pathways. In 2007 an association was reported for TNFAIP3 in RA, an association that was confirmed in additional studies and extended to other autoimmune diseases [22,23]. In mice this gene acts as a major negative regulator of TNF-induced NF-kB signaling pathway. *TRAF1* and *TNFRSF5* (CD40) are two additional genes that have been identified using GWAS, with multiple effects on the immune system mediated through the kB signaling pathways [24]. These are just a few examples from a long list of genes with a wide range of roles including tolerance, barrier function, innate immunity, adaptive immunity, lymphocyte activation, chemotaxis, and many more functions related to different aspects of the immune system.

Single gene diseases may present with clinical symptoms and signs that closely resemble multifactorial disease. Identifying the underlying genes in such cases is relatively simple using positional cloning approaches or exome sequencing [25]. Al-Mayouf et al. [26] recently described a family with a familial form of SLE inherited in an autosomal recessive mode. Positional cloning showed that the disease in these families was caused by mutations in the *DNASE1L3* gene, a gene responsible for the clearance of degraded DNA from cells. Van Eyck et al. [27] performed exome sequencing on a 16 year old girl suffering from SLE and a selective IGA deficiency and found a de novo gain of function mutation in the *IFIH1* gene. Using exome sequencing Chao et al. [28] identified rare variants in the *CYP27B1* gene associated with MS. Interestingly, recessive mutations in this gene result in rickets, thus involving vitamin D metabolism in the pathogenesis of MS.

Not entirely surprising, a large overlap between genes involved in different autoimmune diseases has been found; thus, diseases such as MS, RA, Crohn's, ulcerative colitis and psoriasis have common predisposition genes. Though hundreds of genes have been identified so far, most contribute an extremely small fraction of the overall heritability. In fact, for each of the autoimmune diseases examined to date using GWAS studies, the total heritability of the genes identified does not exceed 50%. For example, for Crohn's disease the genes found so far explain 20% of the disease heritability and for SLE about 16%. Despite the small fraction of heritability discovered, the genes identified until now enhance our understanding regarding the biological pathways involved in the pathogenesis of these diseases, pathways that may serve as therapeutic targets in the future [29-31].

**Correspondence**
**Dr. E. Pras**
Gertner Institute of Human Genetics, Sheba Medical Center, Tel Hashomer52621, Israel
**Fax:** (972-3) 530-2998
**email:** epras@post.tau.ac.il

### References

1. Gregersen PK, Olsson LM. Recent advances in the genetics of autoimmune disease. *Annu Rev Immunol* 2009; 27: 363-91.
2. Sfriso P, Ghirardello A, Botsios C, et al. Infections and autoimmunity: the multifaceted relationship [Review]. *J Leukoc Biol* 2010; 87: 385-95.
3. Dunstan SJ, Hue NT, Han B, et al. Variation at HLA-DRB1 is associated with resistance to enteric fever. *Nat Genet* 2014; 46: 1333-6.
4. Viatte S, Plant D, Han B. Association of HLA-DRB1 haplotypes with rheumatoid arthritis severity, mortality, and treatment response. *JAMA* 2015; 313: 1645-56.
5. Spurlock CF, Tossberg JT, Olsen NJ, Aune TM. Cutting edge: chronic NF-κB activation in CD4+T cells in rheumatoid arthritis is genetically determined by HLA risk alleles. *J Immunol* 2015; 195: 791-5.
6. Hafler DA, Compston A, Sawcer S, et al. Risk alleles for multiple sclerosis identified by a genomewide study. *N Engl J Med* 2007; 357: 851-62.
7. Sebastiani GD, Galeazzi M. Immunogenetic studies on systemic lupus erythematosus. *Lupus* 2009; 18: 878-83.
8. O'Hanlon TP, Li Z, Gan L, Gourley MF, Rider LG, Miller FW. Plasma proteomic profiles from disease-discordant monozygotic twins suggest that molecular pathways are shared in multiple systemic autoimmune diseases. *Arthritis Res Ther* 2011; 13: R181.
9. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2000; 8: 559-75.
10. Liu C, Cui W, Wang L, et al. Klotho gene polymorphisms are related to colorectal cancer susceptibility. *Int J Clin Exp Pathol* 2015; 8: 7446-9.
11. Sadeh M, Benjamin Glazer B, Landau Z. Association of the M315I variant in the transient receptor potential vanilloidreceptor-1(*TRPV1*) gene with type 1 diabetes in an Ashkenazi Jewish population. *IMAJ* 2013; 15: 545-8.
12. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007; 447: 661-78.
13. Seldin MF, Price AL. Application of ancestry informative markers to association studies in European Americans. *PLoS Genet* 2008; 4: e5.
14. Paran D, Naparstek Y. Is B cell-targeted therapy effective in systemic lupus erythematosus? *IMAJ* 2015; 17: 98-103.
15. Smyth D, Cooper JD, Collins JE, et al. Replication of an association between the lymphoid tyrosine phosphatase locus (LYP/PTPN22) with type 1 diabetes, and evidence for its role as a general autoimmunity locus. *Diabetes* 2004; 53: 3020-3.
16. Criswell LA, Pfeiffer KA, Lum RF, 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: 561-71.
17. Vandiedonck C, Capdevielle C, Giraud M, et al. Association of the PTPN22*R620W polymorphism with autoimmune myasthenia gravis. *Ann Neurol* 2006; 59: 404-7.
18. Salmond RJ, Brownlie RJ, Zamoyska R. Multifunctional roles of the autoimmune disease-associated tyrosine phosphatase PTPN22 in regulating T cell homeostasis. *Cell Cycle* 2015; 14: 705-11.
19. Zheng J, Petersen F, Yu X. The role of PTPN22 in autoimmunity: learning from mice. *Autoimmun Rev* 2014; 3: 266-71.
20. Barrett JC, Hansoul S, Nicolae DL, et al. Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat Genet* 2008; 40: 955-62.
21. Simoncic PD, Lee-Loy A, Barber DL, Tremblay ML, McGlade CJ. The T cell protein tyrosine phosphatase is a negative regulator of janus family kinases 1 and 3. *Curr Biol* 2002; 12: 446-53.
22. Plenge RM, Cotsapas C, Davies L, et al. Two independent alleles at 6q23 associated with risk of rheumatoid arthritis. *Nat Genet* 2007; 39: 1477-82.
23. Musone SL, Taylor KE, Lu TT, et al. Multiple polymorphisms in the TNFAIP3 region are independently associated with systemic lupus erythematosus. *Nat Genet* 2008; 40: 1062-4.
24. Dunn IF, Sannikova TY, Geha RS, Tsitsikov EN. Identification and characterization of two CD40-inducible enhancers in the mouse TRAF1 gene locus. *Mol Immunol* 2000; 37: 961-73.
25. Johar AS, Anaya JM, Andrews D, Candidate gene discovery in autoimmunity by using extreme phenotypes, next generation sequencing and whole exome capture. *Autoimmun Rev* 2015; 3: 204-9.
26. Al-Mayouf SM, Sunker A, Abdwani R, Loss-of-function variant in DNASE1L3 causes a familial form of systemic lupus erythematosus. *Nat Genet* 2011; 43: 1186-8.
27. Van Eyck L, De Somer L, Pombal D, et al. Brief Report: IFIH1 mutation causes systemic lupus erythematosus with selective IgA deficiency. *Arthritis Rheum* 2015; 67: 1592-7.
28. Chao MJ, Barnardo MC, Lincoln MR, et al. HLA class I alleles tag HLA-DRB1*1501 haplotypes for differential risk in multiple sclerosis susceptibility. *Proc Natl Acad Sci USA* 2008; 105: 13069-74.
29. Cho JH, Feldman M. Heterogeneity of autoimmune diseases: pathophysiologic insights from genetics and implications for new therapies [Review]. *Nat Med* 2015; 21: 730-8.
30. Lazzaroni MG, Nalli C, Tincani A. What's new in autoimmunity: new antibodies, new therapies, new diseases. *IMAJ* 2015; 17: 71-3.
31. Relle M, Weinmann-Menke J, Scorletti E, Cavagna L, Schwarting A. Genetics and novel aspects of therapies in systemic lupus erythematosus. *Autoimmun Rev* 2015; 11: 1005-18.

**"My entire soul is a cry, and all my work is a commentary on that cry"**

Nikos Kazantzakis (1883-1957), Greek writer, whose most famous novel is *Zorba the Greek*