Abstract
Rheumatoid arthritis (RA) is a heterogeneous and complex disease where both genetic and environmental risk factors are involved. Reviewing the increased familial incidence and increased disease susceptibility demonstrated in twin studies, RA clearly has a genetic component. More evidence for a strong inherited component in RA is supported by association studies of RA with variants in the human leukocyte antigen (HLA) region, particularly the HLA-DRB1 gene in a wide range of populations. After the discovery of HLA associations, traditional linkage and association studies have identified new RA risk loci. In the last few years, the demonstration of genetic associations through the use of genome-wide association studies (GWAS) with a technical capacity to genotype hundreds of thousands of single nucleotide polymorphisms (SNPs) in a cost-effective manner proceeded to uncover new RA susceptibility genes. We review herein recent progress made in the field of RA genetics.

Key words: Rheumatoid arthritis, genetic susceptibility, HLA, linkage analysis, GWAS

Received: 17.06.2009 Accepted: 25.10.2009

Introduction
Rheumatoid arthritis (RA) is a systemic, inflammatory, autoimmune disorder which is associated with significant disability and early mortality; it affects approximately 1% of the population worldwide and is therefore one of the most common autoimmune diseases. RA is considered to be a heterogeneous and complex disease where both genetic and environmental risk factors contribute to the phenotype in multiple different combinations (1). A “complex disease” is a definition used to refer to a phenotype with a genetic etiology that is composed of a multitude of susceptibility genes, each contributing only a small magnitude of the overall risk for the disorder whereas geneticists refer to a “phenotype” as the observed trait (physical, behavioral, biochemical) of the organism under study (2). The polygenic etiology, in common with other complex diseases, presents several challenges to identify genes that confer risk for RA. “Genetic heterogeneity” is one of these challenges, meaning that different genes can result in the same phenotype. Further, each susceptibility gene is likely to have low penetrance, thus not all carriers will develop the disorder. Lastly, specific environmental influences are also more likely to be important risk factors for complex disorders than for simple Mendelian diseases, and gene-environment interactions constitute an important aspect to be considered in the etiology of such disorders (2-5).

Although the exact causes of RA remain largely unknown for the reasons mentioned above, there is strong evidence and it is generally accepted that RA has a genetic component. Twin studies of RA suggest that genetic influences contribute substantially to its etiology, with heritability estimates ranging from 50% to 60% (6).
The risk of disease in siblings of RA patients compared with that of the general population ($\lambda_s$) is estimated as 2-17 and the concordance rates for monozygotic twins versus dizygotic twins range between 15% and 3.6% respectively (7, 8).

More evidence for a strong inherited component in RA is supported by data obtained in association studies of RA with variants in the HLA region, particularly the HLA-DRB1 gene in a wide range of populations (9). The contribution of HLA-DRB1 gene to RA susceptibility was the first identified and remains as the most major determinant to genetic predisposition to RA (10). Studies evaluating the contribution of non-HLA susceptibility genes revealed inconsistent results with the exception of a few genes such as protein tyrosine phosphatase non-receptor 22 (PTPN22) gene in European populations and peptidylarginine deiminase 4 (PADI4) gene in Asian populations indicating the presence of genetic heterogeneity in RA. This genetic heterogeneity is considered to be responsible for the clinical heterogeneity as it is known that there is a wide spectrum of clinical manifestations, great variability in disease severity and progression, and different responses to a range of therapies among RA patients (1).

The identification of the genetic factors contributing to RA is of great importance since it would shed light on RA pathogenesis, diagnostic and prognostic markers, and new therapeutic approaches. Advances in molecular genetic technology particularly GWAS, along with traditional linkage and association studies have accelerated the identification of new RA risk loci. This article comprehensively reviews the latest findings on the genetics of RA, focusing mainly on the identification of novel RA susceptibility genes.

**HLA-DRB1 gene**

HLA-DRB1 gene encodes major histocompatibility complex (MHC) class-II β-chain molecules that take role in the presentation of antigen to CD4+ T helper cells and is known as one of the most polymorphic gene in the human genome (11). Genetic association of RA with a particular set of HLA-DRB1 alleles (DRB1*0401, DRB1*0404, DRB1*0405, DRB1*0408, DRB1*0101, DRB1*0102, DRB1*1001, DRB1*1401) have been consistently shown in a wide range of populations (9). All these alleles encode a conserved amino acid sequence (QKRAA, GRRAA, RRRAA) within the third hypervariable region of the DRB1 molecule which is referred to as the shared epitope (SE). The frequency of RA associated SE alleles differ between ethnic groups; in populations of European ancestry HLA-DRB1*0401 and *0404 are the most common alleles in RA patients, while *0405 is prominent in East Asian populations (12, 13). It is also demonstrated that these SE alleles show a dose effect with homozygotic individuals carrying increased risk over heterozygotes and some compound heterozygotes reveal a greater risk indeed, for example, the heterozygous combination of DRB1*0401/*0404 is identified to be strongly associated with early onset and a more severe form of disease than homozygosity for either allele (13, 14).

In the recent years, a class-II MHC β-chain amino-acid sequence (DERAA) has been identified that is protective against both RA susceptibility and severity in a northern European population (15). Interestingly, it has also been shown that the presence of this protective DRB1 allele in a woman may provide protection to her children even the allele itself is not inherited to them: a phenomenon named non-inherited maternal allele (NIMA) protection (16).

In spite of the significant role of the HLA region to RA genetic predisposition, it is estimated that, in total, the HLA region only contributes approximately 30% of the genetic component for RA, therefore it is believed that other non-HLA genes may confer a relevant role in RA susceptibility and there has been intense search for non-HLA disease associations recently (10).

**Non-HLA Genetic Association**

The PTPN22 gene, located on chromosome 1p13, encodes a lymphoid-specific phosphatase (Lyp) protein which is a negative regulator of T cell activation. A missense (non-synonymous) CT SNP at base position 1856 of this gene results in an arginine (R) to tryptophane (W) substitution at residue 620 of the protein product. This “R620W” substitution causes autoantigen-specific T cells to escape clonal deletion leading to a tendency for increased reactivity of the immune system and consequently increased predisposition to autoimmunity (17-19). The initial report of association of PTPN22 C1858T (R620W) allele with an autoimmune disease came from a study by Bottini et al. in March 2004, defining an association with type 1 diabetes (20). In the same year, a US group identified association of the same variant with RA and since then, numerous confirmations of the RA association have been reported in populations of European descent (21-27). Once again, the association is not seen in all ethnic groups and the R620W variant is rarely found in Asian populations, and therefore contributes to RA in populations of European descent in particular (28).

PADI4 gene encodes an enzyme that is responsible for post-translational conversion of arginine residues to citrulline. It is known that autoantibodies to citrullinated proteins are specific to RA and are associated with severity of the disease, thus, PADI4 which is involved in the citrullination pathway is thought to play a critical role in RA pathogenesis (29). Variants in the PADI4 gene provide convincing results for association among individuals of Asian ancestry but no evidence among individuals of European ancestry suggesting that there might be significant population differences for this gene (29-35).
Recent studies investigating the association of RA with signal transducer and activator of transcription-4 (STAT-4) gene polymorphisms have confirming results in North American, Swedish, Korean and Japanese populations providing the first example of a non-HLA gene that is associated across two major ethnic groups (36-38). STAT4 is a transcription factor that transmits cytokine signals such as interleukin-12, interleukin-23 and type I interferone in T cells and monocytes leading to the production of interferon-γ and interleukin-17; T-helper type 1 and T-helper type 17 differentiation; and monocyte activation. Although there is strong and replicated evidence for this association, it is not yet known how polymorphisms in this gene cause immune dysregulation and autoimmunity (36, 39, 40).

A recent revolution in human disease genetics has been the development of GWAS which allow identification of genetic risk factors without prior hypothesis of the pathogenetic mechanisms (41). With completion of the Human Genome Project and the documentation of patterns of genome wide-variation and linkage disequilibrium by the HapMap Project it is now technically possible and financially feasible to genotype hundreds of thousands of SNPs simultaneously amongst large sample sets (41). In RA, this approach has identified a region of thousands of SNPs simultaneously amongst large sample sets (41). In RA, this approach has identified a region of association on the long arm of chromosome 9 (9q33-34), that habours genes encoding both tumour-necrosis-factor-associated factor-1 (TRAF-1) and C5 complement and a region at 6q23 in proximity to the TNF-α-induced protein 3 (TNFAIP3) gene (42-44).

Although there is substantial evidence that TRAF1 and C5 may both be effective in the proliferation of inflammatory responses, functional studies will be required to further investigate the TRAF-1/C5 association (45).

The SNP with a strong evidence of association that lies in an intragenic region of 6q23 in proximity with TNFAIP3 gene was reported to be associated with RA in an independent study in a US population further underlying the possibility of a disease causing variant. It is estimated that the variant controls the function of the adjacent TNFAIP3 gene however functional studies will be necessary to identify how it affects disease progression (43).

Conclusion

Great effort has been made over the past years to understand the genetic basis of the susceptibility to RA leading to a new picture of disease pathogenesis by highlighting the heterogeneity among individuals and populations. It is expected that within the next few years, the genetic etiology of RA will be substantially resolved by the repeated use of GWAS for populations worldwide and the meta-analysis of these studies. The translation of genetic information to support clinical management in disease diagnosis, treatment and prediction is estimated to become a reality in the near future as a consequence of the recent exciting genetic discoveries.

Conflict of Interest

The authors of this manuscript confirm that there are no competing interests or financial disclosures to declare.

References


