Review

Recent advances in the genetics of RA susceptibility

J. Bowes and A. Barton

RA is a common autoimmune disease with a complex aetiology in which genetic and environmental factors contribute to disease. The genetic component of RA is largely undefined and, up until very recently, there were only two reproducible associations. The strongest of these associations is of genes within the HLA region, particularly the HLA-DRB1 gene. A second, more modest, association identified has been of the protein tyrosine phosphatase non-receptor 22 (PTPN22) gene. Advances in genotyping technology have facilitated the application of whole genome association approaches to identify disease causal variants. This, coupled with the availability of large case and control collections has enabled the identification of low-to-moderate risk loci. These newer study designs combined with traditional linkage and association studies have accelerated the identification of novel risk loci. The past few months alone have witnessed the identification of three new RA risk loci. In this review, we aim to give an update on recent progress in RA genetics, focusing mainly on the identification of novel loci.

Evidence to support genetic component to RA

There is strong evidence to support a significant genetic component to the susceptibility of RA, although this contribution remains poorly defined. Evidence from twin studies demonstrates excess disease concordance between monozygotic (15%) when compared with dizygotic (3.6%) twins [3]. From such studies, the heritability of RA, defined as the extent to which variation in liability to disease in a population can be explained by genetic variation, has been estimated at between 50% and 60% [4]. The increased risk of disease in siblings of patients with RA compared with that of the general population (3%) has been estimated to be between 2 and 17 [5]. These results suggest that genetic factors have a substantial influence on disease susceptibility and account for a major proportion of disease liability within populations.

Indeed, compelling evidence to support the genetic basis for disease is the consistent and reproducible association of RA with variants within the HLA-DRB1 and PTPN22 genes in populations of Northern European descent, and the associations with peptidylarginine deiminase 4 (PADI4) genetic variants in Asian populations. For the purposes of this review, discussion will be restricted to genetic susceptibility variants identified in populations of Northern European descent.

HLA-DRB1 gene

Association with variation in the HLA region on chromosome 6 (6p21.3) was established in the late seventies [6, 7] and is the only region that has been consistently shown to be both linked and associated with RA across all populations. This region extends over 3.6 Mb and is divided into three sub-regions (classes I, II and III). It is a highly gene dense area containing ~220 genes and many of these genes are thought to have immunoregulatory functions [8]. It was subsequently demonstrated that RA is associated with specific alleles of the class II gene, HLA-DRB1, that encode a conserved sequence of amino acids in the third hypervariable region (HVR3) of the class II DRβ1 chain. These associated alleles are collectively referred to as the shared epitope (SE) [9]. The residues occur in the antigen-binding site and may influence the efficiency of antigen presentation. The frequency of RA-associated SE alleles has been found to vary considerably depending on ethnic group. For example, the alleles *0401 and *0404 are predominantly associated with RA in Caucasian populations, *0405 allele in Asian populations and *0101 in Israeli Jews. A further layer of complexity is that combinations of SE alleles can carry a greater risk than homozygosity for those alleles. For example, the heterozygous combination of DRB1*0401/*0404 is strongly associated with early onset and a more severe form of disease than homozygosity for either allele [8].

It remains unclear whether all the genetic contribution to RA susceptibility arising from the HLA locus is accounted for by the HLA-DRB1-SE association or whether other susceptibility loci also reside within the region. However, in total, the HLA region only contributes ~30–50% of the genetic component for RA. Recently, considerable progress has been made in identifying non-MHC-related susceptibility loci.

PTPN22 gene

The minor allele of a non-synonymous single nucleotide polymorphism (SNP) (rs2476601, 1858C-T, R620W) in the protein tyrosine phosphatase non-receptor 22 (PTPN22) gene, located on chromosome 1p13, has been found to be associated with multiple autoimmune diseases. The PTPN22 gene is a compelling biological candidate for involvement in autoimmune diseases. It has been found to be expressed in a range of immunologically relevant tissues [10] and encodes the intracellular protein lymphoid tyrosine phosphatase (LYP). Protein tyrosine phosphatases (PTPs) play an essential role in signal transduction and are integral in the T-cell antigen receptor (TCR) signalling pathway. LYP itself is known to be a powerful inhibitor of T-cell activation [11].

The initial report of association of the minor *T allele with an autoimmune disease came from a candidate-gene study of two independent case–control cohorts of patients with type 1 diabetes [12], and this association with type 1 diabetes has since been replicated in a number of studies [13–16]. However, swiftly afterwards, a US group reported association of the same variant with RA [10]. This was followed by numerous other studies
in which the association with RA has been unequivocally confirmed in populations of European descent, including UK [17], Finnish [18], Swedish [19], German [20], Dutch [21], Spanish [22] and Canadian [23] populations. Interestingly, a study in a Japanese population could not test for association as the causal variant was found to have a very low minor allele frequency [24]. Subsequent studies have extended these observations and have shown that the same variant is also associated with susceptibility to SLE [25], juvenile idiopathic arthritis (JIA) [17], Graves’ disease [26] and generalized vitiligo [27]. However, no associations with multiple sclerosis [28], Crohn’s disease [23], psoriasis [29] or psoriatic arthritis [17] have been detected, suggesting that there may be differences in the aetiology of these subsets of autoimmune diseases [30].

This variant confers the second largest genetic risk to the development of RA (the highest genetic risk being conferred by the HLA-DRB1 gene), with an effect size of \( n = 1.8 \). The associated non-synonymous SNP results in an amino acid substitution of arginine (Arg620) for tryptophan (Trp620) in one of the several proline-rich motifs within the non-catalytic C-terminal end of LYP. This motif enables the physical binding of LYP with the Src homology 3 (SH3) domain of the negative regulatory kinase, C-terminal Src kinase (Csk). The resulting LYP-Csk complex is known to synergistically inhibit the TCR signalling pathway by down-regulating the tyrosine kinase, lymphocyte-specific protein tyrosine kinase (LCK), which is involved in early stages of T-cell activation [31]. Experimental data demonstrate that Trp620 disrupts this motif and implies that the disease predisposing effects are exerted via a reduced binding affinity towards the SH3 domain of Csk for the mutant protein [10, 12]. Although the consequences of Trp620 on the function of the TCR signalling pathway have not yet been fully elucidated, the evidence has led to the speculation that carriers of the Trp620 have a reduced capacity for the down-regulation of activated T-cells due to loss of cooperative inhibition by LYP-Csk complexes, and would therefore be prone to autoimmunity due to over-reactive T-cells following immune stimulation. Interestingly, data from two studies contradict this hypothesis and suggest that the Trp620 allele is gain-of-function mutation conferring enhanced inhibition of both T and B lymphocytes [32, 33]. The Trp620 allele results in increased catalytic activity of the enzyme with increased lymphocyte inhibition leading to an increased threshold level of stimulation required for T- and B-cell activation. The authors speculate that the increased inhibition may lead to thymocyte hyporesponsiveness and a failure to delete autoreactive T-cells during thymic selection.

**STAT4** genetic variation is associated with RA

A recent study investigating a selection of 13 candidate genes within a 2 Mb region under a peak of linkage on chromosome 2q identified association with an SNP mapping near the genes signal transducer and activator transcription (STAT) 4 and STAT4 in a North American cohort \( (n_{cases/controls} = 607/1309) [34]\). Further investigation revealed association with four SNPs in the third intron of the **STAT4** gene, the strongest association being with rs7574865 \( (P = 8.29 \times 10^{-5}; \text{Odds Ratio (OR)} 1.37; 95\% \text{ CI } 1.17, 1.60) \). This association was confirmed in a large replication cohort of North Americans \( (n_{cases/controls} = 1013/1326; P = 6.26 \times 10^{-4}); \text{OR } 1.28; 95\% \text{ CI } 1.11, 1.47) \), although, when tested in a Swedish cohort \( (n_{cases/controls} = 1529/881) \) the association and effect observed were much weaker \( (P = 0.02; \text{OR } 1.18; 95\% \text{ CI } 1.02, 1.36) \). This may be due to demographic heterogeneity between the two populations. The majority of the North American RA patients had long-standing erosive disease, whereas the Swedish cohort generally had more recent disease onset. Association with these four intronic SNPs has subsequently been replicated in a Korean population [35], providing the first example of a non-HLA gene that is associated across two major ethnic groups. Association with rs7574865 was also found in a meta-analysis of three SLE cohorts \( (n_{cases/controls} = 1039/1248; P = 1.78 \times 10^{-7}; \text{OR } 1.55; 95\% \text{ CI } 1.34, 1.79) \) in the original study, suggesting that this locus may harbour a risk allele pre-disposing to susceptibility for multiple autoimmune diseases with a shared aetiology in a similar manner to the PTPN22 gene.

**Whole genome association studies**

The study of human disease genetics has recently undergone a dramatic revolution due to the coming of age of genome-wide association studies (GWAS), the advent of which has been catalysed by a number of recent developments. First, advances in technology have enabled hundreds of thousands of SNPs to be economically genotyped in thousands of samples, thus allowing studies to achieve good coverage of common variation in the human genome. For example, the Affymetrix 500k GeneChip is estimated to capture 65% \( (\text{r}^2 \geq 0.8) \) of the variation represented in the HapMap Phase II data set [36]. Second, there has been a recognition that only large, well-powered studies will be able to detect the genetic effects expected in complex diseases. Hence, the establishment of large and well-characterized sample cohorts such as the 1958 birth cohort [37] represent valuable resources for researchers.

A ground-breaking GWAS was undertaken by The Wellcome Trust Case Control Consortium (WTCCC) within a British population [38]. This is a collaborative effort of over 50 research groups within the UK, enabling the analysis of 500,000 SNPs in 2000 case samples from seven complex human diseases, including RA and 3000 shared controls. This study has proven very successful with the identification of 24 independent regions achieving stringent significance thresholds \( (P < 5 \times 10^{-8}) \) across all diseases. The majority of these regions are suspected to be genuine susceptibility loci as they have been previously identified or have since been independently replicated. The success of this project is considered to represent a thorough validation of the GWAS approach. It has addressed many important methodological issues including quality control, study design and analysis considered essential for the successful application of GWAS. The results provide a valuable resource for researchers in the field, not only as a data repository but also as a practical lesson in the application of such studies.

**Association of the 6q23 region with RA**

The two strongest effects for RA observed from the WTCCC study of 1860 RA cases and 2930 population controls, corresponded to the two well-documented RA susceptibility genes, HLA-DRB1 and PTPN22. A further nine variants were found to be associated with RA at moderate significance levels \( (P = 5 \times 10^{-5} - 1 \times 10^{-7}) \), none of which map to previously recognized disease loci. Subsequently, these nine variants were genotyped in a large independent UK cohort of 5063 cases and 3849 controls [39]. Strong evidence of association was detected with an SNP in an intergenic region of 6q23 \( (rs6920220; P = 1.1 \times 10^{-8}; \text{OR } 1.23; 95\% \text{ CI } 1.15, 1.33) \) and the association was independent of HLA-DRB1 and PTPN22. The associated SNP is flanked by two genes, oligodendrocyte lineage transcription factor 3 (OLIG3) and TNF-α-induced protein 3 (TNFAIP3). Simultaneously, the same region was reported to be associated with RA in an independent study in a US population, providing unequivocal confirmation that the region harbours a disease-associated variant [40]. Fine mapping studies will be required to identify the causal variant and functional studies will be necessary to determine how it affects disease. It is possible that the variant regulates the function of the adjacent TNFAIP3 gene. The latter is an attractive candidate RA susceptibility gene because previous studies have shown that the TNFAIP3 gene acts as a negative regulator of the transcription factor NF-κB in response to TNF-α and toll-like receptor-1, but not IL-1β-induced activation [41, 42].
Furthermore, in a previous study, mice deficient for the TNFAIP3 protein developed multi-organ inflammation that included inflammation of joints [43].

**Association of the TRAF1/C5 locus with RA**

A further GWAS has been conducted in a combined cohort of anti-cyclical citrullinated peptide-positive RA patients and population controls from North America and Sweden, providing a final filtered data set of 297,086 SNPs genotyped in 1493 cases and 1831 controls [44]. As with the WTCCC study, the strongest associations were identified with the **HLA-DRB1** and **PTPN22** genes. One other SNP reached genome-wide significance levels highlighting a region on chromosome 9q33–34 that includes the genes **TRAFL** and **C5** (rs3761847; $P = 2.8 \times 10^{-8}$; OR 1.36; 95% CI 1.23, 1.50). Association was validated in two independent cohorts: in the North American validation samples, rs3761847 was replicated in 485 cases and 1282 control subjects ($P = 1 \times 10^{-7}$; OR 1.37; 95% CI 1.18, 1.58). However, in 568 RA cases and 516 controls from Sweden, no evidence for association was detected ($P = 0.31$; OR 1.11; 95% CI 0.93, 1.32). When the data from all subjects were combined ($n_{cases/controls} = 2575/3648$), a highly significant association with rs3761847 was observed ($P = 4 \times 10^{-12}$; OR 1.32; 95% CI 1.23, 1.42).

An independent candidate-gene study supports the association of this region with RA. In a multi-tiered case-control study involving the use of Dutch, Swedish and North American subjects, association with an SNP in the 5′ region of the **TRAFL** gene has been reported [45]. A highly significant association with rs10818488 when data from all four independent cohorts ($n_{cases/controls} = 2719/1999$) tested were combined ($P = 1.40 \times 10^{-8}$; OR 1.26; 95% CI 1.15, 1.37).

Interestingly, although the most associated **TRAFL/C5** SNPs from the two studies outlined were not genotyped directly in the WTCCC study, highly correlated SNPs (which can therefore act as proxies for the un-genotyped variants) were tested but showed no evidence for association in that data set. Further investigation in larger sample sizes will be necessary to confirm association of this locus with RA in the UK population.

**Summary**

Over the past 12 months, understanding of the genetic basis of the susceptibility to RA has increased dramatically with five susceptibility genes now confirmed in populations of Northern European descent. The major effect resides within the **HLA-DRB1** gene with the other genes making more modest contributions to susceptibility (Fig. 1).

The challenge now is to identify the remaining, probably smaller, genetic effects and explore how these variants interact with each other as well as environmental factors to induce the development of RA. It also provides an exciting opportunity to determine whether knowledge about these genes can help clinically in defining subgroups of patients, for example, with different disease trajectories and outcomes. The translation of genetic information to aid clinical management looks set to become a reality in the not too distant future.

**Rheumatology key messages**

- There are now five confirmed RA susceptibility genes
- **HLA-DRB1** confers the largest effect size and **PTPN22** the second largest
- In 2007, three further genes were identified.

**Acknowledgements**

We thank the Arthritis Research Campaign for their support.

**Funding:** This work was funded by the Arthritis Research Campaign (ARC grant reference no: 17552).

**Disclosure statement:** The authors have declared no conflicts of interest.

**References**

20. Pifer M, Kallenhauser S, Arnold S et al. Association of PTPN22 1858 single-nucleotide polymorphism with rheumatoid arthritis in a German cohort: higher
frequency of the risk allele in male compared to female patients. Arthritis Res Ther 2006;8:R75.


38 The 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.


