Available online at www.sciencedirect.com
ScienceDirect
Biomedicine & Pharmacotherapy 60 (2006) 656–662
BIOMEDICINE & PHARMACOTHERAPY
http://france.elsevier.com/direct/BIOPHA/

Dossier: Rheumatoid arthritis

Genetic basis of rheumatoid arthritis

G. Orozco, B. Rueda, J. Martin*

Instituto de Parasitología y Biomedicina "López-Neyra", Consejo Superior de Investigaciones Científicas (CSIC), Parque Tecnológico de Ciencias de la Salud, Avenida del Conocimiento s/n 18100 Armilla, Granada, Spain

Received 1 August 2006; accepted 21 September 2006
Available online 16 October 2006

Abstract

Rheumatoid arthritis (RA) is a clinically heterogeneous condition with a complex aetiology in which environmental and genetic factors are implicated. The contribution of human leukocyte antigen (HLA) genes, particularly the *HLA-DRB1* gene, to RA genetic predisposition was the first described, and remains as the best characterised single genetic risk factor contributing to RA. However, it has been estimated that only 30% of the genetic contribution to RA can be attributed to HLA genes and it is suggested that other non-HLA genes may play a relevant role in RA susceptibility. Linkage studies and association studies are the two main strategies used in the investigation of genetic factors contributing to complex genetic traits. In this work we review the progress made in the field of RA genetics, focusing mainly on the contribution of candidate gene association studies to the dissection of RA genetic risk factors.

Keywords: Rheumatoid arthritis; Genetic factors

1. Introduction

Rheumatoid arthritis (RA) is the most common chronic autoimmune disorder affecting approximately 1% of the population worldwide. It is characterised by the chronic inflammation and destruction of the synovial joints leading to progressive joint damage and disability. RA is considered a clinically heterogeneous condition with a wide spectrum of clinical manifestations, great variability in severity and disease progression, and different responses to a range of therapies. RA aetiology is complex, in common with other autoimmune disorders, and implies the interaction between environmental and genetic factors.

Evidence for a strong inherited component in RA is supported by data obtained in familial and twin studies, which suggest that a 60% of disease susceptibility is due to genetic factors [1]. RA shows a high concordance in monozygotic twins (12–30%) and is more prevalent in first-degree relatives (2–12%) with an estimated λs coefficient (defined as the ratio of disease prevalence among first-degree relatives to population prevalence) for the disease of 3–15 [2,3].

The contribution of human leukocyte antigen (HLA) genes to RA susceptibility was the first described, and remains as the best characterised single genetic risk factor contributing to RA. Different *HLA-DRB1* alleles (*DRB1*0401*, *DRB1*0404*, *DRB1*0405*, *DRB1*0408*, *DRB1*0101*, *DRB1*0102*, *DRB1*1001*, *DRB1*1401*) are associated with RA in a wide range of populations [4]. All these alleles are characterised by the presence of a conserved short amino acid sequence (QKRAA, GRRAA, RRRAA) defined as shared epitope and located in the third hypervariable region of the DRB1 molecule [5]. Although the contribution of HLA-DRB1 alleles to RA predisposition is clear, the shared epitope hypothesis fails to explain completely the contribution of the HLA region to pathology [6]. The HLA region is characterised by a high degree of linkage disequilibrium and it is hypothesised that in addition to *HLA-DRB1*, this region harbours other genes, which might either modify RA risk or be independently associated with the disease [7–9].

In spite of the relevance of the HLA region to RA genetic predisposition, it is estimated that the degree of familial risk due to the HLA genes is only about 30% [10]. On this basis it is suggested that other non-HLA genes may play a relevant role in RA susceptibility.

This genetic heterogeneity may account for the complexity of RA presentation and progression, and therefore the identifi-

* Corresponding author. Tel.: +34 958 18 1669; fax: +34 958 18 1632.
E-mail address: martin@ipb.csic.es (J. Martin).

0753-3322/$ - see front matter © 2006 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.biopha.2006.09.003

cation of the genetic factors contributing to RA is of great relevance since it would enable a better understanding of RA pathogenic mechanisms, the improvement of diagnostic and prognostic markers and the development of new therapeutic targets.

Two main strategies can be followed to identify genetic factors contributing to a complex trait such as RA: linkage studies and candidate gene association studies.

2. Genetic linkage studies in RA

Linkage analysis aims to identify genomic regions containing disease predisposing genes by observing related individuals. The widely used approach is to study sibling pairs diagnosed for the disease. It is expected that affected relatives will show an excess sharing of haplotypes identical by descendent in the region of a disease-causing variant. Thus, chromosomal regions exhibiting increased allele sharing are likely to contain susceptibility genes [11]. In a linkage study the genome can be screened either partially or completely (whole genome scan, WGS), using multiallelic markers (microsatellites) or biallelic markers (single nucleotide polymorphisms, SNPs).

In the investigation of RA genetic component, several genetic linkage studies have been conducted in different populations from Europe, USA and Japan [12–20]. The most consistent finding is the evidence of linkage with the chromosome 6p21 region where the HLA genes map. In addition, different genomic regions outside the HLA region have shown suggestive linkage (LOD>2.2, $P<0.001$) to RA susceptibility. However, only a few of them show significant linkage (LOD>3.6, $P<2\times10^{-5}$) or have been replicated in independent sets of samples (Table 1). Interestingly, these regions with replicated linkage are as well associated with other autoimmune disorders, suggesting that there may exist a common genetic background predisposing to autoimmunity [21–24].

Due to the relatively large size of the regions identified, it is necessary to carry out fine mapping using dense sets of markers followed by candidate gene analysis in order to identify the real causing variant which is responsible for the association with these regions. To date, only the candidate gene mapping within the 1p13 region, the *PTPN22* gene, has been identified. This gene has become the best characterised RA genetic factor besides the HLA-DRB1 gene, which will be discussed later.

The results obtained in these studies support the notion that there might be several loci implicated in RA genetics with effect sizes less than HLA. To detect loci with such a small effect, linkage studies conducted in a few hundred families are underpowered, and therefore more robust strategies are required. An efficient approach which permits a substantial increment in statistical power of linkage studies is the combination of data from different populations through the realisation of meta-analysis or reanalysis of genotype data. By means of this strategy it has been possible to identify additional susceptibility locus on chromosome 16p as one of the most significant after HLA [25,26].

Table 1
Genomic regions showing significant linkage in RA and replicated in independent populations

Population	Locus	Suggested candidate gene
UK, USA	6p21.3	HLA-DRB1, other MHC genes
UK, USA	18q21	RANK
UK, USA	1q43	Unknown
UK, USA	6q21	Unknown
UK, USA	1p13	PTPN22
Meta-analysis	16p	Unknown

Linkage studies can benefit from the large amount of information available about the variation across the human genome, specially the description and localisation of millions of SNPs. This fact together with the rapid development of high throughput genotyping technologies makes it possible to perform SNP-based WGS, in which a higher number of markers can be analysed generating significantly higher information that allows a more precise definition of locus showing linkage with the disease. At present, only one preliminary RA SNP-based WGS has been published which replicated the strong linkage with 6p21 region [19].

3. Candidate gene association studies in RA

In genetic association studies the main objective is to detect if a genetic marker is implicated in disease susceptibility. The study design is based in the selection of a well define set of affected individuals (cases) and a healthy control group from the same population, to compare the distribution of certain genetic markers between the two groups [27].

This approach has greater power than linkage studies to detect small size effects contributing to disease predisposition. However, a careful study design is required in order to avoid false positive associations and population stratification [28]. In case–control studies patients are compared with matched (age and sex) controls, ideally with the same ethnicity, but differences in the genetic background of the two groups can be difficult to detect leading to the introduction of variables unrelated to the disease and might cause spurious associations due to stratification [29]. Thus, to guarantee an adequate power of association studies, replication of findings in different unrelated populations and large population sizes are required. In addition, it is possible to design an association study based on nuclear families (one patient and their parents) avoiding population stratification [30].

In association studies, candidate genes can be selected on the basis of their implication in the disease pathogenic mechanisms (functional candidate genes) or considering their location within genomic regions, which were previously in linkage to disease (positional candidate genes). Several genetic variants exist in a candidate gene and the selection of a candidate polymorphism to be analysed in an association study is difficult. The most interesting are putative functional genetic variants that might cause an alteration of gene function [31]. Thus, acquire great relevance coding SNPs altering protein structure and function. In addition non-coding variants located in regulatory regions are of importance since they may alter promoter activity or RNA splicing and processing leading to differences in protein levels [31].