Genetics and genomics in rheumatoid arthritis (RA): An update

Ana Karen Rodríguez-Elías1,2, Karina Maldonado-Murillo2,3, Luis Fernando López-Mendoza2,3 and Julián Ramírez-Bello3*

1Experimental Biology Postgraduate Program, UAM-I; 2Biology Undergraduate Program, Facultad de Estudios Superiores Zaragoza, UNAM; 3Laboratory of Genomic Medicine, Research Unit, Hospital Juárez de México, SSA, Mexico City, Mexico

Abstract

Rheumatoid arthritis is a chronic inflammatory autoimmune disease that affects approximately 0.5-1% of the general population and leads to chronic synovial inflammation, destruction of cartilage and bone, and disability. The heritability of rheumatoid arthritis has been estimated to be about 60%, while the contribution of HLA to heritability has been estimated to be 11-37%. Other genes, such as PTPN22, STAT4, CTLA4, TRAF1, PADI4, IRF5, FCRL3, TNFIP3, TNF-α, miRNAs, CD28, CD40, TYK2, etc., have been associated with susceptibility, severity, activity, and treatment response of rheumatoid arthritis. The aim of this review is to describe the role of gene variants located in immune system genes associated with susceptibility to rheumatoid arthritis. (Gac Med Mex. 2016;152:194-203)

Corresponding author: Julián Ramírez-Bello, dr.julian.ramirez.hjm@gmail.com


Introduction

Rheumatoid arthritis (RA) represents the chronic inflammatory disease prototype; it is characterized by synovial membrane inflammation, cartilage destruction, bone erosion, joint deformity and functional disability of the affected individual1,2. It is well documented that when not opportunely treated, RA causes loss of job, decreases the quality of life and is associated with premature death due to cardiovascular disease2,3. Different cells of the innate and adaptive immune system show alterations in the expression of different genes that codify for proteins, such as cytokines, chemokines, receptors, adhesion molecules and genes that synthesize non-coding RNAs, specifically micro-RNAs (miRNA), which have differential expression in this disease4,5. Although RA etiology is not fully known, the interaction between different low penetrance genetic factors and several environmental factors, such as sex hormones and agents that trigger the immune response, such as viruses and bacteria, has been documented to influence on its pathogenesis6. Different evidences have shown that genetic alterations, mainly of the single nucleotide polymorphisms (SNPs) type, located in genes that produce proteins...
and non-coding RNAs (specifically miRNAs), and that regulate the innate and adaptive immune response, are the main genetic risk factor involved in RA. Candidate gene or genome-wide association studies (GWAS) have identified different risk loci associated with the RA etiology. Currently, in RA, about 100 genes associated with susceptibility, protection, severity, activity and treatment response have been described, including genes that codify for class II HLA and several non-HLA genes, such as STAT4, CTLA4, TRAF1, PADI4, FCRL3, TNFIP3 and TNF-α, as well as miRNAs, mainly miR-146a and miR-499. These genes importantly influence on RA pathogenesis; in this review, details will be given on these genetic associations with this autoimmune disease (AD).

Epidemiology

RA affects approximately 1% of the general population. Prevalence differences have been reported in industrialized countries: it can affect from 0.5 to up to 2% of the population, with an incidence of 12-200 cases for each 100,000 inhabitants. The female: male ratio is 2:3:1, and peak age of onset is between 30 and 55 years, although it can occur at any age. In Mexico, this entity affects 1.6% of the population.

Etiology

RA etiology is not fully understood but, as in all multifactorial diseases, its development is known to be strongly influenced by genetic (low penetrance), environmental and intrinsic factors, such as age, gender and ethnic group (Fig. 1). Different environmental factors have been implicated in its pathogenesis: viruses (Epstein-Barr, parvovirus B19), bacteria (Streptococcus, Mycoplasma, Proteus and E. coli), cigarette, silica and hormones, among others. It is well documented that both cigarette and silica strongly contribute to the development of this AD.

On the other hand, the most important genetic risk factors for RA include different alleles of HLA II, PTPN22, STAT4, CTLA4, TRAF1, PADI4, FCRL3, TNFAIP3, TNF-α and miRNAs, among others.
Pathophysiology

RA involves several inflammation cascades that lead to synovial tissue persistent damage, articular cartilage destruction and bone erosion. Different cells, such as B and T cells and macrophages, act as regulators of the innate and adaptive immune system; several cytokines, chemokines, cytokine and chemokine receptors, adhesion and adaptor molecules, among others, have been implicated in RA pathophysiology. One of the main pro-inflammatory cytokines involved in RA pathogenesis is the tumor necrosis factor α (TNF-α). In a murine model, TNF-α overexpression was shown to be sufficient to induce RA. Another study showed that TNF-α can induce the expression of other pro-inflammatory cytokines, such as interleukin 1 (IL-1) and interleukin 6 (IL-6), which play a fundamental role in RA severity and activity. On the other hand, both TNF-α and IL-1B and IL-6 have been shown to be able to induce the expression of protein-coding gene such as intracellular and vascular adhesion molecules (ICAM and VCAM, respectively), which are important in the communication between cells and matrix metalloproteases, fundamental in cartilage destruction and bone erosion, in addition to inducing the synthesis of autoantibodies, which are a factor of severity and bad prognosis in RA. Given the importance of TNF-α in RA, a set of antibodies targeted against this TNF-α has been developed (biologic therapy). On the other hand, pro-inflammatory interleukin 17, mainly produced by Th 17 cells, has been implicated in all development stages of the disease; it has been shown to be an important risk factor that contributes to RA chronicity, since it induces the production of different cytokines in the synovium of patients with RA, has a synergic function with other cytokines that harm the synovial tissue, promotes synoviocytes and inflammatory cells survival, and is involved in their maturation; this way, this cytokine leads to an increase in the number of synoviocytes and inflammatory cells, hyperplasia, and the exacerbated inflammation observed in the joints of patients with RA.

Genetics in RA

Studies conducted in families and twins have shown the importance of the genetic factors in susceptibility and severity of RA. The prevalence of this disease in first-degree relatives (where there is one individual affected by RA) is considerably higher than in the general population, although RA is not transmitted in families with high frequency. Even when concordance is relatively low in comparison with other ADs (30%), the concordance rate for RA has been estimated to be 12-15%. In studies on sibling recurrence risk of affected individuals (I) compared with the general population (unrelated individuals), it is 2-10-fold higher. RA hereditability has been estimated to be 60-70%.

Different protein-coding and non-coding RNAs (specifically miRNAs) genes, which participate in the innate and adaptive immune response have been associated with the RA pathogenesis. These include different HLA classes I, II and III alleles, cytokines, chemokines, adhesion and adaptor molecules, metalloproteases, cytokine and chemokine receptors, Fc-type receptors, integrins, signal transducers (kinases, among others), miR-146a and miR-499, among others, of which further details will be provided later (Table 1).

RA-associated genes

HLA class II

The main genetic risk factor associated with RA is located at the 6p21 cytoband. This region comprises 3.6 Mb and is divided into different class I, II and III HLA genes (HLA class III genes are not involved in antigen presentation). HLA-II has been documented to contribute to up to one third part of the genetic component associated with susceptibility to RA. Recent studies suggest that this percentage is overestimated. Data indicate that HLA-DRB1 contributes only by 11%. HLA-I and II genes are highly polymorphic and encoding for cell-surface heterodimeric proteins and have as primary function binding to own or foreign short peptides and presenting them to CD8+ and CD4+ T cells, respectively. In both cases, HLA-mediated peptide binding and presentation on the cell surface are an indispensable requirement for the formation of the trimolecular peptide-HLA-T cell receptor (TCR) complex, which leads to T cell activation. In 1978, Stasny P., through a candidate gene study, identified that 78% of patients with RA were positive to HLA-DRw4 in comparison with 28% of healthy controls; subsequently, multiple alleles were identified within HLA-DRB1, which were shared by patients with RA; at the amino acid sequence level, which were called the shared epitope (SE). This amino acid sequence located at positions 70-74 of the QKRAA, QRRRA or RRRAA amino acids of the...
DR-β1chain third hypervariable region is associated with high risk for RA. The odds ratio (OR) for this epitope is $4.37^{32}$. On the other hand, several GWAS and meta-analyses have identified and confirmed, respectively, the association of several HLA-DRB1 SNPs with RA; one of them, rs6910071A/G confer susceptibility (OR $2.88^{34}$) and a genetic association of $1 \times 10^{-299}$, whereas rs17878703, located at position 11 of the HLA-DRB1 peptide sequence, shows a p-value $< 10^{-33,34,677}$.

**PTPN22**

The PTPN22 gene, located at the 1p13.3-13.1 cytoband, represents the second most important susceptibility locus associated with RA (this locus is only after the HLA-class gene). PTPN22 (also known as LYP protein), or protein tyrosine phosphatase, non-receptor type 22, belongs to the protein tyrosine phosphatases (PTPs) family, which are implicated in the negative regulation of TCR-mediated signalling. Tyrosine kinases and PTPs regulate the signal transduction of a wide group of physiological processes, including the immune response. PTPs disturbances produce immune anomalies and different human diseases.

A candidate gene study and a GWAS conducted in 2004 identified a non-synonymous C1858T SNP in the codon 620 (R620W) of PTPN22 (changes the arginine amino acid for tryptophan and is located at exon 14) that was associated with type 1 diabetes mellitus, RA.
and systemic lupus erythematosus (SLE)\textsuperscript{38}. Subsequent studies showed an association between this SNP and other ADs\textsuperscript{37}. The substitution of this amino acid occurs in a LYP polyproline domain (involved in PepCsk binding), and as a consequence, the 620W variant shows reduced interaction with Csk\textsuperscript{37}. A recent study suggests how-of-function PTPN22 alleles can lead to the population expansion of effector and/or memory T cells and a predisposition to ADs\textsuperscript{39}.

Multiple studies in different ethnical groups have reported an association of this variant with RA; OR values range from 1.3 to 2.13. T allele occurs more frequently in patients rheumatoid factor (RF)-positive than in patients FR-negative. In European populations, this allele shows OR values of 1.423 and a p-value = 1.0 x 10\textsuperscript{-8}, whereas non-European have an OR of 1.902 and a p-value = 2.8 x 10\textsuperscript{-8}\textsuperscript{37,40,41}. One meta-analysis categorically confirmed the association of T allele with susceptibility to RA (OR: 1.94; p = 91 x 10\textsuperscript{-74})\textsuperscript{34}. In the Mexican population, T allele was associated with susceptibility to RA: OR: 2.5 and p = 0.008\textsuperscript{42}.

**PADI4**

The PADI4 gene, located at the 1p36 region (a region previously linked to RA), encoding citrullinating enzyme peptidylarginine deiminase 4, which catalyzes the proteic conversion of arginine residues into citrulline, generating citrullinated proteins\textsuperscript{43}. This phenomenon can cause the loss of immune tolerance and originate anti-CCP synthesis. Anti-CCP identification has served to provide with accurate diagnosis and prognosis in RA\textsuperscript{44,45}. This enzyme has been observed to be overexpressed in synovial fluid and synovial tissue of patients with RA\textsuperscript{44}. A candidate gene study identified several PADI4 SNPs (PADI4_89, PADI4_90, PADI4_92 and PADI4_104) involved with risk for RA. In addition, a PADI4 haplotype (associated with susceptibility) that affected transcript stability was identified and associated with high levels of anti-citrullinated peptides (anti-CCP) were observed to have higher risk for RA: OR: 2.5 and p = 0.008\textsuperscript{42}.

**CTLA4**

Another gene that has shown association with RA is CTL4; this locus is located in the 2q33 region and encode for cytotoxic T cells antigen 4. The CTL4 protein negatively regulate the T-cell activation by two mechanisms: negative signaling and competitive antagonism of the CD28/B7-mediated co-stimulation pathway; an anti-CTLA4 therapy has been even developed, the goal of which is to bind to B7 (co-stimulating molecule) and prevent T-cells’ activation signal\textsuperscript{53}.

By means of candidate gene studies and later with GWAS, different RA-associated genes were identified, especially in Caucasians\textsuperscript{52,56}. A meta-analysis of GWAS, categorically identified the association of the CTLA4 SNP rs3087243 (CT60) with RA (OR: 0.44; p = 1 x 10\textsuperscript{-8})\textsuperscript{34}. Another CTLA4 SNP that has been constantly
analyzed in different populations is 49A/G (rs231775). Data indicate that the 49A/G SNP is associated with risk for RA in Asians (OR: 1.16; p = 0.002), but not in Europeans57,58. Two studies conducted in RA patients of Western Mexico identified that A allele confers risk for the development of RA (OR: 1.45; p = 0.01), whereas the CT60 SNP has been associated with protection (OR: 0.61; p = 0.024)59,60.

**STAT4**

STAT4, located at genetic cytoband 2q32.2, encode for the transcription factor known as signal transducer and activator of transcription 4, which transmits signals induced by several cytokines, including interleukin 12, interleukin 23 and interferon 1. STAT4 is implicated in Th1 and Th7-cells differentiation and proliferation, which are crucial in the development of chronic and autoimmune inflammatory diseases61. A candidate gene study identified 4 SNPs (all of them with high linkage disequilibrium) located at STAT4 intron 3 that were associated with RA and SLE61. The SNP that showed more evidence of association with RA was rs7574865G/T (OR: 1.32; p = 2.81 x 10-7)61. This and other STAT4 SNPs (e.g., SNP rs134269947A/G confers an OR of 1.15; p = 7.2 x 10-10) have been associated with RA, SLE, systemic sclerosis and Sjögren syndrome; and, by means of GWAS and meta-analyses, their association with RA has been confirmed in different populations62-65. A meta-analysis conducted in 2013 identified that the rs7574865G/T SNP was associated with RA in Latin Americans (OR: 1.36; p = 0.008)66.

**IRF5**

This gene, located at genomic cytoband 7q32, encode for interferon-regulator factor 5 (IRF-5), which belongs to the interferon-regulation factors family. Its functions include cell-cycle regulation, apoptosis and immune and inflammatory response by means of induction of different pro-inflammatory cytokines that are fundamental in the pathophysiology of RA67-69. IRF5 contains several polymorphisms; some of them are rs2004640T/G, rs729302A/C and rs752637A/G (all of them are functional). Several have been identified and associated, by means of candidate gene studies and GWAS, with RA and other ADs, especially with SLE and multiple sclerosis (MS)68,70,71.

The G/G genotype of the SNP rs2004640T/G correlated with an IRF-5 isoform that includes exons 1A and 1C, whereas those carrying T allele are correlated with a transcript that carries exons 1A, 1B and 1C; the constitutive transcripts are those that carry exons 1A and 1B, and are expressed in plasmacytoid dendritic cells and B cells, while the transcripts that carry exon 1C are inducible by type 1 interferon. IRF5 transcripts abnormalities owing to these variants have been proposed to be able to confer risk for developing RA70. This SNP has been analyzed by means of a a meta-analysis, with the results showing this variant to be associated with risk for RA (OR: 1.14; p = 0.003)70. A second meta-analysis in RA, where IRF5 SNPs rs2004640T/G, rs729302A/C and rs752637A/G were assessed, showed an association with susceptibility of each one of them in different ethnic groups, especially in Europeans and Asians69. Another IRF5 SNP, identified through a GWAS’ meta-analysis, reported an association between rs10488631T/C and RA (OR: 1.19; p = 1.2 x 10-6)34.

**FCRL3**

FCRL3 is located at genomic cytoband 1q21-23 and encode for Fc receptor-like protein 3; its function is to regulate B cells activation through two modalities: tyrosine-based activation or inhibition72. Kouchi et al., in 2005, performed a fine mapping of 1q21.23 region, by means of SNPs-type gene markers, in Japanese patients with RA, SLE and autoimmune thyroiditis. The results showed that the FCRL3-169T/C (rs7528684) SNP was associated with susceptibility for the development these three ADs; in RA, it showed an OR of 2.15 and a p-value = 0.0000008573. In spite of the genetic association robust result in this population, other populations failed to replicate the association between this functional SNP that alters the binding affinity to nuclear transcription factor kappa B (NF-kB) and is associated with higher expression levels of FCRL3, and with susceptibility to RA, except in Asian populations. Two recently published meta-analyses show that -169T/C SNP is specifically associated in Asian populations (OR: 1.101; p = 0.002)74,75. One study published in 2013, in our population, showed that the FCRL3-169T/C allele was associated with protection against developing juvenile rheumatoid arthritis (JRA), and was gender-dependent (in males, OR: 0.57; p = 0.003)76. The role of FCRL3-169C allele should be assessed in our population, in order to determine if it's associated with RA, either with susceptibility or protection.
TNFAIP3

The TNFAIP3 gene, which is located in the short arm of chromosome 6, at cytogenetic band q23, encode for TNF-α-induced protein 3 (TNFAIP3, also known as A20); and its function is to negatively regulate NF-kβ in response to multiple stimuli, and inhibits TNF-α-induced inflammation and apoptosis. One study showed the expression of TNFAIP in human synovial membrane and in several cell types that play important roles in RA pathophysiology, such as synoviocytes, lymphocytes and fibroblasts.

Different genetic studies have identified that the TNFAIP gene is associated with RA and other ADs. One meta-analysis conducted in patients with RA shows that the SNP rs6920220A/G is associated with susceptibility (OR: 1.22; p = 1 x 10^-9). Data on European population show this same SNP associated with susceptibility (OR: 1.23; p = 1 x 10^-9). This same meta-analysis shows that the SNP rs10499194C/T is associated with RA, specifically in Asian populations (OR: 1.25; p = 6.7 x 10^-4).

miRNA

Currently, micro-RNAs (miRNAs) have been the subject of study in different diseases, such as different types of cancer, cardiovascular conditions and ADs, including RA. miRNAs are produced from DNA, as long, non-coding RNAs, and are known as primary (pri-miRNA); subsequently, different RNases in the nucleus produce precursor miRNAs (pre-miRNAs) of approximately 70 nucleotides and, finally, other RNases located in the cytoplasm produce the 18-22-nucleotide-long mature forms of miRNA, whose main functions include regulating translation repression and degradation of different miRNAs; this way, miRNAs importantly regulate inflammatory, apoptotic and immune system-activation processes and other biological events. SNPs-type alterations have been reported in these genes; these variants can affect their structure and the processing from pri-miRNA to miRNA, ultimately altering their binding to their miRNA targets and their biological function. Some SNPs located in the miR-146a and miR-499 genes have been associated with RA in different populations. Two meta-analyses assessed the importance of two SNPs in miR-146a (rs2910164G/C) and miR-499 (rs3746444G/A) in RA. The first meta-analysis reported that the rs2910164G/C SNP did not show association with RA, whereas miR499 rs3746444G/A SNP was associated with susceptibility (OR: 1.62; p = 0.001). Second meta-analysis showed an association between the miR499 rs3746444G/A SNP and RA, especially in Asian populations. One study carried out in a Mexican population with JRA reported that miR-146a rs2910164G/C SNP was not associated with this disease, but it did with pediatric asthma. This genetic variant should be assessed in our population in order to determine if it is an important susceptibility factor in the pathogenesis of this AD.

Other genes that have shown genetic associations with RA

Other GWAS have identified the association of different loci that regulate the innate and adaptive immune system with RA, but there are not yet many studies assessing the association and replicating it in other populations. These genes include the following: CD28 (rs1980422C/T; OR: 1.11; p = 1.3 x 10^-9) and CD40.
(rs4810485G/T; OR: 0.87; p = 8.2 x 10^-9)104, FCGR3A (codon 158VF functional SNP, valine for phenylalanine; OR: 1.25; p = 0.01)105, TYK2 (rs34536443C/G; OR: 0.62; p = 2.3 x 10^-14) and IRAK1 (rs13397A/G; OR: 1.27; p = 1.2 x 10^-12), among others33.

Conclusions

The development of RA is strongly influenced by multiple environmental and genetic risk factors. Advances in genetics and genomics over the past decade have been impressive; candidate gene studies and GWAS have helped to identify different susceptibility loci implicated in the pathogenesis of this AD. Different genetic variants, mainly of the SNP type, located in different genes that produce proteins or non-coding RNAs and regulate the innate and adaptive immune system, have been associated with susceptibility to RA. Genes identified as causing susceptibility to RA include class II HLA, PTPN22, STAT4, PADI4, FCRL3, TNFAIP3, CTLA4, TRAF1-C5, TNF-α, HLA, PTPN22, STAT4, PADI4, among others. On the other hand, genetic/genomic studies have helped us to better understand the distribution of certain alleles, genotypes and haplotypes, and how these are associated with RA susceptibility and/or protection in different populations. Finally, functional studies in genes that produce proteins and non-coding RNA and that regulate the innate and adaptive immune response have helped us to better understand the effect of these alleles of different SNPs on gene expression, translation, alternative splicing and stability and degradation of miRNAs or their binding to their targets. It is important assessing the allelic and genotypical distribution of different SNPs in genes that have not yet been analyzed in our population, in order to establish their role in susceptibility to RA.

Acknowledgements

The authors thank the Hospital Juárez de México for all the support provided for the performance of this work.

Declaration of interests

The authors declare not having any conflicts of interest with regard to this review.

References


