Journal of Advanced Research (2016) 7, 1–16

Cairo University

Journal of Advanced Research

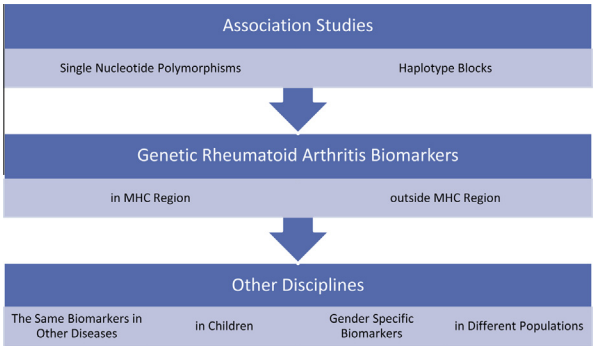
REVIEW

Identification of rheumatoid arthritis biomarkers based on single nucleotide polymorphisms and haplotype blocks: A systematic review and meta-analysis

Mohamed N. Saad [a,*], Mai S. Mabrouk [a], Ayman M. Eldeib [b], Olfat G. Shaker [c]

[a] *Biomedical Engineering Department, Faculty of Engineering, Misr University for Science and Technology, 6th of October City, Egypt*
[b] *Systems and Biomedical Engineering Department, Faculty of Engineering, Cairo University, Giza, Egypt*
[c] *Medical Biochemistry and Molecular Biology Department, Faculty of Medicine, Cairo University, Cairo, Egypt*

GRAPHICAL ABSTRACT



* Corresponding author. Tel.: +20 2 22094734.
E-mail address: m.n.saad@ieee.org (M.N. Saad).
Peer review under responsibility of Cairo University.

Production and hosting by Elsevier

http://dx.doi.org/10.1016/j.jare.2015.01.008

ARTICLE INFO

Article history:
Received 28 September 2014
Received in revised form 13 January 2015
Accepted 20 January 2015
Available online 4 February 2015

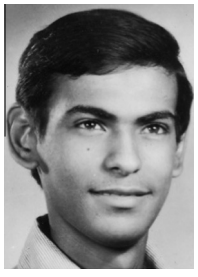
Keywords:
Haplotype block
Linkage disequilibrium
Major histocompatibility complex
Rheumatoid arthritis
Single nucleotide polymorphism

ABSTRACT

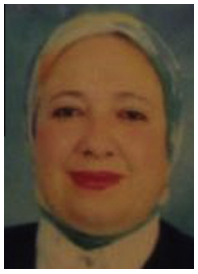
Genetics of autoimmune diseases represent a growing domain with surpassing biomarker results with rapid progress. The exact cause of Rheumatoid Arthritis (RA) is unknown, but it is thought to have both a genetic and an environmental bases. Genetic biomarkers are capable of changing the supervision of RA by allowing not only the detection of susceptible individuals, but also early diagnosis, evaluation of disease severity, selection of therapy, and monitoring of response to therapy. This review is concerned with not only the genetic biomarkers of RA but also the methods of identifying them. Many of the identified genetic biomarkers of RA were identified in populations of European and Asian ancestries. The study of additional human populations may yield novel results. Most of the researchers in the field of identifying RA biomarkers use single nucleotide polymorphism (SNP) approaches to express the significance of their results. Although, haplotype block methods are expected to play a complementary role in the future of that field.

Mohamed N. Saad received the BSc and MSc from Systems and Biomedical Engineering Department, Cairo University, Giza, Egypt, in 2005, 2011, respectively. From 2006 to 2010, he was a clinical engineer in the Department of Medical Equipment Management at the Suez Canal Authority. He is currently an Assistant Lecturer in the Biomedical Engineering Department, Misr University for Science and Technology (MUST). His research interests include Biomedical Image Processing, Bioinformatics, and Biostatistics. He has authored three research papers in the area of Biomedical Image Compression and Bioinformatics.

Mai S. Mabrouk received the BSc degree from Systems and Biomedical Engineering Department, Cairo University, Giza, Egypt, in 2000. She completed her MSc and PhD in Biomedical Engineering from the same school in 2004 and 2008, respectively. She is an assistant professor in the Biomedical Engineering Department, Misr University for Science and Technology (MUST), since August 2008. Her research interests include Biomedical Image Processing, Bioinformatics and Digital Signal Processing in addition to Genomic Signal Processing. She has authored several research papers in the area of Image processing and Bioinformatics.



Ayman M. Eldeib, received the PhD degree in 1995. He is an associate professor at Systems and Biomedical Engineering Department, Faculty of Engineering, Cairo University, Egypt. He has valuable technical, academic, and industrial skills that produced many research papers and three USA patents. He served as a research scientist and the principal investigator of a medical imaging research project at Electrical Engineering Department, University of Louisville (UofL), KY. He served as the scientific program chair of CIBEC 2012 conference that is the 6th Cairo International Conference on Biomedical Engineering, sponsored by the IEEE Engineering in Medicine and Biology Society (EMBS), and was held from December 20–22, 2012 in Cairo, Egypt. He is a senior member of the IEEE.



Olfat G. Shaker, received the M.D degree in 1993. She is a professor at Medical Biochemistry and Molecular Biology Department, Faculty of Medicine, Cairo University, Egypt. She is a member of the European Society of Gene Therapy. She has over hundred international and local specialized publications. She has participated in and attended over hundred conferences. She received The National Prize for Medical Science for the years 1999, 2010, Egypt. She received the Cairo University Prize for Biochemistry for the years 2002, 2006. Also, she received awards from Cairo University for international publications for years 2006, 2007, 2008, 2009, 2010, 2011 and 2012.

Introduction

RA is an autoimmune disease that causes chronic inflammation of the joints and other areas of the body. RA is characterized by periods of disease development and attenuation. RA tends to affect multiple joints usually, but not always, in symmetrical patterns [1].

The US and UK populations are affected by RA disease with 1% approximately. In some other ethnicities, such as China, Japan and some black populations in rural South Africa, assessment of the spread of the disease is as low as 0.2–0.3%. The affected women are approximately twice the affected men. It most often starts within the range of 45–55 years of age [2].

The precise etiology of RA has not been established yet. The cause of RA is a very active area of the worldwide research. It is believed that the tendency to develop RA may be genetically inherited. Also, environmental factors, such as smoking tobacco, may cause the malfunction of the immune system in susceptible individuals [3].

There is no singular test for diagnosing RA. Instead, RA diagnosis is based on a combination of (1) the presentation of the joints involved, (2) the characteristic joint stiffness in the morning, (3) positive rheumatoid factor (RF) and citrulline antibody, and (4) the findings of rheumatoid nodules and radiographic changes. There is no known specific cure for RA. To date, the goal of treatment in RA is to (a) reduce joint inflammation and pain, (b) maximize joint function, and (c)

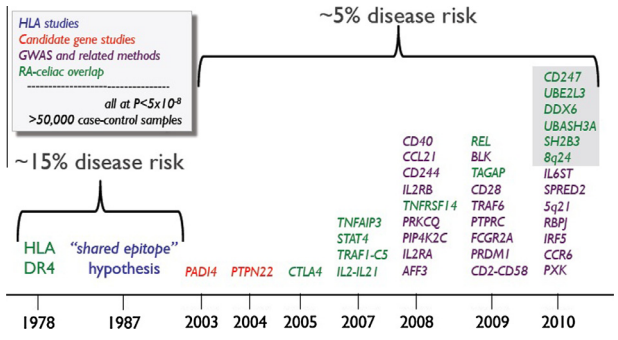


Fig. 1 More than 35 risk loci that have been previously identified as biomarkers for RA disease [10].

prevent joint destruction and deformity. Treatment is customized according to many factors such as disease activity, types of joints involved, general health, age, and patient's occupation.

The first-line of drug treatment, such as cortisone, is used to reduce pain and inflammation in RA patients. The disease-modifying anti-rheumatic drugs (DMARDs), such as methotrexate, promote disease remission and prevent progressive joint destruction. In some cases with severe joint deformity, surgery may be necessary [4].

Biological drugs which are considered other kinds of DMARDs, offer more specific action and provide clues to other biological pathways and biomarkers. They work on the immune system and block signals that lead to inflammation. For example, (etanercept, infliximab, and adalimumab) block tumor necrosis factor alpha (TNFα) which is an important player in RA. A test, predicting the response to anti-TNFα treatment, would be an important tool to rheumatologists. This test will reduce the deterioration of the patient and save time and money by defining the most effective biological drug before its usage [5].

Due to the extremely increase in the diseases, re-characterization of disease in pathological and physiological terms using biomarkers is a turn to the future of medicine. A biomarker is defined as any parameter that can be objectively examined and measured as a marker of (a) normal biological processes, (b) pathogenic processes, or (c) pharmacological response to a therapeutic intervention. These indicators could include a wide range of biochemical materials, such as nucleic acids, proteins, sugars, lipids, and metabolites, as well as whole cells or biophysical characteristics of tissues. Detection of biomarkers, either individually or as larger sets or patterns, can be accomplished by a wide variety of methods, ranging from biochemical analysis of blood or tissue samples to biomedical imaging [6].

SNPs are considered as the most common type of sequence variation in genomes. Most commonly, SNPs can serve as valuable genetic biomarkers; guiding biologists in detecting genes that are related to common diseases [7]. In this review, the SNPs were used as biomarkers for detecting RA. The variation in these nucleotides has higher frequency in affected people than in normal individuals. Most of these nucleotides are located within genes or near genes. Most of those genes are involved in immune regulation. As RA is an autoimmune disease, so those genes suggest an important set of processes involved in RA pathogenesis [8].

Genome-wide association studies (GWAS), using SNPs, have marked a collection of genes that may be associated with RA susceptibility, especially the genes that encode immunoregulatory factors [9]. TNFα is a part of the immunopathogenesis and an early stage biomarker for RA. It is reasonable to assume that TNFα levels are elevated long before the appearance of symptoms on the patient [5]. Fig. 1 shows a historical view of RA genetic biomarkers until 2010. *HLA* (human leukocyte antigen)-*DR4* and shared epitope (*SE*) (multiple alleles at the *HLA–DRB1* locus) represent approximately 15% of RA disease risk factors. The last decade reflects enormous growth in RA biomarker findings [10].

Haplotype block

Linkage Disequilibrium (LD) specifies that the nearer alleles that coexist on the same chromosome tend to be linked to each other. The alleles that are far away from each other are more likely to take place by chance, since the recombination events between such alleles are more likely. The target of association mapping is defining alleles that raise the susceptibility to a disease. These alleles are more frequent among cases than among controls. The SNP associated with a disease (risk or protective) is an evidence for the association of its region with the disease. So, the LD patterns are very useful for the identification of other indirect SNPs at the same region.

If two alleles that coexist on the same chromosome are linked to each other, then a deviation (D) will be presented in the observed frequencies from the expected frequencies. (D) is one of the most common measures of LD [11]. Other two valuable measures of LD are correlation coefficient (r^2) and normalized deviation (D'). (r^2) takes a value from zero to one reflecting the strength of association between pairs of alleles. Generally speaking, the strength of association between SNPs decreases as the genetic distance between these SNPs increases. The perfect LD quantitatively means ($r^2 = 1$) [12]. The strong LD has a (r^2) cut-off of 0.8 as generally seen in published papers [13–15].

The recent sequencing/genotyping technologies allow completely sequencing large DNA segments or genotyping millions of SNPs. However, the presence of LD between SNPs allows reducing the number of genotyped SNPs and, therefore, reducing also the cost of the association study without a significant loss in the power of association [7].

Some studies on different genes of the human genome showed that the structure of the human genome is blocky in nature [16]. The observations of experiments concluded that many chromosomes have blocky patterns [17]. The existence of that blocky structure has a great advantage to the association studies. Haplotype blocks contain the structure of LD in the human chromosomes. These blocks describe the SNP pattern using a somewhat uncomplicated scheme. The main properties of the haplotype block are: (a) the reduced haplotype diversity within the block; (b) absence or very low number of recombination events inside the block i.e. high LD; (c) recombination events present between blocks i.e. low LD [18]. Although recombination events are usually at the boundaries of the haplotype blocks, a homogenous recombination region will deceivingly look like a blocky pattern of haplotypes. The relationship between any SNP in the blocky pattern and all the other SNPs in the same pattern is

statistically significant; as any SNP contributes to the whole block. A reduced set of tagging SNPs can be used to identify all observed haplotypes instead of analyzing all SNPs within the block [19].

Block extent varies greatly with different ethnicities. Depending on this information, the US Nat'l Human Genome Research Institute (USNHGRI) has started an extensive endeavor, called the Int'l HapMap Project in 2002. This project intended to construct a genome-wide map of LD and haplotype blocks among populations. The sampled populations were from European, Asian, and African ancestries. Four countries provided the 270 DNA samples for HapMap project which are US, Japan, China, and Nigeria. The US provided samples of 30 trios (two parents and one adult child) from Utah residents of European ancestry. Japan provided samples of 45 unrelated residents from Tokyo area. Beijing, China provided 45 samples of unrelated Han Chinese. Nigeria provided samples of 30 trios from the Yoruba people of Ibadan. The number of genotyped SNPs is more than one million SNPs with 5 kb (kilo base) inner intervals [20].

The 1000 Genomes Project was launched in 2008. The aim of the project was to identify the genetic variants that have at least 1% allele frequencies in the studied populations. The studied populations were East Asians, South Asians, Africans, Europeans, and Americans. Researchers could benefit from the identified variants by relating them to diseases through association studies. Also, the project targeted the haplotype background and LD patterns of the variant alleles [21].

The challenge of segmenting the genome into blocks of low haplotype diversity is called haplotype block partitioning. The main haplotype block partitioning methods are the four gamete test (FGT) [22] and the confidence interval test (CIT) [23]. There are other approaches to partition haplotype blocks such as solid spine of LD [24], hidden markov model [25,26], dynamic programming-based algorithm [27–29], wavelet decomposition [30], greedy algorithm [31], minimum description length [32,33], and block entropy [34].

The target of haplotype block partitioning is to decrease the complexity of association mapping so as to deal with haplotype blocks instead of individual SNPs. Other important applications of haplotype block partitioning are SNPs tagging, post-GWAS SNP-set analysis, SNPs-to-gene mapping [35].

Haplotype blocks vs individual SNPs

Individual SNP approaches accomplished impressing findings in case of monogenic diseases (such as cystic fibrosis). On the other hand, they did not reach the same success in complex diseases (such as type I diabetes mellitus). Haplotype blocks may capture interaction between SNPs (SNPs that contribute to the disease status together but not separately), which is not possible with individual-SNP tests [35]. Using individual SNP approaches expands the dimension of association testing. While using haplotype block methods reduces it with ensuring reasonable error rates. So, the haplotype block methods are expected to increase the power of association more than the individual SNP approaches. The drawbacks of haplotype blocks methods, (a) haplotype data are more expensive to collect than genotype data; (b) phase (i.e. haplotype) estimation, where the phase of SNPs can be homozygous or heterozygous; (c) different haplotype block methods lead to different haplotype blocks resulting in a problematic decision making in choosing the best method for association mapping; (d) if the number of haplotypes inside a block increases, the degree of freedom of the block increases resulting in reduced power; (e) applying statistical procedures ends up with computational error. Some studies debated that individual SNPs will have at least the same power of association as haplotype blocks. Experimental results showed unclear conclusions in this debate. The contradictory findings support that the performance of the methods may depend on the nature of the studied data.

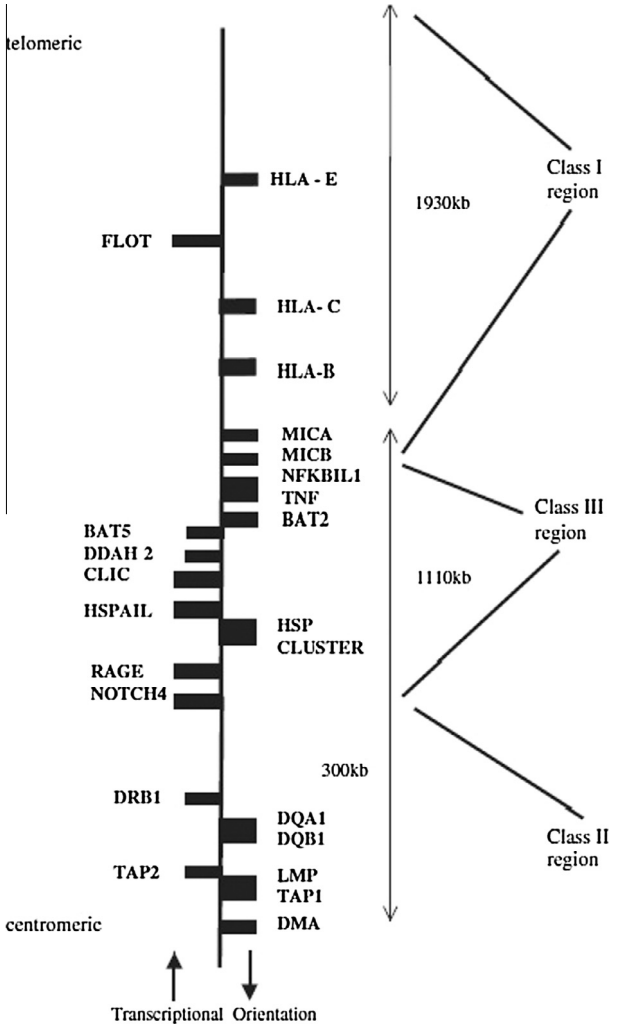


Fig. 2 The MHC region showing class I, class II, and class III regions [39].

Shim et al. [36] aimed to measure the power of association of the two strategies: the individual SNP approaches and the haplotype block methods. They used a dataset from the North American RA Consortium (NARAC) provided by the Genetic Analysis Workshop 16 (GAW16). They tested 513,935 SNPs in 868 cases and 1194 controls. The examined haplotype block methods were FGT and CIT implemented using Haploview program. Haplotype block methods had lower *p*-values than individual SNP approaches. A low *p*-value means that the probability of observing these results by chance is very small. Haplotype block methods reduced the no. of the required tests using individual SNP approaches by a factor of about 0.65.

Some biomarkers were detected by haplotype block methods only. This may be due to the higher ability of haplotype block methods for detecting rare alleles. Other biomarkers were detected by individual SNP approaches only. This may be because of the neighboring SNPs, to the causal SNP in the block, having weakened the strength of the association. Finally, they suggested the use of the two strategies to maximize the detection of RA biomarkers.

RA biomarkers

The main biomarkers that have been considered as risk factors for RA are within the major histocompatibility complex (MHC) region which is positioned on chromosome 6 (6p21.3). The HLA region within the MHC contributes to almost 50% of the genetic susceptibility for RA. Other RA biomarkers outside the MHC region are also considered significant [37,38].

Biomarkers in the MHC region or chromosome 6

The MHC region is highly related to autoimmune diseases. This relation has been shown through many association mapping studies. The MHC region extends over 3.6 Mb, as shown in Fig. 2. The MHC region contains about 220 genes, many of which have immunoregulatory functions [39].

HLA–DR4 and *HLA–DRB1* genes are highly associated with RA. The *HLA–DRB1* associations are intensely detected in the anti-CCP + (anti-cyclic citrullinated peptide positive) antibodies group [40]. TNF locus is one of the most studied loci in the MHC region. TNF locus is located inside the MHC class III region, about 1000 kb from *HLA–DRB1* [39].

HLA locus is associated with RA disease in multiethnic populations. Muazzam et al. [41] confirmed the haplotype association of *DRB1* and *DQB1* variants of HLA class II with RA in Pakistani population. Atouf et al. [42] verified that *HLA–DRB1*04* allele predisposed to RA, while *HLA–DRB1*07* allele had a protective role in Moroccan population.

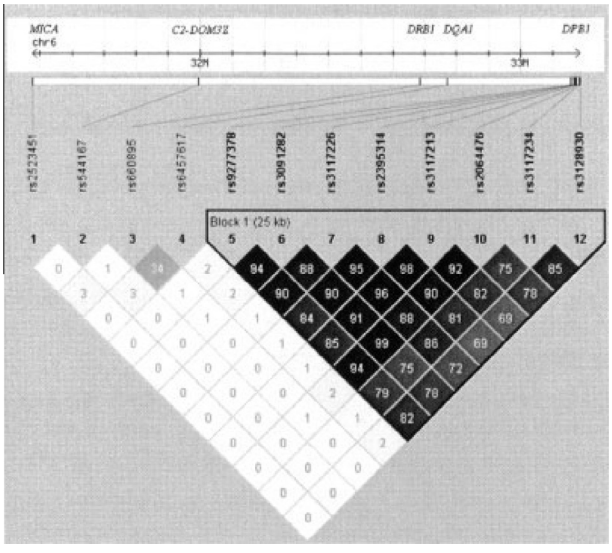


Fig. 3 LD structure for 12 SNPs at *HLA–DRB1* and *HLA–DPB1*. A constructed block was shown including eight SNPs, from SNP 5 to SNP 12 [40].

Al-Swailem et al. [43] provided that *HLA–DRB1*04* prevailed **08* and **10* alleles in association with RA, while *DRB1*06* allele seemed protective to RA in Saudi Arabian population. Ucar et al. [44] confirmed the association of RA *with HLA–DRB1*01*,**04*, and **09* alleles, whereas **13* was the protective allele against RA in Eastern Black Sea Turkish population.

Mourad and Monem [45] detected the association of Syrian RA patients with *HLA–DRB1*01*, **04*, and **10* alleles, while **11* and **13* were the protective alleles against RA. Ben Hamad et al. [46] indicated that *HLA–DRB1*04*, and **10* alleles are related with RA, while patients harboring *DRB1*08* allele had a decreased risk of developing RA in the Southern Tunisian population. HLA alleles had been verified in many populations [47–63].

Ding et al. [40] identified other risk loci in the MHC region. They studied RA patients in two risk groups, defined according to the presence or absence of ACPA. They refined *DRB1* common variants, and detected additional associations with alleles near *HLA–DPB1* for ACPA-positive RA patients, as shown in Fig. 3. For ACPA-negative RA patients, they did not find any linkage with alleles within the MHC region.

Lee et al. [64] provided supplementary risk loci for RA in the MHC region, separated from the class II *HLA–DRB1* locus. This research suggested the existence of two regions of association with RA in the class I region. *HLA-C* locus was associated with RA ($P \sim 5 \times 10^{-5}$). In addition, alleles located near the *ZNF311* (zinc finger protein 311) locus were detected.

A known risk variant at the *TAGAP* (T cell activation RhoGTPase activating protein) gene locus represents a candidate, but not convincing, biomarker for association with RA susceptibility. Chen et al. [65] refined the *TAGAP* risk locus. The SNP (rs212389) demonstrated a potent association with RA disease ($P = 3.9 \times 10^{-8}$). This risk locus prevailed overwhelmingly convincing upon the former RA SNP (rs394581, $P = 2.2 \times 10^{-5}$).

NKAPL (NF-kB (nuclear factor kappa-light-chain-enhancer of activated B cells) activating protein-like) gene is 90% similar to *NKAP* gene. While *NKAPL* functions are still unknown, *NKAP* is a protein implicated in NF-kB-mediated transcriptional activation of *TNF* and *IL-1* (interleukin 1 family). Xie et al. [66] fine-mapped the *NKAPL* gene to verify the association with RA disease. Fine-mapping analyses detected six SNPs in a single haplotype block in Canadian

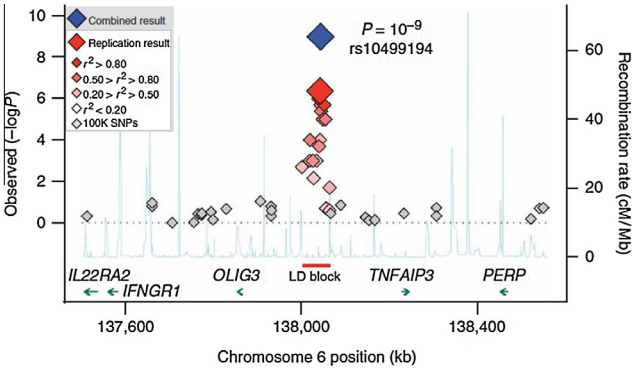


Fig. 4 Case-control association results at 6q23. The associated SNP (rs10499194) was about 165 kb from both *TNFAIP3* and *OLIG3* genes [67].

population. (rs35656932) in the *ZNF193* gene and (rs13208096) in the *NKAPL* gene showed the highest significance of association with RA susceptibility, and were replicated in the US cohort. By illustrating supplementary *NKAPL* alleles, the results confirmed the potent association between *NKAPL* and RA disease. These additional *NKAPL* variants were associated with variants located in *HLA–DRB1* locus. *NKAPL* variants and *HLA–DRB1* variants suggested a synergistic effect between the two regions.

Plenge et al. [67] detected a SNP at 6q23 (rs10499194) approximately 150 kb from (*TNFAIP3* (TNF alpha-induced protein 3), telomeric) and approximately 185 kb from (*OLIG3* (oligodendrocyte transcription factor 3), centromeric), as shown in Fig. 4. In a parallel research, the Wellcome Trust Case Control Consortium (WTCCC) identified potent association of RA to a distinctive SNP (rs6920220) lied 3.8 kb from (rs10499194).

TNFAIP3, which encodes protein A20, is a strong terminator of the NF-kB signaling and is needed for inhibition of TNF-induced signals. TNFα levels are elevated in RA patients. Termination of TNFα is an effective treatment of severe RA. In addition, mice showing shortage in *TNFAIP3* present chronic inflammation. *TNFAIP3* plays a dominant role in autoimmunity. There is a lack of information about *OLIG3*. Mice, with mutation in *OLIG3*, have deficiencies in neuronal development. But, abnormalities are not recognized in the immune system or musculoskeletal system.

Biomarkers outside MHC region

Association mapping studies have led to the detection of genetic biomarkers outside the MHC region. *PTPN22* (protein tyrosine phosphatase non-receptor type 22) is identified as the most statistically significant biomarker for RA disease outside the MHC region in populations of European ancestry. *TRAF1-C5* (TNF receptor-associated factor 1 – complement component 5) region comes next *PTPN22* in the significance of association with RA [68].*PADI4* (peptidyl-arginine deiminases_type 4) appears to have important association with RA in Asian populations. On the other hand, an association between RA and *PADI4* is not confirmed in Caucasian populations. These populations vary in environmental factors which may explain the previous results [69].

Plenge et al. [70] tested 17 SNPs from 14 genes in 2370 RA patients and 1757 controls from the NARAC and the Swedish Epidemiological Investigation of RA (EIRA) datasets. All cases and controls were of European descent. The association of *PTPN22* with ACPA-positive RA was confirmed. Also, an association with *CTLA4* (cytotoxic T-lymphocyte antigen 4) and *PADI4* was provided, but in NARAC dataset only. The results concluded that *PTPN22* is associated with not only RF-positive patients but also ACPA-positive patients. This conclusion was expected, providing the vigorous correlation between RF and CCP situation. Together, these findings gave the most powerful indication of a non-MHC region that influenced the susceptibility to RA.

CD40 (cluster of differentiation 40) signaling plays a very important role in innate and adaptive immunity against microorganisms. *CD40* is a member of the *TNFR* (TNF receptor) family of genes and is expressed on B cells and antigen-presenting myeloid cells. *CD40* exists on chromosome region (20q13.1). Genetic studies on *CD40* showed an association with autoimmune diseases [71].

Raychaudhuri et al. [72] detected an allele at the *CD40* gene locus (rs4810485) which was susceptible for RA. This result showed that *CD40* was a critical player in RA pathogenesis. They also detected another variant at the *CCL21* (chemokine (C–C motif) ligand 21) gene locus (rs2812378). *CCL21* is a gene involved in immunoregulatory functions. Finally, they provided a proof of association at four supplementary gene loci: *MMEL1-TNFRSF14* (membrane metallo-endopeptidase-like 1 – TNFR superfamily member 14) (rs3890745), *CDK6* (cyclin-dependent protein kinase 6) (rs42041), *PRKCQ* (protein kinase C theta type) (rs4750316), and *KIF5A-PIP4K2C* (kinesin family member 5A-phosphatidylinositol-5-phosphate 4-kinase, type II, gamma) (rs1678542).

Gregersen et al. [73] performed a GWAS for RA disease patients from North America on a combined dataset of 2418 cases and 4504 controls. They provided an association at the *REL* (reticuloendotheliosis) locus, which encodes protein c-Rel, on chromosome region 2p13 (rs13031237, rs13017599). The combined dataset also identified other variants at *CTLA4* (rs231735) and *BLK* (B lymphocyte kinase) (rs2736340). c-Rel has biological activity effects on hematopoietic cells, and is an NF-kB family member. c-Rel was associated with RA providing disease tracks that included other known RA susceptibility genes such as *CD40*, *TRAF1*, *TNFAIP3* and *PRKCQ*.

Raychaudhuri et al. [74] tested 22 susceptibility loci in a dataset of 7957 cases and 11,958 controls. Three loci were conclusively approved: (a) *CD2–CD58* (cluster of differentiation 2-cluster of differentiation 58) (rs11586238); (b) *CD28* (cluster of differentiation 28) (rs1980422); and (c) *PRDM1* (PR domain zinc finger protein 1) (rs548234). A supplementary four susceptibility genes were reproduced ($P < 2.3 \times 10^{-3}$): *TAGAP* (rs394581), *PTPRC* (protein tyrosine phosphatase receptor type C) (rs10919563), *TRAF6–RAG1* (TRAF6-recombination activating gene 1) (rs540386) and *FCGR2A* (Fc fragment of IgG, low affinity IIa) (rs12746613). Many of these SNPs reveal some of the shared mechanism of RA pathogenesis as they are also associated with other immunologic diseases.

Kurreeman et al. [75] applied a research plan on distinct populations to confirm the identification of universal RA risk loci. Thirteen known risk variants were tested in different sample sets consisting of overall 4366 cases and 17,765 controls of European, African American, Japanese, and Korean ethnicities. Two alleles (rs3890745 at the 1p36 locus) and (rs2872507 at the 17q12 locus) overstepped genome-wide significance in all 16,659 RA cases and 49,174 controls combined. They used GWAS data to refine these two alleles in Europeans and East Asians, and they confirmed risk association in both ethnic groups. A series of bioinformatics analyses identified *MMEL1–TNFRSF14* at the 1p36 locus and *IKZF3-ORMDL3-GSDMB* (IKAROS family zinc finger 3-ORM1-like 3-gasdermin B) at the 17q12 locus as the genes most likely associated with RA.

RA biomarkers in other diseases

RA and celiac disease (CD) show shared mechanism of disease pathogenesis. They are two distinct autoimmune diseases with co-occurrence in families. GWAS verified the HLA region and

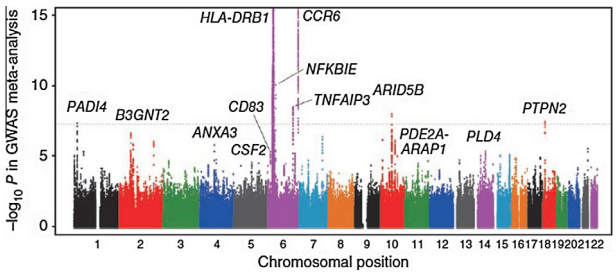


Fig. 5 Manhattan plots of the GWAS meta-analysis for RA in the Japanese population [80].

26 non-HLA genetic variants in each disease. Past studies confirmed six SNPs occurring in both diseases out of the 26 risk loci. Zhernakova et al. [76] thought to enhance the definition of shared disease pathogenesis through identifying new risk loci. They performed a combined analysis of 50,266 samples. The study resulted in the identification of new four SNPs that were not previously verified in either disease: (a) the rs10892279 near the *DDX6* (DEAD (Asp-Glu-Ala-Asp) box helicase 6), (b) the rs864537 near *CD247* (cluster of differentiation 247), (c) the rs2298428 near *UBE2L3* (ubiquitin-conjugating enzyme E2L3), and (d) the rs11203203 near *UBASH3A* (ubiquitin associated and SH3 domain containing A). Four common variants associated in both diseases are confirmed: *SH2B3* (SH2B adaptor protein 3), 8q24, *STAT4*, and *TRAF1-C5*. These results involved genes responsible for immune functions such as antigen presentation and T-cell activation.

KCNB1 (potassium voltage-gated channel, shab-related subfamily, member 1) is a candidate gene for association with RA disease. This candidacy comes from the important function of *KCNB1* in the immune system. Four identical *KCNB1* sub-units are the main components of the functional channel in human T lymphocytes. Autoimmune diseases such as type 1 diabetes mellitus and RA are medicated with several peptide inhibitor of *KCNB1*. Noticeable defect in potassium channels function involving *KCNB1* by autoimmune diseases had been confirmed. Xiao et al. [77] examined the association between *KCNB1* and RA disease in GAW16 dataset. *KCNB1* showed moderate association with RA.

Chung et al. [78] tested common variants associated with RA and GPA (granulomatosis with polyangiitis) (wegener's). They conducted a meta-analysis of GPA showing convincing association with risk loci in *CTLA4*. The studied risk alleles associated with RA were also significantly associated with GPA. RA and GPA may originate from a similar genetic tendency.

Some genes are risk variants for several autoimmune diseases. Li and Begovich [79] stated that the risk variant *TNFAIP3* was the only one that had been recorded in both psoriasis and RA diseases. *TNFAIP3* had also been associated with SLE (systemic lupus erythematosus). These results showed that RA, SLE, and psoriasis may originate from a similar genetic predisposition. Also, *TNFAIP3* was confirmed to play a complex role in different autoimmune diseases.

Okada et al. [80] conducted a GWAS for RA in a Japanese cohort. They confirmed strong association with RA disease at nine loci. The nine variants were as follows:

(1) *B3GNT2* (UDP-GlcNAc: beta-1,3-N-acetylglucosaminyltransferase 2),
(2) *ANXA3* (annexin A3),
(3) *CSF2* (colony stimulating factor 2),
(4) *CD83* (cluster of differentiation 83),
(5) *NFKBIE* (NF-kB inhibitor, epsilon),
(6) *ARID5B* (AT rich interactive domain 5B),
(7) *PDE2A-ARAP1* (phosphodiesterase 2A-ArfGAP with RhoGAP domain, ankyrin repeat and PH domain 1),
(8) *PLD4* (phospholipase D family, member 4) and
(9) *PTPN2* (protein tyrosine phosphatase non-receptor type 2), as shown in Fig. 5.

ANXA3 gene, associated with RA, was also associated with SLE. *B3GNT2* and *ARID5B* were associated with susceptibility to graves' disease.

RA biomarkers in children

RA is an autoimmune disease, generally affects people during middle age. Children with RF and/or ACPA-positive juvenile idiopathic arthritis appear like adults with RA disease, and represent the childhood onset of RA (CORA). Polymorphisms within HLA and many other genes were evaluated for RA risk susceptibility, but had not been investigated intensively in children. To provide evidence that RA risk alleles would also be connected to CORA, Prahalad et al. [81] examined RA SNPs in large set of children with CORA. CORA was most frequent among children of 11 years, and 85% of studied cases were females. CORA and *HLA–DRB1* SNPs revealed a significant association as the situation in RA disease. Genetic studies showed a critical association between CORA and *TNFAIP3*, *PTPN22*, and *STAT4*.

Ezzat et al. [82] studied the susceptibility of Egyptian children to juvenile rheumatoid arthritis (JRA) associated with *HLA–DRB1* locus. They provided that *HLA–DRB1*04* and **14* prevailed *DRB1*01* alleles in the association with JRA in a study of 60 cases and 50 controls. *HLA-DRB1*08* allele seemed to be protective to JRA in Egyptian children.

Gender specific biomarkers

Caliz et al. [83] performed a study to analyze alleles in *Th1* (T helper 1 cells) and *Th17* (T helper 17 cells) which are cell mediated immune response genes. They aimed to investigate whether the studied genes differently control RA susceptibility in females and males. Patients accommodating *Dectin-2* allele (rs4264222T) had a critical RA susceptibility, while *DC-SIGN* (dendritic cell-specific intercellular adhesion molecule-3-grabbing non-integrin) allele (rs4804803G), *MCP-1* (monocyte chemotactic protein-1) allele (rs1024611G), *MCP-1* allele (rs13900T) and *MCP-1* allele (rs4586C) had a protective role against RA.*Dectin-2* allele (rs4264222T) and *Dectin-2* allele (rs7134303G) were associated with RA in females. *MCP-1* alleles (rs1024611G), (rs13900T), and (rs4586C) increased the immunization against RA in females. *DC-SIGN* allele (rs2287886A) was associated with RA in males. *DC-SIGN* allele (rs4804803G) played a protective role in RA in males.

They also concluded that *Dectin-2* SNPs (rs4264222) and (rs7134303) represented a potent two locus interaction model

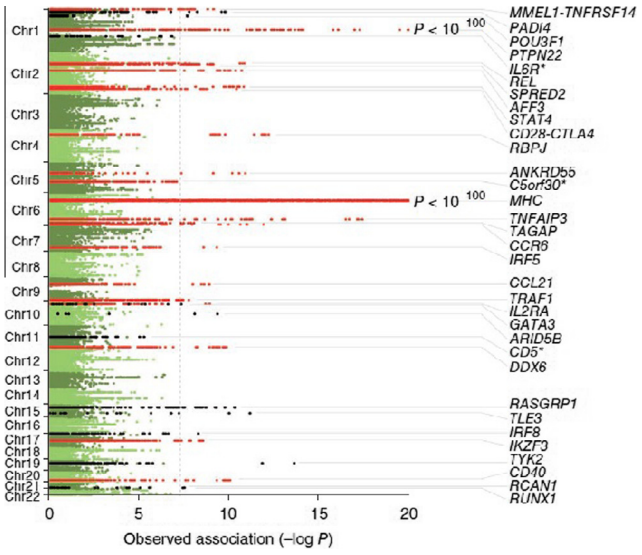


Fig. 6 Manhattan plot of association with RA in the European descents [87].

in females through SNP-SNP interaction analysis of significant SNPs. The last findings were not seen in men.

WTCCC detected a SNP (rs11761231), on chromosome 7q, which presented gender relationship. This SNP showed a critical association with RA susceptibility only in females in a British population. Korman et al. [84] tested the same SNP in a North American population but failed to find any association of the 7q region with RA.

Biomarkers in different populations

Stahl et al. [85] identified seven RA risk loci ($P < 5 \times 10^{-8}$) in a study of all 41,282 samples from Canada, North America, Sweden, Netherlands, UK, and US. The associated variants were located close to genes of immunoregulatory functions, involving the following:

(1) *IL6ST* (interleukin 6 signal transducer),
(2) *SPRED2* (sprouty-related, EVH1 domain containing 2),
(3) *RBPJ* (recombination signal binding protein for immunoglobulin kappa J region),
(4) *CCR6* (chemokine (C–C motif) receptor 6),
(5) *IRF5* (interferon regulatory factor 5),
(6) *C5orf30* (chromosome 5 open reading frame 30), and
(7) *PXK* (PX domain containing serine/threonine kinase).

They also enhanced associations at two RA common variants (*IL2RA* (interleukin 2 receptor, alpha) and *CCL21*) and verified the association at *AFF3* (AF4/FMR2 family, member 3).

Hughes et al. [86] tested whether the validated RA SNPs among people of European ancestry are linked to RA risk loci in an African American population. Twenty-four of the 27 examined SNPs had been confirmed for the association with RA in the European and African American populations. On the contrary, the remaining 3 of the 27 SNPs (*CCR6*, *TAGAP*, and *TNFAIP3* (rs6920220)) failed to represent acceptable association with RA in African American population.

Eyre et al. [87] tested 14 alleles, 5 of which were definitely associated with ACPA-positive RA patients and, 9 of which were associated generally with RA disease. The studied populations were from Canada, North America, Sweden, Spain, Netherlands, UK, and US. The genes involved in RA susceptibility in European descents in that study were shown in Fig. 6.

Many of the identified biomarkers of RA belong to Caucasian populations. Viatte et al. [88] examined the association of Caucasian non-HLA alleles with RA patients in Black African populations. They found weak association between most of the SNPs and West/Central African population. RA susceptibility SNPs, grouped in a set of 28 Caucasian alleles, were highly distinct between the UK and Africa with ($p < 0.001$). They concluded that the genetic risk variants of developing RA are different in Africans from Caucasians. Interestingly, these results confirmed that ethnic group had a great influence on the genetic architecture of RA susceptibility, forcing the researchers to identify the RA SNPs of each ethnic group separately.

Weak association between *PADI4* and RA susceptibility was noticed in Caucasian cohorts. *PADI4* was convincingly associated with RA disease in East Asian populations. Too et al. [89] aimed to verify the association between *PADI4* and RA susceptibility in Malaysian, Chinese, Indians, and other populations from South East Asia. The results presented that *PADI4* and RA were associated in the multiethnic populations from South East Asia and provided supplementary association with *PADI2* risk locus. *PADI2* and *PADI4* genes contributed to enzymes responsible for citrullination. The results thus verified the association of RA with *PADI4* in multiple populations of Asian descent.

The *MTHFR* (Methylenetetrahydrofolate reductase) is a catalytic enzyme which plays an essential role in the conversion of 5,10-methylenetetra-hydrofolate to 5-methyltetrahydrofolate. 5-methyltetrahydrofolate helps in the conversion of methionine from homocystine in vitamin B_{12} dependent pathway. Homocysteine was recorded at high levels in RA patients. In further reactions, methionine is converted to S-adenosylmethionine. S-adenosylmethionine helps in nucleotide methylation in DNA, RNA, and proteins. The *MTHFR* gene is located on 1p36 region. The A1298C is a common polymorphism in the *MTHFR* gene. The A1298C was verified as a biomarker for RA disease in Jewish and Italian populations [90,91]. Contradictory results were shown in American population with Caucasian and African ethnicities [92]. The allele (*1298C*) was found to exhibit lower *MTHFR* enzyme activity, hyperhomocysteinemia, and decreased folate levels.

Okada et al. [93] conducted a GWAS for RA in 29,880 cases and 73,758 controls of European and Asian ancestries. The total number of studied SNPs was nearly 10 million. They identified 42 novel risk loci for association with RA. The results of the study led to the expansion of the detected RA risk loci to 101. They also designed a systematic strategy to identify 98 biological candidate genes at these 101 risk loci. These genes should be targeted for RA drug discovery and further repurpose approved drugs for other diseases for RA treatment.

Haplotype blocks in RA biomarker discovery

Suzuki et al. [94] tested a haplotype block, consisted of 17 SNPs, for association with RA in *PADI4* gene. Expectation

maximization method was used to detect the haplotypes that were expected to have a frequency of more than 0.02 in both patient and healthy individual groups. Four haplotypes, out of 217 possible haplotypes, fulfilled the condition. The most redundant haplotype (haplotype 1) and the second most redundant haplotype (haplotype 2) together represented more than 85% of the observed haplotypes in both groups. Haplotype 1 was mainly detected in healthy individuals and was called the non-susceptible haplotype. Haplotype 2 was mainly detected in patients and was called the susceptible haplotype. Next, they aimed to test the haplotypes affecting the stability of *PADI4* mRNA. They concluded that susceptible mRNA had higher stability than non-susceptible mRNA. Furthermore, the susceptible haplotype was associated with higher levels of antibody to citrullinated peptide in patients' sera.

Ikari et al. [95] tested a haplotype block within *PTPN22* gene, consisted of 8 SNPs spanning 45 kb, for association with RA in Japanese population. Expectation maximization method was used to detect the haplotypes that were expected to have a frequency of more than 0.01 in the studied groups. Finally, they did not detect any association with RA in Japanese population.

Plenge et al. [96] tested a haplotype block in *PHF19* (PHD finger protein 19)-*TRAF1-C5* region, containing nine tag SNPs extends for 100 kb, for association with RA in NARAC II and EIRA II. Omnibus association test was used to test all haplotypes combined for association with RA. The SNP(rs3761847) was identified as a susceptible SNP for RA and verified using logestic regression analyses. Another SNP (rs2900180) was also identified as a susceptible SNP for RA. These two SNPs were in strong LD with each other ($r^2 = 0.62$). So, the causal ungenotyped variant was considered to be in strong LD with these two polymorphisms. Interestingly, they detected a synonymous SNP in *TRAF1* gene (rs2239657) which demonstrated near perfect LD ($r^2 = 0.97$) with (rs2900180).

The detected SNP (rs4810485) in [72] was located in a haplotype block containing about the entire CD40 gene. The SNP(rs1883832) which had been associated with graves' disease, was in a very strong LD ($r^2 = 0.95$) with (rs4810485). Another detected SNP (rs2812378) was located in a haplotype block containing the entire *CCL21* gene.

Plenge et al. [67] tested a haplotype block in 6q23 region, with 20 SNPs extended for 63 kb, for association with RA cases from the Brigham RA Sequential Study (BRASS) and controls from Framingham Heart Study (FHS). Logestic regression analyses and omnibus association test were used to test all haplotypes for association with RA. Six different haplotypes, with five tag SNPs, represented 96% of the total haplotypes with a frequency of more than 0.05. The SNP(rs6920220) was identified as a susceptible SNP for RA, while SNP(rs10499194) showed a protective role against RA. Scherer et al. [97] used the haplotype block, defined in [67], for detecting the linkage between 6q23 region and the rate of joint destruction in early RA Dutch patients. The SNPs(rs675520G) and (rs9376293C) were identified as two susceptible alleles for increased joint destruction in ACPA-positive patients.

Zhang et al. [98] promoted a GWAS based on haplotypes, extended for 1 Mb, to search for risk loci and associated genes for RA. The dataset consisted of 5,393 informative SNPs containing 822 uncorrelated individuals which were obtained from NARAC. They used FGT, CIT, solid spine of LD, and fusion of these methods to identify the haplotype blocks. Density-based clustering algorithm was used to select the final set of risk haplotypes based on the Pearson correlation coefficient for the nearest neighbor method. They detected 25 haplotypes in 18 haplotype blocks. These haplotype blocks contained 33 genes which are highly associated with the risk of RA. The genes *PTPRC* and *F12* (coagulation factor 12) prevailed the other genes in RA susceptibility.

Xie et al. [66] tested three haplotype blocks in *NKAPL* region, consisted of 101 SNPs within 372 kb, for association with RA in Canadian patients. They used the CIT method to identify the haplotype blocks. Benjamini and Hochberg's false discovery rate method showed that there were six statistically significant SNPs associated with RA. These SNPs were all located in the middle haplotype block, across about 70 kb region, which contained *NKAPL*, *ZNF193*, *ZNF307*, and *ZNF187* genes. *ZNF193* (rs35656932) and *NKAPL* (rs13208096) were identified as the highest two susceptible SNPs for RA. This result was verified using stepwise and conditional logestic regression analyses.

SE represented a strong association with ACPA-positive RA patients. *SE* encoded consensus amino acid sequences extended from 70 to 74 positions in *HLA-DRB1*. Raychaudhuri et al. [99] tried to fully explain the association with RA within MHC region in addition to *SE*. They tested 99 classical HLA alleles at two-digit resolution, 164 classical HLA alleles at four-digit resolution, 372 polymorphic amino acid positions, and 3,117 SNPs for association with ACPA-positive RA in BRASS, EIRA, NARAC I, NARAC III, WTCCC, and Canadian datasets using logistic regression. Conditional haplotype analyses uncovered new findings within the MHC region. *HLA-DRB1* codon 11, rs17878703A (a quadrallelic SNP), was identified as the highest susceptible SNP for RA. SNPs at codon 13 were in strong LD (r^2 not specified) with those of codon 11 resulting in a double influence at this region. These findings were verified in a South Korean dataset. The two *SE* positions, 71 and 74, came after the position 11 in association with RA. Furthermore, *HLA-B* codon 9 and *HLA-DPB1* codon 9 were also associated with RA within the MHC region.

Park et al. [100] explored the interaction among haplotypes through two steps. At the first step, they tested the whole genome by individual-SNP methods (codominant and additive models). Then, the haplotype blocks of the significant SNPs were identified. They used the CIT method to identify the haplotype blocks. At the second step, the interactions among haplotypes were detected using expectation maximization method and contingency table. The individual-SNP methods followed by the haplotype block method detected 411 significant SNPs and 146 haplotype blocks. Some previous detected genes that associated with RA were confirmed such as *PTPN22*, *TRAF1*, *NFKBIL1*, *HLA-C*, and *HLA-G*. Two non-synonymous SNPs showed shared mechanism of disease pathogenesis. The SNP(rs2075800) in *HSPA1L* (Heat Shock 70 kDa Protein 1-Like) was associated with both RA and sarcoidosis. The SNP(rs2476601) in *PTPN22* was associated with type I diabetes mellitus, RA, SLE, and hashimoto thyroiditis.

Most of GWAS findings in RA are because of common SNPs that do not affect protein coding regions. Diogo et al. [101] identified SNPs in three genes that encode proteins that involved in RA immunopathogenesis. The three genes were *IL2RA*, *IL2RB*, and *CD2*. Then, they tried to verify the association of *CD2* with RA susceptibility using conditional

Table 1 Detected SNPs associated with RA susceptibility.

SNP	Gene	Position	Method	Population	Comment	Reference
rs3117213	HLA-DPB1	33,172,583	Individual-SNP	EIRA, NARAC	ACPA +	[40]
rs6923005	ZNF311	29,084,051	Individual-SNP	NARAC, Wichita	ACPA +	[64]
rs6930903		29,089,224	Individual-SNP	Rheumatic Disease Data Bank (WRDDB), the National Inception Cohort of RA Patients (NICRAP), Study of New Onset RA (SONORA)		
rs212389	TAGAP	159,068,759	Individual-SNP	BRASS, Canada, EIRA, NARAC I, NARAC III, WTCCC	ACPA + or RF +	[65]
rs2476601	PTPN22	113,834,946	Individual-SNP	EIRA, NARAC		[70]
rs2240340	PADI4	17,336,144	Individual-SNP	NARAC		
rs3087243	CTLA4	203,874,196	Individual-SNP	NARAC		
rs4810485	CD40	44,181,354	Individual-SNP	EIRA, NARAC,	100% (ACPA + or RF+),	[72]
rs1883832	CD40	46,118,343	LD	WTCCC, Nurses Health	except for WTCCC (80%	
rs2812378	CCL21	34,700,260	Individual-SNP	Study (NHS), BRASS,	ACPA+, 84% RF+)	
rs3890745	MMEL1-TNFRSF14	2,585,786	Individual-SNP	NARAC II, NARAC		
rs42041	CDK6	91,891,395	Individual-SNP	III, Genomics		
rs4750316	PRKCQ	6,433,266	Individual-SNP	Collaborative Initiative		
rs1678542	KIF5A-PIP4K2C	56,254,982	Individual-SNP	(GCI), Leiden University Medical Center (LUMC), EIRA-II, Genetics Network Rheumatology Amsterdam (GENRA)		
rs13031237	REL	60,908,994	Individual-SNP	Canada and USA	~95% ACPA+	[73]
rs13017599		60,937,196	Individual-SNP	(European descent)		
rs231735	CTLA4	203,829,153	Individual-SNP			
rs2736340	BLK	11,486,464	Individual-SNP			
rs11586238	CD2-CD58	116,720,516	Individual-SNP	Using GRAIL (Gene		[74]
rs1980422	CD28	203,745,673	Individual-SNP	Relationships Across		
rs548234	PRDM1	106,120,159	Individual-SNP	Implicated Loci)		
rs394581	TAGAP	159,061,489	Individual-SNP			
rs10919563	PTPRC	198,731,313	Individual-SNP			
rs540386	TRAF6-RAG1	36,503,743	Individual-SNP			
rs12746613	FCGR2A	161,497,252	Individual-SNP			
rs3890745	MMEL1-TNFRSF14	2,585,786	Individual-SNP	European, African		[75]
rs2872507	IKZF3-ORMDL3-GSDMB	39,884,510	Individual-SNP	American, Japanese, and Korean ethnicities		
rs10892279	DDX6	118,741,072	Individual-SNP	GWAS Meta-Analysis		[76]
rs864537	CD247	167,442,147	Individual-SNP			
rs2298428	UBE2L3	21,628,603	Individual-SNP			
rs11203203	UBASH3A	42,416,077	Individual-SNP			
rs653178	SH2B3	111,569,952	Individual-SNP			
rs975730	8q24.2	128,303,768	Individual-SNP			
rs1953126	TRAF1	120,878,222	Individual-SNP			
rs7574865	STAT4	191,099,907	Individual-SNP			
rs1051295	KCNB1	49,372,368	Individual-SNP	NARAC		[77]
rs3087243	CTLA4	203,874,196	Individual-SNP	European Descent	GPA	[78]
rs11900673	B3GNT2	62,225,526	Individual-SNP	Japanese	81.4% ACPA+, 80.4% RF+	[80]
rs2867461	ANXA3	78,592,061	Individual-SNP			
rs657075	CSF2	132,094,425	Individual-SNP			
rs12529514	CD83	14,096,427	Individual-SNP			
rs2233434	NFKBIE	44,265,183	Individual-SNP			
rs10821944	ARID5B	62,025,330	Individual-SNP			
rs3781913	PDE2A-ARAP1	72,662,452	Individual-SNP			
rs2841277	PLD4	104,924,668	Individual-SNP			
rs2847297	PTPN2	12,797,695	Individual-SNP			

(*continued on next page*)

Table 1 (*continued*)

SNP	Gene	Position	Method	Population	Comment	Reference
rs2476601	PTPN22	113,834,946	Individual-SNP	Non-hispanic white children	ACPA + or RF+	[81]
rs7574865	STAT4	191,099,907	Individual-SNP			
rs10499194	TNFAIP3	137,681,500	Individual-SNP			
rs4264222	Dectin-2	8,459,172	Individual-SNP	Caucasian (Spain and Portugal)		[83]
rs4804803	DC-SIGN	7,747,847	Individual-SNP			
rs1024611	MCP-1	34,252,769	Individual-SNP			
rs13900		34,256,892	Individual-SNP			
rs4586		34,256,250	Individual-SNP			
rs6859219	IL6ST	56,142,753	Individual-SNP	BRASS, CANADA, EIRA, NARAC I, NARAC III, WTCCC, CANADA II, Dutch, GENRA, GCI, LUMC, NARAC II, United Kingdom RA Genetics (UKRAG), and NHS	ACPA + or RF+	[85]
rs934734	SPRED2	65,368,452	Individual-SNP			
rs26232	C5orf30	103,261,019	Individual-SNP			
rs874040	RBPJ	26,106,575	Individual-SNP			
rs3093023	CCR6	167,120,802	Individual-SNP			
rs10488631	IRF5	128,954,129	Individual-SNP			
rs13315591	PXK	58,571,114	Individual-SNP			
rs706778	IL2RA	6,056,986	Individual-SNP			
rs951005	CCL21	34,743,684	Individual-SNP			
rs11676922	AFF3	100,190,478	Individual-SNP			
rs34536443	TYK2	10,352,442	Individual-SNP	UK, EIRA, US, Dutch, Swedish Umea, Spanish, BRASS, CANADA, NARAC II, and WTCCC		[87]
rs13397	IRAK1	153,982,797	Individual-SNP			
rs8026898	TLE3	69,699,078	Individual-SNP			
rs8043085	RASGRP1	38,535,939	Individual-SNP			
rs2240336	PADI4	17,347,907	Individual-SNP			
rs2228145	IL6R	154,454,494	Individual-SNP			
rs13330176	IRF8	85,985,481	Individual-SNP			
rs12764378	ARID5B	62,040,245	Individual-SNP			
rs9979383	RUNX1	35,343,463	Individual-SNP			
rs12936409	IKZF3	39,887,396	Individual-SNP			
rs2872507		39,884,510	Individual-SNP			
rs883220	POU3F1	38,151,199	Individual-SNP			
rs2834512	RCAN1	34,539,301	Individual-SNP			
rs595158	CD5	61,142,109	Individual-SNP			
rs2275806	GATA3	8,053,377	Individual-SNP			
rs2240340	PADI4	17,336,144	Individual-SNP	Malaysian Epidemiological Investigation of RA (MyEIRA)		[89]
rs1005753	PADI2	17,118,274	Individual-SNP			
rs1801131	MTHFR	11,854,476	Individual-SNP	Jewish and north Italians		[90,91]
rs699738	CD2	116,768,525	Haplotype Block	UK, EIRA, US, Dutch, Swedish Umea, Spanish, BRASS, CANADA, NARAC II, and WTCCC	ACPA +	[101]
rs624988		116,721,168				
rs798036		116,766,208				
rs11203366	PADI4	17,331,039	Haplotype Block	Japanese	75% RF+	[94]
rs11203367		17,331,121				
rs874881		17,334,004				
rs1748033		17,336,167				
rs17878703	HLA–DRB1	32,584,360	Haplotype Block	BRASS, CANADA, EIRA, NARAC I, NARAC III, WTCCC	ACPA+	[99]
rs13195291	ZNF193	28,201,463	Haplotype Block	Canada, USA (European Ancestry)		[66]
rs35656932		28,223,510				
rs13204012		28,233,753				
rs17720293	ZNF307	28,246,920				
rs13208096	NKAPL	28,257,533				
rs67998226	ZNF187	28,270,281				
rs6920220	TNFAIP3-OLIG3	137,685,367	Haplotype Block	BRASS, FHS, NARAC I, EIRA	ACPA +	[67]
rs10499194		137,681,500				
rs3761847	TRAF1	120,927,961	Haplotype Block	NARAC I, NARAC II, EIRA I, EIRA II	ACPA +	[96]
rs2900180	TRAF1-C5	120,944,104				
rs2239657	TRAF1	120,909,242	LD			

Table 2 Studies that agree/disagree with other studies.

Gene	Confirmed in	Not detected in	References
CTLA4	North Americans, European and Asian ancestries	Swedish	[70,73,93]
TAGAP	European ancestry	African Americans	[65,74,86]
FCGR2A	Europeans	Taiwanese, Europeans or Asians	[74,102,103]
MMEL1-TNFRSF14	European ancestry	–	[72,75]
TRAF1	European ancestry	–	[76,104]
STAT4	European ancestry	–	[76,105]
PADI4	North Americans and Asians	Swedish	[70,89,94]
7q	British females	North Americans	[84,106]
CCR6	Europeans	African Americans	[85,86]
TNFAIP3			
MTHFR	Jewish and Italians	Americans (Africans and Caucasians)	[90–92]

haplotype analysis. They detected missense SNPs (rs798036, rs699738) and a noncoding SNP (rs624988) in *CD2* which had the best signal of association in the conditional haplotype analysis.

IL2R consists of IL2Rα (encoded by *IL2RA*), IL2Rβ (encoded by *IL2RB*), and the common gamma chain (CD132). The holding down of IL2Rα or IL2Rβ in mice leads to destructive autoimmunity. *CD2* encodes CD2 protein which is a cell-surface antigen located on T cells. The activation of regulatory T cells (through $CD4^+$ and $IL2R\alpha^+$) coactivates CD2 leading to the suppression of T cells. The regulatory T cells deal with proinflammatory processes. The regulatory T cells are functionally compromised in RA patients.

Table 1 summarized the detected SNPs associated with RA. Table 1 showed the SNP ID, the related gene, the SNP position in the genome, the used method for identification, and the studied population. The used methods for the detection of the associated SNPs were individual-SNP methods, LD, and haplotype block methods.

Studies that agree/disagree with others

This section shows the studies which agree/disagree with others. The study, carried out by Gregersen et al. [73], was in line with that conducted by Plenge et al. [70] verifying the association of *CTLA4* with RA in North Americans. But the study carried out by Plenge et al. [70] did not detect the association of *CTLA4* with RA in Swedish population, while the study carried out by Okada et al. [93] confirmed the association of *CTLA4* with RA in European and Asian ancestries. The study, conducted by Chen et al. [65], supported the results obtained by Raychaudhuri et al. [74] for the association of *TAGAP* with RA susceptibility in European ancestry. However, *TAGAP* did not show any association with RA in African Americans through Hughes et al. [86] study. The results, obtained by Raychaudhuri et al. [74], contradicted the findings of Chen et al. [102] for the association of *FCGR2A* with RA. This contradiction might be due to the different population ancestries (Europeans and Taiwanese) in the two studies. Also, the very small sample size used in Chen et al. study [102] compared with that in Raychaudhuri et al. study [74] might be the reason. Controversially, the findings of Lee et al. [103] supported the findings of Chen et al. [102] for the lack of association of *FCGR2A* with RA in Europeans or Asians.

The study, conducted by Kurreeman et al. [75], agreed with the study carried out by Raychaudhuri et al. [72] for the association of *MMEL1-TNFRSF14* with RA in European ancestry. The findings of Raychaudhuri et al. [72] and Okada et al. [80] supported that RA and graves' disease might have a shared mechanism of disease pathogenesis (*B3GNT2* (rs11900673), *ARID5B* (rs10821944), and *CD40* (rs1883832)). The study, carried out by Zhernakova et al. [76], confirmed the results of Han et al. [104] for the association of *TRAF1* with RA in European ancestry. The findings of both studies (Daha et al. [105] and Zhernakova et al. [76]) confirmed the association of *STAT4* with RA in European ancestry.

The *MTHFR* was confirmed for association with RA susceptibility in Jewish and Italian populations [90,91]. However, the study conducted by Hughes et al. [92] showed contradictory results in American Africans and Caucasians. The negative findings found in the Americans might be due to the enrichment of the flour products in the US with folic acid since 1998 [91]. The *PADI4* was confirmed as an RA biomarker in the Asian populations through Too et al. [89] and Suzuki et al. [94] studies. Also, Plenge et al. [70] detected the association of *PADI4* with RA in North Americans but not in Swedish population. The 7q region showed contradictory results for association with RA in British (females) and North American populations. These findings were found in studies performed by the WTCCC [106] and Korman et al. [84] respectively. The study, conducted by Hughes et al. [86], was not in line with the study of Stahl et al. [85] for the association of *CCR6* and *TNFAIP3* with RA in African Americans and Europeans respectively.

Table 2, summarizing the above findings, showed the studies that confirm/contradict one another in the association of certain genes with RA in different ethnicities. RA biomarkers have been increased through the last decade and still increasing. To conclude, the journey to uncover all RA biomarkers seems to be endless unless large scale trans-ethnic studies take place.

Conclusions

RA is an autoimmune disease that is considerably spread all over the world. Researchers believe that RA has genetic and environmental causes for attacking the body joints. SNPs play a vital role in shedding the light on genes and biological

pathways that contribute to RA. Reviewing the literature, *HLA-DRB1* seems to be the most successful candidate for the title of RA universal biomarker depending on the large number of studies providing its association with RA patients worldwide, and even with JRA.

Most of the applied strategies on the discovery of RA biomarkers are the individual SNP approaches. Later on, the LD mapping techniques are used to detect the correlation among the studied SNP and the neighboring SNPs to identify the causal SNPs. Recently, GWAS have facilitated the empirical study of large dataset of SNPs. Then, the haplotype block methods were introduced to detect the association of RA with a whole block instead of a SNP.

Future work should concentrate on the unstudied populations, comparison among different populations, RF and CCP status, disease severity, gender related genes, JRA biomarkers, disease outcomes, response to therapies, and shared mechanism of disease pathogenesis. Extensive work in these areas should lead to understanding the etiology of RA, identification of further biological pathways, new RA drug discovery, and personal treatment of RA patients.

Conflict of interest

The authors have declared no conflict of interest.

Compliance with Ethics Requirements

This article does not contain any studies with human or animal subjects.

References

[1] Tsokos GC, Virella G. Rheumatoid arthritis, part III: clinical immunology. In: Virella G, editor. Medical immunology. New York: Informa Healthcare; 2007. p. 273.

[2] Mackie S, Quinn M, Emery P. Rheumatoid arthritis, section II: multisystem diseases. In: Rose NR, Mackay IR, editors. The autoimmune diseases. Saint Louis, Missouri: Elsevier Academic Press; 2006. p. 417–8.

[3] Isaacs JD. The changing face of rheumatoid arthritis: sustained remission for all? Nat Rev Immunol 2010;10(8):605–11.

[4] Ruderman E, Tambar S. Rheumatoid arthritis [Internet]. Atlanta, Georgia: American College of Rheumatology; 2012 August [cited 2013 October 30]. <http://www.rheumatology.org/Pratice/Clinical/Patients/Diseases_And_Conditions/Rheumatoid_Arthritis/ra.pdf>.

[5] Biomarkers in rheumatoid arthritis (RA) (white paper) [Internet]. United Kingdom: Quotient Bioresearch; 2010 February [cited 2013 March 20]. <http://landing.quotientbioresearch.com/Portals/24413/docs/quotient%20biomarkers%20in%20rheumatoid%20arthritis%20feb%2010.pdf>.

[6] Nass SJ, Moses HL. Cancer biomarkers: the promises and challenges of improving detection and treatment. Washington, D.C.: National Academies Press; 2007. p. 01.

[7] Fareed M, Afzal M. Single nucleotide polymorphism in genome-wide association of human population: a tool for broad spectrum service. Egypt J Med Hum Genet 2013;14(2):123–34.

[8] Liang Y, Kelemen A. Sequential support vector regression with embedded entropy for SNP selection and disease classification. Stat Anal Data Min 2011;4(3):301–12.

[9] Ruyssen-Witrand A, Constantin A, Cambon-Thomsen A, Thomsen M. New insights into the genetics of immune responses in rheumatoid arthritis. Tissue Antigens 2012;80(2):105–18.

[10] Stahl E, Raychaudhuri S, Plenge R. Polygenic inheritance of rheumatoid arthritis (RA) risk, California, Stanford University, Pharmacogenomics Research Network (PGRN), Research in Progress Seminars (RIPS); 2011. p. 66.

[11] Croucher PJP. Linkage disequilibrium. eLS: John Wiley & Sons, Ltd.; 2013.

[12] Lin W-Y, Schaid DJ. Identifying single-nucleotide polymorphisms responsible for the linkage signal of rheumatoid arthritis on chromosome 6 by joint modeling of linkage and association. BMC Proc 2007;1(Suppl 1):S40.

[13] Sun X, Namkung J, Zhu X, Elston RC. Capability of common SNPs to tag rare variants. BMC Proc 2011;5(Suppl 9):S88.

[14] Schaub MA, Boyle AP, Kundaje A, Batzoglou S, Snyder M. Linking disease associations with regulatory information in the human genome. Genome Res 2012;22(9):1748–59.

[15] Kitajima H, Sonoda M, Yamamoto K. HLA and SNP haplotype mapping in the Japanese population. Genes Immun 2012;13(7):543–8.

[16] Halldórsson BV, Bafna V, Edwards N, Lippert R, Yooseph S, Istrail S. Combinatorial problems arising in SNP and haplotype analysis. In: Calude CS, Dinneen MJ, Vajnovszki V, editors. Discrete mathematics and theoretical computer science. Lecture notes in computer science, 2731. Dijon, France, Berlin Heidelberg: Springer; 2003. p. 26–47.

[17] Gupta A, Manuch J, Stacho L, Zhao X. Algorithm for haplotype inference via galled-tree networks with simple galls. J Comput Biol 2012;19(4):439–54.

[18] Wall JD, Pritchard JK. Haplotype blocks and linkage disequilibrium in the human genome. Nat Rev Genet 2003;4(8):587–97.

[19] Clark AG. The role of haplotypes in candidate gene studies. Genet Epidemiol 2004;27(4):321–33.

[20] Gibbs RA, Belmont JW, Hardenbol P, Willis TD, Yu F, Yang H, et al. The international HapMap project. Nature 2003;426(6968):789–96.

[21] Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, et al. A map of human genome variation from population-scale sequencing. Nature 2010;467(7319):1061–73.

[22] Wang N, Akey JM, Zhang K, Chakraborty R, Jin L. Distribution of recombination crossovers and the origin of haplotype blocks: the interplay of population history, recombination, and mutation. Am J Hum Genet 2002;71(5):1227–34.

[23] Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, et al. The structure of haplotype blocks in the human genome. Science 2002;296(5576):2225–9.

[24] Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 2005;21(2):263–5.

[25] Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES. High-resolution haplotype structure in the human genome. Nat Genet 2001;29(2):229–32.

[26] Kimmel G, Shamir R. A block-free hidden Markov model for genotypes and its application to disease association. J Comput Biol 2005;12(10):1243–60.

[27] Zhang K, Deng M, Chen T, Waterman MS, Sun F. A dynamic programming algorithm for haplotype block partitioning. Proc Natl Acad Sci USA 2002;99(11):7335–9.

[28] Zhang K, Qin ZS, Liu JS, Chen T, Waterman MS, Sun F. Haplotype block partitioning and tag SNP selection using genotype data and their applications to association studies. Genome Res 2004;14(5):908–16.

[29] Zahiri J, Mahdevar G, Nowzari-Dalini A, Ahrabian H, Sadeghi M. A novel efficient dynamic programming

algorithm for haplotype block partitioning. J Theor Biol 2010;267(2):164–70.
[30] Pugach I, Matveyev R, Wollstein A, Kayser M, Stoneking M. Dating the age of admixture via wavelet transform analysis of genome-wide data. Genome Biol 2011;12(2):R19.
[31] Patil N, Berno AJ, Hinds DA, Barrett WA, Doshi JM, Hacker CR, et al. Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. Science 2001;294(5547):1719–23.
[32] Koivisto M, Perola M, Varilo T, Hennah W, Ekelund J, Lukk M, et al. An MDL method for finding haplotype blocks and for estimating the strength of haplotype block boundaries. Pac Symp Biocomput 2003;8:502–13.
[33] Anderson EC, Novembre J. Finding haplotype block boundaries by using the minimum-description-length principle. Am J Hum Genet 2003;73(2):336–54.
[34] Su SC, Kuo CC, Chen T. Inference of missing SNPs and information quantity measurements for haplotype blocks. Bioinformatics 2005;21(9):2001–7.
[35] Su SC, Kuo CC, Chen T. Single nucleotide polymorphism data analysis – state-of-the-art review on this emerging field from a signal processing viewpoint. IEEE Signal Process Mag 2007;24(1):75–82.
[36] Shim H, Chun H, Engelman CD, Payseur BA. Genome-wide association studies using single-nucleotide polymorphisms versus haplotypes: an empirical comparison with data from the North American Rheumatoid Arthritis Consortium. BMC Proc 2009;3(Suppl 7):S35.
[37] Kallberg H, Padyukov L, Plenge RM, Ronnelid J, Gregersen PK, van der Helm-van Mil AH. Gene-gene and gene-environment interactions involving HLA-DRB1, PTPN22, and smoking in two subsets of rheumatoid arthritis. Am J Hum Genet 2007;80(5):867–75.
[38] Perricone C, Ceccarelli F, Valesini G. An overview on the genetic of rheumatoid arthritis: a never-ending story. Autoimmun Rev 2011;10(10):599–608.
[39] Newton JL, Harney SM, Wordsworth BP, Brown MA. A review of the MHC genetics of rheumatoid arthritis. Genes Immun 2004;5(3):151–7.
[40] Ding B, Padyukov L, Lundstrom E, Seielstad M, Plenge RM, Oksenberg JR, et al. Different patterns of associations with anti-citrullinated protein antibody-positive and anti-citrullinated protein antibody-negative rheumatoid arthritis in the extended major histocompatibility complex region. Arthritis Rheum 2009;60(1):30–8.
[41] Muazzam AG, Mansoor A, Ali L, Siddiqi S, Hameed A, Ajmal M, et al. Association of HLA-DRB1 and -DQB1alleles and haplotypes with rheumatoid arthritis in a Pakistani population. Arthritis Res Ther 2013;15(4):R95.
[42] Atouf O, Benbouazza K, Brick C, Bzami F, Bennani N, Amine B, et al. HLA polymorphism and early rheumatoid arthritis in the Moroccan population. Joint Bone Spine 2008;75(5):554–8.
[43] Al-Swailem R, Al-Rayes H, Sobki S, Arfin M, Tariq M. HLA-DRB1 association in Saudi rheumatoid arthritis patients. Rheumatol Int 2006;26(11):1019–24.
[44] Ucar F, Karkucak M, Alemdaroglu E, Capkin E, Yucel B, Sonmez M, et al. HLA-DRB1 allele distribution and its relation to rheumatoid arthritis in eastern Black Sea Turkish population. Rheumatol Int 2012;32(4):1003–7.
[45] Mourad J, Monem F. HLA-DRB1 allele association with rheumatoid arthritis susceptibility and severity in Syria. Rev Bras Reumatol 2013;53(1):47–56.
[46] Ben Hamad M, Mahfoudh N, Marzouk S, Kammoun A, Gaddour L, Hakim F, et al. Association study of human leukocyte antigen-DRB1 alleles with rheumatoid arthritis in south Tunisian patients. Clin Rheumatol 2012;31(6):937–42.
[47] Barnetche T, Constantin A, Cantagrel A, Cambon-Thomsen A, Gourraud PA. New classification of HLA-DRB1 alleles in rheumatoid arthritis susceptibility: a combined analysis of worldwide samples. Arthritis Res Ther 2008;10(1):R26.
[48] Usnayo MJ, Andrade LE, Alarcon RT, Oliveira JC, Silva GM, Bendet I, et al. Study of the frequency of HLA-DRB1 alleles in Brazilian patients with rheumatoid arthritis. Rev Bras Reumatol 2011;51(5):474–83.
[49] Singwe-Ngandeu M, Finckh A, Bas S, Tiercy JM, Gabay C. Diagnostic value of anti-cyclic citrullinated peptides and association with HLA-DRB1 shared epitope alleles in African rheumatoid arthritis patients. Arthritis Res Ther 2010;12(2):R36.
[50] El-Gabalawy HS, Robinson DB, Daha NA, Oen KG, Smolik I, Elias B, et al. Non-HLA genes modulate the risk of rheumatoid arthritis associated with HLA-DRB1 in a susceptible North American Native population. Genes Immun 2011;12(7):568–74.
[51] Yang M, Kuang X, Li J, Pan Y, Tan M, Lu B, et al. Meta-analysis of the association of HLA-DRB1 with rheumatoid arthritis in Chinese populations. BMC Musculoskelet Disord 2013;14:307.
[52] Pedersen M, Jacobsen S, Garred P, Madsen HO, Klarlund M, Svejgaard A, et al. Strong combined gene-environment effects in anti-cyclic citrullinated peptide-positive rheumatoid arthritis: a nationwide case-control study in Denmark. Arthritis Rheum 2007;56(5):1446–53.
[53] van der Woude D, Lie BA, Lundstrom E, Balsa A, Feitsma AL, Houwing-Duistermaat JJ, et al. Protection against anti-citrullinated protein antibody-positive rheumatoid arthritis is predominantly associated with HLA-DRB1*1301: a meta-analysis of HLA-DRB1 associations with anti-citrullinated protein antibody-positive and anti-citrullinated protein antibody-negative rheumatoid arthritis in four European populations. Arthritis Rheum 2010;62(5):1236–45.
[54] Balandraud N, Picard C, Reviron D, Landais C, Toussirot E, Lambert N, et al. HLA-DRB1 genotypes and the risk of developing anti citrullinated protein antibody (ACPA) positive rheumatoid arthritis. PLoS One 2013;8(5):e64108.
[55] de Vries N, Ronningen KS, Tilanus MG, Bouwens-Rombouts A, Segal R, Egeland T, et al. HLA-DR1 and rheumatoid arthritis in Israeli Jews: sequencing reveals that DRB1*0102 is the predominant HLA-DR1 subtype. Tissue Antigens 1993;41(1):26–30.
[56] Salvarani C, Macchioni PL, Mantovani W, Bragliani M, Collina E, Cremonesi T, et al. HLA-DRB1 alleles associated with rheumatoid arthritis in Northern Italy: correlation with disease severity. Br J Rheumatol 1998;37(2):165–9.
[57] Wakitani S, Murata N, Toda Y, Ogawa R, Kaneshige T, Nishimura Y, et al. The relationship between HLA-DRB1 alleles and disease subsets of rheumatoid arthritis in Japanese. Br J Rheumatol 1997;36(6):630–6.
[58] Lee HS, Lee KW, Song GG, Kim HA, Kim SY, Bae SC. Increased susceptibility to rheumatoid arthritis in Koreans heterozygous for HLA-DRB1*0405 and *0901. Arthritis Rheum 2004;50(11):3468–75.
[59] Zuniga J, Yu N, Barquera R, Alosco S, Ohashi M, Lebedeva T, et al. HLA class I and class II conserved extended haplotypes and their fragments or blocks in Mexicans: implications for the study of genetic diversity in admixed populations. PLoS One 2013;8(9):e74442.
[60] Lombard Z, Brune AE, Hoal EG, Babb C, Van Helden PD, Epplen JT, et al. HLA class II disease associations in southern Africa. Tissue Antigens 2006;67(2):97–110.
[61] Batmaz I, Sariyildiz MA, Dilek B, Yildiz I, Nas K, Ayyildiz O, et al. HLA-DRB1 alleles distribution in patients with rheumatoid arthritis in a tertiary center in the southeastern Anatolia region of Turkey. Turk J Phys Med Rehab 2013;59(2):123–7.
[62] Farouk HM, Mansour HE, Rahman SA, Mostafa AA, Shamy HA, Zarouk WA. Effect of the human leukocyte antigen HLA-

DRB1 and anti-cyclic citrullinated peptide on the outcome of rheumatoid arthritis patients. Braz J Med Biol Res 2009;42(9):831–8.
[63] Al-Timimi DJ, Rasool MT, Sulaiman DM. HLA-DR/DQ genotypes in Kurd patients with rheumatoid arthritis: relation to disease activity. J Clin Diagn Res 2014;8(5):CC01–4.
[64] Lee HS, Lee AT, Criswell LA, Seldin MF, Amos CI, Carulli JP, et al. Several regions in the major histocompatibility complex confer risk for anti-CCP-antibody positive rheumatoid arthritis, independent of the DRB1 locus. Mol Med 2008;14(5–6):293–300.
[65] Chen R, Stahl EA, Kurreeman FA, Gregersen PK, Siminovitch KA, Worthington J, et al. Fine mapping the TAGAP risk locus in rheumatoid arthritis. Genes Immun 2011;12(4):314–8.
[66] Xie G, Lu Y, Sun Y, Zhang SS, Keystone EC, Gregersen PK, et al. Identification of the NF-kappaB activating protein-like locus as a risk locus for rheumatoid arthritis. Ann Rheum Dis 2013;72(7):1249–54.
[67] Plenge RM, Cotsapas C, Davies L, Price AL, de Bakker PI, Maller J, et al. Two independent alleles at 6q23 associated with risk of rheumatoid arthritis. Nat Genet 2007;39(12):1477–82.
[68] Kurko J, Besenyei T, Laki J, Glant TT, Mikecz K, Szekanecz Z. Genetics of rheumatoid arthritis – a comprehensive review. Clin Rev Allergy Immunol 2013;45(2):170–9.
[69] Coenen MJ, Gregersen PK. Rheumatoid arthritis: a view of the current genetic landscape. Genes Immun 2009;10(2):101–11.
[70] Plenge RM, Padyukov L, Remmers EF, Purcell S, Lee AT, Karlson EW, et al. Replication of putative candidate-gene associations with rheumatoid arthritis in >4000 samples from North America and Sweden: association of susceptibility with PTPN22, CTLA4, and PADI4. Am J Hum Genet 2005;77(6):1044–60.
[71] Peters AL, Plenge RM, Graham RR, Altshuler DM, Moser KL, Gaffney PM, et al. A novel polymorphism of the human CD40 receptor with enhanced function. Blood 2008;112(5):1863–71.
[72] Raychaudhuri S, Remmers EF, Lee AT, Hackett R, Guiducci C, Burtt NP, et al. Common variants at CD40 and other loci confer risk of rheumatoid arthritis. Nat Genet 2008;40(10):1216–23.
[73] Gregersen PK, Amos CI, Lee AT, Lu Y, Remmers EF, Kastner DL, et al. REL, encoding a member of the NF-kappaB family of transcription factors, is a newly defined risk locus for rheumatoid arthritis. Nat Genet 2009;41(7):820–3.
[74] Raychaudhuri S, Thomson BP, Remmers EF, Eyre S, Hinks A, Guiducci C, et al. Genetic variants at CD28, PRDM1 and CD2/CD58 are associated with rheumatoid arthritis risk. Nat Genet 2009;41(12):1313–8.
[75] Kurreeman FA, Stahl EA, Okada Y, Liao K, Diogo D, Raychaudhuri S, et al. Use of a multiethnic approach to identify rheumatoid- arthritis-susceptibility loci, 1p36 and 17q12. Am J Hum Genet 2012;90(3):524–32.
[76] Zhernakova A, Stahl EA, Trynka G, Raychaudhuri S, Festen EA, Franke L, et al. Meta-analysis of genome-wide association studies in celiac disease and rheumatoid arthritis identifies fourteen non-HLA shared loci. PLoS Genet 2011;7(2):e1002004.
[77] Xiao X, Zhang Y, Wang K. Association of KCNB1 to rheumatoid arthritis via interaction with HLA-DRB1. BMC Proc 2009;3(Suppl 7):S134.
[78] Chung SA, Xie G, Roshandel D, Sherva R, Edberg JC, Kravitz M, et al. Meta-analysis of genetic polymorphisms in granulomatosis with polyangiitis (Wegener's) reveals shared susceptibility loci with rheumatoid arthritis. Arthritis Rheum 2012;64(10):3463–71.
[79] Li Y, Begovich AB. Unraveling the genetics of complex diseases: susceptibility genes for rheumatoid arthritis and psoriasis. Semin Immunol 2009;21(6):318–27.
[80] Okada Y, Terao C, Ikari K, Kochi Y, Ohmura K, Suzuki A, et al. Meta-analysis identifies nine new loci associated with rheumatoid arthritis in the Japanese population. Nat Genet 2012;44(5):511–6.
[81] Prahalad S, Conneely KN, Jiang Y, Sudman M, Wallace CA, Brown MR, et al. Susceptibility to childhood-onset rheumatoid arthritis: investigation of a weighted genetic risk score that integrates cumulative effects of variants at five genetic loci. Arthritis Rheum 2013;65(6):1663–7.
[82] Ezzat MH, Awwad KS, Shaheen KY, Abd El-Wehab SM, Osman TI. The influence of HLA-DRB1 alleles on polyarticular onset juvenile rheumatoid arthritis susceptibility, severity, and prognosis. Egypt J Med Hum Genet 2005;6(2):145–71.
[83] Caliz R, Canet LM, Lupianez CB, Canhao H, Escudero A, Filipescu I, et al. Gender-specific effects of genetic variants within Th1 and Th17 cell-mediated immune response genes on the risk of developing rheumatoid arthritis. PLoS One 2013;8(8):e72732.
[84] Korman BD, Seldin MF, Taylor KE, Le JM, Lee AT, Plenge RM, et al. The chromosome 7q region association with rheumatoid arthritis in females in a British population is not replicated in a North American case-control series. Arthritis Rheum 2009;60(1):47–52.
[85] Stahl EA, Raychaudhuri S, Remmers EF, Xie G, Eyre S, Thomson BP, et al. Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. Nat Genet 2010;42(6):508–14.
[86] Hughes LB, Reynolds RJ, Brown EE, Kelley JM, Thomson B, Conn DL, et al. Most common single-nucleotide polymorphisms associated with rheumatoid arthritis in persons of European ancestry confer risk of rheumatoid arthritis in African Americans. Arthritis Rheum 2010;62(12):3547–53.
[87] Eyre S, Bowes J, Diogo D, Lee A, Barton A, Martin P, et al. High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. Nat Genet 2012;44(12):1336–40.
[88] Viatte S, Flynn E, Lunt M, Barnes J, Singwe-Ngandeu M, Bas S, et al. Investigation of Caucasian rheumatoid arthritis susceptibility loci in African patients with the same disease. Arthritis Res Ther 2012;14(6):R239.
[89] Too CL, Murad S, Dhaliwal JS, Larsson P, Jiang X, Ding B, et al. Polymorphisms in peptidylarginine deiminase associate with rheumatoid arthritis in diverse Asian populations: evidence from MyEIRA study and meta-analysis. Arthritis Res Ther 2012;14(6):R250.
[90] Berkun Y, Levartovsky D, Rubinow A, Orbach H, Aamar S, Grenader T, et al. Methotrexate related adverse effects in patients with rheumatoid arthritis are associated with the A1298C polymorphism of the MTHFR gene. Ann Rheum Dis 2004;63(10):1227–31.
[91] Rubini M, Padovan M, Baricordi O, Fotinidi M, Govoni M, Trotta F. The c.1298A>C polymorphism in the methylenetetrahydrofolate reductase gene is associated with rheumatoid arthritis susceptibility in Italian patients. Clin Exp Rheumatol 2008;26(1):163.
[92] Hughes LB, Beasley TM, Patel H, Tiwari HK, Morgan SL, Baggott JE, et al. Racial or ethnic differences in allele frequencies of single-nucleotide polymorphisms in the methylenetetrahydrofolate reductase gene and their influence on response to methotrexate in rheumatoid arthritis. Ann Rheum Dis 2006;65(9):1213–8.
[93] Okada Y, Wu D, Trynka G, Raj T, Terao C, Ikari K, et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature 2014;506(7488):376–81.
[94] Suzuki A, Yamada R, Chang X, Tokuhiro S, Sawada T, Suzuki M, et al. Functional haplotypes of PADI4, encoding citrullinating enzyme peptidylarginine deiminase 4, are

associated with rheumatoid arthritis. Nat Genet 2003;34(4):395–402.
[95] Ikari K, Momohara S, Inoue E, Tomatsu T, Hara M, Yamanaka H, et al. Haplotype analysis revealed no association between the PTPN22 gene and RA in a Japanese population. Rheumatology (Oxford) 2006;45(11):1345–8.
[96] Plenge RM, Seielstad M, Padyukov L, Lee AT, Remmers EF, Ding B, et al. TRAF1-C5 as a risk locus for rheumatoid arthritis – a genomewide study. N Engl J Med 2007;357(12):1199–209.
[97] Scherer HU, van der Linden MP, Kurreeman FA, Stoeken-Rijsbergen G, Cessie S, Huizinga TW, et al. Association of the 6q23 region with the rate of joint destruction in rheumatoid arthritis. Ann Rheum Dis 2010;69(3):567–70.
[98] Zhang R-J, Jiang Y-S, Liu G-Y, Wang Z, Li X, Chen Z-Q, et al. A systematic method based on haplotype analysis: application to risk alleles and genes mining for RA. In: Proceedings of the 4th international conference on natural computation (ICNC '08). Jinan, China: IEEE Computer Society; 2008. p. 39–43.
[99] Raychaudhuri S, Sandor C, Stahl EA, Freudenberg J, Lee HS, Jia X, et al. Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. Nat Genet 2012;44(3):291–6.
[100] Park J, Namkung J, Jhun M, Park T. Genome-wide analysis of haplotype interaction for the data from the North American Rheumatoid Arthritis Consortium. BMC Proc 2009;3(Suppl 7):S34.
[101] Diogo D, Kurreeman F, Stahl EA, Liao KP, Gupta N, Greenberg JD, et al. Rare, low-frequency, and common variants in the protein-coding sequence of biological candidate genes from GWASs contribute to risk of rheumatoid arthritis. Am J Hum Genet 2013;92(1):15–27.
[102] Chen JY, Wang CM, Wu JM, Ho HH, Luo SF. Association of rheumatoid factor production with FcgammaRIIIa polymorphism in Taiwanese rheumatoid arthritis. Clin Exp Immunol 2006;144(1):10–6.
[103] Lee YH, Ji JD, Song GG. Associations between FCGR3A polymorphisms and susceptibility to rheumatoid arthritis: a metaanalysis. J Rheumatol 2008;35(11):2129–35.
[104] Han TU, Bang SY, Kang C, Bae SC. TRAF1 polymorphisms associated with rheumatoid arthritis susceptibility in Asians and in Caucasians. Arthritis Rheum 2009;60(9):2577–84.
[105] Daha NA, Kurreeman FA, Marques RB, Stoeken-Rijsbergen G, Verduijn W, Huizinga TW, et al. Confirmation of STAT4, IL2/IL21, and CTLA4 polymorphisms in rheumatoid arthritis. Arthritis Rheum 2009;60(5):1255–60.
[106] Burton PR, Clayton DG, Cardon LR, Craddock N, Deloukas P, Duncanson A, et al. Genome-wide association study of 14,000 cases of seven common diseases and 3000 shared controls. Nature 2007;447(7145):661–78.