Undifferentiated and Rheumatoid Arthritis

Arthritis is a group of conditions characterized by inflammation of the joints. This inflammation can lead to breakdown of the cartilage of the joints that can be caused by different mechanisms, e.g. autoimmunity, fractures, wearing, or infection. The different types of arthritis are diagnosed according to disease criteria, leaving cases that cannot be classified in one of the accepted categories of rheumatic diseases (usually referred to as ‘undifferentiated arthritis’ (UA)). The diagnosis of rheumatoid arthritis (RA), an inflammatory autoimmune disorder characterized by a chronic inflammation of the synovial tissue of several joints, is based on a list of seven criteria developed by the American College of Rheumatology (1). These criteria include clinical, radiological and laboratory findings; i.e. morning stiffness, arthritis of three or more joint areas, arthritis of hand joints, symmetric arthritis, serum rheumatoid factor, rheumatoid nodules, and radiographic changes. The RA patient population is clinically heterogeneous since only four of these seven ACR criteria have to be fulfilled for the diagnosis of RA. The occurrence of RA varies among countries and areas over the world, but has a prevalence of approximately 1% in Europe (2;3). In the Dutch population, women are affected by RA approximately two times more frequently than men (4).

In the Leiden Early Arthritis Clinic (EAC), which provides an inception cohort of patients with recent onset arthritis (5), 37% of the patients are diagnosed with UA and about 20% with RA at their first visit. After 1 year, 32% of the UA patients have qualified for the diagnosis of RA, indicating the complexity of a diagnosis at initial presentation.

RA patients can develop different kind of autoantibodies, amongst others against citrullinated proteins (anti-citrullinated protein antibodies (ACPA)). Citrullination is a post-translational conversion (deimination) of Arginine to Citrulline residues performed by the enzyme peptidylarginine deiminase (PAD) (Figure 1) that results in a small change in molecular mass (<1 Da) and the loss of one positive charge. Although citrullination is a common natural process, these ACPA are specific for RA, and can be measured already years before symptomatic disease (6;7). Recently, it has been shown that ACPA⁺ and ACPA⁻ RA patients show a different disease course, probably indicating that ACPA⁺ and ACPA⁻ RA reflect a totally different disease (8-10).
Progression and Severity scoring

RA is characterized by the proliferation of synovium and the destruction of cartilage and bone during the progression of the disease. Destruction of the cartilage is a consequence of pro-inflammatory cytokines and enzymes that are released during the chronic inflammation process inducing enhanced breakdown of cartilage matrix and reduced synthesis of matrix components by the articular chondrocytes (11;12), but joint erosion results more directly from osteoclasts (13). The formation of pannus, which results from the proliferating synovium (14-16), will eventually lead to joint space narrowing and joint erosions. This is, at least in part, mediated by fibroblast-like synoviocytes from the synovium (17).

Radiographic joint damage is an important outcome measure in RA, in addition to assessments of physical function and disease activity, which all associate with each other (18). It reflects cumulative disease activity and is related to overall disability (19;20). Therefore, progression rates are also influenced by the current therapy, i.e. disease modifying anti-rheumatic drugs (DMARDs) (21;22). Several scoring methods for the assessment of radiographic joint damage exist, from which the most well known are the Larsen (23) and Sharp-van der Heijde method (24;25). Both methods score the individual joints of the extremities but the Sharp-van der Heijde method scores hands and feet for the amount and severity of erosions and joint space narrowing separately (25). The Sharp-van der Heijde method is sensitive to detect changes over time and shows reliable results since it has a low measurement error (26). Analyses of radiographic progression can be analyzed on the individual patient or a patient group level. For both applies that the radiological progression is linear in the first

![Figure 1. Citrullination. An Arginine residue is enzymatically converted by peptidylarginine deiminase (PAD) into a Citrulline residue in the presence of Ca^{2+}.](image-url)
(approximately) five years but further in time, the curve levels off to a plateau (18;27-29).

Because nowadays, RA patients are seen in an earlier phase of the disease, before the appearance of well established indicators of poor prognosis such as erosions and nodules, markers which have a good predictive value on radiographic damage in an early phase of the disease will become more important.

**Genetic Risk Factors for RA**

The pathogenesis of RA is, as in many other autoimmune diseases, complex and largely unknown. It is generally accepted that both genetic and environmental factors contribute and probably also interact with each other. It has been described that genetic factors contribute for about 2/3 to the development of RA (30;31). The contributing risk factors can differ for the susceptibility to, and the progression of RA.

The strongest genetic risk factor, both for susceptibility and severity, has been mapped to the HLA-class II region, most probably DRB1. HLA-class II molecules consist of an α- and β-chain which both have constant and variable regions. The variable regions constitute the binding groove for the peptide to be presented by HLA molecules to T cells of the immune system (Figure 2). A particular part of the binding groove, the third hypervariable region, is involved in the susceptibility to RA development. At position 70 to 74 in this third hypervariable region, different variants of amino acid sequences are present. Certain HLA-DRB1 alleles share common epitopes at this position. Regarding the risk for RA development, three variants can be discriminated; either amino acids of the so-called shared epitope (SE), the sequence “DERAA”, or ‘neutral’ amino acids are present. Compared to the ‘neutral’ HLA-DRB1 alleles, carrienship of HLA-DRB1 alleles with the SE increases the risk for RA development, and “DERAA”-containing HLA-DRB1 alleles decrease the risk. Both the SE and the “DERAA”-containing HLA-DRB1 allele effects will be discussed in more detail below.
Shared Epitope and ACPA

HLA-DRB1 molecules containing the amino acid sequence R(Q)K(R)RAA (i.e. the amino acids Arginine, (Glutamine), Lysine, (Arginine), Arginine, Alanine, Alanine) at position 70-74 in the third hypervariable region are belonging to the Shared Epitope (SE) alleles. This epitope is present in the HLA-DRB1*0101, *0102, *0401, *0404, *0405, *0408, *0410, *1001 and *1402 alleles. The SE is associated with an increased risk (about 2.5 times (33)) to develop RA and a more severe disease course. It is postulated that this SE sequence, which is present in the binding cleft of the HLA-DR molecule, is directly related to the binding of RA inducing peptides to the SE-containing HLA-DR molecules. These peptides are then presented to T cells thought to play an important role in the pathogenesis of RA. Since there is high linkage disequilibrium between HLA-DRB1 and HLA-DQB1 alleles, it is also hypothesized that the HLA-DQB1 alleles are associated with RA susceptibility, although these associations cannot be distinguished (34-36). RA-inducing peptides are not identified yet, but several findings indicate new directions for epitope discovery. Recently it has been shown that SE-containing HLA-DRB1 alleles do not confer risk to the development of RA itself, but predispose to the development of anti-citrullinated protein antibodies (ACPA). These antibodies are highly predictive for RA development as discussed in a previous section (6;7). ACPA+ UA Patients have approximately

Figure 2. HLA class II molecule (adapted from Boots et al. (32)). The variable regions of the α- and β-chain build up the peptide binding groove. The rectangle indicated in the figure shows the position of the third hypervariable region (HVR) where the shared epitope (SE) or “DERAA”-sequence can be present.
fifteen times higher odds to develop RA within one year compared to ACPA negative (ACPAnegative) patients. These ACPA are commonly measured in the IgG isotype, but are also present in the IgM and IgA isotype (37).

The presence of IgG ACPA indicates the presence of T cell help. One of the proteins described to be citrullinated in vivo and present in the synovial fluid of RA patients is the cytoskeletal protein vimentin (38;39). We have recently shown that 90% of ACPA positive (ACPAPositive) RA patients recognizing a citrullinated peptide derived from vimentin carry one or two SE-containing HLA-DRB1 alleles (40), suggesting the involvement of helper T cells recognizing a citrullinated epitope from vimentin in the context of the SE-containing HLA-DRB1 alleles. The identification of citrullinated vimentin-derived T cell epitopes recognized by HLA-DR4 positive individuals is described in Chapter 4 of this thesis.

“DERAA”

“DERAA” stands for the amino acid sequence of Aspartic acid-Glutamic acid-Arginine-Alanine-Alanine which is present in the HLA-DRB1 alleles of the subtype HLA-DRB1*0103, *0402, *1003, *1301, *1302 and *1304. These alleles are present in 29% of the population (33;41-43). It has been shown by several groups that the frequency of “DERAA”-containing HLA-DRB1 alleles is reduced in RA patients compared to healthy controls and that the risk to develop RA is reduced by about 40% in DERAA positive individuals (33). Since the “DERAA” sequence is positioned at the same location as the SE in the HLA-DRB1 molecule, the effect of “DERAA” has to be evaluated after stratification for presence of the SE-containing HLA-DRB1 alleles. In this way, it has been shown that the effect of “DERAA”-containing HLA-DRB1 alleles is independent of the SE-containing HLA-DRB1 alleles (33).

“DERAA”-containing HLA-DRB1 alleles can be inherited, but can also be conferred as non-inherited maternal antigens (NIMA). NIMA can be conferred from the mother during and/or shortly after the pregnancy since the immune systems of mother and child are in close contact and cell trafficking will occur (44-47). The phenomenon of NIMA was described for the first time in 1954 for the RhD antigen (48) and is illustrated in Figure 3. The terminology is oriented from the point of view of the child in a family, since most studies are coming from the transplantation field. It has been described that haplo-identical NIMA-mismatched sibling transplants have a graft survival similar to that of HLA-identical siblings in contrast to NIPA-mismatched siblings, indicating tolerance for the HLA-mismatch from the mother (49-51).
The phenomenon of protection against RA by “DERAA”-containing HLA-DRB1 alleles as NIMA is studied in Chapter 2 of this thesis and discussed in comparison with the inherited effect in Chapter 3.

Next to HLA-DRB1 there are several other genes involved in the risk for RA development. Below, only the additional risk alleles studied in the context of this thesis are discussed.

**PTPN22**

The protein Tyrosine phosphatase named Lyp is encoded by the protein Tyrosine phosphatase, non-receptor type 22 (PTPN22) gene and is expressed by many cell types present in haematopoietic tissues, like T cells, B cells, NK cells, monocytes, dendritic cells and neutrophils (52).

Genes can differ in their nucleotide sequence between different individuals in a population. When the frequency of a single nucleotide change is equal or higher than 1%, this is called a single nucleotide polymorphism (SNP). The most studied SNP of the PTPN22 gene, the C1858T missense single-nucleotide polymorphism, is associated with RA, UA (52-56) and several other autoimmune diseases (57-59). Upto now, it seems that the SNP does not have an effect on the severity of RA, only on the susceptibility (53;60).

There is a large variation in allele frequency among different ethnic populations among the world of the T-variant of the allele, with a variation only in Europe from 2-

---

**Figure 3.** Terminology of non-inherited maternal antigen (NIMA) and non-inherited paternal antigen (NIPA). The terminology is orientated from the point of view of the child. ♀ gender can be male or female.
3% in Italians to 15% in Finnish people (61). There are several articles describing the association of the PTPN22 T allele of the C1858T polymorphism with the development of ACPA⁺ RA, therefore implicating a correlation of this SNP with the production of ACPA (62-64).

We investigated in **Chapter 5** whether the C1858T polymorphism of the PTPN22 gene can give additive value to the prediction of progression from UA to RA when it is combined with presence of ACPA.

**Genome wide association studies (GWAS)**

In the past few years several genome wide association studies (GWAS) have been performed to scan the entire genome for common polymorphisms associating with different autoimmune diseases, including RA (65-68). Since thousands of patients and controls from different populations are studied in the GWAS, these studies may identify risk factors with modest effects on the risk for RA development, and also allow one to study subpopulations of patients for specific effects.

From all genes identified to be associated with RA, we studied based on a recent GWAS (67) six newly identified SNPs for their influence on the severity of RA (**Chapter 6**). These SNPs are located around genes that are either involved in activation of the immune response (CD40, TNFRSF14, CCL21 and PRKCQ) or play a role in intracellular processes involving cell cycle and homeostatis (MMEL1, KIF5A and CDK6).

**Statistical modelling**

Several statistical models may be applied to the analysis of genetic associations. The kind of statistical model appropriate for the analysis is dependent on the correlation between the variables and measurement groups, the distribution of the data and the type of study performed. Below, two models are discussed that are used for the different genetic association chapters described in this thesis.

**Theory of Bayes**

To study the effect of numerous factors on the development of a disease, modelling of the risk factors studied is performed to fit the observed data with the expected frequencies. Modelling of all these risk factors preferably results in prediction of the outcome. A theory that calculates the probability that a hypothesis is true based on the available information is the theory of Bayes, which is used for the studies performed in
Chapter 2 and 5 of this thesis. With this theory, a prior probability (e.g. to develop RA), based on the risk of an individual in a certain population, is converted into a posterior probability, calculated on the basis of extra information derived from the risk factors studied. The prior probability is most often the prevalence of e.g. a certain disease (or symptom) in the population (69-71).

Linear Mixed Model

A mixed model is a multiple variance analysis often used to compare groups of individuals including multiple measurements from one individual e.g. in time. The data of every individual are plotted in a linear way, and combined for the whole group, thereby interpolating missing values (72). In a mixed model both random and fixed effects are included. Fixed effects are categorical variables from which all levels are fixed and known, whereas random effects are variables where only a random sample of possible values is measured. A Gaussian distribution is assumed for the variables studied. Since all individual measurements are taken into account in one analysis, it is a powerful method with low intra-individual variability and therefore smaller group sizes are required (73). A linear mixed model was used in Chapter 6 of this thesis to study the effect of different SNPs on the severity of RA.

Outline of the thesis

Chapter 2 reports a family study in which we studied whether “DERAA”-containing HLA-DRB1 alleles protect against the development of rheumatoid arthritis not only when the alleles are inherited but also when they are present as a non-inherited maternal antigen (NIMA).

Chapter 3 is a review that summarizes the epidemiological findings about “DERAA”-containing HLA-DRB1 alleles and RA and the observation we made in Chapter 2. Both associations are compared and possible explanations for the protective phenomenon are described.

On the same location in the HLA-DRB1 molecule as the “DERAA” epitope, the so-called shared epitope (SE) can be present. These SE-containing HLA-DRB1 alleles predispose to the formation of ACPA and therefore the development of RA. Since ACPA-producing B cells are helped by T cells, we studied in Chapter 4 whether we
could identify possible T cell epitopes from an important ACPA target antigen, namely the citrullinated vimentin protein.

Both for ACPA and the C1858T polymorphism of the PTPN22 gene associations with RA development have been described. We studied the individual contribution of ACPA and a PTPN22 SNP on the prediction of RA development from UA in Chapter 5. We also studied the influence of the PTPN22 SNP on the level of ACPA.

Genes can be involved both in the susceptibility and severity of RA. This is the case e.g. for the SE-containing HLA-DRB1 alleles. Therefore, we studied in Chapter 6 the contribution of several SNPs that were newly identified as risk factors for the development of RA and the severity of the disease.

The results obtained in the chapters 2-6 of this thesis are summarized and discussed in Chapter 7.
References

(7) van der Helm-van Mil AH, Verpoort KN, Breedveld FC, Huizinga TW, Toes RE, de Vries RR. The HLA-DRB1 shared epitope alleles are primarily a risk factor for anti-cyclic citrullinated peptide antibodies and are not an independent risk factor for development of rheumatoid arthritis. Arthritis Rheum 2006; 54(4):1117-1121.


Naseem H, Thomson W, Silman A, Worthington J, Symmons D, Barton A. The PTPN22*C1858T functional polymorphism is associated with susceptibility to inflammatory polyarthritis but neither this nor other variants spanning the gene is associated with disease outcome. Ann Rheum Dis 2008; 67(2):251-255.


Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007; 447(7145):661-668.


