RHEUMATOLOGY

Rheumatology 2016;55:199–209
doi:10.1093/rheumatology/keu323
Advance Access publication 18 September 2014

RA: from risk factors and pathogenesis to prevention

The genetics of rheumatoid arthritis: risk and protection in different stages of the evolution of RA

Annie Yarwood[1], Tom W. J. Huizinga[2] and Jane Worthington[1,3]

Abstract

There is now a general consensus that RA has a spectrum of disease stages that can begin many years before the onset of clinical symptoms. It is widely thought that understanding the complex interplay between genetics and environment, and their role in pathogenesis, is essential in gaining further insight into the mechanisms that drive disease development and progression. More than 100 genetic susceptibility loci have now been identified for RA through studies that have focused on patients with established RA compared with healthy controls. Studying the early preclinical phases of disease will provide valuable insights into the biological events that precede disease and could potentially identify biomarkers to predict disease onset and future therapeutic targets. In this review we will cover recent advances in the knowledge of genetic and environmental risk factors and speculate on how these factors may influence the transition from one stage of disease to another.

Key words: rheumatoid arthritis, genetics, environment, autoantibodies, anti-citrullinated autoantibodies, anti-carbamylated autoantibodies, risk prediction, disease phase, inflammatory polyarthritis, undifferentiated arthritis.

Rheumatology key messages

- Cohorts representing all phases of RA are essential for studies to define risk of progression.
- Prospective studies of individuals with genetic/environmental risk factors will help in understanding preclinical RA.
- ACPA-negative RA patients will be further classified by fine specificities of autoantibodies.

Introduction

The heritability of RA has been shown from twin studies to be ~60%. Since 2007, rapid advances in technology underpinning the use of genome-wide association studies (GWASs) have allowed the identification of hundreds of genetic risk factors for many complex diseases. There are now >100 genetic loci that have been associated with RA (Table 1). The heritability of RA suggests that a large proportion of the disease could be the result of contributing environmental risk factors. It is widely thought that understanding the complex interplay between genetics and environment, and their roles in pathogenesis, is essential to gain further insight into the mechanisms that drive disease development and progression.

There is now a general consensus that RA has a spectrum of disease stages that can begin many years before the onset of clinical symptoms. Data showing the presence of autoantibodies and indicators of activation of the immune system years before disease onset indicate the presence of a long preclinical phase of disease potentially influenced by environmental factors [20, 21]. This preclinical phase results in a continuum that eventually crosses a threshold leading to the manifestation of clinical symptoms and ultimately joint damage. It is hypothesized that genetic markers associated with disease, in combination with stochastic environmental risk factors, influence the transition from one disease stage to another.

[1]Arthritis Research UK Centre for Genetics and Genomics, Centre for Musculoskeletal Research, Institute of Inflammation and Repair, Faculty of Medical and Human Sciences, Manchester Academic Health Science Centre, Stopford Building, University of Manchester, Manchester, UK, [2]Department of Rheumatology, Leiden University Medical Centre, Leiden, The Netherlands and [3]NIHR Manchester Musculoskeletal Biomedical Research Unit, Central Manchester NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK

Submitted 3 April 2014; revised version accepted 20 June 2014

Correspondence to: Jane Worthington, Arthritis Research UK Centre for Genetics and Genomics, 2nd floor, Stopford Building, University of Manchester, Oxford Road, Manchester M13 9PT, UK.
E-mail: jane.worthington@manchester.ac.uk

Table 1 Genetic loci associated with susceptibility to RA

Chromosome	SNP	Gene	Reference: association with ACPA-positive RA (f)[a]	Ethnicity	Reference: association with ACPA-negative RA	Reference: association with RA severity	Reference: association with IP (e)	Reference: association in individuals at risk of RA [(a) and (b)]
1	rs2228145	*IL6R*	[1]	Caucasians				
1	rs2240336	*PADI4*	[1]	Caucasians				
1	rs883220	*POU3F1*	[1]	Caucasians				
1	rs798000	*CD2*	[2]	Caucasians				
1	rs2843401	*MMEL1*	[2]	Caucasians				[3]
1	rs2014863	*PTPRC*	[2]	Caucasians				
1	rs2476601	*PTPN22*	[2]	Caucasians	[4]	[5, 6]	[7]	
1	rs10494360	*FCGR2A*	[2]	Caucasians				
1	rs2105325	*loc100506023*	[8]	Caucasians				
1	rs28411352	*MTF1-INPP5B*	[8]	Caucasians				
1	rs227163	*TNFRSF9*	[8]	Asian				
1	rs10430455	*FCRL3*	[9]	Asians				
1	rs6682654	*CD244*	[10]	Asians				
2	rs10209110	*AFF3*	[2]	Caucasians		[5, 6]		
2	rs34695944	*REL*	[2]	Caucasians				
2	rs6546146	*SPRED2*	[2]	Caucasians				
2	rs13426947	*STAT4*	[2]	Caucasians	[4]			
2	rs1980422	*CD28*	[2]	Caucasians				
2	rs11571302	*CTLA4*	[2]	Caucasians				
2	rs11900673	*B3GNT2*	[11]	Asians				
2	rs6732565	*ACOXL*	[8]	Caucasians				
2	rs6715284	*CFLAR-CASP8*	[8]	Caucasians				
2	rs10175798	*LBH*	[8]	Caucasians				
3	rs35677470	*DNASE1L3*	[2]	Caucasians				
3	rs3806624	*EOMES*	[8]	Caucasians				
3	rs9826828	*IL20RB*	[8]	Caucasians				
3	rs4452313	*PLCL2*	[8]	Caucasians				
4	rs78560100	*IL2, IL21*	[2]	Caucasians				
4	rs932036	*RBPJ*	[2]	Caucasians				
4	rs2867461	*ANXA3*	[11]	Asians				
4	rs13142500	*CLNK*	[8]	Caucasians				
4	rs2664035	*TEC*	[8]	Caucasians				
5	rs71624119	*ANKRD55*	[2]	Caucasians	[4]			
5	rs39984	*GIN1*	[2]	Caucasians	[4]			
5	rs657075	*CSF2*	[11]	Asians				
6	rs59466457	*CCR6*	[2]	Caucasians				
6	rs6911690	*PRDM1*	[2]	Caucasians				
6	rs629326	*TAGAP*	[2]	Caucasians		[5, 6]		
6	rs6920220	*TNFAIP3*	[2]	Caucasians	[4]			
6	rs72928038	*BACH2*	[12]	Caucasians				
6	rs12529514	*CD83*	[11]	Asians				
6	rs2233434	*NFKBIE*	[11]	Asians				
6	rs2234067	*ETV7*	[8]	Caucasians				
6	rs9378815	*IRF4*	[8]	Caucasians				
6	rs9373594	*PPIL4*	[8]	Asian				
6	amino acid position 9	*HLA-DPB1*	[13]	Caucasians				
6	amino acid position 9	*HLA-B*	[13]	Caucasians	[14, 15]			
6	amino acid position 74	*HLA-DRB1*	[13]	Caucasians				
6	amino acid position 71	*HLA-DRB1*	[13]	Caucasians				
6	amino acid position 11	*HLA-DRB1*	[13]	Caucasians	[14, 15]			
6	amino acid position 77	*HLA-A*	[15]	Caucasians				
7	rs3807306	*IRF5*	[2]	Caucasians				
7	rs4272	*CDK6*	[8]	Caucasians				
7	rs67250450	*JAZF1*	[8]	Caucasians				
8	rs4840565	*BLK*	[2]	Caucasians	[4]			
8	rs678347	*GRHL2*	[8]	Caucasians				
8	rs1516971	*PVT1*	[8]	Caucasians				
8	rs998731	*TPD52*	[8]	Caucasians				
9	rs10739580	*TRAF1*	[2]	Caucasians		[5, 6]		[3]
9	rs2812378	*CCL21*	[2]	Caucasians				
10	rs12764378	*ARID5B*	[1]	Caucasians				
10	rs2275806	*GATA3*	[2]	Caucasians				
10	rs947474	*PRKCQ*	[2]	Caucasians				
10	rs10795791	*IL2RA*	[2]	Caucasians				
10	rs12413578	*10p14*	[8]	Caucasians				

(continued)

Table 1 Continued

Chromosome	SNP	Gene	Reference: association with ACPA-positive RA (f)[a]	Ethnicity	Reference: association with ACPA-negative RA	Reference: association with RA severity	Reference: association with IP (e)	Reference: association in individuals at risk of RA [(a) and (b)]
10	rs726288	*SFTPD*	[8]	Asian				
10	rs2671692	*WDFY4*	[8]	Caucasians				
10	rs793108	*ZNF438*	[8]	Caucasians				
11	rs595158	*CD5*	[1]	Caucasians				
11	rs4938573	*DDX6*	[1]	Caucasians				
11	rs570676	*TRAF6*	[2]	Caucasians				
11	rs3781913	*PDE2A-ARAP1*	[11]	Asians				
11	chr11:107967350	*ATM*	[8]	Caucasians				
11	rs4409785	*CEP57*	[8]	Caucasians				
11	rs73013527	*ETS1*	[8]	Caucasians				
11	rs968567	*FADS1-FADS2-FADS3*	[8]	Caucasians				
12	rs10683701	*KIF5A*	[2]	Caucasians		[5, 6]		
12	rs773125	*CDK2*	[8]	Caucasians				
12	rs10774624	*SH2B3-PTPN11*	[8]	Caucasians				
13	rs9603616	*COG6*	[8]	Caucasians				
14	rs911263, rs1950897	*RAD51L1/RAD51B*	[12]	Caucasians				
14	rs2841277	*PLD4*	[11]	Asians				
14	rs3783782	*PRKCH*	[8]	Asian				
15	rs8043085	*RASGRP1*	[1]	Caucasians				
15	rs8026898	*TLE3*	[1]	Caucasians				
16	rs13330176	*IRF8*	[1]	Caucasians				
16	rs4780401	*TXNDC11*	[8]	Caucasians				
17	rs12936409	*IKZF3*	[1]	Caucasians				
17	rs72634030	*C1QBP*	[8]	Caucasians				
17	rs1877030	*MED1*	[8]	Caucasians				
18	rs7234029	*PTPN2*	[16]	Caucasians				
18	rs2469434	*CD226*	[8]	Asian				
19	rs34536443	*TYK2*	[1]	Caucasians				
19	chr19:10771941	*ILF3*	[8]	Caucasians				
20	rs6032662	*CD40*	[2]	Caucasians				
21	rs2834512	*RCAN1*	[1]	Caucasians				
21	rs9979383	*RUNX1*	[1]	Caucasians				
21	rs73194058	*IFNGR2*	[8]	Caucasians				
21	rs1893592	*UBASH3A*	[8]	Caucasians				
21	rs2075876	*AIRE*	[17]	Asians				
22	rs3218251	*IL2RB*	[2]	Caucasians				
22	rs909685	*SYNGR1*	[8]	Caucasians				
22	rs1043099	*GATSL3*	[18]	Caucasians				
22	rs5754217	*UBE2L3*	[19]	Caucasians				
X	rs13397	*IRAK1*	[1]	Caucasians				
X	chrX:78464616	*P2RY10*	[8]	Asian				

The table shows genetic loci associated with susceptibility to RA. References are shown for the association of loci with ACPA-positive RA. If loci have also been associated with ACPA-negative RA, RA severity, IP or individuals at risk of disease, the references for these studies are also shown. Ethnicity describes the population in which the association was observed. [a]EULAR disease phases are shown in parentheses. SNP: single nucleotide polymorphism; IP: inflammatory polyarthritis; EULAR: European League Against Rheumatism.

This concept is relatively new, however, and as a result the majority of genetic and environmental studies to date have focused on patients with established RA, according to the 1987 ACR criteria for RA [22], compared with healthy controls, and few studies have used cases in the early stages of disease or even in the preclinical phases of disease onset. In more recent years, research has focused on the earlier stages of disease [20, 21, 23, 24]. This has been helped by the ACR/European League Against Rheumatism (EULAR) classification criteria published in 2010, which include earlier features of the disease, such as autoantibody status [25], although it is only recently that researchers have begun establishing cohorts that will allow investigation of the earliest stages of disease. An initial challenge is definition of the disease phases; thus the EULAR Study Group for Risk Factors for RA recently published recommendations to classify individuals into different stages of disease for recruitment into prospective studies [26].

Prior to the development of the EULAR recommendations, RA patients have typically been split into two groups based on autoantibody status [26]. Evidence from studies of ACPA-positive and negative disease subsets suggest that they are two distinct diseases with different underlying pathogenesis; this is discussed in more detail later. Studying patients who may be at risk of disease due to

genetic and environmental risk factors gives researchers the chance to identify significant biological changes that may occur in the preclinical phase of RA, potentially providing invaluable opportunities for new treatment strategies.

ACPAs are of particular interest, as these autoantibodies are highly specific for RA and can be found in ~50% of early RA patients [27], making ACPA an important early clinical biomarker. In addition, the presence of anti-CCP antibodies has been shown to predict disease severity and radiological damage [28, 29], meaning that ACPA could also be used as a biomarker for patients with a more severe disease phenotype and to select those patients eligible for more aggressive treatment [30].

It is not clear how individuals progress from a pre-RA ACPA-positive state with no symptoms to clinical RA. Not all individuals with ACPA develop RA; therefore other triggers may be required to make the transition.

It has also been shown that the number of citrullinated antigens identified by ACPA increases exponentially leading up to disease onset. This is known as epitope spreading [31, 32]. The expansion of the ACPA repertoire that occurs before the manifestation of clinical disease could potentially be used to predict the impending onset of disease. Although no specific single antigen or target has been identified, it may be that certain antigens contribute to an initial break in immune tolerance, resulting in epitope spreading and an immune response to other antigens. This could then result in the transition from preclinical to clinically apparent disease. In this review we will discuss the recent advances in knowledge of genetic susceptibility to RA and the study of environmental risk factors and will speculate on how these factors may influence the transition from one stage of disease to another. Reference made to EULAR disease phase refers to the EULAR disease classification defined in Gerlag *et al.* [26].

Genetics of seropositive RA [EULAR disease phase (f)]

At least two-thirds of the risk of RA is thought to be conferred by genetic risk factors [33]. The largest genetic risk factor for RA lies within the human leucocyte antigen (HLA) class II region and encodes the HLA-DRB1 molecule [34, 35]. The specific alleles of *HLA-DRB1* that have been associated with RA encode a conserved amino acid sequence that lies in the antigen binding groove of the antigen presenting molecule. This conserved sequence is referred to as the shared epitope (SE) [36]. It is thought that the presence of different alleles in the SE influence the interaction between the class II molecule and the T cell receptor (TCR) or antigen, and could therefore directly affect the efficiency of antigen presentation [36].

A more recent analysis of the HLA region in RA carried out by Raychaudhuri *et al.* [13], showed that three amino acid positions (11, 71 and 74) in HLA-DRB1 and a single amino acid position (position 9) in HLA-B and HLA-DPB1, account for the association of the MHC with anti-CCP positive RA. A key finding of this study was the association of amino acids 11 and 13, both in tight linkage disequilibrium, which lie outside the classical SE region of DRB1 but within the peptide binding grooves. Recent evidence has suggested that the HLA class II locus is not associated with risk of becoming ACPA+, but with the risk of progressing from ACPA+ to ACPA+ RA [37, 38]. This supports our hypothesis that the HLA class II locus is not directly involved in the formation of ACPA responses, but rather in its maturation, possibly via T cells providing help to ACPA-producing B-cells before the precipitation of disease

The largest genetic association with RA outside the HLA region lies within the protein tyrosine phosphatase non-receptor 22 (*PTPN22*) gene [39–44]. With an odds ratio (OR) of ~1.6, *PTPN22* encodes lymphoid tyrosine phosphatase which is a negative regulator of signal transduction from the TCR [45]. Mutations in the *PTPN22* gene are associated with multiple autoimmune diseases and have been shown to confer opposite effects [40, 41, 43–53]. Several studies investigating the function of this variant have had conflicting results with some suggesting it results in a gain of function [54–58]. and others suggesting that the R620W polymorphism actually leads to a loss of function [58, 59].

Since the identification of *PTPN22*, fine mapping of linkage peaks [60], GWAS [11, 18, 61–66] and meta analyses [1, 2] have led to the identification of many more genetic risk factors for RA. There is an emerging picture of shared genetic overlap between autoimmune diseases [67, 68]. A recent large scale fine mapping study called the Immunochip project leveraged this remarkable genetic overlap to its advantage by designing a custom Illumina single nucleotide polymorphism (SNP) genotyping array containing ~200 000 SNPs at 186 loci previously implicated in 12 autoimmune diseases, providing the opportunity for fine mapping of autoimmune disease loci [1]. For RA, the Immunochip study was carried out in 11 475 cases of European descent and 15 870 controls. These data were then combined with previous independent GWAS data (additional 2363 cases, and 17 872 controls) in a meta-analysis. 14 new susceptibility loci were identified; additionally the fine mapping approach allowed the peak of association to be refined to a single gene for 19 loci [1]. A later follow up study of loci reaching suggestive significance in the Immunochip study identified a further two novel loci in independent datasets [12].

The most recent large genetic studies have increased the number of RA susceptibility loci to over 100 across multiple populations (Table 1) [8, 69]. Plenge *et al.* performed a large scale trans-ethnic GWAS Meta analysis of >100 000 individuals of European and Asian ancestry and assessed over 10 million SNPs. Outside of the MHC 100 loci were shown to explain 5.5% and 4.7% of the heritability of RA in Europeans and Asians respectively.

The majority of the genetic loci identified to date confer small effect sizes and therefore a substantial proportion of

the heritability of RA remains undetected. The missing heritability could lie in epigenetic changes (heritable changes in gene activity not related to changes in the DNA sequence), gene-gene or gene-environment interactions, or even in rare variants although recent studies have suggested the contribution of rare variants to complex disease may be negligible [70]. As well as risk variants for RA, it is likely that there may also be protective variants and although some of the identified RA genetic loci confer ORs below one, the protective effect of genetic variants for RA is poorly understood with the exception of specific HLA alleles. A meta-analysis investigating *HLA-DRB1* alleles showed HLA-DRB1*13 to be associated with protection in ACPA-positive RA [71].

Genetic susceptibility to seronegative RA

Genetic studies of RA have largely been carried out in anti CCP positive patients, and many of the RA loci identified to date have stronger effects in this subgroup of patients [1]. Although ACPA negative RA patients make up ~30% of the RA population this subset of patients remains understudied. Many people consider anti-CCP positive and anti-CCP negative RA to be two distinct diseases, however this debate is ongoing [72–74]. The heritability of the two subgroups has been estimated to be similar at 68% and 66% respectively although the contribution of the SE to this heritability was found to be significantly lower in anti-CCP negative disease (18% in anti-CCP positive, 2.4% in anti-CCP negative) [75]

Two recent studies of ACPA negative patients and healthy controls identified a peak of association in the HLA region which was determined to be accounted for by two amino acid positions HLA-DRB1 position 11 and HLA-B position 9 [14, 15]. These two positions have also been shown to be associated with ACPA positive RA; however the specific amino acid residues at each position which confer risk were different between the two disease subsets. Interestingly the presence of a serine residue at HLA-DRB1 position 11 conferred risk of ACPA negative disease but was protective against ACPA positive disease [15]. The study by Raychaudhuri *et al.* also identified a new association at position 77 in HLA-A in ACPA positive patients [15]. These studies increase the evidence that ACPA positive and ACPA negative RA are two distinct diseases.

It is now well known that other ACPA antibodies and fine specificities may contribute to RA, for example Lundberg *et al.* [76] tested four fine specificities of ACPA in RA patients, they observed different associations between genetic and environmental risk factors with RA depending on the ACPA specificity, with the HLA-DRB1 SE, *PTPN22* and smoking showing the strongest association with the RA subset defined by the presence of antibodies to citrullinated α-enolase. Therefore it must be recognised that testing for the presence of anti-CCP antibodies in order to define the ACPA status of a patient does not definitively define an individual as ACPA negative and may result in heterogeneity in so called anti-CCP negative patient cohorts.

Recently it has been identified that breaking tolerance to post-translationally modified proteins in arthritis is not exclusively confined to citrullination. ACPA recognize proteins only after the enzymatic conversion of the amino acid arginine by PAD-enzymes to become the amino acid citrulline. It is likely that proteins that have undergone a different type of posttranslational modification are also recognized by auto-antibodies. One of these other post-translational processes is carbamylation, where the amino acid lysine is changed to become homocitrulline. Smoking can enhance carbamylation [77] and extensive carbamylation is especially thought to occur during (chronic) inflammation [77]. As smoking and chronic inflammation are important in the context of RA it is possible that carbamylation could be taking place in the inflamed synovium.

The post-translationally modified amino acids citrulline and homocitrulline are very similar in structure. The resemblance between the two modifications and the likely presence of carbamylated proteins (CarPs) in the joint was the motivation to test for the presence of antibodies directed against CarPs. Using, a novel assay that specifically detects the presence of antibodies directed against CarPs (anti-CarPs) [78], 43% of RA patients were found to have antibodies directed against CarPs. Importantly these anti-CarP antibodies were not only present in ACPA positive but also in ACPA negative RA patients, suggesting that antibodies recognizing one modification do not necessarily cross-react with the other modification [79]. Anti-CarP antibodies were also associated with a higher rate of joint damage however, further studies are required to determine the contribution of these antibodies to disease. This finding highlights the possibility that seronegative RA does not exist rather that we have as yet failed to identify all possible RA associated autoantibodies. It is not known why some people develop these antibodies and how they may contribute to disease. The question remains if these antibodies could be used to predict progression towards clinical disease in individuals at risk, or if patients should be tested for anti-CarP alongside anti-CCP. As indicated above international replication studies are needed to confirm and expand observations discussed.

Several non HLA RA susceptibility genes identified in anti-CCP positive disease have also been shown to be associated with anti-CCP negative disease (*TNFAIP3*, *GIN1/C5orf30*, *STAT4*, *ANKRD55/IL6ST*, *BLK* and *PTPN22*) [4]. The high heritability of anti-CCP negative RA and the low number of associated loci highlights the need for further large well powered GWAS studies to identify novel seronegative disease specific loci. Further investigation into the overlapping associations will be required in these two subgroups of disease.

Genetics of inflammatory polyarthritis [EULAR disease phase (e)]

It is difficult to speculate on the impact of a locus on transition between disease stages as the mechanism of the majority of RA susceptibility loci has not yet been

investigated. Studying early inflammatory polyarthritis (IP) or patients with undifferentiated arthritis (UA) may help determine the genetic factors that are involved in the pathogenesis and prognosis of IP and UA. Some IP and UA patients will ultimately progress to RA or even PsA, however, some individuals will remain undifferentiated indefinitely. Several disease susceptibility loci have been associated with disease severity (*TRAF1/C5*, *KIF5A*, *PTPN22*, *AFF3*, *TAGAP*) [5, 6], suggesting that these loci may represent a tipping point allowing disease to progress one stage further and that individuals without these loci may maintain mild disease.

There have been a limited number of studies investigating the genetics of IP patients [80–84]. A study of 680 patients with IP from the Norfolk Arthritis Register (NOAR) and 286 controls showed a modest association between the SE alleles of *HLA-DRB1* and IP [85]. The OR (1.8) for the association of the SE alleles with IP was much smaller than that typically observed with RA (OR 2–3).

The modest association of a SNP in the *PTPN22* locus has been shown with UA [86]. A more recent study of this locus showed significant association of the PTPN22*1858T allele with IP and the strength of the effect was similar to that observed in RA [7]. A study in a Dutch population has shown that the PTPN22*1858T allele could not predict progression from UA to RA [87]. The new EULAR recommendations should allow further study of this group of patients.

Studies of individuals at risk of RA [EULAR disease phases (a) and (b)]

The ability to identify individuals in the preclinical phase of RA is challenging. Therefore studying individuals who are at high risk of developing RA in the future due to known risk factors such as family history provides a unique opportunity to study the biological events that precede disease development.

El-Gabalawy *et al*. [3] investigated non-HLA genes for association with RA in a North American Native (NAN) population previously shown to be predisposed to RA due a high prevalence of RA, multicase families and a high background frequency of HLA-DRB1 risk alleles. The authors tested 21 non-HLA SNPs previously associated with RA and showed HLA-DRB1 to be the major genetic risk factor for RA in this population. Additionally, the presence of the SE and the minor allele of *MMEL1-TNFRSF14* significantly reduced RA risk, whereas *TRAF1-C5* increased the risk, showing that additional risk factors outside of the HLA can contribute to disease risk in this predisposed population.

A limited number of studies have been carried out in first-degree relatives (FDRs) of RA patients who could be considered an at-risk population. One study showed that FDRs had a higher prevalence of RA autoantibodies than healthy controls, and individuals who were ACPA positive and rheumatoid factor positive had the highest prevalence of joint symptoms [88]. It is possible that some of these patients, particularly individuals with positive autoantibody status, may develop disease in the next few years [89] and therefore would benefit from symptom monitoring to allow early treatment. A study in the same NAN population showed an increase in the levels of multiple cytokines and CRP in FDRs compared with controls [90]. Other studies in the relatives of RA patients have also identified the presence of ACPAs in healthy relatives [91] and demonstrated that the fine specificity of ACPAs in healthy FDRs and their related RA patient can be different [92].

To gain the true benefit of studying these at-risk populations, prospective cohorts of FDRs will need to be established for long-term follow-up.

Environmental risk factors [EULAR disease phase (b)] and gene–environment interactions in RA

The heritability of RA (50–60%) suggests that a large proportion of disease could be due to environmental risk factors [33, 93]. More studies regarding environmental risk factors are required, although identifying environmental risk factors can be challenging due to recall bias and difficulties distinguishing cause from effect.

However, the identification of environmental risk factors presents an interesting challenge, as prevention strategies based on avoidance of exposure to risk would be attractive. The possibility that RA starts outside the joints raises many questions about the role of environmental factors. Could an environmental influence be responsible for the initiation of autoimmunity? Or could an environmental factor be the trigger to drive the transition to clinical disease? It is likely to be the environmental contribution to disease that interacts with a susceptible individual's genetic component to alter disease course and progression. Understanding the influence of the environment may help us to understand the initial phases of disease and how these can be altered.

Several environmental risk factors for RA have been implicated in the development of disease, although few of these have substantial evidence [94–105]. A recent prospective study identified pack-years of smoking, diabetes mellitus, BMI and parity as risk factors for developing RA or IP, while alcohol, higher social class and breastfeeding were associated with a decreased risk of RA or IP [106], although conflicting results have been reported regarding alcohol [107].

Smoking is the most well-established environmental risk factor for RA [101, 108–110], and several studies have shown a strong interaction between smoking and the alleles of the *HLA-DRB1* SE. Studies in the Swedish Epidemiological Investigation of Rheumatoid Arthritis (EIRA) cohort have shown that individuals who have ever smoked and carry SE alleles are at increased risk of developing RF- or ACPA-positive RA [111–113]. This effect was not observed in ACPA-negative RA. This gene environment interaction was replicated in a Danish case–control study [114]. However, attempts to replicate this interaction in three North American cohorts with RA were unsuccessful and only weak evidence for a gene environment interaction was found in one cohort [North

American Rheumatoid Arthritis Consortium (NARAC)] [115].

Further studies by Klareskog *et al*. [111] have demonstrated the presence of citrullinated proteins in bronchoalveolar lavage cells from smokers, allowing the possibility that citrullination of proteins in the lung caused by smoking may, in the presence of *HLA-DRB1* SE alleles, result in RA-specific immune reactions to citrullinated proteins. However, citrullination has been found throughout the body, so it is unclear how citrullination in the lungs may be important in the development of ACPA.

As smoking has been shown to increase citrullination [111], it is hypothesised that smoking could promote autoimmunity to citrulline residues [116]. Presentation of these citrullinated peptides to T cells would then initiate proliferation, differentiation, cytokine production and the formation of T cell memory against citrullinated peptides.

Predicting disease risk

The identification of genetic and environmental risk factors for RA has raised the question as to whether these risk factors are sufficient to predict individuals who are at risk of disease. Indeed, one of the ultimate aims of genetic research is to predict who will develop disease before symptom onset and joint erosion/destruction begins. However, the modest effects of the loci identified by GWASs have left their predictive ability in question. There have been several studies in recent years that have used genetic risk factors, and in some cases environmental risk factors, to try to predict disease risk [106, 117–123]. In general these models have demonstrated the ability to identify a subset of individuals at high risk with relatively high specificity (80–90%), but with relatively low sensitivity (~40%), resulting in a significant proportion of high-risk individuals being misclassified as not at risk. Stratifying RA into disease stages and profiling the genetics of individuals in each group in order to identify distinct profiles may provide insight into which loci are involved in specific processes and inform targeting of more aggressive treatments to a particular subgroup of patients.

Due to the low prevalence of RA, whole population screening of individuals to identify those at risk is not economical. However, risk prediction may prove more fruitful in targeted subsets of individuals who are classified as at risk due to the presence of a family history of disease or environmental risk factors and is the premise of many primary prevention studies.

It should also be noted that the genetic markers identified to date only account for about half of the heritability of RA. Further studies to account for a larger proportion of the heritability of disease and the addition of further risk factors such as environmental factors, biomarkers, and clinical predictors will almost certainly improve the power of predictive models and may prove more clinically useful.

The genetic susceptibility variants identified to date strongly implicate several immune pathways in the development of RA [1]. In addition to this, there is already significant overlap between the targets of several approved RA therapies and identified susceptibility genes for RA (*CTLA4*: abatacept; *TYK2*: tofacitinib; *IL-6*: tocilizumab) demonstrating the potential of genetics not only to identify biological pathways that lead to RA, but also that these pathways can be targeted and lead to the successful treatment of disease symptoms.

Acknowledgements

A.Y. is funded by the Innovative Medicines Initiative Joint Undertaking (IMI JU)-funded project BeTheCure (contract no. 115142-2).

Funding: This manuscript was also funded by a core programme grant from Arthritis Research UK (17552). We acknowledge support from the National Institute of Health Research (NIHR) Manchester Musculoskeletal Biomedical Research Unit.

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

References

1 Eyre S, Bowes J, Diogo D *et al*. High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. Nat Genet 2012;44:1336–40.

2 Stahl EA, Raychaudhuri S, Remmers EF *et al*. Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. Nat Genet 2010;42:508–14.

3 El-Gabalawy HS, Robinson DB, Daha NA *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;12:568–74.

4 Viatte S, Plant D, Bowes J *et al*. Genetic markers of rheumatoid arthritis susceptibility in anti-citrullinated peptide antibody negative patients. Ann Rheum Dis 2012; 71:1984–90.

5 Plant D, Thomson W, Lunt M *et al*. The role of rheumatoid arthritis genetic susceptibility markers in the prediction of erosive disease in patients with early inflammatory polyarthritis: results from the Norfolk Arthritis Register. Rheumatology 2011;50:78–84.

6 Viatte S, Plant D, Lunt M *et al*. Investigation of rheumatoid arthritis genetic susceptibility markers in the early rheumatoid arthritis study further replicates the TRAF1 association with radiological damage. J Rheumatol 2013; 40:144–56.

7 Naseem H, Thomson W, Silman A *et al*. 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; 67:251–5.

8 Okada Y, Wu D, Trynka G *et al*. Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature 2014;506:376–81.

9 Kochi Y, Yamada R, Suzuki A *et al*. A functional variant in FCRL3, encoding Fc receptor-like 3, is associated with rheumatoid arthritis and several autoimmunities. Nat Genet 2005;37:478–85.

10 Suzuki A, Yamada R, Kochi Y *et al*. Functional SNPs in CD244 increase the risk of rheumatoid arthritis in a Japanese population. Nat Genet 2008;40:1224–9.

11 Okada Y, Terao C, Ikari K *et al*. Meta-analysis identifies nine new loci associated with rheumatoid arthritis in the Japanese population. Nat Genet 2012;44:511–6.

12 McAllister K, Yarwood A, Bowes J *et al*. Identification of BACH2 and RAD51B as rheumatoid arthritis susceptibility loci in a meta-analysis of genome-wide data. Arthritis Rheum 2013;65:3058–62.

13 Raychaudhuri S, Sandor C, Stahl EA *et al*. Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. Nat Genet 2012;44:291–6.

14 Bossini-Castillo L, de Kovel C, Kallberg H *et al*. A genome-wide association study of rheumatoid arthritis without antibodies against citrullinated peptides. Ann Rheum Dis 2014 Feb 14. pii: annrheumdis-2013-204591. doi: 10.1136/annrheumdis-2013-204591. [Epub ahead of print].

15 Han B, Diogo D, Eyre S *et al*. Fine mapping seronegative and seropositive rheumatoid arthritis to shared and distinct HLA alleles by adjusting for the effects of heterogeneity. Am J Hum Genet 2014 Apr 3;94(4):522-32. doi: 10.1016/j.ajhg.2014.02.013. [Epub 20 March 2014].

16 Cobb JE, Plant D, Flynn E *et al*. Identification of the tyrosine-protein phosphatase non-receptor type 2 as a rheumatoid arthritis susceptibility locus in Europeans. PLoS One 2013;8:e66456.

17 Terao C, Yamada R, Ohmura K *et al*. The human AIRE gene at chromosome 21q22 is a genetic determinant for the predisposition to rheumatoid arthritis in Japanese population. Hum Mol Genet 2011;20:2680–5.

18 Orozco G, Viatte S, Bowes J *et al*. Novel rheumatoid arthritis susceptibility locus at 22q12 identified in an extended UK genome-wide association study. Arthritis Rheumatol 2014;66:24–30.

19 Orozco G, Eyre S, Hinks A *et al*. Study of the common genetic background for rheumatoid arthritis and systemic lupus erythematosus. Ann Rheum Dis 2011;70:463–8.

20 Nielen MM, van Schaardenburg D, Reesink HW *et al*. Specific autoantibodies precede the symptoms of rheumatoid arthritis: a study of serial measurements in blood donors. Arthritis Rheum 2004;50:380–6.

21 Rantapaa-Dahlqvist S, de Jong BAW, Berglin E *et al*. Antibodies against cyclic citrullinated peptide and IgA rheumatoid factor predict the development of rheumatoid arthritis. Arthritis Rheum 2003;48:2741–9.

22 Arnett FC, Edworthy SM, Bloch DA *et al*. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum 1988;31:315–24.

23 de Hair MJ, van de Sande MG, Ramwadhdoebe TH *et al*. Features of the synovium of individuals at risk of developing rheumatoid arthritis: implications for understanding preclinical rheumatoid arthritis. Arthritis Rheumatol 2014;66:513–22.

24 Jorgensen KT, Wiik A, Pedersen M *et al*. Cytokines, autoantibodies and viral antibodies in premorbid and postdiagnostic sera from patients with rheumatoid arthritis: case-control study nested in a cohort of Norwegian blood donors. Ann Rheum Dis 2008;67:860–6.

25 Aletaha D, Neogi T, Silman AJ *et al*. 2010 rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Ann Rheum Dis 2010;69:1580–8.

26 Gerlag DM, Raza K, van Baarsen LG *et al*. EULAR recommendations for terminology and research in individuals at risk of rheumatoid arthritis: report from the Study Group for Risk Factors for Rheumatoid Arthritis. Ann Rheum Dis 2012;71:638–41.

27 van der Linden MP, van der WD, Ioan-Facsinay A *et al*. Value of anti-modified citrullinated vimentin and third-generation anti-cyclic citrullinated peptide compared with second-generation anti-cyclic citrullinated peptide and rheumatoid factor in predicting disease outcome in undifferentiated arthritis and rheumatoid arthritis. Arthritis Rheum 2009;60:2232–41.

28 Berglin E, Johansson T, Sundin U *et al*. Radiological outcome in rheumatoid arthritis is predicted by presence of antibodies against cyclic citrullinated peptide before and at disease onset, and by IgA-RF at disease onset. Ann Rheum Dis 2006;65:453–8.

29 Bukhari M, Thomson W, Naseem H *et al*. The performance of anti-cyclic citrullinated peptide antibodies in predicting the severity of radiologic damage in inflammatory polyarthritis: results from the Norfolk Arthritis Register. Arthritis Rheum 2007;56:2929–35.

30 van der Helm-van Mil A, Verpoort K, Breedveld F, Toes R, Huizinga T. Antibodies to citrullinated proteins and differences in clinical progression of rheumatoid arthritis. Arthritis Res Ther 2005;7:R949–R958.

31 van de Stadt LA, van der Horst AR, de Koning MH *et al*. The extent of the anti-citrullinated protein antibody repertoire is associated with arthritis development in patients with seropositive arthralgia. Ann Rheum Dis 2011;70:128–33.

32 van der Woude D, Rantapaa-Dahlqvist S, Ioan-Facsinay A *et al*. Epitope spreading of the anti-citrullinated protein antibody response occurs before disease onset and is associated with the disease course of early arthritis. Ann Rheum Dis 2010;69:1554–61.

33 MacGregor AJ, Snieder H, Rigby AS *et al*. Characterizing the quantitative genetic contribution the rheumatoid arthritis using data from twins. Arthritis Rheum 2000;43:30–7.

34 Gourraud P-A, Dieude P, Boyer J-F *et al*. A new classification of HLA-DRB1 alleles differentiates predisposing and protective alleles for autoantibody production in rheumatoid arthritis. Arthritis Res Ther 2007;9:R27.

35 Stastny P. Mixed lymphocyte cultures in rheumatoid arthritis. J Clin Invest 1976;57:1148–57.

36 Gregersen PK, Silver J, Winchester RJ. The shared epitope hypothesis: an approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis. Arthritis Rheum 1987;30:1205–13.

37 Haj HA, Magnusson PK, Joshua V *et al*. Environmental and genetic factors in the development of anticitrullinated

protein antibodies (ACPAs) and ACPA-positive rheumatoid arthritis: an epidemiological investigation in twins. Ann Rheum Dis 2013 Nov 25. doi: 10.1136/annrheumdis-2013-203947. [Epub ahead of print].

38 Bos WH, Ursum J, de Vries N *et al*. The role of the shared epitope in arthralgia with anti-cyclic citrullinated peptide antibodies (anti-CCP), and its effect on anti-CCP levels. Ann Rheum Dis 2008;67:1347–50.

39 Begovich AB, Carlton VE, Honigberg LA *et al*. A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (PTPN22) is associated with rheumatoid arthritis. Am J Hum Genet 2004;75:330–7.

40 Criswell LA, Pfeiffer KA, Lum RF *et al*. Analysis of families in the multiple autoimmune disease genetics consortium (MADGC) collection: the PTPN22 620W allele associates with multiple autoimmune phenotypes. Am J Hum Genet 2005;76:561–71.

41 Hinks A, Barton A, John S *et al*. Association between the PTPN22 gene and rheumatoid arthritis and juvenile idiopathic arthritis in a UK population. Arthritis Rheum 2005; 52:1694–9.

42 Lee AT, Li W, Liew A *et al*. The PTPN22 R620W polymorphism associates with RF positive rheumatoid arthritis in a dose-dependent manner but not with HLA-SE status. Genes Immun 2005;6:129–33.

43 Orozco G, Sanchez E, Gonzalez-Gay MA *et al*. Association of a functional single-nucleotide polymorphism of PTPN22, encoding lymphoid protein phosphatase, with rheumatoid arthritis and systemic lupus erythematosus. Arthritis Rheum 2005;52:219–24.

44 Viken MK, Amundsen SS, Kvien TK *et al*. Association analysis of the 1858C>T polymorphism in the PTPN22 gene in juvenile idiopathic arthritis and other autoimmune diseases. Genes Immun 2005;6:271–3.

45 Bottini N, Musumeci L, Alonso A *et al*. A functional variant of lymphoid tyrosine phosphatase is associated with type I diabetes. Nat Genet 2004;36:337–8.

46 Canton I, Akhtar S, Gavalas NG *et al*. A single-nucleotide polymorphism in the gene encoding lymphoid protein tyrosine phosphatase (PTPN22) confers susceptibility to generalised vitiligo. Genes Immun 2005;6:584–7.

47 Dieude P, Guedj M, Wipff J *et al*. The PTPN22 620W allele confers susceptibility to systemic sclerosis: findings of a large case-control study of European Caucasians and a meta-analysis. Arthritis Rheum 2008;58: 2183–8.

48 Roycroft M, Fichna M, McDonald D *et al*. The tryptophan 620 allele of the lymphoid tyrosine phosphatase (PTPN22) gene predisposes to autoimmune Addison's disease. Clin Endocrinol 2009;70:358–62.

49 Skorka A, Bednarczuk T, Bar-Andziak E, Nauman J, Ploski R. Lymphoid tyrosine phosphatase (PTPN22/LYP) variant and Graves' disease in a Polish population: association and gene dose-dependent correlation with age of onset. Clin Endocrinol 2005;62:679–82.

50 Vandiedonck C, Capdevielle C, Giraud M *et al*. Association of the PTPN22*R620W polymorphism with autoimmune myasthenia gravis. Ann Neurol 2006;59:404–7.

51 Velaga MR, Wilson V, Jennings CE *et al*. The codon 620 tryptophan allele of the lymphoid tyrosine phosphatase (LYP) gene is a major determinant of Graves' disease. J Clin Endocrinol Metab 2004;89:5862–5.

52 Wang K, Baldassano R, Zhang H *et al*. Comparative genetic analysis of inflammatory bowel disease and type 1 diabetes implicates multiple loci with opposite effects. Hum Mol Genet 2010;19:2059–67.

53 Zhernakova A, Eerligh P, Wijmenga C *et al*. Differential association of the PTPN22 coding variant with autoimmune diseases in a Dutch population. Genes Immun 2005; 6:459–61.

54 Fiorillo E, Orru V, Stanford SM *et al*. Autoimmune-associated PTPN22 R620W variation reduces phosphorylation of lymphoid phosphatase on an inhibitory tyrosine residue. J Biol Chem 2010;285:26506–18.

55 Rieck M, Arechiga A, Onengut-Gumuscu S *et al*. Genetic variation in PTPN22 corresponds to altered function of T and B lymphocytes. J Immunol 2007;179:4704–10.

56 Smyth D, Cooper JD, Collins JE *et al*. Replication of an association between the lymphoid tyrosine phosphatase locus (LYP/PTPN22) with type 1 diabetes, and evidence for its role as a general autoimmunity locus. Diabetes 2004;53:3020–3.

57 Vang T, Congia M, Macis MD *et al*. Autoimmune-associated lymphoid tyrosine phosphatase is a gain-of-function variant. Nat Genet 2005;37:1317–9.

58 Wu DJ, Zhou W, Enouz S *et al*. Autoimmunity-associated LYP-W620 does not impair thymic negative selection of autoreactive T cells. PLoS One 2014;9:e86677.

59 Zhang J, Zahir N, Jiang Q *et al*. The autoimmune disease-associated PTPN22 variant promotes calpain-mediated Lyp/Pep degradation associated with lymphocyte and dendritic cell hyperresponsiveness. Nat Genet 2011;43: 902–7.

60 Remmers EF, Plenge RM, Lee AT *et al*. STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus. N Engl J Med 2007;357:977–86.

61 Gregersen PK, Amos CI, Lee AT *et al*. REL, encoding a member of the NF-κB family of transcription factors, is a newly defined risk locus for rheumatoid arthritis. Nat Genet 2009;41:820–3.

62 Julia A, Ballina J, Canete JD *et al*. Genome-wide association study of rheumatoid arthritis in the Spanish population: KLF12 as a risk locus for rheumatoid arthritis susceptibility. Arthritis Rheum 2008;58: 2275–86.

63 Kochi Y, Okada Y, Suzuki A *et al*. A regulatory variant in CCR6 is associated with rheumatoid arthritis susceptibility. Nat Genet 2010;42:515–9.

64 Plenge RM, Seielstad M, Padyukov L *et al*. TRAF1-C5 as a risk locus for rheumatoid arthritis—a genomewide study. N Engl J Med 2007;357:1199–209.

65 Plenge RM, Cotsapas C, Davies L *et al*. Two independent alleles at 6q23 associated with risk of rheumatoid arthritis. Nat Genet 2007;39:1477–82.

66 Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007;447: 661–78.

67 Eyre S, Hinks A, Bowes J *et al*. Overlapping genetic susceptibility variants between three autoimmune disorders:

rheumatoid arthritis, type 1 diabetes and coeliac disease. Arthritis Res Ther 2010;12:R175.

68 Richard-Miceli C, Criswell LA. Emerging patterns of genetic overlap across autoimmune disorders. Genome Med 2012;4:6.

69 Kim K, Bang SY, Lee HS *et al*. High-density genotyping of immune loci in Koreans and Europeans identifies eight new rheumatoid arthritis risk loci. Ann Rheum Dis 2014 Feb 14. doi: 10.1136/annrheumdis-2013-204749. [Epub ahead of print].

70 Hunt KA, Mistry V, Bockett NA *et al*. Negligible impact of rare autoimmune-locus coding-region variants on missing heritability. Nature 2013;498:232–5.

71 van der Woude D, Lie BA, Lundstrom E *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;62:1236–45.

72 van der Helm-van Mil AH, Huizinga TW. Advances in the genetics of rheumatoid arthritis point to subclassification into distinct disease subsets. Arthritis Res Ther 2008;10:205.

73 Daha NA, Toes RE. Rheumatoid arthritis: are ACPA-positive and ACPA-negative RA the same disease? Nat Rev Rheumatol 2011;7:202–3.

74 Ohmura K, Terao C, Maruya E *et al*. Anti-citrullinated peptide antibody-negative RA is a genetically distinct subset: a definitive study using only bone-erosive ACPA-negative rheumatoid arthritis. Rheumatology 2010;49:2298–304.

75 van der Woude D, Houwing-Duistermaat JJ, Toes RE *et al*. Quantitative heritability of anti-citrullinated protein antibody-positive and anti-citrullinated protein antibody-negative rheumatoid arthritis. Arthritis Rheum 2009;60:916–23.

76 Lundberg K, Bengtsson C, Kharlamova N *et al*. Genetic and environmental determinants for disease risk in subsets of rheumatoid arthritis defined by the anticitrullinated protein/peptide antibody fine specificity profile. Ann Rheum Dis 2013;72:652–8.

77 Wang Z, Nicholls SJ, Rodriguez ER *et al*. Protein carbamylation links inflammation, smoking, uremia and atherogenesis. Nat Med 2007;13:1176–84.

78 Shi J, Knevel R, Suwannalai P *et al*. Autoantibodies recognizing carbamylated proteins are present in sera of patients with rheumatoid arthritis and predict joint damage. Proc Natl Acad Sci USA 2011;108:17372–7.

79 Shi J, Willemze A, Janssen GM *et al*. Recognition of citrullinated and carbamylated proteins by human antibodies: specificity, cross-reactivity and the 'AMC-Senshu' method. Ann Rheum Dis 2013;72:148–50.

80 Willis G, Scott DG, Jennings BA *et al*. HFE mutations in an inflammatory arthritis population. Rheumatology 2002;41:176–9.

81 Barton A, Bowes J, Eyre S *et al*. Investigation of polymorphisms in the PADI4 gene in determining severity of inflammatory polyarthritis. Ann Rheum Dis 2005;64:1311–5.

82 Barton A, Platt H, Salway F *et al*. Polymorphisms in the mannose binding lectin (MBL) gene are not associated with radiographic erosions in rheumatoid or inflammatory polyarthritis. J Rheumatol 2004;31:442–7.

83 Barton A, Platt H, Salway F *et al*. Polymorphisms in the tumour necrosis factor gene are not associated with severity of inflammatory polyarthritis. Ann Rheum Dis 2004;63:280–4.

84 Thomson W, Pepper L, Payton A *et al*. Absence of an association between HLA-DRB1*04 and rheumatoid arthritis in newly diagnosed cases from the community. Ann Rheum Dis 1993;52:539–41.

85 Thomson W, Harrison B, Ollier B *et al*. Quantifying the exact role of HLA-DRB1 alleles in susceptibility to inflammatory polyarthritis: results from a large, population-based study. Arthritis Rheum 1999;42:757–62.

86 Wesoly J, Hu X, Thabet MM *et al*. The 620W allele is the PTPN22 genetic variant conferring susceptibility to RA in a Dutch population. Rheumatology 2007;46:617–21.

87 Feitsma AL, Toes RE, Begovich AB *et al*. Risk of progression from undifferentiated arthritis to rheumatoid arthritis: the effect of the PTPN22 1858T-allele in anti-citrullinated peptide antibody positive patients. Rheumatology 2007;46:1092–5.

88 Smolik I, Robinson DB, Bernstein CN, El-Gabalawy HS. First-degree relatives of patients with rheumatoid arthritis exhibit high prevalence of joint symptoms. J Rheumatol 2013;40:818–24.

89 van de Stadt LA, Witte BI, Bos WH, van Schaardenburg D. A prediction rule for the development of arthritis in seropositive arthralgia patients. Ann Rheum Dis 2013;72:1920–6.

90 El-Gabalawy HS, Robinson DB, Smolik I *et al*. Familial clustering of the serum cytokine profile in the relatives of rheumatoid arthritis patients. Arthritis Rheum 2012;64:1720–9.

91 Young KA, Deane KD, Derber LA *et al*. Relatives without rheumatoid arthritis show reactivity to anti-citrullinated protein/peptide antibodies that are associated with arthritis-related traits: studies of the etiology of rheumatoid arthritis. Arthritis Rheum 2013;65:1995–2004.

92 Ioan-Facsinay A, Willemze A, Robinson DB *et al*. Marked differences in fine specificity and isotype usage of the anti-citrullinated protein antibody in health and disease. Arthritis Rheum 2008;58:3000–8.

93 Frisell T, Holmqvist M, Kallberg H *et al*. Familial risks and heritability of rheumatoid arthritis: role of rheumatoid factor/anti-citrullinated protein antibody status, number and type of affected relatives, sex, and age. Arthritis Rheum 2013;65:2773–82.

94 Balandraud N, Meynard JB, Auger I *et al*. Epstein-Barr virus load in the peripheral blood of patients with rheumatoid arthritis: accurate quantification using real-time polymerase chain reaction. Arthritis Rheum 2003;48:1223–8.

95 Brennan P, Silman A. Breast-feeding and the onset of rheumatoid arthritis. Arthritis Rheum 1994;37:808–13.

96 Pattison DJ, Silman AJ, Goodson NJ *et al*. Vitamin C and the risk of developing inflammatory polyarthritis: prospective nested case-control study. Ann Rheum Dis 2004;63:843–7.

97 Pattison DJ, Symmons DP, Lunt M *et al*. Dietary risk factors for the development of inflammatory polyarthritis: evidence for a role of high level of red meat consumption. Arthritis Rheum 2004;50:3804–12.

98 Pattison DJ, Symmons DP, Young A. Does diet have a role in the aetiology of rheumatoid arthritis? Proc Nutr Soc 2004;63:137–43.

99 Ariza-Ariza R, Mestanza-Peralta M, Cardiel MH. Omega-3 fatty acids in rheumatoid arthritis: an overview. Semin Arthritis Rheum 1998;27:366–70.

100 Silman AJ, Roman E, Beral V, Brown A. Adverse reproductive outcomes in women who subsequently develop rheumatoid arthritis. Ann Rheum Dis 1988;47:979–81.

101 Silman AJ, Newman J, MacGregor AJ. Cigarette smoking increases the risk of rheumatoid arthritis. Results from a nationwide study of disease-discordant twins. Arthritis Rheum 1996;39:732–5.

102 Silman A, Kay A, Brennan P. Timing of pregnancy in relation to the onset of rheumatoid arthritis. Arthritis Rheum 1992;35:152–5.

103 Spector TD, Silman AJ. Is poor pregnancy outcome a risk factor in rheumatoid arthritis? Ann Rheum Dis 1990; 49:12–4.

104 Symmons D, Harrison B. Early inflammatory polyarthritis: results from the Norfolk Arthritis Register with a review of the literature. I. Risk factors for the development of inflammatory polyarthritis and rheumatoid arthritis. Rheumatology 2000;39:835–43.

105 Symmons DP. Environmental factors and the outcome of rheumatoid arthritis. Best Pract Res Clin Rheumatol 2003;17:717–27.

106 Lahiri M, Luben RN, Morgan C *et al*. Using lifestyle factors to identify individuals at higher risk of inflammatory polyarthritis (results from the European Prospective Investigation of Cancer-Norfolk and the Norfolk Arthritis Register—the EPIC-2-NOAR Study). Ann Rheum Dis 2014;73:219–26.

107 Huidekoper AL, van der Woude D, Knevel R *et al*. Patients with early arthritis consume less alcohol than controls, regardless of the type of arthritis. Rheumatology 2013;52:1701–7.

108 Stolt P, Bengtsson C, Nordmark B *et al*. Quantification of the influence of cigarette smoking on rheumatoid arthritis: results from a population based case-control study, using incident cases. Ann Rheum Dis 2003;62: 835–41.

109 Vessey MP, Villard-Mackintosh L, Yeates D. Oral contraceptives, cigarette smoking and other factors in relation to arthritis. Contraception 1987;35:457–64.

110 Hazes JM, Dijkmans BA, Vandenbroucke JP, de Vries RR, Cats A. Lifestyle and the risk of rheumatoid arthritis: cigarette smoking and alcohol consumption. Ann Rheum Dis 1990;49:980–2.

111 Klareskog L, Stolt P, Lundberg K *et al*. A new model for an etiology of rheumatoid arthritis: smoking may trigger HLA-DR (shared epitope)-restricted immune reactions to autoantigens modified by citrullination. Arthritis Rheum 2005;54:38–46.

112 Lundstrom E, Kallberg H, Alfredsson L, Klareskog L, Padyukov L. Gene-environment interaction between the DRB1 shared epitope and smoking in the risk of anti-citrullinated protein antibody positive rheumatoid arthritis: all alleles are important. Arthritis Rheum 2009;60: 1597–603.

113 Padyukov L, Silva C, Stolt P, Alfredsson L, Klareskog L. A gene-environment interaction between smoking and shared epitope genes in HLA-DR provides a high risk of seropositive rheumatoid arthritis. Arthritis Rheum 2004; 50:3085–92.

114 Pedersen M, Jacobsen S, Garred P *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: 1446–53.

115 Lee HS, Irigoyen P, Kern M *et al*. Interaction between smoking, the shared epitope, and anti-cyclic citrullinated peptide: a mixed picture in three large North American rheumatoid arthritis cohorts. Arthritis Rheum 2007;56: 1745–53.

116 Quirke AM, Fisher BA, Kinloch AJ, Venables PJ. Citrullination of autoantigens: upstream of TNFα in the pathogenesis of rheumatoid arthritis. FEBS Lett 2011; 585:3681–8.

117 Chibnik LB, Keenan BT, Cui J *et al*. Genetic risk score predicting risk of rheumatoid arthritis phenotypes and age of symptom onset. PLoS One 2011;6:e24380.

118 Karlson EW, Chibnik LB, Kraft P *et al*. Cumulative association of 22 genetic variants with seropositive rheumatoid arthritis risk. Ann Rheum Dis 2010;69: 1077–85.

119 Karlson EW, Ding B, Keenan BT *et al*. Association of environmental and genetic factors and gene-environment interactions with risk of developing rheumatoid arthritis. Arthritis Care Res 2013;65: 1147–56.

120 McClure A, Lunt M, Eyre S *et al*. Investigating the viability of genetic screening/testing for RA susceptibility using combinations of five confirmed risk loci. Rheumatology 2009;48:1369–74.

121 Scott IC, Seegobin SD, Steer S *et al*. Predicting the risk of rheumatoid arthritis and its age of onset through modelling genetic risk variants with smoking. PLoS Genet 2013;9:e1003808.

122 van der Helm-van Mil AH, Toes RE, Huizinga TW. Genetic variants in the prediction of rheumatoid arthritis. Ann Rheum Dis 2010;69:1694–6.

123 Yarwood A, Han B, Raychaudhuri S *et al*. A weighted genetic risk score using all known susceptibility variants to estimate rheumatoid arthritis risk. Ann Rheum Dis 2013 Oct 3. doi: 10.1136/annrheumdis-2013-204133. [Epub ahead of print].