Looking back: developments in our understanding of the genetic epidemiology of rheumatoid arthritis over the last 50 years

J. Worthington

The contribution of the arc Epidemiology Unit over its 50-yr history to the study of the genetic epidemiology of rheumatoid arthritis (RA) has been reviewed. Early family and population based studies carried out by John Lawrence were important in establishing the role of both genetic and environmental factors in determining susceptibility to RA. More recently, under the leadership of Alan Silman, population-based cohorts, twin- and family-based studies have formed the basis for an extensive programme of research aimed at identifying specific genetic factors that might influence susceptibility and outcome. A review of linkage and association studies is presented.

The Lawrence years (1954–1968)

The history of research into genetic aspects of rheumatoid arthritis (RA) began with the work of the first clinical director of the unit, Dr John Lawrence, appointed just 1 yr after Watson and Crick signalled the commencement of a new era of genetic research with their publication of the double helical structure of DNA [1].

The population survey carried out in Leigh, Lancashire by Lawrence and Jonas Kellgren, the first Professor of Rheumatology in Manchester, had been used to determine the prevalence of RA in a random sample of adults aged between 55 and 64 yr [2]. The cause of RA was unknown but it was suggested that hereditary factors might be involved. The simplest observation that implicates genetics in disease susceptibility is the clustering of cases within families. Lawrence subsequently revisited the RA cases [and individuals positive for rheumatoid factor (RF)] identified in the Leigh study, together with their surviving parents, siblings and children over the age of 14. A sample of 1 in 30 of the households in Leigh, taken from the electoral register, served as the control group. In addition to clinical examination, a blood sample for RF was obtained and X-rays were taken using the mobile X-ray facility. RF was encountered four times as often in the rheumatoid families as in the controls, leading to the conclusion that familial aggregation did exist for RA, but it was not possible to determine if this was caused by genetic or environmental factors [3].

This question is most readily addressed by studying twins, based on the assumption that disease concordance rates will be higher in identical twins or monozygotic (MZ) twins compared with non-identical or dizygotic (DZ) twins. So between 1962 and 1967 Lawrence carried out a survey of twins in the UK and Netherlands. All RA patients attending a rheumatology clinic in 23 UK and four NL centres were asked: ‘are you a twin?’ X-rays and serum RF were used in diagnosis and zygosity was determined using eight blood group markers. Concordance rates were 30% for MZ and 5% for DZ for seropositive RA [4]. In his Heberden Oration in 1969, entitled ‘Rheumatoid arthritis—nature or nurture?’, when discussing this data Lawrence concluded that both environmental and genetic factors were involved and that a monogenic dominant or recessive mode of inheritance was unlikely, polygenic being more likely, and he speculated about the role of mitochondrial versus nuclear genes [5].

The middle years: Philip Wood (1968–1989)

Although researchers at the Epidemiology Unit were not actively involved in genetic research during this period, there were key discoveries in genetics and rheumatology that would impact on future work. Techniques for DNA sequencing led to a greater understanding of gene structure and the invention of the polymerase chain reaction in 1985, is probably the single most significant methodological development in modern genetics [6].

In 1978 Peter Statsny described the association of HLA-DR4 with RA, the first evidence of a genetic association to RA [7]. This was followed by hundreds of publications investigating in detail the relationship between HLA and RA and the identification of numerous associations. In 1987 Peter Gregersen published his unifying shared epitope hypothesis, defining a short amino acid sequence in the third hypervariable region of the HLA*DRB1 gene, common to all RA-associated DRB1 genes [8].

Alan Silman 1988-

Before his move to Manchester, Alan Silman had already recognized the value of combining clinical, genetic and epidemiological skills for the study of common complex diseases. Together with Bill Ollier, he carried out a study of multicase RA families in order to investigate whether the well-recognized phenotypic heterogeneity observed in RA patients was influenced by the same factors that determine disease susceptibility. Clinical and laboratory measures, including age and calendar year at onset, pattern of joint involvement, presence of nodules, RF, ANA titres and HLA type, were assessed in 27 sibling pair and six sibling trio families. No greater concordance for clinical features was observed within families compared with between families [9]. Although the study did not account for differences in treatment and compared only cross-sectional data, the conclusion drawn is consistent with the current thinking that RA susceptibility and outcome are influenced by the interplay of multiple genetic and environmental factors. The results were offered as justification for the use of family-based linkage studies to identify susceptibility factors relevant to the sporadic form of the disease.
National Repository of Multi-case RA Families and Linkage Studies

The pioneering study by John Todd and colleagues in diabetes illustrated the utility of large cohorts of multicase families for genetic linkage studies in complex diseases [13]. Capitalizing on the experience of the nationwide twin study, Alan Silman organized a UK collection of affected sibling pair (ASP) families with RA. Research nurses were used to visit patients in their homes and collect detailed clinical and epidemiological information. Over 400 families with two or more affected siblings were recruited and again lymphocytes were harvested so that EBV cell lines were derived for most participants, ensuring the establishment of long-term resource for RA genetic research [14].

At this time, markers within candidate genes were a rarity, meaning that candidate gene association studies were generally not possible. The arc Repository of Family Material provided the opportunity to investigate candidate genes by microsatellite-based linkage analysis. Using markers mapping close to the candidate genes, promising results were obtained for the natural resistance associated macrophage protein 1 (NRAmP1), oestrogen synthase (CYP19) and corticotrophin releasing hormone (CRH) [15–17]. An investigation of a number of cytokine genes using the same approach revealed evidence of linkage to three genes, IL5R, interferon γ and IL2, but only in clinically defined subsets of ASPs (female pairs, pairs with an affected male, and seronegative pairs respectively) [18].

The main motivation for the collection of ASP families was to allow a microsatellite whole-genome linkage study. This study was carried out collaboratively between the arc Epidemiology Unit and Paul Wordsworth and colleagues at the University of Oxford. One hundred and eighty-two ASP families were genotyped for 365 microsatellite markers. Single-point and multipoint linkage analysis revealed linkage to the HLA region on chromosome 6p and 11 additional loci [19]. The identified loci were then investigated in a second cohort of 217 ASPs. No significant evidence for linkage was detected. The power of the two cohorts was modest but similar. However, the first cohort had an earlier age at disease onset and higher frequency of the shared epitope (SE). Evidence for linkage was detected for a number of loci in one or more subsets, with the strongest effects seen for loci on 6q (female and two copies of the SE) and 16p (age at onset <40 yr) [20].

Although there are no spontaneously occurring rodent models of arthritis, there are a number of methods of inducing an inflammatory arthritis. Many loci have been mapped and a number are common to more than one model. Homologous regions in the human genome can be identified and targeted for investigation in human linkage studies. Using this approach, Anne Barton has identified a potential RA locus on chromosome 17q21 homologous to Oia3 and Cia5 [21]. Studies in three cohorts of ASP and simplex families have confirmed this finding and defined a 5 MB region containing over 40 genes that is currently being studied in detail.

It is a commonly held view that phenotypically overlapping conditions may share some of the same genetic susceptibility factors and that some RA genes may be common inflammatory or autoimmune genes. We have therefore targeted loci identified in type 1 diabetes for investigation in RA. Evidence of linkage and association has been detected for two loci on chromosome 6q (IDDM5 and 8) [22].

Technology has continued to advance at a pace and has in turn driven genetic research. In 2004 the reference sequence of the human genome was published and millions of markers have been identified and recorded on public databases. This brings exciting new possibilities for complex disease genetics, ranging from better-designed candidate gene association studies to, ultimately, systematic whole-genome screens by association based on the analysis of hundreds of thousands of single-nucleotide polymorphisms (SNPs). In collaboration with Affymetrix, we genotyped over 11 000 SNPs in 157 families, generating 5.6 million genotypes. This study confirmed the utility of this technology for SNP-based linkage screening and demonstrated a uniformly higher information and resultant improved peak definition. In terms of RA loci, the results were broadly similar to the microsatellite data, although four additional loci were identified on 13q, 14q, 21 and Xp [23].

HLA

The unit has contributed to many aspects of HLA research and a thorough review of all this work is beyond the scope of this article. Investigations have included molecular-based genotyping studies in many arthritic diseases, HLA-based definition of disease subgroups [24], population studies of RA and HLA associations (reviewed in [25]) in Nigeria [26], Greece [27], UK Caribbean [28] and India [29], and MHC in different species [30]. The HLA SE alleles have generally been assumed to be susceptibility factors for RA; however, the NOAR (Norfolk Arthritis Register) study (described in detail in D. Symmons’ article), provided the opportunity to dissect out the influence of SE alleles on susceptibility and severity or outcome of arthritis. In this population-based study, patients have been recruited at an early stage in disease and followed for a number of years. Only a modest association was observed for the SE and inflammatory polyarthritis (odds ratio 1.8, 95% confidence interval 1.4–2.4) and this increased slightly if only patients satisfying the 1987 ACR criteria were included (odds ratio 2.3, 95% confidence interval 1.7–3.1). Compared with the five-fold increased risk usually quoted for hospital-based studies, these figures are low, suggesting that SE alleles may be more important in influencing disease progression than in susceptibility [31].

Non-HLA candidate gene investigations

In the modern era we have the capacity to design candidate gene investigations to have the power to detect the weak effects we must expect for RA, a typical common complex disease. Large,
clinically well-characterized cohorts and the ability to carry out
high throughput genotyping are essential. Recent studies from the
arc Epidemiology Unit have failed to detect association with
disease susceptibility or severity of outcome for CCR-5 [32], TNF
[33], MBL [34], CTLA-4 [35], PADI4 [36] and SLC22A4 [37].
Association has been detected with two interesting genes: macro-
phage migration inhibitory factor (MIF) and protein tyrosine
phosphatase 22 (PTPN22).

MIF is a ubiquitously expressed protein with proinflammatory,
and enzymatic activities that appears to have a central
role in determining the magnitude of the immune response.
Through investigations of juvenile idiopathic arthritis (JIA),
Rachelle Donn has characterized a functional two-marker haplo-
type in the promoter of MIF comprising the allele with seven
repeats of a tetranucleotide repeat and the mutant allele of ~173
G/C SNP [38]. Carriage of this haplotype was associated with a
three-fold increased risk of developing an inflammatory poly-
arthritis, although it did not appear to influence the severity of
disease assessed at 5yr. This is the same haplotype previously
associated with JIA [39] and also with psoriasis [40].

PTPN22 is a negative regulator of T-cell activation and over
the last year a non-synonymous SNP (R620W) in PTPN22 that
appears to interfere with the negative regulatory role has been
associated with type 1 diabetes [41], systemic lupus erythematosus
[42], autoimmune thyroid disease [43] and with RF-positive RA
[44]. Our investigation of this polymorphism revealed association
with RA, including a subset of RF-negative patients and patients
with JIA; and with all ILAR subgroups except systemic onset and
juvenile psoriatic arthritis, but not with patients with psoriasis,
psoriatic arthritis or multiple sclerosis [45].

Summary
The importance of combining the skills of epidemiologists,
rheumatologists and laboratory-based scientists when studying
the genetic epidemiology of arthritic diseases has been long
recognized in the arc Epidemiology Unit. From the earliest
population surveys, with the introduction of the mobile X-ray
unit and the development of an assay for RF through to the most
recent candidate gene investigations based on large, well-char-
acterized, prospectively ascertained cohorts, taking advantage of
the latest high-throughput genotyping technologies, this multi-
disciplinary approach has yielded high-quality research.

References
1. Watson JD, Crick FH. Molecular structure of nucleic acids; a struc-
2. Kellgren J, Lawrence JS, Aitken-Swan J. Rheumatic complaints in an
8. Gregersen PK, Silver J, Winchester RJ. The shared epitope hypo-
thesis. An approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis. Arthritis Rheum 1987;30:
1205–13.
135–8.
14. Worthington J, Ollier WE, Leach MK et al. The Arthritis and Rheumatism Council’s National Repository of Family Material: pedigrees from the first 100 rheumatoid arthritis families containing
17. Fife MS, Fisher SA, John S et al. Multipoint linkage analysis of a candidate gene locus in rheumatoid arthritis demonstrates significant
19. MacKay K, Eyre S, Myerscough A et al. Whole-genome linkage analysis of rheumatoid arthritis susceptibility loci in 252 affected
21. Barton A, Eyre S, Myerscough A et al. High resolution linkage and
association mapping identifies a novel rheumatoid arthritis suscept-
22. Myerscough A, John S, Barrett JH, Ollier WER, Worthington J. Linkage of rheumatoid arthritis to insulin-dependent diabetes mellitus
loci: evidence supporting a hypothesis for the existence of common
28. MacGregor AJ, Riste LK, Hazes JMW, Silman AJ. Low prevalence of rheumatoid arthritis in Black-Caribbeans compared with whites
32. MacGregor AJ, Riste LK, Hazes JMW, Silman AJ. Low prevalence of rheumatoid arthritis in Black-Caribbeans compared with whites
33. Ollier WE, Stephens C, Azad et al. Is rheumatoid arthritis in Indians associated with HLA antigens sharing a DR beta 1 epitope?
34. Ollier WE, Kennedy LJ, Thomson W et al. Dog MHC alleles containing the human RA shared epitope confer susceptibility to


