

Review

## Genetics of rheumatoid arthritis: confronting complexity

Peter K Gregersen

North Shore University Hospital, Manhasset, New York, USA

Received: 8 October 1999  
Accepted: 14 October 1999  
Published: 26 October 1999  
© Current Science Ltd

### Important note about how to cite this article

This article is also available online in the *Arthritis Research* website. To avoid confusion, please ensure that only the online version of the article is cited in any reference, as follows:  
Gregersen PK: Genetics of rheumatoid arthritis: confronting complexity [review]. <http://arthritis-research.com/26oct99/ar0101r04>

### Abstract

The genetic basis for rheumatoid arthritis (RA) is likely to be extremely complex. Even the role of MHC genes remains to be fully defined, and may involve interactive genetic effects. The difficulty of precisely defining the clinical phenotype, as well as underlying genetic heterogeneity, complicates the problem. In addition, stochastic genetic or physiologic events may contribute to the low penetrance of susceptibility genes. This situation parallels developing paradigms for other autoimmune disorders, in which many different genes each appear to contribute a small amount to overall risk for disease, and where severity and specific phenotypic subtypes are subject to genetic effects. The completion of the human genome project, along with advances in informatics, will be required to reach a deeper understanding of RA. It is likely that this will involve an iterative and interactive process between several different scientific disciplines.

**Keywords:** genetics, rheumatoid arthritis

### Introduction

Now that we are beginning to feel the impact of human genomic research on biomedical science, those of us who are attempting to apply the tools of modern genetics to autoimmune disorders are being forced to confront a sobering reality – genetically complex disorders really are complex! This lesson is also being learned by geneticists working on other polygenic diseases, including psychiatric and metabolic disorders, and cancer. The heyday of identifying the molecular basis for diseases caused by one or a few high-penetrance genes is not yet over, as illustrated by the recent exciting progress in defining the periodic fever syndromes [1,2]. However, the future challenges are largely going to involve understanding the genetic underpinnings of common disorders that are generally ‘sporadic’,

in the sense that they do not exhibit a high degree of familial recurrence. It is likely though that genetics will be important for understanding these diseases.

A fundamental feature of mammalian organisms, not to mention single cells, is that they are very complex. For cell biologists, biochemists, and even scientists working with whole organisms, however, clever experiments can be designed that attempt to control for this complexity, and enable one to manipulate and examine the effects of isolated changes in a parameter of interest. When it comes to the study of genetics and polygenic human disease, however, the experiments have already been done, and we must make do with interpreting the results as they exist in the human population. The task of the researcher

is to define a unifying set of parameters to study (ie carefully define the phenotype), and to seek out and interpret the informative mating experiments that are hidden within modern populations. This requires a group of skills that are rarely found in one individual, and this naturally leads to collaborations among scientists with expertise in clinical disease definition, disease pathogenesis, and molecular and statistical genetics. The experimental designs also generally require large sample sizes, and therefore patients and clinicians must do their part if we are to succeed in bringing such studies to fruition. Finally, technology is a major force behind this rapidly moving field, and has a substantial influence on the kinds of experiments that can be contemplated. This review highlights some of these current trends as they relate to the problem of rheumatoid arthritis (RA).

### Estimating the size of the genetic component in rheumatoid arthritis

RA does not aggregate in families with very high frequency. In addition, concordance rates in identical (monozygotic) twins are relatively low (12–15% see [3,4]) compared with other autoimmune disorders, which generally have monozygotic twin concordance rates in the range of 30% [4]. Nevertheless, the prevalence rates of RA in first-degree relatives of probands with RA are considerably higher than in the general population as a whole. A comparison of disease prevalence rates in populations of individuals with different degrees of genetic relatedness can be used to calculate risk ratios, or  $\lambda$ . The relative risk to siblings of affected probands, compared with that in the general (unrelated) population, is a widely used measure [5], and is calculated as follows:

$$\lambda_s = \frac{\text{Disease prevalence in siblings of affected individuals}}{\text{Disease prevalence in general population}}$$

For many of the common autoimmune disorders,  $\lambda_s$  values appear to be in the range 10–20, or higher [6]. However, for RA the true  $\lambda_s$  is highly uncertain and ranges from a low of 2 to at least 10, although the prevalence in siblings of affected persons is generally agreed to be in the 2–4% range [7]. There are at least two reasons for this uncertainty. First, and probably most important, is the fact that RA is very likely to be a clinical syndrome with a heterogeneous etiology. One has only to consider the difficulty of obtaining meaningful data if the broad category of ‘arthritis’ were used as the defining phenotype for a  $\lambda_s$  calculation. In that case, one would lump in spondyloarthropathies, hemochromatosis and other highly genetic disorders into the calculation, and the  $\lambda_s$  value would be low and misleading. Our inability to define RA precisely makes it difficult to know whether additional highly genetic subtypes of ‘rheumatoid arthritis’ exist. The second problem, related to the first, is the inaccuracy of

defining background prevalence for this disease. A prevalence of 0.8–1% is generally cited [8]. Both under-reporting and over-reporting of ‘true’ prevalence is probably common, however, because of misdiagnosis, lack of current disease activity, and again the problem of disease definition. In addition, severity of disease, which is also likely to have a genetic component [9,10], is not readily assessed in large population prevalence surveys.

The description of the human leukocyte antigen (HLA) associations with RA over two decades ago [11] has been a major source of support for the hypothesis that genetic factors are important for susceptibility to RA. The contribution of HLA to the overall genetic risk has been variously estimated at between 30 and 50% [7]. Depending on your assumptions regarding the  $\lambda_s$ , this estimate for the HLA contribution may be too low or too high. Several recent studies of HLA haplotype sharing among affected sibling pairs with RA [12,13] are in excellent agreement, and indicate that the HLA region contributes a  $\lambda$  of about 1.7 to the total  $\lambda_s$ . If the total  $\lambda_s$  is as low as 2, this leaves little room for a major contribution by other genes. On the other hand, if the  $\lambda_s$  is 10, the majority of genetic susceptibility must be due to non-HLA genes.

### The nongenetic component

Usually a tacit assumption is made that ‘nongenetic’ components of disease really refer to environmental factors. Despite the obvious role of microbial agents in some forms of inflammatory arthritis, the search for infectious causes for RA has been frustrating [14]. Other environmental factors, such as body weight, smoking, and blood transfusion, may make a modest contribution to disease risk [15]. Similar to genetic associations, a role for these factors must be confirmed by replicating these findings in large populations. A second type of ‘nongenetic’ factor relates to chance events, however, some of which may occur either early in development, or before the onset of disease. Clearly, because the concordance rate for monozygotic twins is relatively low (12–15%), and the ‘environmental’ contributions appear to be modest, it is reasonable to explore additional explanations for the relatively low penetrance of disease in individuals who are known to be genetically predisposed (ie unaffected cotwins from discordant monozygotic twin pairs with RA).

There are at least three broad areas in which stochastic processes might influence disease penetrance: somatic genetic events, epigenetic events, and physiologic events at the biochemical and cellular level.

The importance of chance somatic mutation is a widely accepted concept in the field of cancer research. The recent reports of mutations in p53 in synovial tissue are provocative, and raise the possibility that somatic mutations might also play a role in RA [16]. In addition, the

rearrangements of immunoglobulin and T-cell receptors are examples of normal somatic genetic events that can generate diversity between genetically identical individuals. It must be admitted, however, that the importance of these somatic genetic mechanisms for differential disease susceptibility has not been directly demonstrated for RA, or for any other autoimmune disease.

On the other hand, somatic epigenetic events have clearly been shown to influence phenotypes other than cancer. This is well illustrated by the example of X-chromosome inactivation [17]. In females, a choice regarding which X chromosome should be inactivated in each cell is made very early in embryonic life, and not infrequently results in the preferential utilization of either paternally or maternally derived X-linked genes [18]. This can result in discordant expression of X-linked phenotypes in identical twins, due to the epigenetic influence of DNA methylation and chromosomal structure on gene transcription [19]. Interestingly, such monoallelic expression of genes also occurs on autosomes, including several cytokine genes [20,21], as well as other chromosomal regions subject to genomic imprinting [22]. The degree of random variation in this process, or other epigenetic phenomena, is largely unexplored, and might contribute to the discordant expression of autosomal susceptibility genes that are relevant to autoimmune disease [23]. One interesting variant of this line of thought is the idea that epigenetic variation *per se* might be a risk factor for disease. A hypothesis incorporating this idea with respect to X-chromosome inactivation has been proposed to explain the higher prevalence of autoimmune disease in women [24]. So far, our data do not support this hypothesis [25] (Gregersen PK, unpublished data), but the concept remains an intriguing one.

Finally, chance variation, or an individual's tendency toward higher degrees of variation, in normal biochemical and cellular processes might predispose to dysregulation of inflammatory processes. For example, the difference between positive and negative selection of the T-cell repertoire is currently thought to be a result of rather small differences in the binding energy of each T-cell receptor for particular MHC-self peptide combinations [26]. Indeed, similar small differences in binding affinity determine agonist versus antagonist responses by peripheral T cells. This explains how extremely subtle changes in either peptide, MHC, or T-cell receptor structure may have a profound influence on the outcome of T-cell stimulation [26]. There are bound to be numerous aspects of T-cell receptor-MHC peptide interactions in the thymus that are subject to chance; it seems highly unlikely that all thymocytes are exposed to all MHC peptide combinations, in the same density and in the same cellular context. It is reasonable to suppose that there are geographic, circadian, or other types of variation in the selecting environment for each thymocyte. Perhaps this variation increases with age,

as the thymus slowly involutes. At some point, thymocytes that usually would cross the affinity threshold for negative selection may end up being positively selected, or vice versa, resulting in a peripheral repertoire with different tendencies for stimulation by cross-reactive antigens.

Admittedly, this example implies a role for cross-reactive T-cell recognition in the pathogenesis of RA, a hypothesis that remains to be proved. Analogous sources of variation, however, might lead to the chance crossing of thresholds for cytokine feedback amplification or cell migration into sites of injury and inflammation. Obviously, this concept can be also be extended to include variations in the physiologic state of an organism at the time of exposure to environmental risk factors. This kind of variability could partly explain the low penetrance of RA, as reflected in the commonly discordant expression of disease among monozygotic twins.

### **The major histocompatibility complex in susceptibility to rheumatoid arthritis**

Since the late 1980s, a consensus has developed around susceptibility to RA being due to a closely related set of polymorphic sequences (the 'shared epitope') on several different DRB1 alleles, especially certain subtypes of the DR4 and DR1 allelic families [27,28]. It is still unclear how this set of class II alleles operates to confer susceptibility to RA, however. Popular models invoke selective peptide presentation of autoantigens [28], biased thymic selection toward an autoreactive T-cell repertoire [29], direct effects on antigen processing [30], and a direct role for the shared epitope itself as a nominal peptide antigen [31], to name a few. The unifying concept of the shared epitope has considerable appeal for understanding the MHC class II associations with RA, but it is clearly an oversimplification to consider the shared epitope as the only relevant MHC polymorphism for susceptibility to RA [32]. There are important haplotypic influences on the degree of risk conferred by the shared epitope alleles. Certain shared epitope alleles, such as DRB1\*0401, confer much greater risk than others, such as DRB1\*0101 [33]. In addition, homozygosity for particular combinations of haplotypes, such as DRB1\*0401/0404, appear to confer especially high risk or influence disease severity as well as risk [10,34,35]. Differences in female versus male risk have been described for different alleles [36]. Some investigators have proposed a role for DQ alleles on these haplotypes [37], although so far no convincing population data in humans supports this. It must be admitted, however, that it is very difficult to identify large enough populations to do the risk comparisons for interactive effects between DR and DQ, because DR4 and DQ3 alleles are so commonly found on the same haplotypes.

Another recent development has been the sequencing of most of the MHC, and the realization that the 'central'

MHC contains many genes that could be directly involved in disease risk, or might interact with DRB1 alleles to modify risk [38]. Tumor necrosis factor (TNF)- $\alpha$  is a particularly compelling candidate because of the obvious therapeutic importance of this cytokine. Several recent reports suggest that polymorphisms in the TNF region may interact with DR alleles to modify susceptibility to RA [39,40]. These studies will need to be replicated and confirmed using large numbers of patients and well matched control individuals, but they emphasize that our understanding of the MHC influences on susceptibility to RA is far from complete.

### Identifying susceptibility genes outside of the major histocompatibility complex

#### The candidate gene approach

One approach to identifying genetic susceptibility alleles outside the MHC is to analyze polymorphisms in genes that can reasonably be implicated in a pathway of pathogenesis [41]. Obviously, success depends on having a model of pathogenesis that at the very least involves relevant biochemical pathways, and on selecting the right genes in these pathways to study. It is highly uncertain whether either of these prerequisites can be fulfilled by our current knowledge. Most candidate genes that have been addressed to date involve immune recognition, cytokines and their receptors, or genes generally thought to be involved in inflammation. Of course 'inflammation' entails a large number of cellular processes, so that the list of candidates can get very broad, and could reasonably extend to a large fraction of the 100 000–150 000 genes now thought to reside in the human genome.

An additional problem is that the extent of genetic polymorphism is not well defined for many candidate genes of interest, and many polymorphisms exist in untranslated regulatory portions of genes, often in the form of single nucleotide polymorphisms (SNPs) or variable numbers of tandem repeats. Usually it is unclear whether these SNPs and variable numbers of tandem repeats have any functional significance, although there are examples where transcription does appear to correlate with these types of sequence changes [42,43]. Clearly, the premise underlying a candidate gene approach dictates that functional changes should be of most interest. Recent surveys indicate that SNPs within coding regions are biased towards silent, and presumably functionally irrelevant, substitutions [44,45]. This is consistent with evolutionary selection against deleterious mutations. Nevertheless, functionally irrelevant polymorphisms may be in linkage disequilibrium with other sequence differences that do affect function. This implies that a general screen of SNPs or other types of polymorphisms within and around a candidate gene can be a rational approach, regardless of their direct functional effects. As discussed below, however, there is some debate concerning this point [46].

Positive associations between RA and a number of candidate genes have been reported, and a few have been replicated. Variable results have been reported for the T-cell receptor loci [47–50]. Among the cytokines (other than TNF), a recent report of an association with interleukin (IL)-4 is provocative [51]. Studies of IL-1, IL-1 receptor antagonist, and IL-10 show only weak or negative associations [51,52]. IL-10 is of interest because promoter polymorphisms in this gene are associated with differences in levels of transcription [42], and low IL-10 expressing haplotypes may influence the asthma phenotype [53]. Modest associations (estimated relative risks in the 1.3–2 range) with cytotoxic T-lymphocyte associated antigen-4 have been reported in RA [54,55], as well as in type I diabetes [56,57] and Graves' disease [57]. Other candidate genes that do not fall strictly into the immunological category include corticotrophin releasing hormone [58], glutathione S-transferase [59], and Nramp1 [60].

Overall, the evidence for associations with most of these candidates is only suggestive at best. It should be remembered that the sample size requirements are substantial to detect these kinds of modest associations. For example, even for an allele that is quite common in the population (20%), the sample size required to achieve 80% power to detect a relative risk of 1.3 is 1324 individuals per group. For detection of relative risks of 1.5, 1.7 or 2.0, the required sample sizes are 535, 304 and 172 respectively. Many studies do not achieve these sample sizes. Therefore, major resources must be put into clinical recruitment if these observations are to be convincingly demonstrated and confirmed. In addition, stratification of patient populations into groups that carry certain other alleles, such as particular MHC haplotypes, is likely to be required. This can occasionally reveal evidence of striking interactive effects [56].

#### Screening the genome

An alternative and complementary approach is to pursue a genome-wide screening strategy, instead of focusing on particular candidate genes. This has the advantage of avoiding any assumptions about the molecular pathways involved in disease pathogenesis. Currently, the only feasible method of doing this depends on searching for linkage within multiplex families. A preferred approach for RA is to utilize affected sibling pairs (ASP) [7], because this avoids the necessity of defining any family members as unaffected. (Defining a family member as 'unaffected' is particularly difficult for RA, because the age of onset may be late in life; in addition, an unaffected family member may well carry susceptibility genes, due to the low penetrance of RA.) Several major collections of affected sibling pairs are now being carried out [12,61,62]. The European Consortium for Rheumatoid Arthritis Families published the first genome screen of such families (total of 308 sibling pairs) and has reported preliminary evidence for linkage to a number of different chromoso-

mal regions, including regions on chromosomes 3 and 18 [12]. A second major collection of RA sibling pairs in the UK is currently being analyzed [62].

In the USA, the North American Rheumatoid Arthritis Consortium was established in 1997 with the goal of identifying 1000 affected RA sibling pairs [61]. Over 600 families have now been ascertained. A preliminary analysis of 180 affected sibling pairs confirms the estimated  $\lambda_{\text{HLA}}$  of 1.7 which was reported by the European Consortium for Rheumatoid Arthritis Families, but fails to confirm linkage to the region around 3q13 [12,13] (North American Rheumatoid Arthritis Consortium, unpublished observations). These results demonstrate the utility of ASP for mapping susceptibility regions such as the MHC, but also emphasize the importance of having large numbers of such families in order to detect chromosomal regions containing genes which confer only modest risk. Because of the relatively lower statistical power of the ASP linkage-based method compared with association studies [63], both positive and negative findings on a few hundred sibling pairs must be viewed with caution.

In order to foster optimal utilization of the North American Rheumatoid Arthritis Consortium collection of data on RA families, a web site has been established to provide access for the scientific community. This site (<http://narac.patternrx.com>) will go online in November 1999. It contains detailed clinical information on these families (anonymized and coded to preserve patient confidentiality), including digitized and downloadable hand radiographs on all affected sibs, joint evaluations using the joint alignment and mobility score [64], and HLA and serological data. This site will also be a repository for genotyping data. There are several levels of access to the site, and access to genotyping data and DNA samples will be available to investigators. The intent is to provide an information resource that will steadily increase in value as genetic information accumulates on these families to permit genotype-phenotype correlations.

### **Integrating diverse methods and technologies to identify rheumatoid arthritis susceptibility genes**

While the study of RA families will probably lead to the identification of one or more regions outside the MHC that contains susceptibility genes, this method will not permit the identification of these genes. Linkage by ASP will at best permit narrowing a region of interest to 5–10 million base pairs. A region of this size contains a large number of genes. Hopefully most of the genes will be known by virtue of the completion of the Human Genome Project. We will then still be faced with a large number of candidate genes to investigate. In this sense, linkage mapping using affected sibling pairs can be thought of as a method of hypothesis generation, by focusing attention on

a more limited, but still large, set of candidate genes. In this case, the rationale for the hypothesis rests on gene location, rather than on involvement in a pathway of pathogenesis.

Therefore, the end game for gene identification is going to involve the evaluation of candidates. This will likely be an iterative process involving knowledge of gene location, pattern of expression, evidence for involvement in pathways of pathogenesis, and discovery of functionally relevant allelic variants. Any techniques or advances that contribute to these factors will be important components of successful gene identification.

Recently, SNPs have received a lot of attention as a powerful tool for genetic analysis [65]. These polymorphisms may be present in 0.05–0.15% of the human genome, implying the existence of 2–6 million such SNPs. Several recent analyses [44,45] suggested that the majority of these changes are silent within coding regions [44,45]. SNPs in promoter and regulatory regions may affect gene expression [42,66]. Furthermore, even in the absence of function, a dense collection of SNPs in a candidate gene or region of interest may be very useful for genetic mapping by linkage disequilibrium, although this may not be the case for all regions of the genome [46,67]. At present, however, the data on SNPs is limited, and SNP databases are only just beginning to be organized (see <http://www.ncbi.nlm.nih.gov/SNP/>). The development of such resources is critical to allow investigators to pursue the efficient analysis of candidate genes for RA, or any other disease. In addition, reasonably large sample sizes are going to be required to pursue this approach to gene identification [68]. Finally, the best methods for achieving high throughput SNP typing are still being worked out [69,70].

Advances in microarray technology will also permit more intelligent and directed selection of candidate genes [71]. In the next few years, it is likely that massive parallel analysis of gene expression will become accessible to individual investigators. This may permit the identification of characteristic changes in patterns of gene expression in specific cells, tissues and patient populations. Clearly, comprehensive data on patterns of gene expression in rheumatoid synovial cells, monocyte or dendritic cell subsets may lead to reasonable new candidate genes. A database of SNPs in these genes would then permit a rapid analysis of patient populations for evidence of association with SNP haplotypes.

Information technologies and analytic tools are another important component for candidate gene evaluation. Internet access to the clinical and genetic data, as well as physical access to DNA, from well defined patient populations will likely be increasingly valuable to individual investigators who are pursuing a gene of interest. This is

in part because it is likely that candidate gene involvement may only be apparent in subsets of patients who have particular phenotypic and genotypic characteristics. This relates to the analytic problem of gene–gene interaction effects in disease susceptibility, it may be that the effects of some loci are dependent on particular alleles at a second locus, a phenomenon termed epistasis [72]. Thus, in analyzing affected sibling pairs, stratifying the population by sharing at HLA may reveal stronger evidence for linkage at a second non-HLA locus. These types of genetic interactions have clearly been shown in murine models of autoimmunity [73], and evidence for this phenomenon is beginning to emerge in human studies [74]. If such epistatic interactions involve multiple loci, it becomes computationally challenging to examine all the different possible combinations. The continuing development of new analytic methods by statistical geneticists will clearly be important for future progress in this area.

### Rheumatoid arthritis genetic susceptibility in the larger context of human autoimmune disease

A recent analysis by Becker *et al* [75] ignited renewed interest in the question of whether there are common genes underlying multiple forms of autoimmunity. A review of both human and rodent mapping studies indicated that chromosomal regions linked to autoimmunity are not randomly distributed across the genome, but rather appear to cluster in at least 18 regions [75]. Indeed, even within the human MHC, it has been known for many years that certain haplotypes, such as A1-B8-DR3 are associated with different autoimmune phenotypes [38]. A more recent example may be the polymorphisms in cytotoxic T-lymphocyte associated antigen-4, which appear to associate with RA, type I diabetes and autoimmune thyroid disease [54–57].

Another means of assessing this genetic overlap is to examine the aggregation of different autoimmune disorders within families. Perhaps the most compelling evidence involves the aggregation of type I diabetes and autoimmune thyroid disease, particularly Hashimoto's thyroiditis. In addition to the increased prevalence of Hashimoto's in patients with type I diabetes, there is a marked increase of this thyroid disorder in the parents and siblings of diabetic probands. In a large study [76,77], the prevalence of Hashimoto's thyroiditis was 16.2 and 25.3% in male and female siblings under the age of 40 years, respectively. This suggests a  $\lambda_s$  of at least 20 for Hashimoto's thyroiditis in families with a type I diabetic proband.

The clinical association of RA with autoimmune thyroid disease is well known, but there has been little formal investigation of familial aggregation of these two disorders. A recent small study of families of RA probands [78] revealed an increase in both autoimmune thyroid disease

and type I diabetes in the first-degree relatives. Families of patients with inflammatory myopathy also exhibit familial aggregation of RA, thyroid disease, and type I diabetes [79]. Thus, although the data is still relatively sparse, it appears likely that RA, type I diabetes, and autoimmune thyroid disease actually do aggregate in families [80], and this presumably reflects the effects of common genes for these disorders. Whether this aggregation of autoimmune diseases extends to other less common disorders such as multiple sclerosis and lupus is unclear. Convincing epidemiological data on large populations are not currently available.

In September 1999, the National Institute of Immunology, Allergy, and Infectious Diseases awarded a major contract to gather a large number of families in which multiple autoimmune disorders aggregate together. (For more information, readers may contact the author at peterg@nshs.edu, Dr Timothy Behrens at behre001@maroon.tc.umn.edu, or Dr Lindsey Criswell at lac@itsa.ucsf.edu). This database and repository will include multiplex families with RA as well as a number of other autoimmune diseases such as lupus, type I diabetes, multiple sclerosis, and autoimmune thyroid disease. Similar to the major collections of RA families, this new family collection will provide additional resources for individual researchers to investigate the genetic relationships between various forms of autoimmunity.

### Conclusion

This brief review emphasizes the need for major collaborative efforts between clinicians, immunologists, geneticists, and statisticians in order to tease apart the complex genetic factors that underlie RA and other forms of autoimmunity. The completion of the Human Genome Project is going to accelerate and complicate the process of gene discovery. In his recent book, *The Sun, the Genome, and the Internet*, Dyson [81] has articulated the importance of new tools, as well as new ideas, for the craft of doing science. It is increasingly apparent that both the genome and the Internet are new tools that will have a major impact on chasing down the genetic basis of rheumatoid arthritis.

### Acknowledgements

The author is supported by the National Institutes of Health (AR-44422, AR-7-2232).

### References

1. McDermott MF, Aksentjevich I, Galon J, *et al*: **Germline mutations in the extracellular domains of the 55 kDa TNF receptor, TNFR1, define a family of dominantly inherited autoinflammatory syndromes.** *Cell* 1999, **97**:133–144.
2. Centola M, Aksentjevich I, Kastner DL: **The hereditary periodic fever syndromes: molecular analysis of a new family of inflammatory diseases.** *Hum Mol Genet* 1998, **7**:1581–1588.
3. Silman AJ, MacGregor AJ, Thomson W, *et al*: **Twin concordance rates for rheumatoid arthritis: Results from a nationwide study.** *Br J Rheumatol* 1993, **32**:903–907.

4. Jarvinen P, Aho K: **Twin studies in rheumatic diseases.** *Semin Arthritis Rheum* 1994, 24:19–28.
5. Risch N: **Linkage strategies for genetically complex traits. I. Multi-locus models.** *Am J Hum Genet* 1990, 46:222–228.
6. Vyse TJ, Todd JA: **Genetic analysis of autoimmune disease.** *Cell* 1996, 85:311–318.
7. Seldin MF, Amos CI, Ward R, Gregersen PK: **The genetics revolution and the assault on rheumatoid arthritis.** *Arthritis Rheum* 1999, 42:1071–1079.
8. Silman AJ. **Rheumatoid arthritis.** In: *Epidemiology of the Rheumatic Diseases*. Edited by Silman AJ, Hochberg MC. Oxford: Oxford University Press, 1993:7–68.
9. Weyand CM, Hicok KC, Conn DL, et al: **The influence of HLA-DRB1 genes on disease severity in rheumatoid arthritis.** *Ann Intern Med* 1992, 117:801–806.
10. Criswell LA, Mu H, Such CL, King MC: **Inheritance of the shared epitope and long-term outcomes of rheumatoid arthritis among community-based Caucasian females.** *Genet Epidemiol* 1998, 15: 61–72.
11. Stastny P: **Association of the B-cell alloantigen DRw4 with rheumatoid arthritis.** *N Engl J Med* 1978, 298:869–871.
12. Cornelis F, Faure S, Martinez M, et al: **New susceptibility locus for rheumatoid arthritis suggested by a genome-wide linkage study.** *Proc Natl Acad Sci USA* 1998, 95:10746–10745.
13. Jawaheer D, Costello T, Amos C, et al: **Analysis of affected sibling pairs with rheumatoid arthritis: The North American Rheumatoid Arthritis Consortium [abstract].** *Am J Hum Genet* 1999, 65:A276.
14. Firestein GS. **Etiology and pathogenesis of rheumatoid arthritis.** In: *Textbook of Rheumatology*, 5th ed. Philadelphia: W.B. Saunders & Co, 1997:851–897.
15. Symmons DP, Bankhead CR, Harrison BJ, et al: **Blood transfusion, smoking, and obesity as risk factors for the development of rheumatoid arthritis: results from a primary care-based incident case-control study in Norfolk, England.** *Arthritis Rheum* 1997, 40: 1955–1961.
16. Han Z, Boyle DL, Shi Y, Green DR, Firestein GS: **Dominant-negative p53 mutations in rheumatoid arthritis.** *Arthritis Rheum* 1999, 42: 1088–1092.
17. Panning B, Jaenisch R: **RNA and the epigenetic regulation of X chromosome inactivation.** *Cell* 1998, 93:305–308.
18. Monteiro J, Derom C, Vlietink R, et al: **Commitment to X inactivation precedes the twinning event in monozygotic MZ twins.** *Am J Hum Genet* 1998, 63:339–346.
19. Gregersen PK: **Discordance for autoimmunity in monozygotic twins: are 'identical' twins really identical?** *Arthritis Rheum* 1993, 36:1185–1192.
20. Hollander GA, Zuklys S, Morel C, et al: **Monoallelic expression of the interleukin-2 locus.** *Science* 1998, 279:2118–2121.
21. Riviere I, Sunshine ML, Littman DR: **Regulation of IL-4 expression by activation of individual alleles.** *Immunity* 1998, 9:217–228.
22. Tycko B: **Genomic imprinting: mechanism and role in human pathology.** *Am J Pathol* 1994, 144:431–443.
23. Ohlsson R, Tycko B, Sapienza C: **Monoallelic expression: 'there can only be one'.** *Trends Genet* 1998, 14:435–438.
24. Stewart JJ: **The female X-inactivation mosaic in systemic lupus erythematosus.** *Immunol Today* 1998, 19:352–357.
25. Gregersen PK, Chitnis S, Monteiro J, Salmon J: **Increased X-inactivation skewing in SLE? [Letter].** *Immunol Today* 1999, 20:152
26. Manning TC, Kranz DM: **Binding energetics of T-cell receptors: correlation with immunological consequences.** *Immunol Today* 1999, 20:417–422.
27. Gregersen PK, Silver J, Winchester RJ: **The shared epitope hypothesis: an approach to understanding the molecular genetics of rheumatoid arthritis susceptibility.** *Arthritis Rheum* 1987, 30:1205–1213.
28. Nepom GT: **Major histocompatibility complex-directed susceptibility to rheumatoid arthritis.** *Adv Immunol* 1998, 68:315–332.
29. Walsler-Kuntz DR, Weyand CM, Weaver AJ, O'Fallon WM, Goronzy JJ: **Mechanisms underlying the formation of the T cell receptor repertoire in rheumatoid arthritis.** *Immunol* 1995, 597–605.
30. Auger I, Escola JM, Gorvel JP, Roudier J: **HLA-DR4 and HLA-DR10 motifs that carry susceptibility to rheumatoid arthritis bind 70-kD heat shock proteins.** *Nature Med* 1996, 2:306–310.
31. La Cava A, Nelson JL, Ollier WE, et al: **Genetic bias in immune responses to a cassette shared by different microorganisms in patients with rheumatoid arthritis.** *Clin Invest* 1997, 100:658–663.
32. Dizier M-H, Eliaou J-F, Babron M-C, et al: **Investigation of the HLA component involved in rheumatoid arthritis (RA) by using the marker association segregation  $\chi^2$  (MASC) method: rejection of the unifying shared epitope hypothesis.** *Am J Hum Genet* 1993, 53:715–721.
33. Thomson W, Harrison B, Ollier WER, et al: **Quantifying the exact role of HLA-DRB1 alleles in susceptibility to inflammatory polyarthritis.** *Arthritis Rheum* 1999, 42:757.
34. Nepom B, Nepom GT, Schaller J, et al: **Characterization of specific HLA-DR4 associated histocompatibility molecules in patients with juvenile rheumatoid arthritis.** *J Clin Invest* 1984, 74:287–291.
35. Wordsworth P, Pile KD, Buckely JD, et al: **HLA heterozygosity contributes to susceptibility to rheumatoid arthritis.** *Am J Hum Genet* 1992, 51:585–591.
36. Meyer JM, Han J, Singh R, Moxley G: **Sex influences on the penetrance of HLA shared-epitope genotypes for rheumatoid arthritis.** *Am J Hum Genet* 1996, 58:371–383.
37. Zanelli E, Gonzalez-Gay MA, David CS: **Could HLA-DRB1 be the protective locus in rheumatoid arthritis?** *Immunol Today* 1995, 16: 274–278.
38. Price P, Witt C, Allcock R, et al: **The genetic basis for the association of the 8.1 ancestral haplotype (A1, B8, DR3) with multiple immunopathological diseases.** *Immunol Rev* 1999, 167:257–274.
39. Mulcahy B, Waldron-Lynch F, McDermott MF, et al: **Genetic variability in the tumor necrosis factor-lymphotoxin region influences susceptibility to rheumatoid arthritis.** *Am J Hum Genet* 1996, 59: 676–683.
40. Mu H, Chen JJ, Jiang Y, et al: **Tumor necrosis factor  $\alpha$  microsatellite polymorphism is associated with rheumatoid arthritis severity through an interaction with the HLA-DRB1 shared epitope.** *Arthritis Rheum* 1999, 42:438–442.
41. John S, Hajeer A, Marlow A, et al: **Investigation of candidate disease susceptibility genes in rheumatoid arthritis: principles and strategies.** *J Rheumatol* 1997, 24:199–201.
42. Crawley E, Kay R, Sillibourne J, et al: **Polymorphic haplotypes of the interleukin-10 5' flanking region determine variable interleukin-10 transcription and are associated with particular phenotypes of juvenile rheumatoid arthritis.** *Arthritis Rheum* 1999, 42:1101–1108.
43. Vafiadis P, Bennett ST, Todd JA, et al: **Insulin expression in human thymus is modulated by INS VNTR alleles at the IDDM2 locus.** *Nature Genet* 1997, 15:289–292.
44. Halushka MK, Fan JB, Bentley K, et al: **Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis.** *Nature Genet* 1999, 22:239–247.
45. Cargill M, Altshuler D, Ireland J, et al: **Characterization of single-nucleotide polymorphisms in coding regions of human genes.** *Nature Genet* 1999, 22:231–238.
46. Chakravarti A: **It's raining SNPs, hallelujah?** *Nature Genet* 1998, 19: 216–217.
47. Mu H, Charnley P, King MC, Criswell LA: **Synergy between T cell receptor beta gene polymorphism and HLA-DR4 in susceptibility to rheumatoid arthritis.** *Arthritis Rheum* 1996, 39:931–937.
48. Cornelis F, Hardwick L, Flipo RM, et al: **Association of rheumatoid arthritis with an amino acid allelic variation of the T cell receptor.** *Arthritis Rheum* 1997, 40:1387–1390.
49. Ibberson M, Peclat V, Guerne PA, et al: **Analysis of T cell receptor V alpha polymorphisms in rheumatoid arthritis.** *Ann Rheum Dis* 1998, 57:49–51.
50. McDermott M, Kastner DL, Holloman JD, et al: **The role of T cell receptor beta chain genes in susceptibility to rheumatoid arthritis.** *Arthritis Rheum* 1995, 38:91–95.
51. Cantagrel A, Navaux F, Loubet-Lescoulie P, et al: **Interleukin-1beta, interleukin-1 receptor antagonist, interleukin-4, and interleukin-10 gene polymorphisms: relationship to occurrence and severity of rheumatoid arthritis.** *Arthritis Rheum* 1999, 42:1093–100.
52. Hajeer AH, Lazarus M, Turner D, et al: **IL-10 gene promoter polymorphisms in rheumatoid arthritis.** *Scand J Rheumatol* 1998, 27: 142–145.
53. Lim S, Crawley E, Woo P, Barnes PJ: **Haplotype associated with low interleukin-10 production in patients with severe asthma [Letter].** *Lancet* 1998, 352:113.
54. Seidl C, Donner H, Fischer B, et al: **CTLA4 codon 17 dimorphism in patients with rheumatoid arthritis.** *Tissue Antigens* 1998, 51:62–66.
55. Gonzalez-Escribano MF, Rodriguez R, Valenzuela A, et al: **CTLA-4 polymorphisms in Spanish patients with rheumatoid arthritis.** *Tissue Antigens* 1999, 53:296–300.
56. Van der Auwera BJ, Vandewalle CL, Schuit FC, et al: **CTLA-4 gene polymorphism confers susceptibility to insulin-dependent diabetes mellitus (IDDM) independently from age and from other**

- genetic or immune disease markers. The Belgian Diabetes Registry. *Clin Exp Immunol* 1997, **110**:98–103.
57. Donner H, Rau H, Walfish PG, et al: **CTLA4 alanine-17 confers genetic susceptibility to Graves' disease and to type 1 diabetes mellitus.** *J Clin Endocrinol Metab* 1997, **82**:143–146.
  58. Baerwald CG, Panayi GS, Lanchbury JS: **Corticotropin releasing hormone promoter region polymorphisms in rheumatoid arthritis.** *J Rheumatol* 1997, **24**:215–216.
  59. Matthey DL, Hassell AB, Plant M, et al: **Association of polymorphism in glutathione S-transferase loci with susceptibility and outcome in rheumatoid arthritis: comparison with the shared epitope.** *Ann Rheum Dis* 1999, **58**:164–168.
  60. John S, Marlow A, Hajeer A, et al: **Linkage and association studies of the natural resistance associated macrophage protein 1 (NRAMP1) locus in rheumatoid arthritis.** *J Rheumatol* 1997, **24**:452–457.
  61. Gregersen PK: **The North American Rheumatoid Arthritis Consortium: bringing genetic analysis to bear on disease susceptibility, severity, and outcome.** *Arthritis Care Res* 1998, **11**:1–2.
  62. Hay EM, Ollier WE, Silman AJ: **The Arthritis and Rheumatism Council's national family material repository.** *Br J Rheumatol* 1993, **32**:443–444.
  63. Risch N, Merikangas K: **The future of genetic studies of complex human diseases.** *Science* 1996, **273**:1516–1576.
  64. Spiegel TM, Spiegel JS, Paulus HE: **The joint alignment and motion scale: a simple measure of joint deformity in patients with rheumatoid arthritis.** *J Rheumatol* 1987, **14**:887–892.
  65. Kuska B: **Snipping 'SNPs': a new tool for mining gene variations.** *J Natl Cancer Inst* 1999, **91**:1110.
  66. Wilson AG, Symons JA, McDowell TL, McDevitt HO, Duff GW: **Effects of a polymorphism in the human tumor necrosis factor alpha promoter on transcriptional activation.** *Proc Natl Acad Sci USA* 1997, **94**:3195–3199.
  67. Nickerson DA, Taylor SL, Weiss KM, et al: **DNA sequence diversity in a 9.7-kb region of the human lipoprotein lipase gene.** *Nature Genet* 1998, **19**:233–240.
  68. Long AD, Langley CH: **The power of association studies to detect the contribution of candidate genetic loci to variation in complex traits.** *Genome Res* 1999, **9**:720–731.
  69. Holloway JW, Beghe B, Turner S, et al: **Comparison of three methods for single nucleotide polymorphism typing for DNA bank studies: sequence-specific oligonucleotide probe hybridisation, TaqMan liquid phase hybridisation, and microplate array diagonal gel electrophoresis (MADGE).** *Hum Mutat* 1999, **14**:340–347.
  70. Griffin TJ, Hall JG, Prudent JR, Smith LM: **Direct genetic analysis by matrix-assisted laser desorption/ionization mass spectrometry.** *Proc Natl Acad Sci USA* 1999, **96**:6301–6306.
  71. Duggan DJ, Bittner M, Chen Y, Meltzer P, Trent JM: **Expression profiling using cDNA microarrays.** *Nature Genet* 1999, **21** (1 suppl):10–14.
  72. Frankel WN, Schork NJ: **Who's afraid of epistasis?** *Nature Genet* 1996, **14**:371–373.
  73. Morel L, Tian XH, Croker BP, Wakeland EK: **Epistatic modifiers of autoimmunity in a murine model of lupus nephritis.** *Immunity* 1999, **11**:131–139.
  74. Cucca F, Goy JV, Kawaguchi Y, et al: **A male–female bias in type 1 diabetes and linkage to chromosome Xp in MHC HLA-DR3-positive patients.** *Nature Genet* 1998, **19**:301–302.
  75. Becker KG, Simon RM, Bailey-Wilson JE, et al: **Clustering of non-major histocompatibility complex susceptibility candidate loci in human autoimmune diseases.** *Proc Natl Acad Sci USA* 1998, **95**:9979–9984.
  76. Dorman J, Kramer MK, O'Lear LA, et al: **Molecular epidemiology of autoimmune thyroid disease.** *Gac Med Mex* 1997, **133** (suppl 1): 97–103.
  77. Dorman JS. **Epidemiology of autoimmune disease.** In: *Immunologically Mediated Endocrine Diseases*. Edited by Harmon J, Gill R, MacLaren N (in press).
  78. Lin JP, Cash JM, Doyle SZ, et al: **Familial clustering of rheumatoid arthritis with other autoimmune diseases.** *Hum Genet* 1998, **103**: 475–482.
  79. Ginn LR, Lin JP, Plotz PH, et al: **Familial autoimmunity in pedigrees of idiopathic inflammatory myopathy patients suggests common genetic risk factors for many autoimmune diseases.** *Arthritis Rheum* 1998, **41**:400–405.
  80. Torfs CP, King MC, Huey B, Malmgren J, Grumet FC: **Genetic interrelationship between insulin-dependent diabetes mellitus, the autoimmune thyroid diseases, and rheumatoid arthritis.** *Am J Hum Genet* 1986, **38**:170–187.
  81. Dyson F: *The Sun, the Genome, and the Internet*. New York: Oxford University Press, 1999.

**Author address:** Division of Biology and Human Genetics, North Shore University Hospital, Manhasset, New York, USA

**Correspondence:** Peter K Gregersen, MD, Division of Biology and Human Genetics, North Shore University Hospital, 350 Community Drive, Manhasset, NY 11030, USA. Tel: +1 516 562 1134, fax: +1 516 562 1683, e-mail: peterg@nshs.edu