

## REVIEW ARTICLE

# New approaches to investigating heterogeneity in complex traits

R Bompreszi, P E Kovanen, R Martin

*J Med Genet* 2003;40:553–559

Great advances in the field of genetics have been made in the last few years. However, resolving the complexity that underlies the susceptibility to many polygenic human diseases remains a major challenge to researchers. The fast increase in availability of genetic data and the better understanding of the clinical and pathological heterogeneity of many autoimmune diseases such as multiple sclerosis, but also Parkinson's disease, Alzheimer's disease, and many more, have changed our views on their pathogenesis and diagnosis, and begins to influence clinical management. At the same time, more powerful methods that allow the analysis of large numbers of genes and proteins simultaneously open opportunities to examine their complex interactions. Using multiple sclerosis as a prototype, we review here how new methodologies such as gene expression profiling can be exploited to gain insight into complex trait diseases.

number of aetiologically and pathogenetically distinct disorders, but it would also be unlikely to be successful. Similar phenotypes may be produced by different genes in the same pathways as well as by completely unrelated causes. Tumours may in some cases arise from exposure to environmental agents (for example, asbestosis),<sup>4</sup> while in other instances single gene mutations are of primary importance (for example, retinoblastoma).<sup>5</sup> By analogy, it appears imperative to identify biomarkers that would allow a stratification of cases into homogeneous classes for the purpose of dissecting the complexity of the multifactorial diseases.<sup>6</sup> Subsequently, the analyses performed in the so stratified groups will have much greater chances to show the mechanisms in common to each category. In particular, researchers have applied new techniques to achieve a novel classification of apparently homogeneous groups of tumours, thus permitting new insight into the pathogenesis of some malignancies.<sup>7–9</sup> In this review we are referring mainly to multiple sclerosis (MS) as well as to some other autoimmune and neoplastic conditions as models in support of the aforementioned point.

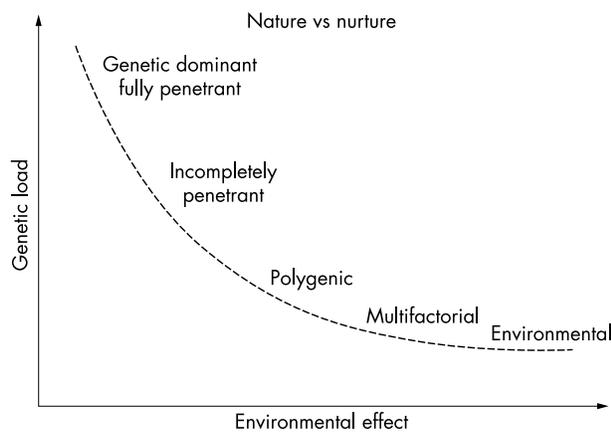
In the view of classical genetics, diseases are divided into Mendelian disorders and complex traits. While the former are attributed to single gene mutations with a simple mode of inheritance, the latter are thought to result from multiple genes, each playing a small and interactive role in the susceptibility to the diseases. From a clinical point of view a continuum of phenotypic presentations may be observed. At one end of the spectrum are disorders caused by fully penetrant deleterious mutations and on the opposite end are environmental diseases. Between these two extremes lie the incompletely penetrant and the polygenic disorders, creating a smooth transition from strictly genetic to multifactorial illnesses (fig 1).<sup>1</sup> The identification of the causative genes for Mendelian disorders has been a major undertaking for researchers, in most cases requiring many years of investigations, for example, tuberous sclerosis.<sup>2</sup> For obvious reasons, the characterisation of the genetics underlying the complex traits poses substantially greater challenges.<sup>3</sup> However, as these are far more common than the Mendelian disorders, tackling the complex traits has substantial socioeconomic impact. Heterogeneity is the common denominator to all of them and can be considered the most important hurdle to overcome. As an example, thinking of cancer as a condition resulting from an alteration of cell cycle control mechanisms would not only oversimplify and unjustly pool together an immense

## THE EARLY STEPS IN MOLECULAR GENETIC INVESTIGATIONS

Before the emergence of the new molecular techniques, epidemiology was the means to evaluate the relative weight of genetic versus environmental factors, and its application to the problem remains very valuable,<sup>9</sup> as shown by our example, multiple sclerosis.<sup>10</sup> A genetic predisposition to this chronic debilitating neurological illness was initially suspected based on a fairly large number of observations both at the population level and by familial aggregation studies.<sup>11</sup> In this context, twin surveys are considered an important source of information.<sup>12–13</sup> However, the concordance rate in monozygotic twins should only be viewed as the upper limit of the genetic influence on the disease susceptibility, as underscored by the fact that the concordance rate for tuberculosis, an environmental disease, is also higher in monozygotic than in dizygotic twins.<sup>14</sup> In MS, the concordance in monozygotic twins varies between <10% to about 30% across different populations.<sup>15–18</sup> This is still far from 100%, a figure only expected yet not always observed in monogenic highly penetrant traits.<sup>19–20</sup> Overall, the epidemiological studies in MS suggest that genetics is relevant, but also that the environment exerts a key influence on its phenotypic expression.<sup>21</sup> In any event, the numerous surveys performed in MS over the course of the last half a century have prompted major efforts of exploiting

See end of article for authors' affiliations

Correspondence to:  
Dr R Bompreszi, Cancer Genetics Branch, National Human Genome Research Institute, National Institutes of Health, 50 South Drive, Room 5150, Bethesda, Maryland 20892, USA; rbompresz@nhgri.nih.gov



**Figure 1** The progressive decrease in the genetic load contributing to the development of a disease creates a smooth transition in the distribution of illnesses on an aetiological diagram. In theory there are no diseases completely free from the influence of both genetic and environmental factors.

the new genotyping techniques to assess linkage in large numbers of MS multiplex families. The eight whole genome screens completed to date<sup>10</sup> and a meta-analysis of the data<sup>22</sup> have provided the concept that MS represents a multifactorial disorder with high levels of complexity. Moreover, linkage studies have proved to be insufficiently powered to detect the small effects of the many genes involved in polygenic disorders, and, while significantly larger studies could be more informative, they are difficult to perform even in disorders that occur at considerable frequencies.<sup>23, 24</sup> In the case of MS the prevalence is up to about 100/100 000 subjects. Currently, researchers have moved on to testing parents to address whether the susceptibility allele is preferentially transmitted to their single affected offspring at a higher percentage rate than the 50% expected for a neutral allele, according to the Transmission Disequilibrium Test (TDT).<sup>25</sup> Many regions of interest (1p, 6p21, 12p, 17q, 19q3) have been identified in MS using this approach.<sup>26</sup> More importantly, however, the large amount of data collected by linkage and TDT underscores the relevance of genetic heterogeneity in MS, identifying it as a key point to be addressed.<sup>27</sup>

Consistent with this view are the findings by Jacobsen *et al.*<sup>28, 29</sup> In a molecular genetic investigation in a German population, these researchers have described a strong association between MS and a point mutation at position 77 of exon 4 of the gene encoding the protein-tyrosine phosphatase, receptor type C (PTPRC), also known as CD45. This polymorphism causes an alternative splicing of the mRNA of this membrane phosphatase, thus altering its isoform expression in immune cells. Heterozygosity for the mutation was initially reported both in linkage and association with the disease in three MS nuclear families.<sup>28</sup> More recently the same authors have identified in a fourth German multiplex family another mutation (59 C-A transversion) in exon 4 of the *PTPRC* gene, responsible for a non-conservative amino acid substitution, which affects maturation, activation, and migration of immune cells.<sup>29</sup> The fact that some healthy first degree relatives of the index cases in those families carry the mutation, in addition to the lack of association between the CD45 polymorphism and MS in two case-control studies in North American and Swedish populations,<sup>30, 31</sup> together point to a basic principle of complex traits: none of the involved genes is per se either necessary or sufficient for the development of the disease.<sup>32</sup>

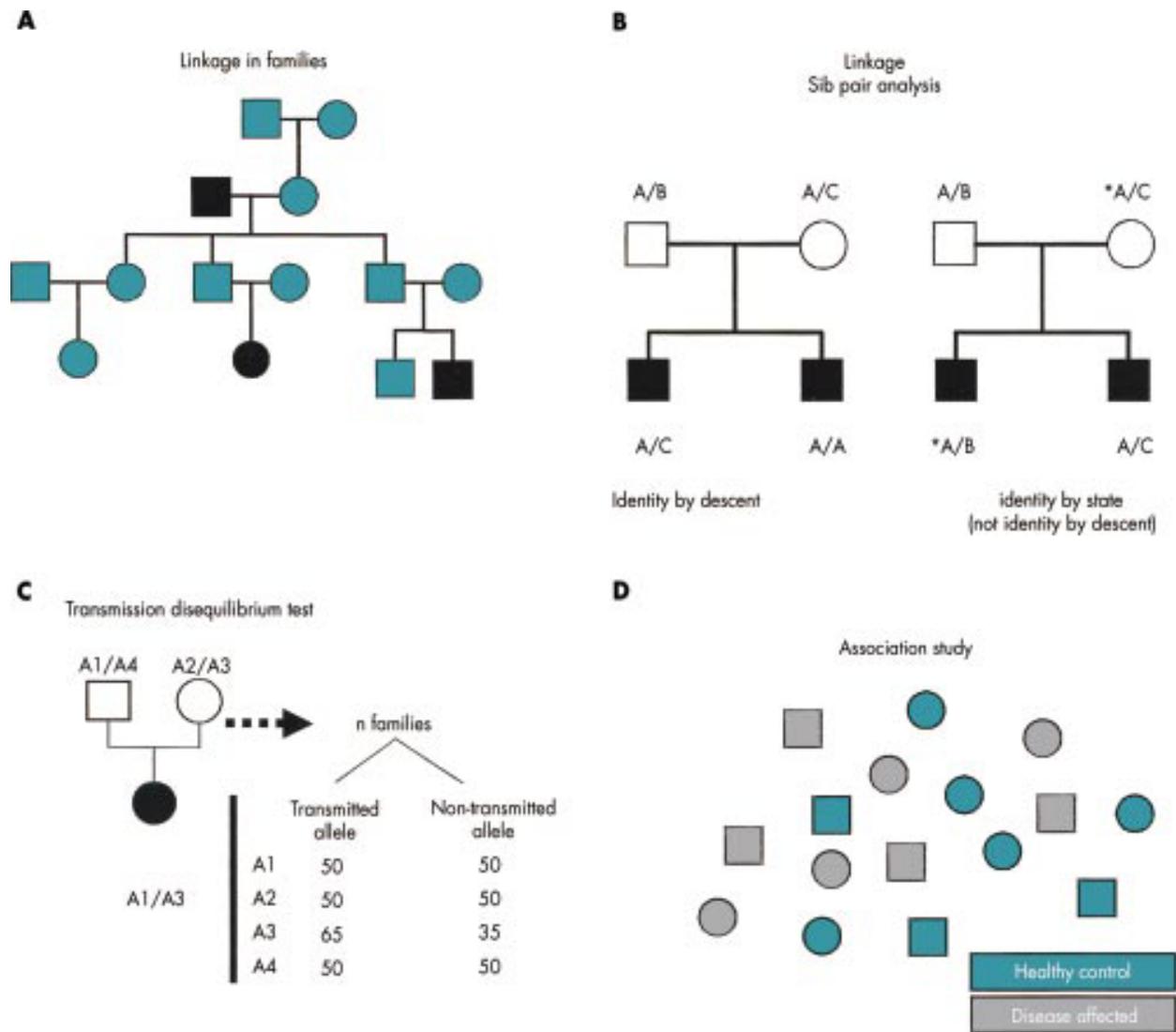
It is not surprising then that numerous case-control studies seeking association to candidate genes performed in MS as well as in many other diseases have often yielded contradictory results.<sup>33</sup>

A further instance with similar conclusions to MS is provided by the search for genetic causes of Parkinson's disease (PD). The identification of a mutation in the alpha-synuclein gene through linkage analysis in large, multi-generational kindreds<sup>34</sup> only accounts for a rare autosomal dominant form of PD, and the genetic heterogeneity creates a situation analogous to Alzheimer's disease where mutations in the presenilin 1 and 2 genes are only responsible for a small minority of the cases.<sup>35</sup> Hence, we believe that new approaches are needed to complement this research. Fig 2 offers a schematic view of the analytical methods mentioned in the paragraph.

### SEEKING "ASSOCIATION" IN AN EXCITING NEW ERA FOR RESEARCH IN GENETICS

Investigators have been puzzled by the question of why two subjects with a similar genetic background would respond so differently, that is, one develops an often serious disorder, while the other remains healthy. In other words, is the major determinant of complex traits a predisposing genetic factor or is an abnormal reaction to external stimuli, acting on epigenetic and stochastic events, more important? This problem is addressed by examining differences between a group of affected subjects and a properly matched healthy control group. These are association type investigations (case-control studies) and are often based on selected hypothesis driven candidate genes.<sup>36</sup> However, the interpretation of the results derived from these studies demands caution to avoid false positive and spurious results.<sup>37</sup> Yet association studies remain a powerful approach.<sup>38</sup> In particular, it is the recent availability of information from the human genome project that has started an era of tremendous opportunities. Initially, it led to the use of multiple tandem repeats (or microsatellites) as markers for disease susceptibility regions, and is now moving forwards to use single nucleotide polymorphisms (SNPs).<sup>39</sup> As the SNPs are evenly distributed over the genome and are far more abundant than the microsatellites, efforts are under way to complete the map of the SNPs in the human genome.<sup>40</sup> Moreover, haplotypes of SNPs in humans may represent a substantial shortcut in the search for susceptibility genes over the single marker based approaches (fig 3).<sup>41</sup> But what is the rationale behind the use of the SNPs in the search for the susceptibility to diseases? The idea is that changes of single base pairs in genomic DNA occur at a predictable rate, that is, the younger a population the fewer the polymorphisms. However, some hot spots in the genome appear to be remarkably more inclined to bear polymorphisms than other more stable areas that are well preserved even throughout many species.<sup>42, 43</sup> While a deleterious mutation would abrogate a gene function, other more "benign" changes in the coding as well as in the non-coding part of a gene may account for a variation in the functional level of the gene product. It is conceivable that the genetic polymorphisms and their multiple interactions are the basis for the quantitative characters<sup>44</sup>; they underlie the biodiversity and are therefore attributable to the success of adaptation. Nevertheless, they represent those common variations in the human genome that may affect the predisposition to a disease, and it is worth mentioning that what may be an advantageous profile in one environment, may become unfavourable or uncontrolled in a different setting. Hence, the goal of an association study is to identify those polymorphisms that only exist in an affected population.

Particularly emblematic of complex traits is Hirschsprung disease, or aganglionic megacolon. Several loci have been implicated in the susceptibility to this disease, characterised by the congenital absence of ganglia in various portions of the intestine. Interestingly, distinct genes with diverse mode of inheritance (from dominant to recessive) as well as a full spectrum of degrees of penetrance have been identified by linkage in families and subsequent mapping. Epistatic



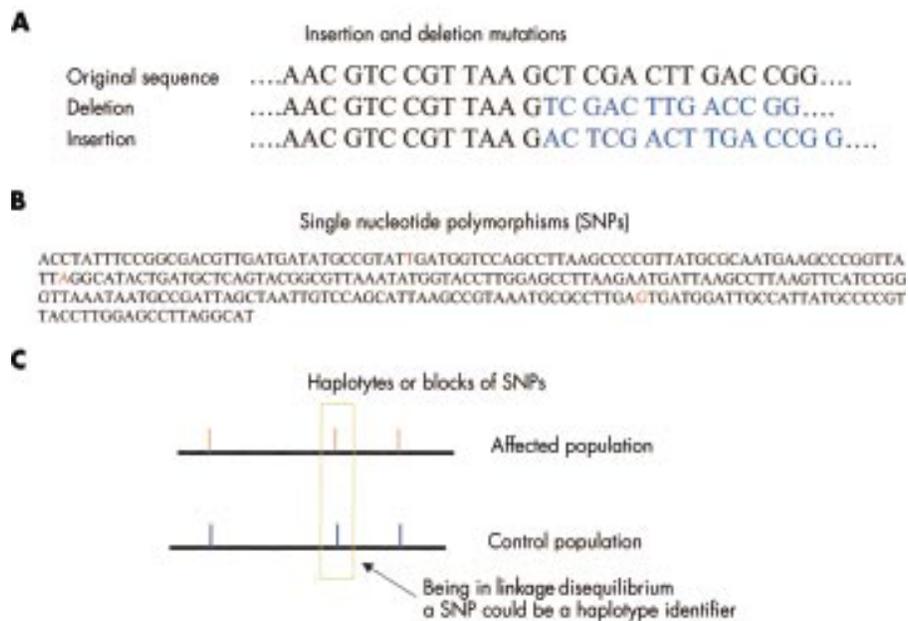
**Figure 2** (A) Linkage analysis, aimed at identifying genetic markers that are coinherited with disease status within a multigeneration family, has largely been successful in Mendelian disorders. (B) Sib pair analysis, extensively used in the 90s to investigate the genetics of complex traits, is designed to show alleles that are shared by two or more affected subjects in the same family. (C) The transmission disequilibrium test looks for an allele of a selected marker that is preferentially transmitted to an affected subject by his/her unaffected parents. To be informative the marker has to be polymorphic. (D) Association studies are performed by assessing the statistically significant recurrence of an allele in an affected population in comparison to a control population.

interactions between genes have also been established to take place, creating a complicated picture of a multiform oligogenic trait that paradigmatically highlights the above notions.<sup>45</sup>

In regard to the search for polymorphisms responsible for the susceptibility to adverse conditions, one of the most sophisticated examples is given by the immune system whose function is under the control of many interactive genes and environmental factors. In fact the human immune system has evolved over a long period of time, constantly challenged and shaped by external agents. Sudden and dramatic changes in the environment have raised problems of adaptation, as indicated by the steep increase of allergies and autoimmune disorders observed in industrialised countries during the last few decades. According to the hygiene hypothesis, an imbalance in the cytokine production by different T lymphocyte subclasses accounts for the higher incidence of allergies.<sup>46</sup> Or, as a more recent theory proposes, increases in both allergy and autoimmunity are attributed to a dysfunction of innate immunity.<sup>47–50</sup> While the true causes are not clear at present, sudden changes (certainly remarkable on the

evolutionary scale) in daily challenges that the immune system faces in modern societies are possibly responsible for these pathologies, thus the hypothetical point of a lack of adaptation remains valid.

Heterogeneity is a major obstacle to the resolution of the aetiologies of complex traits. One of the best arguments in favour of MS being an autoimmune disease is its linkage to the major histocompatibility complex (MHC, HLA in humans) class II region and association with HLA-DR and -DQ alleles.<sup>51–53</sup> It has been estimated that about 10% of the overall genetic susceptibility to MS is conferred by a gene or genes in this area, thus leaving a substantial fraction of susceptibility to be explained by other genes and factors.<sup>54</sup> The abovementioned investigations have shown that MS is indeed a polygenic trait. To address the issue of the interaction between genes, Coraddu *et al*<sup>55</sup> used data from a genome wide scan, in this instance stratifying the patients according to HLA status, and repeating the linkage analysis.<sup>56</sup> This step resulted in significant changes in linkage scores. Some chromosomal regions that had appeared earlier were not evident any more, whereas



**Figure 3** As opposed to base pair insertions and deletions that cause frameshift mutations in a DNA stretch (A), SNPs are single base pair changes that do not necessarily alter the amino acid sequence in the relative gene product, that is, the protein. SNPs (B) recur roughly once in every 1300 bp, and are considered useful markers for genotyping purposes. Haplotypes or blocks of SNPs may be identified in genomic regions where recombination occurs at a low rate and linkage disequilibrium is conserved (C).

new ones emerged. They speculate that these alterations correspond to genes that interact with the MHC region in conferring genetic susceptibility to MS. Furthermore, they conducted an association study in which they observed a correlation between female gender and age of onset and HLA-DR2, but no other features.<sup>57</sup> Many other groups have also looked for a correlation between disease phenotype and surrogate markers with partial success.<sup>58–60</sup> In fact, the relationship between genotype and phenotype is not always immediately evident, owing to the existence of both genetic (different genes-identical phenotypes) and disease heterogeneity (identical genes-different phenotypes).<sup>61</sup> The example of the autoimmune lymphoproliferative syndrome (ALPS) serves to illustrate the case. This disease is characterised by the massive proliferation of germinal centre lymphocytes owing to a defect in the pro-apoptotic signalling molecule FAS (ALPS1a) or its ligand (ALPS1b).<sup>62–65</sup> In addition to and as one consequence of the excessive lymphocyte proliferation, the syndrome has autoimmune characteristics, such as manifestations of haemolytic anaemia and thrombocytopenia.<sup>64</sup> Genetic heterogeneity is observed when considering ALPS2, a syndrome caused by mutations in the Caspase10 gene, whose product acts along the same pathway downstream of FAS.<sup>65</sup> Thus, ALPS provides the case for a multiform disorder in which similar clinical features are determined by single and distinct gene mutations. In addition, even though ALPS may not be considered a classical autoimmune disease, it offers important insights into the pathogenetic mechanisms underlying the breakdown of tolerance.<sup>66</sup>

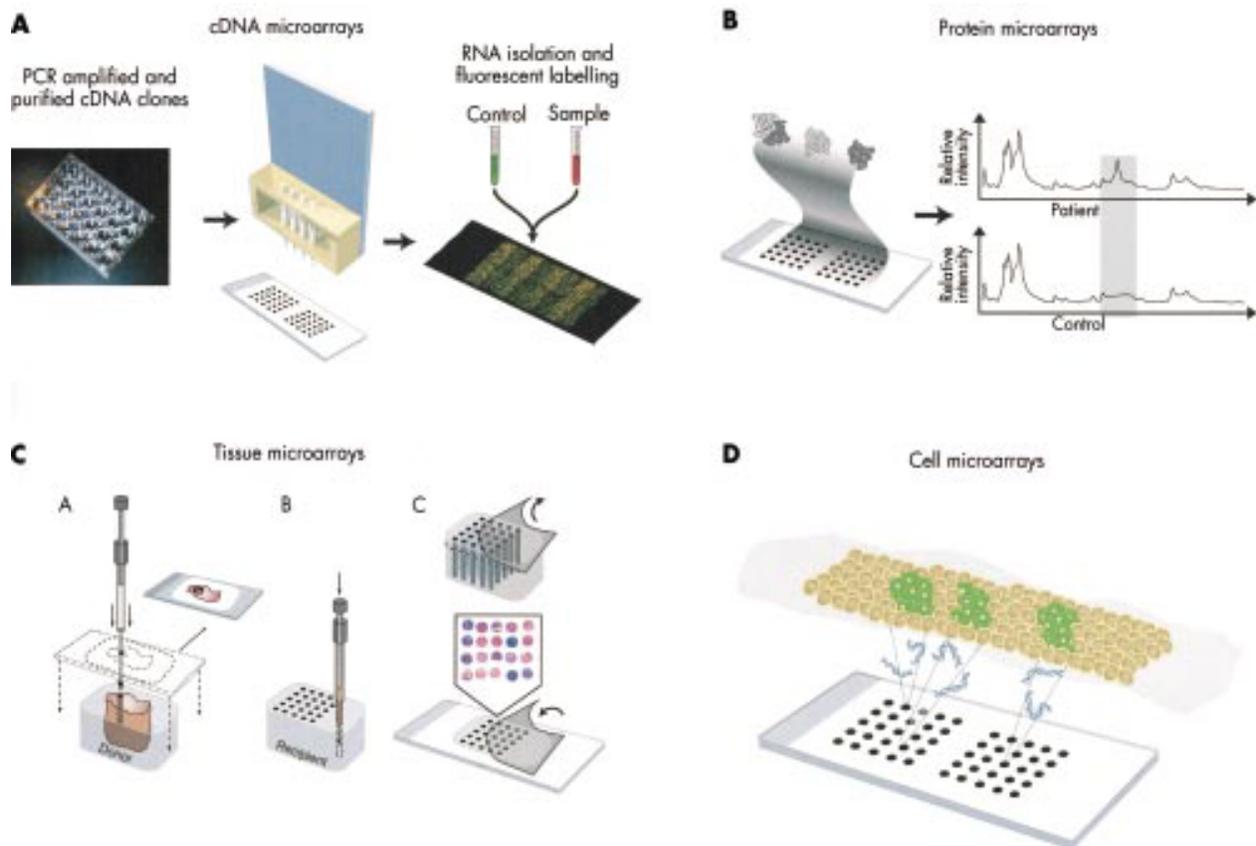
The possibility that excessive lymphocyte proliferation leads to autoimmunity has recently been further explored in animal models of systemic lupus erythematosus (SLE). By knocking out either the Gadd45a, a p53 effector gene, or the p21 cell cycle inhibitor, researchers have generated mice that show features typical of SLE.<sup>67–68</sup> Similarly to ALPS, the phenotypes in these animals involve deregulated cell proliferation as a result of the loss of functions of Gadd45 and p21 proteins. The possible explanation for the autoimmune component observed in these models is an increased number of autoreactive lymphocytes. Small numbers of autoreactive cells are part of the healthy immune system,<sup>69–70</sup> but an excess can overcome

the peripheral control mechanisms, rendering the occurrence of the breakdown of tolerance much more probable. Interestingly, female mice defective for either Gadd45a or p21 are significantly more prone to an aggressive and early onset of disease than their male counterparts, a fact that is consistent with the observation of the predominance of many human autoimmune disorders in the female gender, for example, MS. It also underscores that a single gene mutation requires interaction with other factors (unique to females for instance) for the disease to manifest or at least to contribute to its development, and more data of distinct patterns between genders are starting to be generated.<sup>71</sup>

In light of the above studies, it is clear that cell cycle progression regulators need to be taken into consideration in the context of autoimmune diseases, just as genes controlling lymphocyte activation, cell-cell interaction, or compartment trafficking have been in the past.<sup>72</sup> Recent findings by Maas *et al*<sup>73</sup> in various groups of autoimmune disorders further support this prospect. Accordingly, the involvement of the immune system in a multifactorial trait may complicate the search for individual genes predisposing to that disease, as the regulation of the immune system itself is complex and incompletely understood.

### NEW TECHNOLOGIES PROVIDE NOVEL INSIGHTS

Understanding the specific events and their sequences occurring during a pathological process is important for many reasons, the ultimate being the identification of an aetiological rather than a symptomatic cure. It is our incomplete knowledge of the disease mechanisms that sets the limits to tailoring specific and effective treatments.<sup>74</sup> Yet we believe that the most powerful approach to identifying the causes of diseases is to study well defined categories of patients or subgroups within a disease spectrum. Above, we reviewed how a number of distinct pathogenetic mechanisms may determine apparently analogous conditions, and that defining groups of patients only on the basis of phenotypic characteristics may not suffice. Now, advances in biotechnology provide new tools to researchers at an extraordinary rate,<sup>75–80</sup> and it is through the combination of new techniques and alternative approaches that optimal results can be achieved (fig 4). It has



**Figure 4** (A) Expression microarray techniques exploit complementary hybridisation, which is a basic property of the nucleic acids. Large numbers of DNA probes are deposited in high density on glass slides or membranes. These probes are either robotically spotted PCR amplified products of cloned cDNAs or synthesised oligonucleotides, or alternatively on glass synthesised oligonucleotides (such as Affymetrix Genechips). The RNA sample to be tested (or the genomic DNA in the case of the CGH) can be labelled with various methods and is assayed onto the microarray platform. Only the spots containing the specific sequence present in the samples will “light up” proportionally to the amount in the original sample. (B) Protein arrays are designed to detect the presence in a given biological fluid of proteins that may interact with those deposited on the array, probes. Probes can be antigens, antibodies, other binding proteins, or DNA with a specific protein binding motif. In most cases the identity of the positive spots, whose readout is expressed in terms of peaks, remains unknown. (C) In the tissue microarray, small tissue samples are punched from a pathology specimen embedded in a donor paraffin block (A) and arrayed into a recipient block together with a high number of other tissues to be examined in parallel (B). Subsequently, the recipient block is horizontally sliced and each slice is transferred onto the test slides (C) to yield a histological slide comprising up to 1000 samples in each. (D) The most recent evolution of the array based technology is the cell array. Here molecules such as antisense oligonucleotides or small interfering RNA (siRNA) are deposited onto glass slides. Living cells are then grown on the slides, resulting in gene silencing via reverse transfection that occurs in the area spotted with the gene specific target sequence. Functional assays such as reporter assays involving green fluorescent protein are used to identify the biologically relevant gene.

been shown that gene expression profiling via microarray experiments is capable of reclassifying apparently homogeneous groups of clinical entities.<sup>81–82</sup> Indeed, by interrogating the coordinated expression of large number of genes, this method has already provided new insights into disease mechanisms and led to envisioning strong clinical applications.<sup>83–84</sup> Furthermore, by linking gene expression data to existing databases containing functional annotations (such as GeneOntology at <http://www.godatabase.org/dev/database/>), it is possible to evaluate results from microarray experiments from a functional point of view. Identifying patient groups that are homogeneous at the molecular level, that is, by gene expression profiling, will be facilitated by the creation of large databases containing comprehensive expression data from a variety of tissues and pathological conditions. The importance of this latter goal has convinced and moved investigators and institutions to undertake major efforts to carry out the project and to make the databases immediately available to the public domain.<sup>85–88</sup>

The availability of microarray technology has led our group to study MS genetics by profiling gene expression in peripheral blood mononuclear cells of patients and healthy subjects in order to assess whether it allows the distinction of

patients from controls (unpublished data). The rationale behind the use of peripheral blood cells is in the possibility that primed and activated effector cells relevant to the pathogenesis of MS may circulate in the periphery, resulting in a better understanding of the pathogenesis of the disease, as recently suggested.<sup>89</sup> In line with previous studies by Ramanathan *et al*<sup>90</sup> and others,<sup>73</sup> we found that subtle differences in the level of expression of various immune and cell cycle related genes are detectable in the MS patients when compared to controls. However, as expected for a complex trait, we noticed that none of the genes works as a perfect discriminator of the two groups. Assuming that the gene expression reflects the genetic predisposition to the disease rather than being the altered response to an external agent, it is conceivable to explain our finding on the basis of the existence of genetic heterogeneity in the MS population. This would be in accord with observations in recent years that MS is a heterogeneous group of disorders in many aspects, including clinical presentation, course and severity, histopathological characteristics of lesions, magnetic resonance imaging findings, response to treatment, and immunological observations.<sup>51–91–92</sup> It has been shown that measuring multiple immunological

parameters, stratifying patients according to MRI characteristics, and incorporating clinical parameters allows the recognition of subgroups among MS patients.<sup>93-94</sup> This approach could be further improved by the combined use of microarrays, MRI scans, and in the near future proteomics data.

## CONCLUSIONS

The classification of complex traits into well defined categories is the prerequisite for understanding the underlying pathogenetic mechanisms. This can only be achieved by using the most sophisticated tools available applied to large numbers of samples. Techniques that simultaneously analyse as many genes or proteins as possible have already been put into practice, and as the collection of appropriate sets of samples is an essential step in this endeavour, a close interaction between the medical and the research community is crucial. To accomplish a successful search and to address the complexity of the multifactorial disorders a joint effort is required, particularly to assure a collection of reliable data before they can be used for patient management.<sup>95</sup> Like assessing the degree of polymorphism in a given population, which involves genotyping individual subjects,<sup>96</sup> we have to analyse distinct patient cohorts if we aim at characterising subsets of MS patients to define the disease phenotypes. Or, maybe, the more we comprehend about disease pathogeneses, the clearer it becomes that the need for classification is only ours not Mother Nature's.

## ACKNOWLEDGEMENTS

We would like to thank M Cichanowski of the Scientific Illustration Unit, NIH, for his invaluable help with the illustrations.

## Authors' affiliations

**R Bomprezzi**, Cancer Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA

**P E Kovanen**, Laboratory of Molecular Immunology, National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, Maryland 20892, USA

**R Martin**, Neuroimmunology Branch, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, Maryland 20892, USA

## REFERENCES

- 1 **Strachan T**, Read AP. Genes in pedigrees. In: *Human molecular genetics*. 2nd ed. New York: John Wiley, 1999.
- 2 **Verhoef S**, Bakker L, Tempelaar AM, Hesseling-Janssen AL, Mazurczak T, Jozwiak S, Fois A, Bartolini G, Zonnenberg BA, van Essen AJ, Lindhout D, Halley DJ, van den Ouweland AM. High rate of mosaicism in tuberous sclerosis complex. *Am J Hum Genet* 1999;**64**:1632-7.
- 3 **Threadgill DW**, Hunter KW, Williams RW. Genetic dissection of complex and quantitative traits: from fantasy to reality via a community effort. *Mamm Genome*. 2002;**13**:175-8.
- 4 **Roach HD**, Davies GJ, Athanosi R, Crane M, Adams H, Phillips S. Asbestos: when the dust settles an imaging review of asbestos-related disease. *Radiographics* 2002;**22**, special No:S167-84.
- 5 **Murphree AL**, Benedict WF. Retinoblastoma: clues to human oncogenesis. *Science* 1984;**223**:1028-33.
- 6 **Gould DB**, John SW. Anterior segment dysgenesis and the developmental glaucomas are complex traits. *Hum Mol Genet* 2002;**11**:1185-93.
- 7 **Bittner M**, Meltzer P, Chen Y, Jiang Y, Sefor E, Hendrix M, Radmacher M, Simon R, Yakhini Z, Ben-Dor A, Sampas N, Dougherty E, Wang E, Marincola F, Gooden C, Lueders J, Glatfelter A, Pollock P, Carpten J, Gillanders E, Leja D, Dietrich K, Beaudry C, Berens M, Alberts D, Sondak V, Hayward N, Trent J. Molecular classification of cutaneous malignant melanoma by gene expression profiling. *Nature* 2000;**406**:536-40.
- 8 **Khan J**, Wei JS, Ringner M, Saal LH, Ladanyi M, Westermann F, Berthold F, Schwab M, Antonescu CR, Peterson C, Meltzer PS. Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nat Med* 2001;**7**:673-9.
- 9 **John S**, Jenkins S. Third international meeting on the genetic epidemiology of complex traits, April 4-6, 2002, Cambridge, UK. *Hum Mutat* 2002;**20**:227-9.
- 10 **Compston A**, Coles A. Multiple sclerosis. *Lancet* 2002;**359**:1221-31.
- 11 **Compston A**. The genetic epidemiology of multiple sclerosis. *Philos Trans R Soc Lond B Biol Sci* 1999;**354**:1623-34.
- 12 **Boomsma D**, Busjahn A, Peltonen L. Classical twin studies and beyond. *Nat Rev Genet* 2002;**3**:872-82.

- 13 **Salveti M**, Ristori G, Bomprezzi R, Pozzilli P, Leslie RD. Twins: mirrors of the immune system. *Immunol Today* 2000;**21**:342-7.
- 14 **MacGregor AJ**, Lanchbury J, Rigby AS, Kaprio J, Snieder H. Using twin studies to label disease as genetic or environmental is inappropriate. *BMJ* 2002;**324**:1100-1.
- 15 **Mumford CJ**, Wood NW, Kellar-Wood H, Thorpe JW, Miller DH, Compston DA. The British Isles survey of multiple sclerosis in twins. *Neurology* 1994;**44**:11-15.
- 16 **Sadovnick AD**, Armstrong H, Rice GP, Bulman D, Hashimoto I, Paty DW, Hashimoto SA, Warren S, Hader W, Murray TJ, et al. A population-based study of multiple sclerosis in twins: update. *Ann Neurol* 1993;**33**:281-5.
- 17 **French Group on Multiple Sclerosis**. Multiple sclerosis in 54 twinships: concordance rate is independent of zygosity. French Research Group on Multiple Sclerosis. *Ann Neurol*. 1992;**32**:724-7.
- 18 **Ristori et al**. manuscript in preparation.
- 19 **Abbadi N**, Philippe C, Chery M, Gilgenkrantz H, Tome F, Collin H, Theau D, Recan D, Broux O, Fardeau M, et al. Additional case of female monozygotic twins discordant for the clinical manifestations of Duchenne muscular dystrophy due to opposite X-chromosome inactivation. *Am J Med Genet* 1994;**52**:198-206.
- 20 **Costa T**, Greer W, Rysiecki G, Buncic JR, Ray PN. Monozygotic twins discordant for Aicardi syndrome. *J Med Genet* 1997;**34**:688-91.
- 21 **James WH**. Review of the contribution of twin studies in the search for non-genetic causes of multiple sclerosis. *Neuroepidemiology* 1996;**15**:132-41.
- 22 **The Transatlantic Multiple Sclerosis Genetics Cooperative**. A meta-analysis of genomic screens in multiple sclerosis. The Transatlantic Multiple Sclerosis Genetics Cooperative. *Mult Scler* 2001;**7**:3-11.
- 23 **Risch N**, Merikangas K. The future of genetic studies of complex human diseases. *Science* 1996;**273**:1516-17.
- 24 **Allison DB**, Schork NJ. Selected methodological issues in meiotic mapping of obesity genes in humans: issues of power and efficiency. *Behav Genet* 1997;**27**:401-21.
- 25 **Ewens WJ**, Spielman RS. The transmission/disequilibrium test: history, subdivision, and admixture. *Am J Hum Genet* 1995;**57**:455-64.
- 26 **Sawcer S**, Maranian M, Setakis E, Curwen V, Akesson E, Hensiek A, Coraddo F, Roxburgh R, Sawcer D, Gray J, Deans J, Goodfellow PN, Walker N, Clayton D, Compston A. A whole genome screen for linkage disequilibrium in multiple sclerosis confirms disease associations with regions previously linked to susceptibility. *Brain* 2002;**125**:1337-47.
- 27 **Compston A**, Sawcer S. Genetic analysis of multiple sclerosis. *Curr Neurol Neurosci Rep* 2002;**2**:259-66.
- 28 **Jacobsen M**, Schweer D, Ziegler A, Gaber R, Schock S, Schwinzer R, Wonigeit K, Lindert RB, Kantarci O, Schaefer-Klein J, Schipper HI, Oertel WH, Heidenreich F, Weinschenker BG, Sommer N, Hemmer B. A point mutation in PTPRC is associated with the development of multiple sclerosis. *Nat Genet* 2000;**26**:495-9.
- 29 **Jacobsen M**, Hoffmann S, Cepok S, Stei S, Ziegler A, Sommer N, Hemmer B. A novel mutation in PTPRC interferes with splicing and alters the structure of the human CD45 molecule. *Immunogenetics* 2002;**54**:158-63.
- 30 **Barcellos LF**, Caillier S, Dragone L, Elder M, Vittinghoff E, Bucher P, Lincoln RR, Pericak-Vance M, Haines JL, Weiss A, Hauser SL, Oksenberg JR. PTPRC (CD45) is not associated with the development of multiple sclerosis in US patients. *Nat Genet* 2001;**29**:23-4.
- 31 **Vorechovsky I**, Kralovicova J, Tchilian E, Masterman T, Zhang Z, Ferry B, Misbah S, Chapel H, Webster D, Hellgren D, Anvret M, Hillert J, Hammarstrom L, Beverley PC. Does 77C-G in PTPRC modify autoimmune disorders linked to the major histocompatibility locus? *Nat Genet* 2001;**29**:22-3.
- 32 **Strachan T**, Read AP. Genetic mapping of complex characters. In: *Human molecular genetics*. 2nd ed. New York: John Wiley, 1999.
- 33 **Hirschhorn JN**, Lohmueller K, Byrne E, Hirschhorn K. A comprehensive review of genetic association studies. *Genet Med* 2002;**4**:45-61.
- 34 **Polymeropoulos MH**, Lavedan C, Leroy E, Ide SE, Dehejia A, Dutra A, Pike B, Root H, Rubenstein J, Boyer R, Stenros ES, Chandrasekharappa S, Athanasiadou A, Papepetropoulos T, Johnson WG, Lazzarini AM, Duvoisin RC, Di Iorio G, Golbe LI, Nussbaum RL. Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. *Science* 1997;**276**:2045-7.
- 35 **Scott WK**, Stajich JM, Yamaoka LH, Speer MC, Vance JM, Roses AD, Pericak-Vance MA. Genetic complexity and Parkinson's disease. *Science* 1997;**277**:387-8.
- 36 **Tabor HK**, Risch NJ, Myers RM. Opinion. Candidate-gene approaches for studying complex genetic traits: practical considerations. *Nat Rev Genet* 2002;**3**:391-7.
- 37 **Lander ES**, Schork NJ. Genetic dissection of complex traits. *Science* 1994;**265**:2037-48.
- 38 **Lohmueller KE**, Pearce CL, Pike M, Lander ES, Hirschhorn JN. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat Genet* 2003;**33**:177-82.
- 39 **Nowoiny P**, Kwon JM, Goate AM. SNP analysis to dissect human traits. *Curr Opin Neurobiol* 2001;**11**:637-41.
- 40 **Collins FS**, Brooks LD, Chakravarti A. A DNA polymorphism discovery resource for research on human genetic variation. *Genome Res* 1998;**8**:1229-31.
- 41 **Weiss KM**, Clark AG. Linkage disequilibrium and the mapping of complex human traits. *Trends Genet* 2002;**18**:19-24.
- 42 **Collins A**, Lonjou C, Morton NE. Genetic epidemiology of single-nucleotide polymorphisms. *Proc Natl Acad Sci USA* 1999;**96**:15173-7.

- 43 **Waterston RH**, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P, Antonarakis SE, Aitwood J, Baertsch R, Bailey J, Barlow K, Beck S, Berry E, Birren B, Bloom T, Bork P, Botcherby M, Bray N, Brent MR, Brown DG, Brown SD, Bult C, Burton J, Butler J, Campbell RD, Carninci P, Cawley S, Chiaromonte F, Chinwalla AT, Church DM, Clamp M, et al. Initial sequencing and comparative analysis of the mouse genome. *Nature* 2002;**420**:520-62.
- 44 **Griffiths AJF**, Gelbart WM, Miller JH, Lewontin RC. Genetics begins with variation. In: *Modern genetic analysis*. New York: W H Freeman, 1999.
- 45 **Passarge E**. Dissecting Hirschsprung disease. *Nat Genet* 2002;**31**:11-12.
- 46 **Shirakawa T**, Enomoto T, Shimazu S, Hopkin JM. The inverse association between tuberculin responses and atopic disorder. *Science* 1997;**275**:77-9.
- 47 **Rook GA**, Stanford JL. Give us this day our daily germs. *Immunol Today* 1998;**19**:113-16.
- 48 **Rook GA**. Clean living increases more than just atopic disease. *Immunol Today* 2000;**21**:249-50.
- 49 **Fearon DT**, Locksley RM. The instructive role of innate immunity in the acquired immune response. *Science* 1996;**272**:50-3.
- 50 **Ristori G**, Buttinelli C, Pozzilli C, Fieschi C, Salvetti M. Microbe exposure, innate immunity and autoimmunity. *Immunol Today* 1999;**20**:54.
- 51 **Hemmer B**, Archelos JJ, Hartung HP. New concepts in the immunopathogenesis of multiple sclerosis. *Nat Rev Neurosci* 2002;**3**:291-301.
- 52 **Sawcer S**, Goodfellow PN, Compston A. The genetic analysis of multiple sclerosis. *Trends Genet* 1997;**13**:234-9.
- 53 **Chataway J**, Feakes R, Coraddu F, Gray J, Deans J, Fraser M, Robertson N, Broadley S, Jones H, Clayton D, Goodfellow P, Sawcer S, Compston A. The genetics of multiple sclerosis: principles, background and updated results of the United Kingdom systematic genome screen. *Brain* 1998;**121**:1869-87.
- 54 **Dyment DA**, Sadovnick AD, Ebers GC. Genetics of multiple sclerosis. *Hum Mol Genet* 1997;**6**:1693-8.
- 55 **Sawcer S**, Jones HB, Feakes R, Gray J, Smaldon N, Chataway J, Robertson N, Clayton D, Goodfellow PN, Compston A. A genome screen in multiple sclerosis reveals susceptibility loci on chromosome 6p21 and 17q22. *Nat Genet* 1996;**13**:464-8.
- 56 **Coraddu F**, Sawcer S, Feakes R, Chataway J, Broadley S, Jones HB, Clayton D, Gray J, Smith S, Taylor C, Goodfellow PN, Compston A. HLA typing in the United Kingdom multiple sclerosis genome screen. *Neurogenetics* 1998;**2**:24-33.
- 57 **Hensiek AE**, Sawcer SJ, Feakes R, Deans J, Mander A, Akesson E, Roxburgh R, Coraddu F, Smith S, Compston DA. HLA-DR 15 is associated with female sex and younger age at diagnosis in multiple sclerosis. *J Neurol Neurosurg Psychiatry* 2002;**72**:184-7.
- 58 **Tubridy N**, Coles AJ, Molyneux P, Compston DA, Barkhof F, Thompson AJ, McDonald WI, Miller DH. Secondary progressive multiple sclerosis: the relationship between short-term MRI activity and clinical features. *Brain* 1998;**121**:225-31.
- 59 **Nijeholt GJ**, van Walderveen MA, Castelijns JA, van Waesberghe JH, Polman C, Scheltens P, Rosier PF, Jongen PJ, Barkhof F. Brain and spinal cord abnormalities in multiple sclerosis. Correlation between MRI parameters, clinical subtypes and symptoms. *Brain* 1998;**121**:687-97.
- 60 **Enzinger C**, Ropele S, Strasser-Fuchs S, Kapeller P, Schmidt H, Poltrum B, Schmidt R, Hartung HP, Fazekas F. Lower levels of N-acetylaspartate in multiple sclerosis patients with the apolipoprotein E epsilon4 allele. *Arch Neurol* 2003;**60**:65-70.
- 61 **Oksenberg JR**, Barcellos LF. The complex genetic aetiology of multiple sclerosis. *J Neurovirol* 2000;**6**(suppl 2):S10-14.
- 62 **Straus SE**, Sneller M, Lenardo MJ, Puck JM, Strober W. An inherited disorder of lymphocyte apoptosis: the autoimmune lymphoproliferative syndrome. *Ann Intern Med*. 1999;**130**:591-601.
- 63 **Martin DA**, Zheng L, Siegel RM, Huang B, Fisher GH, Wang J, Jackson CE, Puck JM, Dale J, Straus SE, Peter ME, Krammer PH, Fesik S, Lenardo MJ. Defective CD95/APO-1/Fas signal complex formation in the human autoimmune lymphoproliferative syndrome, type Ia. *Proc Natl Acad Sci USA* 1999;**96**:4552-7.
- 64 **Drappa J**, Vaishnav AK, Sullivan KE, Chu JL, Elkon KB. Fas gene mutations in the Canale-Smith syndrome, an inherited lymphoproliferative disorder associated with autoimmunity. *N Engl J Med* 1996;**335**:1643-9.
- 65 **Wang J**, Zheng L, Lobito A, Chan FK, Dale J, Sneller M, Yao X, Puck JM, Straus SE, Lenardo MJ. Inherited human caspase 10 mutations underlie defective lymphocyte and dendritic cell apoptosis in autoimmune lymphoproliferative syndrome type II. *Cell* 1999;**98**:47-58.
- 66 **Lenardo M**, Chan KM, Hornung F, McFarland H, Siegel R, Wang J, Zheng L. Mature T lymphocyte apoptosis - immune regulation in a dynamic and unpredictable antigenic environment. *Annu Rev Immunol* 1999;**17**:221-53.
- 67 **Salvador JM**, Hollander MC, Nguyen AT, Kopp JB, Barisoni L, Moore JK, Ashwell JD, Fornace AJ Jr. Mice lacking the p53-effector gene Gadd45a develop a lupus-like syndrome. *Immunity* 2002;**16**:499-508.
- 68 **Balomenos D**, Martin-Caballero J, Garcia MI, Prieto I, Flores JM, Serrano M, Martinez-A C. The cell cycle inhibitor p21 controls T-cell proliferation and sex-linked lupus development. *Nat Med* 2000;**6**:171-6.
- 69 **Yan J**, Mamula MJ. Autoreactive T cells revealed in the normal repertoire: escape from negative selection and peripheral tolerance. *J Immunol* 2002;**168**:3188-94.
- 70 **Zhang J**, Markovic-Plese S, Lacet B, Raus J, Weiner HL, Hafler DA. Increased frequency of interleukin 2-responsive T cells specific for myelin basic protein and proteolipid protein in peripheral blood and cerebrospinal fluid of patients with multiple sclerosis. *J Exp Med* 1994;**179**:973-84.
- 71 **Nguyen LT**, Ramanathan M, Weinstock-Guttman B, Baier M, Brownschidle C, Jacobs LD. Sex differences in vitro pro-inflammatory cytokine production from peripheral blood of multiple sclerosis patients. *J Neurol Sci* 2003;**209**:93-9.
- 72 **Balomenos D**, Martinez AC. Cell-cycle regulation in immunity, tolerance and autoimmunity. *Immunol Today* 2000;**21**:551-5.
- 73 **Maas K**, Chan S, Parker J, Slater A, Moore J, Olsen N, Aune TM. Cutting edge: molecular portrait of human autoimmune disease. *J Immunol* 2002;**169**:5-9.
- 74 **Martin R**, Sturzebecher CS, McFarland HF. Immunotherapy of multiple sclerosis: where are we? Where should we go? *Nat Immunol* 2001;**2**:785-8.
- 75 **Schena M**, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 1995;**270**:467-70.
- 76 **Kononen J**, Bubendorf L, Kallioniemi A, Barlund M, Schraml P, Leighton S, Torhorst J, Mihatsch MJ, Sauter G, Kallioniemi OP. Tissue microarrays for high-throughput molecular profiling of tumor specimens. *Nat Med* 1998;**4**:844-7.
- 77 **MacBeath G**, Schreiber SL. Printing proteins as microarrays for high-throughput function determination. *Science* 2000;**289**:1760-3.
- 78 **Ziauddin J**, Sabatini DM. Microarrays of cells expressing defined cDNAs. *Nature* 2001;**411**:107-10.
- 79 **Mousses S**, Kallioniemi A, Kauraniemi P, Elkhoun A, Kallioniemi OP. Clinical and functional target validation using tissue and cell microarrays. *Curr Opin Chem Biol* 2002;**6**:97-101.
- 80 **Voorhoeve PM**, Agami R. Knockdown stands up. *Trends Biotechnol* 2003;**21**:2-4.
- 81 **Alizadeh AA**, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, Powell JJ, Yang L, Marti GE, Moore T, Hudson Jr J, Lu L, Lewis DB, Tibshirani R, Sherlock G, Chan WC, Greiner TC, Weisenburger DD, Armitage JO, Warnke R, Levy R, Wilson W, Greveer MR, Byrd JC, Botstein D, Brown PO, Staudt LM. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 2000;**403**:503-11.
- 82 **Hedenfalk I**, Duggan D, Chen Y, Radmacher M, Bittner M, Simon R, Meltzer P, Gusterson B, Esteller M, Kallioniemi OP, Wilfond B, Borg A, Trent J. Gene-expression profiles in hereditary breast cancer. *N Engl J Med* 2001;**344**:539-48.
- 83 **Davis RE**, Brown KD, Siebenlist U, Staudt LM. Constitutive nuclear factor kappaB activity is required for survival of activated B cell-like diffuse large B cell lymphoma cells. *J Exp Med* 2001;**194**:1861-74.
- 84 **van de Vijver MJ**, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, Schreiber GJ, Peterse JL, Roberts C, Marton MJ, Parrish M, Atsma D, Witteveen A, Glas A, Delahaye L, van der Velde T, Bartelink H, Rodenhuis S, Rutgers ET, Friend SH, Bernards R. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 2002;**347**:1999-2009.
- 85 **Knight J**. Cancer comes under scrutiny in fresh genomic initiative. *Nature* 2001;**410**:855.
- 86 **Dalton R**. Phoenix rises to genomic challenge. *Nature* 2002;**417**:107.
- 87 **Stephenson J**. Human genome studies expected to revolutionize cancer classification. *JAMA* 1999;**282**:927-8.
- 88 **Vastag B**. Gene chips inch toward the clinic. *JAMA* 2003;**289**:155-9.
- 89 **Whitney AR**, Diehn M, Popper SJ, Alizadeh AA, Boldrick JC, Relman DA, Brown PO. Individuality and variation in gene expression patterns in human blood. *Proc Natl Acad Sci USA* 2003;**100**:1896-901.
- 90 **Ramanathan M**, Weinstock-Guttman B, Nguyen LT, Badgett D, Miller C, Patrick K, Brownschidle C, Jacobs L. In vivo gene expression revealed by cDNA arrays: the pattern in relapsing-remitting multiple sclerosis patients compared with normal subjects. *J Neuroimmunol* 2001;**116**:213-19.
- 91 **Lucchinetti C**, Bruck W, Parisi J, Scheithauer B, Rodriguez M, Lassmann H. Heterogeneity of multiple sclerosis lesions: implications for the pathogenesis of demyelination. *Ann Neurol* 2000;**47**:707-17.
- 92 **Lassmann H**, Bruck W, Lucchinetti C. Heterogeneity of multiple sclerosis pathogenesis: implications for diagnosis and therapy. *Trends Mol Med* 2001;**7**:115-21.
- 93 **Bielekova B**, Kadom N, McFarland H, Martin R. Immunoreactivity to different autoantigens in MS patients influences their disease phenotype. International Society of Neuroimmunology - Sixth International Congress. *J Neuroimmunol* 2001;**118**:1:A242.
- 94 **Bielekova B**, Kadom N, Frank J, McFarland H, Martin R. Immunoreactivity to myelin epitopes differ between MS patients subgrouped based on MRI and clinical characteristics of the disease. In: *Proceedings of 2002 AAN Annual Meeting*, S33.001, Denver, Colorado, USA.
- 95 **Miller LD**, Long PM, Wong L, Mukherjee S, McShane LM, Liu ET. Optimal gene expression analysis by microarrays. *Cancer Cell* 2002;**2**:353-61.
- 96 **Lai E**, Bowman C, Bansal A, Hughes A, Mosteller M, Roses AD. Medical applications of haplotype-based SNP maps: learning to walk before we run. *Nat Genet* 2002;**32**:353.