New approaches to investigating heterogeneity in complex traits

R Bomprezzi, P E Kovanen, R Martin

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.
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, multigenerational kindreds 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. 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. However, the interpretation of the results derived from these studies demands caution to avoid false positive and spurious results. Yet association studies remain a powerful approach. 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). 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. 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). But what is the relevance of genetic heterogeneity in MS, identifying it as a key point to be addressed.7

Consistent with this view are the findings by Jacobsen et al.6 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 in linkage and TDT underscores the relevance of genetic heterogeneity in MS, identifying it as a key point to be addressed.27

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.

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.

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, multigenerational kindreds 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. 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. However, the interpretation of the results derived from these studies demands caution to avoid false positive and spurious results. Yet association studies remain a powerful approach. 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). 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. 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). But what is the relevance of genetic heterogeneity in MS, identifying it as a key point to be addressed.7

Consistent with this view are the findings by Jacobsen et al.6 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 in linkage and TDT underscores the relevance of genetic heterogeneity in MS, identifying it as a key point to be addressed.27

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.

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.
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.

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. Or, as a more recent theory proposes, increases in both allergy and autoimmunity are attributed to a dysfunction of innate immunity. 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. 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. The abovementioned investigations have shown that MS is indeed a polygenic trait. To address the issue of the interaction between genes, Coraddu et al. used data from a genome wide scan, in this instance stratifying the patients according to HLA status, and repeating the linkage analysis. This step resulted in significant changes in linkage scores. Some chromosomal regions that had appeared earlier were not evident any more, whereas

---

**Figure 2**
(A) Linkage analysis, aimed at identifying genetic markers that are co-inherited 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.
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. Many other groups have also looked for a correlation between disease phenotype and surrogate markers with partial success. 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). 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). 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. 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.

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 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. 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 and it is through the combination of new techniques and alternative approaches that optimal results can be achieved (fig 4).
been shown that gene expression profiling via microarray experiments is capable of reclassifying apparently homogeneous groups of clinical entities. 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. Further- more, 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.

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. In line with previous studies by Ramanathan et al and others, 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. It has been shown that measuring multiple immunological

**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.
parameters, stratifying patients according to MRI characteristics, and incorporating clinical parameters allows the recognition of subgroups among MS patients. This approach could be further improved by the combination 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 pathogenic 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. Like assessing the degree of polymorphism in a given population, which involves genotyping individual subjects, 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, Neuromuscular Branch, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, Maryland 20892, USA

REFERENCES


47 Rook GA, Stanley JL. Give us this day our daily germs. Immunol Today 1999;20:113-16.


New approaches to investigating heterogeneity in complex traits

R Bomprezzi, P E Kovanen and R Martin

*J Med Genet* 2003 40: 553-559
doi: 10.1136/jmg.40.8.553

Updated information and services can be found at:
http://jmg.bmj.com/content/40/8/553

*These include:*

**References**
This article cites 89 articles, 28 of which you can access for free at:
http://jmg.bmj.com/content/40/8/553#BBL

**Email alerting service**
Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

**Topic Collections**
Articles on similar topics can be found in the following collections

- Immunology (including allergy) (604)
- Multiple sclerosis (20)
- Clinical genetics (256)

**Notes**

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/