

Mapping the new frontier: complex genetic disorders

Richard Mayeux

J Clin Invest. 2005;115(6):1404-1407. <https://doi.org/10.1172/JCI25421>.

Review Series Introduction

The remarkable achievements in human genetics over the years have been due to technological advances in gene mapping and in statistical methods that relate genetic variants to disease. Nearly every Mendelian genetic disorder has now been mapped to a specific gene or set of genes, but these discoveries have been limited to high-risk, variant alleles that segregate in rare families. With a working draft of the human genome now in hand, the availability of high-throughput genotyping, a plethora of genetic markers, and the development of new analytical methods, scientists are now turning their attention to common complex disorders such as diabetes, obesity, hypertension, and Alzheimer disease. In this issue, the *JCI* provides readers with a series dedicated to complex genetic disorders, offering a view of genetic medicine in the 21st century.

Find the latest version:

<https://jci.me/25421/pdf>





Mapping the new frontier: complex genetic disorders

Richard Mayeux

Gertrude H. Sergievsky Center and Taub Institute for Research on Alzheimer's Disease and the Aging Brain, Columbia University, New York, New York, USA.

The remarkable achievements in human genetics over the years have been due to technological advances in gene mapping and in statistical methods that relate genetic variants to disease. Nearly every Mendelian genetic disorder has now been mapped to a specific gene or set of genes, but these discoveries have been limited to high-risk, variant alleles that segregate in rare families. With a working draft of the human genome now in hand, the availability of high-throughput genotyping, a plethora of genetic markers, and the development of new analytical methods, scientists are now turning their attention to common complex disorders such as diabetes, obesity, hypertension, and Alzheimer disease. In this issue, the *JCI* provides readers with a series dedicated to complex genetic disorders, offering a view of genetic medicine in the 21st century.

Unraveling complex genetic disorders

The identification of genes underlying Mendelian disorders, named after Gregor Mendel and defined by the occurrence of a disorder in fixed proportions among the offspring of specific matings, has been greatly enhanced over the last few decades by remarkable achievements in gene mapping and the development of rigorous statistical methods. Most of the progress in human genetics during this time has come from the studies of families with rare segregating high-risk alleles. With at least 30,000 genes in the human genome and the identification and characterization of these genes underway, the challenge now is to dissect common complex genetic disorders such as obesity, diabetes, schizophrenia, and cancer. As a group, the majority of these disorders have a tendency to aggregate in families but rarely in the classical Mendelian fashion. While researchers have made some progress in the genetics of complex disorders over the last decade, gaps clearly remain. It is likely that, with characterization of the genetic influences underlying these complex disorders, there will be even greater opportunities for improving the lives of affected individuals.

The ability to genetically map complex disorders has been facilitated by technological improvement in identifying and genotyping polymorphic DNA markers (Table 1). The current trend is to use single-nucleotide polymorphisms (SNPs), the most frequently seen type of genetic polymorphisms, with an estimated 3 million SNPs present in the human genome. Though somewhat less informative than other types of DNA markers, SNPs are technically easier and less expensive to genotype because they have only 2 alleles and require less DNA. For example, a set of 2 arrays can genotype more than 100,000 SNPs with a single primer.

Most researchers believe that complex disorders are oligogenic, the cumulative result of variants in several genes, or polygenic, resulting from a large number of genetic variants, each contributing small effects. Still others have proposed that these disorders result from an interaction between one or more genetic variants and environmental or nongenetic disease risk factors. The motivation for unraveling these complex genetic disorders is clear.

Not only will this shed new light on disease pathogenesis, but it may also provide potential targets for effective treatment, screening, and prevention and increase the understanding of why some patients do not respond to currently available treatments while others do. The difficulty facing researchers who work on these complex genetic disorders is in designing appropriate studies to merge the richness of modern genome science with the vast potential of population-based, epidemiological research.

In this issue, a series of reviews describes the current state of the art in methods for gene mapping of complex disorders, including statistical methods for association studies and linkage disequilibrium mapping. We also include reviews that offer examples of the application of these methods in 3 complex genetic disorders: diabetes (see Permutt et al., pages 1431–1439; ref. 1), schizophrenia (see Kirov et al., pages 1440–1448; ref. 2), and neurodegeneration (see Bertram and Tanzi, pages 1449–1457; ref. 3). Clearly, this is an exciting and rapidly evolving area of science in which the elucidation of the human genome can now be applied to common complex genetic problems. However, it is also worth briefly reviewing the traditional application of methods to assess genetic contributions to human disorders.

Estimating the heritability of a disorder

A higher concordance of disease among monozygotic compared with dizygotic twins or a higher risk among relatives (e.g., siblings) of patients with disease than among relatives of controls or those in the general population are usually the observations that lead researchers to believe that a disease is familial or at least possibly under genetic influence. However, there are statistical methods available to determine the degree to which a disease or trait is heritable. These estimates reflect the proportion of genetic variance over the total phenotypic variance from members within the family (4–6). The residual variance is the proportion reflecting environmental or nongenetic risk factors. Studies of heritability provide an estimate of the degree to which the variability in the phenotype is related to genetic variation, but it is difficult to separate shared genetic from shared environmental influences. Siblings, especially twins, share their childhood environment in addition to some portion of their genetic background. Genetic epidemiologists view heritability estimates as approximations of the genetic variance in disease risk because heritability depends on all contributing genetic and environmental or nongenetic compo-

Nonstandard abbreviations used: SNP, single-nucleotide polymorphism.

Conflict of interest: The author has declared that no conflict of interest exists.

Citation for this article: *J. Clin. Invest.* 115:1404–1407 (2005).
doi:10.1172/JCI25421.



Table 1
DNA markers for gene mapping

DNA markers	Type of genetic marker	Advantages	Disadvantages	Use
Restriction fragment length polymorphisms	Restriction enzymes at specific sites	Two alleles; site is present or absent	Limited information; expensive; time consuming	Linkage; analysis
Variable number of tandem repeats (VNTR)	Minisatellites	Many alleles; highly informative	Required Southern blotting and radioactive probes; not evenly spread throughout genome	Linkage; analysis
VNTR	Microsatellites	Many alleles; highly informative; widely distributed	Requires polymerase chain reaction (PCR); expensive	Linkage; analysis
Single-nucleotide polymorphisms	SNPs	Two alleles; widely distributed; typing can be automated	Less informative than microsatellites	Linkage and association analyses

nents. As described by Kirov et al. in this series, a high heritability score does not always mean that gene mapping will be easy (2). A change in any one factor can influence the overall estimate. Heritability estimates do not effectively separate shared genetic from shared environmental influences and cannot effectively apportion the degree of gene-environment interaction. This is most certainly true in studies of diabetes (see Permutt et al.; ref. 1).

Segregation analysis is a statistical tool that can model the inheritance pattern. It is useful in the analysis of non-Mendelian or complex genetic disorders that may be polygenic or the result of gene-environment interaction (4). Segregation analysis estimates the appropriate mix of genetic and environmental factors using information from a series of families identified by the researcher. Certain assumptions regarding gene mechanism, the frequency of the variant form of the gene, and its suspected penetrance are provided by the researcher who must also specify the model of inheritance: sporadic, polygenic, dominant, or recessive. A maximum likelihood analysis, the probability of obtaining the observed results given the distribution of data in the population, provides results that reflect the mix of parameters that best fit the observed data compared with a general or mixed model. Segregation analysis estimates genetic contributions by aggregating a set of genes, but it is not specific to a single gene, and the types of families recruited can affect results. For example, very large families will contribute more toward the specified model than smaller ones. Nonetheless, this approach informs the investigator regarding the degree to which the disease is genetic and can also provide some of the parameters of inheritance. For simpler diseases, these genetic parameters can be used in subsequent genetic analyses, such as linkage analysis, to provide greater power in identifying the variant gene or genes.

Identifying disease genes

For the investigation of many inherited Mendelian diseases, researchers have used linkage analysis in families with several affected family members to identify putative involved genes. Linkage analysis attempts to identify a region (locus) of the chromosome or regions (loci) in the genome associated with the disease or trait by identifying which alleles in the loci are segregating with the disease in families. Geneticists use genetic markers that are evenly distributed throughout the genome to reduce the number of chromosomal regions to a handful that may harbor a disease gene. Simply put, this method exploits the biological reality that in meiosis I, genes located

close to each other on the same chromosome are inherited together more often than expected by chance. The genes that are far apart will not inherit together because recombination will break up segments of the chromosome. Thus, if a set of marker alleles are segregating with the disease, those markers are assumed to be located near the disease gene (7, 8). Using linkage analysis, scientists determine the likelihood that the loci (genetic marker and disease gene) are linked by calculating the logarithm of the odds or lod score, which is a ratio of 2 likelihoods: the odds that the loci are linked and the odds that the loci are not linked or are independent. To take into account multiple testing and the likelihood of linkage prior to considering the genetic evidence, a lod score of 3 or more is used as an indication of statistically significant linkage with a 5% chance of error, though more stringent criteria have been recommended for genome-wide scans (9). Two-point (ratio of the likelihoods that 2 loci are linked) and multipoint linkage (ratio of likelihoods at each location across the genome) analyses are standard analyses used in gene mapping. Once a location or set of locations suggestive of linkage are identified, researchers turn to finer mapping methods using either a more dense set of additional microsatellites or SNPs in a smaller region underlying the high lod score.

While linkage analysis remains a mainstay of gene mapping, it does have shortcomings. Both genotyping and phenotyping errors have devastating effects on the validity of the lod score. Locus heterogeneity (more than one causal gene) and clinical heterogeneity (multiple forms of the same disease with different etiology) can also pose serious problems. A pattern of inheritance or model must be assumed, and the researcher must estimate the frequency and penetrance of the disease gene. Therefore, the analysis is parametric. In late onset diseases, additional complications can arise when individuals with putative variant allele develop the disease later in life or in a much milder form (incomplete penetrance). Therefore, linkage analysis is best suited for Mendelian disorders, not common complex genetic disorders, unless the correlation between the genotype and the phenotype is known to be very robust (10). Occasionally, a rare, high risk allele is found in patients with a rare, familial form of a common disorder, such as Alzheimer disease. Though the findings often have implications in the disease pathogenesis, the role of the rare mutation is limited for common sporadic forms of disease in the general population. In this series, an example of this is described by Bertram and Tanzi in their discussion of Alzheimer disease, in which mutations in the amyloid pre-

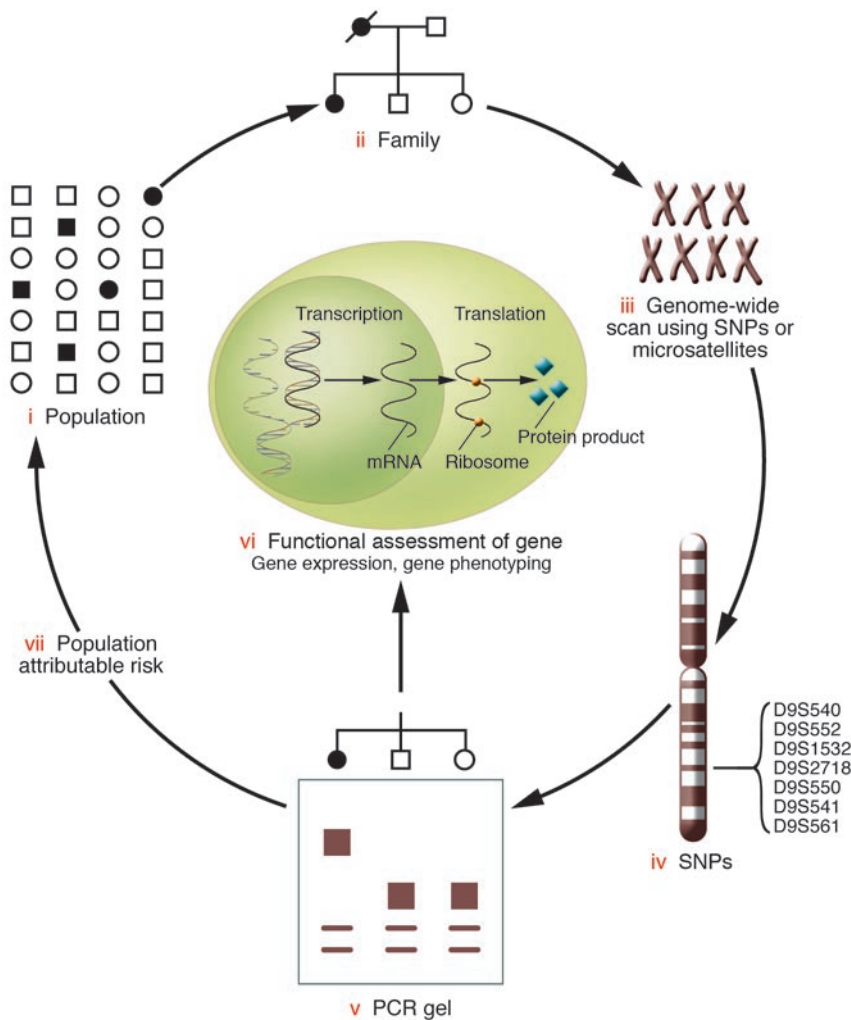


Figure 1

Progression of gene mapping in genetic epidemiological studies. (i) Population from which the complex genetic disorder arose. (ii) One of several families included in the genome-wide scan. However, more recently, genome-wide association studies of unrelated patients and controls have been advocated. (iii) Genome-wide scan using microsatellite DNA markers or SNPs. (iv) Fine mapping using a dense collection of SNPs in a region that segregates with disease. (v) Variant allele detection using sequencing. (vi) Functional assessment of the protein product. (vii) Determination of population attributable risk.

cursor protein and presenilin I and II lead to an overproduction of amyloid β protein, which is deposited in the brains of all patients with Alzheimer disease, regardless of the etiology (3).

While linkage analysis is arguably the most powerful method for identifying rare, high-risk alleles in Mendelian disease, many consider genetic association analysis to be the best method for identifying genetic variants related to common complex diseases (11, 12). In contrast to linkage analysis, which involves scanning the entire genome or a very large segment, association analyses are best suited to interrogating smaller regions or segments of the genome. Association analyses are generally model free, or nonparametric, so the researcher does not have to assume a mode of inheritance is unknown. Unlike linkage analysis, where markers are identified, association studies determine whether or not a specific allele within a marker is associated with disease. Association studies can be conducted in a group of randomly selected patients and controls as well as in small families or affected sibling pairs. Thus, this approach is sometimes added to ongoing epidemiological or clinical trials and can be adapted for use with relatively small-sized families. Association analyses of candidate genes underlying quantitative traits such as body mass index as related to obesity or blood pressure in relation to hypertension are also feasible, as will be clear from the discussion by Majumder and Ghosh in this series (see pages 1419–1424; ref. 13).

There is at least one important similarity between linkage and association analyses. Linkage analysis involves association within families, while genetic association analysis examines whether affected individuals share the common allele more often than do controls. Patients who share the variant allele may also share a common ancestor from whom the allele originated. In reality, researchers often do both linkage and association analyses. Linkage analysis is used for the genome-wide screen to identify candidate loci. The region is subsequently narrowed using linkage disequilibrium mapping, which is reviewed by Morton in this series (see pages 1425–1430; ref. 14). Genome-wide association studies are now feasible and can provide an additional means for identifying genes related to complex disorders. This approach combines the best features of linkage with the strength of association approaches (12). Figure 1 illustrates the progression from the study of a population to the identification of a variant allele and subsequent functional analysis. Genetic epidemiologists often go back to the population in order to determine the population attributable risk, which is defined as the proportion of disease in the population that can be ascribed to the variant allele or risk factor of concern. It is based both on the relative risk (see Gordon and Finch, pages 1408–1418; ref. 15) and the prevalence of the variant in the population.



Association studies also have limitations. Because linkage disequilibrium, cosegregation of a series of genetic markers or alleles, is sustained over only a short chromosomal segment, a large number of loci need to be tested to cover a region (or the genome if a genome-wide association is conducted). This increases the possibility of false-positive findings. Therefore, one cannot rely on the conventional threshold P value of 0.05. With each test, the possibility of a false-positive result increases, requiring the need either for replication in an independent study or computer simulation (11). For complex genetic disease studies, researchers can use computer simulation of 1,000 replicates of the family collection based on observed allele frequencies and recombination fractions to determine the threshold for statistical significance in order to reduce the possibility of false-positive results. For case-control studies, patients with disease and the comparison group of controls can differ in genetic background, introducing variables unrelated to the disease and causing a type of spurious association or confounding termed *population stratification*. Finally, the number of subjects required for these studies can be large, particularly if the heritability or relative risk of the disorder or trait is low. In this series, Gordon and Finch review both the benefits and limitations of using association analysis in family-based and population-based studies to identify genes related to complex disorders (15).

Conclusions

Researchers, clinicians, patients, and their families are likely to reap the benefits of continued application of progress in human genetics to various disciplines in medicine. The relatively new field of genetic epidemiology, a hybrid of genetics and epidemiology, is already capitalizing on this progress by enabling researchers to focus both on genetic variations in different populations (unrelated patients and controls or small families and sibling pairs) and their exposure to environmental or nongenetic risk factors in order to explain how their joint effects lead to disease. With this continued payoff has come the need for a better understanding of the intricacies of genetic exploration and genome science. Designing the appropriate studies, using the correct analytic approach, and appreciating the strengths and weaknesses of genetic methods as applied to common complex disorders is essential. It is our hope that this series, *Complex genetic disorders*, in the *JCI* will facilitate that process for our readers.

Address correspondence to: Richard Mayeux, the Gertrude H. Sergievsky Center, 630 West 168th Street, Columbia University, New York, New York 10032, USA. Phone: (212) 305-2391; Fax: (212) 305-2518; E-mail: rpm2@columbia.edu.

1. Permutt, M.A., Wasson, J., and Cox, N. 2005. Genetic epidemiology of diabetes. *J. Clin. Invest.* **115**:1431–1439. doi:10.1172/JCI24758.
2. Kirov, G., O'Donovan, M.C., and Owen, M.J. 2005. Finding schizophrenia genes. *J. Clin. Invest.* **115**:1440–1448. doi:10.1172/JCI24759.
3. Bertram, L., and Tanzi, R.E. 2005. The genetic epidemiology of neurodegenerative disease. *J. Clin. Invest.* **115**:1449–1457. doi:10.1172/JCI24761.
4. Khoury, M.J., Beaty, T.H., and Cohen, B.H. 1993. *Fundamentals of genetic epidemiology*. Oxford University Press. New York, New York, USA. 383 pp.
5. Grilo, C.M., and Pogue-Geile, M.F. 1991. The nature of environmental influences on weight and obesity: a behavior genetic analysis. *Psychol. Bull.* **110**:520–537.
6. Falconer, D.S. 1989. *Introduction to quantitative genetics*. 3rd edition. John Wiley & Sons. New York, New York, USA. 438 pp.
7. Ott, J. 1999. *Analysis of human genetic linkage*. 3rd edition. Johns Hopkins University Press. Baltimore, Maryland, USA. 382 pp.
8. Terwilliger, J.D., and Ott, J. 1994. *Handbook of human genetic linkage*. Johns Hopkins University Press. Baltimore, Maryland, USA. 320 pp.
9. Lander, E.S., and Schork, N.J. 1994. Genetic dissection of complex traits. *Science*. **265**:2037–2048.
10. Terwilliger, J.D., and Goring, H.H. 2000. Gene mapping in the 20th and 21st centuries: statistical methods, data analysis, and experimental design. *Hum. Biol.* **72**:63–132.
11. Neale, B.M., and Sham, P.C. 2004. The future of association studies: gene-based analysis and replication. *Am. J. Hum. Genet.* **75**:353–362.
12. Hirschhorn, J.N., and Daly, M.J. 2005. Genome-wide association studies for common diseases and complex traits. *Nat. Rev. Genet.* **6**:95–108.
13. Majumder, P.P., and Ghosh, S. 2005. Mapping quantitative trait loci in humans: achievements and limitations. *J. Clin. Invest.* **115**:1419–1424. doi:10.1172/JCI24757.
14. Morton, N.E. 2005. Linkage disequilibrium maps and association mapping. *J. Clin. Invest.* **115**:1425–1430. doi:10.1172/JCI25032.
15. Gordon, D., and Finch, S.J. 2005. Factors affecting statistical power in the detection of genetic association. *J. Clin. Invest.* **115**:1408–1418. doi:10.1172/JCI24756.