The search for allelic variants that cause monogenic disorders or predispose to common, complex polygenic phenotypes

Stylianos E. Antonarakis, MD, DSc

The search for the mutant genes for monogenic disorders has been a spectacular success. This was accomplished because of the mapping and sequencing of the human genome, the determination of the sequence variability, the collection of well-characterized families with mendelian disorders, the development of statistical methods for linkage analysis, and laboratory methods for mutation search. The challenge of the genetic medicine is now to decipher the nucleotide sequence variants that predispose to common complex, polygenic phenotypes. The methodology for this challenge is in development and constant evolution. It is anticipated that, in the next 10 to 20 years, susceptibility alleles for these common disorders will be identified.

The principal aim of the field of genetic medicine is to discover the links between nucleotide sequence variation in the human genome and the various human phenotypes. The methodologies of linkage analysis and mutation detection, along with progress in the mapping and sequencing of the human genome and that of model organisms, resulted in a plethora of exciting discoveries concerning mutant alleles of genes and their related phenotypes. In this review, I will briefly summarize some general principles regarding the search for genes (more specifically, mutant alleles of these genes) that either cause the various human genetic disorders or confer predisposition to common, complex phenotypes.

Monogenic disorders

There are a large number of phenotypes (each of which is rare in the population) due to abnormal mutant alleles of single genes. These disorders are usually called monogenic since there is one gene of paramount importance related to the development of the phenotype; consequently, these phenotypes show a mendelian mode of inheritance. For a particular disorder, the mapping of the responsible gene could easily be determined by studying the transmission of polymorphic markers within a family. Positional candidate gene–cloning strategies could then be employed to identify the responsible gene by virtue of mutations (nucleotide sequence variants) present in a patient's DNA and not in controls. The genetic methodology identified a route to understanding the molecular basis of disease, which otherwise seemed intractable. On May 25, 2001, the knowledge-based database OMIM (Online Mendelian Inheritance in Man)[1,2] contained 1168 mutant genes linked to human monogenic disorders. Several notable examples of neurological disorders are shown in *Table I*, which lists disease genes, their corresponding phenotypes, and the years of the linkage mapping and their positional cloning. The first mutant gene–disease link discovered by positional cloning strategies was that of chronic granulomatous disease in 1986.

Keywords: *monogenic disorder; complex polygenic phenotype; pathogenic allele; linkage analysis; linkage disequilibrium*

Author affiliations: Division of Medical Genetics, University of Geneva Medical School and University Hospitals of Geneva, Switzerland

Address for correspondence: Division of Medical Genetics, University of Geneva Medical School and University Hospitals of Geneva, 1 rue Michel-Servet, 1211 Geneva, Switzerland
(e-mail: Stylianos.Antonarakis@medecine.unige.ch)

Gene	Disease	Mapping	Cloning
HD	Huntington chorea	1983	1993
NF1	Neurofibromatosis, type I	1987	1991
AT1	Ataxia-telangiectasia	1988	1995
SCN4A	Hyperkalemic periodic paralysis	1990	1991
VHL	Von Hippel-Lindau syndrome	1988	1993
DRPLA	Dentatorubral-pallidoluysian atrophy	1993	1993
SMN1	Spinal muscular atrophy	1990	1995
SOD1	Amyotrophic lateral sclerosis, familial	1991	1993
PS1	Alzheimer's disease, early-onset	1992	1995
PS2	Alzheimer's disease, early-onset	1995	1995
FA	Friedreich ataxia	1988	1996
PARK2	Parkinson's disease, juvenile	1997	1998

Table I. Partial list of selected mutant genes that cause monogenic neurological disorders.

The work of numerous investigators in both academia and the biotechnology industry over the last 20 years or so has provided the infrastructure necessary to perform studies of linkage mapping and gene identification of genetic disorders. A large number of polymorphic markers due to variable units of short sequence repeats (SSRs) have been identified throughout the entire genome and used to create linkage maps of all human chromosomes. These highly polymorphic markers in turn provided the tools to localize the unknown disease-related genes to intervals of the genome. The most common of these SSRs are dinucleotide repeats such as $(GT)_n$ or $(AT)_n$. Dense linkage maps that determined the position, order, and distance of adjacent SSRs have been produced by genotyping the DNAs of appropriate large families including those collected and distributed by CEPH (Centre d'Étude de Polymorphism Humain).[3] A "draft" sequence of the entire human genome has been obtained and is publicly available.[4,5] Approximately 40% of this sequence is already finished (ie, of high quality, ordered, and practically gapless), including that of the chromosomes 22 and 21.[6,7] The availability of the nucleotide sequence has two important consequences for identifying disease loci. First, the recognition of a total of approximately 35 000 genes will now greatly facilitate and accelerate gene–disease matchmaking. Second, the discovery of more than 2 million single nucleotide polymorphisms (SNPs)[5,8] will likely result in the recognition of the functional nucleotide sequence variability that is associated with common complex phenotypes.

The story of positional identification of disease-related alleles

Let me now describe a typical project to identify the mutant gene associated with a monogenic disorder *(Figure 1)*, for example, the autosomal dominant Huntington disease gene. Dominant means that a mutation in only *one* allele of an autosomal gene is needed in order to manifest the disorder. In contrast, in a recessive disease, mutations in *both* alleles of an autosomal gene are needed for the phenotypic expression of the disease. Mutations in genes on the X or Y sex chromosomes are associated with X- or Y-linked phenotypes. However, no matter the mode of inheritance, the general strategy to identify the causative gene mutation(s) is similar.

Collection of families

The initial phase of the project is to identify families with the precise phenotypic characteristics of the disease, and establish that the number of individuals available for study provides the appropriate power in linkage analysis to identify the disease gene location. The collection of samples from affected and unaffected members of the families is then justified, after approval of the study by the local human experimentation ethics committees and informed consent. For Huntington disease, members of a large family from Maracaibo, Venezuela, were collected,[9] but the biomedical literature is full of other interesting family collections from different parts of the world and different geoethnic communities. The best population groups for rare autosomal recessive disorders are those in which consanguineous marriages are common, or those originated from a few founders. I wish to emphasize the importance of accurate diagnosis of affected and nonaffected individuals in order to perform errorless linkage analysis. Aspects of the phenotype that have to be taken into account include the age of onset and clinical variation; the participation of a clinical expert is therefore of paramount importance in the development and success of the project.

Linkage analysis

The next phase is the performance of linkage analysis to localize the yet unknown "disease gene" to a small genomic region. This linkage analysis is based on the

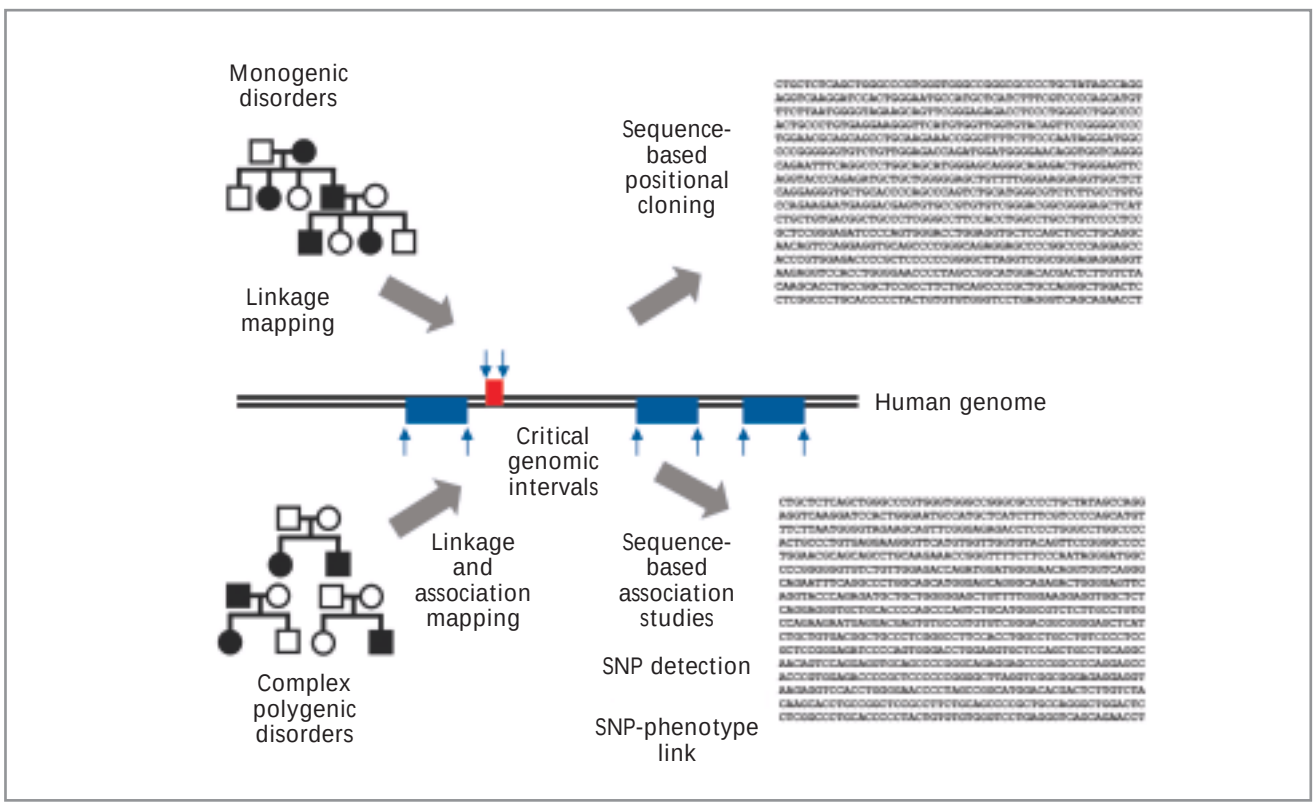

Figure 1. Schematic representation of a genetic (map-based) approach to identify mutant alleles involved in monogenic and complex phenotypes. The human genome is shown as a double straight line in the middle. The top panel shows a simplified strategy for monogenic disorders. The responsible gene is mapped by linkage analysis to a small genomic interval (shown in red); subsequently, positional identification of the pathogenic mutation is based on the nucleotide sequence information. The bottom panel shows a simplified strategy for complex, polygenic disorders and traits. The predisposing alleles are mapped by linkage and association studies to several large regions of the genome (shown in blue). It is expected that sequence-based association studies will uncover some of the predisposing alleles. SNP, single nucleotide polymorphism.

identification of DNA polymorphic markers that *cosegregate* with the disease phenotype. The DNA markers, which constitute part of the normal nucleotide variability of the genome, usually fall into two categories as mentioned: the SSRs and the SNPs. For the linkage analysis studies, the most useful markers are SSRs since they are highly polymorphic. There are more than two different (usually six) alleles per SSR marker in the population, and they are therefore informative in the majority of the families. Most of the successful linkage mapping studies have used approximately 300 such markers equally distributed throughout the genome with an average interval of 10 cM, or 10% recombination between adjacent markers. Note that this distance is measured in genetic terms, ie, in recombination units in human meiosis; 1 cM on average corresponds to approximately 1000 kb or 10^6 nucleotides of DNA. After the use of sufficient markers, the success of a linkage mapping project in a monogenic phenotype depends on:

- *The size of the families and the DNAs available for study.* It is imperative to perform a linkage simulation analysis of the available sample to determine if there is sufficient statistical "power" to detect linkage.
- *The accuracy of the diagnosis.* Problems arise when affected individuals are categorized as normal or vice versa (due to inability to detect the manifestations of the phenotype, or reduced "penetrance," ie, the absence of phenotype in spite of the presence of the mutant gene or late onset of the phenotypic characteristics).
- *The extent of the genetic heterogeneity of the phenotype.* It is much easier to map the disease locus if the phenotype is always due to mutations in the same gene. In contrast, it is much more difficult to map loci for disor-

ders/phenotypes that result from mutant alleles of more than one gene. An example of genetic homogeneity is Huntington disease, in which all affected pedigrees are due to mutations in the same gene on chromosomal region 4p.[10] In contrast, tuberous sclerosis shows genetic heterogeneity. There are two genes, *TSC1* and *TSC2* on chromosomes 9 and 16, respectively; mutations in each result in the same phenotype of tuberous sclerosis.[11]

The next step after the localization of a disease-related locus to a particular genomic interval is to narrow down this region to an area of approximately 1 to 2 megabases (Mb) (1000 000-2000 000 nucleotides). Data provided by the preliminary annotation of the human genome sequence indicate that, on average, there are 9 genes per megabase, with a range between different chromosomes from 5 to 23.[4,5]

The narrowing down of the "critical region" is therefore important and can generally be achieved by two methods. The first is to identify critical recombination events between certain DNA markers and the disease phenotype in the families examined. This is achieved by the addition of affected families and by studying a large number of markers in the critical region. It is advisable to rely mostly on recombinants in the DNAs of affected individuals.

The second approach takes advantage of linkage disequilibrium (LD), that is the historical recombinants between the disease mutation and the polymorphic variants surrounding the mutation. The extent of LD or allelic association usually defines the area of the disease locus. "Old" mutations show a short region of LD; more recent, "young," mutations are obviously associated with a large region of LD because there were only few meioses and generations to restrict the area of LD. LD is useful in autosomal recessive disorders with consanguinity, or founder effect autosomal dominant and X-linked disorders with ancient mutations. In contrast, LD is not contributory in dominant or X-linked disorders with many different and recent (only a few generations) mutant alleles.

Positional identification of the pathogenic allele

The next phase requires a search for mutant alleles of genes that map within the critical interval. The methodology of this search for the elusive gene has changed most dramatically in the last 12 to 15 years. The advances of the human genome project provide a publicly available genomic infrastructure that becomes more detailed every year. In the mid-1980s, it was necessary to complete the physical map of the critical region, ie, to develop an overlapping set of cloned human DNAs that covered the entire critical region. Then, it was necessary to identify portions of all genes in the critical interval, clone the entire cDNAs, and determine the intron–exon junctions and their genomic structure. All of the above steps have now been largely accomplished by the international collaboration and competition that is collectively called the human genome project. This extraordinary project provided a dense linkage map,[3] a complete physical map of the genome,[12,13] a large number of partial gene sequences,[14] and, this year, the almost entire human genome sequence.[4,5] At the time of writing (May 2001), there exists in the public database a sequence of the human genome that consists of about 40% finished high-quality sequence and ~50% draft sequence of lower quality with numerous gaps and unordered DNA fragments. Less than 10% of the human genomic sequence is still unknown. There are two chromosomes, namely 22 and 21, for which the sequence is complete with only minimal gaps.[6,7] There are now catalogues of well-characterized and predicted genes in the entire genomic landscape. These catalogues are freely available and constantly updated, for example, see the websites of the National Center for Biotechnology Information and the Ensembl project.[15,16] It is likely therefore that a substantial number of genes have already been identified in the critical genomic interval that harbors the elusive gene for the monogenic disorder of interest. The list of genes is by no means complete and the characterization of each gene is far from adequate; the genomic efforts in the next couple of years will be directed towards the structural and functional characterization of all human genes. At present, however, the unfinished gene list per genomic interval provides a wealth of candidates to search for mutations in the patients' DNA or RNA. The methodologies for mutation characterization have also changed considerably in the last few years. Direct sequencing after polymerase chain reaction (PCR) amplification of exons, intron–exon junctions, 5'-untranslated region (UTR), and 3'-UTR usually detects the majority of mutations that may cause a disease phenotype. Certainly, there are exceptions, and sequencing after PCR amplification may miss muta-

tions. A large, heterozygous deletion, for example, that eliminates one or more exons, may go undetected without a quantitative PCR, or by hybridization after Southern blotting. In some cases, it is necessary to separate the two different alleles in vitro and search for mutations in isolated alleles.

What is usually the result of a mutation search in an "average" positional cloning project? A considerable number of sequence variants are identified in the genes sequenced. Since the frequency of polymorphic variants is about 1 in 1300 nucleotides between two randomly chosen chromosomes, approximately 40 nucleotide differences are expected to be found in the genes of the 1-Mb critical region between normal and affected individuals. This is based on the fact that about 2% of the genome is a coding region of genes, and with the inclusion of intron–exon junctions and 5'- and 3'-UTR, approximately 4% of the 1-Mb region is usually sequenced in patients. The next important question is: which one of these nucleotide differences is associated with (or is responsible for) the disease phenotype? Several criteria could be employed in order to focus on some of the differences.

- The presence of the variant nucleotide in normal individuals from the same ethnic group as the affected individuals is normally sufficient to consider these changes as nonpathogenic variation. Usually, DNAs from about 100 individuals are examined for rare and common variants.
- The predicted consequence of the mutation is important. Nucleotide differences resulting in nonsense codons or translational frameshifts, or severe splicing defects, are more likely to be pathogenic.
- De novo mutations, not present in the parents, particularly in sporadic cases of dominant or X-linked disorders are more likely to be pathogenic.
- The most difficult to interpret nucleotide changes are those that result in amino-acid substitutions (missense mutations). Substitutions in evolutionarily well-conserved amino acids among homologous proteins in different species are excellent candidates for pathogenic mutations.
- Mutations that are predicted to alter the function of the protein or have been experimentally demonstrated to do so are excellent candidates.
- Certain mutations need to be tested in model organisms in order to study their effect.
- Other mutations require long-term epidemiological studies to prove their involvement with a disease phenotype.

The study of the molecular basis of the disease phenotype in unrelated pedigrees and the demonstration of mutations in the same gene often confirm the involvement of this gene in the disease.

The description of studies to elucidate the function of the disease-related protein and the pathogenetic mechanism of the disease is beyond the scope of this article. It is, however, important to emphasize that the evolutionary conservation of genes makes model organisms (yeast, worm, fruitfly, zebrafish, or mouse) indispensable tools for the functional analysis of human genes.

The methodology described above for gene cloning responsible for monogenic disorders has been repeatedly successful.[2] A considerable number of disease-related genes and alleles have been identified in the last 15 years. The OMIM contains 1168 genes with mutant alleles associated with disease phenotypes. Most of these have been identified using positional cloning efforts without any previous knowledge of the biochemistry or pathophysiology of the disease phenotype.

Functional gene variants for predisposition to common, complex, phenotypes

One of the greatest challenges of this decade for biomedicine is to identify the mutant/polymorphic alleles that cause or predispose to common human disorders with a strong genetic component. It is not far from the truth if we state that the entire effort for the mapping, sequencing, and determination of the normal variability of our genome has been done in order to be able to find the mutant alleles of the common, complex phenotypes. These phenotypes include disorders such as schizophrenia and bipolar disease, diabetes, asthma, atherosclerosis, multiple sclerosis, obesity, hypertension, Alzheimer's disease, aging, and susceptibility to infectious diseases. The tasks appear enormous, but the expected benefits for medicine could be so profound that are certainly worth the effort and expenses from both academia and industry.

The discovery of predisposing mutant alleles for common disorders is nevertheless very difficult. Although we do not understand all the reasons for this difficulty, we could certainly mention the following points.

First, the inheritance of the common complex phenotypes is not clearly mendelian. It is true that there is an aggregation of affected individuals in certain families,

but the mode of inheritance is not compatible with the usual recognizable patterns. On the other hand, there is strong evidence for a genetic predisposition for these complex disorders[17]: (i) the frequency of these disorders in relatives of affected individuals is much greater than in the general population; and (ii) the frequency of concordance in monozygotic twins is always greater than in dizygotic twins. For some disorders, such as schizophrenia, adoption studies are also in favor of genetic predisposition. In all of these studies, the underlying hypothesis in favor of genetic predisposition is that people who share a greater proportion of alleles have a higher probability of manifesting the disease in their lifetimes. Monozygotic twins share 100% of their alleles; siblings and dizygotic twins share 50% of their alleles; first cousins share an eighth of their alleles. Therefore, for a complex disease with genetic predisposition, the probability of developing the disease is greater in a monozygotic twin of an affected individual, less in a sibling of an affected individual, and even less in a first cousin. The fact that the inheritance pattern of these disorders is not mendelian renders the use of parametric linkage analysis difficult or impossible, since these studies require a fixed mode of inheritance.

Second, the phenotype may be uncertain. This is clearly evident in the psychiatric disorders in which the diagnoses are based only on clinical criteria. The danger of misdiagnosis, or misclassification, is therefore considerable. In addition, the age of onset of a phenotype is variable even within the same family; this makes it difficult to categorize unaffected individuals as truly lifetime unaffected. A further problem is that there is likely genetic heterogeneity in affected individuals within the same family. This is because most of these disorders are common and it is therefore possible to have affected individuals due to the contribution of mutant alleles of different genes. For example, it is not unusual to find individuals with breast cancer not related to the *BRCA1* gene belonging to families with well-documented *BRCA1*-related breast cancer. After all, breast cancer is common since it affects approximately 10% of females in their lifetime.

Third, it is possible that each of the predisposing mutant alleles has a minor effect on the phenotype and that several mutant alleles from different genes in concert result in a pathological phenotype. In addition, most of the predisposing mutant alleles may be common polymorphic variants in the population. Unlike the successes of the monogenic disorders, we know of very few mutant alleles that predispose to common, complex polygenic disorders. Such examples include the *APOE4* allele, which predisposes to Alzheimer's disease,[18] and factor V Leiden, which predisposes to deep venous thrombosis.[19]

The most important challenge to the genetic medicine in the next decade is certainly to uncover the mutant alleles that predispose to the complex common disorders. The advances of the sequences of the human and other genomes, and the discovery of the polymorphic variability among human genomes, provide the necessary infrastructure to tackle these problems. In addition, the development of rapid, high-throughput, inexpensive, and reliable methods for mutation detection will also contribute to these discoveries. Finally, the availability of samples from a large number of parents, their family members, and controls is also another necessary component in this endeavor.

The methods that exist today for the mapping of predisposing alleles (PDAs) could be summarized as follows *(Figure 1)*.[20,21]

Linkage analysis methods

The parametric methods with fixed mode of inheritance could still be used in the large rare families segregating a complex phenotype. Furthermore, linkage projects usually involve small families with complex disorders, in which case all possible modes of inheritance should be tested. The nonparametric methods, also known as model-free methods, are certainly more suitable for complex phenotypes. These methods score the "amount" of shared alleles among affected individuals. The most widely used method is that of sibling pairs. In this, potentially interesting alleles are those that are shared in siblings in frequencies statistically different from the expected 50%. A large number of affected siblings are necessary (most studies used 100-200 such pairs) and their power to reveal linkage increases when the DNAs of their parents are available. There are several variations of this method, since all affected relatives could be used and nonaffected individuals could also provide valuable information. Most of the studies with sibpairs use SSR markers because ideally all four parental alleles could be recognized. For these studies, a genome-wide scan usually requires approximately 300 polymorphic markers placed in the average intervals of average size 10 cM.

Transmission disequilibrium test

This test, which is in between linkage and association, estimates the difference between the alleles transmitted and nontransmitted to patients from their parents. The null hypothesis for a noncontributing locus is that there is no difference between transmitted and nontransmitted alleles. In this method, single affected individuals and their parents could be used (usually referred as trios). Most studies have used about 100 such trios. The advantage of this method is that it utilizes a powerful internal control of the nontransmitted alleles from the same population as the affected alleles.

Association studies

The most simple study used to determine the implication of a mutant allele to a phenotype is that of association of the polymorphic allele to the phenotype. The polymorphic allele (usually an SNP) could be PDA, or be within a very short genomic distance from the PDA. The question then is: how close could an SNP be to the PDA in order to still detect an association, or LD? The answer to this question is complex and depends on the number of the different PDAs to the phenotype, the age of the mutant alleles, the frequency of the mutant alleles in the population, the heterogeneity of the phenotype, and the history and size of the population. For example, the average blocks of LD in European populations is approximately 60 kb around common polymorphisms, whereas in the older African human populations these blocks are much smaller.[22] It follows that for an association study in the old populations, one needs many more polymorphic markers than in younger populations. The isolated populations may also provide an advantage since the blocks of LD may be even larger; however, the disadvantage of these populations may be that their PDAs are in different genes than those of the outbred populations.

Most of the markers used in LD/association studies are SNPs. Theoretically, the most useful SNPs are those that change an amino acid, or occur in regions of gene expression regulation. More than 2 million SNPs have now been identified as part of the genome-sequencing effort and a small fraction (0.2%) result in missense codons.[4,7] It has been estimated that the average human gene shows two to four common variants in the population. It is perhaps more advisable at this stage to concentrate on the SNPs within genes and their regulatory regions as markers for LD/association studies. In addition, many investigators recommend the use of two adjacent SNPs in concert as haplotypes (pattern of polymorphic alleles in one chromosome), in order to increase the detective power of each site. The samples of patients and controls used in association studies have to be as identical as possible in terms of genome variability; ideally, the two samples need to be drawn from the same ethnic group. The sample size is also a matter of debate, but larger sample sizes provide more statistical power. It may well be necessary to collect samples of several thousands of individuals per category. The sample size largely depends on the contribution of each mutant allele to the phenotype, the number of genes involved, the age of the mutant alleles, and the frequency of the mutant PDAs in the unaffected population.

Joint linkage and association

In this approach, several genomic regions containing PDAs are first identified by linkage analyses and then LD/association studies are performed in the 10- to 20-Mb critical regions. It is a common finding that the critical intervals from linkage studies of complex disorders are 10 times larger than those of monogenic disorders, even if considerable numbers of samples are used. In addition, linkage and association studies of complex phenotypes reveal several (more than one) areas of suggestive linkage, some of which could be replicated in subsequent studies, but others could not be verified. (For a recent discussion of the status of such studies in schizophrenia and bipolar disease, see references 23 and 24, respectively.)

Determination of the predisposing mutant alleles

How can we identify the predisposing mutant allele after its mapping in the human genome? The answer to this question may well be more difficult than the localization of the PDA. Let us suppose that the PDA maps to a 100-kb region of the genome. Within this interval, there may well exist 100 common polymorphic alleles in the population and a substantial number of these would be present in the affected individuals. The determination of those alleles with a significant contribution to the phenotype may require genotyping of additional affected and unaffected individuals from different geoethnic

groups, the functional analysis of the variants, and large epidemiologic studies. It is also possible that different alleles in the same gene predispose to the phenotype, similar to the situation in which different mutant alleles within one gene cause the same monogenic phenotype. Model organisms could also be used to map and clone PDAs for common phenotypes. Due to space limitations, the experimental strategies using animal models are not discussed here.

Concluding remarks

The identification of mutant genes responsible for monogenic disorders has been a triumph of medical genetics in the last 15 years. These discoveries depended on the successes of the mapping and sequencing of the human genome, and identification of the normal variability. These achievements created an environment of enthusiasm for further developments, high expectations, and underestimation of the difficulties that lie ahead in the complex, common phenotypes.

There is a cautious optimism now, in both academia and industry, for further advances in the identification of these functional sequence variants that predispose to the common human diseases. These will certainly continue to revolutionize medicine and will place genetic medicine at the center of the diagnosis and treatment of human disorders.

I thank Dr Robert Lyle for critical reading of this manuscript; I also thank all the members of our laboratory, past and present, for ideas, debates, curiosity, experiments, enthusiasm, and hard work. The research in our laboratory has been supported over the last 20 years by numerous funding agencies, mainly the NIH, the Swiss National Science Foundation, and the European Union.

La investigación de variantes de alelos que producen trastornos monogénicos o predisponen a fenotipos poligénicos, comunes y complejos

La investigación de genes mutantes para los trastornos monogénicos ha significado un suceso espectacular. Esto se ha podido realizar gracias al mapeo e identificación de la secuencia del genoma humano, la determinación de la variabilidad de la secuencia, la serie de familias bien caracterizadas con trastornos mendelianos, el desarrollo de métodos estadísticos para el análisis de enlaces y los métodos de laboratorio para la investigación de mutaciones. El desafío actual de la medicina genética es descifrar las variantes de las secuencias de nucleótidos que predisponen a fenotipos poligénicos, comunes y complejos. La metodología para este desafío está en desarrollo y evolución constantes. Se puede anticipar que en los próximos 10 a 20 años serán identificados los alelos susceptibles para estos trastornos frecuentes.

La recherche de variants alléliques responsables de maladies monogéniques ou prédisposant aux phénotypes polygéniques courants complexes

La recherche sur les gènes mutants des maladies monogéniques a été un succès spectaculaire. Celle-ci a été rendue possible grâce à l'établissement de la carte et du séquençage du génome humain, à la détermination de la variabilité des séquences, à la collecte de familles bien caractéristiques porteuses de maladies mendéliennes et au développement de méthodes statistiques pour l'analyse de liaison ainsi que de méthodes de laboratoire pour la recherche de mutations. Le défi de la médecine génétique est maintenant de déchiffrer les variants de séquences nucléotides qui prédisposent aux phénotypes polygéniques courants complexes. La méthodologie pour y parvenir est en cours de développement et en constante évolution. Il est à prévoir que les gènes de susceptibilité pour ces maladies courantes seront identifiés dans les 10 à 20 ans à venir.

REFERENCES

1. Antonarakis SE, McKusick VA. OMIM passes the 1000-disease-gene mark. *Nat Genet.* 2000;25:11
2. National Center for Biotechnology Information. Online Mendelian Inheritance in Man. http://www.ncbi.nlm.nih.gov/Omim. Accessed June 18, 2001.
3. NIH/CEPH Collaborative Mapping Group. A comprehensive genetic linkage map of the human genome. *Science.* 1992;258:67-86.
4. Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. *Nature.* 2001;409:860-921.
5. Venter JC, Adams MD, Myers EW, et al. The sequence of the human genome. *Science.* 2001;291:1304-1351.
6. Dunham I, Shimizu N, Roe BA, et al. The DNA sequence of human chromosome 22. *Nature.* 1999;402:489-495.
7. Hattori M, Fujiyama A, Taylor TD, et al. The DNA sequence of human chromosome 21. The chromosome 21 mapping and sequencing consortium. *Nature.* 2000;405:311-319.
8. Sachidanandam R, Weissman D, Schmidt SC, et al. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature.* 2001;409:928-933.
9. Gusella JF, Wexler NS, Conneally PM, et al. A polymorphic DNA marker genetically linked to Huntington's disease. *Nature.* 1983;306:234-238.
10. The Huntington's Disease Collaborative Research Group. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell.* 1993;72:971-983.
11. Cheadle JP, Reeve MP, Sampson JR, et al. Molecular genetic advances in tuberous sclerosis. *Hum Genet.* 2000;107:97-114.
12. Cohen D, Chumakov I, Weissenbach J. A first-generation physical map of the human genome. *Nature.* 1993;366:698-701.
13. McPherson JD, Marra M, Hillier L, et al. A physical map of the human genome. *Nature.* 2001;409:934-941.
14. Deloukas P, Schuler GD, Gyapay G, et al. A physical map of 30 000 human genes. *Science.* 1998;282:744-746.
15. National Center for Biotechnology Information. http://www.ncbi.nlm.nih.gov. Accessed June 18, 2001.
16. The Sanger Center and European Bioinformatics Institute. http://www.ensembl.org. Accessed June 18, 2001.
17. Kendler KS, Diehl SR. The genetics of schizophrenia: a current, genetic–epidemiologic perspective. *Schizophr Bull.* 1993;19:261-285.
18. Roses AD. A model for susceptibility polymorphisms for complex diseases: apolipoprotein E and Alzheimer disease. *Neurogenetics.* 1997;1:3-11.
19. Dahlback B. Resistance to activated protein C caused by the factor VR506Q mutation is a common risk factor for venous thrombosis. *Thromb Haemost*. 1997;78:483-488.
20. Lander ES, Schork NJ. Genetic dissection of complex traits. *Science*. 1994;265:2037-2048.
21. Risch NJ. Searching for genetic determinants in the new millennium. *Nature*. 2000;405:847-856.
22. Reich DE, Cargill M, Bolk S, et al. Linkage disequilibrium in the human genome. *Nature*. 2001;411:199-204.
23. Baron M. Genetics of schizophrenia and the new millennium: progress and pitfalls. *Am J Hum Genet.* 2001;68:299-312.
24. Berrettini WH. Susceptibility loci for bipolar disorder: overlap with inherited vulnerability to schizophrenia. *Biol Psychiatry.* 2000;47:245-251.