2002 ASHG PRESIDENTIAL ADDRESS
The Complexity of Complex Diseases*

P. Michael Conneally

Department of Medical and Molecular Genetics, Indiana University, Indianapolis

P. Michael Conneally

Fellow members of our society, I am deeply appreciative for the honor of president you have bestowed upon me. It has been an exciting year, and my responsibilities as president have been greatly helped by the addition of our new executive vice president, Joann Baughman, and also by Executive Director Elaine Strass, Jane Salomon, Marsha Ryan, and the other able members of our administrative office staff.

My first exposure to human genetics was a course in January 1958 at the University of Wisconsin, which was cotaught by a former president of our society, Jim Crow, and his former student and the first Allen Award winner, Newton Morton. The structure of DNA had been described just five years earlier by Watson and Crick, and just two years earlier, the diploid number of human chromosomes was shown to be 46, to the surprise of most scientists. A year later, it was found that Down syndrome was caused by an extra chromosome. The ensuing 44 years have established human genetics at the forefront of the biological sciences.

The Human Genome Initiative has not only essentially completed the human sequence, but it has also fostered the development of rapid automated genotyping, allowing, among other things, a larger number of polymorphic markers to be typed and, thus, a greater coverage of the genome. One result has been that the vast majority of relatively common Mendelian traits have been mapped, and most of the responsible genes have been cloned. Thus, in the case of Mendelian disorders, the weakest link, the paucity of markers, has been overcome, leading to the elucidation of the genetic etiology of these disorders.

Mendelian disorders, however, are uncommon. Based mainly on twin studies, it was clear that many of the common disorders in humans also have a major genetic component. Buoyed by their success in mapping Mendelian disorders, scientists began to turn their attention to these common disorders with great enthusiasm and optimism. There was an urgency to proceed. Physicians in general clinical practice rarely see a patient with a Mendelian disorder. A large number of their clientele, however, have a serious illness in which genetics plays a major role. HMOs, with an emphasis on prevention, realize the importance of knowing their patients' risk factors for common disorders. It soon became clear that this endeavor, mapping and cloning genes for complex disorders, was far and away more difficult than was first imagined. At this juncture, I would like to make a distinction between two types of multifactorial traits. The first involves mostly common congenital malformations, such as cleft lip and/or palate, pyloric stenosis, and club foot. These are dichotomous (all-or-none) attributes, which often have a major gene involved, but are not inherited in simple Mendelian fashion. These will not be discussed further. The second group is known as complex disorders for good reason, since both their genetic and environmental components would seem to be more complicated.

While multigenerational families were ideal for linkage of Mendelian traits, their use in early linkage studies of bipolar disorder and schizophrenia produced a great deal of publicity for results later found to be incorrect. There was general agreement that new approaches were necessary, but there was no consensus on any one method.

National Institutes of Health (NIH), in particular National Institute of Mental Health, National Institute of Alcohol Abuse and Alcoholism, and National Heart,

Received November 14, 2002; accepted for publication November 14, 2002; electronically published January 23, 2003.

Address for correspondence and reprints: Dr. P. Michael Conneally, Department of Medical and Molecular Genetics, Indiana University, Indianapolis, IN 46202-5251. E-mail: pconneal@iupui.edu

* Previously presented at the annual meeting of The American Society of Human Genetics, in Baltimore, on October 16, 2002.

0002-9297/2003/7202-0003$15.00

Lung, and Blood Institute, realized that the time was ripe to fund initiatives to find causative genes for those common disorders, such as the psychiatric diseases, addiction, hypertension, and adult diabetes. It also became clear that a large number of subjects, whether as families, affected sib pairs, or case-controls, were needed for adequate power to detect linkage to a gene with only a partial affect on the phenotype. Thus, any one researcher would have difficulty in recruiting sufficient subjects, and this naturally led to a large number of collaborative efforts by scientists. The overall budgets for these NIH-funded grants to study complex diseases were quite high, with a large proportion being spent on recruitment and diagnosis. In recent years, this has led a number of institutes to demand that researchers share their resources, DNA and/or immortalized cell lines, and clinical data with other investigators outside the collaboration. This has been a controversial decision, and NIH is still refining the rules for this sharing venture. On the other hand, researchers realize that NIH has invested a considerable amount of money to procure these resources, and it wants to maximize the research potential of the resource rather than invest further money in sample procurement. It may be difficult to share resources previously collected without obtaining institutional review board approval to recontact individuals for their permission to broaden the scope of sharing. Clearly, the best solution to enforced sharing of resources is to enter into a collaborative agreement with the new researchers, and, hopefully, this will become the trend in the future. I believe that collaborative sharing is likely to be commonplace in the near future.

As in most genetic studies, the procedure for a successful study for positional cloning of complex disease genes has a number of components. The first is ascertainment, clinical evaluation, and sample collection from appropriate families or cases and controls. Family ascertainment is becoming more difficult, with new restrictions being placed on obtaining clinical information on an individual from a relative. As you are aware, this can be a major impediment to research, and ASHG is attempting to allay fears of administrators, IRBs, and others that the rights of family members will not be taken into account. It is important that all of us be aware of the sensitive nature of many familial disorders, and respect the privacy of family members. On the other hand, it is important to emphasize that family members are usually very eager to participate and are likely to be concerned if they perceive too many impediments limiting scientific progress.

In general, three types of population samples are in vogue, extended families, affected sib pairs (with or without their parents), and case-control or transmission/disequilibrium studies. Each has its proponents among the genetics community and, due to the many possible genetic mechanisms involved in complex disorders, each may well be optimum for a specific genetic mechanism. Thus, they should complement each other rather than being viewed as competitors.

Probably the single most important aspect of studies of complex disease is the phenotype, which may well be multidimensional. It is essential that a well-thought-out standardized protocol, which has been validated, be used by all clinicians involved. While it may not be possible to foresee important variables that may later be very useful for defining endophenotypes, attempting to revisit individuals for further information at a later time will be both time consuming and expensive.

The next step in gene discovery is genotyping. While restriction fragment length polymorphisms (RFLPs) were useful for positional cloning of Mendelian disorders, their limited number and expensive typing were not very useful for the study of complex disorders. The discovery of dinucleotide repeats by Weber and, later, the identification of other short tandem repeats (microsatellites) allowed a 10-cM genome screen to be performed easily and inexpensively. Thus, sufficient markers were available for gene mappers to realize that the genetic etiology of complex traits could be dissected. Unfortunately, the findings from numerous studies using these new markers were inconclusive. Single-nucleotide polymorphisms (SNPs) may well be the savior for complex disorders. Their almost-limitless abundance in the human genome, and the ever-increasing speed of high-throughput genotyping, bodes well for the mapping community.

The final phase of a linkage study of a complex trait is the statistical analysis and interpretation of the results. Early on, it became apparent that the analytic methods used to localize Mendelian traits were of very limited usefulness for complex disorders. Fortunately, many highly capable genetic epidemiologists developed a wide variety of statistical genetic tools to deal with large numbers of genotypes, multiple phenotypes, and complex disease inheritance. These methods encompass both dichotomous and quantitative traits. In mapping Mendelian traits, nonreplication usually simply meant locus heterogeneity. In complex-trait analysis, nonreplication of a linkage finding may be due to different sets of genes acting in different samples but also may mean that the initial positive results were artifactual, possibly due to the large number of markers being tested. While many studies utilize different methods as a quasireplication, hoping to give increased weight to a specific chromosomal location, this is not a substitute for actual replication.

For the second part of my presentation, I chose an example of a complex disorder that is one of a number that our group has been involved with for a number of years, namely, addictive disorders, and I will focus on alcohol dependence. Though per capita alcohol consumption is declining, it is estimated that 20 million Ameri-

cans have a serious drinking problem leading to early death. Alcohol contributes to up to 50% of fatal automobile accidents.

The concordance rate for alcoholism in MZ twins is twice that in DZ twins. Probably the most compelling evidence that alcohol dependence is to a large extent genetic came from a study of Swedish children adopted at birth. In essence, offspring of an alcoholic parent were at greater risk of alcohol dependence than children of nonalcoholics raised in an adoptive home with an alcoholic adopting parent. Adopted men whose biological fathers were early-onset alcoholics had a ninefold increased risk of alcoholism as compared with those with nonalcoholic fathers. Other twin studies have demonstrated similar findings, although the risk is typically closer to a fourfold increase.

Early studies of alcohol susceptibility focused on the study of candidate genes. The gene for the mitochondrial aldehyde dehydrogenase (ALDH2), which oxidizes acetaldehyde, has a variant, ALDH2*2. The protein encoded by this variant has very low or absent activity. Homozygotes have no enzyme activity, while heterozygotes have much lower activity than the normal homozygote. Individuals with the mutant allele experience a flushing reaction upon ingestion of only a small amount of alcohol (Thomasson et al. 1991). This adverse reaction acts as a protective factor against alcoholism, and the mutant allele has a relatively high prevalence in Asians. There are also differences in allele frequencies at the two alcohol dehydrogenase (ADH) loci in different racial groups, and they also play a role in alcoholism. Extensive association studies of the dopamine D2 receptor (DRD2) have had mixed results. A few studies have found a significant association between the TaqI-A1 polymorphism in the DRD2 gene with alcoholism (Noble et al. 1994), but a majority of researchers have been unable to confirm this finding (Edenberg et al. 1998).

The largest genetic linkage study on alcoholism is the Collaborative Study of the Genetics of Alcoholism (COGA), which was initiated in 1990. The study involves nine centers throughout the U.S. All are responsible for recruiting families and subjects. Two, Indiana University School of Medicine and Washington University, St. Louis, are responsible for molecular genotyping and data management and analysis, and Southwestern Foundation for Biomedical Research is also involved in data analysis. Other sites are State University of New York, Downstate Medical Center (New York), University of Connecticut Health Science Center, University of Iowa School of Medicine, University of California School of Medicine (San Diego), and Howard University (Washington, D.C.). Initially, 105 families were evaluated and genotyped for more than 300 microsatellite markers as part of a genome screen. Individuals were defined as affected if they fulfilled COGA criteria for alcoholism, based on both DSM-III-R and Feighner criteria. A total of 382 sib pairs met these criteria. Potential evidence of linkage was found on chromosomes 7 (LOD score 3.5), 1 (LOD score 2.9), and 2 (LOD score 1.8). Interestingly, additional analysis, using individuals without a diagnosis of alcoholism but with multiple alcoholic relatives, suggested a locus with protective effects on chromosome 4, near the ADH loci (Foroud et al. 2000). Since the COGA sample includes only a few Asians, this finding suggested that the ADH protective factor is not limited to the Asian population. Subsequently, using the phenotype, "maximum number of drinks ever consumed," linkage to chromosome 4 in the ADH region was detected (Saccone et al. 2000). This region is now being extensively studied.

A replication data set of 157 pedigrees was ascertained and evaluated using the same diagnostic criteria and marker set as the initial sample. This sample supported linkage to chromosomes 1 and 7 but not to chromosome 2. A new region on chromosome 3 (LOD = 3.4) was identified. It is typical of replication studies to support some, but not all, of the previous findings. This would suggest either that the two samples may be from different populations or, probably more likely, the disparate results are due to a type 1 error. A more-detailed discussion of problems in replication is given by Suarez et al. (1994). This two-stage approach, an initial and then a replication sample, is ideal for mapping a complex trait, and it also gives independent assessment of marker order. As one might expect, there are numerous candidate genes in these broadly linked chromosomal regions, and the regions must be further refined before embarking on a screen of candidate genes.

It is likely that gene interaction plays a major role in the etiology of complex disorders. To misquote a famous phrase, "no gene is an island." Using the COGA data set to look for interaction (epistasis), one was discovered between loci on chromosomes 1 and 15. An analysis, holding the score on 1 constant, gave a significant score for chromosome 15.

Another phenomenon, being investigated at the Indiana Alcohol Research Center, is acute tolerance to alcohol. Two measures were assessed: sensitivity (which is the initial response to alcohol) and acute adaptation (which is the recovery of a dependent measure toward baseline while the exposure of the brain to alcohol is held constant). Alcohol is infused into family-history-positive (FHP) and family-history-negative (FHN) individuals to obtain a breath alcohol level (BrAc) of 60 mg % over a 20-min period. A scale to measure subjective perceptions is administered to subjects before and after this 20-min period. This scale measures the subjects' perception of their level of intoxication in several dimensions. The BrAc level is then maintained at 60 mg % for 175 min, and the perceptions scale is administered a third time.

The result was that FHP subjects reported greater initial response to alcohol following the initial alcohol infusion than FHN individuals. During the 175 minutes of constant alcohol exposure, the FHP subjects developed more acute tolerance to alcohol, compared with the FHN subjects. The majority of the measures were highly significant (Morzorati et al. 2002).

What can the brain tell us about alcohol dependence? Human brain oscillations, as measured by an electroencephalogram (EEG), are stable and are highly heritable. The average heritability in four frequency bands, delta (1.5–3.5 Hz), theta (4.0–7.5 Hz), alpha (8.0–12.5 Hz), and beta (13–25 Hz) is 76%, 89%, 89%, and 86%, respectively. Porjesz and colleagues (2002) recently studied 1,553 individuals (age range 7–70 years) from 250 COGA families. A whole genome screen, using 351 markers, was performed. The strongest evidence for linkage was observed on chromosome 4p for beta-based traits. Beta was divided into three components and gave LOD scores 3.39, 5.01, and 2.17 for a region encompassing the GABRB1 locus. This region is situated within a cluster of $GABA_A$ receptor genes located on chromosome 4p, the same region implicated, independently, in the susceptibility to alcoholism. Also, significant linkage disequilibrium was found in the COGA sample, across the GABRA2 gene, using DSM-IV criteria to define the phenotype for alcoholism. A significant LOD score (4.12) was also found with a theta wave on chromosome 7 in the region of the acetylcholine muscarinic receptor, which is known to control theta rhythm.

To summarize this brief discussion on the genetics of alcoholism, it was clear from the beginning that multiple centers would need to be involved, both for the necessary broad expertise they provided and also to provide sufficient family data for adequate power. When COGA was first conceived, it was realized that the project would be long term. The protocol was well thought out, with strict adherence to ascertainment, assessment, and data management. Sharing of results and problems was seen to be very important, and members were enthusiastically collaborative.

COGA studies have shown that multiple approaches and multiple phenotypes can also be useful both for replication purposes, and, more importantly, to find subtypes (endophenotypes). The latter is important if there is genetic heterogeneity, for example, a number of major loci each involved in a specific phenotype. In the case of alcoholism, it appears that electrophysiological variables may be determined by some of the same loci involved in alcohol dependence.

A number of other organisms have been used to identify genes for alcoholism. A strain of rats developed at Indiana University by T.-K. Li and colleagues have been divergently selected for their preference (P strain) or lack of preference (NP strain) for alcohol. A QTL for alcohol preference has been mapped to a region on rat chromosome 4 that includes the neuropeptide Y (NPY) locus (Bice et al. 1998). Mutations in NPY cause antisocial behavior in *Caenorhabditis elegans*.

What can we expect in the future? While progress in locating genes for complex disorders has been very slow and frustrating, I believe we are now at a turning point. A major development is the utilization of SNPs. They are abundant in the genome, allowing relatively very fine mapping of genes. They are present in coding and noncoding, as well as regulatory regions, of genes. Highly efficient automated techniques are now available for high-throughput genotyping. I believe SNPs and CHIPS are our salvation. Multiple loci can be assembled into haplotypes, which makes the region highly polymorphic and, in many cases, population specific. A HapMap project is under way to construct haplotypes across the genome for the major races, which will undoubtedly benefit the mapping of regions involved in complex diseases. Regions of interest can be substantially narrowed to limit the number of candidate genes to be sequenced. Case-control studies that have been frowned on by many geneticists due to questionable results, often due to stratification, can now be used in a much more meaningful manner in conjunction with SNP haplotypes. Linkage disequilibrium studies are much more likely to locate susceptibility genes. Witness the success of Graeme Bell's group in type 2 diabetes mellitus by finding an association with the gene encoding calpain-10, using 21 SNPs located in the telomeric region of chromosome 2q (Horikawa et al. 2000).

For most complex disorders, the family data that are available may not always be sufficient to provide enough power, but, on the other hand, most families are being updated, in many cases, with more precise phenotyping and/or new approaches to look at subphenotypes now being termed "endophenotypes." Thus, it is important to have good rapport with families from the outset. It will be also important in the future that more collaborative efforts take place with less secrecy, especially in the commercial field, but this is probably wishful thinking on my part. We must remember that collaborative research is much more efficient than individual scientists "going it alone."

New statistical methodologies will be forthcoming to deal with novel types of populations and samples and, hopefully, will be conservative in their conclusions. This is important, since there will now be many more comparisons (multiple phenotypes and even more multiple genotypes) resulting in many more type 1 errors leading to false conclusions.

In the more distant future, health care providers will be enriching personalized medicine with risk factors, which will be easy to determine, and appropriate interventions will be available to those who wish to partake

of them. Hopefully, such information will be available to all who seek it, at a reasonable cost. This is likely to happen if testing for risk factors is not subject to high-cost royalties. Like many of you, I am very concerned about gene patenting but am hopeful that risk-factor genes will not be granted patents.

Education of the public and health care providers, including genetic counselors, will be essential. This should be an important part of our mission in the future. Our society is planning to substantially upgrade its educational mission. It is critical that all of us be involved in this endeavor.

Finally, I leave you with two thoughts. If you are in doubt about the power of your individual study, *collaborate*. For the public and health care delivery personnel who are very likely to be in doubt on the relevance of risk factors, *educate*.

Acknowledgments

The author would like to thank the members of COGA who have inspired him to extol the benefits of collaborative efforts in science.

References

Bice P, Foroud T, Bo R, Castelluccio P, Lumeng L, Li T-K, Carr LG (1998) Genomic screen for QTLs underlying alcohol consumption in the P and NP rat lines. Mamm Genome 9: 949–955

Edenberg HJ, Foroud T, Koller DL, Goate A, Rice J, Van-Eerdewegh P, Reich T, Cloninger CR, Nurnberger JI, Kowalczuk M, Wu B, Li T-K, Conneally PM, Tischfield JA, Wu W, Shears S, Crowe R, Hesselbrock V, Schucket M, Porjesz B, Begleiter H (1998) A family-based analysis of the association of the dopamine D2 receptor (DRD2) with alcoholism. Alcohol Clin Exp Res 22:505–512

Foroud T, Edenberg HJ, Goate A, Rice J, Flury L, Koller DL, Bierut LJ, Conneally PM, Nurnberger JI, Bucholz KK, Li T-K, Hesselbrock V, Crowe R, Schuckit M, Porjesz B, Begleiter H, Reich T (2000) Alcoholism susceptibility loci: confirmation studies in a replicate sample and further mapping. Alcohol Clin Exp Res 24:933–945

Horikawa Y, Oda N, Cox NJ, Li X, Orho-Melander M, Hara M, Hinokio Y, et al (2000) Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. Nat Genet 26:163–175

Morzorati SL, Ramchandani VA, Flury L, Li T-K, O'Connor S (2002) Self-reported subjective perception of intoxication reflects family history of alcoholism when breath alcohol levels are constant. Alcohol Clin Exp Res 26:1299–1306

Noble EP, Syndulko K, Fitch RJ, Ritchie T, Bohlman MC, Guth P, Sheridan PJ, Montgomery A, Heinzmann C, Sparkes RS, Blum K (1994) D2 dopamine receptor TaqI A alleles in medically ill alcoholic and nonalcoholic patients. Alcohol 29: 729–744

Porjesz B, Almasy L, Edenberg HJ, Wang K, Chorlian DB, Foroud T, Goate A, Rice JP, O'Connor SJ, Rohrbaugh J, Kuperman S, Bauer LO, Crowe RR, Schuckit MA, Hesselbrock V, Conneally PM, Tischfield JA, Li T-K, Reich T, Begleiter H (2002) Linkage disequilibrium between the beta frequency of the human EEG and a $GABA_A$ receptor gene locus. Proc Natl Acad Sci USA 99:3729–3733

Saccone NL, Kwon JM, Corbett J, Goate A, Rochberg N, Edenberg HJ, Foroud T, Li T-K, Begleiter H, Reich T, Rice JP (2000) A genome screen of maximum number of drinks as an alcoholism phenotype. Am J Med Genet 96:632–637

Suarez BK, Hampe CL, Van Eerdewegh PV (1994) Problems of replicating linkage claims in psychiatry. In: Geshon ES, Cloninger CR (eds) Genetic approaches to mental disorders. American Psychiatric Association Press, Washington, DC, pp 23–46

Thomasson HR, Edenberg HJ, Crabb DW, Mai XL, Jerome RE, Li T-K, Wang SP, Lin YT, Ly RB, Yin SJ (1991) Alcohol and aldehyde dehydrogenase genotypes and alcoholism in Chinese men. Am J Hum Genet 48:677–681