

Genetics of Atherosclerosis: Some Strategies for Studies of Apolipoprotein E

PATRICIA PEYSER MOLL

Department of Epidemiology, University of Michigan, Ann Arbor, Michigan 48109

ABSTRACT Many genes are hypothesized to be involved in determining an individual's risk for coronary artery disease (CAD). Recent efforts have focused on the genetic architecture of quantitative traits related to risk for CAD. Studies relating genetic variation in the structural gene for apolipoprotein E to plasma levels of apolipoprotein E illustrate strategies to begin to understand the genetic architecture of plasma levels of apolipoprotein E. Studies using a measured gene approach suggest that variability in the isoforms of apolipoprotein E explain some, but not all, of the variability in plasma levels of apolipoprotein E. Products of other genes may be associated with plasma apolipoprotein E variability. No studies to date have presented findings from an unmeasured gene approach to plasma levels of apolipoprotein E. Models most often used in the unmeasured gene approach are not appropriate for studies of plasma levels of apolipoprotein E and perhaps are inappropriate for the study of other traits. Variations of the models are suggested to ask if a single unmeasured gene is the same gene defined by the apolipoprotein E isoforms. Another model can be used to ask if there is evidence for genes influencing levels of apolipoprotein E after accounting for the effects of the isoforms. Both the measured gene and unmeasured gene strategies have limitations. The failure of most models to allow for the complexity of genotype-phenotype relationships or the heterogeneity of genetic causes will slow the progress to understand the genetic architecture of quantitative traits associated with risk for CAD. © 1993 Wiley-Liss, Inc.

In the United States, coronary artery disease (CAD) is the single largest cause of death, accounting for 24% of all deaths (American Heart Association, 1991). CAD represents the clinical manifestations of coronary atherosclerosis. CAD begins in childhood with the accumulation of lipid in the intima of arteries to form fatty streaks (McGill, 1988). At some sites in the coronary arteries, the fatty streaks thicken by the continued accumulation of abnormal lipid, of smooth muscle cells, and of connective tissue to form fibrous plaques. In young adulthood, fibrous plaques increase in extent and thickness and undergo calcification and vascularization. In middle years and later, qualitative changes in the fibrous plaque may result in hemorrhage, ulceration, and thrombosis, potentially causing arterial occlusion and catastrophic ischemic necrosis of vital organs (McGill, 1988). In the general

population, clinical disease is manifested typically in the fifth and sixth decades as angina pectoris or as sudden plaque rupture with thrombosis, vessel occlusion, and myocardial infarction with or without sudden death. Approximately 800,000 myocardial infarctions occur yearly in the United States, and more than 500,000 deaths and 2 million hospitalizations annually are associated with CAD (Higgins and Thom, 1989).

The major risk factors for CAD include older age, male gender, smoking, family history of premature disease, hypertension, diabetes, obesity, and abnormal levels of lipids (cholesterol and triglyceride) and lipoproteins (low-density lipoprotein cholesterol [LDL-C] and high-density lipoprotein cho-

Received November 30, 1992; accepted March 19, 1993.

lesterol [HDL-C]). Other risk factors have been identified including fat patterning, physical inactivity, dietary excesses or deficiencies, and a stressful lifestyle. Case-control studies, twin studies, and family studies have provided evidence that CAD aggregates in families (Berg, 1983; Reed et al., 1986; Nora et al., 1991). However, in most families, CAD does not segregate as if determined by a single Mendelian gene. This is not unexpected when the risk of developing disease is determined by the levels of many intermediate quantitative traits whose distributions are each influenced by the segregation of many genes and environmental factors (Sing and Moll, 1990a).

Several studies have presented evidence that genetic differences among individuals can explain at least 40% of the variability in LDL-C and 30% of the variability in HDL-C (Sing et al., 1988). These studies provide estimates of the contribution from all of the genes together to variability in a quantitative trait under the assumptions that the contributions from the genes are all equal, small, and additive. There are also rare mutations that can explain Mendelian segregation of very high LDL-C levels and very low HDL-C levels in some families (Sing et al., 1988). Since only a small proportion of families have these rare mutations, they can only explain a small proportion of the genetic variability in these traits and CAD in the population at large.

GENETIC ARCHITECTURE

Many genes are hypothesized to be involved in determining an individual's risk for CAD (Sing and Moll, 1990a). Specific risk factors, like measures of lipid metabolism, are associated with variation in several genes (Sing and Boerwinkle, 1987; Kaprio et al., 1991; Kessling et al., 1991). As a consequence of contributions from many genes and the pleiotropic effect of individual genes, recent efforts have focused on genetic architecture of quantitative traits related to risk for CAD (Sing and Davignon, 1985; Boerwinkle and Sing, 1987; Sing and Boerwinkle, 1987; Kaprio et al., 1991; Boerwinkle, 1992). The genetic architecture of a trait in a particular population is defined by the number of genes involved in the phenotypic expression of the trait, the number of alleles of each gene and their relative frequencies, the impact of each allele on the level and variability of the trait, and the impact of

each allele on the relationship between the trait and other risk factors for the disease of interest (Sing and Moll, 1990b).

Expectations about the relative frequencies of alleles and the size of their effects on CAD risk have come from many studies (Sing and Moll, 1990a,b). These include studies of the gene coding for the LDL receptor and the defects in it which are associated with familial hypercholesterolemia (Brown and Goldstein, 1986), the gene coding for apolipoprotein E and the protein isoforms (Davignon et al., 1988), and the genes and restriction fragment length polymorphism (RFLP) markers in them coding for other proteins hypothesized to be involved in regulation of lipid metabolism (Humphries, 1988; Kessling et al., 1991). It is expected that there will be a few mutations or alleles that will be rare but have large effects on risk factors for CAD; yet, they will have small effects on the risk of CAD. It is expected that most of the genetic effects on CAD risk will probably come from common allelic variation for many genes with small effects. As the number of genes involved in the genetic architecture increases, the number of families segregating for the same genes is expected to decrease (Sing and Moll, 1990b). Studies to understand the genetic architecture of CAD and its risk factors have begun. Among quantitative risk factors for CAD, plasma lipoprotein (a) is an example of a trait with more than 90% of the genetic variation identified (Boerwinkle, 1992; Boerwinkle et al., 1992). In contrast to what is known about the genetic architecture of plasma lipoprotein (a), less is known about the genetic architecture of other risk factors and little is known about the genetic architecture of CAD at this time.

APOLIPOPROTEINS

To begin to understand the genetic architecture of CAD and its risk factors, many studies have focused on measures of lipid metabolism. The two major lipids, cholesterol and triglyceride, are insoluble and circulate between the gut, liver, and peripheral tissues as lipoprotein particles. It is hypothesized that many genes are involved in the regulation of lipid uptake in the gut, metabolism in the plasma, liver, and peripheral cells and removal from the body. One way to look for genetic influence is to study the protein components of these lipoproteins since the proteins are encoded by genes. Apoli-

poproteins, protein constituents of lipoproteins, have various structural and metabolic roles related to the metabolism of lipids. Although all of the apolipoproteins have important structural roles in the transport of lipids, most have specialized functions. Chan et al. (1990) review the structure and function of specific apolipoproteins, while Ferrell (1992) reviews many of the known associations between polymorphic variation in apolipoprotein genes and quantitative variation in lipid and lipoprotein levels.

Assays have been developed to measure the quantitative plasma levels of many apolipoproteins (Kottke et al., 1991). Since the chromosomal location of many of the structural genes for apolipoproteins is known (Chan et al., 1990), variation in these genes can be related to variation in plasma levels of apolipoproteins. There are many studies relating genetic variation in the structural gene for apolipoprotein E to plasma levels of apolipoprotein E (Boerwinkle and Utermann, 1988; Havekes et al., 1988; Hanis et al., 1991; Kaprio et al., 1991; Reilly et al., 1991). These studies, as well as several in progress, illustrate some of the strategies being implemented to begin to understand the genetic architecture of plasma levels of apolipoprotein E.

APOLIPOPROTEIN E

Apolipoprotein E is synthesized in various organs including the liver, brain, spleen, and kidney, and is present in high concentrations in interstitial fluid, where it appears to participate in cholesterol redistribution from cells with excess cholesterol to those requiring cholesterol (Mahley, 1988). The three major isoforms of apolipoprotein E, referred to as apo E2, apo E3 and apo E4, are the products of three alleles ($\epsilon 2$, $\epsilon 3$, $\epsilon 4$) of a single gene on chromosome 19. The isoforms can be determined by isoelectric focusing (Kamboh et al., 1988). Apo E3 differs from apo E2 and apo E4 by one amino acid difference at either residue 112 or 158. Functional differences have been identified among the isoforms. Apo E2 has impaired receptor binding activity compared to apo E3, while apo E4 has enhanced *in vivo* catabolism (Mahley, 1988).

One assumption that is made in studies using the isoforms is that individuals with the same isoform phenotypes have the same genotypes. In a direct comparison of apo E isoforms measured by isoelectric focusing

(Kamboh et al., 1988) and apo ϵ genotypes at the two common polymorphic sites (amino acids 112 and 158), there was only one discordant allele in a sample of 195 unrelated heterozygotes from a population-based sample (Mailly et al., 1992). This finding suggests that apo E isoforms can be used to infer apo ϵ genotypes.

Different apo E isoforms are associated with different risks of atherosclerosis. Two recent cross-sectional studies have shown this association. The Pathobiological Determinants of Atherosclerosis in Youth (PDAY) study by Hixson et al. (1991) considered lesions in autopsied young males, while the other study considered angiographically defined disease in a sample of adult males (Nieminen et al., 1992). In both studies, individuals who were homozygotes or heterozygotes for apo E4 had higher prevalence of disease. In a nested case-control study of participants in the Multiple Risk Factor Intervention Trial, the presence of apo E4 was associated with an increased risk of disease that was most evident for fatal cases (Eichner et al., 1993).

TWO STRATEGIES FOR STUDYING THE RELATIONSHIP BETWEEN GENES AND PLASMA LEVELS OF APOLIPOPROTEINS

There are two strategies that have been used to study the relationship between genes and quantitative traits like apolipoproteins (Boerwinkle et al., 1987; Sing and Boerwinkle, 1987). The first strategy has been called a bottom-up, measured gene approach. It involves obtaining measures of genes and of quantitative trait levels in an unrelated sample and then estimating the effects of genes on the trait of interest. Examples of this approach investigating the association of apo E isoforms with levels of plasma apolipoprotein E are given in the studies by Boerwinkle and Utermann (1988), Kaprio et al. (1991), Hanis et al. (1991), and Reilly et al. (1991). Some of these studies are discussed in detail later.

The second strategy has been called a top-down, unmeasured gene approach. It includes obtaining measures of a quantitative trait in a sample of pedigrees and then applying statistical techniques (i.e., complex segregation analysis [Elston and Stewart, 1971; Lalouel et al., 1983]) to search for evidence that unmeasured genes are associated with variation in the quantitative trait. These unmeasured genes can include

polygenes and/or an unmeasured single gene with a large impact on the level of a quantitative trait. Examples of unmeasured gene approaches in studies of quantitative levels of apolipoproteins in pedigrees are given by Hasstedt et al. (1987), Pairitz et al. (1988), and Moll et al. (1989). If there is statistical evidence that an unmeasured single gene has a large impact on the trait, then specific pedigrees that are most likely to be segregating for this unmeasured single gene can be identified. These pedigrees would then be used in future molecular studies. An unmeasured gene approach can be useful when there are no candidate genes or candidate genes cannot be measured. Such an approach can also be useful to provide evidence for new genes even if some genes have already been identified and measured. Alternatively, an unmeasured gene approach can provide an estimate of the total genetic contribution to the variation in a trait. No studies, to date, have presented an application of a top-down, unmeasured gene approach to plasma levels of apolipoprotein E.

PLASMA APOLIPOPROTEIN E: MEASURED GENE APPROACH

The impact of the common alleles at the structural gene coding for apo E on plasma apolipoprotein E has been estimated in 453 unrelated Caucasians 26–63 years of age in a population-based sample (Kaprio et al., 1991). The relative frequencies of the three alleles that code for the apo E isoforms in this sample (apo $\epsilon_2 = 0.073$, $\epsilon_3 = 0.777$, and $\epsilon_4 = 0.150$) are similar to those reported for other Caucasian populations in North America (Davignon et al., 1988; Gerdes et al., 1992). The distribution of genotypes is not different between males and females and the pooled frequencies are in Hardy-Weinberg equilibrium. In the 227 men in this sample, plasma levels of apolipoprotein E ranged from 1.5–40.7 mg/dl with a mean of 5.5 mg/dl. In the 226 women, the levels ranged from 1.3–13.4 mg/dl with a mean of 5.2 mg/dl. In the men, 4.5% of the sample variability was explained by linear regression of plasma levels of apolipoprotein E on date of assay, age, age squared, height, weight, and smoking status. In women, these same concomitants explained 6.3% of the sample variability. After adjustment for concomitants, Kaprio et al. (1991) estimated the contribution of apo E isoforms to the population variance of apolipoprotein E was

11.6% in males and 16.8% in females. This contribution was estimated using the unbiased statistic given by Boerwinkle and Sing (1986).

A measured gene approach has been applied to plasma apolipoprotein E in two other studies. Hanis et al. (1991) estimated the contribution of apo E isoforms to plasma apolipoprotein E variability at 0.9% in males, 3.9% in postmenopausal females, and 27.5% in premenopausal females sampled from Starr County, Texas. Boerwinkle and Utermann (1988) estimated the contribution to be 20.2% in a sample of male and female blood bank donors from Marburg and Geissen. The findings from these studies suggest that differences in apo E isoforms are associated with differences in plasma apolipoprotein E levels. However, the association varies by population, gender, and menopausal status. These two studies differ from one another and from the study of unrelated adults by Kaprio et al. (1991) in the age range of the participants, the specific concomitants considered as well as the methods used to adjust plasma levels of apolipoprotein E for concomitants, the statistical methods used to estimate the contribution of isoform variability to plasma apolipoprotein E variability, the frequencies of the isoforms, and the frequencies of other genes and environmental exposures. These three studies together support the inference that variability in the isoforms does not explain all of the variability in plasma levels of apolipoprotein E and that the contribution from apo E isoforms to plasma levels of apolipoprotein E varies within and between populations (Boerwinkle and Utermann, 1988; Hanis et al., 1991; Kaprio et al., 1991). These studies also suggest that, although variation in the apo E isoforms explains some of the plasma apolipoprotein E variability, the products of other genes or other genes themselves may also be associated with apolipoprotein E variability.

PLASMA APOLIPOPROTEIN E: UNMEASURED GENE APPROACH

No studies to date have presented the findings from an unmeasured gene approach to plasma levels of apolipoprotein E. Any unmeasured gene study would need to consider the findings from the measured gene studies when selecting the models to be fit to the family data. The models that have been applied most often for quantitative

traits would not be appropriate for plasma levels of apolipoprotein E and would need to be extended. The models most often used have been described in detail elsewhere (Moll et al., 1989, 1991; MacCluer, 1992). Briefly, these models assume that the observed distribution of a quantitative trait is a consequence of the independent contributions from the following: a single genetic or nontransmitted environmental factor with a large effect on the level of the quantitative trait, the effects of a large number of independent polygenes each with a small, additive effect, the effects of shared environments (often defined by household membership), and individual-specific environmental influences. The major factor is modeled as having two alternatives (L for lower and H for higher) that may be of either genetic or nontransmitted environmental origin. These combine to define three distinct groups or classes (also referred to as ousiotypes [Cannings et al., 1978]) of individuals, denoted LL, LH, and HH. The relative frequency of L is denoted p and the relative frequency of H is equal to $1-p$. Most applications of the unmeasured gene approach assume that there is Hardy-Weinberg equilibrium in the population being sampled and so the relative frequency of LL, LH, and HH individuals is p^2 , $2pq$, and q^2 , respectively. Other parameters of the model include a phenotypic mean for each ousiotype (LL, LH, and HH) and a phenotypic variance among individuals with the same ousiotype. It is assumed that this variance is the same for all three ousiotypes. Based on the findings from studies using the measured gene approach, this assumption may not be valid (Reilly et al., 1991). The model partitions this variance into a fraction attributable to polygenes, a fraction attributable to shared environmental effects, and a fraction that is attributable to individual-specific environmental effects and measurement error. τ_1 , τ_2 , and τ_3 are the probabilities that individuals of ousiotypes LL, LH, and HH, respectively, transmit the L factor to their offspring. These transmission probabilities define alternative models. For the general transmission model (Lalouel et al., 1983), these transmission probabilities are each estimated under the constraint that they be inclusive 0–1. A single gene Mendelian model defines these probabilities of transmitting alleles from parents to offspring as being $\tau_1 = 1.0$, $\tau_2 = 0.5$, and

$\tau_3 = 0.0$. The nontransmitted environmental factor model predicts that the probability that an individual is one ousiotype or another is independent of his or her generation and independent of the ousiotypes of his or her parents (Elston and Stewart, 1971). Under this model, each transmission probability is taken to be equal to p , the relative frequency of L.

Testing hypotheses about parameters corresponds to restricting one or more parameters to specific hypothesized values, while estimating the remaining parameters from the data. The model in which all three transmission probabilities are estimated is considered the unrestricted model and is often called the general transmission model. The restricted models include a model with the transmission probabilities constrained to be equal to Mendelian expectation (a single gene model) or a model in which the transmission probabilities are all equal to the relative frequency of L (a nontransmitted environmental model).

For the unrestricted and restricted models, a likelihood is computed that is a function of the observed data and the parameters of the model. These likelihoods are a posteriori probabilities after the data have been observed. Twice the difference between the maximum of the \log_e likelihood of the model with unrestricted parameters and the maximum of the \log_e likelihood of a model with restricted parameters is distributed approximately as a chi-square when the null hypothesis is true. The degrees of freedom for this chi-square test are equal to the number of parameters restricted to values stated by the hypothesis. The likelihoods of the models can be computed using the Pedigree Analysis Package (PAP) (Hasstedt, 1989). The likelihoods for each model considered can be maximized using a quasi-Newton method (Lalouel, 1979). Several other computer programs are available to compute exact likelihoods as well as approximations to the likelihoods for the models described above (MacCluer, 1992).

The models described above, with or without the shared environmental effects, have been used extensively (see MacCluer, 1992) and are currently being used (e.g., Schumacher et al., 1992). However, they are not appropriate for plasma apolipoprotein E given that six genotypes defined by apo $\epsilon 2$, apo $\epsilon 3$, and apo $\epsilon 4$ have been shown in several studies to explain a large proportion of

the variability. The models used in the unmeasured approach would first have to consider three alternatives for the factor, rather than two. A model with three alleles has only been used once before in a study of lipoprotein (a) (Hasstedt and Williams, 1986). A model with three alternatives for the factor would have six means. Under the single gene model, the transmission probabilities would be fixed at their Mendelian expectation. However, under the nontransmitted environmental factor model, there would be two transmission probabilities that would need to be estimated and under the unrestricted model there would be 12 transmission probabilities that would need to be estimated. Given the difficulties in estimating transmission probabilities, it seems unlikely that an unrestricted model with three alternatives for the factor with a large effect on the levels of the trait will be applied in the near future. While it is possible to fit a single gene model and a nontransmitted environmental model with three alternatives for the factor with a large effect, the comparison of these models is complicated by the fact that neither of these models can be considered unrestricted with respect to the other model. That is, these two models are not nested. The inability to be able to compare different models with three alternatives for the factor illustrates the limitations in the models to allow for the complexity of the genotype-phenotype relationships. The potential for future studies to provide statistical evidence for new genes will depend, to some extent, on how much the models evolve in the next few years. In addition to limitations in the models, there are a number of known sources of both Type I and Type II errors that could jeopardize the validity of findings from a top-down, unmeasured gene approach (Moll, 1993).

Although the models used most often are not appropriate and models with three alternatives for the factor introduce estimation and hypothesis testing challenges, there is one specific model that could be applied to plasma levels of apolipoprotein E that would address a very specific question. The question that could be addressed is whether a single unmeasured gene with three alternative factors and a large effect on plasma levels of apolipoprotein E is the same gene defined by the gene that codes for the apo E isoforms. The model would include the estimation of relative frequencies of two alleles

since the relative frequency of the third allele is the difference between 1 and the sum of the estimates of the relative frequencies of the other two alleles. The model would also include the estimation of six genotype-specific means, a phenotypic variance among individuals with the same genotype, a parameter that reflects the effects of polygenes, and a parameter that reflects the effects of shared environmental factors. The most probable genotype could be assigned to each individual in the study and a test of independence between the genotype at the unmeasured gene and the genotypes that code for the apo E isoforms could be made. This application of the unmeasured approach would provide an assessment of whether complex segregation analysis (i.e., an unmeasured gene approach) can identify polymorphic genes that are known to affect levels of a quantitative trait.

Simulation studies have suggested that the top-down, unmeasured gene approach is limited to detecting only those single gene effects that are relatively large (MacLean et al., 1975; Go et al., 1978; Burns et al., 1984). The size of the effects is a function of the relative frequency of the alleles, the means for the genotypes, and the variance within each genotype. In general, simulation studies have suggested that the difference between the means for different genotypes needs to be at least one and a half within genotype standard deviations apart in order to be able to detect the effects of a single gene (MacLean et al., 1975; Go et al., 1978; Burns et al., 1984). The reported difference between means for plasma levels of apolipoprotein E for the six genotypes defined by the apo E isoforms is not each one and a half standard deviations in the study by Hani et al. (1991). Therefore, it is not clear whether the genes that code for the apo E isoforms have effects that are large enough to be detected with the top-down, unmeasured gene approach.

A different application of the unmeasured approach could address the question of whether there is evidence that there are genes, in addition to the gene that codes for the apo E isoforms, that influence the levels of plasma apolipoprotein E. The model to address this question would include unmeasured genes as well as the gene that codes for the apo E isoforms. In the model, the different isoforms would be associated with different levels of plasma apolipoprotein E

and the variability among individuals with the same isoform would be explained by polygenes, shared environments, and residual effects. This model was originally developed for nuclear families (Boerwinkle et al., 1986; Boerwinkle and Sing, 1987).

This model could be extended to allow the effects of the gene that codes for the isoforms to be gender specific. Then the unrestricted model would include 12 means: six genotype-specific means for males and six genotype-specific means for females. The other parameters would include a standard deviation that is equal for all genotypes and both genders as well as parameters for the contributions of polygenes and shared environments to the variability among individuals with the same genotype and gender. Restricted models would be associated with hypotheses that the isoforms do not contribute to variability and/or the unmeasured polygenes do not contribute to variability and/or the shared environments do not contribute to variability. Other unrestricted models could test the hypothesis that the isoform means are equal for males and females. The maximum of the log_e likelihoods for the unrestricted and restricted models could be computed using PAP (Hasstedt, 1989). If the restricted models that did not include isoforms or did not include polygenes were rejected when compared to the unrestricted model, that would be evidence that there are genes, in addition to the gene that codes for the apo E isoforms, that influence the plasma levels of apolipoprotein E. Preliminary studies in a population-based sample of pedigrees suggest that there is evidence for these other, new genes and that the genetic architecture of apolipoprotein E will not be limited to the gene that codes for the apo E isoforms.

FUTURE STUDIES

In thinking about the design of future studies to identify these new genes, one approach would be to limit measured or unmeasured studies to pedigrees in which all individuals are homozygotes for apo ϵ 3. Apo ϵ 3 has the highest relative frequency in every population studied so far (Gerdes et al., 1992). Based on the relative frequencies reported by Kaprio et al. (1991), it is expected that 60.4% of the individuals from the population sampled by that study would be homozygotes for apo ϵ 3. Making assumptions, 36.4% of all matings would be between two

homozygotes for apo ϵ 3. If one were sampling three generation pedigrees, then approximately 13.2% of all these pedigrees would have four grandparents who were homozygotes for apo ϵ 3. While a higher frequency of nuclear families would have all homozygotes for ϵ 3 compared to three generation pedigrees, the three generation pedigrees would most likely provide more power to detect segregation (Burns et al., 1984). Restricting either measured gene or unmeasured gene approaches to the nuclear families or pedigrees in which all individuals are homozygotes for apo ϵ 3 would control for the isoform variability. If the unmeasured approach provided evidence for a single gene with a large effect on plasma levels of apolipoprotein E, then specific pedigrees could be identified, in which all the individuals are ϵ 33 homozygotes, that are the most likely to be segregating for the unmeasured single gene with a large effect. These pedigrees could be the focus of future molecular studies. Alternatively, sibling pairs in all the pedigrees in which all individuals are ϵ 33 homozygotes could be used to detect linkage between a measured gene and quantitative levels of apolipoprotein E (Haseman and Elston, 1972; Amos et al., 1989). This model-free approach has been used for quantitative levels of lipoprotein(a) (Boerwinkle et al., 1992).

CONCLUSIONS

The genetic architecture of plasma apolipoprotein E is just beginning to emerge and will most probably not be simple. Apo E isoforms do explain some of the variability in plasma levels of apolipoprotein E. It is very likely that there are genes other than the gene that codes for the apo E isoforms that also influence plasma apolipoprotein levels. The challenge will be to identify these other genes.

The measured and unmeasured gene approaches provide evidence that genes do play a role in the variability of the levels of many quantitative traits thought to be important for CAD. However, any inferences from either the measured or unmeasured approach will have to be interpreted with caution for any quantitative trait. The impact of variation in any specific gene will depend on a number of host characteristics such as other genes, age, gender, concomitant diseases, and environmental exposures (Sing et al., 1992). There are many examples

in the literature that suggest that gene-gene interaction as well as gene-environment interaction may not be rare phenomena (Nestruck et al., 1987; Reilly et al., 1992; Sing et al., 1992). While stratification might be a good approach for certain situations to control for gene-gene interactions (i.e., consideration of only pedigrees with $\epsilon 33$ homozygotes), it is clear that stratification will not be appropriate as a method to control for many genes simultaneously or to control for many concomitants.

Almost all of the models used so far in the measured and unmeasured approaches are naive representations and fail to allow for complexity of genotype-phenotype relationships or the heterogeneity of genetic causes among individuals and pedigrees that make up the population at large (Sing and Reilly, 1993). Recently, there have been applications using both the measured and unmeasured approach that have begun to incorporate genetic interactions with concomitants like age and gender (Perusse et al., 1991; Rebbeck et al., 1991; Reilly et al., 1991, 1992, 1993). Failure to incorporate these concomitants will slow the progress to understand the genetic architecture of quantitative traits associated with the development and progression of atherosclerosis.

ACKNOWLEDGMENTS

I wish to thank Dr. Charles F. Sing for numerous helpful discussions on many of the ideas presented here. I also thank Dr. Robert Ferrell for his helpful comments on the paper and the many discussions we have had about apolipoprotein E. Two anonymous reviewers provided many helpful suggestions. Support was provided by HL39107 and HL46292.

LITERATURE CITED

- American Heart Association (1991) Heart and Stroke Facts. Dallas: American Heart Association.
- Amos CI, Elston RC, Wilson AF, Bailey-Wilson JE (1989) A more powerful robust sib-pair test of linkage for quantitative traits. *Genet. Epidemiol.* 6:435-449.
- Berg K (1983) Genetics of coronary heart disease. In AG Steinberg, AG Bearn, AG Motulsky, and B Childs (eds.): *Progress in Medical Genetics*. Vol. 5. Philadelphia: Saunders, pp:35-90.
- Boerwinkle E (1992) Genetics of plasma lipoprotein (a) concentrations. *Curr. Opin. Lipidol.* 3:128-136.
- Boerwinkle E, Chakraborty R, Sing CF (1986) The use of measured genotype information in the analysis of quantitative phenotypes in man. I. Models and analytical methods. *Ann. Hum. Genet.* 50:181-194.
- Boerwinkle E, Leffert CC, Lin J, Lackner C, Chiesa G, Hobbs HH (1992) Apolipoprotein (a) gene accounts for greater than 90% of the variation in plasma lipoprotein (a) concentrations. *J. Clin. Invest.* 90:52-60.
- Boerwinkle E, Sing CF (1986) Bias of the contribution of single-locus effects to the variance of a quantitative trait. *Am. J. Hum. Genet.* 39:137-144.
- Boerwinkle E, Sing CF (1987) The use of measured genotype information of quantitative phenotypes in man III. Simultaneous estimates of the frequencies and effects of the apolipoprotein E polymorphism and residual polygenic effects on cholesterol, betalipoprotein and triglyceride levels. *Ann. Hum. Genet.* 51:211-226.
- Boerwinkle E, Utermann G (1988) Simultaneous effects of the apolipoprotein E polymorphism on apolipoprotein E, apolipoprotein B, and cholesterol metabolism. *Am. J. Hum. Genet.* 42:104-112.
- Boerwinkle E, Visvikis S, Welsh D, Steinmetz J, Hanash SM, Sing CF (1987) The use of measured genotype information in the analysis of quantitative phenotypes in man. II. The role of the apolipoprotein E polymorphism in determining levels, variability, and covariability of cholesterol, betalipoprotein, and triglycerides in a sample of unrelated individuals. *Am. J. Med. Genet.* 27:567-582.
- Brown MS, Goldstein JL (1986) A receptor-mediated pathway for cholesterol homeostasis. *Science* 232:34-47.
- Burns TL, Moll PP, Schork MA (1984) Comparisons of different sampling designs for the determination of genetic transmission mechanisms in quantitative traits. *Am. J. Hum. Genet.* 36:1060-1074.
- Cannings C, Thompson EA, Skolnick MH (1978) Probability functions on complex pedigrees. *Adv. Appl. Prob.* 10:26-61.
- Chan L, Boerwinkle E, Li W-H (1990) Molecular genetics of the plasma apoproteins. In S Chien (ed.): *Molecular Biology of the Cardiovascular System*. Philadelphia: Lea and Febiger, pp. 183-219.
- Davignon J, Gregg RE, Sing CF (1988) Apolipoprotein E polymorphism and atherosclerosis. *Arteriosclerosis* 8:1-21.
- Eichner JE, Kuller LH, Orchard TJ, Grandits GA, McCallum LM, Ferrell RE, Neaton JD (1993) The relationship of apolipoprotein E phenotype to myocardial infarction and CHD mortality. *Am. J. Cardiol.* 71:160-165.
- Elston RC, Stewart J (1971) A general model for the genetic analyses of pedigree data. *Hum. Hered.* 21:523-542.
- Ferrell RE (1992) Genetics of the apolipoproteins and the contribution of allelic variation to quantitative variation in lipid and lipoprotein levels in the population. *Curr. Opin. Lipidol.* 3:122-127.
- Gerdes LU, Klausen IC, Sihm I, Faergeman O (1992) Apolipoprotein E polymorphism in a Danish population compared to findings in 45 other study populations around the world. *Genet. Epidemiol.* 9:155-168.
- Go RCP, Elston RC, Kaplan EB (1978) Efficiency and robustness of pedigree segregation analysis. *Am. J. Hum. Genet.* 30:28-37.
- Hanis CL, Hewett-Emmett D, Douglas TC, Bertin TK, Schull WJ (1991) Effects of the apolipoprotein E polymorphism on levels of lipids, lipoproteins, and apolipoproteins among Mexican-Americans in Starr County, Texas. *Arterioscler. Thromb.* 11:362-370.
- Haseman JK, Elston RC (1972) The investigation of linkage between a quantitative trait and a marker locus. *Behav. Genet.* 2:2-19.

- Hasstedt SH (1989) PAPA: Pedigree Analysis Package, rev 3. Department of Human Genetics, University of Utah, Salt Lake City.
- Hasstedt SJ, Williams RR (1986) Three alleles for quantitative Lp(a). *Genet. Epidemiol.* 3:53–55.
- Hasstedt SJ, Wu L, Williams RR (1987) Major locus inheritance of apolipoprotein B in Utah pedigrees. *Genet. Epidemiol.* 4:67–76.
- Havekes LM, de Knijff P, Smit M, Frants R (1988) The effect of apolipoprotein E allele substitutions on plasma lipid and apolipoprotein levels. *Adv. Exp. Med. Biol.* 243:87–93.
- Higgins M, Thom T (1989) Trends in CHD in the United States. *Int. J. Epidemiol.* 18:S58–S66.
- Hixson JE, the PDAY Research Group (1991) Apolipoprotein E polymorphisms affect atherosclerosis in young males. *Arterioscler. Thromb.* 11:1237–1244.
- Humphries SE (1988) DNA polymorphisms of the apolipoprotein genes—their use in the investigation of the genetic component of hyperlipidaemia and atherosclerosis. *Atherosclerosis* 72:89–108.
- Kamboh MI, Ferrell RE, Kottke BA (1988) Genetic studies of human apolipoproteins. V. A novel rapid procedure to screen apolipoprotein E polymorphism. *J. Lipid Res.* 29:1535–1543.
- Kaprio J, Ferrell RE, Kottke BA, Kamboh MI, Sing CF (1991) Effects of polymorphisms in apolipoproteins E, A-IV, and H on quantitative traits related to risk for cardiovascular disease. *Arterioscler. Thromb.* 11:1330–1348.
- Kessling A, Ouellette S, Bouffard O, Chamberland A, Betard C, Selinger E, Xhignesse M, Lussier-Cacan S, Davignon J (1991) Patterns of association between genetic variability in apolipoprotein (Apo) B, Apo AI-CIII-AIV, and cholesterol ester transfer protein gene regions and quantitative variation in lipid and lipoprotein traits: Influence of gender and exogenous hormones. *Am. J. Hum. Genet.* 50:92–106.
- Kottke BA, Moll PP, Michels VV, Weidman WH (1991) Levels of lipids, lipoproteins and apolipoproteins in a defined population. *Mayo Clin. Proc.* 66:1198–1208.
- Lalouel JM (1979) GEMINI: A computer program for optimization of a linear function. Tech rep 13, Department of Medical Biophysics and Computing, University of Utah, Salt Lake City.
- Lalouel JM, Rao DC, Morton NE, Elston RC (1983) A unified model for complex segregation analysis. *Am. J. Hum. Genet.* 35:816–826.
- MacChuer JW (1992) Biometrical studies to detect new genes with major effects on quantitative risk factors for atherosclerosis. *Curr. Opin. Lipidol.* 3:114–121.
- MacLean CJ, Morton NE, Lew R (1975) Analysis of familial resemblance. IV. Operational characteristics of segregation analysis. *Am. J. Hum. Genet.* 27:365–384.
- Mahley RW (1988). Apolipoprotein E: Cholesterol transport protein with expanding role in cell biology. *Science* 240:622–630.
- Maily F, Moll PP, Kottke BA, Kamboh MI, Humphries SE, Ferrell RE (1992) Estimation of frequency of isoform-genotype discrepancies at the apolipoprotein E locus in heterozygotes for the isoforms. *Genet. Epidemiol.* 9:239–248.
- McGill HC Jr (1988) The pathogenesis of atherosclerosis. *Clin. Chem.* 34(Suppl):B33–B39.
- Moll PP (1993) Biometrical searches for new genes with effects on quantitative risk factors for atherosclerosis: Problems and prospects. *Curr. Opin. Lipidol.* 4:144–151.
- Moll PP, Burns TL, Lauer RM (1991) *The genetic and environmental sources of body mass index variability: The Muscatine Ponderosity family study.* *Am. J. Hum. Genet.* 49:1243–1255.
- Moll PP, Michels VV, Weidman WH, Kottke BA (1989) Genetic determination of plasma apolipoprotein AI in a population based sample. *Am. J. Hum. Genet.* 44:124–139.
- Nestruck AC, Bouthillier D, Sing CF, Davignon J (1987) Apolipoprotein E polymorphism and plasma cholesterol response to probucol. *Metabolism* 36:743–747.
- Nieminen MS, Mattila KJ, Aalto-Setälä K, Kuusi T, Kontula K, Kauppinen-Makelin R, Ehnholm C, Jauhainen M, Valle M, Taskinen M-R (1992) Lipoproteins and their genetic variation in subjects with and without angiographically verified coronary artery disease. *Arterioscler. Thromb.* 12:58–69.
- Nora JJ, Berg K, Nora AH (1991) *Cardiovascular Diseases: Genetics, Epidemiology and Prevention.* New York: Oxford.
- Pairitz GL, Davignon J, Maillous H, Sing CF (1988) Sources of interindividual variation in the quantitative levels of apolipoprotein B in pedigrees ascertained through a lipid clinic. *Am. J. Hum. Genet.* 43:311–321.
- Perusse L, Moll PP, Sing CF (1991) Evidence that a single gene with gender- and age-dependent effects influences systolic blood pressure determination in a population-based sample. *Am. J. Hum. Genet.* 49:94–105.
- Rebeck TR, Turner ST, Michels VV, Moll PP (1991) Genetic and environmental sources of interindividual variability in sodium-lithium countertransport in the Rochester, MN population. *Am. J. Hum. Genet.* 48:1092–1104.
- Reed T, Wagener DK, Donahue RP, Kuller LH (1986) Young adult cholesterol as a predictor of familial ischemic heart disease. *Prev. Med.* 15:292–303.
- Reilly SL, Ferrell RE, Kottke BA, Kamboh MI, Sing CF (1991) The gender-specific apolipoprotein E genotype influence on the distribution of lipids and apolipoproteins in the population of Rochester, MN. I. Pleiotropic effects on means and variances. *Am. J. Hum. Genet.* 49:1155–1166.
- Reilly SL, Ferrell RE, Kottke BA, Sing CF (1992) The gender-specific apolipoprotein E genotype distribution of plasma lipids and apolipoproteins in the population of Rochester, Minnesota. II. Regression relationships with concomitants. *Am. J. Hum. Genet.* 51:1311–1324.
- Reilly SL, Ferrell RE, Sing CF (1993) The gender-specific apolipoprotein E genotype distribution of plasma lipids and apolipoproteins in the population of Rochester, Minnesota. III. Pleiotropic effects on covariances and correlations (submitted, *Am. J. Hum. Genet.*).
- Schumacher MC, Hasstedt SJ, Hunt SC, Williams RR, Elbein SC (1992) Major gene effect for insulin levels in familial NIDDM pedigrees. *Diabetes* 41:416–423.
- Sing CF, Boerwinkle E (1987) Genetic architecture of interindividual variability in apolipoprotein, lipoprotein and lipid phenotypes. In G Bock, GM Collins (eds.): *Molecular Approaches to Human Polygenic Disease.* Ciba Foundation Symposium 130. Chichester: Wiley, pp. 99–127.
- Sing CF, Boerwinkle E, Moll PP, Templeton AR (1988) Characterization of genes affecting quantitative traits in humans. In BS Weir, MM Goodman, EJ Eisen, G Namkoong (eds.): *Proceedings of the Second International Conference on Quantitative Genetics.* Sunderland: Sinauer Associates, pp. 250–269.

- Sing CF, Davignon J (1985) Role of the apolipoprotein E polymorphism in determining normal plasma lipid and lipoprotein variation. *Am. J. Hum. Genet.* 37: 268–285.
- Sing CF, Haviland MB, Templeton AR, Zerba KE, Reilly SL (1992) Biological complexity and strategies for finding DNA variations responsible for inter-individual variation in risk of common chronic disease. *Ann. Med.* 24:539–547.
- Sing CF, Moll PP (1990a) Genetics of atherosclerosis. *Annu. Rev. Genet.* 24:171–187.
- Sing CF, Moll PP (1990b) Strategies for unravelling the genetic basis of coronary artery disease. In K Berg, N Retterstol, S Refsum (eds.): *From Phenotype to Gene in Common Disorders*. Copenhagen: Muskgaard, pp. 37–59.
- Sing CF, Reilly SL (1993) Genetics of common diseases that aggregate, but do not segregate, in families. In CF Sing, CL Hanis (eds.): *Genetics of Cellular, Individual, Family and Population Variability*. New York: Oxford University Press, pp. 140–161.