Sociological Methods & Research
Volume 37 Number 2
November 2008 164-200

10.1177/0049124108323538
http://smr.sagepub.com
hosted at
http://online.sagepub.com

The Importance of Gene–Environment Interaction

Implications for Social Scientists

Kari E. North
University of North Carolina at Chapel Hill
Lisa J. Martin
Cincinnati Children's Medical Hospital and the University of Cincinnati School of Medicine, Ohio

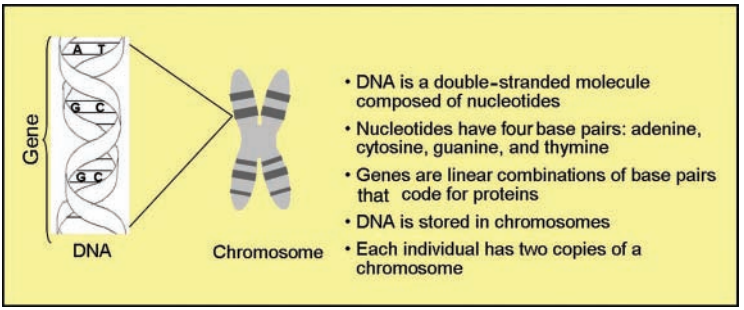
Given recent genetic advances, it is not surprising that genetics information is increasingly being used to improve health care. Thousands of conditions caused by single genes (Mendelian diseases) have been identified over the last century. However, Mendelian diseases are rare; thus, few individuals directly benefit from gene identification. In contrast, common complex diseases, such as obesity, breast cancer, and depression, directly affect many more individuals. Common complex diseases are caused by multiple genes, environmental factors, and/or interaction of genetic and environmental factors. This article provides a framework for the successful conduct of gene–environment studies. To accomplish this goal, the basic study designs and procedures of implementation for gene–environment interaction are described. Next, examples of gene–environment interaction in obesity epidemiology are reviewed. Last, the authors review reasons why epidemiological studies that incorporate gene–environment interaction have been unable to demonstrate statistically significant interactions and why conflicting results are reported.

Keywords: *genetics; epidemiology; complex inheritance; statistics*

Almost all human diseases result from gene-environment interaction. Proving, documenting, and quantifying this statement is a long-sought goal of the scientific community and one that, if achieved,

Authors' Note: Please address correspondence to Kari E. North, Associate Professor, Department of Epidemiology and Carolina Center for Genome Sciences, University of North Carolina at Chapel Hill, Bank of America Center, 137 E. Franklin St., Suite 306, Chapel Hill, NC 27514; e-mail: kari_north@unc.edu.

Figure 1
DNA

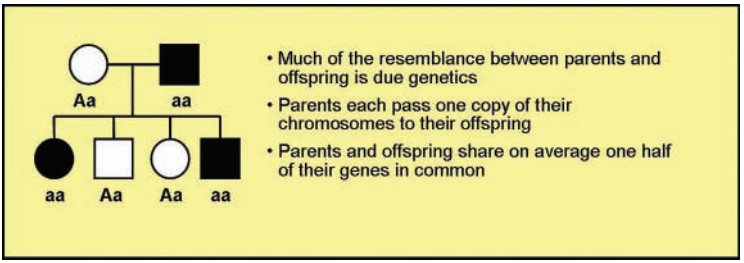


> could provide fundamental insights into the causes, courses, and prevention of many conditions.
>
> Botto and Khoury 2001:1016

Incorporating genetic information in medicine is a major challenge for the 21st century. Although for decades genetics has had a major role in the health care of some patients, soon specific genetic information will be important for delivery of effective health care to the many (Guttmacher and Collins 2002). Over the last 50 years, there has been a rapid accumulation of knowledge on the mechanisms of heredity. Deoxyribonucleic acid (DNA) is a double helix molecule that constitutes the basis of all heredity (see Figure 1). DNA carries the blueprint or set of instructions (often called *genes*) to make the proteins found in our bodies, in addition to having a key role in whether a protein is made and in its quantity and quality. It is in this manner that characteristics of individuals have a genetic basis. DNA is passed from one generation to the next (inheritance); therefore, the transmission of each protein "code" or blueprint is passed on following consistent inheritance patterns (see Figure 2). For example, all children inherit two copies of each gene, one from each parent. By following these simple rules, one can determine if traits of interest follow consistent patterns of transmission and begin to disentangle the complex web of genetic inheritance.

In June of 2000, the Human Genome Project and Celera Genomics jointly announced the first full sequences of the human genome (Lander et al. 2001). The sequence of the human genome is a powerful tool because

Figure 2
Inheritance



it provides information on not only the 3 billion base pairs making up the human genome, but also many of the subtle differences between individuals. This is important because any two unrelated individuals are thought to share 99.9 percent of their genome information in common. However, given that there are 3 billion base pairs in the genome, any two individuals may differ in their genetic sequences by millions of base pairs (Guttmacher and Collins 2002). Thus, the characterization of individual differences is important in understanding the variability seen in humans.

Given advances made in understanding the basic mechanisms of heredity, it is not surprising that genetics information is being used more and more to improve health care. Thousands of conditions caused by mutations in single genes (termed *Mendelian diseases*) have been identified over the last century and are catalogued in an online compendium titled "Online Mendelian Inheritance in Man" (OMIM; http://www.ncbi.nlm.nih.gov/sites/entrez?db=omim). For example, mutations responsible for cystic fibrosis (Buchwald 1996), Progeria syndrome (extremely premature aging; Eriksson et al. 2003), and Huntington's disease (van Dellen and Hannan 2004) have been identified. However, Mendelian (single-gene) diseases are rare, affecting at most one in a few hundred individuals in the United States (Guttmacher and Collins 2002). Thus, although the effect of monogenic (single-gene) conditions on the individual patient is substantial, fewer individuals in the population will likely directly benefit from the identification of the causal mutation.

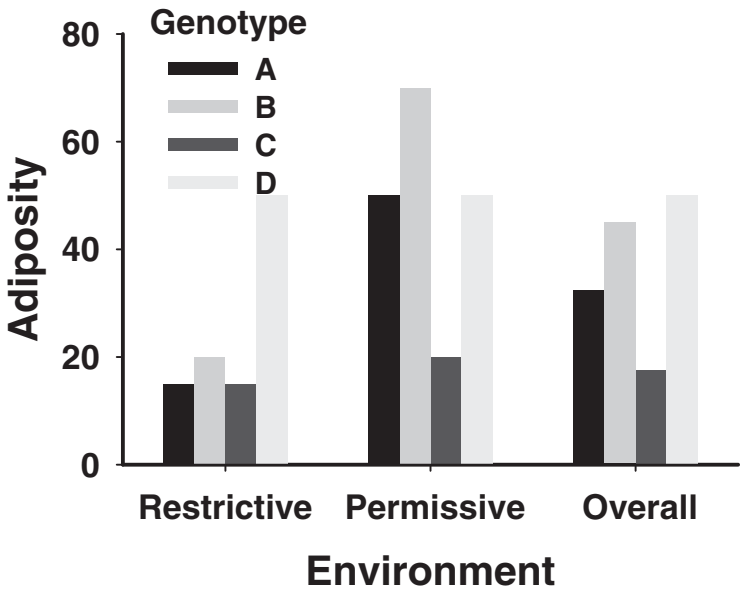
In contrast, common complex diseases, such as obesity, type 2 diabetes, breast cancer, and depression, directly affect many more individuals.

Common complex diseases are caused by mutations in multiple genes, environmental factors, and/or interaction of these genetic and environmental factors. Several high-penetrance (a high frequency of individuals who have a genetic variant will display the characteristics associated with the variant) genes increasing the risk for common complex disease have been identified, including for breast and ovarian cancer (Nathanson et al. 2001), for hereditary nonpolyposis colorectal cancer (Lynch and Smyrk 1999), and for diabetes (Froguel and Velho 1999). Moreover, highly prevalent genes, but with reduced penetrance, have been associated with common complex diseases, including for long QT syndrome (Priori et al. 2003), Alzheimer's disease (Ashford 2004), and myocardial infarction (Yamada et al. 2002). Genomewide association studies (GWAS) are now widely recognized as a powerful tool for identifying genetic variants related to common complex diseases and have demonstrated success in detecting relevant results, particularly related to common complex and polygenic diseases. GWAS were used to identify a link between the fat mass and obesity associate gene (*FTO*) and obesity (Dina et al. 2007; Frayling et al. 2007; Scott et al. 2007; Scuteri et al. 2007), the complement factor H (*CFH*) polymorphism in age-related macular degeneration (Klein et al. 2005; Z. Yang et al. 2006), interleukin 23 receptor (*IL23R*) and inflammatory bowel disease (Duerr et al. 2006), and WW and C2 domain containing 1 (*WWC1*) and human memory performance (Papassotiropoulos et al. 2006). Similarly, a genomewide nonsynonymous single-nucleotide polymorphism (SNP) scan revealed a susceptibility locus in the interferon induced with helicase C domain 1 (*IFIH1*) region for type 1 diabetes (Smyth et al. 2006).

Gene–environment interaction encompasses two different categories of risk, both the ways in which individuals are differentially vulnerable to environmental agents and the ways in which individuals have been affected by exposure to such agents (including the acquisition of genetic susceptibilities). Gene–environment interaction may be measured by the different effects of an exposure on disease risk among individuals with different genotypes or by the different effects of a genotype on disease risk among individuals with different exposures (Shostak 2003; Q. Yang and Khoury 1997). Thus, a trait is a result of not only the sum of environmental and genetic influences, but also the interactions between environment and genes.

Figure 3 graphically demonstrates a possible gene–environment interaction for a highly prevalent common complex disease, obesity. In this model, we present a hypothetical gene-influencing metabolism with four possible

Figure 3
Genotype by Environment Interaction



genotypes: Genotype A, B, C, and D. In a low-calorie environment (restrictive), individuals with Genotypes A, B, and C all have low levels of adiposity and D has a medium level of adiposity. But in a high-calorie environment (permissive), individuals with Genotypes A and D have medium levels of adiposity, B has a high level of adiposity, and C has a low level of adiposity. However, the differential effects of Genotypes A and B are observed only in the context of different environments.

Researchers have been keenly aware of the importance of gene–environment interactions (Andrieu and Goldstein 1998; Brennan, Mednick, and Jacobsen 1996; Cadoret et al. 1985; Cloninger, Bohman, and Sigvardsson 1981; A. C. Heath, Jardine, and Martin 1989; Khoury, Beaty, and Cohen 1993; Khoury and James 1993; Ottman 1990, 1995; Udry, Kovenock, and Morris 1996; Q. Yang and Khoury 1997) in the etiology of common complex phenotypes (the outward physical manifestation of a genotype), and such a phenomenon may help explain the difficulty in identifying common susceptibility genes. Furthermore, the focus of

identifying these gene–environment interactions is part of the long-range goal of the human genome project (Collins, Morgan, and Patrinos 2003). Behaviors or traits that change over a historical period seem to defy the explanation of genetic influences because genes do not change over a short period of time. Although genes or genetic distributions are unlikely to change dramatically, the social and other environmental conditions that regulate the gene expression do. When these conditions change markedly over time, the amount of genetic influences realized could also change markedly over the same period. For this reason, the historical changes in obesity are likely to have been influenced by gene–environment interactions.

In this article, we seek to provide a framework for successfully conducting gene–environment studies for investigators and students interested in social factors but who wish to incorporate genetic factors into their study. To accomplish this goal, we first describe the basic study designs and procedures of implementation that are generally applied in gene–environment interaction. This discussion is limited in its scope and coverage and is meant solely to introduce the interested reader to gene–environment study designs; it is not meant to be comprehensive. Next, we briefly review several examples of gene–environment interaction in the field of obesity epidemiology from the published literature, as a learning exercise. Last, we review the common reasons why epidemiological studies that incorporate gene–environment interaction have been unable to demonstrate statistically significant interactions and why conflicting results are reported in the literature.

Tools for Investigating Gene–Environment Interaction

Definition of Exposure

As with all epidemiological research, an accurate assessment of the exposure is of paramount importance. For the most part, a consideration of the timing, amount, duration, and intensity of exposure is necessary to avoid potential misclassification, which can bias the estimate of the main effect as well as the joint effect of the interaction (N. Rothman et al. 1999; N. Rothman and Hayes 1995). However, for most genetic studies, the characterization of exposure has been minimal. For example, most genetic epidemiology studies collect dichotomous variables such as the presence or absence of environmental and social factors like smoking, physical activity, alcohol consumption, and so forth. But the dichotomization of these clearly quantitative variables provides a substantial loss of power.

Selection of Genes

Knowledge about the response of organisms to changes in their environment may be gathered at the DNA level using DNA arrays, classical sequencing, and analysis of polymorphisms using linkage and association analysis (Daniel 2002), and at the messenger RNA level and/or the protein level by expression arrays and proteome analysis (Trayhurn 2000). This review focuses on the former and thus assumes that the investigator will be collecting DNA. The types of DNA-level information that can be collected include microsatellites, SNPs, and sequence data. Microsatellites or short tandem repeats are highly variable segments of DNA for which multiple forms exist. These DNA markers are found at a rate of 1 in 10,000 base pairs and are often used for linkage studies as the average heterozygosity per locus in 80 percent (Shastry 2002). SNPs are single base pair changes at a specific region (Schork, Fallin, and Lanchbury 2000). SNPs are becoming the polymorphism of choice because of the ease of high-throughput genotyping and the large number of SNPs available, occurring once every 1,000 base pairs (Brookes 1999). Although multiple SNP mapping techniques are available (Shastry 2001), DNA sequencing is still the method of choice for SNP detection.

The choice of markers often depends on the type of study. When conducting a genome scan, the genotyping lab will likely determine the selection of genetic markers. However, the investigator will decide whether to use a 20-centimorgan (cM), 10-cM, 5-cM, or 1-cM map. A frequently used laboratory for high-throughput genomic scans is the National Heart, Lung, and Blood Institute's Mammalian Genotyping Service (Marshfield, WI, http://research.marshfieldclinic.org/genetics/home/index.asp). Association studies typically utilize SNPs, but these types of studies can be candidate gene–wide or genomewide. Although many GWAS studies are now being conducted, the majority of population-based research still likely uses a candidate gene approach. The selection of candidate genes should be made based on a biologically plausible hypothesis, such that the gene product is involved in the phenotype of interest. Information on the function of candidate genes and the identification of important polymorphisms and validated SNPs within these genes can be obtained from many Web sites (see Table 1).

Selection of Polymorphisms

Selecting polymorphisms with known functional effects is advantageous, and SNPs may be considered following criteria for putative functionality

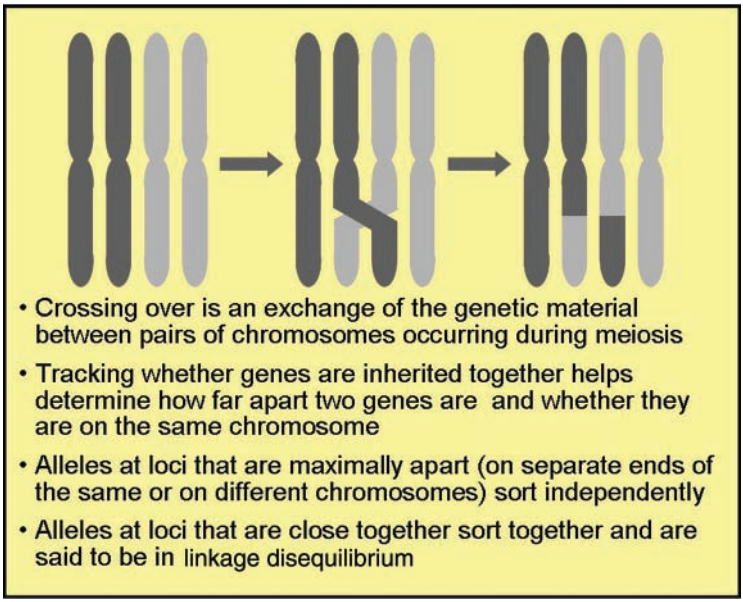
Table 1
Genetic Information Web Sites

Web Site	URL
Online Mendelian Inheritance in Man (OMIM)	http://www.ncbi.nlm.nih.gov/sites/entrez?db = omim
Human Genome Epidemiology (HuGE) Net	http://www.cdc.gov/genomics/hugenet
National Center for Biotechnology Information (NCBI) dbSNPs	http://www.ncbi.nlm.nih.gov/SNP
SNPs Consortium	http://snp.cshl.org
GeneCard	http://bioinfo.weizmann.ac.il/cards/index.html
GeneSNPs	http://www.genome.utah.edu
GENATLAS	http://www.dsi.univ-paris5.fr/genatlas
The Wellcome Trust Sanger Institute	http://www.sanger.ac.uk
Seattle SNPs	http://www.pga.gs.washington.edu
PubMed	http://www.ncbi.nlm.nih.gov/sites/entrez
University of California, Santa Cruz, Genome Bioinformatics	http://genome.ucsc.edu
Celera Genomic Database	http://www.celera.com
RealSNP, Sequenom, Inc.	https://www.realsnp.com/default.asp

(descending order of priority): amino acid change, promoter polymorphism, association with potential splice sites, synonymous coding region variation, and intronic (noncoding region) variants. To account for variation contributed by SNPs contained within a given gene and to sample all regions of each gene being investigated, researchers may evaluate multiple SNPs per gene, depending on the size and the number of variants residing within each gene (Neale and Sham 2004). Because some genes encompass very large spans of DNA sequence, this strategy will ensure that all regions of each gene are investigated.

To avoid redundancies when selecting multiple SNPs in a given gene, linkage disequilibrium (LD) within a given gene can be identified. LD is simply nonrandom segregation of alleles (one of the variant forms of a gene at a particular locus; see Figure 4). The cosegregation of alleles is best captured using haplotypes, which describe an allelic configuration of multiple markers present on a single chromosome of a given individual (van den Oord and Neale 2004). The HapMap Web site (http://www.hapmap.org) supports bulk data downloads so that LD blocks can be estimated, along with identification of tag SNPs, which define or "tag" haplotype blocks within the genes, thus capturing most of the variation across the gene. Moreover,

Figure 4
Linkage Disequilibrium



several groups have identified methods for identifying tag SNPs (e.g., Byng et al. 2003; Lowe et al. 2004; K. Zhang et al. 2004). SNPs of potential functional importance, based on positioning within the gene (e.g., nonsynonymous variation) or previously reported genetic studies may also be selected to supplement the tag SNPs. Ideally, researchers should type SNPs at a high density in a subset of the population to create a haplotype map for the study populations of interest and use these tag SNPs for all subsequent typing (van den Oord and Neale 2004). Thus, the gold standard approach is a gene-based approach that considers all variation within a gene and its regulatory region, yet because of LD, nearly full information (with the exception of rare variants) is often obtained by genotyping only a subset of the common SNPs (Neale and Sham 2004).

Haplotype Estimation

Because multiple SNPs per gene may be examined, information about the intervening SNPs not typed can be inferred, and because of an increased

power to detect association between a phenotype and genetic variation (e.g., E. R. Martin et al. 2000), many investigators are expanding their study design to include haplotypes. Gabriel et al. (2002) described LD and haplotype patterns across 51 autosomal regions, spanning a total of 13 million base pairs of the human genome, and showed that haplotype frameworks provide substantial statistical power in disease association studies in these regions. Similarly, other studies have investigated additional regions of the human genome and described the LD and haplotype structures. Reich et al. (2001) determined LD between SNPs in 160 kilobase (kb) regions around 19 different genes and reported significant LD ranging from 6 to 155 kb, with the average about 60 kb in non-Hispanic Whites. Nonetheless, it is important to note that patterns of LD vary largely by region of the genome and by population (Erichsen and Chanock 2004).

Individual haplotypes can be determined using family- and population-based studies. In family studies with three generations or more of genetic information, haplotypes can be determined directly by following how alleles segregate. In contrast, haplotype estimation in unrelated individuals requires the use of specific algorithms. The most commonly implemented approach to the estimation of haplotypes in unphased data is the expectation maximization (EM) algorithm (Excoffier and Slatkin 1995). The EM algorithm estimates population haplotype probabilities based on maximum likelihood, detecting haplotype probabilities that optimize the probability of the observed data, assuming Hardy–Weinberg equilibrium. This method is computationally intensive, cannot handle a large number of loci, and usually requires various simplification strategies (Qin, Niu, and Liu 2002). Variations of this method are implemented in several statistical packages including PL-EM (Qin et al. 2002), SNPHAP (http://www-gene.cimr.cam.ac.uk/clayton/software), and EM-DeCODER (Niu et al. 2002).

A second method of haplotype estimation is based on maximum parsimony but is limited in scope because haplotype assignments are made only when unambiguous (Clark 1990). Clark's method is advantageous, however, as it can handle a large number of loci when haplotype diversity is limited in the population and it is relatively straightforward (Niu 2004).

A more recent method of haplotype estimation uses a semi-Bayesian or Bayesian method that incorporates prior expectations based on observed population genetics principles (e.g., Niu et al. 2002; Stephens, Smith, and Donnelly [SSD] 2001). For example, Stephens and colleagues' (2001) method utilizes Markov chain Monte techniques in which unknown haplotypes are derived from a conditional distribution that assumes genealogical relationships between individuals based on the neutral coalescent

theory. Lin et al. (2002) recently compared haplotypes estimated using the semi-Bayes algorithm to empirically derived haplotypes and found the SSD method to be more than 95.2 percent accurate for polymorphisms with a minor allele frequency >0.2 over 100 kb of DNA sequence. The PHASE software, available from the Oxford Mathematical Genetics Group with suggested modifications from Lin et al. (2002), provides both the estimated haplotypes as well as the probabilities associated with their accuracy. Several studies have assessed the performance of these various haplotype estimation algorithms and have found that EM and semi-Bayes methods exhibited similar performances in both simulated data (Xu et al. 2002; S. Zhang et al. 2001) and empirical data from the human growth hormone locus (Adkins 2004).

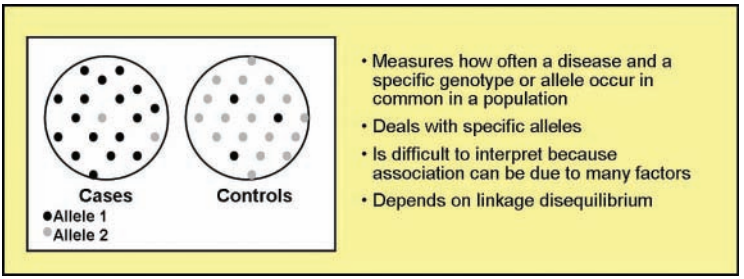
Study Design and Selection of Analytical Approaches

The selection of study design and analytical approach is one of the most important choices an investigator will make when embarking on a gene–environment study. The merits and pitfalls of such study designs have been elegantly described by several authors (Borecki and Suarez 2001; Risch 2000) and are presented here only briefly.

Association Studies: Statistical Models

Association studies measure the statistical dependence or correlation between two or more events, characteristics, or other variables (Tabor, Risch, and Myers 2002; see Figure 5). Association analysis can be conducted using single SNPs, multiple SNPs, and haplotypes and also using a variety of samples, including case-control, case-only, and family data. Gene–environment interactions can be measured on the additive or multiplicative scale. Additive interaction is assessed on the numeric scale and is indicated when the joint effect (i.e., incidence rates) of two exposures (gene and environment) differs (is higher or lower) from the sum of the independent effects of the two exposures alone. In contrast, multiplicative interaction is assessed on the multiplicative scale and measures whether the relative risk for the joint exposure of genotype and environment is statistically different (greater or smaller) from what would be expected by multiplying the relative risks for environmental or genetic exposure alone (Q. Yang and Khoury 1997). The ability to quantify the gene–environment interaction lends itself well to applications in retrospective and prospective epidemiological studies (Piegorsch, Weinberg, and Taylor 1994; Prentice 1995), and standard statistical methods are

Figure 5
Association



available for analyzing interaction in the context of a case-control study (Q. Yang and Khoury 1997) and a case-cohort study (Breslow and Day 1987).

Single SNP analysis. Briefly, in the context of a case-cohort study design, the investigator may choose to apply the Cox proportional hazards (PH) model to analyze the effect of genetic variation within the context of environment exposure data on time to event. The Cox PH model allows for multifactorial designs and the inclusion of continuous (and categorical) covariates and their joint effects, with censored survival time as the outcome variable. In contrast, in the context of a case-control study design, logistic regression is often applied, with multiplicative interactions determined by a Wald χ^2 test for significance of the estimated β coefficient, $\hat{\beta}$, for the interaction term, and an additive interaction assessed by testing the interaction contrast ratio (Li et al. 2000; K. J. Rothman 2002).

Multiple SNP analysis. Studies have shown that incorporating multiple SNPs may be more powerful than incorporating single SNPs (Akey, Jin, and Xiong 2001; Morris and Kaplan 2002) because several variants may act together on the trait of interest. Thus, a variety of multi-SNP analyses that utilize multilocus genotype approaches, as well as haplotype characterization of genetic variation in regions of the genome, has been developed to test for association.

When multiple SNPs in a given gene are typed, the hierarchical linkage disequilibrium (HLD) method can be used to incorporate information from multiple SNPs without having to estimate haplotypes (Conti and Witte 2003). This approach offers an attractive alternative, as statistical estimation of individual haplotypes can be somewhat problematic in

practice, especially when parental information is missing (which is true in the case-control design). The HLD takes into account correlations among markers due to the haplotype block structures as well as the physical distances among markers. The HLD model uses a two-stage hierarchical estimation procedure. In the first stage, the independent effect of each SNP on the phenotype is characterized using a standard statistical model, producing a set of estimates of magnitude of effect and their standard errors. When the phenotype is qualitative (e.g., case-control), a logistic regression model is used, producing a set of odds ratios. When the phenotype is quantitative, a regression/analysis of variance model is used, producing a set of regression coefficients. For the second stage, the first-stage effect estimates are used as input data in a mixed-model procedure. Incorporating a single fixed effect for the block structure of the gene (all SNPs in the same block have the same value of the effect) and a random Gaussian error/variance component (VC), which is correlated across SNPs, the intermarker correlation structure due to physical distance, based on a general exponential decay function, is modeled. Following Conti and Witte, a semi-Bayes estimation procedure is used to derive the second-stage estimates from those in the first stage. Essentially, the mixed model leverages the block information because individual SNP effect estimates from the same block are assumed to be indexing the same block genotype–phenotype effect.

Haplotype analysis. Most haplotype analysis programs assume that LD decays with increasing distance. However, extensive variability in LD has been observed in many other published studies (e.g., Abecasis et al. 2001; Dawson et al. 2002). Nonetheless, recent studies have showed that when haplotype block structure can be reliably constructed in a region, representative common haplotypes can capture most of the variation (Gabriel et al. 2002). A moving window haplotype association approach that sequentially examines adjacent SNPs across the relevant genomic region has recently been developed (Fallin et al. 2001). In this approach, a window is defined (e.g., $2N\ldots$ consecutive SNPs), and haplotype frequency differences between cases and controls are computed iteratively using a permutation algorithm to test for significance. An "omnibus test" is used to determine if multiple haplotypes are associated with the phenotype of interest (Akey et al. 2001; Longmate 2001). As this method assesses overall haplotype frequency profile differences, it can lead to detection of very subtle differences between haplotypes that are manifested in the aggregate rather than individually, and it is also able to detect associations using SNPs surrounding the functional allele even if the functional allele was not typed.

Haplotypes are also analyzed using haplotype similarity programs that compare similarity of haplotypes among cases with similarity of haplotypes among the controls (e.g., Bourgain, Genin, and Clerget-Darpoux 2002; Bourgain, Genin, Holopainen, et al. 2001; Bourgain, Genin, Margaritte-Jeannin, et al. 2001; Bourgain et al. 2000; Tzeng et al., 2003). The underlying assumption is that mutation-carrying haplotypes are inherited from a common ancestor and are therefore identical by descent (IBD) not only at the locus mutation but also within a narrow region flanking the mutation (the haplotype). If a relatively large proportion of the cases are descendants of one (or a few) common founders, then the case sample is expected to demonstrate excessive allele sharing (haplotype similarity) over the control sample in this narrow vicinity around the mutation locus. This approach is limited by founder heterogeneity, which is likely important in large admixed populations.

Case-only analysis. The above methods assume a sample of cases and controls from a population. However, more recently, the case-only design has been popularized as a simple tool to screen for gene–environment interaction (Khoury and Flanders 1996). The interaction can be calculated in the context of a 2×2 table or logistic model and can result in greater precision in estimating interactions in comparison to case-control analysis. There are a few assumptions of the case-only study design, including that the genotype and exposure are independent in controls; there is a population-based series of incident cases; no estimate of main effect of genotype or exposure can be determined; departure from multiplicative interaction only is considered; the genotypic effect must have low penetrance, with an odds ratio attributable to genetics <6; and there is no misclassification of exposure or genotype, no selection bias, and no confounding. Certainly, the case-only approach is attractive, as selecting control participants represents a substantial challenge to many studies. Nonetheless, the main effect of genotype and environment cannot be quantified using this method and is thus a major caveat.

The case-only study design has also been criticized because many of the traditional problems with genetic markers (misclassification, admixture, LD) still apply, the fundamental assumption of independence of genotype and exposure cannot be assessed without a control group, and independence of genotypes is assumed, which needs to be empirically verified (Albert et al. 2001).

It may be that the case-only design is most useful as a preliminary screen (Lawrence and Greenwald, 1977), where limited information is

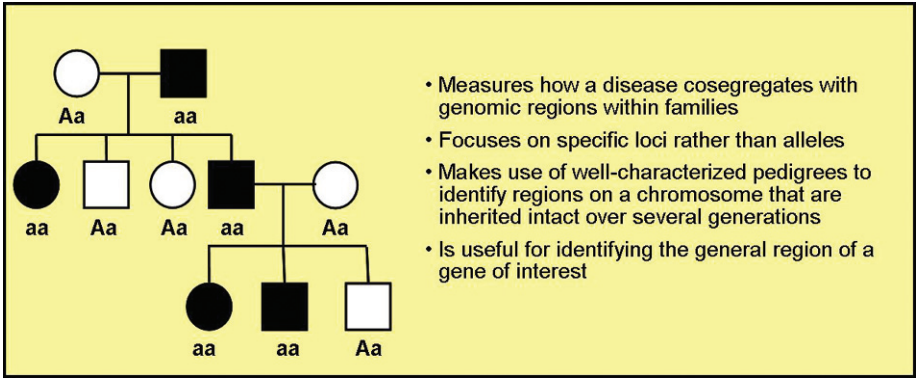
gathered on cases only to test for potential gene–environment interactions, from which a larger study that includes control-series data is designed. Since the sample size requirements for testing interaction in a case-only design can be much lower than those for a full case-control study, this strategy can lead to efficient management of limited resources yet still provide investigators with good power to detect gene–environment interactions.

Linkage Studies

Family-based studies can also be used to investigate gene–environment interaction in disease etiology. Genes that are likely to interact with environmental factors can be identified through genomewide searches, whereby families are genotyped for DNA markers with known locations along the chromosomes, and a disease or trait gene is localized to a specific region through the use of linkage analysis (see Figure 6). These gene mapping strategies are based on the assumption that both the genetic marker and the DNA sequence immediately surrounding it are transmitted intact from one generation to the next (Bray 2000; see Figure 4). Thus, we assume that the genotype at the marker locus represents the genotype at the surrounding sequence, which may harbor the functional variant associated with disease etiology. Significant evidence for linkage to a phenotypic trait is determined either by consistent transmission of a phenotype or disease from parent to offspring, proportion of alleles shared between affected and unaffected siblings, or the relationship between allele sharing and means or differences in phenotypic values for pairs of relatives.

VC models. Robust VC models for linkage analysis (sometimes called model-free or nonparametric) do not make strong assumptions about the latent trait genes themselves and rely on the relationship between the degree of phenotypic similarity and the level of IBD sharing between relatives to infer linkage. One of the strengths of the VC methodology is its ability to be extended to include genotype–environment interaction at a specific locus (Eisen and Legates 1966; Robertson 1959; Towne, Blangero, and Siervogel 1993). In the VC method, one tests for evidence of genotype by partitioning the genetic variance into genetic variance by environmental exposure (usually discrete—such as smoking and nonsmoking) and estimating the correlation between the two genetic effects. If either the genetic variances differ or the correlation between the genetic effects does not equal 1, then there is evidence for a genotype–environment interaction. The genotype by environment approach has been recently implemented in the program package SOLAR, using the methods detailed by Almasy and Blangero (1998).

Figure 6
Linkage



Structural equation modeling. More recently, Province and colleagues (2003) have developed a model to assess the gene–environment interaction using structural equation modeling as implemented in SEGPATH. This methodology is an extension of typical multivariable regression modeling, but instead of considering each outcome variable independently, it simultaneously accounts for correlations among several outcomes and the relative impact that several independent variables have on these outcomes. It will be particularly useful to examine the relationship between genetic variation (polymorphisms) and related metabolic traits simultaneously because the model is defined by genetic inheritance. Indeed, although more simulation work is warranted, Province et al. (2003) provided a simple example that demonstrated the importance of adequately modeling "nuisance effects" in linkage analysis. Indeed, consideration of the gene–environment interaction in quantitative and molecular studies will potentially direct and enhance gene-mapping efforts and lead to better understanding of the complexity of joint genetic and environmental factors in the pathogenesis of disease (Purcell 2002; Purcell and Sham 2002).

Combined Linkage and Association Analysis

Case–parental control studies began to avoid the problem of population stratification (Spielman and Ewens 1996) and compare the genotype of the affected offspring with the genotype of a fictitious control subject carrying the nontransmitted alleles from each parent. The case-parental approach is considered an association and linkage approach because an inference is made about a specific allele in the population of families and also about a location: If there were no linkage between the marker and disease, alleles at the marker locus would segregate independently of disease in the children. This approach has been generalized to allow for multiallelic markers, inclusion of full sibships (both affected and unaffected siblings), missing parental genotypes, and quantitative phenotypes. Case-control parental studies have also been extended to incorporate quantitative variation, as described by Horvath et al. (2001) and implemented in the computer program FBAT. This program has been further extended to include analysis of haplotypes, in HBAT (Horvath et al. 2004). However, because of the limited number of family members examined, the case-parental study design has limited power (Spence et al. 2003).

To test for gene–environment interaction in the context of a case-parental design (using a qualitative trait), investigators stratify case subjects according to their environmental exposure status (presence or absence) and can use the difference of odds ratios derived with and without environmental

exposure as an indication of departure from multiplicative interaction (Q. Yang and Khoury 1997). Several methods combine the advantages of linkage and population association analysis and also take into account the possibility of confounding, including the score test, haplotype relative risk, genotype relative risk, and transmission disequilibrium test (Schaid and Sommer 1993; Spielman and Ewens 1996; Thomson 1995). The score test, as proposed by Clayton (Clayton 1999; Clayton and Jones 1999), and implemented in the computer program TRANSMIT, accommodates full nuclear families with arbitrary missing values patterns and can also test haplotypes. Empirical p values can be calculated by simulation within the program to assess significance.

Sample Size, Statistical Power, and Multiple Comparisons

The most important design consideration when planning a study is the determination of the minimal number of subjects (sample size) required to ensure that a study has sufficient power to detect an interaction, if it exists. Indeed, gene–environment interactions represent effects of a higher order than those associated with either of the separate main effects. As such, greater numbers of subjects are required for assessing interactions than those typical for studying a single disease factor (Greenland 1983). The exact numbers required for each study are difficult to determine; two primary factors that affect the power of a study are the prevalence of the polymorphism and the magnitude of effect modification (N. Rothman et al. 2001). In general, common genetic variants are less likely to exhibit a strong effect, yet there is more statistical power in examining these polymorphisms (Risch 2000). Several different power analysis methods for population-based research have been described in the literature (Foppa and Spiegelman 1997; Garcia-Closas and Lubin 1999; Gauderman 2002; Hwang et al. 1994; Lubin and Gail 1990), and many are publicly available for download from the Internet including the Foppa and Spiegelman software, http://www.hsph.harvard.edu/faculty/spiegelman/foppa.html; Garcia-Closas and Lubin software, http://www.dceg.cancer.gov/POWER/readme.html; and Spiegelman and Logan software, http://www.hsph.harvard.edu/faculty/spiegelman/ge_trend_v2.html.

For family-based studies, simulation experiments are necessary to estimate power. However, many of the major statistical packages, for example, SOLAR and Genehunter, have that capability built in.

The issue of multiple comparisons in genetic studies is one of the largest problems facing genetic epidemiologists today. The traditional family-wise

Type I error rates, often implemented as a Bonferroni correction, assess the probability of making at least one false-positive inference but are too conservative as they do not account for the correlation among multiple SNPs and may lead to an increase in type 2 errors. An alternative to the family-wise Type I error rates correction focuses on controlling the false discovery rate (FDR), defined as the percentage of statistical tests deemed significant that are false positives; (Benjamini et al. 2001). By controlling the FDR, we can assume that on average, only α (typically 5 percent) of the total positive discoveries are false. This is a more realistic goal that preserves greater power to detect true positives than the more traditional Bonferroni-type procedures. FDR approaches are implemented by ranking the statistical tests by significance level and requiring more stringent α levels for significance serially, adjusting each time for the increasing number of tests. Many such tests have recently been published (e.g., Efron and Tibshirani 2002; Keselman, Cribbie, and Holland 2002; Sabatti, Service, and Freimer 2003; Storey and Tibshirani 2003).

Another approach to the multiple comparisons problem is to use permutation testing (Nichols and Holmes 2002). In this approach, the correlation structure between the tests is incorporated in the multiple adjustment procedure, and significance is assessed by counting the number of ways the data can be permuted that produce results more extreme than observed. Because this method accounts for the exact correlation between tests and does not overcorrect by assuming that all tests are independent, it is more powerful than family-wise Type I error and FDR methods of correction.

An alternative approach to the multiple comparisons problem is provided by the sequential multiple-decision procedure (SMDP; Province 2000, 2001), which relies on sequential sampling theory to keep power high and Type I error low. Using this method, two phases of analysis are implemented. In the first phase, the hypothesis generation phase, all SNPs/haplotypes are tested simultaneously in a single multiple-decision test, but the subjects are sequentially sampled, each time looking to separate the signal subset of SNPs/haplotypes from the remaining noise subset. Such sequential sampling continues until the correct signal subset is selected with 95 percent confidence, all the while preserving high power and a low overall Type I error rate. Afterward, the remaining sample is not used in the SMDP hypothesis generation phase but is used in the second-stage hypothesis-testing phase. In this phase, no correction for multiple testing is needed, as the number of hypotheses considered is small and well motivated.

Informed Consent

Informed consent for genetic testing has become a major concern, and much has been written about ethical, legal, and social issues of genetic testing (Anderlik and Rothstein 2001; ASHG 1996) and informed consent for genetic research (ASHG 1996). However, in response to the need for formal guidelines specific to population-based studies of low-penetrant genes, the Centers for Disease Control and Prevention (CDC) formed a multidisciplinary group to develop an informed consent approach for integrating genetic variation into population-based research (Beskow et al. 2001). This group used expert opinion and federal regulations, the National Bioethics Advisory Commission's report research using human biologic materials, existing consent forms, and literature on informed consent to create an informed consent document and supplemental brochure, which is publicly available at the CDC Web site: http://www.cdc.gov/genomics/info/reports/policy/consent.htm.

Handling and Testing of Genetic Data

DNA Collection

There are several options currently available for the collection of DNA samples. In addition to the standard venous blood samples, buccal cell collection brushes (Walker et al. 1999), mouth washes (Garcia-Closas et al. 2001; E. M. Heath et al. 2001), and blood spots (Caggana, Conroy, and Pass 1998; Poyser, Wyatt, and Chambers 1998) have been used and can offer an appealing alternative for studies aimed at populations less likely to provide a blood sample (i.e., children). Although the quantity of the genomic DNA yielded from the buccal cell collection brush and mouthwash methods can be considerably less than that obtained from blood sampling, it is usually sufficient for all but the most demanding genotyping projects. Buccal cell methods routinely obtained 30–40 μg of DNA per individual, which is generally sufficient for 4,000 individual PCRs (Anchordoquy et al. 2003). In comparison, blood extractions typically yield in excess of 100 μg of DNA.

Genotyping Method

There are several methods used to genotype individuals, depending on both the type of polymorphism under examination (e.g., microsatellite markers or SNPs) and the type of sample obtained. DNA sequencing is the best approach, however, at this time; at approximately $83 per kb, such

an approach is often too expensive and time consuming. Restriction fragment length polymorphisms are often used if the polymorphism of interest is known to result in the addition or deletion of a restriction site. High-throughput approaches include 5′–nuclease-based fluorescence assays (Taqman), matrix-assisted laser desorption/ionization time-of-flight mass spectrometry analysis, SNP bead (Illumina, http://www.illumina.com/Products/prod_snp.ilmn), and SNP Chip (Affymetrix, http://www.affymetrix.com/products/application/dna_analysis_products.affx). Choosing an appropriate method and using high standards of quality control are imperative to avoid problems of genotype misclassification, which can bias study results (Garcia-Closas and Lubin 1999; N. Rothman et al. 1999).

Gene–Environment Interaction and Obesity: Examples

Defining interactions between genetic pathways and environmental effects may assist in clarifying the discrepancies among preclinical, epidemiological, and intervention studies on the importance of specific genes influencing a phenotype. The following discussion describes several studies that have sought to identify the gene–environment interaction in the pathogenesis of obesity and will serve as examples of the progress being made in elucidating gene–environment interactions that may be important in the pathogenesis of disease.

Genetic Epidemiology of Obesity

Obesity is a common complex disease with a huge public health burden; in the United States 65 percent of the adult population is either overweight or obese (Manson and Bassuk 2003). Obesity is influenced by both genes and the environment, and most likely an interaction between the two. The understanding of genetic effects on human weight regulation has increased enormously in recent years (Chagnon et al. 2003; Comuzzie 2002; Comuzzie and Allison 1998). Studies have consistently shown that 40 to 70 percent of the variation in obesity-related measures is heritable. More than 200 human quantitative trait loci (QTL; regions of the genome) for obesity-related phenotypes have been identified. Of those, a total of 35 genomic regions have been corroborated among at least two studies (Snyder et al. 2004). Moreover, there are at least 15 candidate genes that are supported by a minimum of five independent studies. GWAS have identified two susceptibility variants for obesity, including the noncoding variants near insulin-induced gene 2

(*INSIG2*) associated with obesity in the Framingham Heart Study with replication in adults and children (Herbert et al. 2006) and the link between the fat mass and obesity associate gene (*FTO*) and obesity (Dina et al. 2007; Frayling et al. 2007; Scott et al. 2007; Scuteri et al. 2007). Environmental factors interact with genetic factors in the etiology of obesity. For example, the amount of caloric intake has a great influence on the level of adiposity, as has been demonstrated in multiple epidemiological studies (e.g., Heitmann et al. 1995) and experimental feeding studies (Bouchard et al. 1990). Thus, the questions are not whether genes or environmental factors matter for obesity, but how they matter and how genetic effects are moderated by environmental influences across the life course.

The past two decades have seen numerous studies that attempt to detect evidence of environmental moderation on the genetic contribution to obesity. Many of these studies use immigration, cohort, gender, or other designs to estimate the interactions indirectly. Studies have shown that Asian American adolescents born in the United States are more than twice as likely to be obese as are adolescents who immigrated into the United States recently (Popkin and Doak 1998). Presumably, the genetic heritage between the Asian immigrants and the U.S.-born Asians is comparable; then the observed differences between the two groups might be reasonably attributed to the different environments in their early lives or a gene–environment interaction. Likewise, the Pima Indians who now reside in Arizona have a much higher body mass index (BMI; mean = 33.4) than their counterparts living in a remote area of Mexico ($M = 24.9$). The difference in obesity between the two groups has been attributed to a gene–environment interaction as these populations presumably have similar genetic backgrounds but very different diets and levels of physical activity. Other studies have also suggested that genetic factors may be important in how body mass responds to changes in energy balance (Bouchard and Tremblay 1997; Bouchard et al. 1990; Poehlman et al. 1986; Poehlman et al. 1987).

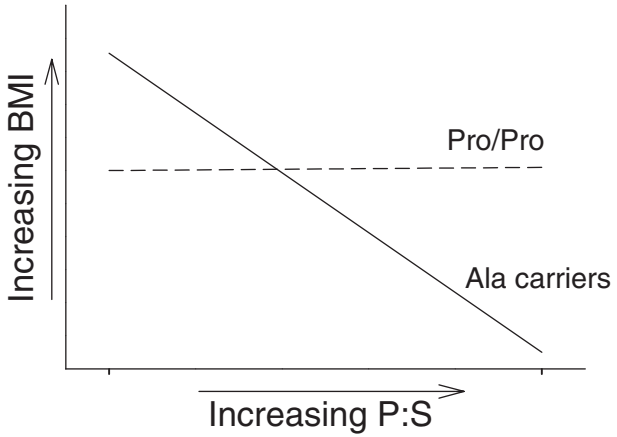
Twin studies have also indicated the importance of gene–environment interactions in the pathogenesis of obesity. Bouchard et al. (1990) conducted a feeding study of adult monozygotic (MZ) male twins to explore possible gene–environment interactions in weight gain. The participants were given the same caloric excess for 84 days and were asked to minimize all physical activity. At follow-up, the amount of weight gain varied considerably across individuals; however, there was a significant intrapair correlation among MZ twins in weight gain, suggesting a differential response of genotype to environmental change. Bouchard et al. (1990)

conducted a similar experiment focused on physical activity and weight loss to determine if similar patterns of gene–environment interaction would be detected. During the experiment, young adult male MZ twins exercised on cycles twice a day, 9 out of 10 days, over 93 days, while keeping a consistent daily energy intake. At follow-up, large within-person differences in weight loss were observed, but not between MZ twin pairs.

Association Studies

As stated previously, the selection of study design and analytical approach is one of the most important choices an investigator will make when embarking on a gene–environment study. In the obesity literature, both association and linkage approaches have been conducted. Specifically, the importance of gene–environment effects on obesity has also been investigated directly, using candidate gene studies and measured environmental factors. *PPAR*-γ is involved in adipocyte differentiation and possibly lipid metabolism and insulin sensitivity (Tontonoz, Hu, Devine, et al. 1995; Tontonoz, Hu, and Spiegelman 1994, 1995). Families with gain-of-function mutations demonstrate its significance in obesity (Hasstedt et al. 2001), while loss-of-function mutations demonstrate its significance in hypertension (Barroso et al. 1999). Luan et al. (2001) examined the role of a common variant in the *PPAR*-γ gene (Pro12Ala), the ratio of dietary polyunsaturated fat to saturated fat (P:S ratio), and their interaction, on BMI. A total of 592 nondiabetic participants in a population-based cohort study of type 2 diabetes were examined (mean age $53.9 + 10.9$ years, 259 men and 333 women). Because the Ala homozygotes were uncommon, they compared Pro homozygotes ($n = 468$) with Ala carriers ($n = 124$). Habitual diet during the previous year was assessed using a self-completion, semiquantitative food frequency questionnaire (Bingham et al. 1994). There were no significant associations between either BMI or fasting insulin and the Pro12Ala polymorphism and a significant inverse relationship between the P:S ratio and fasting insulin, after adjustment for age and sex ($p = .0119$). Because fatty acids may be natural ligands for *PPAR*-γ, they investigated the interaction between the P:S ratio and the Pro12Ala variant. The authors reported that as the P:S ratio increases, there is a tendency for both BMI and insulin to decrease in Ala carriers, but not in Pro homozygotes (interaction between P:S ratio and BMI and fasting insulin, $p = .0038$ and $p = .0097$, respectively), and when the P:S ratio is low, the mean BMI and insulin in Ala carriers are greater than in Pro

Figure 7
Interaction Between the P:S Ratio, BMI, and the Pro12Ala Variant



Note: BMI = body mass index; P:S = ratio of polyunsaturated fat to saturated fat.

homozygotes (Figure 7). Of note, when total dietary fat (expressed as a proportion of total caloric intake) was analyzed instead of the P:S ratio, no detectable interaction with genotype on BMI or fasting insulin was detected. However, the authors suggest that this finding reflects the difficulty of quantifying absolute intake with a food frequency questionnaire rather than speaking to the robustness of their findings. Overall, this study provides a positive example of a gene–environment interaction, and the biological plausibility of this association is high. Nonetheless, future studies need to replicate this association with larger sample sizes, and laboratory studies should be designed to gain an understanding of the biological basis of the observed effect.

Linkage Studies

Studies that use genomewide linkage scans have also demonstrated the importance of gene–environment interactions in the pathogenesis of obesity. Using a large sample of White and African American relative pairs, Lewis et al. (2005) reported detecting separate QTLs influencing obesity-related measures in women (chromosome 12q) and men (chromosome

15q), thus underscoring the importance of considering context-specific effects in the study of obesity.

L. J. Martin and colleagues (2002) also demonstrated the importance of considering genotype–environment interactions when identifying QTLs for obesity-related traits. Using a Mexican American population, they identified a significant genotype–sex interaction for serum leptin levels (a satiety protein whose levels are positively associated with adiposity). By including testosterone, but not other sex hormones, as a covariate they were able to eliminate the genotype–sex interaction, suggesting that the relationship between leptin and testosterone influenced genetic control. Furthermore, when including testosterone as a covariate, they identified a novel QTL on chromosome 22 (L. J. Martin et al. 2002), whereas the previous linkage unadjusted for testosterone was reported on chromosome 2 (Comuzzie et al. 1997). This novel QTL may provide insight into the sexual dimorphism seen in serum leptin levels, and perhaps dimorphism, in obesity-related risk.

Summary on Gene–Environment Studies

In summary, these few studies demonstrate how environmental and genetic factors may interact in the pathogenesis of disease. There is increasing evidence that susceptibility genes to common disease cannot be considered in a vacuum, but rather need to be investigated in the context of environmental factors such as diet, physical activity, stress, and other diseases. Previous epidemiological studies addressing gene–environment interactions in the pathogenesis of disease are lacking. Nonetheless, most of the studies that have been conducted are too small to precisely estimate the main effects of genotype and interaction with environmental and/or other genes and are often inconsistent. Indeed, the need for large epidemiological studies to adequately address gene–environment interactions has become increasingly apparent (Millikan 2002).

Why Conflicting Results for Gene–Environment Interaction Studies?

The Polygenic Nature of Most Common Complex Diseases

The majority of diseases that are considered in gene–environment interaction analyses are common complex conditions, likely resulting from oligogenic or polygenic inheritance, with each gene likely contributing a small

effect. However, most studies have attempted to explore only single gene–environment effects. Unfortunately, the sample size necessary to explore multiple interactions, while maintaining adequate power, is often not accessible. Future studies will likely need to be collaborative, with sample sizes chosen to assume suboptimal conditions such as weak effects and rare alleles in incomplete LD. However, studies with 2,000 (or more) cases and controls are seldom feasible without sharing data, and such meta-analyses may be confounded (e.g., by including low-quality studies or heterogeneity of phenotype; Gambaro, Anglani, and D'Angelo 2000).

Misclassification of Exposure and/or Genotype

The measurement of the exposure should be carefully evaluated to assess the possibility of misclassification (Colhoun, McKeigue, and Davey Smith 2003). The consequences of environmental exposure mismeasurement can lead to bias in the estimation of interaction effects and possible loss of precision and power with which interaction effects are estimated, both away from and toward the null hypothesis (Khoury et al. 1993).

Unfortunately, we do not yet have a full understanding of the genetic variation present in the genome. Therefore, the functional mutations that may be involved in gene–environment interactions are likely still undiscovered. Because of this, researchers may not be able to measure the actual functional polymorphism and choose instead several other SNPs within the gene of interest to measure association. However, because the known polymorphism is in the gene of interest, the assumption is that it will be segregated in the same pattern as the functional polymorphism (this is called LD). However, patterns of LD are highly variable across the genome, and although there are regions of substantial LD, other regions show virtually no LD (Reich et al. 2001). Therefore, misclassification of an individual's genotype for a given gene at the DNA level can occur because of LD (Khoury et al. 1993). Until a comprehensive catalog of all common variants of all genes is developed, investigators must rely on genetic markers in the region of the candidate gene or in a nonexpressed portion of the gene to conduct disease-association studies. Under these circumstances, the observed differences in prevalence of a marker allele between case and control groups would be a result of LD unless the actual functional mutation involved in the disease is targeted (Q. Yang and Khoury 1997). If a polymorphism is selected because of ease of genotyping, rather than probable biological function, then LD between the polymorphism and the functional variant is assumed (Tabor et al. 2002).

As LD varies across populations, different studies might have different findings for the same gene. Therefore, an overreliance on LD to detect association may contribute to nonreplication in findings.

Consideration of Confounders

Differences in demographic, social, and behavioral confounders can also contribute to differences in results. Unmeasured confounders associated with both the disease outcome and the genetic trait may also create spurious associations or mask true biological relationships. This is seen with the ApoE gene and Alzheimer's disease, where the association is global (Farrer et al. 1997), but strongest in Whites and Asians and weaker in Hispanics and African Americans, presumably because of background-confounding factors (either genetic or environmental). Association studies are subject to confounding by population stratification if the gene under investigation shows marked variation in allele frequency across subgroups of the population and if these subgroups also differ in their baseline risk of disease (Thomas and Witte 2002). Indeed, population substructure (or admixture) errors can clearly confound case-control study results, producing spurious associations between a genetic marker and disease. This limitation is particularly difficult when considering one's findings in the context of other published studies. However, the extent to which such stratification introduces quantifiable bias in epidemiological studies is hotly debated (e.g., Thomas and Witte 2002; Wacholder, Rothman, and Caporaso 2002). Nonetheless, exploration of existing data has provided little evidence that admixture is a major threat to epidemiological inference (Wacholder et al. 2002).

LD

LD refers to the nonrandom association between the alleles at two or more loci in a randomly mating population. However, the demonstration of a statistical association between an allele and a disease does not mean causation.

Associations arise when a locus is causally related to the disease, when the true allele of interest lies elsewhere in the genome, in LD with the measured allele, and due to chance or bias. Thus, it is possible that studies have been unable to demonstrate statistically significant gene–environment interactions because they do not exist for the particular gene and environmental variable under consideration.

Conclusions

Although any two unrelated individuals are thought to share 99.9 percent of their genetic information in common, we still know very little about the 0.1 percent of our genome that makes each individual unique. Nonetheless, in the last 50 years, rapid advances in understanding the basic mechanisms of heredity have been made, and this information is being used more and more to examine genetic diversity. Indeed, scientists have begun to characterize the variability seen in humans in an effort to explain human disease and improve preventative, diagnostic, and therapeutic health care. It can be imagined that in the near future health care practitioners will need to account for genetic variability and understand how it interacts with social and environmental factors in the etiology of disease.

In this article, we provided a framework for the successful conduct of gene–environment studies for investigators and students interested in social factors but who want to incorporate genetic factors into their studies. Because our discussion of gene–environment interaction was limited in scope and coverage, it is our hope that we introduced the interested reader to important genetic concepts, varying study designs, and useful examples from the field of obesity epidemiology.

Although many scientific studies have incorporated genetic effects over the last two decades, the number of examples demonstrating consistently positive results is small to moderate. Nonetheless, the confluence of data from in vitro studies, animal studies, and human studies supports that most diseases of public health importance are caused by the complex interaction of genetic and environmental factors. The availability of the sequence of the human genome, coupled with our ability to measure, quantify, and analyze DNA variability, provides us with an enormous opportunity for creating a wealth of information in terms of the pathogenesis of disease (Venter et al. 2001). Future studies need to have much larger population-based samples and carefully defined dietary data and should investigate the effects of polymorphisms in multiple genes instead of the effects of polymorphisms in single genes. Such research can only be accomplished through large interdisciplinary collaborative studies and the use of rather advanced bioinformatic tools (Kelada et al. 2003).

References

Abecasis, G. R., E. Noguchi, A. Heinzmann, J. A. Traherne, S. Bhattacharyya, N. I. Leaves, et al. 2001. "Extent and Distribution of Linkage Disequilibrium in Three Genomic Regions." *American Journal of Human Genetics* 68:191–97.

Adkins, R. M. 2004. "Comparison of the Accuracy of Methods of Computational Haplotype Inference Using a Large Empirical Dataset." *BMC Genetics* 5:22.

Akey, J., L. Jin, and M. Xiong. 2001. "Haplotypes Vs Single Marker Linkage Disequilibrium Tests: What Do We Gain?" *European Journal of Human Genetics* 9:291–300.

Akey, J. M., H. Wang, M. Xiong, H. Wu, W. Liu, M. D. Shriver, et al. 2001. "Interaction Between the Melanocortin-1 Receptor and P Genes Contributes to Inter-Individual Variation in Skin Pigmentation Phenotypes in a Tibetan Population." *Human Genetics* 108:516–20.

Albert, P. S., D. Ratnasinghe, J. Tangrea, and S. Wacholder. 2001. "Limitations of the Case-Only Design for Identifying Gene-Environment Interactions." *American Journal of Epidemiology* 154:687–93.

Almasy, L. and J. Blangero. 1998. "Multipoint Quantitative-Trait Linkage Analysis in General Pedigrees." *American Journal of Human Genetics* 62:1198–211.

Anchordoquy, H. C., C. McGeary, L. Liu, K. S. Krauter, and A. Smolen. 2003. "Genotyping of Three Candidate Genes After Whole-Genome Preamplification of DNA Collected From Buccal Cells." *Behavior Genetics* 33:73–78.

Anderlik, M. R. and M. A. Rothstein. 2001. "Privacy and Confidentiality of Genetic Information: What Rules for the New Science?" *Annual Review of Genomics and Human Genetics* 2:401–33.

Andrieu, N. and A. M. Goldstein. 1998. "Epidemiologic and Genetic Approaches in the Study of Gene-Environment Interaction: An Overview of Available Methods." *Epidemiologic Reviews* 20:137–47.

Ashford, J. W. 2004. "ApoE Genotype Effects on Alzheimer's Disease Onset and Epidemiology." *Journal of Molecular Neuroscience* 23:157–65.

ASHG. 1996. "ASHG Report. Statement on Informed Consent for Genetic Research. The American Society of Human Genetics." *American Journal of Human Genetics* 59:471–77.

Barroso, I., M. Gurnell, V. E. Crowley, M. Agostini, J. W. Schwabe, M. A. Soos, et al. 1999. "Dominant Negative Mutations in Human PPARgamma Associated With Severe Insulin Resistance, Diabetes Mellitus and Hypertension." *Nature* 402:880–83.

Benjamini, Y., D. Drai, G. Elmer, N. Kafkafi, and I. Golani. 2001. "Controlling the False Discovery Rate in Behavior Genetics Research." *Behavioural Brain Research* 125:279–84.

Beskow, L. M., W. Burke, J. F. Merz, P. A. Barr, S. Terry, V. B. Penchaszadeh, et al. 2001. "Informed Consent for Population-Based Research Involving Genetics." *Journal of the American Medical Association* 286:2315–21.

Bingham, S. A., C. Gill, A. Welch, K. Day, A. Cassidy, K. T. Khaw, et al. 1994. "Comparison of Dietary Assessment Methods in Nutritional Epidemiology: Weighed Records V. 24 H Recalls, Food-Frequency Questionnaires and Estimated-Diet Records." *British Journal of Nutrition* 72:619–43.

Borecki, I. B. and B. K. Suarez. 2001. "Linkage and Association: Basic Concepts." *Advances in Genetics* 42:45–66.

Botto, L. D. and M. J. Khoury. 2001. "Commentary: Facing the Challenge of Gene-Environment Interaction: The Two-by-Four Table and Beyond." *American Journal of Epidemiology* 153:1016–20.

Bouchard, C. and A. Tremblay. 1997. "Genetic Influences on the Response of Body Fat and Fat Distribution to Positive and Negative Energy Balances in Human Identical Twins." *Journal of Nutrition* 127 (5 Suppl.): 943S–47.

Bouchard, C., A. Tremblay, J. P. Despres, A. Nadeau, P. J. Lupien, G. Theriault, et al. 1990. "The Response to Long-Term Overfeeding in Identical Twins." *New England Journal of Medicine* 322:1477–82.

Bourgain, C., E. Genin, and F. Clerget-Darpoux. 2002. "Comparison of Family Based Haplotype Methods Using Intragenic SNPs in Candidate Genes." *European Journal of Human Genetics* 10:313–19.

Bourgain, C., E. Genin, P. Holopainen, K. Mustalahti, M. Maki, J. Partanen, et al. 2001. "Use of Closely Related Affected Individuals for the Genetic Study of Complex Diseases in Founder Populations." *American Journal of Human Genetics* 68:154–59.

Bourgain, C., E. Genin, P. Margaritte-Jeannin, and F. Clerget-Darpoux. 2001. "Maximum Identity Length Contrast: A Powerful Method for Susceptibility Gene Detection in Isolated Populations." *Genetic Epidemiology* 21 (Suppl. 1): S560–4.

Bourgain, C., E. Genin, H. Quesneville, and F. Clerget-Darpoux. 2000. "Search for Multifactorial Disease Susceptibility Genes in Founder Populations." *Annals of Human Genetics* 64 (Pt. 3): 255–65.

Bray, M. S. 2000. "Genomics, Genes, and Environmental Interaction: The Role of Exercise." *Journal of Applied Physiology* 88:788–92.

Brennan, P. A., S. A. Mednick, and B. Jacobsen. 1996. "Assessing the Role of Genetics in Crime Using Adoption Cohorts." *Ciba Foundation Symposium* 194:115–23; Discussion 123–28.

Breslow, N. E. and N. E. Day. 1987. "Statistical Methods in Cancer Research. Volume II—The Design and Analysis of Cohort Studies." *IARC Scientific Publications* 82:1–406.

Brookes, A. J. 1999. "The Essence of SNPs." *Gene* 234:177–86.

Buchwald, M. 1996. "Cystic Fibrosis: From the Gene to the Dream." *Clinical and Investigative Medicine* 19:304–10.

Byng, M. C., J. C. Whittaker, A. P. Cuthbert, C. G. Mathew, and C. M. Lewis. 2003. "SNP Subset Selection for Genetic Association Studies." *Annals of Human Genetics* 67 (Pt. 6): 543–56.

Cadoret, R. J., T. W. O'Gorman, E. Heywood, and E. Troughton. 1985. "Genetic and Environmental Factors in Major Depression." *Journal of Affective Disorders* 9:155–64.

Caggana, M., J. M. Conroy, and K. A. Pass. 1998. "Rapid, Efficient Method for Multiplex Amplification From Filter Paper." *Human Mutation* 11:404–9.

Chagnon, Y. C., T. Rankinen, E. E. Snyder, S. J. Weisnagel, L. Perusse, and C. Bouchard. 2003. "The Human Obesity Gene Map: The 2002 Update." *Obesity Research* 11:313–67.

Clark, A. G. 1990. "Inference of Haplotypes From PCR-Amplified Samples of Diploid Populations." *Molecular Biology and Evolution* 7:111–22.

Clayton, D. 1999. "A Generalization of the Transmission/Disequilibrium Test for Uncertain-Haplotype Transmission." *American Journal of Human Genetics* 65:1170–77.

Clayton, D. and H. Jones. 1999. "Transmission/Disequilibrium Tests for Extended Marker Haplotypes." *American Journal of Human Genetics* 65:1161–69.

Cloninger, C. R., M. Bohman, and S. Sigvardsson. 1981. "Inheritance of Alcohol Abuse: Cross-Fostering Analysis of Adopted Men." *Archives of General Psychiatry* 38:861–68.

Colhoun, H. M., P. M. McKeigue, and G. Davey Smith. 2003. "Problems of Reporting Genetic Associations With Complex Outcomes." *Lancet* 361:865–72.

Collins, F. S., M. Morgan, and A. Patrinos. 2003. "The Human Genome Project: Lessons From Large-Scale Biology." *Science* 300:286–90.

Comuzzie, A. G. 2002. "The Emerging Pattern of the Genetic Contribution to Human Obesity." *Best Practice & Research: Clinical Endocrinology & Metabolism* 16:611–21.

Comuzzie, A. G. and D. B. Allison. 1998. "The Search for Human Obesity Genes." *Science* 280:1374–77.

Comuzzie, A. G., J. E. Hixson, L. Almasy, B. D. Mitchell, M. C. Mahaney, T. D. Dyer, et al. 1997. "A Major Quantitative Trait Locus Determining Serum Leptin Levels and Fat Mass Is Located on Human Chromosome 2." *Nature Genetics* 15:273–76.

Conti, D. V. and J. S. Witte. 2003. "Hierarchical Modeling of Linkage Disequilibrium: Genetic Structure and Spatial Relations." *American Journal of Human Genetics* 72:351–63.

Daniel, H. 2002. "Genomics and Proteomics: Importance for the Future of Nutrition Research." *British Journal of Nutrition* 87 (Suppl. 2), S305–11.

Dawson, E., G. R. Abecasis, S. Bumpstead, Y. Chen, S. Hunt, D. M. Beare, et al. 2002. "A First-Generation Linkage Disequilibrium Map of Human Chromosome 22." *Nature* 418:544–48.

Dina, C., D. Meyre, S. Gallina, E. Durand, A. Korner, P. Jacobson, et al. 2007. "Variation in FTO Contributes to Childhood Obesity and Severe Adult Obesity." *Nature Genetics* 39:724–26.

Duerr, R. H., K. D. Taylor, S. R. Brant, J. D. Rioux, M. S. Silverberg, M. J. Daly, et al. 2006. "A Genome-Wide Association Study Identifies IL23R as an Inflammatory Bowel Disease Gene." *Science* 314:1461–63.

Efron, B. and R. Tibshirani. 2002. "Empirical Bayes Methods and False Discovery Rates for Microarrays." *Genetic Epidemiology* 23:70–86.

Eisen, E. J. and J. E. Legates. 1966. "Genotype-Sex Interaction and the Genetic Correlation Between the Sexes for Body Weight in *MUS MUSCULUS*." *Genetics* 54:611–23.

Erichsen, H. C. and S. J. Chanock. 2004. "SNPs in Cancer Research and Treatment." *British Journal of Cancer* 90:747–51.

Eriksson, M., W. T. Brown, L. B. Gordon, M. W. Glynn, J. Singer, L. Scott, et al. 2003. "Recurrent de Novo Point Mutations in Lamin A Cause Hutchinson-Gilford Progeria Syndrome." *Nature* 423:293–98.

Excoffier, L. and M. Slatkin. 1995. "Maximum-Likelihood Estimation of Molecular Haplotype Frequencies in a Diploid Population." *Molecular Biology and Evolution* 12:921–27.

Fallin, D., A. Cohen, L. Essioux, I. Chumakov, M. Blumenfeld, D. Cohen, et al. 2001. "Genetic Analysis of Case/Control Data Using Estimated Haplotype Frequencies: Application to ApoE Locus Variation and Alzheimer's Disease." *Genome Research* 11:143–51.

Farrer, L. A., L. A. Cupples, J. L. Haines, B. Hyman, W. A. Kukull, R. Mayeux, et al. 1997. "Effects of Age, Sex, and Ethnicity on the Association Between Apolipoprotein E Genotype and Alzheimer Disease. A Meta-Analysis. ApoE and Alzheimer Disease Meta Analysis Consortium." *Journal of the American Medical Association* 278:1349–56.

Foppa, I. and D. Spiegelman. 1997. "Power and Sample Size Calculations for Case-Control Studies of Gene-Environment Interactions With a Polytomous Exposure Variable." *American Journal of Epidemiology* 146:596–604.

Frayling, T. M., N. J. Timpson, M. N. Weedon, E. Zeggini, R. M. Freathy, C. M. Lindgren, et al. 2007. "A Common Variant in the FTO Gene Is Associated With Body Mass Index and Predisposes to Childhood and Adult Obesity." *Science* 316:889–94.

Froguel, P. and G. Velho. 1999. "Molecular Genetics of Maturity-Onset Diabetes of the Young." *Trends in Endocrinology and Metabolism* 10:142–46.

Gabriel, S. B., S. F. Schaffner, H. Nguyen, J. M. Moore, J. Roy, B. Blumenstiel, et al. 2002. "The Structure of Haplotype Blocks in the Human Genome." *Science* 296:2225–29.

Gambaro, G., F. Anglani, and A. D'Angelo. 2000. "Association Studies of Genetic Polymorphisms and Complex Disease." *Lancet* 355:308–11.

Garcia-Closas, M., K. M. Egan, J. Abruzzo, P. A. Newcomb, L. Titus-Ernstoff, T. Franklin, et al. 2001. "Collection of Genomic DNA From Adults in Epidemiological Studies by Buccal Cytobrush and Mouthwash." *Cancer Epidemiology, Biomarkers & Prevention* 10:687–96.

Garcia-Closas, M. and J. H. Lubin. 1999. "Power and Sample Size Calculations in Case-Control Studies of Gene-Environment Interactions: Comments on Different Approaches. *American Journal of Epidemiology* 149:689–92.

Gauderman, W. J. 2002. "Sample Size Requirements for Matched Case-Control Studies of Gene-Environment Interaction." *Statistics in Medicine* 21:35–50.

Greenland, S. 1983. "Tests for Interaction in Epidemiologic Studies: A Review and a Study of Power." *Statistics in Medicine* 2:243–51.

Guttmacher, A. E. and F. S. Collins. 2002. "Genomic Medicine—A Primer." *New England Journal of Medicine* 347:1512–20.

Hasstedt, S. J., Q. F. Ren, K. Teng, and S. C. Elbein. 2001. "Effect of the Peroxisome Proliferator-Activated Receptor-gamma 2 Pro(12)Ala Variant on Obesity, Glucose Homeostasis, and Blood Pressure in Members of Familial Type 2 Diabetic Kindreds." *Journal of Clinical Endocrinology and Metabolism* 86:536–41.

Heath, A. C., R. Jardine, and N. G. Martin. 1989. "Interactive Effects of Genotype and Social Environment on Alcohol Consumption in Female Twins." *Journal of Studies on Alcohol* 50:38–48.

Heath, E. M., N. W. Morken, K. A. Campbell, D. Tkach, E. A. Boyd, and D. A. Strom. 2001. "Use of Buccal Cells Collected in Mouthwash as a Source of DNA for Clinical Testing." *Archives of Pathology & Laboratory Medicine* 125:127–33.

Heitmann, B. L., L. Lissner, T. I. Sorensen, and C. Bengtsson. 1995. "Dietary Fat Intake and Weight Gain in Women Genetically Predisposed for Obesity." *American Journal of Clinical Nutrition* 61:1213–17.

Herbert, A., N. P. Gerry, M. B. McQueen, I. M. Heid, A. Pfeufer, T. Illig, et al. 2006. "A Common Genetic Variant Is Associated With Adult and Childhood Obesity." *Science* 312:279–83.

Horvath, S., X. Xu, and N. M. Laird. 2001. "The Family Based Association Test Method: Strategies for Studying General Genotype–Phenotype Associations." *European Journal of Human Genetics* 9:301–6.

Horvath, S., X. Xu, S. L. Lake, E. K. Silverman, S. T. Weiss, and N. M. Laird. 2004. "Family-Based Tests for Associating Haplotypes With General Phenotype Data: Application to Asthma Genetics." *Genetic Epidemiology* 26:61–69.

Hwang, S. J., T. H. Beaty, K. Y. Liang, J. Coresh, and M. J. Khoury. 1994. "Minimum Sample Size Estimation to Detect Gene-Environment Interaction in Case-Control Designs." *American Journal of Epidemiology* 140:1029–37.

Kelada, S. N., D. L. Eaton, S. S. Wang, N. R. Rothman, and M. J. Khoury. 2003. "The Role of Genetic Polymorphisms in Environmental Health." *Environmental Health Perspectives* 111:1055–64.

Keselman, H. J., R. Cribbie, and B. Holland. 2002. "Controlling the Rate of Type I Error Over a Large Set of Statistical Tests." *British Journal of Mathematical and Statistical Psychology* 55 (Pt. 1): 27–39.

Khoury, M. J., T. H. Beaty, and B. Cohen. 1993. *Fundamentals of Genetic Epidemiology*. New York: Oxford University Press.

Khoury, M. J. and W. D. Flanders. 1996. "Nontraditional Epidemiologic Approaches in the Analysis of Gene-Environment Interaction: Case-Control Studies With No Controls!" *American Journal of Epidemiology* 144:207–13.

Khoury, M. J. and L. M. James. 1993. "Population and Familial Relative Risks of Disease Associated With Environmental Factors in the Presence of Gene-Environment Interaction." *American Journal of Epidemiology* 137:1241–50.

Klein, R. J., C. Zeiss, E. Y. Chew, J. Y. Tsai, R. S. Sackler, C. Haynes, et al. 2005. "Complement Factor H Polymorphism in Age-Related Macular Degeneration." *Science* 308:385–89.

Lander, E. S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, et al. 2001. "Initial Sequencing and Analysis of the Human Genome." *Nature* 409:860–921.

Lawrence, C. and P. Greenwald. 1977. "Epidemiologic Screening: A Method to Add Efficiency to Epidemiologic Research." *American Journal of Epidemiology* 105:575–81.

Lewis, C. E., K. E. North, D. Arnett, I. B. Borecki, H. Coon, R. C. Ellison, et al. 2005. "Sex-Specific Findings From a Genome-Wide Linkage Analysis of Human Fatness in Non-Hispanic Whites and African Americans: The HyperGEN Study." *International Journal of Obesity* 29:639–49.

Li, R., E. Boerwinkle, A. F. Olshan, L. E. Chambless, J. S. Pankow, H. A. Tyroler, et al. 2000. "Glutathione S-Transferase Genotype as a Susceptibility Factor in Smoking-Related Coronary Heart Disease." *Atherosclerosis* 149:451–62.

Lin, S., D. J. Cutler, M. E. Zwick, and A. Chakravarti. 2002. "Haplotype Inference in Random Population Samples." *American Journal of Human Genetics* 71:1129–37.

Longmate, J. A. 2001. "Complexity and Power in Case-Control Association Studies." *American Journal of Human Genetics* 68:1229–37.

Lowe, C. E., J. D. Cooper, J. M. Chapman, B. J. Barratt, R. C. Twells, E. A. Green, et al. 2004. "Cost-Effective Analysis of Candidate Genes Using htSNPs: A Staged Approach." *Genes and Immunity* 5:301–5.

Luan, J., P. O. Browne, A. H. Harding, D. J. Halsall, S. O'Rahilly, V. K. Chatterjee, et al. 2001. "Evidence for Gene-Nutrient Interaction at the PPARgamma Locus." *Diabetes* 50:686–89.

Lubin, J. H. and M. H. Gail 1990. "On Power and Sample Size for Studying Features of the Relative Odds of Disease." *American Journal of Epidemiology* 131:552–66.

Lynch, H. T. and T. C. Smyrk. 1999. "Hereditary Colorectal Cancer." *Seminars in Oncology* 26:478–84.

Manson, J. E. and S. S. Bassuk. 2003. "Obesity in the United States: A Fresh Look at Its High Toll." *Journal of the American Medical Association* 289:229–30.

Martin, E. R., E. H. Lai, J. R. Gilbert, A. R. Rogala, A. J. Afshari, J. Riley, et al. 2000. "SNPing Away at Complex Diseases: Analysis of Single-Nucleotide Polymorphisms Around ApoE in Alzheimer Disease." *American Journal of Human Genetics* 67:383–94.

Martin, L. J., M. C. Mahaney, L. Almasy, J. E. Hixson, S. A. Cole, J. W. Maccluer, et al. 2002. "A Quantitative Trait Locus on Chromosome 22 for Serum Leptin Levels Adjusted for Serum Testosterone." *Obesity Research* 10:602–7.

Millikan, R. 2002. "The Changing Face of Epidemiology in the Genomics Era." *Epidemiology* 13:472–80.

Morris, R. W. and N. L. Kaplan. 2002. "On the Advantage of Haplotype Analysis in the Presence of Multiple Disease Susceptibility Alleles." *Genetic Epidemiology* 23:221–33.

Nathanson, K. L., R. Wooster, B. L. Weber, and K. N. Nathanson. 2001. "Breast Cancer Genetics: What We Know and What We Need." *Nature Medicine* 7:552–56.

Neale, B. M. and P. C. Sham. 2004. "The Future of Association Studies: Gene-Based Analysis and Replication." *American Journal of Human Genetics* 75:353–62.

Nichols, T. E. and A. P. Holmes. 2002. "Nonparametric Permutation Tests for Functional Neuroimaging: A Primer With Examples." *Human Brain Mapping* 15:1–25.

Niu, T. 2004. "Algorithms for Inferring Haplotypes." *Genetic Epidemiology* 27:334–37.

Niu, T., Z. S. Qin, X. Xu, and J. S. Liu. 2002. "Bayesian Haplotype Inference for Multiple Linked Single-Nucleotide Polymorphisms." *American Journal of Human Genetics* 70:157–69.

Ottman, R. 1990. "An Epidemiologic Approach to Gene-Environment Interaction." *Genetic Epidemiology* 7:177–85.

———. 1995. "Gene-Environment Interaction and Public Health." *American Journal of Human Genetics* 56:821–23.

Papassotiropoulos, A., D. A. Stephan, M. J. Huentelman, F. J. Hoerndli, D. W. Craig, J. V. Pearson, et al. 2006. "Common Kibra Alleles Are Associated With Human Memory Performance." *Science* 314:475–78.

Piegorsch, W. W., C. R. Weinberg, and J. A. Taylor. 1994. "Non-Hierarchical Logistic Models and Case-Only Designs for Assessing Susceptibility in Population-Based Case-Control Studies." *Statistics in Medicine* 13:153–62.

Poehlman, E. T., A. Tremblay, J. P. Despres, E. Fontaine, L. Perusse, G. Theriault, et al. 1986. "Genotype-Controlled Changes in Body Composition and Fat Morphology Following Overfeeding in Twins." *American Journal of Clinical Nutrition* 43:723–31.

Poehlman, E. T., A. Tremblay, M. Marcotte, L. Perusse, G. Theriault, and C. Bouchard. 1987. "Heredity and Changes in Body Composition and Adipose Tissue Metabolism After Short-Term Exercise-Training." *European Journal of Applied Physiology and Occupational Physiology* 56:398–402.

Popkin, B. M. and C. M. Doak. 1998. "The Obesity Epidemic Is a Worldwide Phenomenon." *Nutrition Reviews* 56 (4 Pt. 1): 106–14.

Poyser, K. H., H. A. Wyatt, and S. M. Chambers. 1998. "Multiplex Genotyping for Cystic Fibrosis From Filter Paper Blood Spots." *Annals of Clinical Biochemistry* 35 (Pt. 5): 611–15.

Prentice, R. L. 1995. "Design Issues in Cohort Studies." *Statistical Methods in Medical Research* 4:273–92.

Priori, S. G., P. J. Schwartz, C. Napolitano, R. Bloise, E. Ronchetti, M. Grillo, et al. 2003. "Risk Stratification in the Long-QT Syndrome." *New England Journal of Medicine* 348:1866–74.

Province, M. A. 2000. "A Single, Sequential, Genome-Wide Test to Identify Simultaneously All Promising Areas in a Linkage Scan." *Genetic Epidemiology* 19:301–22.

———. 2001. "Sequential Methods of Analysis for Genome Scans." *Advances in Genetics* 42:499–514.

Province, M. A., T. K. Rice, I. B. Borecki, C. Gu, A. Kraja, and D. C. Rao. 2003. "Multivariate and Multilocus Variance Components Method, Based on Structural Relationships to Assess Quantitative Trait Linkage via SEGPATH." *Genetic Epidemiology* 24:128–38.

Purcell, S. 2002. "Variance Components Models for Gene-Environment Interaction in Twin Analysis." *Twin Research* 5:554–71.

Purcell, S. and P. Sham. 2002. "Variance Components Models for Gene-Environment Interaction in Quantitative Trait Locus Linkage Analysis." *Twin Research* 5:572–76.

Qin, Z. S., T. Niu, and J. S. Liu. 2002. "Partition-Ligation-Expectation-Maximization Algorithm for Haplotype Inference With Single-Nucleotide Polymorphisms." *American Journal of Human Genetics* 71:1242–47.

Reich, D. E., M. Cargill, S. Bolk, J. Ireland, P. C. Sabeti, D. J. Richter, et al. 2001. "Linkage Disequilibrium in the Human Genome." *Nature* 411:199–204.

Risch, N. J. 2000. "Searching for Genetic Determinants in the New Millennium." *Nature* 405:847–56.

Robertson, A. 1959. "The Sampling Variance of the Genetic Correlation Coefficient." *Biometrics* 15:469–85.

Rothman, K. J. 2002. *Epidemiology: An Introduction*. New York: Oxford University Press.

Rothman, N., M. Garcia-Closas, W. T. Stewart, and J. Lubin. 1999. "The Impact of Misclassification in Case-Control Studies of Gene-Environment Interactions." *IARC Scientific Publications* 148:89–96.

Rothman, N. and R. B. Hayes. 1995. "Using Biomarkers of Genetic Susceptibility to Enhance the Study of Cancer Etiology." *Environmental Health Perspectives* 103 (Suppl. 8): 291–95.

Rothman, N., S. Wacholder, N. E. Caporaso, M. Garcia-Closas, K. Buetow, and J. F. Fraumeni Jr. 2001. "The Use of Common Genetic Polymorphisms to Enhance the Epidemiologic Study of Environmental Carcinogens." *Biochimica et Biophysica Acta* 1471 (2): C1–10.

Sabatti, C., S. Service, and N. Freimer. 2003. "False Discovery Rate in Linkage and Association Genome Screens for Complex Disorders." *Genetics* 164:829–33.

Schaid, D. J. and S. S. Sommer. 1993. "Genotype Relative Risks: Methods for Design and Analysis of Candidate-Gene Association Studies." *American Journal of Human Genetics* 53:1114–26.

Schork, N. J., D. Fallin, and J. S. Lanchbury. 2000. "Single Nucleotide Polymorphisms and the Future of Genetic Epidemiology." *Clinical Genetics* 58:250–64.

Scott, L. J., K. L. Mohlke, L. L. Bonnycastle, C. J. Willer, Y. Li, W. L. Duren, et al. 2007. "A Genome-Wide Association Study of Type 2 Diabetes in Finns Detects Multiple Susceptibility Variants." *Science* 316:1341–45.

Scuteri, A., S. Sanna, W. M. Chen, M. Uda, G. Albai, J. Strait, et al. 2007. "Genome-Wide Association Scan Shows Genetic Variants in the FTO Gene Are Associated With Obesity-Related Traits." *PLoS Genetics* 3 (7): E115.

Shastry, B. S. 2001. "Molecular and Cell Biological Aspects of Alzheimer Disease." *Journal of Human Genetics* 46:609–18.

———. 2002. "SNP Alleles in Human Disease and Evolution." *Journal of Human Genetics* 47:561–66.

Shostak, S. 2003. "Locating Gene-Environment Interaction: At the Intersections of Genetics and Public Health." *Social Science & Medicine* 56:2327–42.

Smyth, D. J., J. D. Cooper, R. Bailey, S. Field, O. Burren, L. J. Smink, et al. 2006. "A Genome-Wide Association Study of Nonsynonymous SNPs Identifies a Type 1 Diabetes Locus in the Interferon-Induced Helicase (*IFIH1*) Region." *Nature Genetics* 38:617–19.

Snyder, E. E., B. Walts, L. Perusse, Y. C. Chagnon, S. J. Weisnagel, T. Rankinen, et al. 2004. "The Human Obesity Gene Map: The 2003 Update." *Obesity Research* 12:369–439.

Spence, M. A., D. A. Greenberg, S. E. Hodge, and V. J. Vieland. 2003. "The Emperor's New Methods." *American Journal of Human Genetics* 72:1084–87.

Spielman, R. S. and W. J. Ewens. 1996. "The TDT and Other Family-Based Tests for Linkage Disequilibrium and Association." *American Journal of Human Genetics* 59:983–89.

Stephens, M., N. J. Smith, and P. Donnelly. 2001. "A New Statistical Method for Haplotype Reconstruction From Population Data." *American Journal of Human Genetics* 68:978–89.

Storey, J. D. and R. Tibshirani. 2003. "Statistical Significance for Genomewide Studies." *Proceedings of the National Academy of Sciences of the United States of America* 100:9440–45.

Tabor, H. K., N. J. Risch, and R. M. Myers. 2002. "Opinion: Candidate-Gene Approaches for Studying Complex Genetic Traits: Practical Considerations." *Nature Reviews: Genetics* 3:391–97.

Thomas, D. C. and J. S. Witte. 2002. "Point: Population Stratification: A Problem for Case-Control Studies of Candidate-Gene Associations?" *Cancer Epidemiology, Biomarkers & Prevention* 11:505–12.

Thomson, G. 1995. "Mapping Disease Genes: Family-Based Association Studies." *American Journal of Human Genetics* 57:487–98.

Tontonoz, P., E. Hu, J. Devine, E. G. Beale, and B. M. Spiegelman. 1995. "PPAR gamma 2 Regulates Adipose Expression of the Phosphoenolpyruvate Carboxykinase Gene." *Molecular and Cellular Biology* 15:351–57.

Tontonoz, P., E. Hu, and B. M. Spiegelman. 1994. "Stimulation of Adipogenesis in Fibroblasts by PPAR gamma 2, a Lipid-Activated Transcription Factor." *Cell* 79:1147–56.

———. 1995. "Regulation of Adipocyte Gene Expression and Differentiation by Peroxisome Proliferator Activated Receptor gamma." *Current Opinion in Genetics & Development* 5:571–76.

Towne, B., J. Blangero, and R. M. Siervogel. 1993. "Genotype by Sex Interaction in Measures of Lipids, Lipoproteins, and Apolipoproteins." *Genetic Epidemiology* 10:611–16.

Trayhurn, P. 2000. "Proteomics and Nutrition—A Science for the First Decade of the New Millennium." *British Journal of Nutrition* 83:1–2.

Tzeng, J. Y., B. Devlin, L. Wasserman, and K. Roeder. 2003. "On the Identification of Disease Mutations by the Analysis of Haplotype Similarity and Goodness of Fit." *American Journal of Human Genetics* 72:891–902.

Udry, J. R., J. Kovenock, and N. M. Morris. 1996. "Early Predictors of Nonmarital First Pregnancy and Abortion." *Family Planning Perspectives* 28:113–16.

Van Dellen, A. and A. J. Hannan. 2004. "Genetic and Environmental Factors in the Pathogenesis of Huntington's Disease." *Neurogenetics* 5:9–17.

Van Den Oord, E. J. and B. M. Neale. 2004. "Will Haplotype Maps Be Useful for Finding Genes?" *Molecular Psychiatry* 9:227–36.

Venter, J. C., M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, et al. 2001. "The Sequence of the Human Genome." *Science* 291:1304–51.

Wacholder, S., N. Rothman, and N. Caporaso. 2002. "Counterpoint: Bias From Population Stratification Is Not a Major Threat to the Validity of Conclusions From Epidemiological Studies of Common Polymorphisms and Cancer." *Cancer Epidemiology, Biomarkers & Prevention* 11:513–20.

Walker, A. H., D. Najarian, D. L. White, J. F. Jaffe, P. A. Kanetsky, and T. R. Rebbeck. 1999. "Collection of Genomic DNA by Buccal Swabs for Polymerase Chain Reaction-Based Biomarker Assays." *Environmental Health Perspectives* 107:517–20.

Xu, C. F., K. Lewis, K. L. Cantone, P. Khan, C. Donnelly, N. White, et al. 2002. "Effectiveness of Computational Methods in Haplotype Prediction." *Human Genetics* 110:148–56.

Yamada, Y., H. Izawa, S. Ichihara, F. Takatsu, H. Ishihara, H. Hirayama, et al. 2002. "Prediction of the Risk of Myocardial Infarction From Polymorphisms in Candidate Genes." *New England Journal of Medicine* 347:1916–23.

Yang, Q. and M. J. Khoury. 1997. "Evolving Methods in Genetic Epidemiology. III. Gene-Environment Interaction in Epidemiologic Research." *Epidemiologic Reviews* 19:33–43.
Yang, Z., N. J. Camp, H. Sun, Z. Tong, D. Gibbs, D. J. Cameron, et al. 2006. "A Variant of the *HTRA1* Gene Increases Susceptibility to Age-Related Macular Degeneration." *Science* 314:992–93.
Zhang, K., Z. S. Qin, J. S. Liu, T. Chen, M. S. Waterman, and F. Sun. 2004. "Haplotype Block Partitioning and Tag SNP Selection Using Genotype Data and Their Applications to Association Studies." *Genome Research* 14:908–16.
Zhang, S., A. J. Pakstis, K. K. Kidd, and H. Zhao. 2001. "Comparisons of Two Methods for Haplotype Reconstruction and Haplotype Frequency Estimation From Population Data." *American Journal of Human Genetics* 69:906–14.

Kari E. North is an associate professor of genetic epidemiology at the University of North Carolina at Chapel Hill. Her PhD training is in genetic epidemiology, and she has played key roles in projects aiming to detect, measure, and characterize the effects of genes and environmental factors on variation in risk factors for cardiovascular disease.

Lisa J. Martin is an associate professor in the Division of Biostatistics and Epidemiology at Cincinnati Children's Hospital and the University of Cincinnati School of Medicine in Ohio. She has worked as an expert in genetic epidemiology for several projects characterizing the genetic etiology of obesity-related traits and cardiac malformations.