Genetic Linkage and Linkage Disequilibrium Analysis

Kongming Wang
Department of Statistics, Pfizer Inc., New York, New York, U.S.A.

Bernice Porjesz
Department of Psychiatry, HSCB, State University of New York, Downstate Medical Center, Brooklyn, New York, U.S.A.

Kevin Jones
Department of Psychiatry, HSCB, State University of New York, Brooklyn, New York, U.S.A.

Henri Begleiter
Department of Psychiatry, HSCB, State University of New York, Downstate Medical Center, Brooklyn, New York, U.S.A.

Abstract
Genetic linkage and linkage disequilibrium analyses, and in general pharmacogenetics, can help biopharmaceutical research and development in identification of drug targets, disease diagnoses, prediction of efficacy and side-effects of drugs, and prescription of personalized medicine. In this paper we review two common approaches, regression and variance component methods, for linkage and linkage disequilibrium analyses. Combined linkage and linkage disequilibrium analysis will be discussed. In addition, we will also review the endophenotype approach to mapping complex disorders such as alcoholism.

INTRODUCTION

Genetic mapping of human traits (phenotypes) aims to identify chromosomal regions that contain genes affecting traits of interest and especially genes that affect human susceptibility to particular diseases. There are many single-gene diseases (e.g., Huntington's disease) that are caused by a change of a single gene. There are also many polygenic diseases resulting from the combined action of alleles of more than one gene (e.g., heart disease and diabetes). A great deal of attention has been focused on identifying disease genes in order that they may be used for disease diagnoses, personalized treatment, and the prediction of treatment outcomes.

Pharmacogenetics can aid biopharmaceutical research and development in many ways. First, it is applied in diagnosis of disease. An important example is prenatal genetic screening, such as amniocentesis for the Down syndrome and newborn genetic screening for inherited metabolic disorders. Second, it is applied in drug efficacy prediction. A typical clinical trial would have a third or more patients who do not respond to the test drug. By identifying the genes that are associated with drug response, it is possible to improve drug response rates in clinical trials by selecting targeted patient populations. See Ref.[1] for a list of genes that have been significantly associated with drug response. Third, similar to predicting drug efficacy, it is applied in side effects prediction of a drug (common side effects in clinical trials and/or rare side effects of marketed drug). Fourth, it would be applied in market expansion in the sense of identifying patients who are currently unsuited to the drug but potentially responsive to the drug when dosage or formula is selected based on their genotypes. Finally, clinical, genetic, and molecular phenotypes can be integrated to identify drug targets.

In this article we will describe the basic concepts of and commonly used methods in genetic mapping [linkage and linkage disequilibrium (LD) analysis] of quantitative trait loci (QTL). Interested readers can go through the references cited here or elsewhere for more information on mapping QTL. Classical linkage analysis is described in section "Classical Linkage Analysis," and it can be used to construct a genetic map according to the distances between marker pairs. The idea of linkage and LD analysis is to scan the genome to detect the correlation between phenotype and the markers. Two commonly used methods used for linkage and LD analysis, regression and variance-component methods, are described in section "Mapping QTL." Section "Genetics of Complex Diseases: The Endophenotype Approach" describes some of our work on mapping QTL of neurological and psychiatric disorders. The last section is devoted to discussion.

Encyclopedia of Biopharmaceutical Statistics DOI: 10.1081/E-EBS-120024876

CLASSICAL LINKAGE ANALYSIS

Chromosomes are long strands of deoxyribonucleic acid (DNA) made up of four bases or nucleotides: A (adenine), C (cytosine), T (thymine), and G (guanine). Each individual has 23 pairs of chromosomes. Each chromosome pair has a paternal homologue (a copy carried by the sperm) and a maternal homologue (a copy carried by the egg). Genes are basic units of inheritance (short segments of DNA) and act as if linearly arranged at field places (loci) on a chromosome transmitted in the gametes from parent to offspring. It is now believed that the human genome has around 20,000–25,000 genes. A genetic marker is a segment of DNA with an identifiable physical location on a chromosome and whose inheritance can be followed. A marker can be a gene or a group of genes, or it can be some section of DNA with no known function. One type of marker is a Short Tandem Repeat, which is a sequence of DNA nucleotides (dinucleotides, trinucleotides, etc.) that varies among individuals in terms of how many times the sequence is repeated (AT, ATAT, ATATAT, etc.). A more recently developed marker is the single nucleotide polymorphism (SNP) in which individuals differ in a single nucleotide at that location (e.g., some individuals have the sequence ATAGCTA, while others have the sequence ATAGGTA). Markers are selected that vary among individuals, i.e., the selected markers are informative. Because DNA segments that lie near each other on a chromosome tend to be inherited together, markers are often used as indirect ways of tracking the inheritance pattern of a gene that has not yet been identified but whose approximate location is known. An allele is one of several alternative forms of a marker occupying the same locus on a particular chromosome. Genotype data at a marker refer to the maternal and paternal alleles at the marker.

The process of forming germ cells is called meiosis. During meiosis, the chromosome pairs are split and the maternal and paternal homologues recombine (crossover), forming two new homologues. This random shuffling of genetic material is called recombination. Every germ cell (sperm or egg) contains one homologue from each of the 23 chromosome pairs. During fertilization, a sperm and an egg combine, and the paternal and maternal homologues form a full chromosomal set. As a result of crossover, each offspring will carry genes from all four grandparents. If no crossover had occurred, the offspring would have only inherited genes from two grandparents.

The genetic distance between two markers is defined as the expected number of crossovers between them. Note that genetic distance is not a measure of the physical distance between the two markers on a chromosome. Haldane's model of recombination is that crossover occurs according to a Poisson process, leading to the mapping function

$$\theta = (1 - e^{-2x})/2$$

where θ is the recombination probability between two markers and x is the genetic distance in Morgans (expected number of crossovers) between the two markers. See Ref.[2] for other mapping functions. With all distances between marker pairs being computed, a genetic map can be constructed to show the relative locations of the markers along the chromosome. Genetic maps are used in linkage analysis to identify markers for disease traits. New techniques have produced dense maps with more markers (e.g., SNPs), and hence enable fine mapping of disease genes.

MAPPING QTL

A qualitative trait is expressed qualitatively, which means that the phenotype (what you see) falls into different categories (e.g., blood types) and is a clear representation of its genotype. A quantitative trait shows continuous variation (e.g., height or blood pressure) because the trait is under the influence of many genes. Most of these genes have a small effect on the total phenotype value. In addition to the genetic effects associated with the quantitative trait, the phenotype is directly influenced by environmental factors. This results in the observed phenotypic variation appearing continuous in nature. This is further complicated by gene–gene interactions and gene–environment interactions. The observed phenotype for a quantitative trait is the sum of all genetic effects across all involved genes and environmental effects.

Mapping of qualitative and QTL (genes) is used to study the relation between a phenotypic value and one or more genetic markers. A large correlation between a phenotype value and a marker genotype indicates that the marker is close to a gene that influences the phenotype. Various statistical models and methods have been developed to study this correlation. Some methods are based on pedigree data (linkage analysis) while other methods are based on population and pedigree data (association methods). Combined linkage and association methods have improved statistical power. We will focus on statistical methods for mapping QTL, which are more complicated than mapping qualitative trait loci. These methods are also applicable or can be adapted to map qualitative trait loci.

Heritability

The concept of heritability was introduced to quantify the level of predictability of passage of a biologically interesting phenotype from parent to offspring.[3]

The phenotype is seen as a function of environment and genotype. Assuming additive and independent effects of genes and environment (a simplified model), phenotypic variance (V_P) of a trait in a population may be expressed by one component of genetic variance (V_G) and one of environmental variance (V_E), so that $V_P = V_G + V_E$.

Heritability measures how the genetic contribution to a trait might vary in a population, and the heritability coefficient, H^2, is given by the ratio of the total genetic variance to the phenotypic variance: $H^2 = V_G/V_P$. This is the broad heritability. Another measure, the narrow heritability, is only based on the additive genetic variation. We will discuss additive genetic variance and dominant genetic variance in the next section when we describe the variance-component model. The heritability can be used to predict the effects of searching for genes. It is easier to detect genes determining a trait if heritability of the trait is stronger. But heritability analysis does not provide any information about the locations of the contributing genes. This is the purpose of linkage/LD analysis.

The phenotypic variance (V_P) is observed, but the genetic variance (V_G) has to be estimated. If genotype data are available, marker-based methods can be employed to estimate heritability.[4,5] Note that heritability can be estimated without knowing genotypes. To estimate heritability from phenotypes without genotype data, one would need to make additional assumptions or focus on specific pairs of relatives so that phenotype variance is not completely described by environmental effects. A popular approach is to use twins[6] and assume equal environmental effects.

Linkage Analysis

As QTL are unknown, linkage analysis studies the coinheritance of phenotypes and marker genotypes within families. One expects relatives who have similar phenotypes to have similar genotypes at marker loci close to genetic loci that influence the trait, while markers at distant loci behave stochastically according to the rules of Mendelian inheritance. We will review two of the most commonly used methods and their extensions: Haseman–Elston regression[7] and variance-component model.[4,8]

Two genes are identical by descent (IBD) if one is a copy of the other or if both are physical copies of the same ancestral gene. Thus a parent and a child always share one allele IBD at each locus. Two siblings can share 0, 1, or 2 alleles IBD at each locus with probabilities 1/4, 1/2, and 1/4, respectively. For a distant pair of relatives in a complex pedigree, it is impossible to enumerate all the possibilities of allele sharing and all the breeding types, to calculate the allele sharing IBD probabilities between the pair. It is even more difficult for multipoint analysis. Statistical methods have been developed to estimate the allele sharing IBD probabilities between any pair of a pedigree.[4,9,10] Some methods are deterministic while most are stochastic (Monte Carlo techniques).

If linkage exists between a trait locus and a marker, then alleles sharing IBD at the two loci are correlated. The Haseman and Elston[7] method (H–E method) is derived for sib pairs and for regressing the squared sib-pair trait phenotype difference on the proportion of alleles the sibs are estimated to share IBD at a marker locus. Let y_{1j} and y_{2j} be the normalized phenotype values of the jth sib pair such that the population mean is zero. Let π_j be the estimated proportion of alleles the jth sib pair shared IBD at a marker locus. The slope β from the regression line

$$(y_{1j} - y_{2j})^2 = \beta_0 + \pi_j\beta + \varepsilon_j$$

indicates whether a linkage exists between the trait locus and the marker locus. The intercept β_0 is of no particular interest, and ε_j is the residual. The slope β is negatively proportional to the genetic part of the phenotype variance. If β is negative then there exists linkage between a trait locus and the marker locus, because it correlates similarity at a trait locus with similarity at the marker locus (sharing more alleles IBD makes the phenotype difference between the sib pair smaller). For each marker on a genetic map, one can carry out the regression and find out which markers are statistically significantly linked to the trait loci. The trait loci are close to these chromosome locations. With more dense maps available because of advanced technology, one can do two sweeps to save computational time: The first sweep is on a relatively sparse map to find the rough locations of the trait loci, and the second sweep is on all the available markers in a neighborhood of those rough locations. Note that this is a multitest problem, but the p-values are not adjusted.

Extensions of the H–E method have been developed to increase statistical power. Wright[11] pointed out that sib-pair QTL linkage information is not fully utilized by considering only phenotype differences of sib pairs. Further information can be obtained from phenotype sums of sib pairs. Sham and Purcell[12] proposed a combined H–E method. This method regresses a linear combination of $(y_{1j} + y_{2j})^2$ and $(y_{1j} - y_{2j})^2$ on π_j, the estimated proportion of alleles shared IBD at a marker locus by sib pairs. Elston et al.[13] introduced an improved H–E method which regresses the products of the sib phenotypes, $y_{1j}\, y_{2j}$, on the estimated proportion of alleles shared IBD. Olson and Wijsman[14] extended the H–E method for use with general pedigrees, considering $(y_{1j} - y_{2j})^2$ for all relative pairs. These H–E methods typically use the Wald statistic based on the SE from ordinary least squares.

For the variance-component model,[4,8] the phenotype values of a pedigree are assumed to follow a multivariate normal distribution. Let $Y_i = (y_{i1}, y_{i2}, \ldots)'$ be the phenotypes of the ith pedigree. If there are n trait loci influencing the trait, then the variance-component model can be written as

$$y_{ik} = \mu + x_{ik}\beta + \sum_{j=1}^{n} + \gamma_{ikj} + \varepsilon_{ik} \quad (1)$$

where μ is the grand mean, x_{ik} is a vector of covariates (e.g., age, gender), β represents fixed effects corresponding to the covariates, γ_{ikj} is the effect of the jth QTL, and ε_{ik} represents a random environmental deviation. Assume γ_{ikj} and ε_{ik} are uncorrelated random variables with mean 0 and variance σ_j^2 and σ_ε^2, respectively. With both additive and dominant effects, $\sigma_j^2 = \sigma_{\mathrm{aj}}^2 + \sigma_{\mathrm{dj}}^2$, where σ_{aj}^2 is the additive genetic variance and σ_{dj}^2 is the dominant genetic variance. The variance of y_{ik} is

$$\sigma_y^2 = \sum_{j=1}^{n}\sigma_j^2 + \sigma_\varepsilon^2$$

and the phenotype covariance between a pair of relatives indexed by k and m is

$$\begin{aligned}\mathrm{COV}(y_{ik}, y_{im}) &= E[(y_{ik} - \mu)(y_{im} - \mu)] \\ &= \sum_{j=1}^{n}[\pi_{kmj}\sigma_{\mathrm{aj}}^2 + \kappa_{kmj}\sigma_{\mathrm{dj}}^2]\end{aligned}$$

where π_{kmj} is the proportion of alleles shared IBD at the jth QTL by the relative pair and κ_{kmj} is the jth QTL-specific probability of the pair of relatives sharing two alleles IBD. Both π_{kmj} and κ_{kmj} can be estimated from genetic marker data.

To search for the QTL that influence the phenotype values, one would focus on one QTL at a time. If we focus on the jth QTL, then model (1) can be written as

$$y_{ik} = \mu + x_{ik}\beta + \gamma_{ikj} + g_{ik} + \varepsilon_{ik}$$

where g_{ik} represents the random genetic effect by all QTL other than the jth QTL. The covariance between a pair of relatives indexed by k and m is then

$$\begin{aligned}\mathrm{COV}(y_{ik}, y_{im}) &= \pi_{kmj}\sigma_{\mathrm{aj}}^2 + \kappa_{kmj}\sigma_{\mathrm{dj}}^2 + 2\phi_{km}\sigma_{\mathrm{g}}^2 \\ &\quad + \delta_{km}\sigma_{\mathrm{d}}^2 \quad (2)\end{aligned}$$

where $\phi_{km} = E[\pi_{kmj}]/2$ is the expected kinship coefficient over the genome and $\delta_{km} = E[\kappa_{kmj}]$ is the expected probability of sharing two alleles IBD.

The variance-component model uses likelihood ratios to test significance of each marker. Let $\Delta_i = Y_i - (\mu, \mu, \ldots)' - [(x_{i1}, x_{i2}, \ldots)\beta]'$ be the centered phenotype values of the ith pedigree and Ω_i be its covariance matrix at the jth QTL with elements given by Eq. (2). Assume Y_i follows a multivariate normal distribution. The log-likelihood of the ith pedigree with t individuals at the jth QTL is

$$\begin{aligned}&\ln L_i(\mu, \beta, \sigma_{\mathrm{aj}}^2, \sigma_{\mathrm{dj}}^2, \sigma_{\mathrm{g}}^2, \sigma_{\mathrm{d}}^2 | Y_i, (x_{i1}, x_{i2}, \ldots)) \\ &\quad = -\frac{t}{2}\ln(2\pi) - \frac{1}{2}\ln(|\Omega_i|) - \frac{1}{2}\Delta_i'\Omega_i^{-1}\Delta_i \quad (3)\end{aligned}$$

If there are N pedigrees, then the log-likelihood of all pedigrees is simply the sum of the log-likelihood for each pedigree. Likelihood estimation results in consistent parameter estimates even when multivariate normality assumption is violated.[8] Simulations confirm the consistency of variance-component estimates of genetic effect size.[15]

LOD Score

If the jth QTL in formula (2) does not contribute to the trait under study, then Eq. (2) becomes a reduced model

$$\mathrm{COV}(y_{ik}, y_{im}) = 2\phi_{km}\sigma_{\mathrm{g}}^2 + \delta_{km}\sigma_{\mathrm{d}}^2 \quad (4)$$

and the log-likelihood (3) is reduced to

$$\begin{aligned}&\ln L_i(\mu, \beta, \sigma_{\mathrm{g}}^2, \sigma_{\mathrm{d}}^2 | Y_i, (x_{i1}, x_{i2}, \cdots)) \\ &\quad = -\frac{t}{2}\ln(2\pi) - \frac{1}{2}\ln(|\Omega_i^r|) - \frac{1}{2}\Delta_t'(\Omega_i^r)^{-1}\Delta_i \quad (5)\end{aligned}$$

where Ω_i^r is a covariance matrix at the jth QTL with elements given by Eq. (4). If there are N pedigrees, then the log-likelihood of all pedigrees for the reduced model is the sum of the log-likelihood for each pedigree.

The hypothesis of no effect of jth QTL can be tested by likelihood ratio test. To this end, we calculate the logarithm of the odds (to the base 10) (LOD) score

$$\mathrm{LOD} = -\log_{10}\frac{\widehat{L}_{\mathrm{r}}}{\widehat{L}}$$

Here $\widehat{L}$ is the likelihood of full model (3) with parameters estimated by maximum-likelihood method, and $\widehat{L}_{\mathrm{r}}$ is the likelihood of reduced model (5) with parameters estimated by maximum-likelihood method.

It is typical to interpret a LOD score above 3 as evidence of linkage. Roughly speaking, a LOD score of 3 corresponds to a p-value of 0.0001. This is much more stringent than the usual significance level of 0.05 and is for controlling the false positive rate when performing multiple tests. To find a chromosome location

(marker) close to a QTL, we need to scan the genome at many markers.

IBD Estimation

Both H–E and variance-component methods rely on IBD sharing probabilities between relatives within a pedigree, but it is impossible to obtain IBD sharing probabilities directly at QTL. Many statistical methods have been developed to estimate IBD allele sharing at QTL from genetic marker data. The idea is that if a QTL is close to a marker, then recombination between the QTL and the marker is rare. Consequently, IBD allele sharing is almost identical at the QTL and the marker. More accurate estimate of IBD allele sharing at a QTL can be obtained by using multiple markers in the vicinity of the QTL.

Davis et al.[16] proposed an algorithm to calculate IBD probabilities for complicated pedigrees when there is no missing genetic marker information. The method of Fulker, Cherny, and Gardon[17] is a multipoint estimation of IBD allele sharing for sib pairs, and Almasy and Blangero[4] extend this multipoint approach to arbitrary pairs in a pedigree. Jung, Fan, and Jin[18] estimate IBD allele sharing at a QTL by a linear combination of IBD allele sharing at multiple markers in the vicinity of the QTL.

LD and Fine Mapping

Linkage analysis is a pedigree-based method that relies on the phenomena of recombination (each chromosome that an offspring receives from a parent consists of segments derived from both grandparents). In practice, one can only obtain DNA samples from a few generations of a pedigree. It is unlikely to observe recombination events between two closely linked markers from pedigree data. Therefore linkage analysis has limited resolution and is usually not able to localize genes within 1.0 cM.[19]

As a complement to linkage analysis, association mapping or LD mapping is a population-based analysis and offers an approach for obtaining more precise location of disease genes—the so-called fine mapping. LD refers to the statistical dependence of alleles at different loci, i.e., $P(\mathrm{AB}) \neq P(\mathrm{A})P(\mathrm{B})$. A frequently used measure is the correlation coefficient Δ^2.[20] Usually fine mapping uses dense markers of SNP. A SNP gives rise to two alleles. Consider two biallelic loci with alleles A and a at one locus and alleles B and b at the other, where labeling is arbitrary. Let π_A, π_a, π_B, and π_b be the allele frequencies, and let π_{AB}, π_{aB}, π_{Ab}, and π_{ab} be the four haplotype frequencies. Then

$$\Delta^2 = (\pi_{AB} - \pi_A \pi_B)^2/(\pi_A \pi_a \pi_B \pi_b)$$

A Chi-square test can be constructed to test the hypothesis of independence between marker pairs.[19] Note that linkage always implies association, but association (correlation) between two markers does not necessarily imply linkage, because the correlation could be because of population stratification. So association mapping has high resolution but with a high false positive rate. One can use linkage analysis to scan the genome for markers affecting a trait, and then use association mapping in the neighborhoods of those markers for precise locations of the QTL.

For mapping QTL one has to detect correlation between markers and quantitative traits. Luo, Tao, and Zeng[21] proposed a likelihood analysis method to estimate LD between a marker and a QTL using unrelated individuals in a natural population. With this maximum-likelihood analysis, the QTL genotypes are considered as missing data and an expectation-maximization algorithm is applied to infer the genotype information. Then QTL allele frequencies, LD between the QTL and the marker under consideration, and QTL effects are estimated simultaneously. Thus the association analysis does not require pedigree data, though pedigree information can be used to avoid bias because of population stratification and to have additional power.

Association studies can also be regarded as identity by state (IBS) analysis. Two alleles are IBS if they are of the same type but are not necessarily pure copies. Two individuals sharing alleles IBS at a marker are not necessarily sharing alleles IBD. The regression and variance-component methods can be applied to both pedigree data and population data using allele sharing IBS. LD between an SNP and a QTL can be achieved by incorporation of the mean effects model as a measured covariate[22–25] with likelihood approach. Consider a QTL with alleles A and a that contribute to the phenotype y. There are three genotypes, G = AA, Aa, and aa, at the SNP. Because the contribution of the SNP to trait y depends on the genotype G, the mean of the trait $E(\mathrm{y})$ will depend on the genotype G as well. Assuming additive effects of SNP alleles, the mean of the trait can be modeled as

$$E(\mathrm{y}|\mathrm{G}) = \mu + \beta N_A$$

where N_A is the number of A alleles corresponding to genotype G, and β is the regression coefficient describing the strength of the influence of the genotype on the trait.[25] It is possible to test whether the phenotypic mean differs between individuals with different genotypes by comparing the likelihood of a model with genotype-specific trait means to that of a model with equal means for all genotypes.

Combined Linkage and LD Analysis

Family pedigree data can be used for both linkage and LD analysis, and population data can be used in LD analysis. Combined linkage and LD mapping would allow the use of both pedigree and population data, reduce the false positive linkage because of the use of pedigree allele sharing IBD information, and at the same time allow the use of more dense maps (e.g., SNPs) for fine mapping because of LD analysis. Combined linkage and LD analysis will have more power in detecting QTL.

Almasy et al.[26] proposed an approach to use allele sharing IBS information between unrelated individuals to augment the linkage information obtained from allele sharing IBD among relatives. This is an extension of the variance-component method for linkage analysis of Almasy and Blangero.[4] Fulker et al.[23] proposed a combined linkage and LD analysis using sib pairs. Their likelihood approach models linkage parameters in the covariance structure while they model LD parameters on the means of the phenotype. The idea is that population stratification will influence sib-pair means of the phenotype but not sib-pair differences. So modeling the gene effect differently for pair means and pair differences will reduce spurious LD. Fan and Xiong[24] proposed a two-point method based on variance-component models. In their model, linkage information is contained in the variance–covariance matrix based on family data, and the LD information is contained in the mean parameter based on both family pedigree and population data.

Software

The literature on linkage and LD analysis is very rich, and software has been developed for QTL mapping in abundance. The webpage http://www.nslij-genetics.org/soft/ has a comprehensive list of software for genetic linkage analysis, marker mapping, LD mapping, and pedigree drawing. Selecting an analysis method and software depends on the experimental design (twin study, sib-pair data, general pedigree, natural population, etc.).

Regression and variance-component methods are commonly used for QTL mapping. The H–E regression method using sib-pair data is available in GENEHUNTER, and the regression method is extended to general pedigree data in Multipoint Engine for Rapid Likelihood INference (MERLIN). The variance-component method is available in GENEHUNTER, MERLIN, and Sequential Oligogenic Linkage Analysis Routines (SOLAR). We have been using SOLAR[4] for combined linkage and LD mapping of QTL of psychiatric disorders. Variance-component methods are more powerful than regression methods but with more restricted assumptions such as normality of the phenotypes. SOLAR is robust in phenotype distribution and it has a t-distribution option.

GENETICS OF COMPLEX DISEASES: THE ENDOPHENOTYPE APPROACH

Linkage analysis techniques have been successfully utilized to locate genes of rare Mendelian diseases; however, these methods have had only marginal success in locating genes associated with more common complex (non-Mendelian) disorders such as neurological and psychiatric diseases. Some of the reasons for these difficulties include the disorders resulting from small effects of many genes (polygenic disorders), incomplete or low penetrance, clinical and genetic heterogeneity, the presence of phenocopies, and diagnosis uncertainty. However, some recent successes have resulted from the use of methods, which efficiently utilize full pedigree structures in extended pedigrees, and which employ QTL underlying disease phenotypes (endophenotypes or intermediate phenotypes).[27] These quantitative biological markers serve as covariates that correlate with the main trait of interest (diagnosis) and serve to better define that trait or its underlying genetic mechanism.[28] Endophenotypes therefore represent the genetic liability of the disorder among non-affected relatives of affected individuals.

To qualify as an endophenotype of a disorder, the trait in question must meet a number of criteria:

1. The trait must be present in affected individuals and correlate with diagnosis and severity of disease or age of onset.
2. The trait must reflect susceptibility and not be the consequence of transient states.
3. The trait must be present in unaffected relatives of affected controls with levels significantly higher than in random controls.
4. The trait must be heritable.

Here we provide a brief overview of work being undertaken as part of the Collaborative Study on the Genetics of Alcoholism (COGA) project. The overarching goal of COGA is to identify genes that affect the susceptibility to develop alcohol dependence and related phenotypes. The gene mapping strategy employed by COGA integrates quantitative endophenotypes derived from human electroencephalogram (EEG) recording along with extensive psychiatric diagnostic tools. This methodology has resulted in the identification of several susceptibility genes, two of which are summarized here.

The COGA

Alcoholism is a complex disease influenced by underlying biological susceptibility factors, by environmental factors, and by complex interactions among genes and between genes and the environment. Since its inception COGA has carefully ascertained and assessed a large sample of densely affected alcoholic families, a sample of less densely affected families, and a sample of comparison families unselected for psychiatric status. This task was conducted to obtain highly informative families and to obtain reasonable power for genetic analyses. A unique approach taken by COGA is the adoption of neuroelectric measures as quantitative endophenotypes to the study of genetic risk for the development of alcoholism and related disorders. Inherent in this approach is the ability to study neuroelectric phenomenon as phenotypes of cognition, which has the potential to elucidate the genetic underpinnings of human brain oscillations.

COGA has reported several important linkage findings for the major clinical phenotype, alcohol dependence, as well as for related clinical features of alcohol dependence such as smoking, depression, suicidal behavior, and conduct disorder.[29,30] A number of the significant COGA linkage findings for alcoholism have been consequent to highly significant linkage findings using neuroelectric endophenotypes.[31,32] These significant linkage findings have been followed up with LD studies to identify specific genes within these regions of linkage.

The Utility of EEG as a Quantitative Endophenotype

Recording brain electrical activity using scalp electrodes provides a non-invasive, sensitive measure of brain function in humans. These neuroelectric phenomena may be recorded during the continuous EEG when the subject is at rest, or one may record the time-specific event-related potentials (ERPs) and event-related oscillations (EROs) during specific cognitive tasks. These neuroelectric phenomena represent important measurable correlates of human information processing and cognition. They represent traits less complex and more proximal to gene function than either diagnostic labels or traditional cognitive measures do.[33] The human brain response, as measured by recording and measuring EEG features, may be utilized as phenotypes of cognition, and as valuable tools for the understanding of some complex genetic disorders.[34]

Two examples of the use of neuroelectric endophenotypes from the COGA project are summarized below. These neuroelectric measures were obtained using two types of EEG experiment: eyes closed resting EEG and the visual oddball ERP. The neuroelectric endophenotypes used in the examples have been shown to adhere to the endophenotype qualification criteria outlined above. The first example uses resting EEG β band power which has been shown in several studies to have increased power in alcoholics, offspring of alcoholics, and high-risk individuals, particularly during the resting condition.[34] Using twin studies, resting EEG β power has been found to be highly heritable with estimates between 80% and 90%.[35] The second example uses measures derived from the P300 wave evoked during the visual oddball event-related experiment; these measures were designed to quantify the EROs underlying the P300 potential. Several studies have demonstrated that reduced P300 amplitude is found in alcoholics and their offspring, and is associated with the risk for alcoholism[34,36,37] suggesting that the P300 is a useful endophenotype indexing familial risk for alcoholism. A P300 amplitude heritability estimate of 60% has been established through a meta-analysis of five P300 twin studies.[38]

Example 1: Spontaneous β Band (16–20 Hz) EEG Oscillations during the Eyes Closed Resting State

EEG recordings were obtained with non-invasive scalp electrodes in the individuals who were awake but eyes closed. The filtered, artifact-free data were transformed into horizontal bipolar derivations. EEG absolute power between 3 and 28 Hz were subdivided into θ (3.0–7.0 Hz), $\alpha 1$ (7.5–9.0 Hz), $\alpha 2$ (9.5–12.0 Hz), $\beta 1$ (12.5–16.0 Hz), $\beta 2$ (16.5–20.0 Hz), and $\beta 3$ (20.5–28.0 Hz). A singular value decomposition procedure[39] was utilized to obtain phenotypic data for each of the six EEG power bands.

Using the EEG-$\beta 2$ phenotypic data, a variance-component linkage analysis[4] revealed significant linkage on a region of chromosome 4p, with a LOD score of 5 in alcoholic families (Fig. 1). Furthermore, a microsatellite marker in *GABRB1* in the center of this region showed LD with EEG-$\beta 2$. A combined linkage/LD analysis[26] using this marker resulted in an increased LOD score of 6.5, which equates to an association p-value of 0.004.[31] The estimated disequilibrium parameter ($\rho_d = 0.57$) indicated LD between the *GABRB1* microsatellite marker and the functional QTL. In addition, a novel non-parametric multipoint linkage technique also gave strong evidence for linkage in this region ($P < 1.0 \times 10^{-6}$) using the same EEG-$\beta 2$ phenotype.[40]

These results implicated the $GABA_A$ gene cluster as containing the causative functional genetic variants influencing the phenotype. This cluster includes the *GABRA2*, *GABRA4*, *GABRG1*, and *GABRB1* genes.

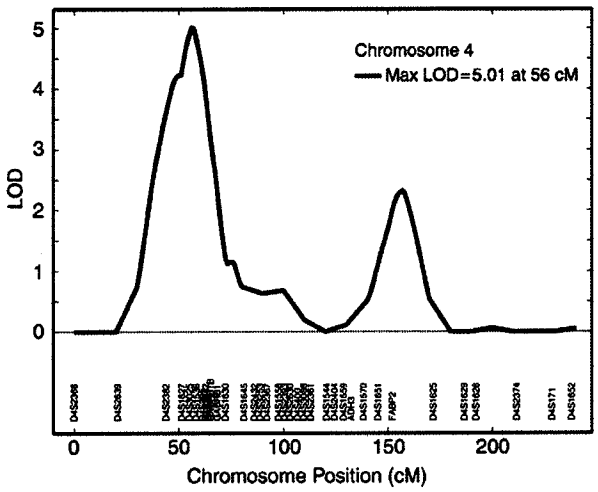

Fig. 1 Linkage curve for the $\beta 2$ band resting eyes closed EEG phenotypic data on chromosome 4. The maximum LOD score of the linkage curve is noted near the location of the *GABRB1* microsatellite marker and a cluster of $GABA_A$ receptor genes. The dataset consisted of approximately 1500 individuals from 250 families.

The actions of $GABA_A$ receptors are believed to be a fundamental requirement for the generation of high-frequency EEG oscillations (β and γ band), and blockade of these receptors has been observed to result in the loss of synchronization of these oscillations.[41] Also, a number of important effects of alcohol are mediated by GABA transmission,[42] making $GABA_A$ receptors excellent candidate genes that may affect the risk of alcoholism.

Subsequent SNP analyses indicated that the *GABRA2* $GABA_A$ receptor gene is associated with both the Diagnostics and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) diagnosis of alcohol dependence and the EEG-$\beta 2$ phenotype data.[28] Thus with the use of the EEG endophenotype, a gene was found relating central nervous system (CNS) disinhibition to genetic risk for alcohol dependence and perhaps related disorders. For the most significant SNP, individuals with a particular risk genotype had significantly elevated β power compared to individuals with other genotypes. An association between the *GABRA2* $GABA_A$ receptor gene and diagnosis of alcohol dependence has recently been replicated in independent studies.[43,44]

Example 2: Evoked θ Band (3–7 Hz) EEG Oscillations during a Visual Oddball Task

EROs elicited during the target condition of a visual oddball task were analyzed using the S-transform time-frequency decomposition method[45] and utilized as phenotypes of cognition.[32] A time window corresponding to the occurrence of the P300 ERP (300–700 msec) and frequency windows corresponding to the δ (1–3 Hz) and θ (3–7 Hz) bands were used to derive phenotypes from the single trial evoked response EEG data. Fig. 2 depicts a grand mean evoked response potential for a frontal and posterior scalp electrode (Fig. 2A) with corresponding time-frequency distributions (Fig. 2B and C). It can be observed in the head plot insets of Fig. 2B and C that the θ band ERO component of the P300 response has a

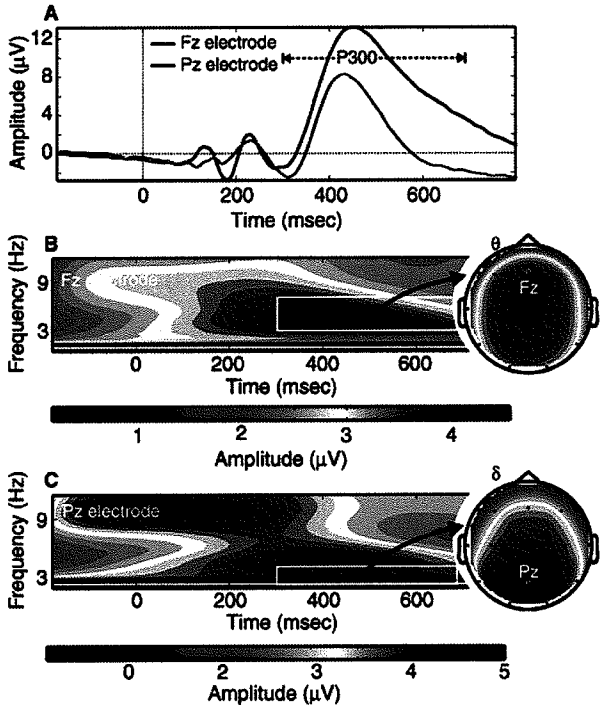

Fig. 2 Illustration of the ERO phenotypes used in the genetic analysis of visual oddball data. Traditional grand averaged evoked response potentials are depicted in (A) corresponding to target condition data for a frontal (Fz) and posterior (Pz) electrode. The P300 event is observed as a positive-going deflection starting approximately 300 msec after the target stimulus presentation. Time-frequency representations (TFR) of the target condition data (calculated using the S-transform) are depicted in (B) for the Fz electrode and (C) for the Pz electrode. These representations are obtained by averaging the instantaneous amplitude of the individual trial TFR data, and hence incorporate both stimulus phase locked and nonphase locked energy. Mean values calculated from individual subjects within time-frequency windows of interest were used as phenotypes in the genetic analysis. The inset head plots in (B) and (C) illustrate the distribution of these phenotypic measures across the head averaged within the θ and δ frequency bands and during the "P300" poststimulus time window (300–700 msec). *(View this art in color at www.informaworld.com)*

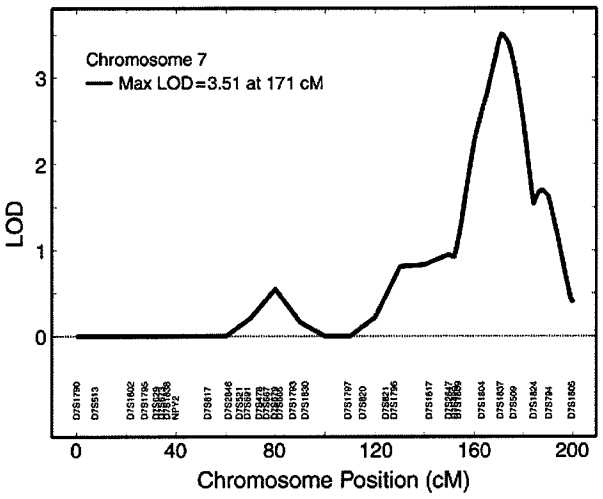

Fig. 3 Linkage curve for the target condition visual evoked θ band oscillation phenotypic data (300–700 msec and θ band TFROI) on chromosome 7. The maximum LOD score of the linkage curve is noted near the location of the cholinergic muscarinic receptor gene *CHRM2*. The dataset consisted of approximately 1300 individuals from 250 families.

frontocentral distribution on the head, whereas the δ component shows a more posterior distribution.

A genome-wide variance-component linkage analysis of frontal channel θ band ERO data revealed significant linkage (LOD = 3.5) on chromosome 7q at 171 cM (Fig. 3). A cholinergic muscarinic receptor gene, *CHRM2*, resides near this locus and is a likely candidate to account for these linkage findings. The results of subsequent association analysis of the phenotype using SNPs genotyped within and flanking the *CHRM2* gene are provided in Table 1. Significant association is observed for SNPs located in the upstream regulatory regions of the gene. It is hypothesized that the genetic underpinnings of these EROs and ERPs are likely to stem from regulatory genes that control the neurochemical processes of the brain and hence influence neural function. Significant linkage and association were only obtained for the target and not for the nontarget condition of the visual evoked response experiment, suggesting a functional significance associated with cognitive processing of the target condition in the visual oddball paradigm for the *CHRM2* gene.[32]

These findings implicate the possible role of *CHRM2* in the generation and modulation of evoked oscillations and the evoked response. Muscarinic receptors influence many effects of acetylcholine in the central and peripheral nervous system and hence are expected to have a direct influence on P300 generation.[46] Moreover, the cholinergic muscarinic genes have a major role in memory and cognition,[47] and the function of acetylcholine has been demonstrated with regard to stimulus significance,[48] selective attention,[49] and P300 generation.[50]

Recent evidence from the COGA project indicates that the *CHRM2* gene is not only related to the EROs associated with P300, but also clinical diagnosis. Significant linkage and association were reported for the *CHRM2* gene with a diagnosis of alcohol dependence and depression.[30] Thus genes important for the expression of the endophenotype (brain oscillations) help in identification of genes that increase the susceptibility for risk of alcohol dependence and related disorders.

CONCLUSIONS

A few of the current relevant issues regarding genetic mapping of complex traits will now be discussed.

Table 1 Genetic association *p*-values (uncorrected) for the frontal electrode visual oddball θ phenotype with *CHRM2* SNPs

	Measured genotype association		Gamete competition
***CHRM2* SNPs (allele frequencies)**	**Additive model—*p*-value**	**Dominant model—*p*-value**	***p*-Value**
rs1824024 (G: 0.33/T: 0.67)	0.0008***	0.09	0.007**
rs2350786 (A: 0.29/G: 0.71)	0.015*	0.11	0.02*
rs8191992 (A: 0.41/T: 0.59)	0.53	0.48	0.14
rs1378650 (C: 0.58/T: 0.42)	0.2	0.68	0.015*

These results were obtained using a Caucasian-only subset of the original data to curtail possible stratification effects. The additive model measured genotype results (implemented in SOLAR) were obtained using a population-based measured genotype SNP model in which AA SNP genotypes are coded as −1, Aa genotypes are coded as 0, and aa genotypes are coded as 1. The dominant model measured genotype results were obtained using a population-based measured genotype SNP model in which AA and aa SNP genotypes are coded as 1 and Aa SNP genotypes are coded as 0. Age, gender, age-squared, and age by gender were included in the measured genotype models as fixed effects on the phenotype trait mean. The gamete competition association method (implemented in MENDEL) tests whether transmission of an allele is correlated with higher (or lower) trait values in the recipient of the allele. Age and gender effects on the quantitative traits were removed by linear regression prior to the gamete competition analysis.

*$p < 0.05$.

**$p < 0.01$.

***$p < 0.001$.

One important question is how large a sample size is required to have a reasonable power to detect linkage/association if it exists. Pfeiffer and Gail[51] calculated sample size for population- and family-based case–control association studies using additive scores for various choices of marker allele frequencies (Table 2 in Ref.[51].). Olson and Wijsman[52] considered several designs and sample size requirements in detecting LD. The simulation of Ekstrøm[53] shows that the LOD score with 300 sib pairs is much smaller than the LOD score with 100 four-sib nuclear families without information on parental genotypes using variance-component linkage models. This confirms that a smaller number of larger pedigrees are more informative than a larger number of smaller pedigrees for linkage analysis. Multipoint analysis gives increased power over single-point analysis to map QTL. Similarly, multitrait analysis provides accurate estimates of QTL positions.[54] Multitrait methods use multivariate statistical models (multitrait single-point and multitrait multipoint). The methods described in section "Mapping QTL" can be extended to multitrait cases.

Sample size depends on required level of statistical significance, required power, design, test statistic to be used, and factors such as magnitude of effect. For linkage and LD analysis, such factors include the proportion of variance explained by QTL, gene action, marker heterozygosity, and density. The closed-form formula for sample size calculations is available in the literature. For the variance-component model, the generalized likelihood ratio test has a central Chi-square distribution under null hypothesis (no genetic component in phenotype variance) and has a non-central Chi-square distribution under alternative hypothesis. For sibship data and the variance-component model, Sham et al.[55] derived closed-form expressions of the non-centrality parameter, λ_S, per sib pair. A critical value is obtained based on the required significance level and the central Chi-square distribution. Then the required total non-centrality, λ_T, can be calculated based on the required power, the critical value, and the non-central Chi-square distribution under alternative hypothesis. The number of required sib pairs is given by

$$N = \lambda_T/\lambda_S$$

The non-centrality parameter per sib pair is:

$$\begin{aligned}\lambda_S \approx\ & V_A^2/8 + 3V_D^2/16 + V_A V_D/4 \\ & + 7V_A^4/64 + 63V_D^4/512 + 45V_A^2V_D^2/63 \\ & + 7V_A^3V_D/16 + 31V_AV_D^3 \\ & + V_S[3V_A^3/8 + 15V_D^3/32 + 9V_A^2V_D/8 + 21V_AV_D^2/8] \\ & + V_S^2[3V_A^2/8 + 9V_D^2/16 + 3V_AV_D/4]\end{aligned}$$

for linkage analysis

$$\lambda_S = (3V_A/2 + 5V_D/4)/(V_N + 2V_S)$$

for association test based on difference between-sibship means, and

$$\lambda_S = (V_A/2 + 3V_D/4)/V_N$$

for association test based on difference within sibship, where V_A, V_D, V_S, and V_N are the additive, dominant, residual shared, and residual nonshared proportions of variance. If the association test is based on individual differences, then the non-centrality per sib pair is the sum of the non-centrality parameters per sib pair of the between-sibships and within-sibship tests. The critical value is 13.8 for a required significance of LOD = 3. With this critical value, the total non-centrality to have 80% power is 20.8. So given V_A, V_D, V_S, and V_N, one can calculate N, the required number of sib pairs. Purcell, Cherny, and Sham[56] developed software for sample size calculation, which is available online (http://pngu.mgh.harvard.edu/~purcell/gpc/). The closed-form formula of sample size calculations for a popular association test, the transmission disequilibrium test that is not discussed in this article, can be found in Iles.[57]

Finally, it is expected that the future of genetic mapping will be influenced by the emerging microarray technologies. Genotypes and gene expressions of thousands of genes can be obtained simultaneously using microarrays. This provides huge amounts of data. Some statistical issues in microarray data analysis are discussed in Smyth, Yang, and Speed.[58] One of the biggest issues with such data is that of multiple statistical testing. A possible solution is to use linkage analysis to locate the approximate locations of QTL and then use microarray to study genes around those locations. Linkage/LD analysis methods can also be extended to microarray data.[59]

ACKNOWLEDGMENTS

The Collaborative Study on the Genetics of Alcoholism (COGA) (Principal Investigator: H. Begleiter; Co-Principal Investigators: L. Bierut, H. Edenberg, V. Hesselbrock, B. Porjesz) includes nine different centers where data collection, analysis, and storage take place. The nine sites and Principal Investigators and Co-Investigators are: University of Connecticut (V. Hesselbrock); Indiana University (H. Edenberg, J. Nurnberger Jr., P.M. Conneally, T. Foroud); University of Iowa (S. Kuperman, R. Crowe); SUNY Downstate Medical Center (B. Porjesz, H. Begleiter); Washington University in St. Louis (L. Bierut,

A. Goate, J. Rice); University of California at San Diego (M. Schuckit); Howard University (R. Taylor); Rutgers University (J. Tischfield); Southwest Foundation (L. Almasy). Zhaoxia Ren serves as the NIAAA Staff Collaborator. This national collaborative study is supported by the NIH Grant U10AA008401 from the National Institute on Alcohol Abuse and Alcoholism (NIAAA) and the National Institute on Drug Abuse (NIDA).

In memory of Theodore Reich, M.D., Co-Principal Investigator of COGA since its inception and one of the founders of modern psychiatric genetics, we acknowledge his immeasurable and fundamental scientific contributions to COGA and the field.

ARTICLES OF FURTHER INTEREST

Analysis of Heritability, p. 36.
Drug Development, p. 317.
Microarray Gene Expression, p. 599.
Sample Size Determination, p. 899.
Statistical Tests for Biomarker Development with Applications to Genetics Data, p. 001.
Surrogate Endpoint, p. 967.

REFERENCES

1. Goldstein, D.; Tate, S.; Sisodiya, S. Pharmacogenetics does genomic. Nat. Rev. Genet. **2003**, *4*, 937–947.
2. Ott, J. *Analysis of Human Genetic Linkage*; Johns Hopkins University Press: Baltimore, London, 1991.
3. Feldman, M.W. Heritability: some theoretical ambiguities. In *Keywords in Evolutionary Biology*; Lloyd, E.A., Keller, E.F., Eds.; Harvard University Press: Cambridge, 1992; 151–157.
4. Almasy, L.; Blangero, J. Multipoint quantitative-trait linkage analysis in general pedigrees. Am. J. Hum. Genet. **1998**, *62*, 1198–1211.
5. Ritland, K. Marker-inferred relatedness as a tool for detecting heritability in nature. Mol. Ecol. **2000**, *9*, 1195–1204.
6. Neale, M.; Cardon, L. *Methodology for Genetic Studies of Twins and Families; NATO ASCI Series D: Behavioural and Social Sciences*; Kluwer Academic Publishers, 1992.
7. Haseman, J.K.; Elston, R.C. The investigation of linkage between a quantitative trait and a marker locus. Behav. Genet. **1972**, *2*, 3–19.
8. Amos, C.I. Robust variance-components approach for assessing genetic linkage in pedigrees. Am. J. Hum. Genet. **1994**, *54*, 535–543.
9. Amos, C.I.; Dawson, D.V.; Elston, R.C. The probabilistic determination of identity-by-descent sharing for pairs of relatives from pedigrees. Am. J. Hum. Genet. **1990**, *47*, 842–853.
10. Pong-Wong, R.; George, A.W.; Woolliams, J.A.; Haley, C.S. A simple and rapid method for calculating identity-by-descent matrices using multiple markers. Genet. Sel. Evol. **2001**, *33* (5), 453–471.
11. Wright, F.A. The phenotypic difference discards sib-pair QTL linkage information. Am. J. Hum. Genet. **1997**, *60*, 740–742.
12. Sham, P.; Purcell, S. Equivalence between Haseman-Elston and variance-components linkage analyses for sib pairs. Am. J. Hum. Genet. **2001**, *68*, 1527–1532.
13. Elston, R.C.; Buxbaum, S.; Jacobs, K.B.; Olson, J.M. Haseman and Elston revisited. Genet. Epidemiol. **2000**, *19*, 1–17.
14. Olson, J.M.; Wijsman, E.M. Linkage between quantitative trait and marker loci: methods using all relative pairs. Genet. Epidemiol. **1993**, *10*, 87–102.
15. Blangero, J.; Almasy, L. Multipoint oligogenic linkage analysis of quantitative traits. Genet. Epidemiol. **1997**, *14*, 959–964.
16. Davis, S.; Schroeder, M.; Goldin, L.; Weeks, D. Nonparametric simulation-based statistics for detecting linkage in general pedigrees. Am. J. Hum. Genet. **1996**, *58*, 867–880.
17. Fulker, D.; Cherny, S.; Gardon, L. Multipoint interval mapping of quantitative trait loci, using sib pairs. Am. J. Hum. Genet. **1995**, *56*, 1224–1233.
18. Jung, J.; Fan, R.; Jin, L. Combined linkage and association mapping of quantitative trait loci by multiple markers. Genetics **2005**, *170*, 881–898.
19. Weir, B. *Genetic Data Analysis II*; Sinauer Associates: Sunderland, MA, 1996.
20. Pritchard, J.; Przeworski, M. Linkage disequilibrium in humans: models and data. Am. J. Hum. Genet. **2001**, *69*, 1–14.
21. Luo, Z.L.; Tao, S.H.; Zeng, Z.-B. Inferring linkage disequilibrium between a polymorphic marker locus and a trait locus in natural populations. Genetics **2000**, *156*, 457–467.
22. Boerwinkle, E.; Chakraborty, R.; Sing, C.F. The use of measured genotype information in the analysis of quantitative phenotypes in Man. I. Models and analytical methods. Ann. Hum. Genet. **1986**, *50*, 181–194.
23. Fulker, D.; Cherny, S.; Sham, P.; Hewitt, J. Combined linkage and association sib-pair analysis for quantitative traits. Am. J. Hum. Genet. **1999**, *64*, 259–267.
24. Fan, R.; Xiong, M. Combined high resolution linkage and association mapping of quantitative trait loci. Eur. J. Hum. Genet. **2002**, *11*, 125–137.
25. Iachine, I. Master of Applied Statistics, 2004. Available at: http://statmaster.sdu.dk/courses/st115.
26. Almasy, L.; Williams, J.T.; Dyer, T.D.; Blangero, J. Quantitative trait locus detection using combined linkage/disequilibrium analysis. Genet. Epidemiol. **1999**, *17*, S31–S36.
27. Almasy, L.; Blangero, J. Endophenotypes as quantitative risk factors for psychiatric disease: rationale and study design. Am. J. Med. Genet. **2001**, *105*, 42–44.
28. Gottesman, I.; Gould, T.D. The endophenotype concept in psychiatry: etymology and strategic intentions. Am. J. Psychiatr. **2003**, *160*, 636–645.
29. Edenberg, H.J.; Dick, D.M.; Xuei, X.; Tian, H.; Almasy, L.; Bauer, L.O.; Crowe, R.R.; Goate, A.; Hesselbrock, V.; Jones, K.; Kwon, J.; Li, T.K.; Nurnberger, J.I.; O'Connor, S.J.; Reich, T.; Rice, J.; Schuckit, M.A.; Porjesz, B.; Foroud, T.; Begleiter, H. Variations in

GABRA2, encoding the α2 subunit of the GABA(A) receptor, are associated with alcohol dependence and with brain oscillations. Am. J. Hum. Genet. **2004**, *74*, 705–714.
30. Wang, J.C.; Hinrichs, A.L.; Stock, H.; Budde, J.; Allen, R.; Bertelsen, S.; Kwon, J.M.; Wu, W.; Dick, D.M.; Rice, J.; Jones, K.; Nurnberger, J., Jr.; Tischfield, J.; Porjesz, B.; Edenberg, H.J.; Hesselbrock, V.; Crowe, R.; Schuckit, M.; Begleiter, H.; Reich, T.; Goate, A.M.; Bierut, L.J. Evidence of common and specific genetic effects: association of the muscarinic acetylcholine receptor *M2* (*CHRM2*) gene with alcohol dependence and major depressive syndrome. Hum. Mol. Genet. **2004**, *13*, 1903–1911.
31. Porjesz, B.; Almasy, L.; Edenberg, H.; Wang, K.; Chorlian, D.B.; Foroud, T.; Goate, A.; Rice, J.P.; O'Connor, S.; Rohrbaugh, J.; Kuperman, S.; Bauer, L.; Crowe, R.R.; Schuckit, M.A.; Hesselbrock, V.; Conneally, P.M.; Tischfield, J.A.; Li, T.K.; Reich, T.; Begleiter, H. Linkage disequilibrium between the beta frequency of the human EEG and a GABA(A) receptor gene locus. Proc. Natl. Acad. Sci. U.S.A. **2002**, *99*, 3729–3733.
32. Jones, K.A.; Porjesz, B.; Almasy, L.; Bierut, L.; Goate, A.; Wang, J.C.; Wu, W.; Bauer, L.O.; Chorlian, D.B.; Dick, D.M.; Edenberg, H.J.; Foroud, T.; Hesselbrock, V.; Kuperman, S.; Nurnberger, J., Jr.; O'Connor, S.J.; Schuckit, M.A.; Stimus, A.T.; Reich, T.; Begleiter, H. Linkage and linkage disequilibrium of evoked EEG oscillations with *CHRM2* receptor gene polymorphisms: implications for human brain dynamics and cognition. Int. J. Psychophysiol. **2004**, *53*, 75–90.
33. Tsuang, M.T.; Faraone, S.V. The future of psychiatric genetics. Curr. Psychiatr. Rep. **2000**, *2*, 133–136.
34. Porjesz, B.; Rangaswamy, M.; Kamarajan, C.; Jones, K.A.; Padmanabhapillai, A.; Begleiter, H. The utility of neurophysiological markers in the study of alcoholism. Clin. Neurophysiol. **2005**, *116*, 993–1018.
35. Van Beijsterveldt, C.E.M.; Boomsma, D.I. Genetics of the human electroencephalogram (EEG) and event-related brain potentials (ERPs): a review. Hum. Genet. **1994**, *94*, 319–330.
36. Begleiter, H.; Porjesz, B.; Bihari, B.; Kissin, B. Event-related potentials in boys at risk for alcoholism. Science **1984**, *225*, 1493–1496.
37. Polich, J.; Pollock, V.E.; Bloom, F.E. Meta-analysis of P300 amplitude from males at risk for alcoholism. Psychol. Bull. **1994**, *115*, 55–73.
38. Van Beijsterveldt, C.E.M.; Van Baal, G.C.M. Twin and family studies of the human electroencephalogram: a review and a meta-analysis. Biol. Psychol. **2002**, *61*, 111–138.
39. Wang, K.; Begleiter, H.; Porjesz, B. Trilinear modeling of event-related potentials. Brain Topogr. **2000**, *12*, 263–271.
40. Ghosh, S.; Begleiter, H.; Porjesz, B.; Edenberg, H.; Foroud, T.; Reich, T. Linkage mapping of beta 2 EEG waves via non-parametric regression. Am. J. Med. Genet. **2003**, *118B*, 66–71.
41. Haenschel, C.; Baldeweg, T.; Croft, R.J.; Whittington, M.; Gruzelier, J. Gamma and beta frequency oscillations in response to novel auditory stimuli: a comparison of human electroencephalogram (EEG) data with in vitro models. Proc. Natl. Acad. Sci. U.S.A. **2000**, *97*, 7645–7650.
42. Harris, R.A.; Mihic, S.J.; Valenzuela, C.F. Alcohol and benzodiazepines: recent mechanistic studies. Drug Alc. Dep. **1998**, *51*, 155–164.
43. Covault, J.; Gelernter, J.; Hesselbrock, V.; Nellissery, M.; Kranzler, H.R. Allelic and haplotypic association of *GABRA2* with alcohol dependence. Am. J. Med. Genet. **2004**, *129B*, 104–109.
44. Xu, K.; Westly, E.; Taubman, J.; Astor, W.; Lipsky, R.H.; Goldman, D. Linkage disequilibrium relationships among GABRA cluster genes located on chromosome 4 with Alcohol dependence in two populations. Alcohol Clin. Exp. Res. **2004**, *28*, 48A.
45. Stockwell, R.G.; Mansinha, L.; Lowe, R.P. Localization of the complex spectrum: the S-transform. IEEE Trans. Sig. Proc. **1996**, *44*, 998–1001.
46. Frodl-Bauch, T.; Bottlender, R.; Hegerl, U. Neurochemical substrates and neuroanatomical generators of the event-related P300. Neuropsychobiology **1999**, *40*, 86–94.
47. Comings, D.E.; Wu, S.; Rostamkhani, M.; McGue, M.; Lacono, W.G.; Cheng, L.S.-C.; MacMurray, J.P. Role of cholinergic muscarinic 2 receptor (*CHRM2*) gene in cognition. Mol. Psychiatr. **2003**, *8*, 10–13.
48. Perry, E. Cholinergic mechanisms and cognitive decline. Eur. J. Anaesthesiol. **1998**, *15*, 768–773.
49. Mitrofanis, J.; Guillery, R.W. New views of the thalamic reticular nucleus in the adult and in the developing brain. Trends Neurosci. **1993**, *16*, 240–245.
50. Callaway, E. The pharmacology of human information processing. Psychopharmacologia **1983**, *20*, 359–370.
51. Pfeiffer, R.; Gail, M. Sample size calculations for population- and family-based case-control association studies on marker genotypes. Genet. Epidemiol. **2003**, *25*, 136–148.
52. Olson, J.M.; Wijsman, E.M. Design and sample-size considerations in the detection of linkage disequilibrium with a disease locus. Am. J. Hum. Genet. **1994**, *55* (3), 574–580.
53. Ekstrøm, C.T. Power of multipoint identity-by-descent methods to detect linkage using variance component models. Genet. Epidemiol. **2001**, *21*, 285–298.
54. Lund, M.; Sørensen, P.; Guldbrandtsen, B.; Sorensen, D. Multitrait fine mapping of quantitative trait loci using combined linkage disequilibria and linkage analysis. Genetics **2003**, *163*, 405–410.
55. Sham, P.; Cherny, S.; Purcell, S.; Hewitz, J. Power of linkage versus association analysis of quantitative traits, by use of variance-components models, for sibship data. Am. J. Hum. Genet. **2000**, *66*, 1616–1630.
56. Purcell, S.; Cherny, S.; Sham, P. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. Bioinformatics **2003**, *19* (1), 149–150.
57. Iles, M. On calculating the power of a TDT study—comparison of methods. Ann. Hum. Genet. **2002**, *66*, 323–328.
58. Smyth, G.; Yang, Y.; Speed, T. Statistical issues in cDNA microarray data analysis. In *Functional Genomics: Methods and Protocols*; Brownstein, M.J., Khodursky, A.B., Eds.; Humana Press: Totowa, NJ, 2003; Vol. 224 —*Methods in Molecular Biology*, 111–136.
59. Ott, J.; Hoh, J. Set association analysis of SNP case-control and microarray data. J. Comput. Biol. **2003**, *10*, 569–574.