BMC Genetics

Proceedings

Description of the data from the Collaborative Study on the Genetics of Alcoholism (COGA) and single-nucleotide polymorphism genotyping for Genetic Analysis Workshop 14

Howard J Edenberg*[1], Laura J Bierut[2], Paul Boyce[3], Manqiu Cao[4], Simon Cawley[4], Richard Chiles[4], Kimberly F Doheny[3], Mark Hansen[5], Tony Hinrichs[2], Kevin Jones[6], Mark Kelleher[4], Giulia C Kennedy[4], Guoying Liu[4], Gregory Marcus[4], Celeste McBride[5], Sarah Shaw Murray[5], Arnold Oliphant[5], James Pettengill[3], Bernice Porjesz[6], Elizabeth W Pugh[3], John P Rice[2], Todd Rubano[5], Stu Shannon[5], Rhoberta Steeke[5], Jay A Tischfield[7], Ya Yu Tsai[3], Chun Zhang[4] and Henri Begleiter[6]

Address: [1]Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, 635 Barnhill Drive, Indianapolis, IN, USA, [2]Department of Psychiatry, Washington University, 660 South Euclid, St. Louis, MO 63110-1093, USA, [3]Center for Inherited Disease Research, Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA, [4]Affymetrix, Inc., 3380 Central Expressway, Santa Clara CA 95051, USA, [5]Illumina, Inc., 9885 Towne Centre Drive, San Diego, CA 92121-1975, USA, [6]Department of Psychiatry, State University of New York, Downstate Medical Center, 450 Clarkson Avenue Box 1203, Brooklyn, NY 11203-2098, USA and [7]Department of Genetics, Rutgers University, 604 Allison Road, Piscataway, NJ 08854-8082, USA

Email: Howard J Edenberg* - edenberg@iupui.edu; Laura J Bierut - bierutl@psychiatry.wustl.edu; Paul Boyce - pboyce@illumina.com; Manqiu Cao - Manqiu_Cao@affymetrix.com; Simon Cawley - Simon_Cawley@affymetrix.com; Richard Chiles - Richard_Chiles@affymetrix.com; Kimberly F Doheny - kdoheny@cidr.jhmi.edu; Mark Hansen - mhansen@illumina.com; Tony Hinrichs - tony@silver.wustl.edu; Kevin Jones - k.jones@cns.hscbklyn.edu; Mark Kelleher - Mark_Kelleher@affymetrix.com; Giulia C Kennedy - Giulia_Kennedy@affymetrix.com; Guoying Liu - Guoying_Liu@affymetrix.com; Gregory Marcus - Gregory_Marcus@affymetrix.com; Celeste McBride - cmcbride@illumina.com; Sarah Shaw Murray - smurray@illumina.com; Arnold Oliphant - aoliphant@illumina.com; James Pettengill - petteja@yahoo.com; Bernice Porjesz - bp@cns.hscbklyn.edu; Elizabeth W Pugh - ewp@cidr.jhmi.edu; John P Rice - john@zork.wustl.edu; Todd Rubano - trubano@illumina.com; Stu Shannon - sshannon@illumina.com; Rhoberta Steeke - rsteeke@illumina.com; Jay A Tischfield - jay@nel-exchange.rutgers.edu; Ya Yu Tsai - ytsai@cidr.jhmi.edu; Chun Zhang - Chun_Zhang@affymetrix.com; Henri Begleiter - hb@cns.hscbklyn.edu

* Corresponding author

Abstract

The data provided to the Genetic Analysis Workshop 14 (GAW 14) was the result of a collaboration among several different groups, catalyzed by Elizabeth Pugh from The Center for Inherited Disease Research (CIDR) and the organizers of GAW 14, Jean MacCluer and Laura Almasy. The DNA, phenotypic characterization, and microsatellite genomic survey were provided by the Collaborative Study on the Genetics of Alcoholism (COGA), a nine-site national collaboration funded by the National Institute of Alcohol and Alcoholism (NIAAA) and the National Institute of Drug Abuse (NIDA) with the overarching goal of identifying and characterizing genes that affect the susceptibility to develop alcohol dependence and related phenotypes. CIDR, Affymetrix, and Illumina provided single-nucleotide polymorphism genotyping of a large subset of the COGA subjects. This article briefly describes the dataset that was provided.

Background

Complex diseases, such as alcohol dependence, are influenced by genetic susceptibility, environmental factors, and by interactions among genes and between genes and environment. The Collaborative Study on the Genetics of Alcoholism (COGA) has utilized a multidisciplinary approach, bringing together expertise in many domains to study this complex and important health problem. COGA has been committed to sharing data with researchers in this field to expedite progress in understanding alcoholism and related phenotypes. COGA has also provided data to Genetic Analysis Workshop 11 (GAW11) [1], and has created an archival database of these families, with both phenotypic data and immortalized cell lines; these data are accessible to investigators for further study through NIAAA http://www.niaaa.nih.gov/ResearchInformation/ExtramuralResearch/SharedResources/proj coga.htm.

COGA was designed as a family study, incorporating detailed assessments of the participants in many domains to allow derivation and study of endophenotypes along with diagnostic phenotypes. Genome surveys, using microsatellite markers, have been performed on both an initial dataset of 105 multigenerational pedigrees and a replication dataset with 157 multigenerational pedigrees. The results of genome surveys on these datasets have been published [e.g., [2-6]], along with analyses that combined the two [e.g., [7-12]]. Linkage studies of clinical phenotypes and electrophysiological endophenotypes have led to identification of genes involved in brain function as well as genes involved in alcohol dependence and related disorders. COGA has moved beyond identifying regions of linkage and is now identifying individual genes within those regions, using targeted single-nucleotide polymorphism (SNP) genotyping in which multiple SNPs were analyzed for each regional candidate gene. Genes identified include *GABRA2* [9], *GABRG3* [12], and *CHRM2* [10,11].

To test the relative merits of SNPs and microsatellites for localizing genes that contribute to complex diseases and their risk factors, COGA has collaborated with GAW and Center for Inherited Disease Research (CIDR), who enlisted two companies (Affymetrix and Illumina) to generate genome screens using SNPs. CIDR has provided high throughput genotyping of short tandem repeats (STR) markers since 1997, currently providing 11 million STR genotypes per year. As SNP genotyping methods have become more affordable and more amenable to genotyping large numbers of SNPs and samples, CIDR wished to address both a perceived need in the statistical genetics community for additional research related to the analysis of large amounts of SNP data in pedigrees, and the need for CIDR to test high throughput SNP platforms to make

an informed decision regarding SNP genotyping services. To address these needs CIDR, in conjunction with Affymetrix and Illumina, provided SNP genotyping of the COGA dataset for GAW14.

The Affymetrix mapping 10k assay [13-15] is an innovative approach that enables rapid typing of 11,560 SNP markers on an array using a single PCR primer and only 250 ng of genomic DNA. The Affymetrix assay uses allele specific hybridization.

The Illumina SNP detection assay [16,17] utilizes allele-specific extension and ligation chemistries. Total genomic DNA is bound to paramagnetic beads. For each SNP, three oligonucleotides are used to interrogate the locus. Two allele-specific oligonucleotides (ASO) each incorporate one of the two possible nucleotides. The third locus-specific oligonucleotide (LSO) anneals 1 to 20 bases downstream of the SNP. This LSO contains a locus-specific address that binds to a complementary address on beads contained in a Sentrix Array Matrix. Specific extension of the complementary ASO occurs joining to the LSO by ligation. Three universal PCR primers are used to amplify the ligated product and incorporate allele-specific fluorescent dyes. Up to 1,536 loci may be multiplexed in one reaction. The Linkage III Panel contains over 4,600 SNP markers distributed evenly across the human genome.

Methods

COGA ascertainment and assessment

Initial ascertainment of alcohol-dependent probands (designated Stage I) was performed by screening consecutive admissions at treatment facilities. Probands were assessed with the Semi-Structured Assessment for the Genetics of Alcoholism (SSAGA), a comprehensive diagnostic instrument developed for this study and now widely used [18,19]. Extensive histories of substance use and abuse were gathered along with diagnostic information for multiple Axis I disorders and antisocial personality disorder. To be recruited into the COGA study, probands had to meet both the diagnostic criteria for alcohol dependence (by DSM-III-R criteria [20] and the criteria for definite alcoholism specified by Feighner et al. [21]); thus, the COGA sample is representative of a severely alcohol-dependent population. All first degree relatives of the probands were invited to participate. Children and adolescents in the families were assessed with complementary age-appropriate instruments (C-SSAGA, child and adolescent versions). A set of control families was ascertained to provide normative measures; they were not screened to eliminate those with psychiatric disorders, and are similar to a general population sample. Written informed consent was obtained from all subjects, and the Institutional Review Boards (IRB) of each collaborative site approved all procedures. A more complete description
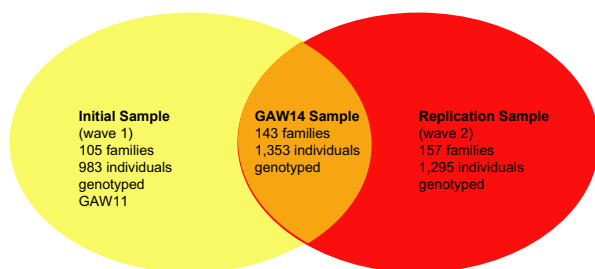
Figure 1
COGA data subset for GAW 14. The GAW 14 dataset was drawn partly from Wave 1 and partly from Wave 2.

of the recruitment procedures can be found in Begleiter et al. [1,22]. Over 13,000 individuals have been interviewed to date.

A subset of COGA families with at least three alcohol-dependent first degree relatives (designated Stage II) was identified as suitable for a genetic linkage study [1]. These families were extended by diagnostic assessment of more distant relatives in branches reached through an affected member. The Stage II families participated in a more comprehensive multi-domain assessment with an electrophysiologic evaluation of event-related potentials (ERP), event-related oscillations (EROs) and resting electroencephalogram (EEG), endophenotypes associated with alcohol dependence [23,24] that are more proximal to genes and may provide measures of the liability underlying a predisposition to alcohol dependence and related disorders.

COGA whole-genome survey
A subset of the Stage II families was selected for an initial and a replication genome survey using microsatellite markers. These pedigrees were pruned to eliminate uninformative individuals and branches from the genotyped sample. The initial sample (Wave 1) included 105 multigenerational pedigrees that include 1,214 members, of whom 983 individuals were genotyped. The replication dataset (Wave 2) included 157 multigenerational pedigrees (1,295 individuals).

Microsatellite genotyping started well before standard genome survey sets of markers were available, and therefore markers were drawn from a variety of sources [2]. Initially data were generated manually using agarose gels; later genotyping switched to automated DNA sequencers (ABI373, ABI377). Allele frequencies were estimated with the USERM13 program [25] and the CRIMAP program [26] were used to estimate marker order and distances. Maps were generated from these data.

GAW 14 dataset
Limitations on the number of individuals who could be genotyped for GAW14 led us to construct a family sample of 1,353 individuals drawn from both the initial and replication datasets (Figure 1). With non-genotyped individuals included for linking in the pedigrees, the 143 families totalled 1,614 individuals. We selected the sample starting with a core of informative families with at least 6 members who had been interviewed and genotyped, even if they did not have electrophysiological data. Family size ranged from 6 to 30. Other phenotypes forwarded for analysis included alcohol dependence, habitual smoking, and the maximum number of drinks in a 24-hour period. Of the 1,353 individuals selected for genotyping, 1,005 subjects had eyes closed EEG data available, while 905 subjects had Visual Oddball ERP/ERO data available.

It should be pointed out that there are differences in the COGA clinical and electrophysiological datasets to be analyzed by the GAW14 participants and the datasets in the previously published COGA papers [e.g., [4-12]]. These published papers use a greater number of subjects than that provided to the GAW14 participants. Hence, results found by the GAW14 participants will not be identical to those previously published. The reason for the discrepancy in the subject numbers provided to GAW14 is explained in the following points:

1) Due to budgetary constraints, not all of the COGA data were able to be genotyped for GAW14; therefore a subsample of Wave 1 and Wave 2 data was selected for genotyping. This subsample selection was based on large family size and interview status to obtain informative pedigrees. Electrophysiologic measurement was not a criterion for selection. This resulted in the selection of 1,353 subjects for genotyping in GAW14. (Note that the GAW sample consists of 1,614 individuals, because additional non-genotyped individuals were included for linking in the pedigrees).

2) In the COGA project, people who underwent the clinical questionnaire and had blood drawn for genotyping did not always undergo the electrophysiological battery. Only a subset of the total Wave 1 and Wave 2 COGA data have corresponding electrophysiological data available. The subset of the COGA Wave 1 and Wave 2 subjects with resting eyes closed EEG data is 1,553 (as published in Porjesz et al. [8]). The subset of the COGA Wave 1 and Wave 2 subjects with Visual Oddball ERP data is 1337 (as published in Jones et al. [10]).

3) The subset of the COGA dataset selected for GAW14 genotyping and the subset of the COGA dataset with people having electrophysiology data do not overlap completely. This means that out of the 1,353 subjects selected

for GAW14 genotyping 1,005 subjects had eyes closed EEG data available, while 905 subjects had Visual Oddball ERP data available.

CIDR assembly of genotyping plates

For this project, 1,396 samples were received from the COGA DNA and Cell Repository, part of the Rutgers University DNA and Cell Repository. The samples were assigned arbitrary identification numbers. Five percent of the samples were chosen to serve as internal blind duplicates and given new identification numbers. Ninety-two samples and four duplicate DNA samples were placed on each 96-well plate.

DNA was quantified using standard PicoGreen protocols from Molecular Probes. Two of the 96-well plates were randomly chosen for the replication experiment at CIDR. Identical daughter plates were robotically generated (at concentrations appropriate for each technology, 50 ng/μl for Affymetrix and 100 ng/μl for Illumina). Because the performance of the Illumina assay is sensitive to low DNA concentration, replacements for 25 samples that contained less than 50 ng/μl of DNA were requested from Rutgers and included along with the original samples sent to Illumina. Plates were shipped to Affymetrix and Illumina for genotyping in their laboratories.

Affymetrix SNP genotyping

Affymetrix was supplied with 1,396 samples (the 1,350 COGA samples along with some blind duplicates) to analyze on the GeneChip® Mapping 10k Array [27]. Total genomic DNA was digested with the restriction enzyme *Xba*I, followed by ligation of adaptors. A single primer recognizing the adaptor sequence is used to amplify the ligated DNA fragments and PCR conditions were set to preferentially amplify fragments in the range of 250–1000 bp. The amplified DNA was labeled and hybridized to GeneChip® arrays containing 25-mer DNA probes designed to hybridize with target sequence corresponding to 11,560 SNPs which are known to be located within fragments which will be amplified by the assay.

Each SNP is represented by 40 unique 25-mer DNA probes scattered throughout the array: 20 probes designed against the A allele and 20 against the B allele. Each set of 20 allele-specific probes interrogate the DNA composition at and immediately surrounding the polymorphic site. Relative allele signals are computed from the probe intensities and are used as the input to a classification scheme [28] that produces high-confidence genotype calls for each SNP.

Illumina SNP genotyping

Illumina received a DNA manifest listing the plate number, well position, and DNA concentration deter-

mined using PicoGreen for 1,396 samples. Twenty-five of these samples were below the Illumina concentration specifications. CIDR provided a second submission for each of these low-concentration samples. Both the original and replacement DNAs were genotyped and genotypes were reported for the higher quality sample. Illumina received a revised DNA manifest document listing the plate barcode, well position, concentration, and indicating the replacement relative to the original sample. Illumina received 16 96-well DNA plates containing 1,421 samples that included 25 second aliquots. CIDR placed 92 samples per plate, leaving wells A01–A04 empty for Illumina DNA control samples. Sixteen DNA plates were accessioned into the Laboratory Information Management System (LIMS) using uniquely barcoded plates.

All DNA samples were quantitated in the production lab using PicoGreen. The quantitation results were very similar to those obtained at the CIDR facility. The plates were assigned to the GAW linkage project created in the LIMS database, thus restricting their use to the assays in the Linkage set.

Fan et al. [16] and Gunderson et al. [17] provide a detailed description of the Illumina genotyping platform. All samples were genotyped using the Linkage III Panel containing 4,763 SNP markers. All genotypes were evaluated using a quantitative quality score called GenCall score. A GenCall score ranges from 0 to 1 and reflects the proximity within a cluster plot of the intensities of that genotype to the centroid of the nearest cluster. In addition, we compared the 25 original and replacement paired DNA samples using the GenCall score metric and selected a single sample in each pair of first and second aliquots. Using the GenCall score, we also identified 20 samples with very poor genotyping quality in relation to controls and all other samples. Poorly performing samples were removed from the genotyping report files, and individual genotypes with GenCall scores below 0.25 were assigned a no-call.

Affymetrix CIDR replicate

Because CIDR hoped to determine if the Affymetrix 10k assay could be used to quickly genotype large numbers of samples, standard Affymetrix 10k protocols were adapted in collaboration with Affymetrix to include automation for all liquid handling steps in a 96-well plate format. Barcode tracking was used to assign plate and well positions to specific Affymetrix GeneChips. Standard Affymetrix protocols were used for chip handling, scanning, and data analysis.

Illumina CIDR replicate

The replication genotyping experiment performed by CIDR was done using the Illumina BeadLab system. The BeadLab system incorporates automation of all DNA and liquid handling steps. The included LIMS incorporates Illumina's protocols as well as tracking and enforcing workflow. Standard Illumina protocols and reagents were used. Data analysis was performed at CIDR using Illumina's Gentrain and GTS Reports software. The cluster definitions were defined independently on the replicate data set.

CIDR quality control and data release

Data received from Affymetrix and Illumina were checked for a variety of quality control measures, then combined with COGA family file and formatted for release to GAW14. Quality control calculations included: missing data rates, error rates (based on lack of concordance of the 5% internal blind duplicate and the between lab replicate genotypes), and Mendelian inconsistencies.

Results

Affymetrix SNP genotyping

Each sample in the COGA dataset was analyzed with the standard Mapping 10k assay. Of the 1,396 samples supplied, 1,381 yielded enough DNA to analyze [The following 15 samples did not yield enough DNA to genotype: CR1371, CR1315, CR1259, CR1169, CR0959, CR0967, CR0859, CR0789, CR0563, CR0370, CR0269, CR0227, CR0150, CR0047, CR0062]. The median call rate over all 1,381 samples was over 95% with an estimated accuracy of greater than 99%. Two forms of quality control were performed on the samples before the final submission of the dataset to GAW14: calls on the X chromosome were checked against the labeled sex of each sample and families within the study were checked for Mendelian inheritance errors. These quality controls revealed 10 problematic samples (sex or pedigree inconsistencies) [The following 10 samples exhibited gender and/or Mendelian inconsistencies: sex errors: CR1234, CR1112, CR1125, CR1037, CR0728, CR0542; Mendelian inconsistencies: CR1224, CR0221; sex and Mendelian inconsistencies: CR1337, CR0538] in the samples supplied by COGA.

Genetic maps were supplied with the Affymetrix Gene-Chip® Mapping 10k Array data set to CIDR. SNPs were mapped to unique physical positions on NCBI genome build 34 and interpolated onto one of two framework genetic maps: deCode [29] and Marshfield [30]. Because those framework maps contain multiple microsatellite markers at the same genetic location, interpolation onto these maps can cause non-unique SNP positions. We therefore removed all but one microsatellite marker at each genetic location to create a non-redundant frame-

work, allowing all SNPs with unique physical positions to also have unique interpolated genetic positions [31]. These maps are periodically updated on new versions of the NCBI human genome sequence and are located on the Affymetrix NetAffx website for download [32].

Illumina SNP genotyping

Production genotyping began on January 16, 2004 and the genotyping report files were delivered on March 5, 2004, a time span of 56 days from the receipt of DNA to data delivery. The files contained the DNA barcode ID, the locus ID, the genotypes, and the GenCall score for each genotype. As genotypes are designated by alleles A and B, an allele key file was provided with context sequence for each SNP and the designation for the nucleotides that represent alleles A and B.

Genotypes were reported for a total of 1,376 DNA samples. Of the 4,763 SNP markers, Illumina reported genotypes for 4,752 resulting in a locus conversion rate of 99.77%. In addition, genetic map positions for each SNP were provided from observed meiotic recombination in 28 CEPH reference pedigrees as described in Murray et al. [33].

CIDR quality control and data release

Currently the laboratory methods used for the 10k Affymetrix assay include multiple manual steps involving the movement of DNA and reagents via a single-channel or multi-channel pipettes. As a result, one sample swap occurred in the CIDR lab. This sample problem was detected when checking for Mendelian inconsistencies, sex, and replicate errors. Identity by state sharing was calculated within families and across all samples in the dataset as an additional method to screen for problematic samples. Suspect samples were re-genotyped. Confirmation that the problems were resolved was achieved by checking lab-to-lab and within-lab replicates as well as confirming genotype concordance for 26 SNPs in common between the Affymetrix and Illumina datasets.

Two versions of the data were provided to GAW (Tables 1 and 2): raw data and clean data with Mendelian inconsistencies removed. Genotyping data was ordered using the maps provided by the companies and merged with the COGA family data to make the raw comma delimited data files. PEDCHECK [34] was used to detect Mendelian inconsistencies for each SNP. Level 0 and 1 checks were run, and Mendelian inconsistencies were removed in each nuclear family according to the rules listed below:

1) If a parent or two parents are inconsistent with a child, the genotype of the child will be zeroed out.

Table 1: Summary statistics: GAW14 dataset

	Illumina III	Affymetrix 10k
Total samples	1,376	1,380
No. SNPs	4,752	11,560
Total genotypes	6,538,752	15,952,800
Missing data rate	0.05%	5.25%
Error rate	0.005%	0.052%
Inconsistency rate	0.09%	0.12%

Table 2: Summary statistics: CIDR independent replicate data

	Illumina III	Affymetrix 10k
Total samples	192	192
No. SNPs	4,782	11,560
Total genotypes	875,106	2,127,040
Missing data rate	0.13%	6.05%
Between lab consistency rate	99.95%	99.91%*

*Error rate after correction of one laboratory sample swap

2) If a specific parent is inconsistent with more than 1 child, the genotype of that specific parent will be zeroed out.

3) If two parents are inconsistent with more than 2 children, the genotypes of the nuclear family will be zeroed out.

After removing Mendelian inconsistencies, files in the format of MERLIN and Linkage PRE-MAKEPED were generated for each chromosome with 250 SNPs per file. SNPs within all chromosome data files were ordered according to the genetic map.

Conclusion

Because the families and individuals selected for genotyping as part of GAW14 were a subset of families from both the initial and replication datasets used for COGA's published analyses, we do not expect that results will be identical to those previously published. This Genetic Analysis Workshop provides a remarkable opportunity to compare genome surveys using microsatellites to those using SNPs, in a very rich dataset that has both qualitative (e.g., diagnosis) and quantitative (e.g., electrophysiological) phenotypes reflecting a common, complex disease, alcoholism. We hope that the data we have provided will serve as a stimulus for progress in the genetic analysis of complex diseases.

Abbreviations

ASO: Allele-specific oligonucleotide

CIDR: Center for Inherited Disease Research

COGA: Collaborative Study on the Genetics of Alcoholism

EEG: Electroencephalogram

ERP: Event-related potentional

ERO: Event-related oscillation

GAW: Genetic Analysis Workshop

IRB: Institutional Review Boards

LIMS: Laboratory Information Management System

LSO: Locus-specific oligonucleotide

NIAAA: National Institute of Alcohol and Alcoholism

NIDA: National Institute of Drug Abuse

SNP: Single-nucleotide polymorphism

SSAGA: Semi-Structured Assessment for the Genetics of Alcoholism

STR: Short tandem repeat

Authors' contributions

COGA authors (HJE, LJB, TH, KJ, BP, JPR, JAT, HB) were responsible for the ascertainment, assessment, initial DNA preparation, and microsatellite genotyping. CIDR authors (PB, KFD, JP, EWP, YYT) were responsible for all DNA sample formatting and distribution, genotyping of replication sample subset on each platform at CIDR, quality control of all SNP data, and formatting of all SNP data released to GAW 14 participants. Affymetrix authors (MC, RC, MK, GCK, GL, GM, SC, CZ) and Illumina authors (MH, CM, AO, SSM, TR, SS, RS) were responsible for the SNP genotyping done with their platforms.

Acknowledgements

References

1. Begleiter H, Reich T, Nurnberger J Jr, Li TK, Conneally PM, Edenberg H, Crowe R, Kuperman S, Schuckit M, Bloom F, Hesselbrock V, Porjesz B, Cloninger CR, Rice J, Goate A: **Description of the Genetic Analysis Workshop 11 Collaborative Study on the Genetics of Alcoholism.** *Genet Epidemiol* 1999, **17(Suppl 1):**S25-S30.
2. Reich T, Edenberg HJ, Goate A, Williams JT, Rice JP, Van Eerdewegh P, Foroud T, Hesselbrock V, Schuckit MA, Bucholz K, Porjesz B, Li TK, Conneally PM, Nurnberger JII Jr, Tischfield JA, Crowe RR, Cloninger CR, Wu W, Shears S, Carr K, Crose C, Willig C, Begleiter H: **Genome-wide search for genes affecting the risk for alcohol dependence.** *Am J Med Genet* 1998, **81:**207-215.
3. Begleiter H, Porjesz B, Reich T, Edenberg HJ, Goate A, Blangero J, Almasy L, Foroud T, Van Eerdewegh P, Polich J, Rohrbaugh J, Kuperman S, Bauer LO, O'Connor SJ, Chorlian DB, Li TK, Conneally PM, Hesselbrock V, Rice JP, Schuckit MA, Cloninger R, Nurnberger J Jr, Crowe R, Bloom FE: **Quantitative trait loci analysis of human event-related brain potentials: P3 voltage.** *Electroencephalogr Clin Neurophysiol* 1998, **108:**244-250.
4. Almasy L, Porjesz B, Blangero J, Goate A, Edenberg HJ, Chorlian DB, Kuperman S, O'Connor SJ, Rohrbaugh J, Bauer LO, Foroud T, Rice JP, Reich T, Begleiter H: **Genetics of event-related brain potentials in response to a semantic priming paradigm in families with a history of alcoholism.** *Am J Hum Genet* 2001, **68:**128-135.
5. Foroud T, Edenberg HJ, Goate A, Rice J, Flury L, Koller DL, Bierut LJ, Conneally PM, Nurnberger JI, Bucholz KK, Li TK, Hesselbrock V, Crowe R, Schuckit M, Porjesz B, Begleiter H, Reich T: **Alcoholism susceptibility loci: confirmation studies in a replicate sample and further mapping.** *Alcohol Clin Exp Res* 24:933-945.
6. Edenberg HJ: **The Collaborative Study on the Genetics of Alcoholism: an update.** *Alcohol Res Health* 2002, **26:**214-218.
7. Saccone NL, Kwon JM, Corbett J, Goate A, Rochberg N, Edenberg HJ, Foroud T, Li TK, Begleiter H, Reich T, Rice JP: **A genome screen of maximum number of drinks as an alcoholism phenotype.** *Am J Med Genet* 2000, **96:**632-637.
8. Porjesz B, Almasy L, Edenberg HJ, Wang K, Chorlian DB, Foroud T, Goate A, Rice JP, O'Connor SJ, Rohrbaugh J, Kuperman S, Bauer LO, Crowe RR, Schuckit MA, Hesselbrock V, Conneally PM, Tischfield JA, Li TK, Reich T, Begleiter H: **Linkage disequilibrium between the beta frequency of the human EEG and a GABAA receptor gene locus.** *Proc Natl Acad Sci USA* 2002, **99:**3729-3733.
9. Edenberg HJ, Dick DM, Xuei X, Tian H, Almasy L, Bauer LO, Crowe RR, Goate A, Hesselbrock V, Jones K, Kwon J, Li TK, Nurnberger JI Jr, O'Connor SJ, Reich T, Rice J, Schuckit MA, Porjesz B, Foroud T, Begleiter H: **Variations in GABRA2, encoding the alpha 2 subunit of the GABA(A) receptor, are associated with alcohol dependence and with brain oscillations.** *Am J Hum Genet* 2004, **74:**705-714.
10. Jones KA, Porjesz B, Almasy L, Bierut L, Goate A, Wang JC, Dick DM, Hinrichs A, Kwon J, Rice JP, Rohrbaugh J, Stock H, Wu W, Bauer LO, Chorlian DB, Crowe RR, Edenberg HJ, Foroud T, Hesselbrock V, Kuperman S, Nurnberger J Jr, O'Connor SJ, Schuckit MA, Stimus AT, Tischfield JA, 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.
11. Wang JC, Hinrichs AL, Stock H, Budde J, Allen R, Bertelsen S, Kwon JM, Wu W, Dick DM, Rice J, Jones K, Nurnberger JII Jr, Tischfield J, Porjesz B, Edenberg HJ, Hesselbrock V, Crowe R, Schuckit M, Begleiter H, Reich T, Goate AM, Bierut LJ: **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.
12. Dick DM, Edenberg HJ, Xuei X, Goate A, Kuperman S, Schuckit M, Crowe R, Smith TL, Porjesz B, Begleiter H, Foroud T: **Association of GABRG3 with alcohol dependence.** *Alcohol Clin Exp Res* 2004, **28:**4-9.
13. Kennedy GC, Matsuzaki H, Dong S, Liu WM, Huang J, Liu G, Su X, Cao M, Chen W, Zhang J, Liu W, Yang G, Di X, Ryder T, He Z, Surti U, Phillips MS, Boyce-Jacino MT, Fodor SP, Jones KW: **Large-scale genotyping of complex DNA.** *Nat Biotechnol* 2003, **21:**1233-1237.
14. **Affymetrix 10k Mapping Assay Manual** [http://www.affymetrix.com/Auth/support/downloads/manuals/10k_manual.pdf]
15. Matsuzaki H, Loi H, Dong S, Tsai YY, Fang J, Law J, Di X, Liu WM, Yang G, Liu G, Huang J, Kennedy GC, Ryder TB, Marcus GA, Walsh PS, Shriver MD, Puck JM, Jones KW, Mei R: **Parallel genotyping of over 10,000 SNPs using a one-primer assay on a high-density oligonucleotide array.** *Genome Res* 2004, **14:**414-425.
16. Fan JB, Oliphant A, Shen R, Kermani BG, Garcia F, Gunderson KL, Hansen M, Steemers F, Butler SL, Deloukas P, Galver L, Hunt S, McBride C, Bibikova M, Rubano T, Chen J, Wickham E, Doucet D, Chang W, Campbell D, Zhang B, Kruglyak S, Bentley D, Haas J, Rigault P, Zhou L, Stuelpnagel J, Chee MS: **Highly parallel SNP genotyping.** *Cold Spring Harb Symp Quant Biol* 2003, **68:**69-78.
17. Gunderson KL, Kruglyak S, Graige MS, Garcia F, Kermani BG, Zhao C, Che D, Dickinson T, Wickham E, Bierle J, Doucet D, Milewski M, Yang R, Siegmund C, Haas J, Zhou L, Oliphant A, Fan JB, Barnard S, Chee MS: **Decoding randomly ordered DNA arrays.** *Genome Res* 2004, **14:**870-877.
18. Bucholz KK, Cadoret R, Cloninger CR, Dinwiddie SH, Hesselbrock VM, Nurnberger JII Jr, Reich T, Schmidt I, Schuckit MA: **A new, semi-structured psychiatric interview for use in genetic linkage studies: a report on the reliability of the SSAGA.** *J Stud Alcohol* 1994, **55:**149-158.
19. Hesselbrock M, Easton C, Bucholz KK, Schuckit M, Hesselbrock V: **A validity study of the SSAGA – a comparison with the SCAN.** *Addiction* 1999, **94:**1361-1370.
20. American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders (DSM-III-R)* Washington, D.C; 1987.
21. Feighner JP, Robins E, Guze SB, Woodruff RA Jr, Winokur G, Munoz R: **Diagnostic criteria for use in psychiatric research.** *Arch Gen Psychiatry* 1972, **26:**57-63.
22. Begleiter H, Reich T, Hesselbrock V, Porjesz B, Li T-K, Schuckit MA, Edenberg HJ, Rice AP: **The Collaborative Study on the Genetics of Alcoholism.** *Alcohol Health Res World* 1995, **19:**228-236.
23. Porjesz B, Begleiter H, Wang K, Almasy L, Chorlian DB, Stimus AT, Kuperman S, O'Connor SJ, Rohrbaugh J, Bauer LO, Edenberg HJ, Goate A, Rice JP, Reich T: **Linkage and linkage disequilibrium mapping of ERP and EEG phenotypes.** *Biol Psychol* 2002, **61:**229-248.
24. Porjesz B, Jones K, Begleiter H: **The genetics of oscillations in the human brain.** *Adv Clin Neurophysiol* 2004, **57:**437-445.
25. Boehnke M: **Allele frequency estimation from pedigree data.** *Am J Hum Genet* 1991, **48:**22-25.
26. Lander ES, Green P: **Construction of multilocus genetic linkage maps in humans.** *Proc Natl Acad Sci USA* 1987, **84:**2363-2367.
27. **Affymetrix Mapping 10k Array** [http://www.affymetrix.com/products/arrays/specific/10k.affx]
28. Liu WM, Di X, Yang G, Matsuzaki H, Huang J, Mei R, Ryder TB, Webster TA, Dong S, Liu G, Jones KW, Kennedy GC, Kulp D: **Algorithms for large-scale genotyping microarrays.** *Bioinformatics* 2003, **19:**2397-2403.
29. Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, Richardsson B, Sigurdardottir S, Barnard J, Hallbeck B, Masson G, Shlien A, Palsson ST, Frigge ML, Thorgeirsson TE, Gulcher JR, Stefansson K: **A high-resolution recombination map of the human genome.** *Nat Genet* 2002, **31:**241-247.
30. Broman KW, Murray JC, Sheffield VC, White RL, Weber JL: **Comprehensive human genetic maps: individual and sex-specific variation in recombination.** *Am J Hum Genet* 1998, **63:**861-869.
31. John S, Shephard N, Liu G, Zeggini E, Cao M, Chen W, Vasavda N, Mills T, Barton A, Hinks A, Eyre S, Jones KW, Ollier W, Silman A, Gibson N, Worthington J, Kennedy GC: **Whole-genome scan, in a complex disease, using 11,245 single-nucleotide polymorphisms: comparison with microsatellites.** *Am J Hum Genet* 2004, **75:**54-64.
32. Liu G, Loraine AE, Shigeta R, Cline M, Cheng J, Valmeekam V, Sun S, Kulp D, Siani-Rose MA: **NetAffx: Affymetrix probesets and annotations.** *Nucleic Acids Res* 2003, **31:**82-86.
33. Murray SS, Oliphant A, Shen R, McBride C, Steeke RJ, Shannon SG, Rubano T, Kermani BG, Fan JB, Chee MS, Hansen MS: **A highly

informative **SNP linkage panel for human genetic studies.**
Nat Methods 2004, **1:**113-117.

34.   O'Connell JR, Weeks DE: **Pedcheck: A program for identification of genotype incompatibilities in linkage analysis.** *Am J Hum Genet* 1998, **63:**259-266.