Toward understanding the genetics of alcohol drinking through transcriptome meta-analysis


*Waggoner Center for Alcohol and Addiction Research and Sections of †Neurobiology and **Molecular Genetics and Microbiology, University of Texas, Austin, TX 78712; ‡Department of Veterans Affairs Medical Center, Portland Alcohol Research Center, and Department of Behavioral Neuroscience, Oregon Health and Science University, Portland, OR 97239; §University of Colorado Health Sciences Center, Aurora, CO 80045; ¶University of Colorado School of Medicine, Indianapolis, IN 46202; and ††The Scripps Research Institute, La Jolla, CA 92037

Edited by Floyd E. Bloom, The Scripps Research Institute, La Jolla, CA, and approved February 9, 2006 (received for review November 30, 2005)

Much evidence from studies in humans and animals supports the hypothesis that alcohol addiction is a complex disease with both hereditary and environmental influences. Molecular determinants of excessive alcohol consumption are difficult to study in humans. However, several rodent models show a high or low degree of alcohol preference, which provides a unique opportunity to approach the molecular complexities underlying the genetic predisposition to drink alcohol. Microarray analyses of brain gene expression in three selected lines, and six isogenic strains of mice known to differ markedly in voluntary alcohol consumption provided >4.5 million data points for a meta-analysis. A total of 107 arrays were obtained and arranged into six experimental data sets, allowing the identification of 3,800 unique genes significantly and consistently changed between all models of high or low amounts of alcohol consumption. Several functional groups, including mitogen-activated protein kinase signaling and transcription regulation pathways, were found to be significantly overrepresented and may play an important role in establishing a high level of voluntary alcohol drinking in these mouse models. Data from the general meta-analysis was further filtered by a congenic strain microarray set, from which cis-regulated candidate genes for an alcohol preference quantitative trait locus on chromosome 9 were identified: Arhgef12, Carm1, Cryn, Cox5a, Dlat, Fxyd6, Limd1, Nic1n, Nmnat3, Pkn0x2, Rp1d1, Sc5d, Scn4b, Tcf12, Vps11, and Zfp291 and four ESTs. The present study demonstrates the use of (i) a microarray meta-analysis to analyze a behavioral phenotype (in this case, alcohol preference) and (ii) a congenic strain for identification of cis regulation.

alcoholism | gene expression | microarray

An understanding of the molecular mechanisms underlying the genetic propensity toward excessive alcohol consumption is crucial for the development of new treatments for alcoholism. Animal models for alcohol-related traits provide an important opportunity to explore mechanisms responsible for different aspects of the uniquely human disease (1). In particular, selected lines and inbred strains of mice with a divergence in voluntary alcohol drinking represent valuable tools to dissect the genetic components of alcoholism (2, 6, 8, 9, §§). However, each model has different advantages. Inbred strains are homogeneous at all alleles, and genetic (allelic) differences between strains lead to observed phenotypic differences. Several inbred strains have been well characterized for numerous alcohol-related traits, but any two strains will differ for numerous other phenotypes (including gene expression), making identification of the specific genes for a given trait difficult. Selected lines exploit homozgyosity for large-effect and some small-effect allelic variants that contribute to the divergent preference for drinking. Although the resulting genome remains relatively polymorphic, as selection continues, fixation of genes unrelated to the selected phenotype occurs because of drift.

We propose to identify candidate genes that contribute to the genetic propensity to drink by combining whole-brain gene expression databases of genetic mouse models with differences in voluntary alcohol consumption. Most importantly, our approach should allow identification of small- or moderate-effect genes. Three different sets of oppositely selected lines bred for high and low amounts of alcohol drinking, five inbred strains known to differ in voluntary alcohol consumption, and a hybrid strain recently shown to have the greatest degree of voluntary alcohol consumption of any known mouse genotype (4) were used (Table 1). Global measurement of brain gene expression in each of the mouse models allowed us to explore which transcripts are consistently changed in different genetic models of high and low amounts of alcohol intake. It is important to note that the mice used for array analysis were not exposed to alcohol; thus, we are defining the transcriptional signatures of genetic predisposition to high and low levels of alcohol consumption. Expression analysis of an additional mouse line congenic for a section of chromosome 9, which is known to contain genes that regulate alcohol consumption (3), allowed us to define candidate quantitative trait genes (QTBs) for this quantitative trait locus (QTL).

Analysis across data sets for common changes in expression provides more statistical power to detect small and reliable changes than analysis of any one or two of the data sets. Meta-analyses of diverse gene expression data sets were recently used to successfully uncover genes related to carcinogenic phenotypes (10), but a similar approach has not been used for a neurobehavioral trait (for a review of meta-analyses of microarray data, see ref. 11). In this study, we employ a meta-analysis of microarray data from different genetic models of high and low levels of alcohol consumption to define functional pathways and individual genes that may determine a predisposition for a high degree of alcohol intake.

Results and Discussion

Use of Cohen’s d as an Effect-Size Measure Allows Comparison of Microarray Experiments from Diverse Sources. Five experimental microarray (Affymetrix and custom cDNA) data sets of mice genetically divergent in voluntary alcohol consumption were analyzed for differential gene expression with the Cohen’s d statistic (12). The initial meta-analysis comprised 13 individual data sets from three groups of selected lines and one group of six isogenic
Table 1. Mouse models and microarrays used

<table>
<thead>
<tr>
<th>Data set</th>
<th>Study</th>
<th>Platform</th>
<th>Strain/line</th>
<th>Sex</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>STS (B6.D2)</td>
<td>Affymetrix</td>
<td>High STS</td>
<td>F/M</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Low STS</td>
<td>F/M</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>HAP/LAP1</td>
<td>Affymetrix</td>
<td>HAP1</td>
<td>M</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>LAP1</td>
<td>M</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>HAP/LAP2</td>
<td>Affymetrix</td>
<td>HAP2</td>
<td>M</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>LAP2</td>
<td>M</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>Inbred strains</td>
<td>Affymetrix</td>
<td>B6</td>
<td>M</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>BalbC</td>
<td>M</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>D2</td>
<td>M</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>LP/NJ</td>
<td>M</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>Inbred strains</td>
<td>Custom cDNA</td>
<td>B6</td>
<td>F</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>FVB/NJ</td>
<td>F</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>FVB.B6 hybrid</td>
<td>F</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>B6.D2 congenic 9</td>
<td>Custom cDNA</td>
<td>B6</td>
<td>F/M</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>FVB.B6 hybrid</td>
<td>F</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>B6.D2 Congenic 9</td>
<td>F/M</td>
<td>12</td>
</tr>
</tbody>
</table>

Three selected lines (data sets 1–3) and the inbred strains (data sets 4 and 5) were used to form four groups for the meta-analysis. The congenic line (data set 6) was used as a filter to select genes from the meta-analysis. See Materials and Methods for details. STS, short-term selection; F, female; M, male. Data set 1 was contributed by R.J.H., J.C.C., and J.K.B. Data sets 2 and 3 were contributed by B.T. Data set 4 was contributed by R.J.H. and J.K.B. Data sets 5 and 6 were contributed by S.E.B.

strains (Table 1). A data set consisting of C57BL/6J (B6), DBA/2J (D2) congenic 9 animals is used later as a filter for the initial meta-analysis. Two advantages of the Cohen’s d approach are important. First, data from different platforms and laboratories can be combined without the use of normalization. Second, small differences that are consistent in direction of change can be detected, even though these changes might not be detected in any single data set.

After calculating Cohen’s d for each data set (Fig. 1A), a pairwise comparison of common changes was applied to test the compatibility of the four data sets. The number of transcripts regulated in the same direction between each pair of data sets is significantly greater than the number ascribed to chance for five of six pairs. When all four data sets are compared, the number of transcripts regulated in the same direction is nearly 2-fold over the chance level ($P < 0.00001, \chi^2$). The lower similarity seen between selected lines and inbred strains, compared with the pairs of selected lines, likely reflects differences in genetic background, an enrichment of alleles linked to alcohol preference in selected lines, and a relative randomness of such alleles in inbred strains.

An overlap in regulation between mouse models sharing the same alcohol phenotype is readily detectable despite the use of data sets from different animal models and platforms. The distribution of z test $P$ values for the meta-analysis of the four data sets is skewed to the left, indicating a high number of significantly coexpressed genes detected by the meta-analysis (Fig. 1B) and suggesting a low incidence of false positives consistent with the measured $Q$ values. Thus, the use of Cohen’s $d$ in our meta-analysis allows for the detection of transcripts significantly divergent between genetically distinct mouse models displaying opposite levels of alcohol consumption.

Meta-Analysis Identifies Candidate Genes for High and Low Amounts of Alcohol Consumption. Use of Cohen’s $d$ values between the mouse strains and lines displaying high and low levels of alcohol drinking across data sets resulted in the identification of 5,182 significant ($|d| < 0.5$ and $Q < 0.05$) differentially expressed transcripts, representing 3,800 unique genes (data for all transcripts is shown in Table 2, which is published as supporting information on the PNAS web site). The highly significant gene expression changes ($Q < 0.01$) across the four data sets are noticeably consistent when seen in pseudocolor raster display (Fig. 2A). Red and green indicate transcripts overexpressed and underexpressed in animals displaying high degrees of alcohol consumption, respectively. In general, more transcripts appear to be up-regulated in models with high levels of consumption relative to models with low levels of consumption.

The sheer number of significant differences suggests that numerous pathways in the brain may be distinct between mouse models with different levels of alcohol consumption. Indeed, the 75 genes with the largest effect size ($|d|$) and lowest $Q$ values (Fig. 2B) fall into broad categories of cellular homeostasis and neuronal function. Several genes of interest expressed higher in preferring models include: B2m, Man2h1, Scn4b, Mapre1, Prkce, and Sts, which function in immunity/cellular defense, glycosylation, ion channel activity, microtubule binding/dynamics, intracellular signaling cascades, and neuronal signaling, respectively. $\beta$-microglobulin may play a role in neuronal plasticity because B2m knockout mice exhibit abnormal synaptic connections as well as altered patterns of long-term potentiation and long-term depression (13). Scn4b is an auxiliary $\beta$-subunit of voltage-gated sodium channels, which can alter channel function and interaction between the ion channel and other proteins (14). In cerebellar Purkinje cells, Scn4b may be required for high-frequency firing of neurons containing voltage-gated sodium channels (15). A role for Prkce (PKC-\$) in ethanol consumption is supported by the observation that Prkce knockout mice consume less alcohol compared with controls in a two-bottle choice drinking paradigm, and introduction of conditional Prkce expression in these mice was sufficient to restore alcohol drinking levels (16). Gnb1, Hmgn2, and Hyou function in GTPase activity/signal transduction, DNA binding/packaging, and the cellular response to stress, respectively, and are significantly down-regulated in mice preferring alcohol. The overall results of the meta-analysis show the complexity of the alcohol drinking phenotype as indicated by the surprisingly large number of consistent, small increases or decreases in gene expression between the mouse models.

In Silico Promoter and Functional Analyses Uncover Potential Significant Regulation and Pathways, Respectively. Nearly 10% of the genes in the mammalian genome were detected as divergent between the alcohol preference lines and strains of mice. One possible explanation is linkage disequilibrium, the tendency of closely linked genes to be coexpressed. Another explanation is that some of the risk-conferring QTG are transcriptional regulators. Although current mouse in silico promoter analysis tools are not yet

Mulligan et al.
Fig. 2. Visual representation of microarray data. Columns represent microarray databases used in the meta-analysis, and rows represent transcripts. The filtered criterion was $Q < 0.01$, and genes were sorted by effect size. (A) Transcripts (2,697) are listed from positive to negative Cohen’s $d$ values (effect size). Red indicates a positive effect size and higher expression in preferring mice, and green indicates a negative effect size and lower expression in preferring mice. Black indicates an effect size near zero. (B) The 75 unique transcripts with the highest absolute average effect size ($Q < 0.05, |d/| > 1.94$). STS, short-term selection; IS, inbred strains.

comprehensive on a genomic scale, the oPOSSUM database (17) is a valuable tool for identifying overrepresented transcription factor (TF) binding sites (TFBS) in promoter regions near a gene’s transcription start site. oPOSSUM analysis is very conservative and has strict requirements; simulation studies have suggested that it detects few false positives (17).

TFBS for Zfp143 (Staf, selenorysteine RNA gene transcription activating factor) were identified as overrepresented in the up-regulated genes (for high amounts of alcohol drinking). The Staf TFBS was present in the upstream (2,000 bp) promoter regions of 64 genes. Importantly, Zfp143 expression was also found to be significantly up-regulated across the models with high drinking levels. Similarly, the TFBS for the fork-head box TF Foxa2 (HNF-3β, hepatocyte nuclear factor 3β) was detected as overrepresented across the down-regulated gene group and present in 146 genes, and Foxa2 was found to be significantly down-regulated in the meta-analysis. In both cases, the overrepresented TF has the same pattern of expression as its target genes. It is easy to imagine, as evidenced by the number of gene targets for Zfp143 and Foxa2, how small changes in the levels of a TF could cause profound differences in the brain transcriptome and may account for the large number of gene expression changes (3,800 genes) detected in this meta-analysis. Although neither Zfp143 nor Foxa2 were identified as cis-regulated and are not primary candidate genes for alcohol preference, an understanding of how changes in the expression level and/or activity of transcriptional regulators affects the predisposition toward alcohol consumption phenotypes warrants further investigation.

To better understand possible global interactions of the divergent genes, overrepresentation analyses for function were completed. Such analyses apply statistical methods to estimate whether some biological function/pathway is represented in a given data set more than expected by chance. Using the webGASTALT search engine, which queries different databases, including KEGG, BioCarta, and GO (Gene Ontology) (18–20), functional group analysis revealed that kinase and signaling pathways were overrepresented in genes divergent between alcohol-prefering and nonpreferring genotypes. The result supports previous suggestions that the mitogen-activated protein kinase, NF-κB, and IL-6 signaling pathways (and CREBBP) are sensitive to alcohol (21). Known functional interactions (summarized in Fig. 3) show that kinase and signaling pathways are overrepresented in the overrepresented categories and apoptotic/antiapoptotic, neurodegeneration, and neuroplasticity pathways were also functionally overrepresented (for all results, see Fig. 5 and Table 3, which are published as supporting information on the PNAS web site). In silico functional analyses are powerful tools for hypothesis generation, especially for complex phenotypes, such as predisposition toward alcohol drinking. It should now be feasible, for example, to test the effect of candidate genes in Fig. 3 for their impact on alcohol consumption.

Several Genes Identified in the Meta-Analysis Are Located Within QTLs for Human Alcoholism Susceptibility. A two-pronged approach was taken to use known QTLs in both human and rodent research. First, the list of ∼3,800 significant genes was annotated using SOURCE (22) for mouse and human chromosomal location. Two separate lists were created by using Feighner, DSM-III-R, and ICD-10 criteria (23); one list for genes found within the boundaries of mouse QTLs for preference drinking (3) and another for genes within the boundaries of alcoholism susceptibility QTLs in humans. The lists were used as filters to identify overlap of significant genes from the meta-analysis with QTLs between the two species (Table 4, which is published as supporting information on the PNAS web site). Thirty-six genes met criteria as candidate genes with 11 (Gstm1, Gstm2, Gstm5, Il1f5, Il1f6, Il1f8, Il1m, S100a1, S100a10, and S100a13) belonging to only three gene families: GSTM (glutathione S-transferase activity), S100 class calcium-binding, and cytokine/proinflammatory activity (IL-1) (Table 5, which is pub-
lized as supporting information on the PNAS web site). GSTM1 alleles have been shown to be associated with increased alcoholism and alcoholic liver disease (24), and recent results from mice with mutant chemokine receptors suggest a role of inflammatory pathways in alcohol consumption (25). QTL analyses for alcohol preference in rats (26, 27) (not listed in Table 5) show two syntenic QTLs with mouse, and the chromosome 2 QTL is syntenic between all three species, including humans. GST8-8, the implicated QTG within the QTL on rat chromosome 8 (28), is syntenic on mouse chromosome 9. However, Gsta4, the mouse alias for GST8-8, was not significantly divergent in this meta-analysis. Interestingly, several other significantly regulated glutathione S-transferase genes found on mouse chromosome 1 and human chromosome 3 were in syntenic QTL regions for both species (again, see Table 5).

It is important to note that, although there is considerable synteny between the genomes of rodents and man, the degree of evolutionary overlap between polymorphisms or alleles of genes is probably much lower. Thus, translation of data from rodent models to humans is more likely to apply on a pathway level rather than as exact mutations in specific genes. The species differences in glutathione S-transferase may, in fact, be such an example.

**Use of a B6.D2 Congenic 9 Data Set Filter Identifies cis-Regulated Genes Occupying a QTL for Strong Alcohol Preference in Mice.** QTL mapping and construction of congenic mice based on such maps provides another genetic approach to defining changes in gene expression linked to propensity for high levels of alcohol consumption. Several mouse QTL experiments show that chromosome 9 contains genes that contribute to the genetic propensity toward increased alcohol intake. This QTL was found to be the largest of several identified in B6- and D2-derived populations (3). Therefore, B6 congenic mice containing a region that captured the D2 alleles between 9 and 58 cm on chromosome 9 were analyzed for gene expression and compared against the B6 background strain. The differentially expressed genes located on chromosome 9 are by default cis-regulated and thus are QTG candidates.

To identify cis-acting candidate genes for the chromosome 9 alcohol preference QTL, the current meta-analysis results were filtered with the congenic data set. Significant genes detected by the meta-analysis that passed the congenic filter were mapped on chromosome 9 along with their relative effect size (Fig. 4). These genes include 16 known genes (Arhgef12, Carm1, Cryab, Cox5a, Dlat, Fxydb, Limd1, Ncn1, Nmnnat3, Pknos2, Rbp1, Sc5d, Scn4b, Tcf12, Vps11, and Zfp291) and four expressed sequence tags (2810423O19Rik, 1110032A03Rik, 5730439E10Rik, and 9030425E11Rik). Four of these (Pknos2, Scn4b, 9030425E11Rik, and 1110032A03Rik) are correlated with alcohol preference in a BXD recombinant inbred population (5). Interestingly, some primary candidate genes also have putative transcriptional activity: Carm1, Pknos2, Tcf12, Zfp29, and Limd1. Although their targets are not yet characterized, each may explain some of the consistently divergent gene expression observed in the meta-analysis. The 20 genes identified with the congenic filter are potential QTGs underlying alcohol preference and provide exciting avenues of investigation into the complex behavior of alcohol preference.

**Conclusions**

Predisposition to excessive alcohol consumption is likely a key aspect of human alcoholism, but molecular determinants of this trait are difficult to study in humans. Our work provides a microarray-based meta-analysis of a behavioral phenotype and uses effect-size measures. The Cohen’s d metric eliminated the need for standardization across array platforms and experiments, allowing identification of consistently different transcripts between alcohol-preferring and nonpreferring genotypes, even if the transcriptional differences were small with respect to magnitude or significance or both.

The meta-analysis provided 5,182 targets, including 3,800 nonoverlapping genes for further study. The significant differences between the genotypes for divergent amounts of alcohol drinking represent numerous functional groups covering a large range of cellular pathways as well as many genes (~25%) whose functions are not yet characterized. Thus, some genes whose functions are currently unknown are also likely to be important determinants of alcohol consumption. The 75 genes with the largest effect size (Q < 0.05, |d| > 1.94) (see Fig. 2) fall into the broad categories of cellular homeostasis and neuronal function. Differences in the ability to
Importantly, the opportunities to reveal mechanisms through large-scale genomic screening approaches provides specific candidate genes for one QTL for alcohol preference. These genes and most of the key functional groups revealed by the current study indicate our ignorance about the molecular mechanisms driving alcohol consumption and, most importantly, the opportunities to reveal mechanisms through large-scale genomic screening approaches.

Materials and Methods

Animal Husbandry. Naïve, adult mice (60–100 d old) had 24 h ad libitum access to rodent chow, water, and 12 h:12 h lighting. All animals were housed and treated according to the National Institutes of Health guidelines for the use and care of laboratory animals (36) and approved Institutional Animal Care and Use Committee protocols at each respective institution. Whole-brain total RNA from naïve mice was used for all array analyses. Microarray hybridizations for the HAP/LAP lines, inbred strains, and congenic 9 strains were completed on individual mice. The lines for high and low levels of short-term selection were hybridized with a pool of three samples for each replicated N value (n = 10 mice per line). Details for each strain are listed below.

Short-Term Selection Lines. Reciprocally crossed, individually housed B6×D2 and D2×B6 F1 mice (8–10 weeks of age) were initially given water in a classical two-bottle choice paradigm. One tube was then replaced with a 3% (vol/vol) solution of ethanol in tap water followed by 10% ethanol (4 d each). The selection index for drinking was based on the 10% concentration expressed in grams per kilogram per day for each of the last days immediately before a bottle position change (i.e., the average of days 8 and 10). The 10 highest-scoring and 10 lowest-scoring mice of each gender were used as breeding pairs to initiate the selection lines with high and low amounts of drinking, respectively. Bidirectional selection continued similarly for each of three generations; the highest of the high line and the lowest of the low line were used to breed the next generation of the high and low selection lines, respectively. Naïve mice with high and low levels of drinking from selection generation S3 were used in the array experiment.

Inbred Strains. Male BALB/cJ, B6, D2, and LP/NJ inbred mice were raised at the Veterans Affairs Medical Center in Portland, OR. LP/NJ mice were tested in an accelerating concentration consumption (6, §§). Male HAP1/LAP1 and HAP2/LAP2 mice (70–90 d old) were used from generations 26 and 19, respectively, for Affymetrix array analysis.

HAP/LAP. Two replicate lines of HAP and LAP mice were originally generated from HS/Ibg mice (2) based on bidirectional selection outcome of two-bottle choice 10% alcohol solution vs. water consumption (6, §§). Male HAP1/LAP1 and HAP2/LAP2 mice (70–90 d old) were used from generations 26 and 19, respectively, for Affymetrix array analysis.

Congenic Strain. A D2 QTL region for alcohol preference from 9 to 58 cM on chromosome 9 was introgressed onto a B6 genomic background through ≈12 generations of breeder selection by Massachusetts Institute of Technology microsatellite marker genotyping and backcrossing to B6. The B6.D2309 mouse drink significantly less than B6 control mice (difference ≈25%, T.J.P., unpublished data). mRNA was isolated from jointly housed and age-matched (70–100 days) congenic and B6 mice.

Oligonucleotide Microarrays. Samples containing at least 10 μg of total RNA were hybridized after manufacturer’s protocol (Affymetrix, Santa Clara, CA). Details for HAP/LAP mice have been published in ref. 29. RNA from STS mice and mice from inbred...
strains BALB/cJ, B6, D2, and LP/NJ arrays were hybridized to Affymetrix 430 A and B chips and analyzed at the Oregon Health and Science University Gene Microarray Shared Resource Facility (for more details, see Supporting Materials and Methods, which is published as supporting information on the PNAS web site).

Custom cDNA Microarrays. PCR products of clones from several cDNA libraries (see Supporting Materials and Methods for details) were printed in the University of Texas array core facility on poly-L-lysine-coated (Sigma) microscope slides (Erie Scientific, Portsmouth, NH) using a custom-built robotic assembler as described in ref. 37. Microarray hybridizations were performed with the Array 350 microarray labeling kit according to the manufacturer’s protocol (Genisphere, Hatfield, PA).

Meta-analysis. Five data sets identified 1–5 in Table 1 were used for meta-analysis. Alcohol-prefering and nonpreferring mice were first compared by using a parametric statistical test. Student’s t test for independent samples was used to compare high and low selected lines (studies 1–3). Both sets of inbred strains (studies 4 and 5) were analyzed independently by an F test using single contrasts (see Supporting Materials and Methods for additional details). The resulting t and F values for each transcript were then used to estimate effect size by calculating a Cohen’s d statistic, where the effect size difference between the strains expressed in pooled within-group data (12). The resulting formulas were applied: d = 2/SD = 2V(F/df), where df represents the degrees of freedom. The direction of change was coded in the resulting Cohen’s d values so that (positive) values indicated an up-regulation and (negative) values indicated a down-regulation of transcripts in alcohol-prefering genotypes compared with nonpreferring genotypes.

Data from the two inbred strain studies (studies 4 and 5 in Table 1) were collapsed by averaging Cohen’s d values for each transcript (for rationale, see Supporting Materials and Methods). To minimize any outlier effects, extremely deviant effect sizes were adjusted to a maximum absolute value of |d| = 4. Finally, the Cohen’s d values generated for the three pairs of selected lines and the combined set of inbred strains were averaged, and a z test was used to test the significance of deviation of the mean effect size from zero. QVALUE software was used to estimate the false discovery rate (Q) for the meta-analysis results (38, 39), which estimates the proportion of all declared significant results that are expected to be false positives.

Cochglen 9 Data Filter. We used microarray data obtained from the B6.D2 congenic strain to filter transcripts detected by meta-analysis as regulated significantly between ethanol-prefering and nonpreferring genotypes and located within the introgressed region containing known ethanol preference QTLs on mouse chromosome 9. Specifically, chromosome 9 transcripts significantly regulated between the congenic and control lines (P < 0.05, t test) were identified and matched with data from the meta-analysis. Only genes detected by both approaches were selected as putative candidates for cis-regulation of alcohol preference.

TFBS Overrepresentation Analysis. The oPOSSUM database (www.cisreg.ca/cgi-bin/oPOSSUM/opossum) (17) was used to analyze TFBS in genes that were significantly (Q < 0.05 and |d| ≥ 0.5) up- or down-regulated in alcohol-prefering vs. nonpreferring mice. The two lists had 2,388 and 1,580 genes with 1,011 and 546 recognized as orthologs in oPOSSUM for up- or down-regulation respectively. TFBS overrepresentation was determined based on one-tailed Fisher exact probabilities and the ranking of z scores. After the initial analysis, the genes for the TFs themselves were individually checked to see whether (i) they were present on the arrays and (ii) transcription differences were consistent with the promoter TFBS analysis.

Functional Overrepresentation Analysis. Transcripts passing statistical thresholds (Q < 0.05, |d| ≥ 0.5) were functionally annotated with WEBGASTALT (web-based gene set analysis toolkit, http://genereg.ornl.gov.webgestalt) (19). A hypergeometric test was used to detect functional groups overrepresented in a selected gene set (meta-regulated genes) compared with a reference set (all genes in meta-analysis). See Supporting Materials and Methods for additional details.

This study was a collaborative effort by investigators affiliated with the Integrative Neuroscience Initiative on Alcoholism Consortium and was supported by National Institute on Alcohol Abuse and Alcoholism Grants AA013182 and AA013404 (to S.E.B.); AA013475, AA006243, and AA010760 (to J.K.B.); AA013520 (to Y.A.B.); AA013519 (to J.C.C.); AA013483 (to N.J.G.); AA013518 (to R.A.H.); AA013484 and AA11034 (to R.J.H.); AA13517 (to G.F.K.); and AA013489 (to B.T.); and by five Merit Review Programs from the Department of Veterans Affairs (to J.K.B., J.C.C., D.A.F., R.J.H., and T.J.P.).