

Alcohol Sensitivity in *Drosophila*: Translational Potential of Systems Genetics

Tatiana V. Morozova,^{*,†,1} Julien F. Ayroles,^{†,‡,1} Katherine W. Jordan,^{†,‡} Laura H. Duncan,^{†,‡}
Mary Anna Carbone,^{†,‡} Richard F. Lyman,^{†,‡} Eric A. Stone,^{†,§} Diddahally R. Govindaraju,^{**}
R. Curtis Ellison,^{††} Trudy F. C. Mackay^{†,‡} and Robert R. H. Anholt^{*,†,‡,2}

^{*}Department of Biology, [†]W. M. Keck Center for Behavioral Biology, [‡]Department of Genetics, and [§]Department of Statistics, North Carolina State University, Raleigh, North Carolina 27695, ^{**}Department of Neurology and ^{††}Institute on Lifestyle and Health, Boston University School of Medicine, Boston, Massachusetts 02118

Manuscript received July 17, 2009
Accepted for publication July 30, 2009

ABSTRACT

Identification of risk alleles for human behavioral disorders through genomewide association studies (GWAS) has been hampered by a daunting multiple testing problem. This problem can be circumvented for some phenotypes by combining genomewide studies in model organisms with subsequent candidate gene association analyses in human populations. Here, we characterized genetic networks that underlie the response to ethanol exposure in *Drosophila melanogaster* by measuring ethanol knockdown time in 40 wild-derived inbred *Drosophila* lines. We associated phenotypic variation in ethanol responses with genomewide variation in gene expression and identified modules of correlated transcripts associated with a first and second exposure to ethanol vapors as well as the induction of tolerance. We validated the computational networks and assessed their robustness by transposon-mediated disruption of focal genes within modules in a laboratory inbred strain, followed by measurements of transcript abundance of connected genes within the module. Many genes within the modules have human orthologs, which provides a stepping stone for the identification of candidate genes associated with alcohol drinking behavior in human populations. We demonstrated the potential of this translational approach by identifying seven intronic single nucleotide polymorphisms of the *Malic Enzyme 1 (ME1)* gene that are associated with cocktail drinking in 1687 individuals of the Framingham Offspring cohort, implicating that variation in levels of cytoplasmic malic enzyme may contribute to variation in alcohol consumption.

IT is generally recognized that studies on genetically tractable model organisms can provide information that is relevant for human disorders (BILEN and BONINI 2005; MACKAY and ANHOLT 2006; CRABBE 2008; FLINN *et al.* 2008; WOODRUFF-PAK 2008). Nonetheless, attempts to identify risk alleles for human diseases have relied almost exclusively on linkage studies in human families or association analyses in human populations without reliance on translational approaches based on model systems. Genomewide association studies (GWAS) have become increasingly popular, but, despite some impressive successes, have fallen short of expectations. Especially, GWAS performed to date for behavioral and neuropsychiatric disorders have yielded few and often nonreproducible results (NICA and DERMITZAKIS 2008; O'DONOVAN *et al.* 2008; FEULNER *et al.* 2009; JOHNSON and O'DONNELL 2009; MOSKVINA *et al.* 2009; PANKRATZ

et al. 2009; PSYCHIATRIC GWAS CONSORTIUM STEERING COMMITTEE 2009).

There are many, nonmutually exclusive reasons for the difficulty in identifying risk alleles for human behavioral traits. It is often difficult to precisely quantify a heterogeneous spectrum of phenotypes often encountered in human behavioral phenotypes and neuropsychiatric disorders. Ethnic stratification in nonhomogeneous populations can lead to spurious associations. Gene-by-gene and gene-by-environment interactions can be confounding factors. Association study designs have low power to detect low frequency alleles with moderate effects, even in samples of several thousand individuals. The power of single nucleotide polymorphisms (SNPs) to detect causal variants is reduced further if the true causal variant is not in perfect linkage disequilibrium with the tagging SNP. Furthermore, failure of GWAS to identify alleles with large effects has generated the notion that the underlying basis of neuropsychiatric disorders may be determined by many genes of small effects, and thus many studies have been seriously underpowered. Finally, human association studies typically implicate regions containing many genes, rather than individual genes or the causal variant; and the significance of those genes that have been identified in terms of their

Supporting information is available online at <http://www.genetics.org/cgi/content/full/genetics.109.107490/DC1>.

¹These authors contributed equally to this work.

²Corresponding author: Department of Biology, North Carolina State University, Campus Box 7617, Raleigh, NC 27695.
E-mail: anholt@ncsu.edu

mechanistic contributions to the phenotype is often unclear. It is now generally recognized both from studies in model organisms and people that complex traits cannot be understood in terms of single genes, but must be investigated at the level of genetic networks (JOOPER *et al.* 2002; SIEBERTS and SCHADT 2007; CHEN *et al.* 2008; EMILSSON *et al.* 2008; AYROLES *et al.* 2009; BUCHANAN *et al.* 2009; COOKSON *et al.* 2009).

One untested strategy for improving the power to detect alleles affecting behavioral traits in human association studies is to identify networks of genes affecting a homologous behavior in a genetically tractable model organism and to perform association tests on only human orthologs of these genes. This strategy ameliorates the statistical penalty for multiple tests in an unbiased genome scan and enables tailoring the choice of molecular markers used in the association tests to the fine structure of linkage disequilibrium at each candidate gene.

Here, we present a proof of principle of this approach by using *Drosophila melanogaster* as a model system for the ultimate identification of polymorphisms in human genes that are associated with alcohol drinking behavior. Although addiction is difficult to model in animal systems, *D. melanogaster* presents an excellent system for genetic studies of alcohol sensitivity and tolerance (MOORE *et al.* 1998; SCHOLZ *et al.* 2000, 2005; SINGH and HEBERLEIN 2000; CHENG *et al.* 2001; ROTHENFLUH and HEBERLEIN 2002; WOLF *et al.* 2002; GUARNIERI and HEBERLEIN 2003; MOROZOVA *et al.* 2006, 2007; ROTHENFLUH *et al.* 2006). Most previous studies on alcohol sensitivity in *Drosophila* have characterized the effects of induced mutations. In human populations, we are interested in the effects of naturally occurring genetic polymorphisms on phenotypic variation. Therefore, we decided to study alcohol sensitivity and tolerance in a panel of *D. melanogaster* inbred lines derived from a natural population (AYROLES *et al.* 2009). Here, individuals within each line are genetically identical, but the phenotypic variation observed in nature is preserved among the lines. We adopted a two-step approach. First, we characterized genotype-phenotype relationships in these lines by identifying and validating networks of correlated transcripts associated with acute sensitivity to alcohol exposure or tolerance. Next, we identified human orthologs in these networks as candidate genes for alcohol sensitivity in human populations. On the basis of information derived from the *Drosophila* model, we selected *Malic enzyme* (*Men*) as a focus for our proof-of-principle translational study. *Men* represents a central metabolic switch that links the glycolytic pathway to the tricarboxylic acid cycle and generates NADPH, an essential cofactor for fatty acid biosynthesis, while converting malate into pyruvate. *Men* is associated with alcohol resistance in flies (MOROZOVA *et al.* 2006) and in humans is likely to represent a critical metabolic juncture that enables development of fatty liver syndrome in heavy drinkers. Here, we demonstrate that

our translational approach indeed has led to the discovery of SNPS in *ME1* that are associated with drinking behavior in a human population, with an effect size that could not have been resolved with large-scale unbiased GWAS.

MATERIALS AND METHODS

Drosophila stocks: The 40 inbred lines were derived by 20 generations of full-sib mating from isofemale lines that were collected from the Raleigh, NC farmer's market in 2003 (AYROLES *et al.* 2009). Homozygous *P*-element insertion lines containing *P*[*GTL*]-elements in or near candidate genes in a co-isogenic Canton-S (*A*, *B*, *F*) background were generated as part of the Berkeley *Drosophila* Genome Project (BELLEN *et al.* 2004). Fly stocks were reared under standard culture conditions on cornmeal-molasses-agar medium at 25°, 60–75% relative humidity, 12 hr light-dark cycle. Flies were not exposed to CO₂ anesthesia for at least 24 hr prior to the assay.

Quantitative assay for alcohol sensitivity and tolerance: We assessed ethanol sensitivity and tolerance for the 40 wild-derived inbred lines in blocks of approximately eight lines and a control line (Canton-S *B*). Each block was tested over a 2-week period. There were two replicate measurements of each sex ($N = 70$ per replicate) per line; the replicates for each line were assessed on different days. We placed the flies in an inebriometer (WEBER 1988) pre-equilibrated with ethanol vapor, and collected them at 1-min intervals as they eluted. We recorded elution times from the initial exposure to ethanol (*E1*) and a second exposure of the same flies (*E2*) 2 hr later. The mean elution time (MET) is a measure of alcohol sensitivity, and the scaled difference of MET between the second and first exposures is a measure of tolerance, *i.e.*, $(E2 - E1) / 0.5 (E1 + E2)$. Sensitivity from a single exposure to ethanol was similarly tested for 11 *P*-element insertional mutations in candidate genes and their co-isogenic control line, but only for males, with 4–5 replicates/line, $N = 60$ –70/replicate.

Whole genome expression analysis: The gene expression analysis has been described previously (AYROLES *et al.* 2009). Briefly, RNA was extracted from two independent pools of 25 3- to 5-day-old flies/sex/line that were frozen at the same time of day, labeled, and hybridized to Affymetrix *Drosophila* 2.0 arrays, using a strictly randomized experimental design. The raw array data were normalized using a median standardization. The measure of expression was the median log₂ signal intensity of the probes in the perfect match probe sets, after removing probes containing single feature polymorphisms (SFPs) between the wild-derived lines and the reference strain sequence used to design the array. Negative control probes were used to estimate the level of background intensity; probe sets with expression levels below this threshold were considered to be not expressed.

Quantitative genetic and statistical analysis: We expressed individual elution times of the 40 wild-derived inbred lines as deviations from the control mean for the appropriate block and sex. We used mixed model factorial analysis of variance (ANOVA) to partition variance in ethanol sensitivity after the first and second exposures among the inbred lines, according to the model $Y = \mu + S + L + S \times L + Rep(L \times S) + W$, where μ is the overall mean, *S* is the fixed main effect of sex, *L* is the random main effect of line, $S \times L$ is the random effect of the sex by line interaction, $Rep(L \times S)$ is the random effect of replicate, and *W* is the within-replicate variance. Parentheses indicate nested effects. Similarly, we analyzed the data across both exposures by introducing a third cross-classified effect,

exposure (E) into the ANOVA model: $Y = \mu + S + E + L + S \times E + S \times L + E \times L + S \times E \times L + Rep(S \times E \times L) + W$. Significance of the $E \times L$ term indicates genetic variation in the difference in MET between E1 and E2; *i.e.*, genetic variation in the induction of tolerance. A false discovery rate (FDR) for the L term of $FDR < 0.001$ was used in the analysis of natural variation in gene expression to account for multiple tests (HOCHBERG and BENJAMINI 1990). The total genotypic variance among lines was estimated as $\sigma_G^2 = \sigma_L^2 + \sigma_{LS}^2$, where σ_L^2 is the among-line variance component and σ_{LS}^2 is the variance attributable to the $L \times S$ interaction. The total phenotypic variance was estimated as $\sigma_P^2 = \sigma_G^2 + \sigma_E^2$, where σ_E^2 is the environmental variance component. We estimated broad sense heritabilities as $H^2 = \sigma_G^2/\sigma_P^2$. We estimated cross-trait genetic correlations as $r_G = cov_{ij}/\sigma_i\sigma_j$, where cov_{ij} is the covariance of line means between trait i and trait j , and σ_i and σ_j are the square roots of the among-line variance components for the two traits. ANOVA models ($Y = \mu + L + Rep(L) + \varepsilon$) were used to assess differences in ethanol sensitivity between P -element insertion lines and their co-isogenic control. Linear regressions and ANOVA models were used to identify quantitative trait transcripts (QTTs) and SFPs significantly associated ($P < 0.01$) with variation in ethanol sensitivity and tolerance across the 40 lines, as described previously (AYROLES *et al.* 2009). We used modulated modularity clustering (STONE and AYROLES 2009) to derive modules of genetically correlated transcripts associated with ethanol phenotypes. Briefly, we compute the correlation between all pairs of transcripts that vary significantly among the lines and transform them to define the edge weights in a graph of genes. Modulated modularity clustering seeks modules of tightly intercorrelated genes by identifying the graph partition that maximizes the modulated modularity function defined in STONE and AYROLES (2009). Transcripts with spurious association to a phenotype are unlikely to correlate with biologically relevant transcripts after removing the source of the association; conversely, transcripts under coordinated control are likely to exhibit correlated abundance patterns even after removing the effect of their common relationship to a phenotype.

Functional gene annotations are based on FlyBase (<http://flybase.bio.indiana.edu>); DRYSDALE and CROSBY (2005) and AFFYMETRIX, NETAFFX ANALYSIS CENTER (2008 <http://www.affymetrix.com>) compilations, using FlyBase release version 5.2. Gene Ontology enrichment analysis used the DAVID program (DENNIS *et al.* 2003), and tissue-specific gene expression data were obtained from FlyAtlas (CHINTAPALLI *et al.* 2007).

Quantitative RT-PCR and enzyme assays: We quantified mRNA levels by quantitative RT-PCR with the SYBR Green detection method (SYBR GREEN PCR master mix, Applied Biosystems/ABI, Foster City, CA) according to the protocol from ABI, using the ABI PRISM 7900 Sequence Detection System (ABI). We used the *glyceraldehyde-3-phosphate dehydrogenase* gene for the internal standard. Five independent replicates of total RNA were isolated from the Canton-S B control and the *pipsqueak* (*psq*) and *Men P*-element insertion lines using the Trizol reagent (GIBCO-BRL, Gaithersburg, MD) and cDNA was generated from 350 ng of total RNA by reverse transcription using the High Capacity cDNA Reverse Transcription kit (ABI). Primer Express software (ABI) was used to design transcript-specific primers to amplify up to 100-bp regions of all genes of interest. Primers were designed to encompass common regions of alternative transcripts. Negative controls without reverse transcriptase were used for all genes to exclude potential genomic DNA contamination. Samples in each run were normalized relative to a control sample (using $2^{-\Delta\Delta Ct}$ values, according to ABI User Bulletin no. 2) (APPLIED BIOSYSTEMS 2001). Statistical analyses for differences in gene expression levels between P -element insertion lines and the

control line were determined by two-tailed Student's t -tests on ΔCt values.

To examine developmental stage-specific gene expression levels in the *psq* or *Men P*-element insertion lines, relative levels of expression were analyzed in the same way as above after extraction of triplicate RNA samples from embryos between 5 and 8 hr after oviposition, third instar larvae, pupae, whole adult flies, and adult heads and bodies, separately.

Malic enzyme activity was assayed in triplicate for males and females separately by measuring the rate of generation of NADPH at 340 nm, as described by MERRITT *et al.* (2009), using a 96-well plate spectrophotometer (Power Wave X, BioTek Instruments, Winooski, VT) in which absorbance was measured every 11 sec over 3 min. Samples were assayed twice and the average of each measurement was used for analysis. Enzyme activities were standardized by soluble protein and expressed as micromoles of NADP⁺ reduced per minute per microgram soluble protein $\times 10,000$. The relationship between Malic enzyme activity after a single exposure to ethanol and phenotypic variation in second MET was assessed using the linear model: $Y = C + S + L(C) + C \times S + S \times C + S \times L(C) + \varepsilon$, where Y is the Malic enzyme activity, C is the phenotypic class (short or long MET after a second exposure to ethanol), S is the effect of sex, $L(C)$ is the effect of line nested within phenotypic class, and ε is the residual error term.

Study subjects and ascertainment of alcohol consumption:

The Framingham Heart Study is a population-based cohort study started in 1948 in Framingham, Massachusetts. The original cohort included 5209 participants. In 1971, children of the original cohort and their spouses were invited to participate in a prospective study called the Offspring Study. Detailed descriptions of the Framingham Heart Study have been published (DAWBER *et al.* 1952; KANNEL *et al.* 1979). Information on alcohol consumption has been collected repeatedly from both the original and offspring cohorts. For the present study we used only unrelated individuals of the offspring cohort. At all examinations (*i.e.*, every 4 years), subjects were asked how many 1.5-oz cocktails, 12-oz glasses (or cans) of beer, and 5-oz glasses of wine they consumed in a week. Details on alcohol assessment in the Framingham Heart Study have been described previously (DJOUSSE *et al.* 2002).

SNP genotyping and association analysis: For genotyping the human *ME1* gene (chr 6: 83,976,829–84,197,498), genotype data spanning the coding region were downloaded from the International HapMap Project (GABRIEL *et al.* 2002; FRAZER *et al.* 2007). Data from the Centre d'Etude du Polymorphisme Humain (CEPH) sample (Utah residents with northern and western European ancestry) were used. The program Tagger was used to select an optimal set of available SNPs for genotyping, at an r^2 threshold of 0.9 (DE BAKKER *et al.* 2005). We excluded SNPs with a minor allele frequency of < 0.2 since we expected the alleles that contribute to alcohol-related phenotypes in this data set to be common, and we chose one tagging SNP from each LD block. To assess whether LD structure in our population might be different from the CEPH sample, we chose two SNPs in the same LD block (rs13215578, rs1170347). Selected SNPs were scored by the Illumina Customer Support Center for expected performance on the Illumina platform. A total of 24 SNPs in the human *ME1* gene were genotyped (see supporting information, Table S1) using the Illumina Golden Gate genotyping platform in samples from 1717 unrelated subjects from the offspring cohort of the Framingham Heart Study, with 250 ng of each sample DNA per well and concentrations adjusted to 50 ng/ μ l. Genotype calls were made using the BeadStudio Genotyping Module v3.2 (Illumina, San Diego, CA) software package. Three SNPs did not cluster well and were not called (rs3798886, rs1145908, and rs1180186). The call rates for the remaining 21

SNPs were within 99.7–100% confidence with 99.99% concordance among the replicates. The call rates for 30 individuals were lower than 98.9% and they were discarded from the analysis.

We used a repeated measures ANOVA over age to identify SNPs associated with variation in drinking behavior. We fitted the following full factorial model using PROC MIXED: $Y = \mu + A + G + S + G \times S + A \times G + A \times S + A \times G \times S + \varepsilon$, where μ represents the grand mean, A is the effect of age, used as a continuous covariate, G is the genotypic effect, S is the sex effect, and ε is the error. A first-order autoregressive covariance structure for the error was selected on the basis of a likelihood ratio test averaged across SNPs. We adjusted for multiple testing using 1000 permutations. For each permutation we recorded the minimal P -value and the 95th and 99th percentile of the resulting distribution.

RESULTS

Variation in alcohol sensitivity and tolerance among 40 wild-derived inbred *Drosophila* lines: Ethanol sensitivity in *Drosophila* can be quantified by exposing groups of flies of the same genotype to ethanol vapors in an “inebriometer,” a 4-foot-long glass tube with slanted mesh partitions to which flies can adhere (WEBER 1988). As flies lose postural control they fall through the inebriometer, from where they are collected at 1-min intervals. The MET is a quantitative measure of ethanol sensitivity. A second exposure 2 hr later extends the MET and reflects the development of tolerance (SCHOLZ *et al.* 2000; MOROZOVA *et al.* 2006). Thus, behavioral responses of flies to alcohol show similarities to the effects of alcohol on people.

We measured METs of 40 wild-derived inbred lines of *D. melanogaster* and found substantial genetic variation in sensitivity following one and two exposures to ethanol, with broad sense heritabilities of 0.244 and 0.216, respectively (Figure 1, A and B; Table S2). Indeed, the range of variation among the inbred lines is comparable to that achieved after 25 generations of selection (MOROZOVA *et al.* 2007). Although females are on average more sensitive than males, there is no genetic variance in sex dimorphism for alcohol sensitivity (Table S2). The MET averaged over all lines was greater after the second than the first exposure, indicating development of tolerance (Table S2). The genetic correlation between the METs after one and two exposures was high ($r_G = 0.84$, $P < 0.0001$), but significantly different from unity (Figure 1B; Table S2), indicating significant genetic variation among the lines in the magnitude of induction of tolerance. The significant line by exposure interaction was entirely attributable to changes in rank order of the MET among the lines in the two exposures (Figure S1). We quantified tolerance (T) as the standardized shift in MET between the first (E1) and second (E2) exposures, scaled by their average (*i.e.*, $T = (E2 - E1)/0.5(E1 + E2)$; Figure 1C). Tolerance has a negative genetic correlation with the initial MET (Figure 1C; $r_G =$

-0.57 ; $P < 0.001$), indicating that tolerance is proportionately greater for sensitive than resistant lines.

Transcriptional networks for alcohol sensitivity and tolerance in *Drosophila*: A previous study quantified genetic variation for 10,096 transcripts and 3136 SFPs among these lines (AYROLES *et al.* 2009). We identified candidate genes affecting alcohol sensitivity and tolerance by regressing phenotypic values on transcript abundance and assessing differences in mean between SFP classes. We identified 1133 transcripts associated with one or more alcohol phenotypes with a nominal P -value of 0.01 (Table S3). This threshold corresponds to a 0.27 false discovery rate, a liberal criterion chosen deliberately for subsequent clustering analysis. We found 295, 592, and 410 candidate genes associated with METs from the first and second exposures and tolerance, respectively (Table S4). In addition, we identified 432 SFPs for a total of 1455 candidate genes. Consistent with the genetic correlations among alcohol phenotypes, there are more transcripts than expected by chance in common between the first and second exposures (105, $\chi^2_1 = 481$, $P < 0.0001$) and the first exposure and tolerance (42, $\chi^2_1 = 78.1$, $P < 0.0001$) but not between tolerance and the second exposure (19, $\chi^2_1 = 0.95$, $P = 0.33$) (Figure 2A). Thus, the genomic signatures for alcohol sensitivity shift dramatically as flies are subjected to repeated exposure to alcohol.

The transcriptome is highly genetically intercorrelated in these lines (AYROLES *et al.* 2009), enabling us to group transcripts significantly associated with each trait into genetically correlated modules (STONE and AYROLES 2009). This unbiased, self-organizing procedure generated eight transcriptional modules associated with variation in MET upon the initial exposure (Figure 2B), six associated with variation in MET after the second exposure (Figure 2C), and five associated with tolerance (Figure 2D; Table S4).

We analyzed gene ontologies, tissue-specific expression patterns, and shared transcription factor binding motifs to interpret the functions of the modules. Genetic variation in initial sensitivity to ethanol exposure is associated with transcriptional variation in genes with a wide range of functions, including defense responses to environmental chemicals and maintenance of cellular integrity through transcriptional regulation, intracellular protein synthesis, and trafficking (Table S5). Gene ontology categories comprised correlated transcripts encoding glutathione transferases (module 5), epithelial fluid transport and voltage-gated calcium channel activity (module 7), and polysaccharide metabolism (module 8) (Table S5). Genetic variation for MET following the second exposure to ethanol is associated with correlated transcripts affecting protein localization and transport (module 3) and mitochondrial protein synthesis, including mitochondrial ribosomal proteins (module 4). Genes implicated in nervous system function do not stand out among the transcripts associated

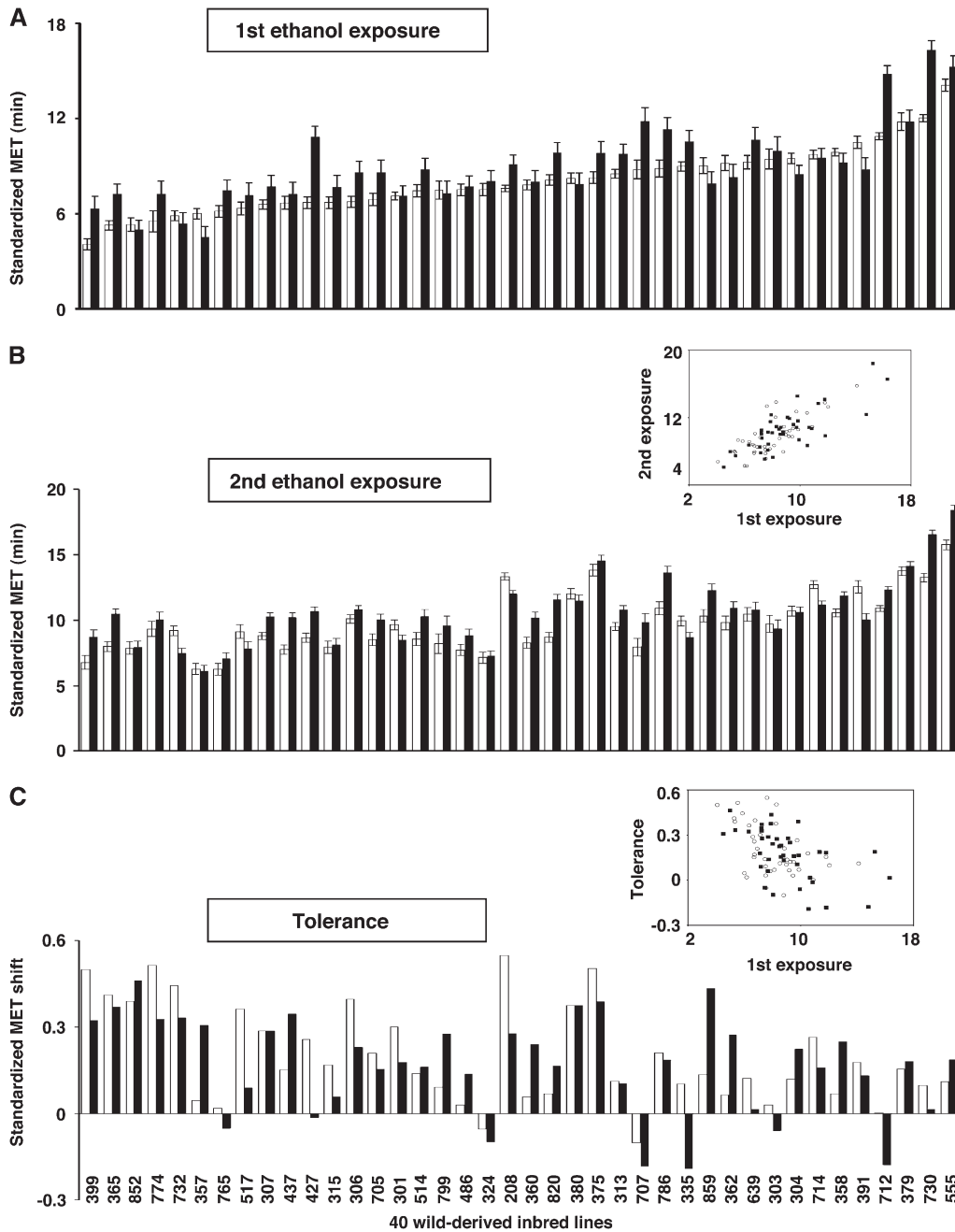


FIGURE 1.—Variation in MET among the 40 wild-derived inbred lines. (A) Variation in response to initial alcohol exposure. Data are the deviation of the MET of each line from a *Canton S* (B) standard \pm SE. (B) Variation in response to the second alcohol exposure. Data are the deviation of the MET of each line from a *Canton S* (B) standard \pm SE. (C) Variation in tolerance. Open bars represent females; solid bars males. The insets in panels B and C depict, respectively, the positive correlation between the METs from the first and second exposures, and the negative correlation between the sensitivity to the initial alcohol exposure and induction of tolerance. Open circles and the solid line represent females; males are shown as solid squares with a dashed line.

with genetic variation in MET following one or two exposures to alcohol. However, natural variation in tolerance is associated with correlated transcripts affecting nervous system function (modules 2 and 4; Table S5). Genetic variation in tolerance is also associated with variation in correlated transcripts affecting oxidative phosphorylation (module 2) and metabolism (modules 4 and 5; Table S5).

Consistent with the wide range of functional annotations of transcripts associated with ethanol phenotypes, we find expression of these transcripts in all tissues, with strong enrichment observed in some modules for expression in the midgut, Malpighian tubules, and testis (Figure S2). Modules of correlated transcripts associated with

variation in ethanol phenotypes were also enriched for binding motifs of several transcription factors (Table S6).

Analysis of the connectivity of the genes in tolerance module 2 reveals two highly connected gene clusters, one associated with synaptic function, and the other with electron transport and oxidative phosphorylation (Figure 3). Gene products associated with muscle contraction and carbohydrate biosynthesis, processes dependent on ATP, are also contained within this cluster. Genetically correlated transcripts associated with ATP synthesis are linked to the cluster of transcripts associated with synaptic function through *comatose* (*comt*), which encodes an *N*-ethylmaleimide sensitive fusion protein that mediates ATP-dependent synaptic vesicle release.

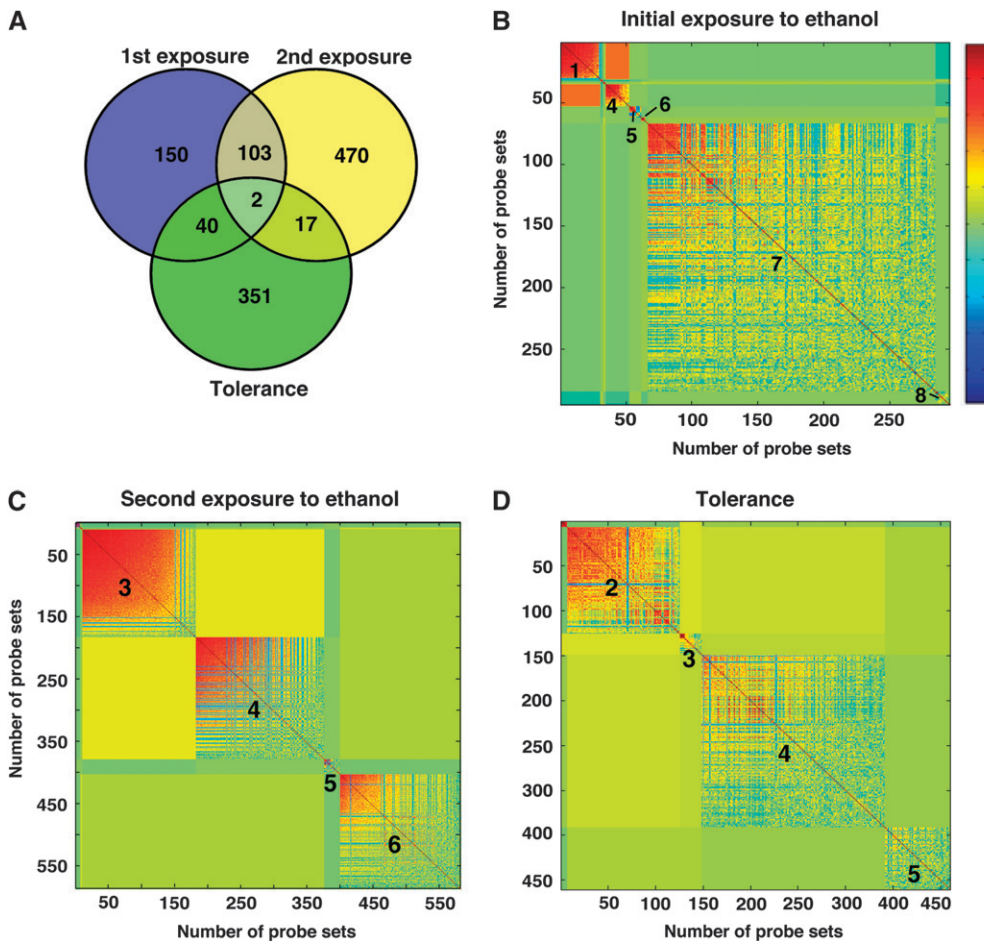


FIGURE 2.—Transcripts associated with variation among wild-derived inbred lines for ethanol sensitivity and tolerance. (A) Venn diagram of the overlap between transcripts associated with the MET from the first and second exposures to ethanol, and tolerance. (B–D) Modules of genetically correlated transcripts associated with variation in alcohol-associated traits. Modules are ordered along the diagonal on the basis of the average module correlation, with the strongest correlations in the upper left corner of each diagram. The colored blocks on the off-diagonal represent the cross-module correlations. Dark red indicates a positive correlation close to 1 and dark blue a negative correlation close to -1 . (B) The 295 transcripts associated with variation in MET after the initial exposure to ethanol clustered into eight modules. (C) The 592 transcripts associated with variation in MET after the second ethanol exposure clustered into six modules. (D) The 410 transcripts associated with tolerance clustered into five modules.

Many transcripts associated with ethanol phenotypes are predicted genes with unknown function. However, their connectivity with other transcripts across the whole genome and within modules of correlated transcripts associated with a specific phenotype can generate testable hypotheses regarding their functions. Transcripts of *CG16743* and *CG5704* are associated with all three ethanol phenotypes. GO analysis of the 100 transcripts most highly genetically correlated with each of these genes indicates that *CG16743* is most likely a peptidase and *CG5704* a transcription factor (Table S7). Similarly, *CG1882* is associated with tolerance and five other quantitative traits, including recovery from chill coma and waking activity (Table S4), and is highly correlated with transcripts affecting oxidative phosphorylation and phototransduction. *CG7990* and *CG9350* are in tolerance module 2 and expressed in the brain. *CG7990* probably affects synaptic transmission and *CG9350* oxidative phosphorylation, consistent with other transcripts in tolerance module 2 (Table S7).

Validation of transcriptional networks: To validate these networks we identified 11 *P*-element insertions in

a common co-isogenic background (BELLEN *et al.* 2004) that correspond to transcripts in modules associated with the initial exposure to alcohol. Seven of these indeed affect alcohol sensitivity (Figure 4A). One of these genes, *psq*, belongs to a family of helix-loop-helix transcription factors that is essential for neurodevelopment (NORGA *et al.* 2003), affects olfactory behavior (SAMBANDAN *et al.* 2006), and shows a correlated transcriptional response to selection for ethanol sensitivity and resistance (MOROZOVA *et al.* 2007). Compared to the co-isogenic control, expression of *psq* is enhanced in the *psq* mutation (Figure 4B; Table S8). We predicted that transcripts genetically correlated with *psq* in E1 module 7 would also show altered transcript levels in the *psq* mutant background, and found that expression of 17 of 23 such transcripts tested (74%) is altered in the *psq* mutant line (Figure 4C; Table S8). Since *P*-element mutations often cause widespread changes in transcript abundance (ANHOLT *et al.* 2003), we also expect to find altered expression of genes that do not belong to correlated transcriptional modules associated with ethanol phenotypes in the *psq* mutant line, but that the

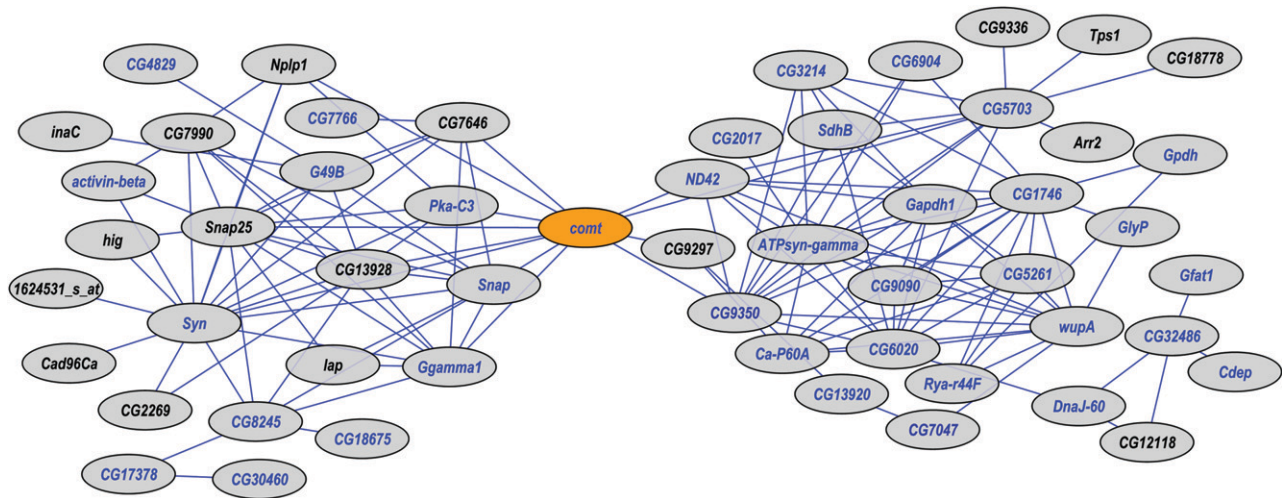


FIGURE 3.—Highly connected gene clusters within tolerance module 2. The gene cluster to the left is associated with synaptic function, whereas transcripts associated with oxidative phosphorylation are enriched in the cluster depicted on the right. Both clusters are connected via *comt*. *Drosophila* genes with human orthologs (68.5%) are indicated in blue type. All genes depicted have transcriptional genetic correlations ≥ 0.7 .

proportion of such genes will be lower than that of genes in the module. Indeed, transcript levels of only 2 of 11 genes (18%) not genetically correlated with *psq* show altered expression in the *psq* mutation (Fisher's exact test $P < 0.0005$; Table S8).

To further examine the effects of *psq* on gene expression we surveyed genes in E1 module 7 for Psq binding GAGAG motifs (SCHWENDEMANN and LEHMANN 2002) in promoter regions and found such motifs in *bi*, *phol*, *Cbl*, and *simj*. Multiple GAGAG repeats 5' to the ATG translation initiation codon in the gene that encodes the *bifid* (*bi*, a.k.a. *optomotor blind*) T-Box type transcription factor were especially prominent (PORSCH *et al.* 2005). Moreover, among the 17 genes analyzed by quantitative RT-PCR, *bi* showed the greatest reduction in transcript abundance in the *psq* *P*-element mutant (Table S8). These findings indicate that the genetic correlations with the phenotype within the *psq* module presented in Figure 4C do not originate from direct transcriptional effects of *psq* on each correlated gene, but rather arise from a complex cascade of indirect effects.

Next, we examined network connectivity with *Men* in E2 module 6, which contains a large number of metabolic enzymes (Table S5), and validated the connectivity using a line in which a *P*-element has inserted in the *Men* gene. This *P*-element insertion line is resistant to alcohol exposure compared to the coisogenic Canton-S (*B*) control by an average MET of 2.8 min ($P < 0.0001$). Analysis of *Men* expression across all developmental stages shows that this *P*-element insertion mutant is essentially a null mutant (Figure 5A). We analyzed expression levels of 16 genes connected in E2 module 6 with *Men* (Figure 5B) and found that 12 (75%) of the transcripts tested indeed showed altered transcript levels measured by RT-PCR in the *P*-element insertion mutant (Figure 5B; Table S8). At

the same time, only 3 of 11 transcripts not genetically correlated with *Men* showed altered expression in the *P*-element insertion mutant (Fisher's exact test $P = 0.02$; Table S8), once again validating the computationally derived covariant module.

We selected four lines with the shortest and four lines with the longest second ethanol exposure METs averaged across sexes and measured Malic enzyme activities after a single exposure to ethanol. We observed significant differences in the means of enzyme activities between the lines, consistent with enhanced Malic enzyme activity contributing to subsequent induction of tolerance and providing further validation for the involvement of Malic enzyme in alcohol sensitivity (Figure 6). Differences in Malic enzyme activity were transient and no longer apparent after a second exposure to ethanol. We also found a significant correlation between variation in transcript level of flies that had not been exposed to ethanol and Malic enzyme activity measured after a single exposure to ethanol during induction of tolerance ($r = 0.49$, $P = 0.0004$). This correlation was no longer evident when Malic enzyme activity was measured after a second exposure to ethanol, after tolerance had been established ($r = 0.098$; $P = 0.50$).

Translational potential: *Variation in alcohol drinking behavior in the Framingham Heart Study cohort:* Individual drinking behavior was assessed over a period of 25 years (KANNEL *et al.* 1979). Each individual was evaluated on average five to six times during the course of the study. Analysis of variance revealed a strong effect of sex for all types of alcoholic drinks, with men overall drinking more beer than women ($P < 0.0001$). There was also a strong effect of age ($P < 0.0001$), with individuals drinking less alcohol as they age. We did not find a correlation between beer, wine, or cocktail drinking (Figure S3). A more extensive description of the

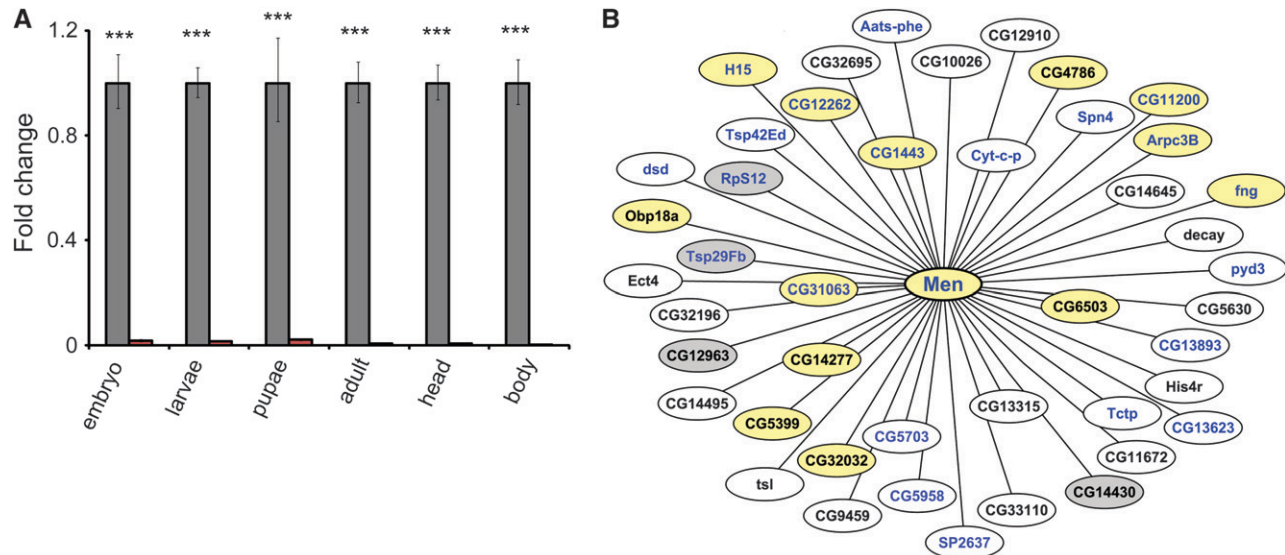


FIGURE 5.—Validation of genes in a transcriptional network with *Malic enzyme*. (A) Relative fold changes in mRNA levels for the *Men* transcript in the *P*-element insertion line. Gray bars show expression of *Men* in the control; red bars indicate *Men* expression in the *Men* mutant line at different developmental stages. The *P*-element insertion causes downregulation of *Men* in all stages, concomitant with greater resistance to ethanol exposure ($***P < 0.001$). (B) Transcripts correlated with *Men* expression and associated with natural variation in ethanol sensitivity (E2 module 6). The strength of correlation is reflected by the distance of each gene from *Men* (Table S8) with greater correlations being closer to the focal gene. Transcripts that were analyzed by RT-PCR and showed significant alterations in transcript abundance in the *Men* mutant line, are indicated with a yellow background, while transcripts that were analyzed and did not show significant differences in transcript abundance are indicated with a gray background (Table S8). *Drosophila* genes with annotated human orthologs (49%) are indicated in blue type.

with cocktail drinking (Figure 7). Six SNPs (rs1180242, rs1144188, rs1145913, rs1144185, rs1144184, rs6941094) survived a permutation threshold of 0.05 after correction for multiple testing, and one SNP (rs1145916) was significant at a permutation threshold of 0.01. Although the tagging SNPs within *ME1* were not in linkage disequilibrium in the HapMap population, we found significant and often strong pairwise linkage disequilibrium among the 21 SNPs in *ME1* genotyped in the Framingham cohort (Figure 7; Table S1). The significant SNPs are located in intron 1, intron 6, and intron 13 of *ME1*, and are all in strong linkage disequilibrium with each other (Table S1 and Table S9). Although significant, the effects of the SNPs in *ME1* on cocktail drinking are small, consistent with the emerging view of the genetic architecture of human complex traits. The effects ranged from 0.35 to 0.64 drinks per day (0.085–0.156 phenotypic standard deviations), and accounted for 0.36–0.9% of the phenotypic variation in cocktail drinking in this population (Table S9).

DISCUSSION

Studies on model organisms have provided a wealth of insights in universal biological processes; yet, the advantage model organisms offer for translational studies that can lead to discovery of genes that harbor risk alleles for human disorders remains remarkably underexplored. The WORLD HEALTH ORGANIZATION (2004) has estimated that alcohol abuse disorders affect

~76.3 million people worldwide. Despite the identification of some candidate genes (RADEL and GOLDMAN 2001; MULLIGAN *et al.* 2003; SINHA *et al.* 2003; FLATSCHER-BADER *et al.* 2005; GOLDMAN *et al.* 2005; RADEL *et al.* 2005; EDENBERG *et al.* 2006; BIRLEY *et al.* 2009), the genetic underpinnings of sensitivity to alcohol remain largely unknown. Here, we demonstrate that findings from studies on alcohol sensitivity in *Drosophila* can be applied directly to the identification of genetic factors associated with human alcohol intake.

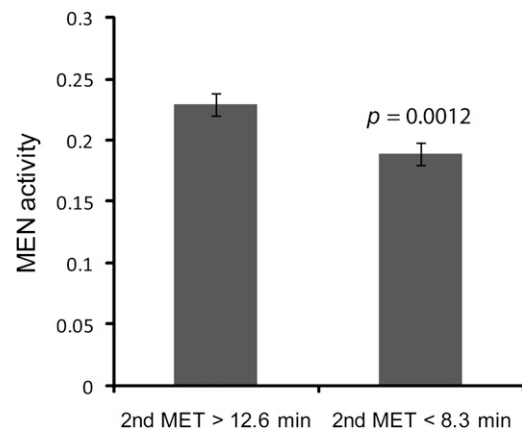


FIGURE 6.—Malic enzyme activities after a single pass through the inebriometer in lines with short and long second exposure METs. Bars represent the averages and SEM for sexes combined of four lines, each measured in triplicate. Enzyme activity is expressed as $10,000 \times \mu\text{mol NADPH}/\text{min}/\mu\text{g}$ protein.

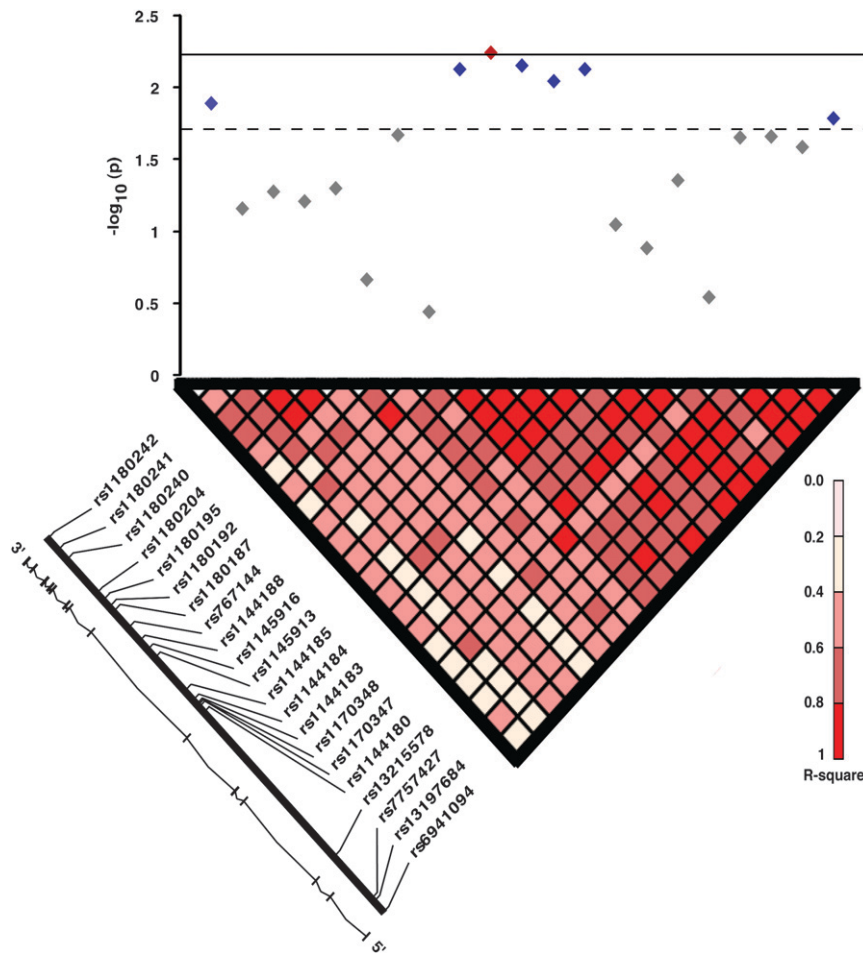


FIGURE 7.—Association between cocktail drinking and polymorphisms in the *Malic Enzyme 1 (ME1)* gene. Single nucleotide polymorphisms associated with cocktail drinking at a permutation threshold of $P < 0.05$ (lower horizontal dash line) are indicated as blue diamonds, whereas the SNP that exceeds the $P < 0.01$ permutation threshold (upper horizontal solid line) is indicated in red. Linkage disequilibrium among the SNPs, measured as r^2 , is illustrated in the diagram below the x -axis of the graph. SNPs, identified from the HapMap, are ordered according to their chromosomal locations and are represented to scale on the black bar below the SNP names. The corresponding gene model of the *ME1* gene is derived from ENSEMBL and shows 14 exons (vertical black boxes) and 13 intervening introns.

Systems genetics in *Drosophila*: coregulated transcriptional networks for alcohol sensitivity and tolerance: For our gene discovery approach to be successful, it was essential first to characterize the genetic underpinnings of alcohol sensitivity and tolerance in *Drosophila*. We capitalized on natural variation among 40 wild-derived inbred lines to identify correlated transcriptional networks associated with response to ethanol exposure. It should be noted that genes that contribute to the response to ethanol exposure, but do not vary in transcript abundance, would not be included in these networks. Transcripts that are present at levels below the detection limit of the Affymetrix microarrays would also go undetected. Furthermore, the exact composition of the transcriptional networks may differ for different populations, and genotype-by-environment interactions may cause shifts in the transcriptional modules under different environmental conditions (FAY *et al.* 2004; LANDRY *et al.* 2006; SAMBANDAN *et al.* 2006; MONROY *et al.* 2007). Nonetheless, it is clear that the transcriptional network associated with acute exposure to ethanol shifts dramatically as flies develop tolerance. This is in agreement with a previous study, which showed that in an isogenic laboratory strain (Canton-S) induction of tolerance is accompanied by rapid downregulation of a

suite of chemosensory genes and modulation of expression of detoxification enzymes, followed by altered regulation of genes that encode metabolic enzymes (MOROZOVA *et al.* 2006). The modules of correlated transcripts that we identified here mimic some aspects of these previous observations. Initial sensitivity to ethanol exposure is associated with correlated transcripts that encode gene products that mediate defense responses to environmental chemicals (*e.g.*, glutathione S transferases), and tolerance involves transcriptional networks associated with intermediary metabolism, including generation of ATP necessary for synaptic exocytosis. Perturbations of the computed modules by introduction of a transposon in a different, Canton-S (*B*), genetic background results in altered levels of transcripts predicted to covary with the disrupted gene, providing direct validation of the modules and indicating significant robustness of the observed modules associated with response to alcohol exposure across genetic backgrounds.

The modular organization of the transcriptional networks associated with alcohol-related phenotypes can provide information about the possible functions of computationally predicted transcripts in the modules based on the “guilt by association” principle and, thus,

can serve as an efficient method for gene annotation. If modules associated with alcohol-related phenotypes in *Drosophila* contain transcripts with human orthologs associated with alcohol-related phenotypes in people, it is reasonable to predict that the latter are likely to preserve a similar modular connectivity. One can then nominate genes of which a functional context is already known as candidates for association studies in human populations. This strategy avoids the large multiple testing penalty inherent in unbiased large scale GWAS and confers power to detect significant associations in a moderate size population. We have used the gene that encodes the cytosolic form of human Malic enzyme to demonstrate proof of principle for this approach.

Identification of human orthologs as candidate genes for drinking behavior: the *MEI* gene as proof of principle: Alcohol intake in the Framingham population was assessed as three different behavioral phenotypes: beer drinking, wine drinking, and cocktail drinking. It should be noted that the vast majority of individuals in this cohort were light to moderate drinkers, with a smaller number of heavy drinkers (Figure S3). Consequently, we do not make inferences about alcohol addiction or alcohol abuse, but instead focus on normal tendencies toward alcohol consumption. Drinking behaviors show sexual dimorphism with men being more inclined to consuming beer. Previous studies have identified association of *Alcohol dehydrogenase* (*Adh*) alleles with alcohol intake (EHRIG *et al.* 1990; AGARWAL and GOEDDE 1992; WHITFIELD 1994; EDENBERG *et al.* 2006; BIRLEY *et al.* 2009). Different *Adh* alleles differ in enzymatic effectiveness and are correlated with the extent of tolerance to alcohol consumption. High tolerance provides increased risk for addiction. Asian populations have a generally lower risk for developing alcohol dependence than European populations due to differences in frequencies of alleles for *Adh* and *Aldehyde dehydrogenase*, designated *ADH1B* and *ALDH2*, respectively (MULLIGAN *et al.* 2003; EDENBERG *et al.* 2006; BIRLEY *et al.* 2009). However, *Adh* is only one of many likely factors that contribute to facilitating alcohol consumption.

In populations where there is no variation for the *Adh* alleles, other factors must contribute to phenotypic variation in drinking behavior. On the basis of the present study and previous studies in *Drosophila* (MOROZOVA *et al.* 2006, 2007), Malic enzyme emerges as a significant facilitator of alcohol intake as it provides a critical metabolic link between the glycolytic pathway, the tricarboxylic acid cycle, and fatty acid biosynthesis, which can shunt alcohol-derived excess energy toward lipid biosynthesis. Indeed, we found significant associations between cocktail drinking and polymorphisms in the *MEI* gene, which encodes cytosolic Malic enzyme. The alcohol content in cocktails is greater than in beer or wine, which may account for our ability to detect associations only with cocktail drinking, but not beer and wine drinking, given the sample size.

It is of interest to note that SNPs associated with cocktail drinking are intronic and, thus, do not change the amino acid sequence of the protein, but might affect enzyme level. Due to the linkage disequilibrium structure in the human genome (GABRIEL *et al.* 2002; FRAZER *et al.* 2007) we cannot determine unambiguously which polymorphism is causal. Moreover, historical recombination within the *MEI* gene differs between the Framingham study cohort and the HapMap CEPH population. Thus, it is possible that SNPs in *MEI* in different populations may contribute to phenotypic variation in drinking behavior to different extents, and in different populations different SNPs within *MEI* might be associated with alcohol intake.

The lack of power of many human genetics studies in the past has often led to poorly supported results and, consequently, replication in independent study populations has been widely adopted as a general standard for reliability. While this ensures that genes with risk alleles with consistent effects can be identified with confidence, it discards risk alleles with effects that are sensitive to genetic background and genotype-by-environment interactions in specific populations, even when such studies are executed with good statistical power. Here we show that translational studies that exploit model organisms can provide a viable alternative to replicate human populations and allow the detection of risk alleles with small effects. *Men* has been associated with alcohol resistance in both flies (MOROZOVA *et al.* 2006) and mice (YIN *et al.* 2007), and the current study shows that translational approaches from findings in model organisms can be extended to human candidate genes. Perhaps the most convincing demonstration of the power of the translational approach advocated in this article is the estimate of the number of individuals that would need to be genotyped for 500,000 SNPs in a GWAS aimed at identifying alleles associated with drinking behavior in people to detect the same size of effect resolved in our study. We would need 903 individuals in the rarer homozygous SNP genotype class to detect our largest effect (0.311 phenotypic standard deviations) following a Bonferroni correction for multiple tests (for a nominal *P*-value of 10^{-7}) (FALCONER and MACKAY 1996). Since the frequency of the rarer homozygote genotype is ~ 0.118 , a total of 7653 individuals would need to be both genotyped and phenotyped to detect an effect of this magnitude.

This work was supported by grants AA016569 (T.F.C.M.), GM45146 (T.F.C.M.), GM59469 (R.R.H.A.), and AA013304 (R.C.E.) from the National Institutes of Health.

LITERATURE CITED

- AFFYMETRIX, NETAFFX ANALYSIS CENTER, 2008 <http://www.affymetrix.com>.
 AGARWAL, D. P., and H. W. GOEDDE, 1992 Pharmacogenetics of alcohol metabolism and alcoholism. *Pharmacogenetics* 2: 48–62.

- ANHOLT, R. R. H., C. L. DILDA, S. CHANG, J. J. FANARA, N. H. KULKARNI *et al.*, 2003 The genetic architecture of odor-guided behavior in *Drosophila*: epistasis and the transcriptome. *Nat. Genet.* **35**: 180–184.
- APPLIED BIOSYSTEMS, A., 2001 Relative quantitation of gene expression. User Bulletin No. 2: ABI PRISM 7700 Sequence Detection System. Foster City, CA.
- AYROLES, J. F., M. CARBONE, E. A. STONE, K. W. JORDAN, R. F. LYMAN *et al.*, 2009 Systems genetics of complex traits in *Drosophila melanogaster*. *Nat. Genet.* **41**: 299–307.
- BELLEN, H. J., R. W. LEVIS, G. LIAO, Y. HE, J. W. CARLSON *et al.*, 2004 The BDGP gene disruption project: single transposon insertions associated with 40% of *Drosophila* genes. *Genetics* **167**: 761–781.
- BILEN, J., and N. M. BONINI, 2005 *Drosophila* as a model for human neurodegenerative disease. *Annu. Rev. Genet.* **39**: 153–171.
- BIRLEY, A. J., M. R. JAMES, P. A. DICKSON, G. W. MONTGOMERY, A. C. HEATH *et al.*, 2009 ADH single nucleotide polymorphism associations with alcohol metabolism *in vivo*. *Hum. Mol. Genet.* **18**: 1533–1542.
- BUCHANAN, A. V., S. SHOLTIS, J. RICHTSMIEER and K. M. WEISS, 2009 What are genes “for” or where are traits “from?” What is the question? *Bioessays* **31**: 198–208.
- CHEN, Y., J. ZHU, P. Y. LUM, X. YANG, S. PINTO *et al.*, 2008 Variations in DNA elucidate molecular networks that cause disease. *Nature* **452**: 429–435.
- CHENG, Y., K. ENDO, K. WU, A. R. RODAN, U. HEBERLEIN *et al.*, 2001 *Drosophila fasciclinII* is required for the formation of odor memories and for normal sensitivity to alcohol. *Cell* **105**: 757–768.
- CHINTAPALLI, V. R., J. WANG and J. A. DOW, 2007 Using FlyAtlas to identify better *Drosophila melanogaster* models of human disease. *Nat. Genet.* **39**: 715–720.
- COOKSON, W., L. LIANG, G. ABECASIS, M. MOFFATT and M. LATHROP, 2009 Mapping complex disease traits with global gene expression. *Nat. Rev. Genet.* **10**: 184–194.
- CRABBE, J. C., 2008 Review. Neurogenetic studies of alcohol addiction. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **363**: 3201–3211.
- DAWBER, T. R., W. B. KANNEL, D. E. LOVE and R. B. STREPPER, 1952 The electrocardiogram in heart disease detection; a comparison of the multiple and single lead procedures. *Circulation* **5**: 559–566.
- DE BAKKER, P. I., R. YELENSKY, I. PE’ER, S. B. GABRIEL, M. J. DALY *et al.*, 2005 Efficiency and power in genetic association studies. *Nat. Genet.* **37**: 1217–1223.
- DENNIS, JR., G., B. T. SHERMAN, D. A. HOSACK, J. YANG, W. GAO *et al.*, 2003 DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.* **4**: 3.
- DJOUSSE, L., J. F. DORGAN, Y. ZHANG, A. SCHATZKIN, M. HOOD *et al.*, 2002 Alcohol consumption and risk of lung cancer: the Framingham Study. *J. Natl. Cancer Inst.* **94**: 1877–1882.
- DRYSDALE, R. A., and M. A. CROSBY, 2005 FlyBase: genes and gene models. *Nucleic Acids Res.* **33**: D390–D395.
- EDENBERG, H. J., X. XUEI, H. J. CHEN, H. TIAN, L. F. WETHERILL *et al.*, 2006 Association of *alcohol dehydrogenase* genes with alcohol dependence: a comprehensive analysis. *Hum. Mol. Genet.* **15**: 1539–1549.
- EHRIG, T., W. F. BOSRON and T. K. LI, 1990 Alcohol and *aldehyde dehydrogenase*. *Alcohol Alcohol.* **25**: 105–116.
- EMILSSON, V., G. THORLEIFSSON, B. ZHANG, A. S. LEONARDSON, F. ZINK *et al.*, 2008 Genetics of gene expression and its effect on disease. *Nature* **452**: 423–428.
- FALCONER, D. S., and T. F. C. MACKAY, 1996 *Introduction to Quantitative Genetics*, Ed. 4. Addison-Wesley Longman, Harlow, UK.
- FAY, J. C., H. L. MCCULLOUGH, P. D. SNIEGOWSKI and M. B. EISEN, 2004 Population genetic variation in gene expression is associated with phenotypic variation in *Saccharomyces cerevisiae*. *Genome Biol.* **5**: R26.
- FEULNER, T. M., S. M. LAWS, P. FRIEDRICH, S. WAGENPFEL, S. H. WURST *et al.*, 2009 Examination of the current top candidate genes for AD in a genome-wide association study. *Mol. Psychiatry* (in press).
- FLATSCHER-BADER, T., M. VAN DER BRUG, J. W. HWANG, P. A. GOCHEE, I. MATSUMOTO *et al.*, 2005 Alcohol-responsive genes in the frontal cortex and nucleus accumbens of human alcoholics. *J. Neurochem.* **93**: 359–370.
- FLINN, L., S. BRETAUD, C. LO, P. W. INGHAM and O. BANDMANN, 2008 Zebrafish as a new animal model for movement disorders. *J. Neurochem.* **106**: 1991–1997.
- FRAZER, K. A., D. G. BALLINGER, D. R. COX, D. A. HINDS, L. L. STUVE *et al.*, 2007 A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**: 851–861.
- GABRIEL, S. B., S. F. SCHAFFNER, H. NGUYEN, J. M. MOORE, J. ROY *et al.*, 2002 The structure of haplotype blocks in the human genome. *Science* **296**: 2225–2229.
- GOLDMAN, D., G. OROSZI and F. DUCCI, 2005 The genetics of addictions: uncovering the genes. *Nat. Rev. Genet.* **6**: 521–532.
- GUARNIERI, D. J., and U. HEBERLEIN, 2003 *Drosophila melanogaster*, a genetic model system for alcohol research. *Int. Rev. Neurobiol.* **54**: 199–228.
- HOCHBERG, Y., and Y. BENJAMINI, 1990 More powerful procedures for multiple significance testing. *Stat. Med.* **9**: 811–818.
- JOHNSON, A. D., and C. J. O’DONNELL, 2009 An open access database of genome-wide association results. *BMC Med. Genet.* **10**: 6.
- JOOPER, R., P. BOKSA, C. BENKELFAT and G. ROULEAU, 2002 Genetics of schizophrenia: from animal models to clinical studies. *J. Psychiatry Neurosci.* **27**: 336–347.
- KANNEL, W. B., M. FEINLEIB, P. M. MCNAMARA, R. J. GARRISON and W. P. CASTELLI, 1979 An investigation of coronary heart disease in families. The Framingham Offspring Study. *Am. J. Epidemiol.* **110**: 281–290.
- LANDRY, C. R., J. OH, D. L. HARTL and D. CAVALIERI, 2006 Genome-wide scan reveals that genetic variation for transcriptional plasticity in yeast is biased towards multi-copy and dispensable genes. *Gene* **366**: 343–351.
- MACKAY, T. F. C., and R. R. H. ANHOLT, 2006 Of flies and man: *Drosophila* as a model for human complex traits. *Annu. Rev. Genomics Hum. Genet.* **7**: 339–367.
- MERRITT, T. J., C. KUCZYNSKI, E. SEZGIN, C. T. ZHU, S. KUMAGAI *et al.*, 2009 Quantifying interactions within the NADP(H) enzyme network in *Drosophila melanogaster*. *Genetics* **182**: 565–574.
- MONROY, A. F., A. DRYANOVA, B. MALETTE, D. H. OREN, M. RIDHA FARAJALLA *et al.*, 2007 Regulatory gene candidates and gene expression analysis of cold acclimation in winter and spring wheat. *Plant Mol. Biol.* **64**: 409–423.
- MOORE, M. S., J. DEZAZZO, A. Y. LUK, T. TULLY, C. M. SINGH *et al.*, 1998 Ethanol intoxication in *Drosophila*: genetic and pharmacological evidence for regulation by the cAMP signaling pathway. *Cell* **93**: 997–1007.
- MOROZOVA, T. V., R. R. H. ANHOLT and T. F. C. MACKAY, 2006 Transcriptional response to alcohol exposure in *Drosophila melanogaster*. *Genome Biol.* **7**: R95.
- MOROZOVA, T. V., R. R. H. ANHOLT and T. F. C. MACKAY, 2007 Phenotypic and transcriptional response to selection for alcohol sensitivity in *Drosophila melanogaster*. *Genome Biol.* **8**: R231.
- MOSKVIN, V., N. CRADDOCK, P. HOLMANS, I. NIKOLOV, J. S. PAHWA *et al.*, 2009 Gene-wide analyses of genome-wide association data sets: evidence for multiple common risk alleles for schizophrenia and bipolar disorder and for overlap in genetic risk. *Mol. Psychiatry* **14**: 252–260.
- MULLIGAN, C. J., R. W. ROBIN, M. V. OSIER, N. SAMBUUGHIN, L. G. GOLDFARB *et al.*, 2003 Allelic variation at alcohol metabolism genes (*ADH1B*, *ADH1C*, *ALDH2*) and alcohol dependence in an American Indian population. *Hum. Genet.* **113**: 325–336.
- NICA, A. C., and E. T. DERMITZAKIS, 2008 Using gene expression to investigate the genetic basis of complex disorders. *Hum. Mol. Genet.* **17**: R129–R134.
- NORGA, K. K., M. C. GURGANUS, C. L. DILDA, A. YAMAMOTO, R. F. LYMAN *et al.*, 2003 Quantitative analysis of bristle number in *Drosophila* mutants identifies genes involved in neural development. *Curr. Biol.* **13**: 1388–1396.
- O’DONOVAN, M. C., N. CRADDOCK, N. NORTON, H. WILLIAMS, T. PEIRCE *et al.*, 2008 Identification of loci associated with schizophrenia by genome-wide association and follow-up. *Nat. Genet.* **40**: 1053–1055.
- PANKRATZ, N., J. B. WILK, J. C. LATOURELLE, A. L. DE’STEFANO, C. HALTER *et al.*, 2009 Genomewide association study for sus-

- ceptibility genes contributing to familial Parkinson disease. *Hum. Genet.* **124**: 593–605.
- PORSCH, M., M. SAUER, S. SCHULZE, A. BAHLO, M. ROTH *et al.*, 2005 The relative role of the T-domain and flanking sequences for developmental control and transcriptional regulation in protein chimeras of *Drosophila* OMB and ORG-1. *Mech. Dev.* **122**: 81–96.
- PSYCHIATRIC GWAS CONSORTIUM STEERING COMMITTEE, 2009 A framework for interpreting genome-wide association studies of psychiatric disorders. *Mol. Psychiatry* **14**: 10–17.
- RADEL, M., and D. GOLDMAN, 2001 Pharmacogenetics of alcohol response and alcoholism: the interplay of genes and environmental factors in thresholds for alcoholism. *Drug Metab. Dispos.* **29**: 489–494.
- RADEL, M., R. L. VALLEJO, N. IWATA, R. ARAGON, J. C. LONG *et al.*, 2005 Haplotype-based localization of an alcohol dependence gene to the 5q34 [gamma]-aminobutyric acid type A gene cluster. *Arch. Gen. Psychiatry* **62**: 47–55.
- ROTHENFLUH, A., and U. HEBERLEIN, 2002 Drugs, flies, and videotape: the effects of ethanol and cocaine on *Drosophila* locomotion. *Curr. Opin. Neurobiol.* **12**: 639–645.
- ROTHENFLUH, A., R. J. THRELKELD, R. J. BAINTON, L. T. TSAI, A. W. LASEK *et al.*, 2006 Distinct behavioral responses to ethanol are regulated by alternate RhoGAP18B isoforms. *Cell* **127**: 199–211.
- SAMBANDAN, D., A. YAMAMOTO, J. J. FANARA, T. F. C. MACKAY and R. R. H. ANHOLT, 2006 Dynamic genetic interactions determine odor-guided behavior in *Drosophila melanogaster*. *Genetics* **174**: 1349–1363.
- SCHOLZ, H., M. FRANZ and U. HEBERLEIN, 2005 The *hangover* gene defines a stress pathway required for ethanol tolerance development. *Nature* **436**: 845–847.
- SCHOLZ, H., J. RAMOND, C. M. SINGH and U. HEBERLEIN, 2000 Functional ethanol tolerance in *Drosophila*. *Neuron* **28**: 261–271.
- SCHWENDEMANN, A., and M. LEHMANN, 2002 Pipsqueak and GAGA factor act in concert as partners at homeotic and many other loci. *Proc. Natl. Acad. Sci. USA* **99**: 12883–12888.
- SIEBERTS, S. K., and E. E. SCHADT, 2007 Moving toward a systems genetics view of disease. *Mamm. Genome* **18**: 389–401.
- SINGH, C. M., and U. HEBERLEIN, 2000 Genetic control of acute ethanol-induced behaviors in *Drosophila*. *Alcohol Clin. Exp. Res.* **24**: 1127–1136.
- SINHA, R., C. R. CLONINGER and A. PARSIAN, 2003 Linkage disequilibrium and haplotype analysis between *serotonin receptor 1B* gene variations and subtypes of alcoholism. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **121**: 83–88.
- STONE, E. A., and J. F. AYROLES, 2009 Modulated modularity clustering as a tool for systems genetics. *PLOS Genet.* **5**: e1000479.
- WORLD HEALTH ORGANIZATION, 2004 Global Status Report: Alcohol Policy. Department of Mental Health and Substance Abuse. WHO Press, Geneva.
- WEBER, K. E., 1988 An apparatus for measurement of resistance to gas-phase reagents. *Drosophila Inform. Serv.* **67**: 90–92.
- WHITFIELD, J. B., 1994 ADH and ALDH genotypes in relation to alcohol metabolic rate and sensitivity. *Alcohol Alcohol. Suppl.* **2**: 59–65.
- WOLF, F. W., A. R. RODAN, L. T. TSAI and U. HEBERLEIN, 2002 High-resolution analysis of ethanol-induced locomotor stimulation in *Drosophila*. *J. Neurosci.* **22**: 11035–11044.
- WOODRUFF-PAK, D. S., 2008 Animal models of Alzheimer's disease: therapeutic implications. *J. Alzheimers Dis.* **15**: 507–521.
- YIN, H. Q., M. KIM, J. H. KIM, G. KONG, K. S. KANG *et al.*, 2007 Differential gene expression and lipid metabolism in fatty liver induced by acute ethanol treatment in mice. *Toxicol. Appl. Pharmacol.* **223**: 225–233.

Communicating editor: N. PERRIMON

GENETICS

Supporting Information

<http://www.genetics.org/cgi/content/full/genetics.109.107490/DC1>

Alcohol Sensitivity in *Drosophila*: Translational Potential of Systems Genetics

**Tatiana V. Morozova, Julien F. Ayroles, Katherine W. Jordan, Laura H. Duncan,
Mary Anna Carbone, Richard F. Lyman, Eric A. Stone, Diddahally R. Govindaraju,
R. Curtis Ellison, Trudy F. C. Mackay and Robert R. H. Anholt**

Copyright © 2009 by the Genetics Society of America
DOI: 10.1534/genetics.109.107490

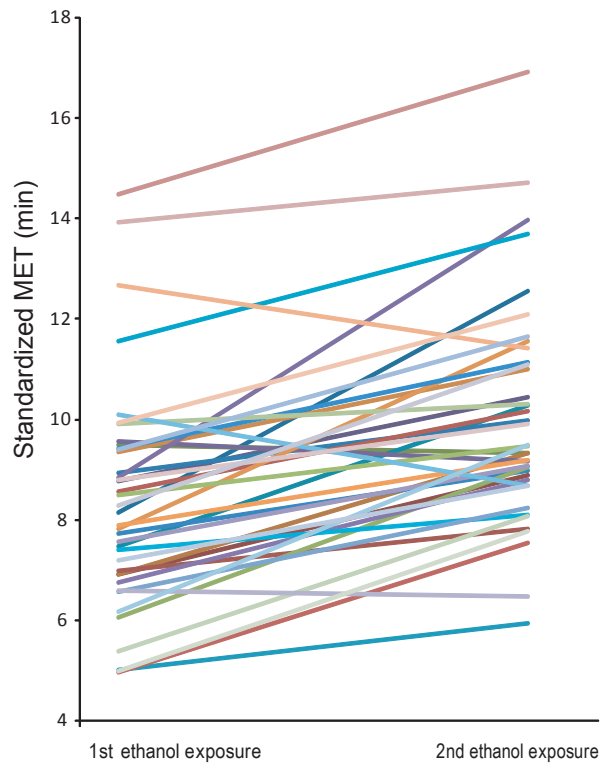


FIGURE S1.—Reaction norms for ethanol sensitivity after one and two exposures. The lines depict the MET of each of the 40 wild-derived inbred lines for the initial (E1) and second (E2) exposure to ethanol. The extensive crossing of reaction norms shows genetic variation in development of tolerance.

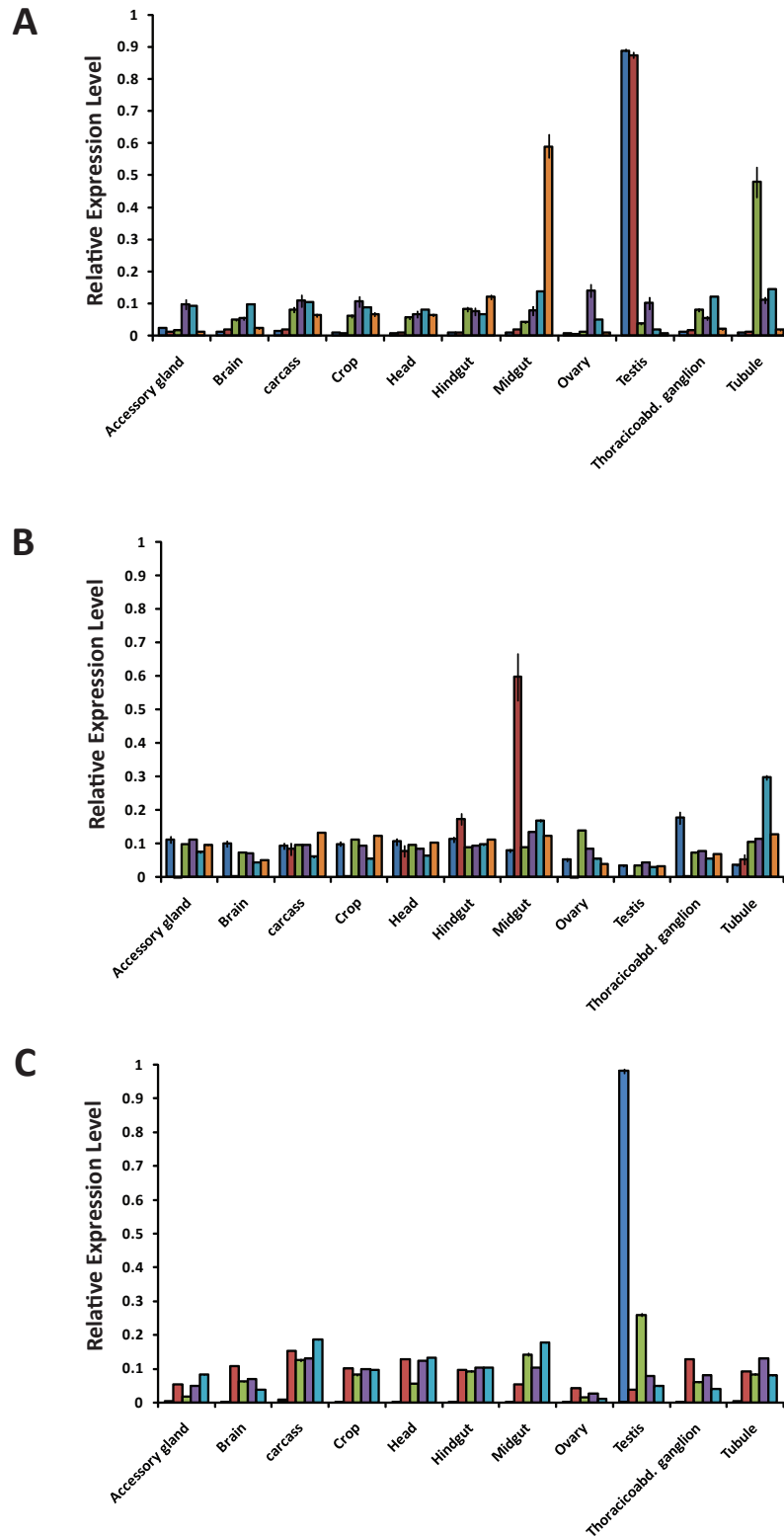


FIGURE S2.—Relative expression of transcripts in each module associated with ethanol phenotypes. Only modules with more than three transcripts are depicted. (A) Transcript modules associated with MET after one exposure to ethanol. (B) Transcript modules associated with MET after two exposures to ethanol. (C) Transcript modules associated with tolerance.

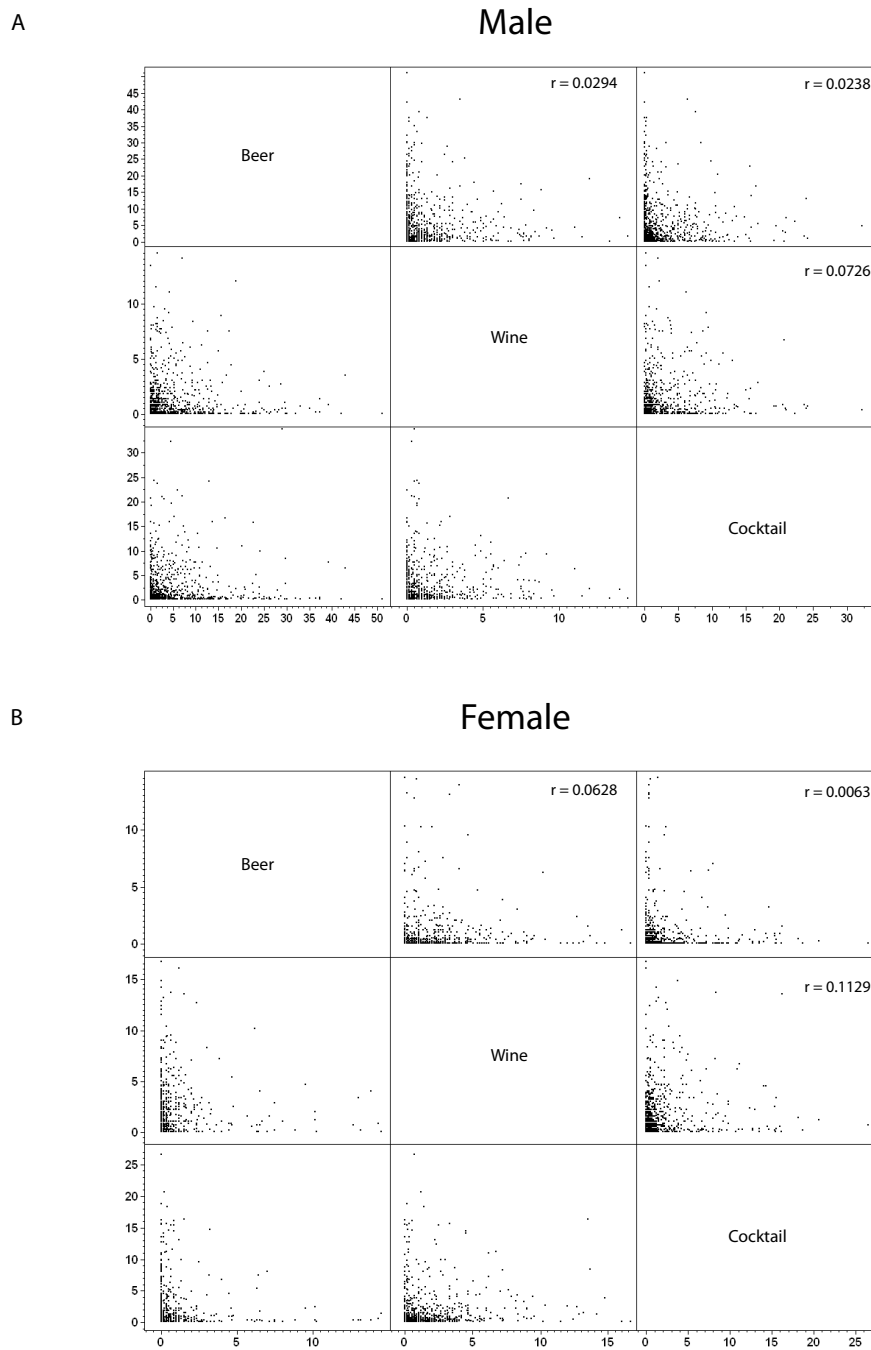


FIGURE S3.—Phenotypic correlations among beer, wine and cocktail intake (A) Male drinking behavior averaged across exams. (B) Female drinking behavior averaged across exams. Units on the axis represent the number of drinks consumed per week.

TABLE S1

**Linkage disequilibrium among 21 SNPs in ME1 in 1,687 individuals of the Framingham Heart Study
Offspring Cohort**

Table S1 is available for download as an Excel file at <http://www.genetics.org/cgi/content/full/genetics.109.107490/DC1>.

TABLE S2

Analyses of variance of mean elution time after one (E1) or two (E2) exposures to ethanol, and pooled across both exposures (E1 + E2)

Analysis	Source	d.f.	MS	F	<i>P</i>	Variance Component
E1 + E2	Line (<i>L</i>)	39	1680.625	5.48	<0.0001	$\sigma_L^2 = 3.660$
	Exposure (<i>E</i>)	1	10338	52.96	<0.0001	Fixed
	Sex (<i>S</i>)	1	27936	18.12	<0.0001	Fixed
	<i>L</i> × <i>E</i>	39	197.491	4.08	<0.0001	$\sigma_{LE}^2 = 0.652$
	<i>L</i> × <i>S</i>	39	157.550	1.34	0.1935	$\sigma_{LS}^2 = 0.088$
	<i>E</i> × <i>S</i>	1	157.8	0.33	0.5670	Fixed
	<i>L</i> × <i>E</i> × <i>S</i>	39	48.468	0.64	0.9349	$\sigma_{LES}^2 = 0$
	Rep (<i>L</i> × <i>S</i>)	77	140.606	1.91	0.0025	$\sigma_{R(LS)}^2 = 0.757$
	Rep (<i>L</i> × <i>S</i> × <i>E</i>)	77	73.687	5.07	<0.0001	$\sigma_{R(LES)}^2 = 1.040$
Within (<i>W</i>)	15695	14.532			$\sigma_W^2 = 14.533$	
E1	<i>L</i>	39	981.7	8.88	<0.0001	$\sigma_L^2 = 4.20$
	<i>S</i>	1	1731.5	15.95	0.0003	Fixed
	<i>L</i> × <i>S</i>	39	110.5	0.99	0.4892	$\sigma_{LS}^2 = 0.033$
	Rep (<i>L</i> × <i>S</i>)	77	108.4	8.33	<0.0001	$\sigma_{R(LS)}^2 = 1.80$
	<i>W</i>	8330	13.01			$\sigma_W^2 = 13.01$
E2	<i>L</i>	39	892.0	9.33	<0.0001	$\sigma_L^2 = 4.48$
	<i>S</i>	1	1122.2	12.04	0.0013	Fixed
	<i>L</i> × <i>S</i>	39	95.6	0.86	0.6953	$\sigma_{LS}^2 = 0$
	Rep (<i>L</i> × <i>S</i>)	77	108.5	6.68	<0.0001	$\sigma_{R(LS)}^2 = 1.88$
	<i>W</i>	7365	16.25			$\sigma_W^2 = 16.25$

TABLE S3

SFPs associated with genetic variation in mean elution time after one (E1) and two (E2) exposures to ethanol, ethanol tolerance (ET), starvation resistance (SR), chill coma recovery (CC), fitness (FT), life span (LS), locomotor reactivity (LR), copulation latency (CL), aggression (AG), day sleep bout number (DB), day sleep time (DS), night sleep bout number (NB), night sleep time (NS) and waking activity (WA).

Table S3 is available for download as an Excel file at <http://www.genetics.org/cgi/content/full/genetics.109.107490/DC1>.

TABLE S4

Modules of correlated transcripts associated with genetic variation in mean elution time among 40 inbred wild derived lines after one (E1) and two (E2) exposures to ethanol, and tolerance (ET).

Table S4 is available for download as an Excel file at <http://www.genetics.org/cgi/content/full/genetics.109.107490/DC1>.

TABLE S5

Over-representation of Gene Ontology Categories, Pathways and Keywords for transcripts associated with ethanol sensitivity after an initial (E1) and second (E2) exposure to alcohol, and ethanol tolerance (ET).

Table S5 is available for download as an Excel file at <http://www.genetics.org/cgi/content/full/genetics.109.107490/DC1>.

TABLE S6

Enrichment of transcription factor motifs in 5' UTR sequences of genes in modules of correlated transcripts associated with sensitivity to ethanol after the initial (E1) and second (E2) exposure, and ethanol tolerance (ET).

Table S6 is available for download as an Excel file at <http://www.genetics.org/cgi/content/full/genetics.109.107490/DC1>.

TABLE S7**Over-representation of Gene Ontology Categories and KEGG Pathways for the 100 transcripts most strongly genetically correlated with focal computationally predicted genes of unknown function.**

Table S7 is available for download as an Excel file at <http://www.genetics.org/cgi/content/full/genetics.109.107490/DC1>.

TABLE S8

RT-PCR analysis of gene expression in *psq* (*psq*^{BG02818}) and *Men* (*Men*^{BG02365}) mutant lines and their co-isogenic control line (*Canton S B*).

Table S8 is available for download as an Excel file at <http://www.genetics.org/cgi/content/full/genetics.109.107490/DC1>.