DNA Microarray and Proteomic Strategies for Understanding Alcohol Action

James M. Sikela, Erik J. MacLaren, Young Kim, Anis Karimpour-Fard, Wei-Wen Cai, Jonathan Pollack, Robert Hitzemann, John Belknap, Shannon McWeeney, Robnet T. Kerns, Chris Downing, Thomas E. Johnson, Kathleen J. Grant, Boris Tabakoff, Paula Hoffman, Christine C. Wu, and Michael F. Miles

This article summarizes the proceedings of a symposium presented at the 2005 annual meeting of the Research Society on Alcoholism in Santa Barbara, California. The organizer was James M. Sikela, and he and Michael F. Miles were chairs. The presentations were (1) Genomewide Surveys of Gene Copy Number Variation in Human and Mouse: Implications for the Genetics of Alcohol Action, by James M. Sikela; (2) Regional Differences in the Regulation of Brain Gene Expression: Relevance to the Detection of Genes Associated with Alcohol-Related Traits, by Robert Hitzemann; (3) Identification of Ethanol Quantitative Trait Loci Candidate Genes by Expression Profiling in Inbred Long Sleep/Inbred Short Sleep Congenic Mice, by Robnet T. Kerns; and (4) Quantitative Proteomic Analysis of AC7-Modified Mice, by Kathleen J. Grant.

Key Words: Array-Based Comparative Genomic Hybridization, Gene Copy Number, Microarrays, Gene Expression Profiling, Alcohol-Related QTL, Proteomics, Adenylyl Cyclase.

WITH THE HUMAN genome project serving as a catalyst, high-throughput technologies are now providing insights into genetic and biological processes at a level of unprecedented detail and scope. Typical of this trend, we are seeing the emergence and increasing utilization of a "genomewide" mindset where methods are designed to simultaneously survey data sets that are representative of whole genomes. Three of the most promising of these technology platforms are array-based comparative genomic hybridization (aCGH), gene expression profiling using DNA microarrays, and the analysis of translated gene products via proteomic platforms.

Interestingly these 3 areas of focus touch on the 3 molecular levels that are part of the original central dogma of information flow in molecular biology, namely that "DNA makes RNA makes protein," and thus encompass 3 of the major target areas of alcohol and other biomedical research: the genome, transcriptome, and proteome. The talks in this symposium have been designed to be representative examples of each of these technologies and specifically to demonstrate some of the ways in which high-throughput biology can be used to advance our understanding of the molecular and cellular basis of alcohol action.

GENOMEWIDE SURVEYS OF GENE COPY NUMBER VARIATION IN HUMAN AND MOUSE: IMPLICATIONS FOR THE GENETICS OF ALCOHOL ACTION

James M. Sikela, Erik J. MacLaren, Young Kim, Anis Karimpour-Fard, Wei-Wen Cai, and Jonathan Pollack

Changes in gene copy number can have important phenotypic effects both within a species and between species (Li, 1997; Mazzarella and Schlessinger, 1998). Such changes have been shown to be important to individual variation in drug response (McLellan et al., 1997), disease risk (Gonzalez et al., 2005), and disease etiology (Singleton et al., 2003; Van Esch et al., 2005), suggesting that this mechanism may also be a contributing factor in other variable human traits, including phenotypic variation related to alcohol action in humans. One of the most powerful technologies for the detection of such copy number variation (CNV) is aCGH, which typically uses arrays of large insert genomic clones such as bacterial artificial chromosomes (BACs) (Pinkel et al., 1998) or PCR-amplified inserts of cDNA clones (Pollack et al., 1999).

Regarding between-species applications, it is well established that gene duplication is a key driving force of evolutionary change and among the most important

From the University of Colorado at Denver and Health Sciences Center, Aurora, Colorado, (EJM, AK, KJG, BT, PH, CCW, JMS); Virginia Commonwealth University Richmond, Virginia, (MFM, RTK): Baylor College of Medicine (WC), Houston, Texas; Stanford University School of Medicine, Stanford, California, (JP); Oregon Health & Science University; Portland, Oregon; and the University of Colorado (CD, TEJ), Boulder, Colorado, (RH, JB, SM).

Received for publication October 25, 2005; accepted November 28, 2005.

Supported in part by National Institutes of Health Grants AA11853 and AA08940 (JMS); AA 06243, AA 10760, AA 11034 and AA 13484 and grants from the Department of Veterans Affairs (RH); AA13678 (MFM); AA08940 and AA011984 (TEJ); AA 013489 (BT); K22 AI059076 (CCW); and the Children's Hospital Research Institute TCHRI 1009000 (JMS).

DOI: 10.1111/j.1530-0277.2006.00081.x

mechanisms for the development of new genes and biochemical processes (Li, 1997). We have previously shown how it is possible to use cDNA aCGH to carry out a genomewide survey of gene copy number differences between human and great ape lineages (Fortna et al., 2004). Using microarrays that contained 39,000 human cDNAs representing 29,000 genes, over 1,000 genes were identified that showed lineage-specific increases or decreases in copy number among these hominoid species. Among 134 genes showing human lineage-specific increases in copy number were a number thought to be important in brain structure and function.

With regard to within-species studies, cDNA aCGH was first applied to search for gene copy number changes associated with cancer (Clark et al., 2002; Pollack et al., 2002). More recently genomewide array-based and fosmid-based methods have been applied to the normal human population and these studies have shown that CNVs is much more common in the human genome than originally thought (Iafrate et al., 2004; Sebat et al., 2004; Sharp et al., 2005; Tuzun et al., 2005). Although these methods were genomewide they did not use gene-based arrays and therefore did not provide direct assessments of gene copy number change. Given the success of cDNA aCGH in cross-species studies, we have begun to explore the possibility that cDNA aCGH may also be capable of directly detecting gene CNVs within the human population. As part of a pilot study we applied cDNA aCGH to several human samples using similar arrays and methods as those used in the primate comparisons (Fortna et al., 2004). All experiments used sex-matched test and reference DNA samples except for one experiment in which the test sample was female and the reference sample male. Because of the differences in sex chromosome number, the sex-unmatched comparison contains internal controls, i.e., in this comparison, genes on the X-chromosome and Y-chromosome (but not those on the autosomes) should show copy number imbalances reflective of a single copy change. We showed that the sample that is not sex-matched had readily detectable differences in aCGH signals for genes on the X and Y chromosomes. No such patterns were evident for the autosomes of the sex unmatched individuals or for the sex chromosomes of the sex matched samples. As an additional control, the same sample was used as the reference and test in the same experiment, i.e., self-versus-self. Complementary DNAs giving signals exceeding a log2 fluorescence ratio of 0.4 were scored as positives. Of 1,463 X-chromosome genes, 46% were positive for copy number differences in the sex-unmatched sample, while $<3\%$ (42/1,463) of X-chromosome cDNAs in the self-versus-self comparison exceeded this threshold.

Using these selection criteria we found a number of genes that were promising candidates to be CNVs that have been previously reported to be CNVs in the human population (Gonzalez et al., 2005; Groot et al., 1991; Tayebi et al., 2000). Overall, our preliminary findings confirmed genes CNVs reported in the literature at a statistically significant rate compared with randomly chosen clones ($p<0.01$). These preliminary results suggest that aCGH using cDNA arrays is as effective at identifying CNVs in genomewide scans as large construct-based methods, but has the additional advantage of being gene-based. This method, therefore, generates higher resolution results than are obtainable using BACs or fosmids and allows CNVs to be defined in greater detail. Once optimized, this methodology has the potential to be able to quickly carry out genomewide searches for CNVs between human populations that differ in any genetically influenced trait, including those related to alcohol action and response. While these results are promising, we are currently testing the method on additional samples, independently confirming the results we have obtained, and evaluating additional aCGH platforms and methods for selection of positives to further optimize this technology.

Finally, we asked if aCGH could be used to identify copy number changes between the ILS and ISS lines of mice, which have been selectively bred for differences in alcohol sensitivity. To address this possibility a pilot study was carried out using murine microarrays that contained $>19{,}000$ BAC probes. Hybridization conditions were as described previously (Li et al., 2004) with C57BL/6J (B6) used as the reference DNA. Suggestive CNVs between ILS and ISS mice were identified for several regions in the mouse genome, and several of these were also found to have CNVs in some of the LS/SS progenitor strains, e.g., DBA/2J and B6. These results suggest that mouse BAC aCGH should provide a valuable compliment to quantitative trait loci (QTL)-mapping and single nucleotide polymorphism-based strategies that seek to identify important genomic differences related to alcohol action.

REGIONAL DIFFERENCES IN THE REGULATION OF BRAIN GENE EXPRESSION: RELEVANCE TO THE DETECTION OF GENES ASSOCIATED WITH ALCOHOL-RELATED TRAITS

Robert Hitzemann, John Belknap, and Shannon McWeeney

Hitzemann et al. (2003) used the BXD recombinant inbred (RI) series to examine the genetic relationship(s) between the D_2 dopamine receptor and ethanol preference phenotypes. There was considerable evidence suggesting that a relationship would be detected. For example, D_2 receptor–deficient mice show marked reductions in ethanol preference, ethanol self-administration and ethanol-conditioned place preference (Cunningham et al., 2000; Phillips et al., 1998; Risinger et al., 2000). The consumption-preference phenotype has been repeatedly mapped to a region on mouse chromosome 9 that includes the *Drd2* locus (Belknap and Atkins, 2001); the place preference phenotype maps to a similar region (Cunningham, 1995). Three independent studies have shown in both mice and rats that using viral transfer to increase D_2 receptor

density reduces ethanol preference suggesting (when compared with the knockout data) that there is an inverted U relationship between ethanol consumption and receptor density (Thanos et al., 2001, 2004, 2005). Hitzemann et al. (2003) included in the RI data analysis 2 independently obtained data sets on D_2 receptor density (one from Jones et al., 1999) and a publicly available data set on *Drd2* expression (www.WebQTL.org). In summary, the results obtained indicated that *Drd2* expression but not D2 receptor density (striatum, accumbens, midbrain) was significantly correlated to both ethanol preference (2-bottle choice) and ethanol-conditioned place preference. The latter observation has been independently confirmed (Chesler et al., 2005); in addition these authors found that *Drd2* expression was partially *cis*-regulated suggesting a point of integration between the preference QTL and gene expression data. However, the disconnect between the gene expression and receptor density data which has been frequently observed (see, e.g., Qian et al., 1993) confounds our interpretation of the expression-preference correlation. After careful consideration of this issue and building especially upon the results of both Walker et al. (2004) and Chesler et al. (2005), we concluded (a) that the *Drd2* expression-preference correlation was not a chance event and (b) that the expression data must have additional information content relevant to understanding the expression-preference correlation. To begin addressing this second point, we initiated a series of experiments to characterize the gene network that is coregulated with *Drd2*. We assumed that the network was likely to contain elements of the D_2 receptor signaling pathway(s). Pilot data supporting this hypothesis were obtained in a whole brain B6D2 F_2 gene expression dataset which in turn prompted characterizing the gene network in a striatal data set; the results from both data sets are presented here. To determine the specificity of the *Drd2* data, we also characterized the *Drd1a* expression network. Overall, the results obtained illustrate that for this particular application the striatal data set is far superior.

The data sets entered in the analyses may be summarized. The whole brain data set consisted of 6 each brains from male C57BL/6J (B6) and DBA/2J (D2) mice and 56 brains from a reciprocal B6D2 F_2 intercross; the striatal data set was composed of 5 each male and female samples from the B6 and D2 strains and 58 B6D2 F_2 intercross samples. The B6 and D2 striatal samples were each pooled from 2 brains. The striatum was dissected from a 1.5-mm tissue slice; the caudal boundary of the slice was the optic chiasm; and thus, the sample contained both the caudate-putamen and the nucleus accumbens. For the whole brain samples the Affymetrix 430 A and B arrays were used; for the striatal data set the 430 2.0 array was used. For the low-level analysis of the array data both Robust Microarray Average (Irizarry et al., 2003) and Position Dependent Nearest Neighbor (PDNN) (Zhang et al., 2003) were used. The data discussed here are from the PDNN analysis; the RMA analysis was quantitatively very similar. Additional details of the analysis strategy are found in Hitzemann et al. (2004). The F_2 animals were genotyped using a standard panel of microsatellites (see Demarest et al., 2001). eQTL data were calculated using R/qtl (Broman et al., 2003). The genomewide LOD threshold for a significant eQTL was 3.7; the threshold for a suggestive eQTL was 2.3. Statisitical analysis of the data used standard procedures. The genomewide threshold for a significant correlation between 2 transcripts was set at $r = 0.635$ ($p < 2 \times 10^{-6}$) and assumed that for the 430 array(s) there are 25,000 independent tests.

Several features of the expression data sets can be summarized:

1. There was no difference between the B6 and D2 strains in whole brain or striatal *Drd1a* or *Drd2* gene expression. Combining male and female striatal samples, *Drd2* expression trended to a significant difference (genomewide false discovery rate < 0.054) with expression higher in the B6 strain. These data are consistent with the observation that D_2 receptor density is modestly higher in the B6 strain (Kanes et al., 1993).
2. In contrast to the strain differences, the ranges of differences of *Drd1a* and *Drd2* expression in the F_2 data sets were marked. The ranges of variation in the whole brain *Drd1a* and *Drd2* expression were 50 and 85%, respectively. The ranges of variation in striatal *Drd1a* and *Drd2* expression were 190 and 285%, respectively. As assessed by a *F* test, the increased variances in the striatal sample were significantly different from those in the whole brain sample; the associated *p* values were 3.0×10^{-17} for *Drd1a* and 1.6×10^{-12} for *Drd2*. Some of this increased variation likely results from technical factors such as the variance associated with tissue dissection; however, given the data described below, most of the increased variance must be biological.
3. In neither the whole brain nor the striatal data set did the correlation between *Drd1a* and *Drd2* expression reach the genomewide threshold for significance. In the striatum the correlation was 0.535. Thus, we conclude that to a large extent *Drd1a* and *Drd2* expression are regulated by independent factors.
4. The *Drd1a* and *Drd2* expression data were entered into a genomewide eQTL analysis. In neither the whole brain nor the striatal sample did a *Drd1a* or *Drd2* eQTL meet the genomewide level of significance (LOD $\geq$ 3.7). Suggestive eQTLs (LOD $\geq$ 2.3) were detected in the whole brain sample for *Drd1a* (Chr 17) and for *Drd2* (Chr 5, 7, and 9). For the striatal sample, suggestive eQTLs were detected for *Drd1a* (Chr 10 and 11) and for *Drd2* (Chr 9). Using a 2-site model, the entire genome (striatal data) was scanned for modifier loci to the eQTL for *Drd2* on Chr 9. A significant interaction ($p < 10^{-4}$) was found on Chr 15 (*D15Mit8*) such that the *Drd2* LOD increased to >4 when the

modifier genotype is B6B6 but decreased to <1 for the B6D2 and D2D2 genotypes. The distribution of the maximum eQTLs for all transcripts on the 430 array was also determined. In the whole brain, the maximum LOD scores were clustered on Chr 12, 19, and X; however, in the striatal sample, nearly 50% of the maximum LOD scores were clustered on Chr 11.

To characterize the associated gene networks, the *Drd1a* and *Drd2* F_2 expression data were entered into genomewide correlation analyses. In the whole brain sample, there were 37 nonredundant transcripts significantly associated with *Drd1a* expression and 22 significantly associated with *Drd2* expression. Notable genes associated with *Drd1a* expression included *Strn* ($r = 0.654$), *Adora2A* ($r = 0.688$), and *Adcy5* ($r = 0.833$); genes significantly associated with *Drd2* expression included *Adcy5* ($r = 0.741$), *Adora2A* ($r = 0.776$), *Penk1* ($r = 0.805$), *Rgs9* ($r = 0.770$), *Pde1b* ($r = 0.885$), and *Ppp1r1b* ($r = 0.760$). It was of some interest to note that the expression of *Ppp1r1b* (which encodes for the protein commonly known as DARPP-32) was significantly associated with *Drd2* but not *Drd1a* expression. For the striatal F_2 sample the number of nonredundant transcripts significantly associated with *Drd1a* and *Drd2* expression were 295 and 480, respectively. There were only 31 transcripts that showed an overlap between *Drd1a* and *Drd2* expression clusters, again illustrating the independence of the factors that regulate *Drd1a* and *Drd2* expression. Analysis of the striatal sample detected 16 transcripts that were associated with *Drd2* expression at $r > 0.90$; known genes included in this group were *Ptpn5, Pde1b, Ppp1r1b, Adcy5, Gng7, Rgs9, Actn2, Foxo1, Smpd3,* and *Kcnip2*. Overall and as predicted, the striatal *Drd2* expression network contained numerous members of the dopamine signaling cascade.

The correlations associated with the *Drd2* gene cluster are phenotypic correlations and reflect the joint action of genetic and environmental effects. It was possible to estimate the genetic correlation between *Drd2* expression and the members of the gene cluster by using the genomewide LOD scores (from Chr 1 to X); this is a genetic correlation because the LOD scores represent QTL effects on expression of both traits. These genetic correlations agreed with the phenotypic correlations at $r = 0.986$; the average phenotypic correlation was 0.721 and the average genotypic correlation was 0.648. For the *Drd1a* gene cluster the phenotypic/genotypic correlation was 0.957; the average phenotypic correlation was 0.705 and the average genetic correlation was 0.486. The difference between the *Drd1a* genetic and phenotypic correlations was significant ($p < 0.01$).

The distribution of the max eQTL LODs associated with the *Drd1a* and *Drd2* gene clusters was determined. For the *Drd1a* cluster, the highest density was found on Chr 11 (33%); however, this was a significant decrease ($p < 7 \times 10^{-8}$) from the genomewide distribution of all transcripts; other chromosomes showing a higher density than the genomewide average were Chr 10 (14.9% vs 2.9%) and Chr 17 (13.0% vs 2.7%). The highest density of peak LOD scores associated with the *Drd2* cluster was found on Chr 9 (29.8%), which was significantly different from the distribution for all transcripts (2.9%, $p \ll 10^{-10}$). Of the 148 transcripts which mapped to Chr 9, only 18 were *cis*-regulated; i.e., 130 of the transcripts were located on chromosomes other than Chr 9. Members of this trans-regulated; cluster included *Actn2, Adcy5, Adra2c, Gripap1, Pde10a, Pde1b, Penk1, Ptpn5*, and *Rgs9*. A high density of peak *Drd2* associated LOD scores was also found on Chr 11 (22.3%) but this was significantly less than the density for all transcripts ($p \ll 10^{-10}$). Similarly, a significant enrichment of *Drd2* associated peak LOD scores was found on Chr 13 [13.5% vs 2.9% (for all transcripts), $p \ll 10^{-10}$].

The study described here was initiated with the observation that ethanol preference was associated with *Drd2* expression (Chesler et al., 2005; Hitzemann et al., 2003). Further, preference QTLs had been repeatedly detected on Chr 9 in the general region of the *Drd2* locus (Belknap and Atkins, 2001). We asked what is the significance of the variation in *Drd2* expression and what are the characteristics of the genes that cluster with *Drd2* expression? The data described above briefly illustrate that *Drd2* expression is part of much larger gene network that contains numerous members of the signaling cascade(s) associated with dopaminergic systems and the related DARPP-32 signaling network. Some evidence of these associations was detected in the whole brain data set. However, the vast scope of the gene cluster was evidenced only in the striatal data set. The question naturally arises as to whether or not if one focused on more discrete regions within the striatum, e.g., nucleus accumbens versus dorsomedial striatum or accumbens shell versus core, would one detect different *Drd2* gene clusters? To answer this question, reliable methods that uniformly amplify small amounts of mRNA will be needed. Overall, we conclude that the observation that ethanol preference is associated with *Drd2* expression must now be viewed in the context of the *Drd2* expression network.

IDENTIFICATION OF ETHANOL QUANTITATIVE TRAIT LOCI CANDIDATE GENES BY EXPRESSION PROFILING IN INBRED LONG SLEEP/INBRED SHORT SLEEP CONGENIC MICE

Robnet T. Kerns. Chris Downing, Thomas E. Johnson, and Michael F. Miles

The long-sleep and short-sleep (LS and SS) mice are a well-characterized model selectively bred for differential sensitivity to a hypnotic dose of ethanol (McClearn and Kakihana, 1981). Using an F_2 intercross derived from inbred LS and SS mice (ILS and ISS), we identified several quantitative trait loci (QTLs) mediating loss of

righting because of ethanol (LORE; Markel et al., 1997). Four of the largest effect *Lore* QTLs were confirmed by creating reciprocal congenic strains (Bennett et al., 2002a) and subsequent fine mapping using interval-specific congenic recombinants (ISCRs: Bennett et al., 2002b). However, limited progress has been made using classical genetic techniques (i.e., sequencing coding regions of candidate genes) in the search for genes underlying the *Lore* QTLs (Ehringer et al., 2001, 2002).

Alternatively, many QTLs may result from polymorphisms in regulatory regions of genes, as opposed to coding regions. These candidate genes can be identified by quantifying mRNA levels. However, this search has been slow because traditional molecular biology approaches to quantify mRNA levels generally target 1 or a few genes. The recent development of DNA microarrays has allowed for the nonbiased, simultaneous measurement of the expression of thousands of genes or expressed sequence tags (ESTs). Array studies provide a method for identifying genes that are differentially expressed in a QTL region on a large scale that also allows for the detection of gene networks. It is this approach that might provide insight to neurobiological pathways and mechanisms of action at the molecular level and therefore assign putative biological function to the QTL region.

The *Lore5* QTL region on mouse chromosome 15 (between *D15Mit150*, 12.10 cM or 14.04 Mb; and *D15Mit79*, 66.2 cM or 103.5 Mb), was introgressed onto both ILS and ISS backgrounds to create the ILS.ISS-Lore5 (ISS QTL on ILS background) and ISS.ILS-Lore5 (ILS QTL on an ISS background) congenics, respectively (Bennett et al., 2002a, 2002b). We used Affymetrix oligonucleotide microarrays (GeneChip^{TM} MG_U74Av2, Bv2, and Cv2) to examine expression profiles of $>35{,}000$ genes in ISS, ILS, and *Lore5* congenics in the prefrontal cortex (PFC), to identify potential candidate genes for the *Lore5* QTL. We compared both the basal expression differences and the gene expression 4 hours after 4.1 g/kg ethanol injection of the progenitor and congenic strains to identify candidate quantitative trait genes (QTGs) and gene networks that might be related to biological mechanism(s) for the behavioral QTL.

Using a stringent, stepwise methodology (Kerns et al., 2005) we identified 573 genes in PFC with either differential basal expression between strains or differential expression in response to an acute sedative dose of ethanol. We found that 21 of these 573 selected genes were located within the *Lore5* region. Hierarchical clustering of these candidate genes showed distinct patterns of gene expression consistent with the *Lore5* genotype. We used a simple filtering method to summarize changes in expression of the 573 selected genes. Genes with an average S-score (Kerns et al., 2003; Zhang et al., 2002) over 3 observations of ≥ 1.5 or -1.5 or less were counted as changed in a category, e.g., ILS ethanol response. We observed a larger general ethanol response in ILS than ISS mice. This was expected because the ILS strain is more sensitive than the ISS strain to the sedative effects of ethanol. We have previously observed that more genes were regulated by 2 g/kg acute ethanol in DBA/2J mice, which are more sensitive to the acute activating effects of ethanol, than C57BL/6J mice (Kerns et al., 2005). A larger number of changes in basal gene expression were observed between the progenitor strains than between the congenics and their background controls. This is expected, as the congenic strains are genetically identical to the progenitor strain, except for the *Lore5* region. Using this filtering approach, we also counted the number of genes with differential basal expression between the progenitor strains, and between both congenic strains compared with their respective background controls, i.e., ILS versus ISS, ISS.ILS-*Lore5* versus ISS, and ILS.ISS-*Lore5* versus ILS. We found only 11 genes with differences in gene expression consistent with the *Lore5* genotype in each comparison, and only 1, N-myc downstream-regulated gene 1 (*Ndrg1*), located in the *Lore5* region. Multivariate analysis (k-means clustering) was used to group genes based on common expression patterns and identified a subcluster of genes with basal expression correlating to the genotype of *Lore5*, but with ethanol responsiveness correlating to the background genotype. *Ndrg1* was the only gene present in this subcluster and located in the *Lore5* interval. Two myelin-related genes, proteolipid protein (*Plp*) and myelin-associated oligodendrocytic basic protein (*Mobp*), coexpressed consistent with the *Lore5* genotype in each strain comparison and with *Ndrg1*. In our previous study in DBA/2J PFC we found that *Ndrg1* and myelin-related genes are coregulated by 2 g/kg acute ethanol. Thus, Ndrg1 and myelin gene expression may represent a common signaling mechanism in ethanol behaviors of animals with high sensitivity to both low and high doses of ethanol. We confirmed the expression of Ndrg1 using Northern blot analysis and showed this gene has qualitatively increased abundance in saline-treated ILS.ISS-*Lore5* and ISS, compared with ISS.ILS-*Lore5* and ILS PFC.

We postulated that expression profiling of ethanol responses in *Lore5* congenic and progenitor strains may also help identify candidate genes for the *Lore5* QTL. Instead of examining the ethanol responsive genes in each strain, we used Post hoc Template Matching (PTM) (Pavlidis, 2003) to identify genes with ethanol responsive patterns with high Pearson correlations ($p<0.001$) to the *Lore5* genotype. Of the 573 selected genes, we identified 29 with patterns of ethanol responsive expression linked to the *Lore5* genotype. Only 1 of these 29 genes, RAN GTPase–activating protein 1 (*Rangap1*), is located in the *Lore5* interval.

The WebQTL resource (http://www.webqtl.org/) includes expression data of $\sim$100 forebrain samples from >30 B6$\times$D2 RI lines and can be used to link expression profiles with ethanol behaviors by correlating gene expression to genetic markers and behavioral QTL data in the same database. We found that *Rangap1* had a significant

positive Spearman's correlation ($p = 0.00114$, $R = 0.633$) with high-dose ethanol loss-of-righting-reflex phenotype (Rodriguez et al., 1995) in the WebQTL database (U74Av2 (Mar04) RMA Orig.). Thus, *Rangap1* is located in the *Lore5* interval and has ethanol responsive expression that correlates with the *Lore5* genotype in ILS and ISS progenitor and congenic strains, and its basal expression correlates with the LORE phenotype in the BXD panel.

We used functional class scoring (FCS) (Pavlidis et al., 2004) to identify potential biological networks associated with ethanol-regulated gene patterns correlating to the *Lore5* genotype in an attempt to assign biological function to the *Lore5* QTL. Functional class scoring uses the statistical distribution of expression information of genes present on an array, instead of only significant genes selected by an arbitrary methodology, and therefore detects subtle changes in functional classes that might otherwise be overlooked. To identify functional classes of genes whose ethanol response correlates with the *Lore5* region, we applied FCS to *p* values from PTM of ethanol responsive gene expression to the *Lore5* genotype. Major significant groups identified by FCS included ubiquitin-dependent protein catabolism, neurotransmitter secretion, and mRNA splicing.

Using a combination of functional genomics and bioinformatics methods, 2 candidate genes for the *Lore5* were identified: *Ndrg1* and *Rangap1*. Both genes are located in the *Lore5* support interval, and *Rangap1* expression correlates with LORE behaviors in the BXD RI panel. Bioinformatics methods were used to assign putative biological functions to the *Lore5* QTL. Overall, these studies identified individual candidate genes and gene networks for the *Lore5* QTL and describe the application of expression profiling for the identification of genes underlying behavioral QTLs.

QUANTITATIVE PROTEOMIC ANALYSIS OF AC7-MODIFIED MICE

Kathleen J. Grant, Boris Tabakoff, Paula Hoffman, and Christine C. Wu

Currently, 3 candidate systems are known to contribute to the etiology of generalized anxiety and alcoholism including the $GABA_A$ receptor system, CRF systems, and glutamate receptor systems. We are interested in the role of adenylyl cyclase signaling in these systems, especially those signals serving to modulate the receptors/ion channels of the glutamate system (e.g., NMDA).

Previously, an association study (subjects recruited in Bethesda, MD, M-A. Enoch, NIAAA) was conducted which indicated that a tetranucleotide repeat polymorphism in the 3′ untranslated region of the adenylyl cyclase 7 (ADCY7) gene was associated with risk for anxiety in males, depending on its length. Animal models were then created, in which mice were genetically modified to over- or underexpress ADCY7 in the brain, and these animals show behavioral phenotypes that reflect the situation found in humans.

We know that alcohol is an anxiolytic in humans as well as other animals, and we also know that anxiety may promote alcohol intake (Carrigan and Randall, 2003) and thus serve as a form of negative reinforcement, such that someone who is predisposed to anxiety may drink to alleviate the anxiety that he or she is feeling. B. Tabakoff (unpublished data) has indicated differences in the untreated, baseline anxiety behavior of the AC7 transgenic (tg) and wild-type (wt) mice as well as suggesting differences in severity of withdrawal following chronic ethanol treatment.

Anxiety is a complex behavior with multiple genes contributing to the resulting behavioral phenotype. The objective of the current work was to investigate and characterize the anxiety-related behavioral differences seen in these animals, before and following chronic ethanol treatment, by utilizing 3 global, nonbiased, investigation platforms.

First, QTL analysis was used to identify QTLs for anxiety using current published literature and yielded several loci mapping to 9 different chromosomes (Henderson et al., 2004; Flint et al., 1995). This information combined with the second global approach, microarray analysis, yielded a list of differentially expressed genes which mapped to the identified QTL regions for anxiety. Using the MAGIC-B (Microarray Ascertainment of Genes Involved in Complex Behaviors) method to compare the naïve C57BL/6 wt and AC7 tg mice allowed identification of differentially expressed genes by statistical methods and mapping of the gene location on the mouse genome. The obtained QTL information was then used as a "filter" to choose differentially expressed genes that map to the QTL regions for anxiety and these genes were then hypothesized to be predisposing to the anxiety behavior phenotype and could be utilized in a "Pathway Assist" or other pathway analysis.

The results of the microarray analysis yielded several hundred differentially expressed genes and when these results were correlated with the QTL data, 8 of these differentially expressed genes were found to map back to a QTL for anxiety with a statistical significance of $p < 0.05$ (B. Tabakoff, unpublished data). *In silico* analysis revealed several interesting functional nodes, including calcium and glutamate signaling, and one primary subcellular protein localization, the synaptic membrane.

The third, global, nonbiased approach that was used to investigate this complex behavior was shotgun proteomics. Enriched synaptosomal samples prepared from the brains of naïve wt and AC7 tg mice were digested using proteases and the resulting peptides were analyzed using Multidimensional Protein Identification Technology (MudPIT) (Wu et al., 2003). The proteins that were identified included 5 of the 8 proteins that mapped back to the QTL for anxiety at a $p < 0.05$. The remaining 3 unidentified proteins are not predicted to have synaptosomal localization and

thus we would not expect them to show enrichment in our samples.

We then wanted to use a nonbiased, quantitative proteomic technique, which would enable us to evaluate quantitative differences in protein levels between our wt and tg samples. Using metabolic labeling with ^{15}N would accomplish this and allow us to identify proteins already known to be differentially expressed as well as having the potential to identify proteins which are not yet known to be differentially expressed and potentially involved in the behavior differences that we see between the 2 phenotypes. Wild-type mice were fed a diet enriched in ^{15}N-labeled proteins to generate ^{15}N-labeled brain tissue for use as internal standards. Animals were then sacrificed, and labeled brains were cofractionated with both unlabeled wt and tg brains. Mixed labeled/unlabeled synaptosomal samples resulting from each fractionation were analyzed using MudPIT. Once a peptide was identified in both samples and the internal standard, the relative abundance ratio of wt sample to standard (wt/std) and also tg sample to standard (tg/std) was measured. The ratio of these 2 ratios [(wt/std)/(tg/std)] was taken, resulting in the fractional change between samples (Wu et al., 2004). Two MudPIT runs were used to create the resulting list of proteins, in which a minimum of 2 unique peptides were required for identification and the program RelEx (MacCoss et al., 2003), was used to automatically generate the list. A total of 151 proteins were quantified between the wt and tg samples. Of these, 26 showed significant differences between wt and tg samples with $p < 0.1$. Five of the 26 proteins were found to have a decreased level in AC7 tg animals compared with wt, whereas 21 of the 26 were found to have an increased level compared with wt.

From this preliminary, quantitative, proteomic analysis, several interesting proteins were identified, including at least 3 related to calcium transport and signaling. These proteins may prove to be functionally significant especially when considering the functional nodes predicted to be involved by *in silico* analysis. Even more promising was the identification of a protein that was differentially expressed, can be mapped back to the QTL and was previously shown to be involved in anxiety-related behavior. In conclusion, 3 global, nonbiased approaches were used to study the behavior of the naïve, AC7 genetically modified mice. The ability to relate all 3 methodologies will be the key in the future for determination of pathways contributing to complex phenotypes such as anxiety.

CONCLUSIONS

In summary, the results of this symposium illustrate how a number of high-throughput, genomewide technologies are being applied to biological questions related to the action of alcohol. Sikela et al. describe a genomewide method for identifying gene and segmental CNVs both within species and between species. Using aCGH, structural variations have been discovered in the genome of a mouse model of ethanol sensitivity, and this technology shows promise for identifying CNPs related to alcohol phenotypes in humans as well. Hitzemann et al. use genomewide expression assays in conjunction with computational analysis to explore the relationship between dopamine receptors and alcohol. They identify an extensive network of genes linked to *Drd2* expression and thereby demonstrate how seemingly contradictory experimental data can be reconciled by taking a more global view of expression data. Kerns et al. address the challenge of finding genes underlying the *Lore* QTLs. Using expression arrays, congenic mouse lines and bioinformatic tools, they propose 2 new candidate genes for *Lore5* and describe gene networks linked to the QTL. Lastly, Grant et al. combine *in silico*, genomic and proteomic strategies to study genetically modified mice that model anxiety phenotypes in humans. Their study demonstrates how global, unbiased strategies can be used to identify several candidate genes and proteins with potential links to anxiety-related behaviors, pathways and QTLs.

Overall, the methodologies used here promise to generate a bounty of useful data regarding the biological effects of ethanol, and making the most effective use of this wealth of data will require correspondingly robust and innovative bioinformatics resources. These will not only need to efficiently handle each of the individual data sets but, perhaps more importantly, will need to be able to integrate them with one another in meaningful ways. This type of integration, which to be effective will require a level of sophistication that matches the level of biological complexity of the system under study, is likely to be among the most important and challenging issues to emerge from high-throughput biology in the coming years.

REFERENCES

Belknap JK, Atkins AL (2001) The replicability of QLTs for murine alcohol preference drinking behavior across eight independent studies. Mamm Genome 12:893–899.

Bennett B, Beeson M, Gordon L, Carosone-Link P, Johnson TE (2002b) Genetic dissection of quantitative trait loci specifying sedative/hypnotic sensitivity to ethanol: mapping with interval-specific congenic recombinant lines. Alcohol Clin Exp Res 26:1615–1624.

Bennett B, Beeson M, Gordon L, Johnson TE (2002a) Reciprocal congenics defining individual quantitative trait loci for sedative/hypnotic sensitivity to ethanol. Alcohol Clin Exp Res 26:149–157.

Broman KW, Wu H, Sen S, Churchill GA (2003) R/qtl: QTL mapping in experimental crosses. Bioinformatics 1:889–890.

Carrigan MH, Randall CL (2003) Self-medication in social phobia: a review of the alcohol literature. Addict Behav 28:269–284.

Chesler EJ, Lu L, Shou S, Qu Y, et al (2005) Complex trait analysis of gene expression uncovers polygenic and pleitropic networks that modulate nervous system function. Nat Genet 37:233–242.

Clark J, Edwards S, John M, Flohr P, Gordon T, Maillard K, Giddings I, Brown C, Bagherzadeh A, Campbell C, Shipley J, Wooster R, Cooper CS (2002) Identification of amplified and expressed genes in breast cancer by comparative hybridization onto microarrays of

randomly selected CDNA clones. Genes Chromosomes Cancer 34: 104–114.

Cunningham CL (1995) Localization of genes influencing ethanol–induced conditioned place preference and locomotor activity of BXD recombinant inbred mice. Psychopharmacology 120:28–41.

Cunningham CL, Howard MA, Gill SJ, Rubenstein M, et al (2000) Ethanol-conditioned place preference is reduced in dopamine D2 receptor deficient mice. Pharmacol Biochem Behav 67:693–699.

Demarest K, Koyner J, McCaughran J, Cipp L, Hitzemann R (2001) Further characterization and high-resolution mapping of quantitative trait loci for ethanol-induced locomotor activity. Behav Gen 31: 79–91.

Ehringer MA, Thompson J, Conroy O, Xu Y, Yang F, Canniff J, Beeson M, Gordon L, Bennett B, Johnson TE, Sikela JM (2001) High-throughput sequence identification of gene coding variants within alcohol-related QLTs. Mamm Genome 12:657–663.

Ehringer MA, Thompson J, Conroy O, Yang F, Hink R, Bennett B, Johnson TE, Sikela JM (2002) Fine mapping of polymorphic alcohol-related quantitative trait loci candidate genes using interval-specific congenic recombinant mice. Alcohol Clin Exp Res 26:1603–1608.

Flint J, Corley R, DeFries JC, Fulker DW, Gray JA, Miller S, Collins AC (1995) A simple genetic basis for a complex psychological trait in laboratory mice. Science 269:1432–1435.

Fortna A, Kim Y, MacLaren E, Marshall K, Hahn G, Meltesen L, Brenton M, Hink R, Burgers S, Hernandez-Boussard T, Karimpour-Fard A, Glueck D, McGavran L, Berry R, Pollack J, Sikela JM (2004) Lineage-specific gene duplication and loss in human and great ape evolution. PLOS Biol 2:937–954.

Gonzalez E, Kulkarni H, Bolivar H, Mangano A, Sanchez R, Catano G, Nibbs RJ, Freedman BI, Quinones MP, Bamshad MJ, Murthy KK, Rovin BH, Bradley W, Clark RA, Anderson SA, O'Connell RJ, Agan BK, Ahuja SS, Bologna R, Sen L, Dolan MJ, Ahuja SK (2005) The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. Science 1101–160.

Groot PC, Mager WH, Frants RR (1991) Interpretation of polymorphic DNA patterns in the human alpha-amylase multigene family. Genomics 10:779–785.

Henderson ND, Turri MG, DeFries JC, Flint J (2004) QTL analysis of multiple behavioral measures of anxiety in mice. Behav Genet 34: 267–293.

Hitzemann R, Hitzemann B, Rivera S, Gately J, Thanos P, Shou LL, Williams RW (2003) Dopamine D2 receptor binding, Drd2 expression and the number of dopamine neurons in the BXD recombinant inbred series: genetic relationships to alcohol and other drug associated phenotypes. Alcohol Clin Exp Res 27:1–11.

Hitzemann R, Reed C, Malmanger B, Lawler M, Hitzemann B, Cunningham B, McWeeney S, Belknap J, Harrington C, Buck K, Phillips T, Crabbe J (2004) On the integration of alcohol-related quantitative trait loci and gene expression analysis. Alcohol Clin Exp Res 28: 1437–1448.

Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C (2004) Detection of large-scale variation in the human genome. Nat Genet 36:949–951.

Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 4:249–264.

Jones BC, Tarantino LM, Rodriguez LA, Reed CL, McClearn GE, Plomin R, Erwin GV (1999) Quantitative-trait loci analysis of cocaine-related behaviors and neurochemistry. Pharmacogentics 9:607–617.

Kanes S, Hitzemann BA, Hitzemann RJ (1993) On the relationship between D2 receptor density and neuroleptic-induced catalepsy among eight inbred mouse strains. J Pharmacol Exp Ther 267:538–547.

Kerns RT, Ravindranathan A, Hassan S, Cage MP, York T, Sikela JM, Williams RW, Miles MF (2005) Ethanol-responsive brain region expression networks: implications for behavioral responses to acute ethanol in DBA/2J versus C57BL/6J mice. J Neurosci 25:2255–2266.

Kerns RT, Zhang L, Miles MF (2003) Application of the *S*-score algorithm for analysis of oligonucleotide microarrays. Methods 31: 274–281.

Li WH (1997) *Molecular Evolution*. Sinauer Associates Inc., Sunderland MA.

Li J, Jiang T, Mao JH, Balmain A, Peterson L, Harris C, Rao PH, Havlak P, Gibbs R, Cai WW (2004) Genomic segmental polymorphisms in inbred mouse strains. Nat Genet 36:952–954.

MacCoss MJ, Wu CC, Liu H, Sadygov R, Yates JR III (2003) A correlation algorithm for the automated quantitative analysis of shotgun proteomics data. Anal Chem 75:6912–6921.

Markel PD, Bennett B, Beeson M, Gordon L, Johnson TE (1997) Confirmation of quantitative trait loci for ethanol sensitivity in long-sleep and short-sleep mice. Genome Res 92:92–99.

Mazzarella R, Schlessinger D (1998) Pathological consequences of sequence duplications in the human genome. Genome Res 8:1007–1021.

McClearn GE, Kakihana R (1981) Selective breeding for ethanol sensitivity: short-sleep and long-sleep mice, in *Development of Animal Models as Pharmacogenetic Tools*. DHHS Publication No. ADM 81–113 (McClearn GE, Deitrich RA, Erwin VG eds), pp 147–159. U.S. Government Printing Office, Washington, DC.

McLellan RA, Oscarson M, Seidegard J, Evans DA, Ingelman-Sundberg M (1997) Frequent occurrence of CYP2D6 gene duplication in saudi arabians. Pharmacogenetics 7:187–191.

Pavlidis P (2003) Using ANOVA for gene selection from microarray studies of the nervous system. Methods 31:282–289.

Pavlidis P, Qin J, Arango V, Mann JJ, Sibille E (2004) Using the gene ontology for microarray data mining: a comparison of methods and application to age effects in human prefrontal cortex. Neurochem Res 29:1213–1222.

Phillips TJ, Brown KJ, Burkhart-Kasch S, Wenger CD, Kelly MA, Rubinstein M, Grandy DK, Low MJ (1998) Alcohol preference and sensitivity are markedly reduced in mice lacking dopamine D2 receptors. Nat Neurosci 1:610–615.

Pinkel D, Segraves R, Sudar D, Clark S, Poole I, Kowbel D, Collins C, Kuo WL, Chen C, Zhai Y, Dairkee SH, Ljung BM, Gray JW, Albertson DG (1998) High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. Nat Genet 20:207–211.

Pollack JR, Perou CM, Alizadeh AA, Eisen MB, Pergamenschikov A, Williams CF, Jeffrey SS, Botstein D, Brown PO (1999) Genome-wide analysis of DNA copy-number changes using cdna microarrays. Nat Genet 23:41–46.

Pollack JR, Sorlie T, Perou CM, Rees CA, Jeffrey SS, Lonning PE, Tibshirani R, Botstein D, Borresen-Dale AL, Brown PO (2002) Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. Proc Natl Acad Sci U S A 99:12963–12968.

Qian Y, Hitzemann B, Yount G, White J, Hitzemann R (1993) D1 and D2 dopamine receptor turnover and D2 mRNA levels in the neuroleptic responsive (NR) and neurolpetic non-responsive (NNR) lines of mice. J Pharmacol Exp Ther 267:1582–1590.

Risinger FO, Freeman PA, Rubenstein M, Low MJ, Grandy DK (2000) Lack of operant ethanol self-administration in dopamine D2 receptor deficient mice. Psychopharmacology 152:343–350.

Rodriguez LA, Plomin R, Blizard DA, Jones BC, McClearn GE (1995) Alcohol acceptance, preference, and sensitivity in mice. II. Quantitative trait loci mapping analysis using BXD recombinant inbred strains. Alcohol Clin Exp Res 19:367–373.

Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Maner S, Massa H, Walker M, Chi M, Navin N, Lucito R, Healy J, Hicks J, Ye K, Reiner A, Gilliam TC, Trask B, Patterson N, Zetterberg A, Wigler M (2004) Large-scale copy number polymorphism in the human genome. Science 305:525–528.

Sharp AJ, Locke DP, McGrath SD, Cheng Z, Bailey JA, Vallente RU, Pertz LM, Clark RA, Schwartz S, Segraves R, Oseroff VV, Albertson

DG, Pinkel D, Eichler EE (2005) Segmental duplications and copy-number variation in the human genome. Am J Hum Genet 77:78–88.

Singleton AB, Farrer M, Johnson J, Singleton A, Hague S, Kachergus J, Hulihan M, Peuralinna T, Dutra A, Nussbaum R, Lincoln S, Crawley A, Hanson M, Maraganore D, Adler C, Cookson MR, Muenter M, Baptista M, Miller D, Blancato J, Hardy J, Gwinn-Hardy K (2003) Alpha-Synuclein locus triplication causes parkinson's disease. Science 302:841.

Tayebi N, Park J, Madike V, Sidransky E (2000) Gene rearranagement on 1q21 introducing a duplication of the glucocerebrosidase pseudogene and a metaxin fusion gene. Hum Genet 107:400–403.

Thanos PK, Rivera SN, Weaver K, Grandy DK, Rubinstein M, Umegaki H, Wang GJ, Hitzemann R, Volkow ND (2005) Dopamine D2R DNA transfer in dopamine D@ receptor deficient mice: effects on ethanol drinking. Life Sci 27:130–139.

Thanos PK, Taintor NB, Rivera SN, Umegaki H, Ikari H, Roth G, Ingram DK, Hitzemann R, Fowler JS, Gatley SJ, Wang GJ, Volkow ND (2004) DRD2 gene transfer into the nucleus accumbens core of the alchol preferring and non-preferring rats attenuates ethanol consumption. Alcohol Exp Clin Res 28:720–728.

Thanos PK, Volkow ND, Freimuth P, Umegaki H, Ikari H, Roth G, Ingram DK, Hitzemann R (2001) Overexpression of dopamine D2 receptors reduces alcohol self-administration. J Neurochem 78:1094–1103.

Tuzun E, Sharp AJ, Bailey JA, Kaul R, Morrison VA, Pertz LM, Haugen E, Hayden H, Albertson D, Pinkel D, Olson MV, Eichler EE (2005) Fine-scale structural variation of the human genome. Nat Genet 37:727–732.

Van Esch H, Bauters M, Ignatius J, Jansen M, Raynaud M, Hollanders K, Lugtenberg D, Bienvenu T, Jensen LR, Gecz J, Moraine C, Marynen P, Fryns JP, Froyen G (2005) Duplication of the MECP2 region is a frequent cause of severe mental retardation and progressive neurological symptoms in males. Am J Hum Genet 77:442–453.

Walker JR, Su AI, Self DW, Hogenesch JB, et al (2004) Applications of a rat multiple tissue gene expression data set. Genome Res 14: 742–749.

Wu CC, MacCoss MJ, Howell KE, Matthews DE, Yates JR III (2004) Metabolic labeling of mammalian organisms with stable isotopes for quantitative proteomic analysis. Anal Chem 76:4951–4959.

Wu CC, MacCoss MJ, Howell KE, Yates JR III (2003) A method for the comprehensive proteomic analysis of membrane proteins. Nat Biotechnol 21:532–538.

Zhang L, Miles MF, Aldape KD (2003) A model of molecular interactions on short oligonucleotide microarrays. Nat Biotechnol 21:818–821.

Zhang L, Wang L, Ravindranathan A, Miles MF (2002) A new algorithm for analysis of oligonucleotide arrays: application to expression profiling in mouse brain regions. J Mol Biol 317:225–235.