Incorporating Mouse Genome
Mammalian Genome
Genes and Phenotypes

Expression profiling identifies novel candidate genes for ethanol sensitivity QTLs

Erik J. MacLaren,[1] Beth Bennett,[2] Thomas E. Johnson,[2,3] James M. Sikela[1,2,4]

[1]Department of Pharmacology, University of Colorado at Denver and Health Sciences Center, Aurora, Colorado 80045, USA
[2]Institute for Behavioral Genetics, University of Colorado, Boulder, Colorado 80309, USA
[3]Department of Integrative Physiology, University of Colorado, Boulder, Colorado 80309, USA
[4]Human Medical Genetics Program, University of Colorado at Denver and Health Sciences Center, Aurora, Colorado 80045, USA

Received: 6 May 2005 / Accepted: 27 October 2005

Abstract

The Inbred Long Sleep (ILS) and Inbred Short Sleep (ISS) mouse strains have a 16-fold difference in duration of loss of the righting response (LORR) following administration of a sedative dose of ethanol. Four quantitative trait loci (QTLs) have been mapped in these strains for this trait. Underlying each of these QTLs must be one or more genetic differences (polymorphisms in either gene coding or regulatory regions) influencing ethanol sensitivity. Because prior studies have tended to focus on differences in coding regions, genome-wide expression profiling in cerebellum was used here to identify candidate genes for regulatory region differences in these two strains. Fifteen differentially expressed genes were found that map to the QTL regions and polymorphisms were identified in the promoter regions of four of these genes by direct sequencing of ILS and ISS genomic DNA. Polymorphisms in the promoters of three of these genes, *Slc22a4*, *Rassf2*, and *Tax1bp3*, disrupt putative transcription factor binding sites. *Slc22a4* and another candidate, *Xrcc5*, have human orthologs that map to genomic regions associated with human ethanol sensitivity in genetic linkage studies. These genes represent novel candidates for the LORR phenotype and provide new targets for future studies into the neuronal processes underlying ethanol sensitivity.

Correspondence to: James Sikela; E-mail: james.sikela@uchsc.edu

Introduction

The Inbred Long Sleep (ILS) and Inbred Short Sleep (ISS) mice have a marked phenotypic difference in their hypnotic sensitivity to ethanol as measured by their loss of righting response (LORR) and have been widely used as a genetic model of intrinsic ethanol sensitivity (Collins 1981; Deitrich 1990). Quantitative trait loci (QTLs) that harbor genetic differences influencing LORR, called *Lores* (for loss of righting due to ethanol), have been mapped in these strains (Markel et al. 1997), and four of these QTLs have been subsequently confirmed by their capture in congenic strains (Bennett et al. 2002b). Although these four genomic regions have been linked to the phenotype, further refinement to specific candidate genes is required to identify the genetic determinants of ethanol sensitivity in these mice.

Genetic differences underlying the QTLs can fall into one of two broad categories: coding region polymorphisms that affect the amino acid sequence of the translated protein or regulatory region polymorphisms that affect the level or pattern of expression of a gene. The coding regions of many candidate genes mapping to the *Lore* QTL regions have been sequenced to address the first possibility (Ehringer et al. 2001, 2002). This study is intended to address the second possibility by examining the transcript levels of *Lore* genes in these mouse strains to determine which are differentially expressed (DE). Any DE genes in QTL regions are excellent candidates to influence ethanol sensitivity in these mice; they thus merit further study.

Transcript levels were assayed in the cerebella of these strains using two different array platforms: Affymetrix oligonucleotide arrays and cDNA arrays manufactured at the University of Colorado Health Sciences Center (UCHSC) from a set of approxi-

mately 15,000 murine cDNA clones distributed by the National Institute of Aging, the NIA Mouse 15K clone set (Tanaka et al. 2000). The cerebellum was chosen for the study both because it is a major target of ethanol in the brain (Fadda and Rossetti 1998; Wong et al. 2003), and because the sensitivity of cerebellar neurons to ethanol correlates very highly with the LORR phenotype in a number of inbred mouse strains that vary with respect to hypnotic sensitivity phenotypes (Spuhler et al. 1982).

In addition to falling within a QTL region, true DE candidate genes would harbor polymorphisms between the ILS and ISS strains in their regulatory regions that are responsible for their differential expression. In order to investigate this possibility, the genomic regions upstream of the transcriptional start sites of several DE *Lore* genes were sequenced. The upstream sequences were then analyzed to identify putative transcription factor binding sites that would be disrupted in one strain or the other by the SNPs that were identified. These polymorphisms are possibly the quantitative trait nucleotides (QTNs) underlying the QTL linkage scores in the genomic region and ultimately influence ethanol sensitivity in these mouse strains. Differentially expressed genes were also analyzed to identify their expression QTLs (eQTLs) to determine if any DE genes are likely to be *cis*-regulated, thus bolstering the case for functional effects of these promoter region SNPs.

In addition, linkage and association studies that have identified human genomic regions related to ethanol sensitivity in humans were queried to assess human orthologs of *Lore* genes that are DE in the cerebella of the ILS and ISS. Any genes implicated in both human and rodent genetic studies are excellent candidates to be involved in influencing ethanol sensitivity in humans.

Materials and methods

Tissue collection and RNA isolation. Adult male Inbred Long Sleep (ILS) and Inbred Short Sleep (ISS) mice between 4 and 10 weeks old were obtained from the Institute for Behavioral Genetics (IBG, Boulder, CO). After arriving at the University of Colorado Health Sciences Center (Denver, CO), the animals were housed five to a cage for two weeks during which time they were provided food and water *ad libitum* in a 12-h light/dark cycle. No experimental manipulations or measurements were carried out on the mice during this time. After the acclimation period, the mice were sacrificed by cervical dislocation without prior anesthesia. Immediately after sacrifice, the brain was surgically removed and the cerebellum was dissected out and flash frozen in liquid N_2. Dissection and freezing was completed within 5 min from the time of sacrifice. The tissue samples were stored at −80°C until RNA isolation was carried out. The cerebella were homogenized individually in Buffer RLT (Qiagen, Valencia, CA) using a Fisher PowerGen 125 and disposable OMNI-Tips (Fisher Scientific, Hampton, NH). Total RNA was extracted from the homogenized samples using an RNeasy Midi Kit (Qiagen) according to the manufacturer's instructions including the optional on-column DNA digestion with RNase-free DNase I. The concentration of each sample was determined by absorbance at 260 nm (A260) and purity by the ratio of A260 to A280. A range of 1.9−2.1 was considered adequately pure.

Affymetrix arrays

Labeling and hybridization. Total RNA from four different mice, two ILS and two ISS, was used in four hybridizations, two replicates per strain, to Mouse Expression Set 430 (MOE430) A and B arrays (Affymetrix, Santa Clara, CA). Five micrograms of total RNA isolated from a single ILS or ISS cerebellum were reverse transcribed into double-stranded cDNA using the SuperScript Choice system (Invitrogen, Carlsbad, CA) and an oligo-dT primer containing a T7 RNA polymerase promoter (Proligo Primers & Probes, Boulder, CO). The ds-cDNA was isolated and purified with the GeneChip Sample Cleanup Module (Affymetrix). The cDNA was next transcribed into biotin-labeled cRNA by incubating at 37°C for 4 h with HY Reaction Buffer, biotin-labeled ribonucleotides, DTT, RNase Inhibitor Mix, and T7 RNA Polymerase (Enzo Life Sciences, Farmingdale, NY). The labeled cRNA was purified using the GeneChip Sample Cleanup Module following the manufacturer's instructions. At this point, the cRNA was quantified and checked for quality using a ''Lab on a Chip'' 2100 Bioanalyzer (Agilent, Palo Alto, CA). Next, 20 μg of cRNA were fragmented into pieces 50-200 bases in length by incubation at 94°C for 35 min with high Mg^{2+} Fragmentation Buffer (Affymetrix). The sample was then added to a hybridization solution containing 100 mM MES, 1 M Na^+, and 20 mM EDTA in the presence of 0.01% Tween 20. The final concentration of the fragmented cRNA was 0.05 μg/μL. Next, 200 μL of the sample were hybridized to the array at 45°C for 16 h using a GeneChip Hybridization Oven 640 (Affymetrix). After hybridization, the hybridization solutions were removed and the arrays were washed and stained with Streptavidin-phycoerythrin using a GeneChip Fluidics Station 450 (Affymetrix). The

arrays were then read at a resolution of 2.5–3 μm using an Affymetrix GeneChip Scanner 3000 to collect the hybridization data.

Data collection and analysis. The statistical expression algorithm of the GeneChip Operating Software (GCOS) was used to scale overall chip intensities to the same level and for probe level analysis of the hybridization data. Affymetrix arrays use probe pairs consisting of a perfect match (PM) and mismatch (MM) in order to subtract background and cross-hybridization artifacts. The MM and PM probes have identical sequences except for one base change in the middle of the MM sequence. The statistical expression algorithm was used to make an absent, present, or marginal call for each probe set on the arrays. Absent calls were made for probe sets that did not have hybridization intensities above the background intensity of the array and also for probe sets with an insignificant hybridization intensity difference between the PM and MM probes.

The hybridization data were then loaded into the GeneSpring software package (Agilent). Only data from probe sets identified as "present" in at least half of the hybridizations were considered during subsequent analysis. Each measurement on the arrays was globally normalized to the 50th percentile value of all measurements on the array to normalize each chip and make the data across chips comparable. As a per-gene normalization step, the hybridization value for each gene was normalized to the median value of the gene in the ILS samples, which were arbitrarily assigned as the reference samples. Genes were then eliminated from further analysis if they had a coefficient of correlation greater than 0.95 to a hypothetical gene profile that was constant for all experiments regardless of the mouse strain used. This step was taken because genes that did not vary in expression level between the two strains were unlikely to be of interest, and so the number of genes tested was minimized to increase the power of statistical analyses. The hybridization data from each strain were then grouped and Welch's *t*-test was used to determine which genes were different in their hybridization ratios between the two strains, to a significance level $p \leq 0.05$. Because of multiple testing, this significance threshold resulted in an expected false positive rate of 11.2% in our results.

cDNA microarrays

Labeling and hybridization. The cDNA microarrays used in these experiments were manufactured in-house at UCHSC by the Gene Expression Core. The cDNA chips were constructed using 15,512 mouse clones from the National Institute of Aging 15K Mouse cDNA Clone Set (Tanaka et al. 2000), and these chips are referred to as the "NIA15K" arrays. The average probe size was 1500 bases, and the full list of clones can be accessed online through the National Institute of Aging (http://lgsun.grc.nia.nih.gov/cDNA/15k.html).

A total of nine hybridization experiments were conducted using these cDNA arrays. Pooled samples of total RNA were used for hybridizations to the cDNA microarrays in order to mask intrastrain variability. Each pool included RNA from five ILS mice or five ISS mice, and no individual mouse was included in more than one pool. Pool 1 was used for four hybridizations, two in which the ILS sample was labeled with cyanine-3 dye (Cy-3) and the ISS sample with cyanine-5 dye (Cy-5), and two hybridizations with the opposite dye orientation. Pool 2 was used in two hybridizations, one in each dye orientation. Pool 3 was used for three hybridizations with the same dye-labeling orientation, ILS labeled with Cy-5 and ISS labeled with Cy-3.

Equal amounts of total RNA from five individual mice of the same strain were pooled together and then directly labeled in a one-step reverse transcription reaction. Total RNA (20 μg) was combined with 5 × first-strand buffer, 1 μg oligo-dT20mer, 0.1 M DTT (Gibco Invitrogen, Carlsbad, CA), low dTTP/dNTP mix (Amersham Biosciences, Piscataway, NJ), RNasin (Promega Biosciences, San Luis Obispo, CA), and either cyanine 3- or cyanine 5-dUTP (PerkinElmer, Torrance, CA) in 500-μL tubes. The reactions were heated to 65°C for 5 min, and then SuperScript II RNase H$^-$ Reverse Transcriptase (Gibco) was added and the mix was incubated at 42°C for 90 min. Five microliters of 0.5 M EDTA were added to stop the reaction and then the mixture was heated to 65°C for 30 min with 10 μL of 1 M NaOH to hydrolyze the RNA, followed by the addition of 25 μL of 1 M Tris to neutralize the NaOH. The Cy-3- and Cy-5-labeled probes were combined and isolated by running them through a Microcon YM-30 size-exclusion column (Millipore), washed with TE buffer, and resuspended in 11 μL of TE. After adding 10 μg of mouse COT-1 DNA (Gibco), 8 μg poly(A) RNA (Amersham Biosciences), 4 μg yeast tRNA (Sigma-Aldrich, St. Louis, MO), 3.1 μL 20 × SSC, and 0.5 μL 10% SDS, the probe mixture was hybridized to the arrays at 42°C for 16 h. The microarrays were washed using dilute SSC solution to remove debris and hybridization buffer, and then scanned with a GenePix4000A scanner (Axon Instruments, Union City, CA). The Cy-3 and Cy-5 fluorescence was measured for each cDNA element, and any probes with a fluorescence level below background or less than 60% of the original

spotting area were flagged as ''absent'' while the rest were noted as ''present'' by the GenePix software.

cDNA microarray data analysis. The hybridization intensity data were loaded directly into the GeneSpring software package (Agilent). All analyses used only genes flagged as ''present'' in at least half of the hybridizations during scanning. All arrays were then normalized using the Lowess normalization feature in GeneSpring (Yang et al. 2002). Briefly, a Lowess curve was fitted to the log-intensity versus log-ratio plot using 20% of the data to calculate the fit at each point. This curve was used to adjust the control value for each measurement and minimize intensity-dependent artifacts. The control value was set to 10 for all measurements with a postnormalization value of less than 10. Next, the chips were grouped by dye-orientation and the two groups were considered separately. Genes that had a statistically significant variation ($p \leq 0.05$) in their hybridization ratios among different experiments in the same dye orientation, as determined by Welch's *t*-test, were considered unreliable and removed from further analysis. The two dye-orientation groups were then compared and Welch's *t*-test was used to determine which genes had statistically significant differences in their hybridization ratios between the two groups. The Benjamini and Hochberg False Discovery Rate (FDR) was used for multiple test correction at an overall error rate of 5%. Probes that passed this test were considered significantly differentially expressed (DE) between ILS and ISS cerebella.

Mapping DE genes in the mouse genome Lore QTL boundaries. The probe sets for the set of DE genes from the cDNA arrays were mapped to the Mouse May 2004 (mm5) assembly of the mouse genome using the UCSC Genome Browser (http://genome.ucsc.edu) and the Blast-like Alignment Tool (BLAT) (Kent 2002). Alignments were required to have a BLAT score ≥200 and a percent identity score ≥98 to be considered valid. In case of multiple position assignments, the gene was assigned to the genomic location corresponding to its highest BLAT score. Differentially expressed probe sets on the Affymetrix arrays were assigned to genes and genomic positions based on annotation available at the NetAffx Analysis Center (Affymetrix; http://www.Affymetrix.com/analysis/netaffx/index.affx).

The genomic boundaries used here to define the *Lores* are based on data from ISCR lines made with ILS donor regions on an ISS background as of March 2004. Some of the intervals vary from those

Table 1. *Lore* intervals used for differentially expressed gene analysis

QTL (Chr)	*Flanking markers*	*Physical position (Mb)*	*Interval size (Mb)*	*Genes in interval*[a]	*MOE430 arrays*[b]	*NIA15K array*[c]
Lore1 (1)	*D1Mit78-D1Mit23*	71.4–75.7	4.3	84	69	73 (30)
Lore2 (2)	*D2Mit258-D2Mit166*	130.2–140.8	16.1	137	163	155 (71)
	D2Mit259-D2Mit57	142.5–148.0				
Lore4 (11)	*D11Mit108-D11Mit316*	35.7–59.0	35.8	635	673	412 (329)
	D11Mit351-D11Mit8	67.1–79.6				
Lore5 (15)	*D15Mit84-D15Mit35*	40.1–104.0	63.9	736	884	500 (386)

The number of known and predicted genes within the intervals is listed along with the number of unique transcripts represented on the arrays from the intervals. Positions are derived from the UCSC mm5 assembly.
[a]Gene numbers are based on known and putative genes annotated on the Gene_Sequence map at NCBI (http://www.ncbi.nlm.nih.gov/genome/guide/mouse/).
[b]These numbers reflect the number of unique Locus Link clusters represented by probe sets on the array plus the number of probe sets not assigned to any Locus Link cluster.
[c]Probe sequences were aligned to the mouse genome to determine position and redundancy. The number of probes matching MOE430 probe sets is shown in parentheses.

published (Bennett et al. 2002a) because of more recent testing (Bennett and Johnson, unpublished) and are shown in Table 1.

Preparation and sequencing of PCR products. The transcriptional start site (TSS) of each DE gene was determined using the Database of Transcriptional Start Sites (DBTSS; http://dbtss.hgc.jp) (Suzuki et al. 2004), except for *Atf1*, *Cthrc1*, *Krt2-8*, *Myo1d*, *Rassf2*, and *Scrt1*, because there was no information available for these genes. The 5′ end of the most upstream oligo-capped cDNA sequence was defined as the TSS. In those cases where the gene was not available in the DBTSS, the 5′ end of the gene's corresponding RefSeq sequence (Maglott et al. 2000) was presumed to be the TSS.

After removing the cerebellum for RNA preparation, the remaining brain tissue of individual ILS and ISS mice was used for DNA isolation. Genomic DNA was obtained using the DNeasy Tissue Kit (Qiagen) following the manufacturer's instructions, including the optional RNase digestion. Primers were designed to amplify approximately 600 base pairs (bp) upstream and 100 bp downstream of the TSS for each gene using the program Primer3 (Rozen and Skaletsky 2000). Polymerase chain reactions (PCR) were carried out using Ready-to-Go PCR Beads (Amersham Biosciences, Piscataway, NJ) with typical cycling parameters: 4 min at 94°C, 35 cycles of 15 sec at 94°C, 75 sec at 58°C, 75 sec at 72°C, followed by 10 min at 72°C. The reaction products were separated and visualized using ethidium bromide-stained 1.8% agarose gels and ultraviolet light. Products of appropriate size were excised from the gel and purified using a QIAquick Gel Extraction Kit (Qiagen).

The purified PCR products were directly sequenced using a BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA) according to the manufacturer's instructions. Precipitated samples were loaded onto an ABI PRISM 3100 Genetic Analyzer (Applied Biosystems). Resulting sequence data were analyzed and aligned using CONSED, Phred, and Phrap software (Gordon et al. 1998). A minimum of 4 × coverage was generated for each reported sequence from each strain.

Transcription factor binding site prediction. The promoter sequences of DE genes were analyzed using the AliBaba 2.1 application (Grabe 2002) available from the Gene Regulation website (http://www.gene-regulation.com). This program predicts transcription factor (TF) binding sites in unknown DNA sequences using empirically derived binding site data in the TRANSFAC database (Wingender et al. 2001). The two alternative promoter sequences (ILS and ISS) were analyzed separately and the TF binding sites predicted for each sequence were then compared. TF binding sites predicted for a promoter sequence in one strain, but not predicted in the sequence for the other strain, are reported as disruptions in the results. Our promoter analyses used the following values for the adjustable parameters: Pairism to known sites = 64; Matrix width = 10; Minimum number of sites = 4; Minimum matrix conservation = 80%; Similarity of sequence to matrix = 1%; Factor class level = 4. These parameters are more stringent than the default settings and were meant to favor sites strongly similar to known TF binding sites while excluding weaker matches.

Expression QTL analysis. The WebQTL Project (www.genenetwork.org) is an online database with genotype and phenotype information from C57BL/6J (B6) and DBA/2J (D2) inbred mouse strains as well as the derived recombinant inbred (RI) strains, the BXD panel, and the LXS panel, from ILS and ISS. In addition, there is basal expression data derived from microarray studies using oligo arrays and cerebellar tissue from B6, D2, and the BXD strains. This expression data and genotype data can be used to identify genomic markers that are linked to the expression level of a gene in the B6 and D2 strains. The regions identified delimit likely locations of *cis* and *trans* acting factors that affect that gene's expression and are termed expression QTLs, or eQTLs (Bystrykh et al. 2005; Chesler et al. 2005; Hubner et al. 2005).

Using the most recent data set for cerebellar tissue [SJUT Cerebellum mRNA M430 (Oct04) MAS5] and the ''Marker Regression'' tool available from WebQTL, single-marker eQTLs were individually mapped for all DE *Lore* genes. The ten markers with the best likelihood ratio statistic (LRS) scores were returned for each DE gene and the p value of each linkage was determined by permutation test. Only markers with significant ($p \leq 0.05$) (Lander and Kruglyak 1995) linkage to a DE gene are reported. A marker linked to a DE gene and mapping within 10 Mb from its position in the mouse genome (UCSC mm5 assembly) is hereafter termed a ''*cis*-eQTL.'' This window is used to account for the fact that linkage extends several Mb in each direction on the chromosome, and the marker showing the highest linkage score is not necessarily the closest on the physical map. Multiple markers linked to the same DE gene and mapping within 25 Mb of each other are considered

Table 2. Differentially expressed cerebellar genes mapping to *Lore* QTL intervals

Symbol	*Gene description*	*Fold-change*	*Lore*
More highly expressed in ILS			
Pcsk2	Proprotein convertase subtilisin/kexin type 2	2.9	2
Slc22a4[a]	Solute carrier family 22, member 4	2.4	2
BG075643	—	6.2	4
Myo1d	Myosin ID	6.2	4
Cthrc1	Collagen triple helix repeat containing 1	3.2	5
Krt2-8	Keratin complex 2, basic, gene 8	2.3	5
Lgals2	Lectin, galactose-binding, soluble 2	2.0	5
Scrt1	Scratch homolog 1, zinc finger protein (*Drosophila*)	2.4	5
More highly expressed in ISS			
Xrcc5[a]	X-ray repair complementing defective repair in Chinese hamster cells 5	2.5	1
Rassf2	*Ras* association (RalGDS/AF-6) domain family 2	2.4	2
Ebf1	Early B-cell factor 1	2.1	4
Stx8	Syntaxin 8	2.0	4
Tax1bp3	Tax1 (human T-cell leukemia virus type I) binding protein 3	2.2	4
Tnfaip1	Tumor necrosis factor, alpha-induced protein 1	2.0	4
Atf1	Activating transcription factor 1	2.0	5

[a]These DE genes have human orthologs in genomic regions associated with LR.

to be part of the same eQTL, and only the most significant marker in each eQTL is reported.

Results

Differentially expressed genes. Gene expression levels in the cerebella of ILS and ISS mice were assayed with Affymetrix MOE430 arrays that contain probe sets for 1789 genes in the *Lore* intervals. Of this total, 15 genes (<1%) were differentially expressed, 8 being more highly expressed in ILS cerebellum and 7 in ISS cerebellum as shown in Table 2. Relative expression levels in ILS and ISS cerebella were also compared using cDNA microarrays made using the NIA15K mouse clone set that included probes for 1033 genes in the *Lore* QTLs. Only one of these *Lore* genes on the NIA15K arrays, *Myo1d*, was identified as differentially expressed between the two strains. It was more highly expressed in ILS and was also found by the MOE430 arrays. The complete data set from these hybridization experiments has been submitted to the Gene Expression Omnibus database (Edgar et al. 2002; Barrett et al. 2005) and is freely available online (http://www.ncbi.nlm.nih.gov/geo/, accession numbers GSE3071 and GSE3114).

Expression QTL mapping for the DE Lore genes. Using the WebQTL database (Wang et al. 2003), expression QTLs (eQTLs) were identified for each of the 15 DE *Lore* candidate genes, and the locations of all linkage scores that were statistically significant were compared to the genes' locations in the mouse genome. Although these eQTLs were identified by analysis of B6, D2, and recombinant inbred strains (BXD) derived from them, these eQTLs obviously identify regulatory regions that are likely to be involved in all mouse strains. Moreover, B6 and D2 are two of the eight progenitor strains used to create the ILS and ISS strains (McClearn and Kakihana 1981) and therefore are likely to apply directly to a subset of genes whose regulatory alleles have been carried through the selection process and are maintained in the ILS and ISS strains. The differential expression of three genes, *Slc22a4*, *Pcsk2*, and the clone BG075643, had significant linkage scores associated with markers near to their genomic locations and thus are potentially *cis*-regulated. The expression linkage results for *Slc22a4* across the mouse genome are shown in Fig. 1 as an example. There is no evidence from the WebQTL database to suggest that the remaining *Lore* DE genes are *cis*-regulated, although that cannot be ruled out.

Sequencing promoter regions of DE Lore genes with evidence of cis-regulation. The promoter regions of *Slc22a4*, *Pcsk2*, and BG075643 were next examined to determine if any sequence polymorphisms exist between the strains that may be responsible for the expression differences that were observed. BG075643 proved to be unsuitable for promoter region sequencing because it was not assigned to any known gene or gene model in the mouse genome, nor could it be aligned to any human gene or gene model using the standard sequence alignment tools BLAST or BLAT (Altschul et al. 1997; Kent 2002). As a transcribed clone without a corresponding gene model, it was not possible to

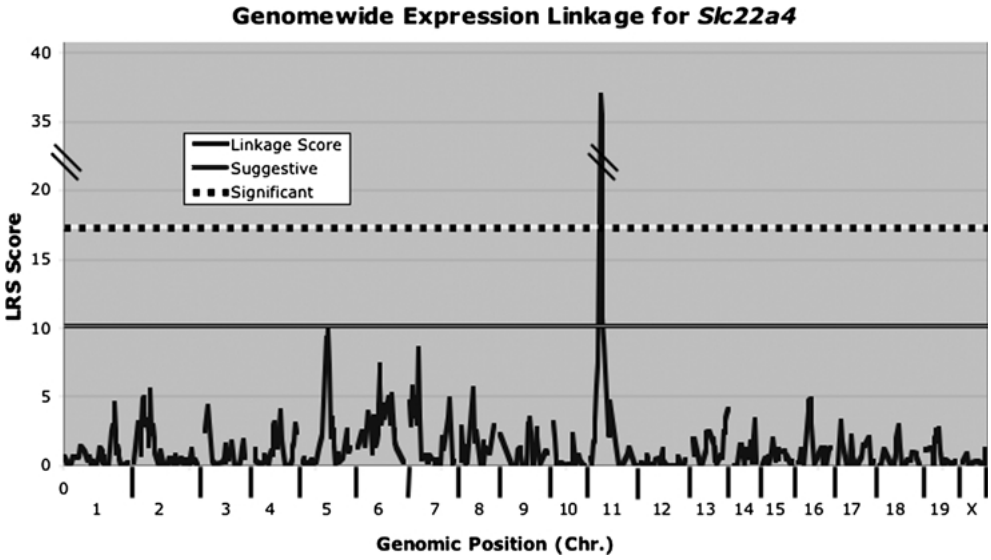

Fig. 1. Graph of linkage scores for *Slc22a4* expression in the B6 and D2 strains for markers distributed across the entire mouse genome. The threshold values for ''suggestive'' and ''significant'' linkage scores are shown. The only peak reaching significance is on Chr 11 near the *Slc22a4* gene, thus implying *cis*-regulation of this transcript.

determine the locations of the transcriptional start site (TSS) or the promoter. Empirical evidence of the TSS was available for the other two *Lore* DE genes at the Database of Transcriptional Start Sites (DBTSS, http://dbtss.hgc.jp). Despite repeated attempts to sequence the promoter region of *Pcsk2*, the proper region could not be amplified from ILS or ISS genomic DNA and so comparative sequence was not generated for this gene promoter.

The 600-bp region immediately upstream of the *Slc22a4* TSS was amplified from ILS and ISS genomic DNA and sequenced. Two single nucleotide polymorphisms (SNPs) were identified between the strains in this region. One of these SNPs, shown in Table 3, disrupts a putative binding site for the transcription factor Sp-1 in the ISS sequence. This matches with the expression data because Sp-1 acts to enhance transcription and the gene is more highly expressed in ILS mice (Schmidt et al. 1989). When promoter sequences for *Slc22a4* from the ILS and ISS strains were compared with the same sequences for various mouse strains in the Celera database, the ILS allele matched B6 sequences while the ISS sequence matched those from D2. This lends further support to the importance of these polymorphisms since eQTL analysis shows that B6 sequences in this region correlate to higher *Slc22a4* expression, as observed in ILS.

Promoter sequencing of other DE Lore genes. Although there was no evidence for *cis*-eQTLs for the 12 other DE *Lore* genes in the B6 and D2 mouse strains, this did not exclude the possibility that these genes are regulated in *cis* in the ILS and ISS mouse strains. The promoter regions of these genes were also PCR amplified and sequenced from ILS and ISS genomic DNA. Polymorphisms were identified in the promoter sequences of *Rassf2*, *Stx8*, and *Tax1bp3*, and several putative TF binding sites were affected by these changes as summarized in Table 3. In *Rassf2*, two A → G transitions disrupt two predicted Sp-1 binding sites in the ILS promoter, in agreement with the expression data showing that this gene is more highly expressed in ISS. Another A → G transition in *Tax1bp3* disrupts a putative binding site for NF-κB in the ILS promoter. This TF stimulates transcription (Molitor et al. 1990), and *Tax1bp3* is more highly expressed in ISS mice.

The promoter region of *Ebf1* contained a polyadenosine [poly(A)] tract that prevented accurate sequencing through the entire region, but no polymorphisms either 5′ or 3′ of this poly(A) region were

Table 3. Polymorphisms in the promoter regions of differentially expressed *Lore* genes

Gene symbol	*Polymorphisms*[a]	*TF binding sites*[b]
High in ILS		
Slc22a4	−327 G → A	
	−457 G → A	Sp1 (ILS)
High in ISS		
Rassf2	−237 A → G	
	−318 G del in ILS	Sp1 (ISS)
	−376 C → T	
	−415 G → A	
	−484 C → T	
	−564 C → G	Sp1 (ISS)
Stx8	−246 T → C	
	−396 ATGT del in ISS	
Tax1bp3	+11 T → C	
	−51 T → C	
	−96 A → G	NF-κB (ISS)
	−200 T → C	

[a]Polymorphisms listed as ILS → ISS; del = deletion.
[b]The strain with the intact TF binding site is indicated in parentheses.

Table 4. *Lore* genes with coding region polymorphisms

QTL (Chr)	*Gene symbol*	*Position (Mb)*
Lore1 (1)	*Ptprn*[a]	75.5
	Znf142[a]	74.9
Lore2 (2)	*Ptpra*	130.2
	Plcb4	135.6
	Znf133[a]	143.9
Lore4 (11)	–	–
Lore5 (15)	*4921531G14Rik*	43.5
	Rad21	52.0
	Prkm8ip[a]	89.5

From Ehringer et al. (2001).
[a]Polymorphisms causing changes in the predicted amino acid sequence of the protein.

found between the ILS and ISS mice. Although multiple primers and pair combinations were tried, PCR products could not be obtained for *Scrt1*, *Cthrc1*, *Myo1d*, or *Atf1*.

Comparison to human studies. Genes that are identified as being important to ethanol sensitivity in mice are also good candidates to be examined in human populations in which a similar phenotype, level of response (LR) (Schuckit 1988; Schuckit and Smith 1996; Schuckit 1998), has been examined. Of the 15 DE *Lore* genes, two have human orthologs that map to a region of the human genome linked to human ethanol sensitivity, as noted in Table 2. A study using data from over 700 individuals collected by the Collaborative Study on the Genetics of Alcoholism (COGA) (Begleiter et al. 1995) identified several markers on Chromosome 2 linked to LR in this population. *XRCC5*, the human ortholog of *Xrcc5*, maps within 2 Mb of the marker with the highest linkage score, D2S434 (Schuckit et al. 2001). In a study using a different human sample, a modest linkage was identified on Chromosome 5 for LR as assessed by body sway measurements (Wilhelmsen et al. 2003). The human gene *SLC22A4*, which is orthologous to *Slc22a4*, is on human Chromosome 5 within the region of linkage reported.

Discussion

The ILS and ISS strains were created as a model system to study hypnotic sensitivity to ethanol, and previous efforts have identified four QTLs, *Lores 1, 2, 4*, and *5*, which are largely responsible for the heritable component of this trait (Markel et al. 1997; Bennett and Johnson 1998). Previous work identified eight candidate genes in these regions that contain sequence differences between the ILS and ISS strains (Ehringer et al. 2001) and are shown in Table 4. Complementing these results, 15 new candidate genes for regulatory region polymorphisms have been identified for this trait by comparing the expression levels of transcripts derived from these QTL regions between the two mouse strains in a major target of ethanol action, the cerebellum.

It is interesting to note that there is no overlap between the coding region candidates and expression level candidates. This reinforces the importance of a two-pronged approach to candidate gene identification because either a protein sequence or expression change may contribute to the phenotypic difference. In this case at least, it appears that these two classes of candidates do not overlap, so a survey of the QTL intervals that searched for only coding region or only expression changes would entirely miss many plausible candidates.

It is also noteworthy that the two platforms used in our experiments, the Mouse Expression Set 430 (Affymetrix) and the NIA 15K arrays, both identified only one *Lore* gene, *Myo1d*, that passed the selection criteria for differential expression. One major reason for this is that 9 of the 15 genes identified by the MOE430 arrays were not represented on the 15K arrays, thus it was impossible to confirm them on this platform. The remaining six genes had hybridization ratios on the cDNA chips that showed higher expression in the same strains identified in the Affymetrix arrays (data not shown); however, these genes did not pass the DE threshold on the 15K arrays.

The majority of the 15 expression candidates map to the *Lore4* and *Lore5* regions, partly because of the comparatively larger sizes of these intervals. Several of these new candidates have supporting evidence that makes them compelling candidates. These include the transcription factor genes *Atf1* and *Scrt1*, which have been shown to have profound influences on brain function and development (Nakakura et al. 2001; Pittenger et al. 2002). Others, including the transcribed clone BG075643, are less well characterized and have less obvious roles in neuronal function but should be examined in future studies because of their demonstrated differences in expression in these mice.

In contrast, there are only four differentially expressed candidate genes from *Lore1* and *Lore2*, both of which have been significantly narrowed using interval-specific congenic recombinant lines (Bennett et al. 2002a). *Xrcc5* and *Slc22a4* are attractive candidates because their human orthologs map to genomic regions linked to human ethanol sensitivity. In *Lore2*, *Rassf2* is a Ras effecter protein thought to promote apoptosis, although its precise function is unclear (Vos et al. 2003), and *Pcsk2* is a protease that processes peptide hormone precursors (Laurent

et al. 2004) that are expressed in the brain, making them attractive candidates as well.

This work has generated the most complete gene expression profiles of cerebellum in these two mouse strains ever reported to date, and it combines gene expression, sequencing, and comparative genomic techniques to advance our understanding of the genetic underpinnings of ethanol sensitivity. It builds on previous efforts that have used the ILS and ISS model of hypnotic sensitivity to find DNA polymorphisms influencing this complex trait. Fifteen novel candidates for having regulatory polymorphisms affecting the LORR phenotype have been presented, and these data complement other studies that sequenced genes in these intervals to generate candidates with coding region differences (Ehringer et al. 2001). These two approaches have thus produced a comprehensive list of candidates to be biologically tested for roles for influence on ethanol sensitivity in mice and humans.

Acknowledgments

These investigation were supported by the following NIH Grants: NRSA AA13786 (EJM), AAO8940 (TEJ) and AAO11853 (JMS).

References

1. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25, 3389–3402
2. Barrett T, Suzek TO, Troup DB, Wilhite SE, Ngau WC, et al. (2005) NCBI GEO: mining millions of expression profiles—database and tools. Nucleic Acids Res 33, D562–D566
3. Begleiter H, Reich T, Hesselbrock V, Porjesz B, Li TK, et al. (1995) The Collaborative Study on the Genetics of Alcoholism. Alcohol Health Res World 19, 228–236
4. Bennett B, Johnson TE (1998) Development of congenics for hypnotic sensitivity to ethanol by QTL-marker-assisted counter selection. Mamm Genome 9, 969–974
5. Bennett B, Beeson M, Gordon L, Carosone-Link P, Johnson TE (2002a) 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
6. Bennett B, Beeson M, Gordon L, Johnson TE (2002b) Reciprocal congenics defining individual quantitative trait loci for sedative/hypnotic sensitivity to ethanol. Alcohol Clin Exp Res 26, 149–157
7. Bystrykh L, Weersing E, Dontje B, Sutton S, Pletcher MT, et al. (2005) Uncovering regulatory pathways that affect hematopoietic stem cell function using 'genetical genomics'. Nat Genet 37, 225–232
8. Chesler EJ, Lu L, Shou S, Qu Y, Gu J, et al. (2005) Complex trait analysis of gene expression uncovers polygenic and pleiotropic networks that modulate nervous system function. Nat Genet 37: 233–242
9. Collins AC (1981) A review of research using short-sleep and long-sleep mice. In: McClearn GE, Deitrich RA, Erwin VG (eds.), *Development of Animal Models as Pharmacogenetic Tools* (Washington, DC: U.S. Government Printing Office), pp 161–170
10. Deitrich RA (1990) Selective Breeding of Mice and Rats for Initial Sensitivity to Ethanol: Contributions to Understanding Ethanol's Actions In: Deitrich Pawlowski RA AA Initial Sensitivity to Alcohol. NIAAA Res Monogr Rockville, MD, National Institute of Alcohol Abuse and Alcoholism, 7–60
11. Edgar R, Domrachev M, Lash AE (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res 30, 207–210
12. Ehringer MA, Thompson J, Conroy O, Xu Y, Yang F, et al. (2001) High-throughput sequence identification of gene coding variants within alcohol-related QTLs. Mamm Genome 12, 657–663
13. Ehringer MA, Thompson J, Conroy O, Yang F, Hink R, et al. (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
14. Fadda F, Rossetti ZL (1998) Chronic ethanol consumption: from neuroadaptation to neurodegeneration. Prog Neurobiol 56, 385–431
15. Gordon D, Abajian C, Green P (1998) Consed: a graphical tool for sequence finishing. Genome Res 8, 195–202
16. Grabe N (2002) AliBaba2: context specific identification of transcription factor binding sites. In Silico Biol 2, S1–15
17. Hubner N, Wallace CA, Zimdahl H, Petretto E, Schulz H, et al. (2005) Integrated transcriptional profiling and linkage analysis for identification of genes underlying disease. Nat Genet 37, 243–253
18. Kent WJ (2002) BLAT—the BLAST-like alignment tool. Genome Res 12, 656–664
19. Lander E, Kruglyak L (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet 11, 241–247
20. Laurent V, Jaubert-Miazza L, Desjardins R, Day R, Lindberg I (2004) Biosynthesis of proopiomelanocortin-derived peptides in prohormone convertase 2 and 7B2 null mice. Endocrinology 145, 519–528
21. Maglott DR, Katz KS, Sicotte H, Pruitt KD (2000) NCBI's LocusLink and RefSeq. Nucleic Acids Res 28, 126–128
22. 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 7, 92–99
23. McClearn GE, Kakihana R (1981) Selective breeding for ethanol sensitivity: Short-sleep and long-sleep mice. In: McClearn GE, Deitrich RA, Erwin VG (eds.)

Development of Animal Models as Pharmacogenetic Tools (Washington, DC: U.S. Government Printing Office), pp 147–159
24. Molitor JA, Walker WH, Doerre S, Ballard DW, Greene WC (1990) NF-kappa B: a family of inducible and differentially expressed enhancer-binding proteins in human T cells. Proc Natl Acad Sci USA 87, 10028–10032
25. Nakakura EK, Watkins DN, Schuebel KE, Sriuranpong V, Borges MW, et al. (2001) Mammalian Scratch: a neural-specific Snail family transcriptional repressor. Proc Natl Acad Sci USA 98, 4010–4015
26. Pittenger C, Huang YY, Paletzki RF, Bourtchouladze R, Scanlin H, et al. (2002) Reversible inhibition of CREB/ATF transcription factors in region CA1 of the dorsal hippocampus disrupts hippocampus-dependent spatial memory. Neuron 34, 447–462
27. Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 132, 365–386
28. Schmidt MC, Zhou Q, Berk AJ (1989) Sp1 activates transcription without enhancing DNA-binding activity of the TATA box factor. Mol Cell Biol 9, 3299–3307
29. Schuckit MA (1988) Reactions to alcohol in sons of alcoholics and controls. Alcohol Clin Exp Res 12, 465–470
30. Schuckit MA (1998) Biological, psychological and environmental predictors of the alcoholism risk: a longitudinal study. J Stud Alcohol 59, 485–494
31. Schuckit MA, Smith TL (1996) An 8-year follow-up of 450 sons of alcoholic and control subjects. Arch Gen Psychiatry 53, 202–210
32. Schuckit MA, Edenberg HJ, Kalmijn J, Flury L, Smith TL, et al. (2001) A genome-wide search for genes that relate to a low level of response to alcohol. Alcohol Clin Exp Res 25, 323–329
33. Spuhler K, Hoffer B, Weiner N, Palmer M (1982) Evidence for genetic correlation of hypnotic effects and cerebellar Purkinje neuron depression in response to ethanol in mice. Pharmacol Biochem Behav 17, 569–578
34. Suzuki Y, Yamashita R, Sugano S, Nakai K (2004) DBTSS, DataBase of Transcriptional Start Sites: progress report 2004. Nucleic Acids Res 32 Database issue, D78–81
35. Tanaka TS, Jaradat SA, Lim MK, Kargul GJ, Wang X, et al. (2000) Genome-wide expression profiling of midgestation placenta and embryo using a 15,000 mouse developmental cDNA microarray. Proc Natl Acad Sci U S A 97, 9127–9132
36. Vos MD, Ellis CA, Elam C, Ulku AS, Taylor BJ, et al. (2003) RASSF2 is a novel K-Ras-specific effector and potential tumor suppressor. J Biol Chem 278, 28045–28051
37. Wang J, Williams RW, Manly KF (2003) WebQTL: web-based complex trait analysis. Neuroinformatics 1, 299–308
38. Wilhelmsen KC, Schuckit M, Smith TL, Lee JV, Segall SK, et al. (2003) The search for genes related to a low-level response to alcohol determined by alcohol challenges. Alcohol Clin Exp Res 27, 1041–1047
39. Wingender E, Chen X, Fricke E, Geffers R, Hehl R, et al. (2001) The TRANSFAC system on gene expression regulation. Nucleic Acids Res 29, 281–283
40. Wong DF, Maini A, Rousset OG, Brasic JR (2003) Positron emission tomography—a tool for identifying the effects of alcohol dependence on the brain. Alcohol Res Health 27, 161–173
41. Yang YH, Dudoit S, Luu P, Lin DM, Peng V, et al. (2002) Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. Nucleic Acids Res 30, e15