Expression Quantitative Trait Loci Mapping Identifies New Genetic Models of Glutathione S-Transferase Variation

Kevin R. Hayes,1 Brandon M. Young, and Mathew T. Pletcher2

Department of Molecular Therapeutics, The Scripps Research Institute, Scripps Florida, Jupiter, Florida

Received January 22, 2009; accepted March 23, 2009

ABSTRACT:

Expression quantitative trait loci (eQTL) mapping can be used to identify the genetic variations that underlie inherited differences in gene transcription. We performed eQTL mapping by combining whole genome transcriptional data from the hypothalami of 33 strains of inbred mice with a detailed haplotype map of those same strains, revealing 10,655 trans associations and 31 cis eQTLs. One of the cis associations was found to be driven by strain-specific variation in the expression of Glutathione S-transferase, mu 5 (Gstm5). Gstm5 is one of seven members of the glutathione S-transferase, Mu family of genes. The glutathione S-transferases are phase II metabolic enzymes and are key regulators of drug and toxin clearance. In mouse, all seven family members are tightly clustered on mouse chromosome 3. Investigation of the Gstm5 cis association in multiple tissues types revealed that an 84-kilobase region on MMU3 acts as a haplotype-specific locus control region for the glutathione S-transferase, Mu cluster. In the strains that share the minor haplotype, drastic reductions in mRNA levels in multiple members of the Gst Mu family were observed. The strain-specific differences in Gst Mu transcription characterized here accurately model the human population, in which extreme variations in expression of GST Mu family members have been observed. Furthermore, the reduction in Gst Mu levels has important relevance for pharmacology and toxicology studies conducted in these strains. For instance, the reduced levels of Gst Mu in general and Gstm5 in particular have implications in models of dopamine metabolism, Parkinson’s disease, and chemical neurotoxicity.

A panel of common laboratory inbred mouse strains contains genetic and phenotypic diversity on par with that found in the human population. This diversity was recently exploited to provide new models of human metabolic syndrome (Svenson et al., 2007). Because of the unique features of the constituent strains of this inbred mouse panel, genetic and genomic methodologies can be applied to it as a means to conduct multilevel, systems biology investigations of complex phenotypes and diseases. For instance, because the inbred mouse strains were developed from a limited founder population, their genome consists of a mosaic structure that is more amenable to genetic dissection than that of humans (Pletcher et al., 2004). By correlating phenotypic values from inbred strains with detailed single nucleotide polymorphism (SNP)-based maps of their genomes, a technique known as haplotype association mapping, genes regulating such phenotypes as drug metabolism, diet-induced obesity, tumor susceptibility, and liver fibrosis have been uncovered (Hillebrandt et al., 2005; Guo et al., 2006; Liu et al., 2006, 2007). This methodology can also identify regulators of transcriptional pathways, or expression quantitative trait loci (eQTL). eQTL analysis has been used to correlate genome-wide gene expression data with genetic variation to identify master-regulator transcription factors and their downstream targets in the brain and, in a different study, a novel component of the oxidative phosphorylation pathway (Lum et al., 2006; Wu et al., 2008).

Dense SNP-based haplotype maps have been generated for the purpose of conducting haplotype association and eQTL mapping using mouse inbred lines. One shortcoming of these maps and the haplotype association algorithms that use these maps is that they fail to consider a common form of genetic variation, copy number variations (CNVs). More recently, CNVs have generated significant interest as genetic markers of variability and disease. In fact, genetic investigations have linked drug sensitivity, cancer, mental retardation, autism, and schizophrenia with these microdeletions, insertions, and duplications (McLeod and Keith, 1996; de Vries et al., 2005; Frank et al., 2007; Szatmari et al., 2007; International Schizophrenia Consortium, 2008). Furthermore, CNVs have been shown to play a substantial role in the genetic regulation of transcription. In one study of human lymphoblastoid cell lines, it was established that CNVs underlie 17.7% of the variance observed in gene expression (Stranger et al., 2007). An initial characterization of CNVs in 21 inbred mouse strains indicates that CNVs probably play a role in the phenotypic diversity observed across the strains (Graubert et al., 2007). Ignoring

ABBREVIATIONS: SNP, single nucleotide polymorphism; eQTL, expression quantitative trait loci; CNV, copy number variation; MIP, molecular inversion probes; Gst/GST Mu, glutathione S-transferase, Mu; PCR, polymerase chain reaction; MMU, Mus musculus chromosome; Gstm5, Glutathione S-transferase, mu 5; bp, base pair; LCR, locus control region; GC-RMA, GeneChip-Robust Multichip Average.

This work was supported by a grant from the state of Florida.
K.R.H. and B.M.Y. contributed equally to this work.
1 Current affiliation: Pioneer Hi-Bred International, Johnston, Iowa.
2 Current affiliation: Investigative Toxicology, Pfizer Global Research and Development, Groton, Connecticut.

Article, publication date, and citation information can be found at http://dmd.aspetjournals.org.
[The online version of this article (available at http://dmd.aspetjournals.org) contains supplemental material.]

0090-9556/09/3706-1269–1276$20.00
Copyright © 2009 by The American Society for Pharmacology and Experimental Therapeutics
the impact of CNVs in genetic studies risks missing biologically relevant associations.

Multiplexed molecular inversion probe (MIP) assays are a recent technology used to detect both nucleotide identity and allelic frequency at a single SNP (Hardenbol et al., 2005; Wang et al., 2007). The use of MIP-based microarray genotyping platforms provides the potential for including CNVs in transcriptional analyses using inbred mouse strain genomic methodologies. In addition to reporting categorical genotyping data, the microarrays also generate quantitative intensity values for each genotyping probe that is relatively proportional to the copy number of the target sequence. Simplistically, the experimental consequence of an SNP, CNV, or other haplotypic variation would be the aberration of signal generated from probes meant to genotype the region. Correlation of intensity values from genotyping arrays with intensity values from whole genome transcriptional analysis could potentially find haplotype-induced alterations in gene expression.

In this study, we describe the first use of quantitative genotyping data generated from MIP-based genotyping technology to correlate genetic variability among inbred mouse strains with variations in gene expression data to identify eQTLs. By doing so, we provide, for the first time with haplotype association mapping, a method that allows for the potential inclusion of CNVs in the genetic analysis. Our investigation identified 10,665 trans eQTL associations between genotype intensity variation and expression data generated from hypothalamic tissue along with 31 cis eQTL associations that met statistical significance. Further examination of one of the cis relationships confirmed a haplotype-specific regulation of a cluster of closely spaced and functionally related genes, the glutathione S-transferase, Mu (Gst Mu) family of drug metabolizers. The observation of genetically regulated transcriptional variation of the Gst Mu family members across inbred strains mirrors the extensive genetic and transcriptional variation of this gene cluster previously characterized in the human population. This result demonstrates the ability of inbred strain genomics to reveal biologically important alterations in transcriptional regulation and to identify new rodent models for the study of variable drug efficacy, metabolism, and toxicity in the human population.

Materials and Methods

Genotyping of Mouse Inbred Strains. Genotyping of inbred mouse DNA samples was conducted using the MegAllele mouse 5K SNP reagent kit and the GeneChip ParAllele TrueTag 5K microarrays from Affymetrix (Santa Clara, CA) according to the manufacturer’s instructions. In brief, the platform consists of small oligonucleotides capable of undergoing circularization. These oligonucleotides are complementary to a genomic sequence with a 1-base pair (bp) gap at the site of the SNP being interrogated. Genomic DNA is hybridized to the oligonucleotides in four separate reactions, one for each base. Only the reaction for the complementary nucleotide will complete the circularization. The DNA in each reaction is then enzymatically selected for circularized DNA, and the remaining “completed” probes are amplified and hybridized against a microarray for detection and quantification.

Genomic DNA samples (The Jackson Laboratory, Bar Harbor, ME) from 33 inbred mouse strains were normalized to 150 ng/μl using nuclease-free water, and 2 μg of total DNA was used for the labeling reaction. All reactions were performed on a GeneAmp PCR System 9700 Thermocycler (Applied Biosystems, Foster City, CA) according to the Affymetrix protocol. An aliquot of 4 μl of the first-stage PCR product was used as a template for a second round of PCR. After the second labeling reaction, a 2% agarose gel was run to ensure the successful amplification. The PCR products were digested using a restriction endonuclease treatment and checked on a second 2% agarose gel. The reactions were added to the 5K ParAllele TrueTag arrays and placed in the Affymetrix 640 hybridization oven at a speed of 25 rpm for 16 h. After hybridization, the arrays were stained and washed using a GeneChip Fluidics Station 450 (Affymetrix). The microarrays were then scanned with a GeneChip Scanner 3000 MegAllele System running GeneChip Operating Software version 1.4 (Affymetrix) and analyzed by Target Genotyping Analysis Software version 7.1 (Affymetrix).

Confirmation of Presumptive CNVs. The PCR reactions were performed using 1 μl of genomic DNA (10 ng/μl) in a final volume of 20 μl, which included 25 mM concentrations of each dNTP, 1.5 mM MgCl₂, 2 U of HotStarTaq DNA polymerase (Qiagen, Valencia, CA) and 1× PCR buffer (100 mM Tris-HCl, pH 9.0, containing 500 mM KCl and 1.5 mM MgCl₂), and 0.1 mM concentrations of each primer. A total of 35 cycles were run on a PE 9700 GeneAmp PCR System Thermocycler, under the following cycling conditions: predenaturation, 15 min at 95°C, denaturation, 30 s at 95°C, annealing, 30 s at 60°C, and primer extension, 60 s at 72°C. The PCR product was then analyzed on a 2% agarose gel, with subsequent purification of PCR reactions using a DNeasy kit (Qiagen). Purified reactions were sequenced by the University of Chicago DNA Sequencing Core Facility, and resulting data were analyzed using the LaserGene software package (DNASTAR, Madison, WI). Sequences were then aligned with the MegAlign function and analyzed for the presence of strain-specific polymorphisms.

Real-Time PCR Confirmation of Expression Variation. Male inbred mice, age 3 to 5 weeks, were obtained from The Jackson Laboratory and maintained on a 12:12 light/dark cycle with lights on at 7:00 AM for approximately 5 weeks before euthanasia and harvesting of tissue samples. Total RNA was isolated from liver samples from 30 inbred strains, with the liver homogenate extracted using the standard TRizol (Invitrogen, Carlsbad, CA) protocol and then further purified with the RNaseasy kit (Qiagen). Independent pools of three animals were constructed for each inbred strain. Reverse transcription was performed on each pool according to the manufacturer’s instructions using the SuperScript III kit (Invitrogen) with random hexamers in a 50-μl reaction volume. cDNA was then carried forward for analysis by real-time PCR, iTaq supermix with Rox (Bio-Rad, Hercules, CA) was used with specific TaqMan probes from ABI for each gene target (Mm00833915_g1, Mm00725711_s1, Mm00833923_m1, Mm00728197_x1, Mm00515890_m1, Mm00656783_g1, Mm00499573_g1, and Mm00607939_g1) and 120 ng of cDNA per pool per reaction. Real-time reactions were carried out in triplicate on the ABI 7900HT system (Applied Biosystems) for 40 cycles under standard conditions. β-Actin was used as the internal control. The three replicates were averaged together to obtain an average Cq value, which was then compared with the β-actin average value. Average Cq values for each strain were subtracted from the β-actin control to obtain a fold change. All animals were handled according to The Scripps Research Institute Scripps Florida Institutional Animal Care and Use Committee approved protocols.

Microarray-Based Expression Analysis. Liver tissue was lysed using a TissueLyser (Qiagen). Total RNA was isolated from each homogenate using the standard TRIzol protocol and then further purified with an RNaseasy kit (Qiagen). Samples were quantified using the NanoDrop ND-1000 spectrophotometer. Double-stranded cDNA was prepared from 1 μg of total RNA using the SuperScript IV cDNA Synthesis kit and then was in vitro-transcribed using an iVT labeling kit (Affymetrix), with the cRNA product purified using a GeneChip Sample Cleanup Module (Affymetrix). Twenty micrograms of biotin-labeled cRNA was fragmented and hybridized to an Affymetrix Mouse 430 2.0 microarray overnight in the Affymetrix 640 hybridization oven with a speed of 60 rpm for 16 h. Microarrays were washed and stained using an Affymetrix Fluidics Station FS400. GeneChip arrays were scanned using a GeneChip Scanner 3000 (Affymetrix). The probe set intensities were quantified using the GeneChip Operating Software and analyzed with GeneChip Robust Multichip Average (GC-RMA) normalization using Array Assist Software (Stratagene, La Jolla, CA). All hybridized chips met standard quality control criteria, and mean fluorescence values of each array were scaled to a mean intensity of 500. Microarray data are available from the Gene Expression Omnibus under accession number GSE10377.

Sequenom Genotyping. All of the genotyping was accomplished by use of a single-base extension reaction using the Sequenom genotyping platform. This is a two-step process. First, the region containing the SNP is amplified. Then, a primer ending at the polymorphic site is used for the single-base extension reaction. The products are sorted by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. Primers for PCR and single base extension reactions were designed by using the MassARRAY Assay Design 3.0 software package (Sequenom, Inc., San Diego, CA). One milliliter of 2.5
to 10 ng/ml genomic DNA was combined with 1.85 ml of water, 0.1 ml of 25 mM dNTPs (Invitrogen), 0.1 ml of 5 U/ml HotStarTaq, 0.625 ml of 10× HotStar PCR buffer containing 15 mM MgCl₂, 1 ml of PCR primers mixed together at a concentration of 500 nM for multiplexed reactions, and 0.325 ml of 25 mM MgCl₂. Reactions were heated at 95°C for 15 min followed by 45 cycles at 95°C for 20 s, 56°C for 30 s, and 72°C for 1 min and a final incubation at 72°C for 3 min. After PCR amplification, the remaining dNTPs were dephosphorylated by adding 1.5 ml of water, 0.17 ml of 10× SAP buffer (Sequenom), and 0.3 U of shrimp alkaline phosphatase (Sequenom). The reaction was placed at 37°C for 20 min, and the enzyme was deactivated by incubating at 85°C for 5 min. After shrimp alkaline phosphatase treatment, the genotyping reaction was combined with 0.76 ml of water, 0.2 ml of iPLEX Termination Mix (Sequenom), 0.04 ml of iPLEX Enzyme (Sequenom, Inc.), 0.2 ml of 10× iPLEX Buffer, and 0.81 ml of 7 to 14 mM multiplexed extension primers. The MassEXTEND reaction was carried out at 94°C for 2 min and then 99 cycles at 94°C for 5 s, 52°C for 5 s, and 72°C for 5 s. The reaction mix was desalted by adding 3 mg of a cationic resin, SpectroCLEAN (Sequenom), and resuspended in 30 ml of water. ComPLETED genotyping reactions were spotted in nanoliter volumes onto a matrix arrayed into 384 elements on a silicon chip (SpectroCHIP; Sequenom), and the allele-specific mass of the extension products were determined by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. Data were analyzed using SpectroTyPER software.

**Statistical Analysis.** All statistical calculations were done in the R programming language (http://www.r-project.org). Genotype signals were generated with the Affymetrix GeneChip Genetyped Genotyping Analysis Software. Values from each of the four nucleotide calls were summed to create a “SumSignal” for further analysis. Expression data for the Mouse Diversity Panel from the Genomics Institute of the Novartis Research Foundation was downloaded from the Gene Expression Omnibus (accession number GSE5961). Raw CEL files were processed in R using GC-RMA and culled for those probe sets whose average signal intensity met a floor of 100. All 5071 probe sets from the GeneChip ParAllele TrueTag 5K microarrays and all 19,820 probe sets from the Genomics Institute of the Novartis Research Foundation microarray experiment across the 24 strains common to both sets were imported into R. Pearson correlation coefficients were generated for each MegAllele-expression pair. Because this procedure generated >100 million Pearson coefficients, significance analysis was only performed on those coefficients that exceeded 0.6. Significance was assessed by a permutation test, sampling 100,000 times, with replacement, creating an empirical cumulative distribution function, and then applying the true value to that function to gain significance.

**Results**

**Genomic Association.** To identify regions of the mouse genome that potentially contain differential haplotypic structures or CNVs, 33 inbred mouse strains were assayed using the GeneChip ParAllele TrueTag 5K platform (Supplemental Table 1). This MIP-based platform provides signal intensity for each of the four nucleic acids possible at a given polymorphic site. The intensity values of the four nucleic acid probes were combined to generate a SumSignal of each SNP for further analyses. Loss or gain of signal intensity between samples is assumed to mark the presence of a sequence anomaly at the site of the microarray probe. However, other factors may contribute to a loss of probe signal. For example, in this format of the microarray platform, the existence of nearby additional polymorphisms can affect probe binding efficiency and signal.

eQTLs were generated from previously published hypothalamus expression data of inbred mouse strains, 33 of which overlapped the strains of the genotype study. The microarray-based expression data are freely available from the National Center for Biotechnical Information Gene Expression Omnibus (accession number GSE5961). Raw CEL files were retrieved, recalled using GC-RMA, and filtered for expressed genes. Of the 31,373 probes, 19,820 exceeded a minimal expression value of 100 and accordingly were used in further analyses.

A full correlation analysis was performed between the SumSignal genotyping probes and gene expression probes. Pearson coefficients were generated for each of the ~100 million SumSignal-expression variable pairs. Hits were identified as having a Pearson coefficient of greater than 0.7 or less than −0.7. Significance of each hit was determined by a 100,000 permutation analysis. With greater than 100 million comparisons made, a false discovery rate correction was applied at Q < 0.05. A total of 10,012 SumSignal-expression pairs met all criteria considered as having a significant SumSignal-expression correlation, a hit rate of 0.01% (Supplemental Table 2).

The significant associations were plotted by genomic location to determine nodes of association (Fig. 1). Greater than 99% of all significant associations act in trans, with several clearly demarcated nodes. These trans eQTLs are associations between a polymorphism and a transcript located at disparate loci in the genome. Although these nodes could arise from chance, the probability of such an event (as corrected by false discovery rate) is low, leading to the possible identification of transcriptional factors that regulate expression of a wide variety of mRNAs. Functional analysis of trans-bands showed that one band, centering around mouse chromosome 6qE1 (MMU6), was significantly enriched for genes tagged by Gene Ontology as acting in cell cycle pathways (p = 6E−8). The other major trans-band, also centered on MMU6, showed no enrichment for any functional category.

A detected cis genotype-expression relationship is less likely the result of a false-positive than a trans relationship (McClurg et al., 2007). In addition to meeting the same statistical criteria as the trans association, a cis eQTL must also show an overlap in genomic position of the genotype and the associated transcript. Given the size of our data sets and the statistical cutoffs we assigned, there is p = 4E−8 that this would happen by random chance. Hence, cis QTLs are described as highly enriched for true positives with a relatively low false-positive rate (McClurg et al., 2007). We queried the significant data (Pearson >0.7, p < 0.05, Q <0.02) to look for cis-acting regions, defined here as the SNP interrogated lying less than 5 megabases away from the expression probe. Thirty-one SumSignal-Expression pairs met these criteria (Table 1).
**Gstm5 cis eQTL.** A cis eQTL containing the Glutathione S-transferase, mu 5 (Gstm5) locus was selected as a top candidate for further studies because the variation in gene expression was observed across multiple strains. Gstm5 is a member of the Mu subfamily of glutathione S-transferases that is expressed in high levels in the brain, testis, muscle (Fulcher et al., 1995; Listowsky, 2005). The genetic correlation was driven by six strains, NZW/LacJ, RIIIS/J, MA/MyJ, JF1/Ms, C57BL/6J, DBA/2J, and FVB/NJ. Those strains that shared the variant haplotype pattern and would be predicted to have a haplotype that was associated with decreased expression over a portion of the Gst Mu family members was performed using microarray analysis on liver samples from two strains containing the functional haplotype, and three strains with the wild-type Gstm5 haplotype were assayed using the MIP format of detection.

**Expansion of the cis Effect to the Gstm5 Mu Family.** Gstm5 is part of a gene cluster containing all seven members of the Gst Mu family, encompassing a region of 150 kilobases on MMU3. Because the functional haplotype was found to extend across most of the gene family members, examination of transcript levels for the other Gst Mu family members was performed using microarray analysis on liver samples from two strains containing the functional haplotype, and those showing wild-type signal levels (BALB/cByJ, BTBR T<+ >tJ, C57BL/6J, DBA/2J, and FVB/NJ). Those strains that produced a loss of probe signal all showed identical sequences but contained a number of polymorphisms that distinguished them from the wild-type strains and the mouse genome reference sequence (Supplemental Table 3). The reduced SumSignal of the outlier strains is probably explained by the presence of this unique haplotype interfering with either the amplification or hybridization steps of the genotyping methodology.

To more fully characterize the haplotype structure underlying the altered Gstm5 gene expression, an additional 153 SNP-based markers were selected for fine mapping of a panel of 50 inbred strains (Supplemental Table 1). These markers showed clear extension of the variant haplotype structure beyond the Gstm5 locus to include additional Gst Mu family members located on MMU3. NZO/HILJ also shared the variant haplotype pattern and would be predicted to have reduced Gstm5 expression (Fig. 4). Wild-derived PWK/PhJ, which produced a reduced SumSignal value, contained the haplotype that was associated with decreased expression over a portion of the Gst Mu family members, although not over Gstm5 (Fig. 4). At the Gstm5 locus, PWK/PhJ exhibited a third haplotype structure that could be traced to a Mus musculus musculus heritage. Additional rare haplotypes were identified at the Gstm5 locus including a SPRET/EiJ-only structure. Together, PWK/PhJ and SPRET/EiJ, both of which generated a greatly decreased SumSignal for the Gstm5-associated SNP, provide further strength to the argument that surrounding genetic diversity can significantly affect the performance of genotyping assays using the MIP format of detection.

**TABLE 1**

cis eQTLs

<table>
<thead>
<tr>
<th>Gene</th>
<th>Probes ID</th>
<th>SNP ID</th>
<th>Pearson</th>
<th>p</th>
<th>Chr</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zfp106</td>
<td>gnf1m16196_s_at</td>
<td>rs13476746</td>
<td>0.7208</td>
<td>4.60E-05</td>
<td>2</td>
<td>123148841</td>
</tr>
<tr>
<td>Zfp106</td>
<td>gnf1m16196_s_at</td>
<td>rs13476758</td>
<td>0.7030</td>
<td>7.20E-05</td>
<td>2</td>
<td>126983455</td>
</tr>
<tr>
<td>Ensa</td>
<td>gnf1m17496_a_at</td>
<td>rs13477279</td>
<td>0.7193</td>
<td>1.10E-05</td>
<td>3</td>
<td>100870621</td>
</tr>
<tr>
<td>Gstm5</td>
<td>gnf1m02321_a_at</td>
<td>rs8261742</td>
<td>0.7130</td>
<td>2.75E-04</td>
<td>3</td>
<td>111150836</td>
</tr>
<tr>
<td>C60658</td>
<td>gnf1m11435_s_at</td>
<td>rs3808636</td>
<td>0.7599</td>
<td>5.55E-03</td>
<td>3</td>
<td>142709218</td>
</tr>
<tr>
<td>C60658</td>
<td>gnf1m11435_s_at</td>
<td>rs13477456</td>
<td>0.7043</td>
<td>7.78E-03</td>
<td>3</td>
<td>146217869</td>
</tr>
<tr>
<td>A63007B10Rik</td>
<td>gnf1m20422_at</td>
<td>rs13477457</td>
<td>0.7024</td>
<td>1.44E-04</td>
<td>3</td>
<td>146408674</td>
</tr>
<tr>
<td>C509023A16Rik</td>
<td>gnf1m26632_at</td>
<td>rs8274321</td>
<td>0.7008</td>
<td>5.40E-05</td>
<td>4</td>
<td>129079328</td>
</tr>
<tr>
<td>2410081M15Rik</td>
<td>gnf1m10324_a_at</td>
<td>rs13477994</td>
<td>0.7691</td>
<td>7.01E-04</td>
<td>4</td>
<td>133594418</td>
</tr>
<tr>
<td>Tmem39b</td>
<td>gnf1m16524_s_at</td>
<td>rs13477994</td>
<td>0.7044</td>
<td>7.98E-03</td>
<td>4</td>
<td>133594418</td>
</tr>
<tr>
<td>573048SN10Rik</td>
<td>gnf1m07480_a at</td>
<td>rs13478048</td>
<td>0.7495</td>
<td>1.23E-03</td>
<td>4</td>
<td>148422515</td>
</tr>
<tr>
<td>Cav1</td>
<td>gnf1m13227_a_at</td>
<td>rs13478643</td>
<td>0.7058</td>
<td>3.82E-03</td>
<td>6</td>
<td>165261199</td>
</tr>
<tr>
<td>Cap2a</td>
<td>gnf1m23213_a_at</td>
<td>rs13478659</td>
<td>0.7403</td>
<td>4.98E-03</td>
<td>6</td>
<td>23719203</td>
</tr>
<tr>
<td>Cap2a</td>
<td>gnf1m23213_a_at</td>
<td>rs13478660</td>
<td>0.7085</td>
<td>1.02E-02</td>
<td>6</td>
<td>23849922</td>
</tr>
<tr>
<td>Slc2a3</td>
<td>gnf1m00559_a_at</td>
<td>rs13479048</td>
<td>0.7621</td>
<td>4.20E-05</td>
<td>6</td>
<td>132747781</td>
</tr>
<tr>
<td>Em1q</td>
<td>gnf1m03141_a_at</td>
<td>rs13479048</td>
<td>0.7202</td>
<td>3.30E-03</td>
<td>6</td>
<td>132747781</td>
</tr>
<tr>
<td>Mif2</td>
<td>gnf1m06480_a_at</td>
<td>rs13479048</td>
<td>0.7233</td>
<td>5.06E-04</td>
<td>6</td>
<td>132747781</td>
</tr>
<tr>
<td>2410127E16Rik</td>
<td>gnf1m16200_a at</td>
<td>rs4226547</td>
<td>0.7584</td>
<td>7.03E-03</td>
<td>7</td>
<td>29741607</td>
</tr>
<tr>
<td>9630047A9Rik</td>
<td>gnf1m21858_at</td>
<td>rs6230901</td>
<td>0.7216</td>
<td>1.71E-04</td>
<td>7</td>
<td>32095986</td>
</tr>
<tr>
<td>2410127E16Rik</td>
<td>gnf1m16200_a at</td>
<td>rs6226814</td>
<td>0.7224</td>
<td>5.38E-03</td>
<td>7</td>
<td>33164014</td>
</tr>
<tr>
<td>Ofyl464</td>
<td>gnf1m07159_a at</td>
<td>rs4226825</td>
<td>0.7298</td>
<td>2.30E-05</td>
<td>7</td>
<td>92158415</td>
</tr>
<tr>
<td>4930507A01Rik</td>
<td>gnf1m16966_a at</td>
<td>rs3696853</td>
<td>0.7262</td>
<td>1.70E-05</td>
<td>8</td>
<td>49044494</td>
</tr>
<tr>
<td>2310021F11Rik</td>
<td>gnf1m13321_a at</td>
<td>rs13480830</td>
<td>0.7037</td>
<td>2.80E-05</td>
<td>10</td>
<td>69713270</td>
</tr>
<tr>
<td>Mef2d</td>
<td>gnf1m29767_a at</td>
<td>rs13480877</td>
<td>0.7372</td>
<td>3.20E-05</td>
<td>10</td>
<td>86902245</td>
</tr>
<tr>
<td>2000001E21Rik</td>
<td>gnf1m14364_a at</td>
<td>rs13480711</td>
<td>0.7012</td>
<td>1.22E-04</td>
<td>10</td>
<td>95961668</td>
</tr>
<tr>
<td>Cyp9</td>
<td>gnf1m27917_s_at</td>
<td>rs13480799</td>
<td>0.7385</td>
<td>2.11E-04</td>
<td>10</td>
<td>125495706</td>
</tr>
<tr>
<td>4933439C20Rik</td>
<td>gnf1m18300_a at</td>
<td>rs13480836</td>
<td>0.7945</td>
<td>1.14E-03</td>
<td>11</td>
<td>3452106</td>
</tr>
<tr>
<td>1200002022Rik</td>
<td>gnf1m16716_s at</td>
<td>rs13481546</td>
<td>0.7407</td>
<td>1.20E-05</td>
<td>12</td>
<td>73732627</td>
</tr>
<tr>
<td>Rain9p</td>
<td>gnf1m03849_a at</td>
<td>rs13481795</td>
<td>0.7361</td>
<td>5.40E-05</td>
<td>13</td>
<td>45949979</td>
</tr>
<tr>
<td>6430695P16Rik</td>
<td>gnf1m19479_s at</td>
<td>rs4230598</td>
<td>0.7691</td>
<td>5.00E-06</td>
<td>14</td>
<td>50719913</td>
</tr>
<tr>
<td>Cab291</td>
<td>gnf1m13054_a at</td>
<td>rs4230594</td>
<td>0.7502</td>
<td>3.30E-04</td>
<td>14</td>
<td>50719913</td>
</tr>
</tbody>
</table>

ID, identification number; Chr, chromosome.
and exhibit the strain-specific variations in transcription. Using these data, we were also able to confirm that the transcriptional effects of the functional haplotype did not extend far beyond the Gst Mu cluster as the liver-expressed genes flanking the cluster, *Ahcyl1* and *Gnai3*, did not show the same level of transcriptional variability (Fig. 5).

These changes were confirmed via quantitative real-time PCR on RNA obtained from the liver tissue of 30 inbred strains (Supplemental Table 1). The results showed that the members of the Gst Mu family exhibit a reduced expression from a small but statistically significant 1.5-fold (*Gstm1*) to a phenotypic null 50-fold (*Gstm3*). The pattern of expression across the family members followed the haplotypic structure over the locus. The relative amounts of each isoform also confer a genetic fingerprint, with similar patterns of expression evident among the haplotypic groups (Fig. 6). Haplotypic structure correctly predicted the reduced Gst Mu transcription pattern observed in NZO/HILtJ. It is interesting to note that PWK/PhJ shares a haplotype structure with these strains over a more limited area and also clusters with other lower expressing strains. This minimal region of overlap appears as the causative factor for the altered expression (Fig. 6). These data represent a multimember gene family eQTL, with clearly delineated borders distinguishing both the gene family and the eQTL.

**Discussion**

We have used gene expression and genotyping microarray data from multiple inbred mouse strains to inform genetic variation that affects gene expression. The particular design and methodology associated with this microarray analysis identified a unique haplotype structure. In addition to the *Gstm5* eQTL that was examined in depth, our data contain other such haplotype structures, as well as potential CNVs, that significantly correlate with gene expression variation. These data provide a rich source of hypotheses for future examination. The successful identification of a true eQTL with confirmation and predictive biology, indicate the potential utility of the method.

Thirty-one associations were identified as having both a significant correlation and a colocalization of the probes for expression and SumSignal. Attempts to confirm these hits led to the identification of a haplotype structure that drastically alters the expression of the Gst Mu family of genes. This haplotype exists in eight laboratory mouse strains, NZO/HILtJ, PWK/PhJ, MSM/Ms, JF1/Ms, MOLF/EiJ, NZW/LacJ, RIIIS/J, and MA/MyJ, with expression changes confirmed in NZO/HILtJ, PWK/PhJ, NZW/LacJ, RIIIS/J, and MA/MyJ. Seven of these eight strains share the same haplotype over the entire *Gstm* locus, whereas PWK/PhJ mice only share the functional haplotype between *Gstm6* and *Gstm1*. Based on the genotyping of the wild-derived strains, the functional haplotype seems to be derived from the *Mus musculus molossinus* and *M. m. musculus* lineages despite the fact that PWK was the only *M. m. musculus* strain whose expression was confirmed.

The Gst Mu eQTL encompasses multitissue alleles for *Gstm2*, *Gstm3*, *Gstm5*, and *Gstm6* and, to a smaller extent, for *Gstm1*. Both liver and hypothalamus tissues exhibited significant reduction in expression. Because the relative amounts of GST Mu family members are consistent within haplotypic groups, we argue that the region is probably under the control of a locus control region (LCR). It is more likely that the controlling SNPs lay in one central region as opposed to individual changes in each of the Gst Mu family member promoters. Further experiments are needed to map the LCR and to determine which transcriptional factors are important for the control of Gst Mu expression. Based on the haplotype map presented here, we hypothesize that the LCR lies between *Gstm6* and *Gstm4*, a distance of approximately 84 kilobases.

The reduction in expression of the Mu class is specific and not related to a general down-regulation of glutathione S-transferases. Members of the Alpha and Theta classes of glutathione S-transferase showed no difference between these strains in the liver microarray data. Other glutathione-dependent enzymes, such as glutathione peroxidase, also were unaffected. Thus, it follows that only those metabolic pathways that are mediated solely by the Gst Mu class are
affected. It remains to be determined whether free glutathione levels are changed because of the reduction in Gst Mu levels in the livers or brains of the underexpressing strains, but it has been previously demonstrated that changes in transcription of Gst Mu genes directly affect Gst Mu protein levels and glutathione S-transferase enzymatic activity (Shao et al., 2007).

The alleles for Gstm2 and Gstm3 (and to a smaller extent for Gstm1) could have drastic effects in hepatic models of pharmacology/toxicology. Drug disposition, metabolism, and clearance are all dependent on both phase I and phase II metabolic enzymes; glutathione S-transferases are phase II enzymes (Hayes et al., 2005). Reduced expression of Gst Mu family members could lead to significant changes in clearance of drugs and toxins. Many of the Gst Mu family members are induced in response to metabolic or xenobiotic clues, and it remains to be seen whether such induction is changed in the affected strains (Hayes et al., 2005). In addition, those strains may be sensitive to accumulation of reactive oxygen species and oxidative damage.

The variable expression of the Gst Mu genes in inbred strains might provide a better model of human biology. It has been estimated that more than 50% of certain human populations are null for the GSTM1 locus, but not for the other GST Mu family members (Hayes and Strange, 2000). The types of copy number variation at GSTM1 are not limited to deletions; GSTM1 has been shown to be duplicated in...
certain Middle Eastern populations (McLellan et al., 1997). In fact, other glutathione S-transferases have copy number variations in human populations that are not found in mice. \( \text{GSTT1} \) is also homozygous null in roughly 15% of white populations and more than 64% of certain Asian populations (Hayes and Strange, 2000).

The null effect of \( \text{GSTM1} \), the ortholog of \( \text{Gstm1} \), has physiologic consequences in human populations (Bolt and Thier, 2006). The null genotype has been associated with increased risks of cancer of the lung, breast, and urogenital tract (Helzlsouer et al., 1998; Ateçs et al., 2005; Singh et al., 2008). In addition, it also has been linked to an increased risk for periodontitis, drug-induced hepatotoxicity, and atherosclerosis (de Waart et al., 2001; Concolino et al., 2007; Huang et al., 2007; Ueda et al., 2007). It is noteworthy that there have been associations drawn between the \( \text{GSTM1} \) null polymorphism and early onset Parkinson’s disease (Perez-Pastene et al., 2007). The high prevalence of the human null polymorphism, as well as a natural phenocopy in mice, lends support to possible benefits from lower GST Mu protein levels in addition to these clear deleterious consequences. These benefits could, in part, include longer half-lives of hormones and protein cofactors. Further characterization is needed to deduce both the beneficial and untoward effects of reduced GST Mu expression.

Dopamine and its metabolic products are proposed to play key roles in the etiology of Parkinson’s disease (Stokes et al., 1999). Dopamine produced by neural cells is either sequestered into synaptic vesicles by the vesicular monoamine transporters, undergoes oxidative deamination by monoamine oxidases, or undergoes auto-oxidation. Dopamine metabolism can produce free radicals, hydrogen peroxide, and reactive oxygen species such as quinones (Hastings et al., 1996). \( \text{Gst} \) enzymes are directly involved in the neutralization and excretion of these toxic compounds via their conjugation to glutathione, an important antioxidant. These species demonstrate cytotoxicity, with this toxicity being inhibited by antioxidant addition (Chiu et al., 2000). In addition, MAO inhibitors are used as a monotherapy in the treatment of Parkinson’s disease (Fernandez and Chen, 2007). Glutathione levels in the substantia nigra are depleted in patients with Parkinson’s disease, strengthening the link between oxidative stress, GST activity, and Parkinson’s disease (Bharath et al., 2002).

It is interesting to note that the addition of \( \text{GSTM2} \) to neuronal cell culture models confers a significant resistance to the cytotoxic effects of dopamine (Dagnino-Subiabre et al., 2000). In our experiments in brain tissue, \( \text{Gst Mu} \) isoforms that are normally expressed at substantial levels are reduced in expression in animals containing the affected haplotype. The probable consequence is that the neurons of these
brains are sensitive to oxidative damage not only from neurotoxins but also from increased neurodegeneration due to dopamine cytotoxicity. Expression of Gstm5 in the hypothalamus is essentially abolished in particular inbred strains. In a survey of all of the Gst isoforms (Alpha, Kappa, Mu, Omega, Pi, Sigma, Theta, and Zeta), Gstm5 was the only Gst highly expressed in the brain (Knight et al., 2007). Because it is the most prevalently expressed Gst isoform in brain tissues, it would follow that the consequences of Gstm5 ablation would be more severe. Because the Gstm5 transcript is reduced in approximately 90% in the strains with the affected haplotype, we propose that these strains will prove to be sensitive to \( \sigma \)-DOPA toxicity. Furthermore, we hypothesize that these strains should be sensitive to chemically induced parkinsonism from compounds such as 1-methyl-\( \sigma \)-phenyl-1,2,3,6-tetrahydropyridine.

We have identified haplotype structures associated with variable gene expression in inbred mouse strains using microarray-generated intensity values as the genotypic input. Furthermore, we identified several cis-acting QTLs affecting gene expression in inbred mouse hypothalami. Further characterization was performed on one of these alleles, covering the entire glutathione S-transferase Mu family of phase II detoxification enzymes. The drastic reduction of levels of these enzymes, particularly in the liver and brain, could have profound effects on both pharmacology and toxicology in these animals. Furthermore, as segments of this gene family are deleted in large portions of the human population, these inbred strains provide new genetic models of human biology, particularly as it relates to Parkinson’s disease.

Acknowledgments. We acknowledge Elizabeth Dominguez for her assistance with the PCR confirmation studies and Cam Hoang and Brian Hayes for critical review of the manuscript.

References


Address correspondence to: Dr. Mathew T. Pletcher, Investigative Toxicology, Pfizer Global Research and Development, Groton, CT 06340. E-mail: mathew.pletcher@pfizer.com