

# Genetic Control of Human Brain Transcript Expression in Alzheimer Disease

Jennifer A. Webster,<sup>1,2,3,16</sup> J. Raphael Gibbs,<sup>4,5,16</sup> Jennifer Clarke,<sup>6</sup> Monika Ray,<sup>7</sup> Weixiong Zhang,<sup>7,8</sup> Peter Holmans,<sup>9</sup> Kristen Rohrer,<sup>4</sup> Alice Zhao,<sup>4</sup> Lauren Marlowe,<sup>4</sup> Mona Kaleem,<sup>4</sup> Donald S. McCorquodale III,<sup>10</sup> Cindy Cuello,<sup>10</sup> Doris Leung,<sup>4</sup> Leslie Bryden,<sup>4</sup> Priti Nath,<sup>4</sup> Victoria L. Zismann,<sup>1,2</sup> Keta Joshipura,<sup>1,2</sup> Matthew J. Huentelman,<sup>1,2</sup> Diane Hu-Lince,<sup>1,2</sup> Keith D. Coon,<sup>1,2,11</sup> David W. Craig,<sup>1,2</sup> John V. Pearson,<sup>1,2</sup> NACC-Neuropathology Group,<sup>12</sup> Christopher B. Heward,<sup>13,17</sup> Eric M. Reiman,<sup>1,2,14</sup> Dietrich Stephan,<sup>1,2,14</sup> John Hardy,<sup>4,5</sup> and Amanda J. Myers<sup>10,15,\*</sup>

We recently surveyed the relationship between the human brain transcriptome and genome in a series of neuropathologically normal postmortem samples. We have now analyzed additional samples with a confirmed pathologic diagnosis of late-onset Alzheimer disease (LOAD; final  $n = 188$  controls, 176 cases). Nine percent of the cortical transcripts that we analyzed had expression profiles correlated with their genotypes in the combined cohort, and approximately 5% of transcripts had SNP-transcript relationships that could distinguish LOAD samples. Two of these transcripts have been previously implicated in LOAD candidate-gene SNP-expression screens. This study shows how the relationship between common inherited genetic variants and brain transcript expression can be used in the study of human brain disorders. We suggest that studying the transcriptome as a quantitative endo-phenotype has greater power for discovering risk SNPs influencing expression than the use of discrete diagnostic categories such as presence or absence of disease.

## Introduction

Our hypothesis is that genetic variation driving gene expression is integral to the processes involved in normal and pathogenic aging of the human cortex. In contrast to polymorphisms that change the amino acid sequence of genes, many single-nucleotide polymorphisms (SNPs) that are responsible for changes in gene expression might have more subtle phenotype effects that better explain diseases with complex modes of inheritance, including many diseases of late onset. We have performed whole-genome genotyping and transcriptome-expression analysis on a series of control and neuropathologically confirmed late-onset Alzheimer disease (LOAD) human brain samples to uncover human-cortex expression quantitative trait loci (eQTLs) that are relevant to both health and disease in the aged brain.

eQTLs are genomic sequence variants (as determined by genome-wide SNP analysis) that correlate with gene-expression differences (as determined by transcriptome-wide microarray experiments). eQTL studies are similar to traditional genetic-association studies, but instead of

associating genetic markers with discrete traits such as disease status, eQTL studies correlate genetic markers with quantitative gene-expression levels. It has been suggested that this could be a powerful and complementary approach to detection of novel risk loci.<sup>1</sup> The outputs from this type of screen are SNP-transcript pairs in which gene expression is correlated with genotype in a dose-dependent fashion.

Human genome- and transcriptome-wide eQTL analysis has been employed in the study of diseases such as asthma, cardiovascular disease, hypertension, and obesity with the use of human EBV-transformed immortalized B cell lines (ECLs<sup>2,3,4,5</sup>), as well as adipose<sup>5</sup> and liver<sup>6</sup> tissue. To our knowledge, other than our screen of control brain,<sup>7</sup> no eQTL studies using human cortical tissues have been published. Recent tissue microarray surveys report that a significant proportion of the human transcriptome is unique to the cortex;<sup>8,9</sup> therefore, defining the transcriptome-genome relationship in health and disease through the use of brain tissue is crucial for unlocking novel risk loci implicated in neurologic and psychiatric disorders.

<sup>1</sup>Neurogenomics Division, Translational Genomics Research Institute, Phoenix, AZ 85004, USA; <sup>2</sup>Arizona Alzheimer's Consortium, Phoenix, AZ 85006, USA; <sup>3</sup>Department of Statistics, Texas A & M University, College Station, TX 77843, USA; <sup>4</sup>Laboratory of Neurogenetics, National Institute on Aging, Porter Neuroscience Building, NIH Main Campus, Bethesda, MD 20892, USA; <sup>5</sup>Reta Lila Weston Institute and Department of Molecular Neuroscience and Neurodegenerative Disease, Institute of Neurology, UCL, Queen Square, London WC1N 3BG, UK; <sup>6</sup>Department of Epidemiology and Public Health, University of Miami School of Medicine, Clinical Research Building, Miller School of Medicine, Miami, FL 33133, USA; <sup>7</sup>Department of Computer Science and Engineering, Washington University, St. Louis, MO 63130, USA; <sup>8</sup>Department of Genetics, Washington University School of Medicine, St. Louis, MO 63110, USA; <sup>9</sup>Biostatistics and Bioinformatics Unit, Department of Psychological Medicine, Cardiff University, Cardiff, Wales CF14 4XN, UK; <sup>10</sup>Department of Psychiatry and Behavioral Sciences, Division of Neuroscience, Miller School of Medicine, University of Miami, Miami, FL 33136, USA; <sup>11</sup>Division of Thoracic Oncology Research, St. Joseph's Hospital and Medical Center, Phoenix, AZ 85013, USA; <sup>12</sup>Subset of the Alzheimer's Disease Neuropathology core group, various sites; see Acknowledgments; <sup>13</sup>Kronos Science Laboratory, Phoenix, AZ 85016, USA; <sup>14</sup>Banner Alzheimer's Institute, Phoenix, AZ 85006, USA; <sup>15</sup>Johnnie B. Byrd Sr. Alzheimer's Center and Research Institute, Tampa, FL 33613, USA

<sup>16</sup>These authors contributed equally to this work

<sup>17</sup>Deceased

\*Correspondence: [amyers@med.miami.edu](mailto:amyers@med.miami.edu)

DOI 10.1016/j.ajhg.2009.03.011. ©2009 by The American Society of Human Genetics. All rights reserved.

## Subjects and Methods

### Subjects

Our LOAD cohort was obtained concurrent to our control cohort<sup>7</sup> from 20 National Alzheimer's Coordinating Center (NACC) brain banks and from the Miami Brain Bank (see [Supplemental Data](#), available online, for a list of sites). Our criteria for inclusion were as follows: self-defined ethnicity of European descent, neuropathologically confirmed LOAD or no neuropathology present, and age of death greater than 65. We restricted the series to samples with European ethnicity because there are known allele-frequency differences between groups with different ethnicities and these differences might distort our genomic SNP-screen results. Neuropathological diagnosis was defined by board-certified neuropathologists as per standard NACC protocols. Samples that had a clinical history of stroke, cerebrovascular disease, Lewy bodies, or comorbidity with any other known neurological disease were excluded. LOAD or control neuropathology was confirmed by plaque and tangle assessment, 43% of the entire series having Braak staging.<sup>10</sup> Samples were de-identified before receipt, and the study met local human studies institutional review board and HIPAA regulations. This work is declared not human-subjects research and is IRB exempt under regulation 45 CFR 46. We received a total of 1174 tissue samples from the sites that replied to our request. Of those samples, ~65% that met criteria were prepared for DNA and RNA. Total sample counts equaled 486 LOAD cases and 279 neuropathologically confirmed controls. Of the plated DNA and RNA, we chose ~60% for inclusion in this pilot study. Inclusion at this stage was determined by assessment of RNA and DNA quality through visual inspection on gels.

### Genotyping and Expression Profiling

DNA was hybridized to the Affymetrix GeneChip Human Mapping 500K Array Set (502,627 SNPs) as previously described.<sup>11,12</sup> Genotypes were extracted with the use of both SNIper-HD<sup>13</sup> and BRLMM (Affymetrix, Santa Clara, CA) algorithms. Genotypes that exhibited less than 98% concordance between calls were excluded. SNPs with call rates less than 90% were excluded from the analysis. Hardy-Weinberg equilibrium (HWE) was assessed with exact tests and the PLINK analysis toolset.<sup>14</sup> SNPs with HWE exact-test *p* values less than 0.05, as well as SNPs with minor-allele frequencies less than 1%, were excluded. Allele calls had a mean of 97% and a range of 90%–99%. cRNA was hybridized to Illumina Human Refseq-8 Expression BeadChip (24,357 transcripts) via standard protocols. Expression profiles were extracted and rank invariant normalized<sup>15–17</sup> with the use of the BeadStudio software available from Illumina, with the Illumina custom error model used. Rank-invariant-normalized expression data were log<sub>10</sub> transformed, and missing data were encoded as missing, rather than as a zero level of expression. Chips with average detection scores less than 0.99 (5% of control chips, 8% of LOAD chips) were excluded from the analysis. Transcripts that were detected in less than 90% of the case or 90% of the control series were excluded from our study. This was a modified criterion from our original report and included 36% of probes on the entire chip. This was done in an effort to enrich for primary effects, particularly in the disease cohort. We assessed correlations between 380,157 SNPs and 8650 transcripts.

### Analysis

As in our prior report,<sup>7</sup> we ran the genotypes from our cohort concurrently in STRUCTURE,<sup>18,19</sup> with seven other populations.

To determine the degree of relatedness between the samples in our cohort, we used the pairwise IBS and IBD calculators available in the PLINK analysis toolset.<sup>14</sup> We plotted sample relatedness data in R, and we used the *cmdscale* function within the *stats* package to perform a multidimensional scaling analysis of the matrix of genome-wide IBS pairwise distances. For verification of the reported gender of samples, the sex-check functionality in the PLINK analysis toolset<sup>14</sup> was used. Samples identified as having a gender mismatch or gender ambiguity were excluded from analysis.

### Single-Locus Analysis

Prior to our analysis, transcript profiles were corrected for several covariates by use of the R suite of programs. Expression profiles were regressed against the sample data for several biological covariates (gender, apolipoprotein E [APOE; MIM +10774], allele dose, age at death, and cortical region) and several methodological covariates (day of expression hybridization, institute source of sample, postmortem interval [PMI], and a covariate based on the total number of transcripts detected in each sample). SNP genotype was not included in this model, because we were interested in covariate correction a priori to running our analysis. All covariates except age at death, APOE  $\epsilon$ 4 allele dose, and postmortem interval were coded as factors for this analysis. All samples had complete information for all covariates except PMI (*n* = 10 LOAD samples, 1 control sample). Average PMI values were used for these samples. Residuals from this analysis were employed in the single-locus eQTL analysis.

To determine our single-locus eQTL effects, we regressed the expression level of each transcript per sample on the number of minor alleles (0, 1, or 2) for the 380,157 SNPs that met the cutoff criteria. To capture both the effects that were independent of disease status and the eQTL effects in which there was an effect of disease, we included an interaction term for diagnosis in our model. To allow for situations in which the eQTL effect differs between LOAD cases and controls, we performed a two degree of freedom (2 df) test comparing a model in which transcript expression depends on the additive effects of SNP and disease status, together with their interaction term, against a model in which expression depends only on disease status. The main effect of diagnosis was included in the model for prevention of spurious associations occurring when SNP allele frequencies and overall expression levels differ independently between cases and controls. Including the interaction term allows the slope of the regression of expression on minor-allele dosage to differ between cases and controls. Allowing for differences between cases and controls via a 2 df test slightly reduces power for detection of SNP effects that are independent of disease status; however, this method greatly increases power for detection of effects that are disease specific. Furthermore, analyzing the entire cohort together considerably increases power as compared to performing separate tests in cases and controls. Uncorrected *p* values were retained from this analysis.

Genome-wide significance levels for each transcript were calculated by permutation of the subject labels, allowing for both non-normality of the transcript data and linkage disequilibrium (LD) among the SNPs. All permuted data sets were generated in the same manner, through randomly permuting sample identifiers in the transcript data (thus permuting all transcripts simultaneously). The replicate data sets were analyzed in the same way as the real data. Permutation analysis was performed on the Translational Genome Research Institute's IBM System Cluster 1350, which contains a total of 1024 computing nodes and is housed on the Arizona State University campus.

To capture the relative contributions of our case and control cohorts to the effects recovered in our 2 df model, we performed individual analysis on each series. We analyzed the same model without the diagnosis interaction term separately in cases and controls. Significance of effects was not determined in this modeling; therefore, permutation corrections were not performed. The correlation coefficients from these analyses were extracted with the use of the linear modeling (lm) functionality in R. The adjusted correlation coefficients ( $r^2_{\text{adj}}$  CO [controls] and  $r^2_{\text{adj}}$  CA [cases]) represent the portion of the expression variance explained by the SNP in each cohort, and the beta values ( $\beta$  cases and  $\beta$  controls) give the magnitude and direction of effects.

Fisher's exact tests<sup>20</sup> were employed for assessing associations between SNPs and diagnosis status without considering expression profiles. These were implemented in PLINK<sup>14</sup> and should correct for cases in which the allele frequencies of tested SNPs are sufficiently small as to make a chi-square test inappropriate.

To determine whether known SNPs mapped to the transcript probes, thus creating a possible false positive through non-biologically relevant differential hybridization, we used the Hapmap CEU data. Hapmap SNPs that were variable in the CEU cohort were mapped with respect to the transcript probes. LD was determined by the  $r^2$  statistic, and the  $r^2$  values between the significant SNP within our screen and the SNP within the probe from the CEU Hapmap data set were annotated within our results. It is presumed that SNP pairs with higher LD are more likely to be putative false positives if it is true that single SNP changes can affect the hybridization of the 50-mer transcript probes on the Illumina chips in a way that is not biologically relevant. However, all data are reported, because we have not made an independent assessment of the SNP-within-probe effect.

#### Network Analysis

We incorporated both transcripts that were significantly associated with allele dosage and those that were differentially expressed between LOAD cases and controls in our analysis. Our eQTL data set was defined as transcripts with a SNP-transcript correlation  $p$  value  $\leq 0.01$  after 1000 permutations and a SNP effect that was not in LD ( $r^2 \geq 0.8$ ) with a polymorphism in the probe. Our differentially expressed data set was defined as transcripts identified through significance analysis of microarrays (SAM<sup>21</sup>) with a false discovery rate (FDR)  $< 1\%$ . We used HQCUT<sup>22</sup> to form the networks from residual corrected transcript data, the Database for Annotation, Visualization and Integrated Discovery (DAVID)<sup>23,24</sup> to map gene ontologies to the clusters found by HQCUT, and the Network Edge Orienting program (NEO)<sup>25</sup> to incorporate SNPs into these networks. We determined hub genes by examining the distribution of the number of links per gene ( $n = 1697$ ). The cutoff level for hub genes was determined by the following formulae: Cutoff threshold =  $\mu + (2 \times S)$ , in which  $\mu$  is the mean number of links and  $S$  is the standard deviation.

#### Assessment of Disease-Specific Effects

We wanted to assess whether profiling a specific neuropathological disease was important in our recovery of new effects or whether these findings were simply a result of the increase in power through the doubling of our sample size. This is an important distinction, because for some diseases it might be difficult to sample from individuals with disease profiles; thus, knowing whether disease effects can be captured in nondisease tissue is important. To test this, we created a bootstrap control sample with 364 subjects by using R to sample the genotype and covariate data from the control portion of our data set with replacement and introducing a small amount of random variation (smoothed boot-

strap). We used PLINK<sup>14</sup> to evaluate the association between the minor-allele dose of each SNP and the transcript-expression level in the bootstrap sample, using the same 2 df model that we employed in the original analysis. Only SNP-transcript pairs that were significant (2 df test,  $\alpha = 0.05$ ) in our original analysis were analyzed, and these were stratified on the basis of those that had a significant interaction  $p$  value (interaction term  $\alpha = 0.05$ ,  $n = 665$ ) and those that did not (interaction term  $\alpha > 0.05$ ,  $n = 1820$ ). We then evaluated the proportion of SNP-transcript pairs that were detected in the LOAD and control data set and could be detected in the bootstrap control sample.

#### Replication

A portion of our SNP-transcript effects were replicated in our entire series through Taqman real-time PCR.<sup>26</sup> Multiplex single-tube relative-quantification measurements were obtained with the assays Hs01028610\_m1 (*ITPKB* [MIM 147522]) and Hs00907680\_g1 (*DKFZp566B183* [MIM 611623]). Primer-limited beta-2-microglobulin was used as the endogenous control. Efficiencies from simplex reactions were 97% for Hs01028610\_m1 and 100% for Hs00907680\_g1. Comparison of target and endogenous delta-CT values in simplex reactions showed similar efficiencies; the absolute values for slopes from delta-CT plots versus log input plots were less than  $< 0.1$ , justifying our use of the comparative delta-delta-CT method. All values were obtained from averages of three replicate reactions per sample. No-RT reactions for all assays were negative. Fold-change values were calculated relative to a normalizer sample present on all plates. Fold changes were corrected for all covariates with the use of R. Analysis was performed in PLINK as before, with the use of the corrected residuals. Multiple test adjustments were computed with a Dunn-Sidak correction.<sup>27</sup>

## Results

### Sample

Table S1 gives all sample information. Two LOAD samples (WGAAD-75 and WGAAD-270; see Figure S1) were found to be ethnic outliers and were excluded from our analysis. Samples were also screened for assessment of whether they had some degree of relatedness, and all LOAD samples were found to be unrelated. The following samples were removed from the analysis because their genetically predicted gender did not match their reported gender: WGACON-105 (reported female appears to be male), WGACON-108 (reported male appears ambiguous), WGACON-212 (reported male appears ambiguous), WGAAD-252 (reported male appears to be female), WGAAD-465 (reported male appears to be female), and WGAAD-479 (reported female appears to be male). Our total cohort included 364 human brain tissue samples; 176 LOAD cases and 188 controls after all exclusions.

Assessment of the expression profiles in the LOAD cohort found that approximately 58% of the transcripts on the Illumina Human RefSeq-8 Chip were expressed in at least 5% of our LOAD cohort. This is the same portion of the transcriptome that was found expressed in brain in the control series, suggesting that there is not a major skew in our LOAD-series profiles. Figure S2 graphs the portion of significant transcript effects (uncorrected  $\alpha = 0.05$ ) for each covariate in both the LOAD and the control

cohorts. Note that our largest effects were with hybridization date (methodological covariate) and *APOE* status (biological covariate), other covariates having less of an effect. PMI was the covariate that had the smallest effect on our data.

### Single-Locus Analysis

*Cis* effects and *trans* effects were defined as described previously;<sup>7</sup> *cis* associations were those that involved SNPs within 1 Mb of the 5' end of the gene or 1 Mb of the 3' end of the gene and within the gene. This criterion was utilized for consistency with our previous study. From a practical perspective, it means that for each transcript, ~1% of the genome is defined as *cis* and 99% as *trans*. It is important to note that all analyses were performed genome-wide, such that all SNPs irrespective of location were run against each transcript and permutations were performed to obtain genome-wide significance. Thus, for any of our SNP-transcript pairs, other definitions of *cis* versus *trans* localizations can be applied, because our significance calculations do not rely on distance.

After correction with 1000 permutations, approximately 9% of the transcripts showed a genome-wide significant correlation with SNP genotype in the combined sample of cases and controls ( $p \leq 0.05$ , 1000 permutations). We identified 1829 SNP-transcript pairs with effects in *cis* in the total cohort and 656 *trans* pairs, in which the SNP mapped greater than 1 Mb from the 5' or 3' end of the gene. We found that 27% of all eQTL had a significant interaction with diagnosis. For each phase of our study, 37% of *cis* LOAD-interacting candidate transcripts in this screen overlap with those found in our first report, which considered only control brains.<sup>7</sup> The overlap with a large subset of candidates found in the analysis of disease-free controls suggests that end-stage neuropathological tissue does not create a bias in expression profiling and that contributions stemming from signal degradation and potential sampling differences are minimal. This also suggests that within our study, several of the SNP-transcript LOAD effects are risk factors, much like the *APOE* E4 allele, and appear in a portion of controls. For these effects, we assume that within LOAD, there are additional environmental or unassessed biological risk factors that cause predisposition to LOAD. There is an 18% overlap with our previous report considering both *cis* and *trans* results with no interaction with diagnosis, and there is a 9% overlap with our previous report considering all results with a positive interaction term.

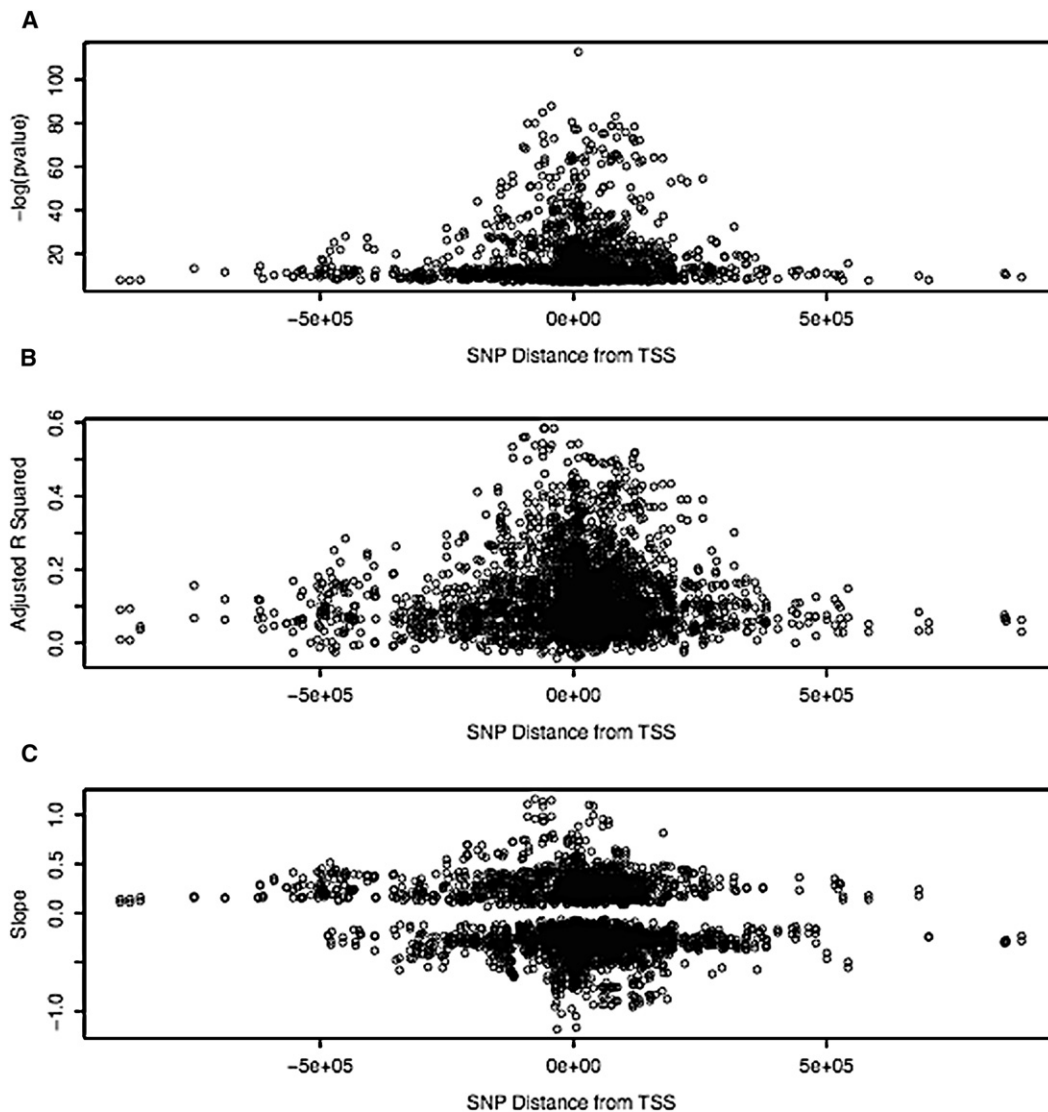
Figure 1 plots the  $-\log$  of the uncorrected p value ( $-\log$  (p value)), the  $\beta$  values (slope), and the adjusted correlation coefficients (adjusted  $r^2$ ) versus distance from the transcription start site (TSS) for all of our *cis* effects. Many of our *cis* effects mapped far downstream or upstream of the TSS (Figure 1A). Although it is possible that there are regulatory elements that act over such long distances, finer mapping will need to be performed for determination of the extent to which they are true effects and of the extent

to which they reflect LD in these regions. In Figures 1B and 1C, it is of note that the size of effect is inversely proportional to the distance from the TSS, as measured by either the proportion of variance explained (Figure 1B) or the slope (Figure 1C). This suggests that it is probable that true effects map close to the TSS. We found approximately as many *cis* negative correlations (minor allele is high expresser) as *cis* positive correlations (major allele is high expresser), greater differences being found between the alleles mapping closer to the TSS (see Figure 1C). Table 1 lists the top eQTL results in which there was a significant interaction with diagnosis, and Table 2 lists the top eQTL results in which there appeared to be no diagnosis effect. All results are listed in Tables S5 and S6.

Although each SNP-transcript pair is unique to either the LOAD-interacting data set (Table 1, Table S5) or the noninteraction data set (Table 2, Table S6), some of our transcripts overlap between these lists. For our top results in Tables 1 and 2, abhydrolase domain-containing 12 (*ABDH12*), ankyrin repeat domain 27 (*ANKRD27*), ninein-like protein (*KIAA0980*, [MIM \*609580]), kinesin family member 1B (*KIF1B*, [MIM \*605995]), inorganic pyrophosphatase 2 (*PPA2*, [MIM \*609988]), and peptidylprolyl isomerase-like protein 3 (*PPIL3*) all occur on both Table 1 (diagnosis interaction effect) and Table 2 (no interaction). To investigate why this overlap is occurring, we examined the SNPs mapping to these transcripts and found two effects. In the first case, the main effect appears to be with diagnosis, but some SNPs just missed our cutoff for having a significant interaction term ( $\alpha = 0.05$ ). For SNPs in *ABDH12*, *ANKRD27*, *KIAA0980*, and *KIF1B*, as the LD between the main-effect SNPs in Table 1 and the additional SNPs in the gene gets weaker, the significance of the interaction term erodes until it slips below our  $\alpha = 0.05$  cutoff, causing these subthreshold lower-LD SNPs to occur in Table 2. In the second case, for the transcripts *PPA2* and *PPIL3*, there are actually separate LD blocks that both contain eQTL SNPs, but only one of the blocks has a significant interaction term. In this situation, two separate eQTL effects are observed for the single transcript, and only one of them has a significant interaction with LOAD. This effect is biologically plausible in that SNPs in each block might be changing different transcription-factor binding sites, one of which is implicated in LOAD risk and the other of which is not.

### Network Analysis

In our original report, each transcript was treated as a single unit and run against the SNP profiles for the entire genome. This type of analysis does not consider the interdependence of transcript profiles or the network of effects that might occur within the human brain. To capture this additional information, we performed an initial network analysis of a subset of our eQTL transcripts. For this analysis, we also included transcripts for which there was no correlation of expression with genotype but was a significant difference in transcript expression between cases and controls. This would help in determining whether



**Figure 1. *Cis* Association Metrics Versus Distance from Transcription Start Site**

Plotted on the x axis is the distance from the TSS mapped on the human genome build 36. y axis metrics are given, with slope in Figure 1C corresponding to the  $\beta$  values from the correlations. Note that  $-\log(\text{p value})$  values were calculated in the 2 df model with the use of the entire series, whereas the  $\beta$  values and the adjusted correlation coefficients were calculated separately in cases and controls, but all values are plotted.

there were additional effects of our eQTL on downstream expression that were relevant to disease and would not be captured in a single-locus analysis.

Of the transcripts we used for our network analysis, ~27% were significantly correlated with minor-allele dose (eQTL data set defined by 1000 permutations,  $\alpha = 0.01$ ). The remaining transcripts were included because they were defined as being differentially expressed through significance analysis of microarrays (SAM<sup>21</sup>) with an FDR cutoff  $\leq 0.1\%$  (differentially expressed [DE] data set). The HQCUT algorithm recovered nine clusters in this data set. Ten transcripts did not cluster within the network. These transcripts were all from our eQTL data set. Figure 2 shows the resulting adjacency matrix. Table 3 gives summary statistics for each cluster, including the major

ontology group mapped with the DAVID annotation for each cluster.<sup>23,24</sup> The full list of top ontology keywords for each cluster is given in Tables S2A and S2B. Several of these ontology groupings matched results from previous LOAD-expression-only screens.<sup>28</sup>

To further assess our network, we examined whether any transcripts within the network had an overrepresentation of links to other transcripts (“hub genes”), suggesting that they were controlling many downstream effects. The essential nature of hub genes has been demonstrated within studies using yeast.<sup>29</sup> In our data set, the mean number of links in our network was 56 and the standard deviation was 38.78; thus, our cut off was 134 links, which is indicated in Figure S3. Within our eQTL and DE network, there were 62 hub genes, which are shown in Table 4.

**Table 1. Top Transcript-SNP Pairs with a Significant Effect of Diagnosis**

Probe ID <sup>a</sup>	Symbol <sup>b</sup>	Chr. <sup>c</sup>	Bp Position <sup>d</sup>	Best SNP <sup>e</sup>	S_Chrom. <sup>f</sup>	S_Pos. <sup>g</sup>	Other <sup>h</sup>	Loc. <sup>i</sup>	Raw p Value <sup>j</sup>	1K p Value <sup>k</sup>	Inter p Value <sup>l</sup>	r <sup>2</sup> CA <sup>m</sup>	r <sup>2</sup> CO <sup>n</sup>	Dir. <sup>o</sup>
GI_34147330-S	ABDH12	20	25.2–25.3	rs6050598	20	25.3	40	<i>cis</i>	$1.540 \times 10^{-65}$	$9.990 \times 10^{-4}$	$3.770 \times 10^{-2}$	0.44	0.38	+
GI_14149802-S	ANKRD27	19	37.8–37.9	rs259228	19	37.8	13	<i>cis</i>	$1.190 \times 10^{-31}$	$9.990 \times 10^{-4}$	$4.160 \times 10^{-2}$	0.17	0.31	-
GI_14916517-S	AP3M1	10	75.6	rs26133	5	14.2	1	<i>trans</i>	$3.280 \times 10^{-9}$	$1.500 \times 10^{-2}$	$7.080 \times 10^{-7}$	0.13	-0.05	+/-
GI_38261968-S	APG12L	5	115.2	rs2483589	10	116.4	1	<i>trans</i>	$7.930 \times 10^{-9}$	$2.000 \times 10^{-2}$	$9.490 \times 10^{-4}$	0.12	-0.04	-
GI_29739147-S	ARL16	17	77.3	rs6565624	17	77.3	3	<i>cis</i>	$4.360 \times 10^{-12}$	$9.990 \times 10^{-4}$	$4.520 \times 10^{-2}$	0.04	0.13	-
GI_38327635-S	C14orf4	14	76.6	rs1120277	14	76.5	0	<i>cis</i>	$1.840 \times 10^{-27}$	$9.990 \times 10^{-4}$	$2.900 \times 10^{-2}$	0.08	0.27	-
GI_23346408-S	C20orf111	20	42.3	a-2038955	1	19.6	0	<i>trans</i>	$2.040 \times 10^{-9}$	$2.000 \times 10^{-2}$	$3.440 \times 10^{-6}$	0.14	-0.05	-/+
GI_25092657-S	C7orf27	7	2.5–2.6	rs3936086	16	13.6	1	<i>trans</i>	$5.520 \times 10^{-9}$	$1.600 \times 10^{-2}$	$1.050 \times 10^{-6}$	0.15	-0.05	-/+
GI_37059735-S	CWF19L1	10	102	rs11594333	10	101.9	6	<i>cis</i>	$4.650 \times 10^{-11}$	$9.990 \times 10^{-4}$	$7.120 \times 10^{-2}$	-0.01	0.13	+
GI_39995081-S	FLJ20303	5	6.7	rs4701742	5	66.7	10	<i>cis</i>	$1.180 \times 10^{-51}$	$9.990 \times 10^{-4}$	$5.820 \times 10^{-4}$	0.24	0.41	+
GI_32698961-S	GDPD1	17	54.7	rs7223491	17	54.6	5	<i>cis</i>	$5.770 \times 10^{-15}$	$9.990 \times 10^{-4}$	$3.110 \times 10^{-2}$	0.15	0.01	-
GI_30061497-I	GGTL3	20	32.9	rs6087619	20	32.8	23	<i>cis</i>	$2.220 \times 10^{-33}$	$9.990 \times 10^{-4}$	$4.230 \times 10^{-2}$	0.14	0.33	-
GI_22095346-S	IGSF4	11	114.5–114.9	rs7125361	11	114.6	5	<i>cis</i>	$3.020 \times 10^{-26}$	$9.990 \times 10^{-4}$	$1.270 \times 10^{-2}$	0.24	0.14	-
GI_29171690-I	ILVBL	19	15.1	rs11851301	14	94.2	0	<i>trans</i>	$1.780 \times 10^{-9}$	$5.990 \times 10^{-3}$	$2.990 \times 10^{-3}$	0.12	-0.05	-
GI_37059782-S	KIAA0980	20	25.4–25.5	rs6132819	20	25.2	35	<i>cis</i>	$1.750 \times 10^{-29}$	$9.990 \times 10^{-4}$	$2.580 \times 10^{-2}$	0.29	0.15	+
GI_41393558-I	KIF1B	1	10.2–10.4	rs12120191	1	10.3	22	<i>cis</i>	$1.530 \times 10^{-79}$	$9.990 \times 10^{-4}$	$3.530 \times 10^{-2}$	0.5	0.41	-
GI_42658911-S	LOC401470	8	93.2	rs11171739	12	54.8	5	<i>trans</i>	$4.540 \times 10^{-58}$	$9.990 \times 10^{-4}$	$3.120 \times 10^{-2}$	0.48	0.27	+
GI_38455399-S	MADH5	5	135.5	rs2483589	10	116.4	1	<i>trans</i>	$3.010 \times 10^{-8}$	$2.100 \times 10^{-2}$	$1.090 \times 10^{-3}$	0.12	-0.04	-
GI_5174552-S	MFAP1	15	41.9	rs17546037	5	7.3	1	<i>trans</i>	$7.340 \times 10^{-9}$	$5.990 \times 10^{-3}$	$1.050 \times 10^{-3}$	0.13	-0.04	-
GI_7662347-S	MONDOA	12	121.0–121.2	rs3741452	12	121.2	1	<i>cis</i>	$2.620 \times 10^{-28}$	$9.990 \times 10^{-4}$	$3.550 \times 10^{-5}$	0.04	0.32	-
GI_31881619-A	PPA2	4	106.5–106.6	rs2866799	4	106.7	19	<i>cis</i>	$1.260 \times 10^{-24}$	$9.990 \times 10^{-4}$	$3.380 \times 10^{-2}$	0.27	0.1	-
GI_45439315-I	PPIE	1	40	rs1180341	1	39.8	2	<i>cis</i>	$3.260 \times 10^{-14}$	$9.990 \times 10^{-4}$	$2.800 \times 10^{-2}$	0.21	0.02	-
GI_19557635-A	PPIL3	2	201.4–201.5	rs6715380	2	201.9	14	<i>cis</i>	$2.340 \times 10^{-12}$	$9.990 \times 10^{-4}$	$2.470 \times 10^{-2}$	0.02	0.13	-
GI_15011935-S	RPS26	12	54.7	rs11171739	12	54.8	5	<i>cis</i>	$1.150 \times 10^{-70}$	$9.990 \times 10^{-4}$	$2.740 \times 10^{-2}$	0.51	0.33	+
GI_7657430-S	SND1	7	127.1–127.5	rs1408015	13	106.9	1	<i>trans</i>	$3.150 \times 10^{-8}$	$3.800 \times 10^{-2}$	$1.790 \times 10^{-3}$	0.14	-0.05	+
GI_5730084-S	TCTEL1	6	159	rs10843090	12	28.1	0	<i>trans</i>	$1.430 \times 10^{-8}$	$3.000 \times 10^{-2}$	$1.760 \times 10^{-3}$	0.13	-0.04	+/-
GI_8922751-S	TYW1	7	66.1–66.3	rs6966142	7	66	8	<i>cis</i>	$1.180 \times 10^{-12}$	$9.990 \times 10^{-4}$	$2.330 \times 10^{-2}$	0	0.14	+
GI_41281748-I	ZNF323	6	28.4	rs149970	6	28.1	8	<i>cis</i>	$8.350 \times 10^{-10}$	$6.990 \times 10^{-3}$	$2.510 \times 10^{-2}$	0.12	-0.02	+

This table shows the list of transcript-SNP pairs that had transcript-specific empirically significant p values by our 2 df model and where there was a significant diagnosis interaction term. Listed is a subset of the effects, generated from the full list of associated transcript-SNP pairs (Table S5). Criteria for generation of subset includes: transcript-specific 2 df model empirical p value from 1000 simulations  $\leq 0.05$ , gene-expression detection rate within both cases and controls  $\geq 99\%$ , SNP call rate  $\geq 99\%$  overall, minor-allele frequency  $\geq 5\%$  in cases and controls, no LD between eQTL SNP and any polymorphism in probe ( $r^2 < 0.8$ ), and SNP accounting for 12% or greater of the variance in transcript expression in either cases or controls (adjusted  $r^2$  in cases or controls  $\geq 0.12$ ). Genes are listed in alphabetical order. All data are also represented in Table S5.

<sup>a</sup> GI illumina probe identifiers.

<sup>b</sup> Correlated transcripts.

<sup>c</sup> Transcript chromosome.

<sup>d</sup> Transcript position (listed in Mb).

<sup>e</sup> SNP with the smallest 2 df model-corrected (1000 permutations) p value for that transcript (dbSNP or Affymetrix identifiers listed).

<sup>f</sup> SNP chromosome.

<sup>g</sup> SNP position (listed in Mb).

<sup>h</sup> Number of other SNPs with a 2 df model p value for the transcript that was significant after 1000 permutations ( $\alpha = 0.05$ ).

<sup>i</sup> Where the SNP was in *cis* or *trans* to the transcript.

<sup>j</sup> Uncorrected p value for the 2 df model.

<sup>k</sup> p value after correction of the 2 df model using 1000 permutations.

<sup>l</sup> p values for the interaction term in the 2 df model.

<sup>m</sup> Adjusted correlation coefficients for the 2 df model in cases.

<sup>n</sup> Adjusted correlation coefficients for the 2 df model in controls.

<sup>o</sup> Direction of the effect as indicated by the sample-specific  $\beta$  values. A plus sign indicates positive correlation and a major allele associated with high expression, and a minus sign indicates negative correlation and a minor allele associated with high expression. "+/-" or "-/+" indicates that direction of effect is different in case and controls, with the data given as direction in cases/direction in controls. Note that the latter effects were seen in *trans* and suggest a more complex model for effect whereby the uncaptured intermediary acts differently within case and control samples.

Further examination of our network analysis found that proportionally, most of our eQTLs were located in clusters 5 (44%), 2 (42%), and 3 (40%). Thirty-four percent of cluster 2 transcripts had a significant interaction with diagnosis with at least one SNP, a higher proportion than any other cluster (see Table 3). From our list of hub effects (see Table 4), 6% of LOAD eQTL transcripts in cluster 2 had greater than 134 genes linked to them, a greater portion of hub transcripts than in any other cluster. Thus, for clarity in this preliminary report, we focused on cluster 2; however, all data are available for analysis on the additional clusters (see Web Resources). To analyze this cluster, we used the program NEO.<sup>25</sup> We used SNPs to anchor the edges of the network and infer causal relationships between pairs of coregulated transcripts. A portion of the resulting network is diagrammed in Figure 3.

### Taqman Replication of Effects

For our data sets, we have replicated several effects via other methodologies. First, within these samples, we had previously shown that expression of the Microtubule-Associated Protein Tau gene (*MAPT* [MIM +157140]) is affected by SNPs mapping to the gene, by using qRT-PCR.<sup>30</sup> Analysis of our data from the genome-wide screen is consistent with our previous data on these samples, in that alleles that occur on the major haplotype of *MAPT* (H1) are associated with higher Tau transcript expression. This provides an internal positive control within our full-genome screen in that looking genome-wide, we can find effects that we have seen in candidate-gene analysis of our samples. Additionally, we have replicated our SNP-transcript-diagnosis relationship for *ITPKB* and rs12125724 (Dunn-Sidák-adjusted 2 df test  $p$  value =  $2.76 \times 10^{-4}$ ) by using Taqman assays and the same model that we applied to the whole-genome and transcriptome data; however, for *DKFZp566B133* and rs9309291, our Taqman analysis results did not survive multiple testing correction.

### Candidates—*cis* eQTL

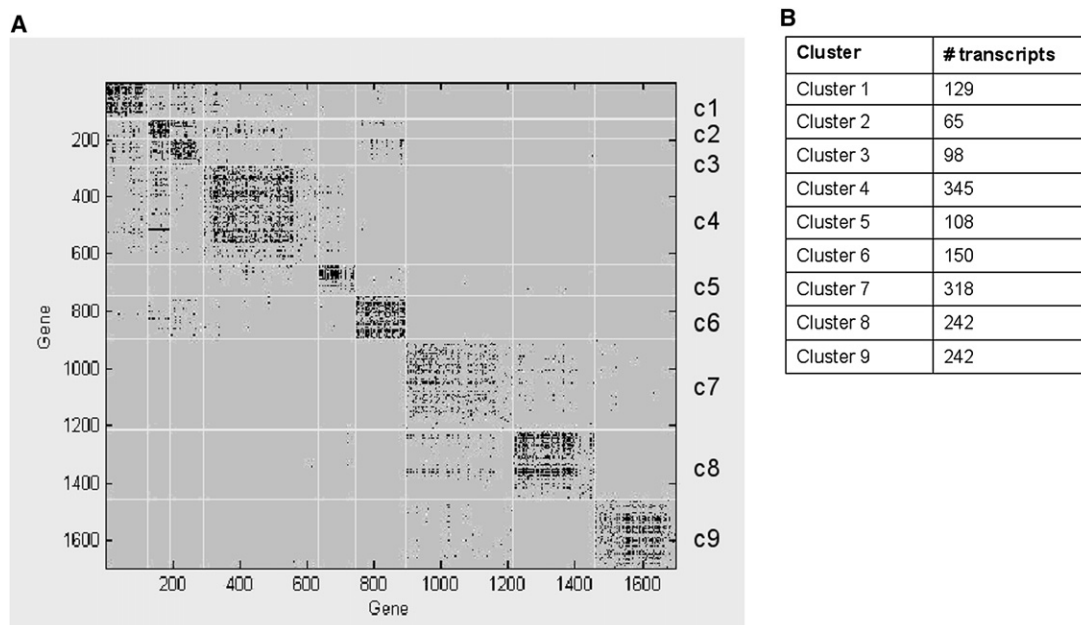
In terms of LOAD-specific effects, two LOAD candidate-gene studies have been performed, examining SNP variation and gene expression. We have replicated both of these results; however, more work clearly needs to be done in independent series for determination of whether a greater portion of our results replicate. In the first candidate study, Lambert et al.<sup>31</sup> examined the Transcription Factor CP2 gene (*TFCP2/LBP-1c/CP2/LSF* [MIM 189889]) and found that a SNP within the 3' UTR of the gene had a genetic association with LOAD risk. Additionally, this report examined expression in lymphoblasts from patients genotyped for the 3' UTR polymorphism and found that absence of the A allele at this site was associated with a lower transcript expression. In our screen, although we did not obtain a significant association by considering only SNP genotypes localized to this gene, we did obtain a significant eQTL effect when considering expression in combination with genotype and diagnosis.

**Table 2. Top Transcript-SNP Pairs without a Significant Diagnosis Interaction**

Probe ID	Symbol	Chr.	Bp Position	Best SNP	S_Ch	S_Pos	Other	Loc.	Raw p Value	1K p Value	Inter p Value	r <sup>2</sup> CA	r <sup>2</sup> CO	Dir.
GI_34147330-S	ABDH12	20	25.2–25.3	rs4815412	20	25.3	40	<i>cis</i>	$6.890 \times 10^{-64}$	$9.990 \times 10^{-4}$	$5.2030 \times 10^{-2}$	0.43	0.38	+
GI_8923854-A	AD-017	3	52.7	rs1961958	3	52.6	31	<i>cis</i>	$1.450 \times 10^{-53}$	$9.990 \times 10^{-4}$	$7.0820 \times 10^{-1}$	0.34	0.32	-
GI_22748838-S	AHSA2	2	61.3	rs720201	2	61.2	19	<i>cis</i>	$8.940 \times 10^{-24}$	$9.990 \times 10^{-4}$	$2.0130 \times 10^{-1}$	0.25	0.13	-
GI_14149802-S	ANKRD27	19	37.8–37.9	rs438268	19	37.8	13	<i>cis</i>	$1.840 \times 10^{-36}$	$9.990 \times 10^{-4}$	$1.3840 \times 10^{-1}$	0.23	0.31	+
GI_23238227-S	CHST7	X	46.3	rs760697	X	46.3	3	<i>cis</i>	$1.890 \times 10^{-23}$	$9.990 \times 10^{-4}$	$3.374 \times 10^{-1}$	0.23	0.19	+
GI_25306276-S	EFG1	3	159.8–159.9	rs9832803	3	159.9	7	<i>cis</i>	$1.420 \times 10^{-34}$	$9.990 \times 10^{-4}$	$5.860 \times 10^{-1}$	0.29	0.20	-
GI_31542792-S	FN3K	17	78.3	rs11078015	17	78.4	16	<i>cis</i>	$3.130 \times 10^{-79}$	$9.990 \times 10^{-4}$	$6.905 \times 10^{-1}$	0.48	0.45	+
GI_24431963-S	HBS1L	6	135.3–135.4	rs4646871	6	135.3	15	<i>cis</i>	$1.080 \times 10^{-27}$	$9.990 \times 10^{-4}$	$7.053 \times 10^{-1}$	0.24	0.19	-
GI_37059782-S	KIAA0980	20	25.4–25.5	rs2258719	20	25.2	35	<i>cis</i>	$1.860 \times 10^{-28}$	$9.990 \times 10^{-4}$	$5.377 \times 10^{-2}$	0.27	0.17	+
GI_21314694-S	KIAA1449	3	39.1	rs811971	3	39.1	22	<i>cis</i>	$5.080 \times 10^{-39}$	$9.990 \times 10^{-4}$	$2.428 \times 10^{-1}$	0.31	0.27	+
GI_41393558-I	KIF1B	1	10.2–10.4	rs10492972	1	10.3	22	<i>cis</i>	$1.650 \times 10^{-66}$	$9.990 \times 10^{-4}$	$8.820 \times 10^{-2}$	0.43	0.40	-
GI_4826813-A	KTN1	14	55.1–55.2	rs10137340	14	55.1	4	<i>cis</i>	$2.650 \times 10^{-23}$	$4.995 \times 10^{-3}$	$2.312 \times 10^{-1}$	0.12	0.21	-
GI_24308385-S	LOC115294	8	52.9	rs17212214	8	53	21	<i>cis</i>	$2.020 \times 10^{-28}$	$9.990 \times 10^{-4}$	$7.537 \times 10^{-1}$	0.24	0.20	-
GI_37546921-S	LOC339804	2	61.2	rs720201	2	61.2	13	<i>cis</i>	$1.590 \times 10^{-25}$	$9.990 \times 10^{-4}$	$2.113 \times 10^{-1}$	0.25	0.15	+
GI_40786403-S	LOC374969	1	43.0–43.1	rs323720	1	43	12	<i>cis</i>	$6.260 \times 10^{-37}$	$9.990 \times 10^{-4}$	$4.611 \times 10^{-1}$	0.26	0.31	+
GI_28603839-S	MGC10120	10	102	rs6584356	10	102	3	<i>cis</i>	$9.020 \times 10^{-37}$	$1.998 \times 10^{-3}$	$7.180 \times 10^{-2}$	0.27	0.26	-
GI_33859747-S	MGC2752	19	63.8	rs10413586	19	63.8	1	<i>cis</i>	$3.940 \times 10^{-32}$	$9.990 \times 10^{-4}$	$3.401 \times 10^{-1}$	0.23	0.23	-
GI_31652223-A	MRPL21	11	68.4	rs478647	11	68.5	10	<i>cis</i>	$1.850 \times 10^{-35}$	$9.990 \times 10^{-4}$	$2.100 \times 10^{-1}$	0.31	0.19	-
GI_28872733-I	MRPL43	10	102.7	rs4917916	10	102.7	5	<i>cis</i>	$1.520 \times 10^{-23}$	$9.990 \times 10^{-4}$	$6.832 \times 10^{-1}$	0.22	0.18	-
GI_29826284-A	MRPL47	3	180.8	rs10513762	3	180.8	2	<i>cis</i>	$1.750 \times 10^{-25}$	$9.990 \times 10^{-4}$	$4.383 \times 10^{-1}$	0.23	0.12	-
GI_6912539-S	NUBP2	16	1.8	rs3848348	16	1.8	18	<i>cis</i>	$9.170 \times 10^{-31}$	$9.990 \times 10^{-4}$	$7.720 \times 10^{-1}$	0.20	0.24	+
GI_40288186-A	PDLIM2	8	22.5	rs2291232	8	22.5	11	<i>cis</i>	$8.470 \times 10^{-43}$	$9.990 \times 10^{-4}$	$6.109 \times 10^{-1}$	0.25	0.31	-
GI_30179910-A	PILRB	7	99.8	rs1000215	7	99.4	6	<i>cis</i>	$4.330 \times 10^{-28}$	$9.990 \times 10^{-4}$	$9.318 \times 10^{-1}$	0.24	0.18	+
GI_31881619-A	PPA2	4	106.5–106.6	a-1897043	4	106.5	19	<i>cis</i>	$3.770 \times 10^{-28}$	$9.990 \times 10^{-4}$	$8.078 \times 10^{-1}$	0.22	0.20	-
GI_19557635-A	PPIL3	2	201.4–201.5	rs10205561	2	201.5	14	<i>cis</i>	$2.210 \times 10^{-32}$	$9.990 \times 10^{-4}$	$6.970 \times 10^{-1}$	0.29	0.19	-
GI_28274698-S	RAB3IP	12	68.4–68.5	rs11177823	12	68.4	3	<i>cis</i>	$1.280 \times 10^{-44}$	$9.990 \times 10^{-4}$	$1.424 \times 10^{-1}$	0.34	0.26	-
GI_14149701-S	RNF167	17	4.8	rs238237	17	4.4	4	<i>cis</i>	$2.600 \times 10^{-29}$	$9.990 \times 10^{-4}$	$1.721 \times 10^{-1}$	0.32	0.14	+
GI_4506738-S	RPS6KB2	11	67	rs1476792	11	67	2	<i>cis</i>	$1.990 \times 10^{-21}$	$9.990 \times 10^{-4}$	$5.155 \times 10^{-2}$	0.25	0.09	-
GI_34147665-S	STK25	2	242.1	rs3755400	2	242	1	<i>cis</i>	$2.170 \times 10^{-31}$	$9.990 \times 10^{-4}$	$6.325 \times 10^{-1}$	0.25	0.23	-
GI_6912707-S	TIMM10	11	57.1	rs2729371	11	57.1	6	<i>cis</i>	$1.590 \times 10^{-72}$	$9.990 \times 10^{-4}$	$2.506 \times 10^{-1}$	0.42	0.40	-
GI_15451941-S	UBA52	19	18.5	rs1468475	19	18.5	6	<i>cis</i>	$3.550 \times 10^{-81}$	$9.990 \times 10^{-4}$	$3.116 \times 10^{-1}$	0.43	0.46	+
GI_4557872-S	UROS	10	127.5	rs1571278	10	127.5	14	<i>cis</i>	$7.880 \times 10^{-28}$	$9.990 \times 10^{-4}$	$9.659 \times 10^{-1}$	0.20	0.21	-
GI_37622342-A	ZNF266	19	9.4	rs4804436	19	9.4	23	<i>cis</i>	$6.870 \times 10^{-44}$	$9.990 \times 10^{-4}$	$5.462 \times 10^{-1}$	0.33	0.26	-

This table lists transcript and SNP pairs that had transcript-specific empirically significant p values by our 2 df model and where there was a nonsignificant diagnosis interaction term. Criteria for selection of the subset is as described in Table 1, except that the SNP had to account for greater than 20% of the variance in transcript expression in either cases or controls. The full data set is available in Table S6. All headers are as described in Table 1.





**Figure 2. Adjacency Matrix of the eQTL and DE Network**

Figure 2A shows the adjacency matrix for our HQCUT network analysis. Modules in this figure are labeled c1–c9. Genes are listed by map number (1–1697) that was assigned during the clustering procedure and is arbitrary. Dots represent inter- and intramodule relationships between transcripts. Figure 2B shows the number of transcripts per cluster.

Our second LOAD hit that has been implicated in prior studies is the glutathione-S-transferase (*GSTO 2* [MIM 612314]) locus on chromosome 10q. This gene maps within a linkage peak found in many family-based LOAD linkage screens (for meta-analysis, see Hamshere et al.<sup>32</sup>) and thus was studied as a LOAD candidate gene by Li and colleagues.<sup>33</sup> Their approach was slightly different than a standard eQTL, in that they did not map expression profiles directly in the linkage pedigrees but instead performed expression profiling on separate samples and then looked for differentially expressed genes that mapped within linkage peaks and followed those candidates. They found associations with SNPs both in *GSTO 1* ( $p = 0.007$  [MIM 605482]) in and *GSTO 2* ( $p = 0.005$ ), both of which are tightly linked on chromosome 10, but only *GSTO 1* was differentially expressed. We did not see a significant eQTL result with *GSTO 1*, and our analysis examining only expression failed to detect differential expression at this locus. We did see a *GSTO 2* effect that had a significant interaction with diagnosis. Running the *GSTO 2* Illumina Chip hybridization probe through Blat to determine whether there was cross-talk between *GSTO 1* and *GSTO 2* with respect to our hybridization signal yielded a single hit in the entire genome, mapping specifically to the 3' UTR of *GSTO 2*. Given that the Li et al. study expression profiling was performed in a separate cohort that was just 3% of the size of this one ( $n = 6$  LOAD, 2 control brains) and that we performed our expression and genotyping using the same cohort, we feel that it is likely that the true effect at this locus on chromosome 10q is with *GSTO 2*. This speaks to the power of examining both

expression and genotypes in the same cohort because correlations will detect linear relationships with gene dosage, a method more powerful and accurate than simply looking at gene expression alone. Additionally, *trans* effects will be captured only by examining the relationship between SNPs and expression together and will be missed if SNPs and transcripts are assessed as separate data sets.

Finally, in both our control screen and our analysis of the entire cohort, we replicated a result from one of the first published eQTL screens using EBV-transformed lymphoblasts from the Centre d'Etude du Polymorphisme Humaine (CEPH) cohort.<sup>34</sup> This study, using different platforms for whole-genome and transcriptome analysis than those used in our analysis, found one transcript that had a similar *cis* association. This transcript was ribosomal protein S26 (*RPS26* [MIM 603701]), which encodes a ribosomal protein that is a component of the 40S subunit. For *RPS26*, Cheung and colleagues reported significant association with marker rs2271194, which is in complete LD with two out of our six associated SNPs. This replicated result suggested that this SNP-transcript pair is unlikely to be a false positive. Additionally, two more recent screens have also found an effect with *RPS26*, further validating the result.<sup>2,6</sup> One of those studies, using human liver samples, has implicated *RPS26* in diabetes pathways,<sup>6</sup> and we have found it to have a significant interaction with diagnosis in our screen. These observations suggest that the function of ribosomal proteins might not be limited to stability and efficiency of the ribosome. Ribosomal proteins may participate in the regulation of specific transcripts or families of transcripts being processed by the

**Table 3. Summary Statistics for the eQTL and DE Network**

Cluster	Total	eQTL	LOAD	Top Process	Score	Count	p Value
1	8%	16%	7%	*organ morphogenesis	1.72	8	$2.900 \times 10^{-3}$
2	4%	42%	34%	zinc finger, C2H2 type	0.99	8	$4.100 \times 10^{-3}$
3	6%	40%	12%	cell growth	1.12	5	$2.400 \times 10^{-2}$
4	20%	33%	16%	*negative regulation of transcription, DNA dependent	2.05	13	$2.400 \times 10^{-3}$
5	6%	44%	19%	protein kinase cascade	2.06	9	$2.200 \times 10^{-3}$
6	9%	12%	9%	KRAB box	2.11	9	$5.500 \times 10^{-5}$
7	19%	21%	10%	ligase activity, forming carbon-nitrogen bonds	1.78	12	$6.400 \times 10^{-3}$
8	14%	35%	15%	mitochondrion	3.02	33	$8.200 \times 10^{-6}$
9	14%	19%	10%	glycoprotein	3.42	49	$3.900 \times 10^{-6}$

Shown are the proportion of transcripts, out of the total number of transcripts networked, that map to each cluster (Total), that had a significant (corrected p value < 0.05 after 1000 permutations) correlation with allele dose (eQTL), and that had significant interaction term in the 2 df model (LOAD). Top processes were defined as ontology groups that had the lowest p value within the annotation cluster with the highest enrichment score. The exception to this definition is indicated by an asterisk; for clusters 1 and 4, all of the ontology groups within the top enrichment-score bin were the same for these two clusters (see Table S2B for ontology groupings in this set). Therefore, for these clusters, top processes are listed within the next most significant grouping of ontologies. Enrichment scores (Score), numbers of transcripts within the top annotation cluster (Count), and p values for the EASE scores for each process (p Value) are given. Table S2A lists all top ontologies for each cluster.

ribosome, and many of those might be integral to disease processes.<sup>35</sup>

### Candidates—*trans* eQTL

Although it is more difficult to capture validated *trans* effects because of the fact that there are no distance limitations placed upon their discovery, some *trans* effects are likely to be real and of great interest for disease. Conceptualization of how *trans* effects could occur is less obvious than that of *cis* effects, but it is important to remember that our screen captures only the expression profiles of poly-A tailed mRNA. Thus, *trans* effects could be occurring because the *trans* SNP is affecting an intermediary (such as miRNA) that we have not captured within our screen. Another possible scenario is that *trans* effects occur through interchromosomal relationships in transcription of mRNA. Nunez and colleagues used chromatin-conformation capture to demonstrate that interchromatin pairing of disparate DNA sequences (in their example genes on chromosomes 21 and 2) is a common feature of estrogen receptor- $\alpha$ -mediated transcription.<sup>36</sup> Thus, there could be direct action of *trans* SNPs on target transcripts located on different chromosomes.

Our network analysis is informative for assessing the *trans* effects in our screen. In our preliminary analysis we focused on cluster 2, because this cluster contained the

highest proportion of transcripts that had LOAD eQTL relative to the other clusters. Within this network, the main effects were with the transcript C6orf29 (171 links total, 58 links within cluster 2, accounting for 90% of the cluster [OMIM 606107]) and the SNP rs9309291 (significant effect with 63 transcripts; all effects had a significant interaction with diagnosis, and ten of those transcript mapped to cluster 2). C6orf29 is a member of a family of three transporters, one of which was cloned because of its ability to suppress a yeast choline transport mutation.<sup>37</sup> Although currently the best LOAD biomarker is beta-amyloid (for review see study by J. Hardy<sup>38</sup>), it has been proposed that cholinergic systems are involved in Alzheimer disease.<sup>39</sup> It is known that cholinergic neurons are particularly susceptible to Alzheimer disease pathogenesis, further implicating this network in the pathogenesis of LOAD. Thus, C6orf29 is a putative *trans* target in our screen.

### Discussion

We have performed a genome- and transcriptome-wide screen to assess the downstream effects of risk variation for LOAD on mRNA expression. The importance of work and data is clear.

First, by focusing on a quantitative endo-phenotype (expression) as opposed to a discrete phenotype (diagnosis), we have considerable power to detect subtle regulatory effects in a relatively small cohort. Statistical calculations have predicted that sample sizes on the order of 100 are sufficient for 80% power for eQTL studies.<sup>40</sup> This is an order of magnitude less than that predicted for genome-wide association studies (GWAS).<sup>41</sup> We can clearly see this by comparing this current analysis to our previous GWAS.<sup>11</sup> The cohort in this report was a part of our larger GWAS sample and is approximately one-quarter the sample size of that report ( $n = 1411$  in Reiman et al.,<sup>11</sup> 364 samples in this study). In our GWAS screen with an unstratified analysis, our only result after correction for multiple testing was with APOE.<sup>12</sup> With rank ordering of our uncorrected GWAS p values, it is of note that 6% of our permutation-corrected eQTL SNPs fall within the top 5% of our uncorrected GWAS effects. Given that these effects were statistically significant in the smaller eQTL cohort but not in the larger GWAS screen, eQTL quantitative analysis might be more powerful for detection of expression effects than discrete trait analysis. Of course, this does not obviate the need for large GWAS studies, because our approach would not capture coding changes or other types of risk variants not associated with changes in expression; however, it is clear that eQTL approaches in smaller cohorts can capture regulatory risk effects that would be found only in GWAS screens of considerably larger size.

Second, assaying the tissue relevant to the disease of interest is crucial. Expression studies profiling different

**Table 4. Table of Hub Transcripts in the eQTL and DE Network**

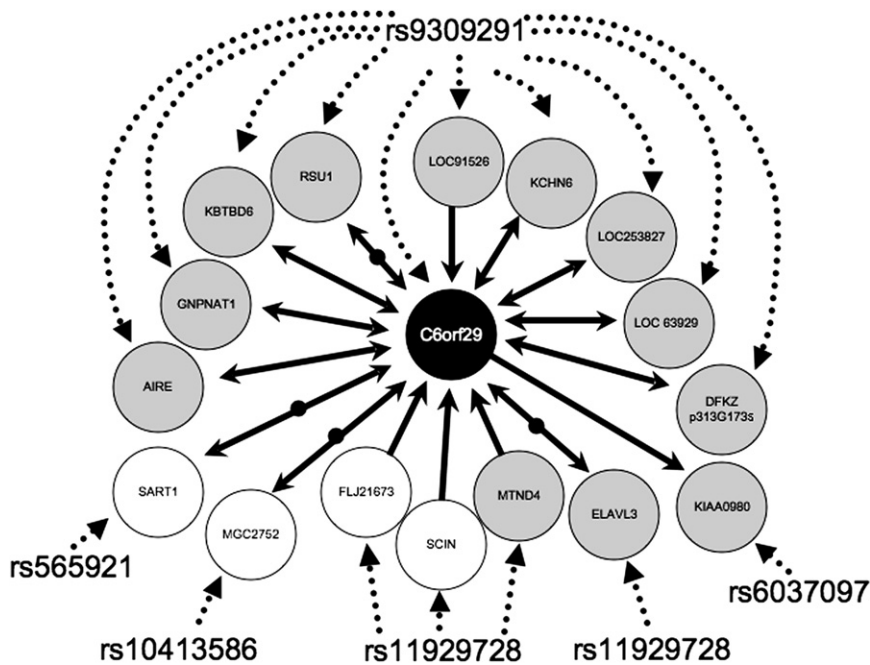
Cluster	Probe ID	Name	Entrez	No. Links
2	<b>GI_37620215-S</b>	<b>LOC253827</b>	<b>253827</b>	<b>134</b>
4	GI_21314672-S	CLST11240	51751	134
2	<b>GI_34577082-A</b>	<b>RSU1</b>	<b>6251</b>	<b>136</b>
4	GI_4505876-A	PLEC1	5339	136
7	GI_29740938-S	KIAA1136	57512	136
7	GI_4504482-S	HPRT1	3251	136
8	GI_4505718-S	PEX11B	8799	136
7	GI_4503064-S	CRYM	1428	137
2	GI_21389426-S	MGC29891	126626	138
4	GI_34577104-A	RNPC1	55544	138
4	GI_4758085-S	CSRP1	1465	138
7	<b>GI_27734858-S</b>	<b>LOC285533</b>	<b>285533</b>	<b>139</b>
7	<b>GI_37221176-S</b>	<b>NUDT11</b>	<b>55190</b>	<b>139</b>
3	GI_34222376-S	MGC44287	340547	142
4	GI_41152085-S	SERPINB6	5269	142
4	<b>GI_4508040-S</b>	<b>ZNF91</b>	<b>7644</b>	<b>142</b>
7	<b>GI_21264573-S</b>	<b>TM45F13</b>	<b>27075</b>	<b>143</b>
4	<b>GI_13899238-S</b>	<b>MGC10812</b>	<b>54858</b>	<b>144</b>
3	GI_37538660-S	IPLA2(GAMMA)	50640	146
4	GI_27754195-S	MTND3	4537	146
4	<b>GI_31982916-S</b>	<b>FLJ23042</b>	<b>5829</b>	<b>146</b>
4	GI_34147573-S	ITGB5	3693	146
4	GI_8922566-S	FLJ10647	55194	146
4	GI_31543199-S	C7orf21	83590	147
2	<b>GI_33589825-S</b>	<b>APG10L</b>	<b>819941</b>	<b>148</b>
3	GI_31543151-S	MGC14817	84298	148
4	<b>GI_38569399-S</b>	<b>ITPKB</b>	<b>3707</b>	<b>148</b>
4	GI_4755133-S	DOK1	1796	149
4	GI_6005923-S	TU3A	11170	149
1	GI_37550007-S	KIAA0121	9686	150
3	GI_41393596-S	CSAD	51380	150
1	<b>GI_38788287-S</b>	<b>ZNF160</b>	<b>90338</b>	<b>151</b>
4	GI_19923329-S	CTDSP2	10106	151
1	GI_28558992-I	TG737	8100	153
4	GI_24308106-S	DKFZp566C0424	26099	153
7	GI_21361148-S	RGS7	6000	155
4	GI_38176295-I	LASS1	10715	156
8	GI_34335244-A	NMNAT2	23057	156
4	GI_11968040-S	DCLRE1C	64421	158
4	<b>GI_30795118-A</b>	<b>FBX018</b>	<b>84893</b>	<b>159</b>
8	GI_19913427-S	ATP6V1B2	526	159
1	GI_8922963-S	SYNJ2BP	55333	163
4	GI_13376551-S	FLJ14346	80097	163
4	GI_24308252-S	DTX2	113878	163
4	GI_25453471-A	EEF1D	1936	164
4	<b>GI_4755141-S</b>	<b>INPPL1</b>	<b>3636</b>	<b>168</b>
2	<b>GI_14249467-I</b>	<b>C6orf29</b>	<b>80736</b>	<b>171</b>
1	GI_14249445-S	FLJ14431	84869	175
8	GI_23111022-S	SNX10	29887	177
8	GI_10863934-S	RTN1	6252	178
4	<b>GI_19923612-S</b>	<b>FLJ21128</b>	<b>80153</b>	<b>179</b>
3	GI_42655861-S	KIAA0492	57238	180
4	<b>GI_19913359-S</b>	<b>TBL1X</b>	<b>6907</b>	<b>180</b>
4	GI_5803097-S	MYST3	7994	182
8	<b>GI_31542527-S</b>	<b>DKFZP566B183</b>	<b>25977</b>	<b>186</b>
1	GI_36030972-S	UBXD4	165324	188
4	GI_38569463-S	SASH1	23328	195
8	GI_20357538-A	ATP6V1G2	534	202
8	GI_5454069-S	SIC9A6	10479	202
4	GI_28558997-A	PTBP1	5725	209
8	<b>GI_45238856-S</b>	<b>LOC114928</b>	<b>114928</b>	<b>214</b>
4	GI_27478393-S	LOC285919	285919	230

Listed are all of the transcripts with greater than 134 links (see Supplemental Data). Transcripts correlated with SNP genotype are highlighted in bold. Cluster number (cluster), probe number from the Illumina RefSeq-8 chip (Probe ID), transcript name from the Illumina RefSeq-8 chip (Name), Entrez gene identity number (Entrez), and the number of transcripts linked to the target (No. Links) are given. All eQTL transcripts highlighted in bold have a significant diagnosis interaction ( $\alpha = 0.05$ ), have SNPs mapping in *trans*, and are differentially expressed between cases and controls (SAM FDR < 0.1%), with the exception of MGC10812, which met other the criteria but was not differentially expressed.

human tissues have shown that the human brain transcriptome is unique.<sup>8,9</sup> Additionally, recent eQTL work examining phenotypes involved in body mass found that there was an order of magnitude difference between the effects that they found by transcriptome sampling in blood verses adipose tissue, many more significant traits being found in adipose, their tissue of interest.<sup>5</sup> We are focused on neurodegenerative disease; therefore, we have sampled brain tissue. A possible argument could be put forth that postmortem tissue does not have relevance for disease in living subjects; however, recent *in vivo* work profiling *MAPT* in human cerebral spinal fluid and examining the relationship between Tau levels and *MAPT* genotype has recapitulated our findings in postmortem samples,<sup>30,42,43</sup> suggesting that postmortem findings are relevant to *in vivo* disease processes. This last point is crucial, because these eQTL targets could be the next generation of biomarkers. eQTL biomarkers have the added benefit of not only being molecular but having mapped genotype profiles, which might be easier to assay than a molecular marker. Identifying biomarkers that could further classify preclinical subgroups and identify subclasses of rapid converters would help significantly reduce the cost of drug trials.

Third, by coupling SNP variation with changes in expression, we are able to map the neurobiological consequences of our genetic findings. We began this screen because we were frustrated with our own and others' efforts in mapping novel risk factors for late onset Alzheimer disease. As of 2006, 355 additional candidate genes for LOAD have been examined.<sup>44</sup> Of the subset of genes providing enough information for meta-analysis, only ~20% had variants that replicated. The majority of the replicated risk variants were noncoding; thus, the biological consequences of those changes are unknown. We felt that examining downstream functionality of putative risk variants would help in distinguishing between true risk variants and those that were merely markers of disease. Clearly, our efforts have had some success, in that we are able to distinguish the genetic effects of *GSTO 2*, the true risk gene, from *GSTO 1*, which most likely associated with risk only because of high LD.

Fourth, although not crucial for detection of all effects, it was important that our study sampled disease tissue. In this study, over half the LOAD-associated effects that we detected were not seen in our screen using control tissue.<sup>7</sup> This could be due to differences in power between the first



**Figure 3. Cluster 2 C6orf29 Network**  
 Diagrammed in the figure are the predicted transcript relationships given from our analysis using NEO<sup>25</sup> to impart directionality on the transcript-transcript correlations that we predicted from HQCUT<sup>22</sup> for the transcripts in cluster 2. For clarity, only transcripts that had correlations with SNP genotype and an interaction with C6orf29 are listed (25% of cluster 2). In the figure, SNP-transcript relationships are given by dashed arrows and transcript-transcript relationships by solid arrows. Transcripts are indicated by circles, gray circles indicating transcripts where there was a significant interaction with diagnosis in the eQTL analysis and white circles indicating no significant interaction term. Bidirectional arrows indicate a feedback loop between the two transcripts analyzed as predicted by NEO analysis. Bidirectional arrows divided by black circles indicate that NEO predicted an unmapped hidden confounder in the analysis of the relationship between the two transcripts.

screen and our second screen, which was approximately double the sample size; however, there is the possibility that there are effects specific to disease tissue. To examine these outcomes, we performed a bootstrap analysis to create from the control cohort data a data set that was equal to the total cohort sample size in our LOAD screen ( $n = 364$ ). Comparison of the permuted control data set to our entire cohort showed that there were few differences in the profiles of the two samples (see Table S3). We then tested whether within this larger control cohort we could detect our LOAD effects. We stratified the results into those that had a significant interaction term in our 2 df model (LOAD set  $n = 665$ ,  $\alpha = 0.05$ ) and those that were significant for allele dosage but did not have a diagnosis effect (non-LOAD set  $n = 1820$ ,  $\alpha = 0.05$ ). The bootstrap sample captured all of the non-LOAD set effects at an  $\alpha = 0.001$  and 95% at an  $\alpha = 1 \times 10^{-7}$ , which was the minimum p value that is able to survive the permutation correction. We were able to detect a much smaller portion of the effects with a significant interaction term, capturing only 66% of the LOAD set at an  $\alpha = 0.001$ . Results are shown in Figure 4 and Tables S3 and S4. This suggests that mapping disease-associated regulatory loci is more efficient and effective if the sampling cohort includes samples with known disease phenotypes; however, it is of note that we did capture LOAD effects in control tissue; thus, for preliminary screens, disease phenotypes might not be an issue.

Finally, the full public release of these data constitutes a resource to the community of researchers working on this prevalent and devastating disease. The data files and sample information used in generating the analysis for this paper are available at the website of A.J.M. Additional

genotype files are available at the Translational Genomics Research Institute website. Additional expression data and information have been deposited in NCBI's Gene Expression Omnibus (GEO GSE15222). DNA from the samples employed in this screen is available on request through the National Cell Repository for Alzheimer's Disease for fine mapping of particular effects.

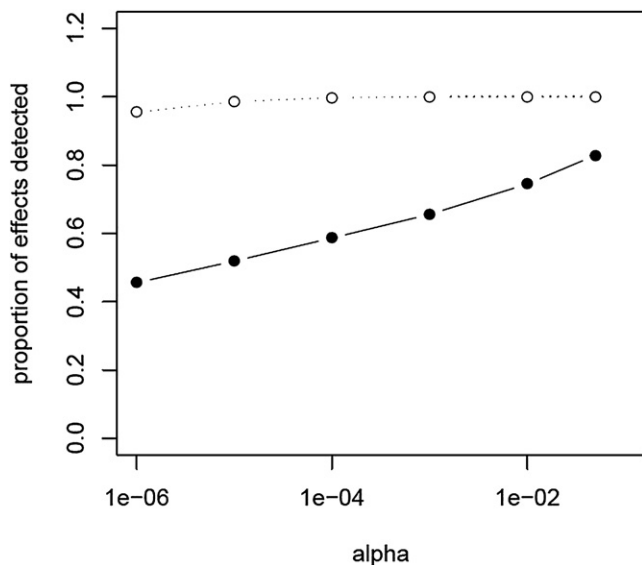
### Supplemental Data

Supplemental Data include Supplemental Acknowledgments, three figures, and four tables and can be found with this article online at <http://www.ajhg.org/>.

### Acknowledgments

We would like to dedicate this work to our colleague Christopher Heward, who died during this study. We thank the brain donors and their families. Many data and biomaterials were collected from several National Institute on Aging (NIA)- and National Alzheimer's Coordinating Center (NACC, grant U01 AG016976)-funded sites. Marcelle Morrison-Bogorad, Tony Phelps, Marilyn Miller and Walter Kukull are thanked for helping to coordinate the brain and data collection.

The directors, pathologists, and technicians involved include: Ruth Seemann, Juan C. Troncoso, Olga Pletnikova, Harry Vinters, Justine Pomakian, Christine Hulette, John F. Ervin, Dikran Horoupi, Ahmad Salehi, Jean Paul Vonsattel, Katerina Mancevska, E. Tessa Hedley-Whyte, Karlotta Fitch, Roger Albin, Andrew Lieberman, Lisa Bain, Eszter Gombosi, William Markesbery, Sonya Anderson, Dennis W. Dickson, Natalie Thomas, Carol A. Miller, Jenny Tang, Dimitri Diaz, Dan McKeel, John C. Morris, Eugene Johnson, Virginia Buckles, Deborah Carter, Thomas Montine, Aimee Schantz, John Q. Trojanowski, Virginia M. Lee, Viviana



**Figure 4. SNP-Transcripts Specific to Disease Are More Likely to Be Discovered in a Cohort with Disease-Specific Samples**

The proportion of effects that we could detect from our current study, utilizing brains from LOAD cases and nondemented controls, in a sample of equal size ( $n = 364$ ) generated by bootstrapping our control data is plotted. The alpha cutoff is plotted on the x axis and the proportion of effects detected is on the y axis. The proportion of eQTLs detected of those that showed a significant diagnosis interaction ( $\alpha = 0.05$  for the interaction term in our 2 df model) is shown by the filled circles and solid line. The proportion detected of those that did not have a significant diagnosis effect (interaction term p value  $\geq 0.05$ ) is shown by the open circles and dotted line. Counts and alphas used to generate this graph are given in Table S4.

Van Deerlin, Terry Schuck, Ann C. McKee, Carol Kubilus, Joseph Rogers, Thomas G. Beach, Lucia I. Sue, Bruce H. Wainer, Marla Gearing, Charles L. White, III, Roger Rosenberg, Marilyn Howell, Joan Reisch, William Ellis, Mary Ann Jarvis, David A. Bennett, Julie A. Schneider, Karen Skish, Wayne T Longman, Deborah C. Mash, Margaret J Basile, and Mitsuko Tanaka.

Funding sources and site affiliations listed in the [Supplemental Acknowledgements](#). None of the sponsors were involved in the design or conduct of the study, in the collection, analysis, and interpretation of the data, or in the preparation, review, or approval of the manuscript.

Received: October 3, 2008

Revised: March 2, 2009

Accepted: March 17, 2009

Published online: April 9, 2009

## Web Resources

The URLs for data presented herein are as follows:

A.J.M.'s personal website, <http://labs.med.miami.edu/myers/>

Alzgene database, <http://www.alzgene.com>

Blat, <http://genome.ucsc.edu/cgi-bin/hgBlat?command=start>

DAVID suite of programs, <http://david.abcc.ncifcrf.gov/home.jsp>

Entrez, <http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene>

Hapmap, <http://www.hapmap.org>

National Cell Repository for Alzheimer's Disease, <http://ncrad.iu.edu/>

NACC protocols, <http://www.alz.washington.edu>

NCBI's Gene Expression Omnibus, <http://www.ncbi.nlm.nih.gov/geo/>

Online Mendelian Inheritance in Man (OMIM), <http://www.ncbi.nlm.nih.gov/omim/>

PLINK analysis toolkit, <http://pngu.mgh.harvard.edu/~purcell/plink/index.shtml>

R suite of software, <http://www.r-project.org/>

Translational Genomics Research Institute website, <http://www.tgen.org/>

## References

- Schadt, E.E. (2005). Exploiting naturally occurring DNA variation and molecular profiling data to dissect disease and drug response traits. *Curr. Opin. Biotechnol.* 16, 647–654.
- Dixon, A.L., Liang, L., Moffatt, M.F., Chen, W., Heath, S., Wong, K.C., Taylor, J., Burnett, E., Gut, I., Farrall, M., et al. (2007). A genome-wide association study of global gene expression. *Nat. Genet.* 39, 1202–1207.
- Göring, H.H., Curran, J.E., Johnson, M.P., Dyer, T.D., Charleworth, J., Cole, S.A., Jowett, J.B., Abraham, L.J., Rainwater, D.L., Comuzzie, A.G., et al. (2007). Discovery of expression QTLs using large-scale transcriptional profiling in human lymphocytes. *Nat. Genet.* 39, 1208–1216.
- Moffatt, M.F., Kabisch, M., Liang, L., Dixon, A.L., Strachan, D., Heath, S., Depner, M., von Berg, A., Bufer, A., Rietschel, E., et al. (2007). Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature* 448, 470–473.
- Emilsson, V., Thorleifsson, G., Zhang, B., Leonardson, A.S., Zink, F., Zhu, J., Carlson, S., Helgason, A., Walters, G.B., Gunnarsdottir, S., et al. (2008). Genetics of gene expression and its effect on disease. *Nature* 452, 423–428.
- Schadt, E.E., Molony, C., Chudin, E., Hao, K., Yang, X., Lum, P.Y., Kasarskis, A., Zhang, B., Wang, S., Suver, C., et al. (2008). Mapping the genetic architecture of gene expression in human liver. *PLoS Biol.* 6, e107.
- Myers, A.J., Gibbs, J.R., Webster, J.A., Rohrer, K.C., Zhao, A.S., Marlowe, L., Kaleem, M., Leung, D., Bryden, L., Nath, P., et al. (2007). A Survey of Genetic Human Cortical Gene Expression. *Nat. Genet.* 39, 1494–1499.
- Ge, X., Yamamoto, S., Tsutsumi, S., Midorikawa, Y., Ihara, S., Wang, S.M., and Aburatani, H. (2005). Interpreting expression profiles of cancers by genome-wide survey of breadth of expression in normal tissues. *Genomics* 86, 127–141.
- Shyamsundar, R., Kim, Y.H., Higgins, J.P., Montgomery, K., Jorden, M., Sethuraman, A., van de Rijn, M., Botstein, D., Brown, P.O., and Pollack, J.R.A. (2005). DNA microarray survey of gene expression in normal human tissues. *Genome Biol.* 6, R22.
- Braak, H., and Braak, E. (1995). Staging of Alzheimer's disease-related neurofibrillary changes. *Neurobiol. Aging* 16, 271–278.
- Reiman, E.M., Webster, J.A., Myers, A.J., Hardy, J., Dunckley, T., Zismann, V.L., Joshipura, K.D., Pearson, J.V., Hu-Lince, D., Huentelman, M.J., et al. (2007). GAB2 alleles modify Alzheimer's risk in APOE epsilon4 carriers. *Neuron* 54, 713–720.
- Coon, K.D., Myers, A.J., Craig, D.W., Webster, J.A., Pearson, J.V., Lince, D.H., Zismann, V.L., Beach, T.G., Leung, D.,

- Bryden, L., et al. (2007). A High Density Whole-Genome Association Study reveals that APOE is the major susceptibility gene for sporadic late-onset Alzheimer's disease. *J. Clin. Psychiatry* 68, 613–618.
13. Hua, J., Craig, D.W., Brun, M., Webster, J., Zismann, V., Tembe, W., Josphipura, K., Huentelman, M.J., Dougherty, E.R., and Stephan, D.A. (2007). SNIPer-HD: Improved genotype calling accuracy by an expectation-maximization algorithm for high-density SNP arrays. *Bioinformatics* 23, 57–63.
  14. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., et al. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575.
  15. Workman, C., Jensen, L.J., Jarmer, H., Berka, R., Gautier, L., Nielser, H.B., Saxild, H.H., Nielsen, C., Brunak, S., and Knudsen, S. (2002). A new non-linear normalization method for reducing variability in DNA microarray experiments. *Genome Biol.* 3, research0048.
  16. Schadt, E.E., Li, C., Ellis, B., and Wong, W.H. (2001). Feature extraction and normalization algorithms for high-density oligonucleotide gene expression array data. *J. Cell. Biochem. Suppl.* 37, 120–125.
  17. Tseng, G.C., Oh, M.K., Rohlin, L., Liao, J.C., and Wong, W.H. (2001). Issues in cDNA microarray analysis: Quality filtering, channel normalization, models of variations and assessment of gene effects. *Nucleic Acids Res.* 29, 2549–2557.
  18. Pritchard, J.K., Stephens, M., Rosenberg, N.A., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959.
  19. Falush, D., Stephens, M., and Pritchard, J.K. (2003). Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* 164, 1567–1587.
  20. Fisher, R.A. (1922). On the interpretation of  $\chi^2$  from contingency tables, and the calculation of P. *J. R. Stat. Soc. [Ser A]* 85, 87–94.
  21. Tusher, V.G., Tibshirani, R., and Chu, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. USA* 98, 5116–5121.
  22. Ruan, J., and Zhang, W. (2008). Identifying network communities with a high resolution. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 77, 016104.
  23. Dennis, G., Jr., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C., and Lempicki, R.A. (2003). DAVID: Database for annotation, visualization, and integrated discovery. *Genome Biol.* 4, 3.
  24. Hosack, D.A., Dennis, G., Jr., Sherman, B.T., Lane, H.C., and Lempicki, R.A. (2003). Identifying biological themes within lists of genes with EASE. *Genome Biol.* 4, 4.
  25. Aten, J.E., Fuller, T.F., Lusic, A.J., and Horvath, S. (2008). Using genetic markers to orient the edges in quantitative trait networks: the NEO software. *BMC Syst Biol.* 2, 34.
  26. Muller, P.Y., Janovjak, H., Miserez, A.R., and Dobbie, Z. (2002). Processing of gene expression data generated by quantitative real-time RT-PCR. *Biotechniques* 32, 1372–1379.
  27. Ludbrook, J. (1991). On making multiple comparisons in clinical and experimental pharmacology and physiology. *Clin. Exp. Pharmacol. Physiol.* 18, 379–392.
  28. Miller, J.A., Oldham, M.C., and Geschwind, D.H. (2008). A systems level analysis of transcriptional changes in Alzheimer's disease and normal aging. *J. Neurosci.* 28, 1410–1420.
  29. Zotenko, E., Mestre, J., O'Leary, D.P., and Przytycka, T.M. (2008). Why do hubs in the yeast protein interaction network tend to be essential: reexamining the connection between the network topology and essentiality. *PLoS Comput Biol.* 4, e1000140.
  30. Myers, A.J., Pittman, A.M., Zhao, A.S., Rohrer, K., Kaleem, M., Marlowe, L., Lees, A., Leung, D., McKeith, I.G., Perry, R.H., et al. (2007). The MAPT H1c risk haplotype is associated with increased expression of tau and especially of 4 repeat containing transcripts. *Neurobiol. Dis.* 25, 561–570.
  31. Lambert, J.C., Goumidi, L., Vrièze, F.W., Frigard, B., Harris, J.M., Cummings, A., Coates, J., Pasquier, F., Cotel, D., Gaillac, M., et al. (2000). The transcriptional factor LBP-1c/CP2/LSF gene on chromosome 12 is a genetic determinant of Alzheimer's disease. *Hum. Mol. Genet.* 9, 2275–2280.
  32. Hamshere, M.L., Holmans, P.A., Avramopoulos, D., Bassett, S.S., Blacker, D., Bertram, L., Wiener, H., Rochberg, N., Tanzi, R.E., Myers, A., et al. (2007). Genome-wide linkage analysis of 723 affected relative pairs with late-onset Alzheimer's disease. *Hum. Mol. Genet.* 16, 2703–2712.
  33. Li, Y.J., Oliveira, S.A., Xu, P., Martin, E.R., Stenger, J.E., Scherzer, C.R., Hauser, M.A., Scott, W.K., Small, G.W., Nance, M.A., et al. (2003). Glutathione S-transferase omega-1 modifies age-at-onset of Alzheimer disease and Parkinson disease. *Hum. Mol. Genet.* 12, 3259–3267.
  34. Cheung, V.G., Spielman, R.S., Ewens, K.G., Weber, T.M., Morley, M., and Burdick, J.T. (2005). Mapping determinants of human gene expression by regional and genome-wide association. *Nature* 437, 1365–1369.
  35. Wool, I.G. (1996). Extraribosomal functions of ribosomal proteins. *Trends Biochem. Sci.* 21, 164–165.
  36. Nunez, E., Kwon, Y.S., Hutt, K.R., Hu, Q., Cardamone, M.D., Ohgi, K.A., Garcia-Bassets, I., Rose, D.W., Glass, C.K., Rosenfeld, M.G., et al. (2008). Nuclear receptor-enhanced transcription requires motor- and LSD1-dependent gene networking in interchromatin granules. *Cell* 132, 996–1010.
  37. O'Regan, S., Traiffort, E., Ruat, M., Cha, N., Compaore, D., and Meunier, F.M. (2000). An electric lobe suppressor for a yeast choline transport mutation belongs to a new family of transporter-like proteins. *Proc. Natl. Acad. Sci. USA* 97, 1835–1840.
  38. Hardy, J. (2006). A hundred years of Alzheimer's disease research. *Neuron* 52, 3–13.
  39. Bartus, R.T., Dean, R.L. 3rd, Beer, B., and Lippa, A.S. (1982). The cholinergic hypothesis of geriatric memory dysfunction. *Science* 217, 408–414.
  40. Schliekelman, P. (2008). Statistical power of expression quantitative trait loci for mapping of complex trait loci in natural populations. *Genetics* 178, 2201–2216.
  41. Wray, N.R., Goddard, M.E., and Visscher, P.M. (2007). Prediction of individual genetic risk to disease from genome-wide association studies. *Genome Res.* 17, 1520–1528.
  42. Laws, S.M., Friedrich, P., Diehl-Schmid, J., Müller, J., Eisele, T., Bäuml, J., Förstl, H., Kurz, A., and Riemenschneider, M. (2007). Fine mapping of the MAPT locus using quantitative trait analysis identifies possible causal variants in Alzheimer's disease. *Mol. Psychiatry* 12, 510–517.
  43. Kauwe, J.S., Cruchaga, C., Mayo, K., Fenoglio, C., Bertelsen, S., Nowotny, P., Galimberti, D., Scarpini, E., Morris, J.C., Fagan, A.M., et al. (2008). Variation in MAPT is associated with cerebrospinal fluid tau levels in the presence of amyloid-beta deposition. *Proc. Natl. Acad. Sci. USA* 105, 8050–8054.
  44. Bertram, L., McQueen, M.B., Mullin, K., Blacker, D., and Tanzi, R.E. (2007). Systematic meta-analyses of Alzheimer disease genetic association studies: The AlzGene database. *Nat. Genet.* 39, 17–23.