

# Patterns of Gene Expression in the Frontal Cortex Discriminate Alcoholic from Nonalcoholic Individuals

Jianwen Liu<sup>1</sup>, Joanne M Lewohl<sup>2</sup>, R Adron Harris<sup>1</sup>, Vishwanath R Iyer<sup>3</sup>, Peter R Dodd<sup>4</sup>, Patrick K Randall<sup>5</sup> and R Dayne Mayfield<sup>\*1</sup>

<sup>1</sup>Waggoner Center for Alcohol and Addiction Research, University of Texas at Austin, Austin, TX, USA; <sup>2</sup>Genomics Research Centre, School of Health Sciences, Griffith University, Qld, Australia; <sup>3</sup>Section of Molecular Genetics and Microbiology, Institute for Cellular and Molecular Biology, University of Texas at Austin, Austin, TX, USA; <sup>4</sup>Department of Biochemistry, University of Queensland, Qld, Australia; <sup>5</sup>Charleston Alcohol Research Center, Medical University of South Carolina, Charleston, SC, USA

Alcohol dependence is characterized by tolerance, physical dependence, and craving. The neuroadaptations underlying these effects of chronic alcohol abuse are likely due to altered gene expression. Previous gene expression studies using human post-mortem brain demonstrated that several gene families were altered by alcohol abuse. However, most of these changes in gene expression were small. It is not clear if gene expression profiles have sufficient power to discriminate control from alcoholic individuals and how consistent gene expression changes are when a relatively large sample size is examined. In the present study, microarray analysis (~47 000 elements) was performed on the superior frontal cortex of 27 individual human cases (14 well characterized alcoholics and 13 matched controls). A partial least squares statistical procedure was applied to identify genes with altered expression levels in alcoholics. We found that genes involved in myelination, ubiquitination, apoptosis, cell adhesion, neurogenesis, and neural disease showed altered expression levels. Importantly, genes involved in neurodegenerative diseases such as Alzheimer's disease were significantly altered suggesting a link between alcoholism and other neurodegenerative conditions. A total of 27 genes identified in this study were previously shown to be changed by alcohol abuse in previous studies of human post-mortem brain. These results revealed a consistent re-programming of gene expression in alcohol abusers that reliably discriminates alcoholic from non-alcoholic individuals.

*Neuropsychopharmacology* (2006) **31**, 1574–1582. doi:10.1038/sj.npp.1300947; published online 2 November 2005

**Keywords:** alcoholism; microarray; human brain; cell adhesion; neural disease

## INTRODUCTION

Chronic alcohol abuse produces persistent changes in brain function which are manifested as tolerance, physical dependence, craving, and other behavioral changes. These changes in brain function likely emanate from alterations in gene expression, which in turn may underlie the cellular adaptations to chronic alcohol abuse (Nestler, 2000). The importance of understanding changes in gene expression in alcoholism can be appreciated by the impact of gene expression profiling in other diseases, most notably cancer, where studies have led to improved pharmacotherapies (Okutsu *et al*, 2002; Taxman *et al*, 2003; Zembutsu *et al*, 2002) and to a molecular classification of disease which

promises to be more accurate and informative than traditional diagnostic tests (Bueno *et al*, 2004; Kim *et al*, 2002; Mor *et al*, 2003). Gene expression profiling is only beginning to be applied to psychiatric illnesses (Geschwind, 2003; Marcotte *et al*, 2003; Mirnics *et al*, 2000; Tkachev *et al*, 2003). Initial studies indicate that changes in brain gene expression in schizophrenia or alcoholism are much smaller than are seen in cancer (Bull *et al*, 2001; Lewohl *et al*, 2001; Mayfield *et al*, 2002; Mirnics *et al*, 2000; Smith, 2002) and it is possible that individual differences due to other factors will overpower the differences due to the disease. A relatively large sample size is critical to identify genes with consistent expression changes at a higher confidence level.

Previously, we and others applied cDNA and oligonucleotide microarray techniques to identify genes with altered expression following long-term alcohol consumption (Flatscher-Bader *et al*, 2005; Lewohl *et al*, 2000; Liu *et al*, 2004; Mayfield *et al*, 2002; Sokolov *et al*, 2003). Genes involved in myelination, trafficking, ubiquitination, and mitochondrion function showed consistent changes at the transcriptional level. However, these studies used either pooled RNA (Lewohl *et al*, 2000; Mayfield *et al*, 2002) or

\*Correspondence: Dr RD Mayfield, Waggoner Center for Alcohol and Addiction Research, University of Texas at Austin, 2500 Speedway, MBB 1.124, Austin, TX 78712, USA, Tel: +1 512 232 2512, Fax: +1 512 232 2525, E-mail: dayne.mayfield@mail.utexas.edu

Received 8 March 2005; revised 10 August 2005; accepted 16 September 2005

Online publication: 21 September 2005 at <http://www.acnp.org/citations/Npp092105050163/default.pdf>

relatively small sample sizes (seven cases/group) (Liu *et al*, 2004). Individual variability could not be assessed at all or with certain confidence. To address these concerns, we combined cDNA microarrays comprised of ~47 000 elements with a partial least squares (PLS) statistical procedure to identify consistent changes in alcoholics (14 cases) compared with controls (13 cases). The current study focused on the superior frontal cortex. The frontal cortex has been an interesting area of research due to the most severe brain damage in this region of alcoholic human brain. This region is closely related to the reward system involved in development of alcohol dependence. This region is also crucial for cognitive, executive and other important functions that are impaired in alcoholics. Our previous study compared frontal cortex and motor cortex (Liu *et al*, 2004) and found that gene expression changes were more consistent across individuals in alcoholic frontal cortex than motor cortex. The goal of the present study was to identify genes or families of genes that show altered expression level in the alcoholic human frontal cortex, and to evaluate the potential influence of individual variation on gene expression patterns.

## MATERIALS AND METHODS

### Case Selection

Brain samples (superior frontal cortex) were obtained from the Brisbane Node of the National Health and Medical Research Council (NHMRC) Brain Bank and the Tissue Resource Centre at the University of Sydney, Australia (<http://www.braindonors.org/>). A total of 27 cases including 14 alcoholic and 13 control cases (matched as closely as possible by age and post-mortem interval (PMI)) were selected (Table 1; see Supplementary Table 4 for detailed clinical information). There were no significant differences between the control and alcoholic groups for either age or PMI (Student's *t*-test). The 27 cases represent two cohorts of donors (C = control; A = alcoholic). Cases C1–C7 and A1–A7 were included in a previous study (Liu *et al*, 2004), whereas C8–C13 and A8–A14 were additional cases obtained from the Australian brain banks. Neither age nor PMI differed significantly between the two sets of controls or alcoholics (Student's *t*-test).

Alcoholics were classified based on the quantity of alcohol consumed, according to the NHMRC (> 80 g of alcohol per day), instead of the criteria established by the American Psychiatric Association (DSM-IV) or the World Health Organization (ICD-10). Many alcoholic patients in this study consumed significantly more than 80 g/day for most of their adult life (Supplemental Table 4). Cerebral atrophy was observed in three alcoholic cases. All alcoholic cases included in our study were actively drinking until death except case A7 who was abstinent for 2 years prior to death. Control cases either abstained from alcohol completely or were social drinkers. Alcoholics with medical complications such as liver cirrhosis, Wernicke Korsakoff Syndrome, or other concomitant diseases were not included in the study. Brain samples were acquired from qualified pathologists under full ethical clearance #97/36, and informed written consent was obtained from the next of kin.

### RNA Extraction and Amplification

Total RNA was extracted from the superior frontal cortex of each case using a modified guanidine isothiocyanate extraction method (Chomczynski and Sacchi, 1987). RNA quality was assessed using an Agilent 2100 Bioanalyzer (Agilent, Palo Alto, CA). In all, 2 µg of extracted total RNA was used for the following amplification procedure. Universal Human Reference RNA (comprised of 10 different human cell lines) was purchased from Stratagene (La Jolla, CA). Both brain and reference RNA was amplified as described (Mayfield *et al*, 2003). Briefly, total RNA (1 µg) was denatured together with T7-oligo dT primer (0.2 µg) at 70°C for 5 min and then chilled on ice. Reverse transcription (20 µl reaction) was catalyzed by SuperScript II (Invitrogen, Carlsbad, CA) at 42°C for 2 h. Reagent mix (80 µl) for the second cDNA strand synthesis was added and incubated at 16°C for 2 h. Double-stranded cDNA was purified and dissolved in 8 µl of DEPC-H<sub>2</sub>O. *In vitro* transcription was performed in a 20 µl reaction at 37°C for 6 h using T7 MegaScript kit (Ambion, Austin, TX). Amplified antisense RNA was purified using the RNeasy Mini kit (QIAGEN, Valencia, CA). The quality and concentration of amplified antisense RNA was determined using the Agilent 2100 Bioanalyzer. Amplified RNA (2 µg) was utilized for each hybridization.

**Table 1** Case information. PMI, Post-Mortem Interval in Hours; 'C' in Case IDs for Control Cases and 'A' for Alcoholic Cases

Case ID	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	
Age (years)	85	55	54	56	58	62	78	70	32	79	61	59	67	
Sex	M	M	M	M	M	M	F	F	F	M	M	M	M	
PMI (h)	38	38	56	10	27	18	22	18	48.4	8.5	27	12	34	
Case ID	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14
Age (years)	34	29	70	36	58	79	70	44	59	43	57	67	49	42
Sex	M	M	M	M	M	M	F	M	M	M	M	M	M	M
PMI (h)	31	24	15	29	28.5	20	17	22	15	21.5	67	46	16	35

Neither PMI nor age is different between control and alcoholic group ( $p > 0.05$ , two-sample student *t*-test).

## Array Fabrication

The cDNA microarray slides were prepared at the microarray core facility at the University of Texas at Austin. A set of ~47 000 sequence-verified IMAGE cDNA clones (ResGen, Carlsbad, CA) were PCR-amplified and spotted onto poly-L-lysine-coated microscope glass slides. The arrays were post-processed using succinic anhydride as described except that the UV crosslink step was skipped (DeRisi *et al*, 1997).

## Array Hybridization and Normalization

Array hybridization and normalization was performed as described previously (Mayfield *et al*, 2003). Briefly, amplified RNA (2 µg) was denatured and reverse transcribed in the reaction containing amino allyl-modified dUTP. Then the RNA strand was degraded and subsequently cDNA was purified and concentrated. Amplified brain RNA samples were always labeled with cyanine 3 (Cy3; channel 1) while the amplified universal human reference RNA was labeled with cyanine 5 (Cy5; channel 2). Hybridization of labeled cDNA with array was carried out in the dark at 65°C for about 16 h. Slides were washed and then scanned with an Axon GenePix 4000 scanner (Axon, Union City, CA). Images were subsequently analyzed using Axon GenePix 4.0 software (Axon, Union City, CA).

Genepix 4.0 default feature detection parameters were used for spot detection. Pixel homogeneity within a spot was determined by calculating the Pearson correlation coefficient of pixel intensities from both channels in each spot on the array. The resulting regression correlation was set at 0.6 and above. Channel 2 (Cy5) was normalized for each array in the Longhorn Array Database (LAD) at the University of Texas at Austin (Killion *et al*, 2003). The normalization was based on the assumption that the log-transformed ratio of net channel 2 intensity to net channel 1 intensity for each spot was centered on zero. The normalized ratio of medians was the raw ratio of medians divided by the calculated normalization factor. Expression values on different batches of arrays were normalized to have the same median expression level across batches. Undetected spots on the array were flagged and excluded from the analysis. In most cases a spot/feature on the array was flagged when the spot intensity was below the detection limit of the scanner. Occasionally, spots had high local background or ran together with adjacent spots on the array. Such spots were also flagged.

## Data Reduction and Statistical Analysis

PLS analysis was employed to derive expression-based latent variables (linear combinations of gene expression ratios) that covary maximally with a categorical variable coding alcoholic *vs* nonalcoholic samples. The latent variable values derived from the expression data using the coefficients of the model were used to indicate group separation. The loadings of the individual gene expression ratios on the best discriminating latent variables were used as indicators of the importance (weight) of each gene in the relationship. For class prediction procedure, a linear discriminant analysis (LDA) procedure was carried out for

prediction analysis. In this procedure, all cases except one were used as a training data set to identify predictor genes. The number of predictors (genes) used in the LDA analysis can not exceed number of cases (26 cases in cross-validation analysis). Genes were rank-ordered by PLS loading scores and sequentially reduced to determine the number of genes used for this analysis. The top 15 genes provided maximum discrimination. The 'left-out' case was then classified based on the expression levels of the 15 selected genes. PLS and LDA analysis were performed using in JMP IN software (SAS Institute Inc., Cary, NC). For clustering analysis by principal component analysis (PCA), the first two principal components were plotted to indicate group separation. PCA analysis was done using STATISTICA software (StatSoft Inc., Tulsa, OK). The *k*-nearest neighbor method was used to compute missing values (using 10 genes as neighbors).

## Gene Selection

A classification system was developed in which differentially expressed genes were divided into two classes. Class I genes were qualitatively different, that is, they were predominantly detected in one group but not the other, while class II genes were consistently detected in both groups. Thus, class I genes represent those that were more likely 'turned off' or 'turned on' as a result of alcohol abuse, whereas class II genes represent consistently expressed genes for which quantitative differences in expression could be determined.

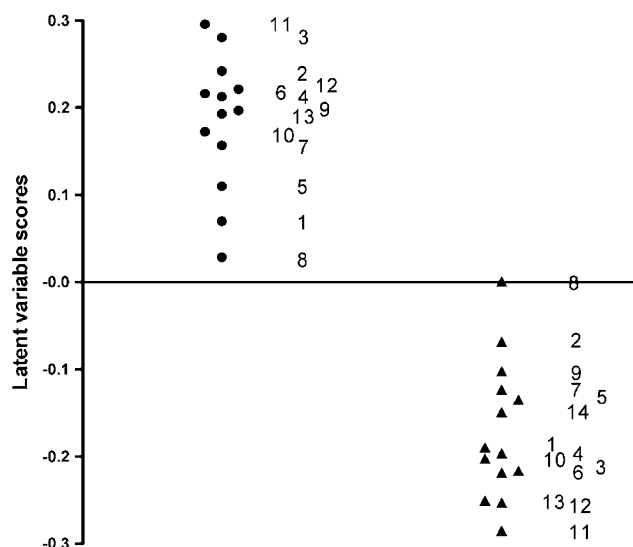
An empirically derived rule was established for defining Class I genes. Genes were selected that were consistently expressed in the controls but not the alcoholics, or were consistently expressed in the alcoholics but not the controls. 'Consistently' detected genes in a given group were defined as those detected on over 50% of the arrays of that group. Genes were then selected that were consistently detected in one group but not the other. In addition, only those genes that were detected over twice as frequently in one group compared to the other were considered Class I. Thus, Class I genes were those detected on at least eight more arrays in one group than the other. A higher threshold ( $n > 8$ ) would risk missing true positives, whereas a lower threshold ( $n < 8$ ) might include false positives. Supplementary Figure 5 shows the distribution of the number of genes as a function of the difference in the frequency of detection between the two groups (control minus alcoholic).

Multiple criteria were used to define Class II genes. First, only the genes detected on over 80% of the arrays in both groups (at least 11 alcoholics and 10 controls) (15 762 genes) were included. The rationale for using an 80% detection rate was based on studies demonstrating that accurate estimations of missing values typically requires at least 80% of present calls (Kim *et al*, 2005; Troyanskaya *et al*, 2001). Class II were then defined as those consistently detected in both groups having absolute PLS loadings scores of  $> 2.0$  and *t*-probabilities of  $< 0.05$ . In addition, genes that were  $< 20\%$  differentially expressed between the groups were excluded from further analysis. Genes that were not consistently detected across all cases could be due to pre-existing individual variability and/or potential variability in detection. This was particularly true for low abundance transcripts.

## RESULTS

DNA microarrays were used to compare gene expression profiles in the frontal cortex between control and well-characterized alcoholic cases (Table 1). We have devoted considerable effort to obtain high-quality RNA from human autopsy brain tissue and all samples were tested for quality and purity as described previously (Lewohl *et al*, 2001; Mayfield *et al*, 2002; Mayfield *et al*, 2003). As described in the Materials and Methods section, we utilized two data sets in the current study. Cases C1–C7 and cases A1–A7 were included in a previous report (Liu *et al*, 2004), whereas C8–C13 and A8–A14 represented additional cases in the current database. Prior to collapsing the two data sets we first determined their similarity. There were no significant differences in age or PMI between the two data sets (see ‘Case Selection’ in Materials and Methods section). We also compared the direction of expression changes (alcoholic compared to control) and found that approximately 70% of the detected genes common to both data sets (~20 000) changed in the same direction. This degree of overlap differed significantly from that expected by chance ( $p < 0.0001$ ,  $\chi^2$  test). In addition, we compared t-probability distributions between the two groups of controls and between the two groups of alcoholics (Supplementary Figure 6a and b, respectively). Both plots illustrate that the number of null comparisons increases as a function of  $p$ -value (an opposite trend would indicate a high number of differentially expressed genes) (Storey and Tibshirani, 2003), suggesting that there was a high level of similarity between groups of controls and alcoholics. Finally, we examined the differentially expressed genes in the 14 additional cases to determine if the major findings from Liu *et al* (2004) could be replicated. Using the same analysis procedures (Bayesian analysis) as reported previously (Liu *et al*, 2004), we identified a total of 2816 differentially expressed genes. Significant overlap was observed in the functional annotations of the differentially expressed genes between the two data sets (Supplementary Table 5). Also, the differentially expressed genes identified in the additional set of cases accurately distinguished between control and alcoholic cases as observed previously (Liu *et al*, 2004; Supplementary Figure 7). Based on these similarities we combined the two data sets for comprehensive analysis.

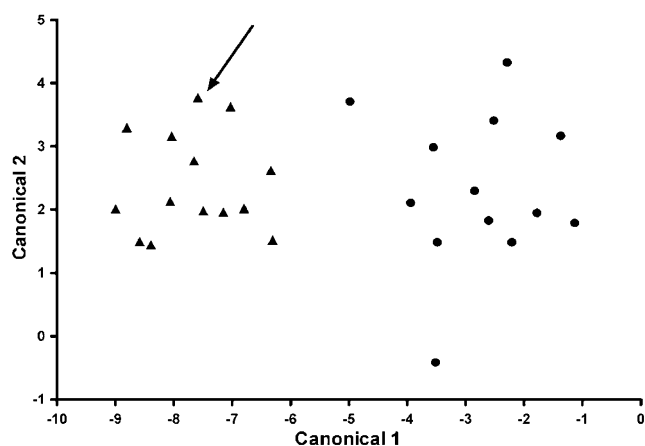
The average number of genes (cDNA spots) detected in the alcoholic group was significantly less than the control group (15%; see Supplementary Figure 8). A total of 15 762 genes were identified that were consistently expressed in at least 10 out of 13 controls and 11 out of 14 alcoholic cases. PLS analysis was used for data reduction and to identify the set of genes that best distinguish control from alcoholic cases. Figure 1 shows the discrimination between control and alcoholic cases based on the first latent variable scores in PLS analysis. The mean latent variable scores derived from this analysis differed significantly between control and alcoholic cases ( $p < 0.001$ , two-sample Student's  $t$ -test). Although the greatest differences in expression between control and alcoholic cases tend to be associated with high loading scores (absolute value), a large set of genes with relative small differences in expression have high loadings (Supplementary Figure 9).



**Figure 1** Discrimination of control and alcoholic cases. PLS analysis was performed on 15 762 consistently expressed genes. Latent variable scores discriminate control from alcoholic cases (approximately 85% of the variation in the groups (control vs alcoholic) is explained by the first latent variable). Filled circles indicate control cases and filled triangles indicate alcoholic cases. The numbers correspond to the case identifications in Table 1.

As in other psychiatric diseases, changes in gene expression observed in the brain of alcoholics are relatively small (20–50%) (Supplementary Figure 10). A leave-one-out procedure was performed to determine if gene expression profiles have sufficient power to discriminate control from alcoholic individuals. For this analysis, one of the 27 cases was left out as a test case and the remaining 26 cases were used as a training set. Based on the PLS loadings, the top 15 genes were selected and used to predict the group classification of the left-out case. The same process was used for each case. Among 27 predictions, only one case was misclassified. An example of one of the 27 predictions is shown in Figure 2. The unknown sample (A14 in this case) was accurately predicted as an alcoholic case.

To define cellular pathways altered in alcoholism, both Class I and Class II genes were identified (see Materials and Methods for classification criteria). A total of 46 genes were consistently expressed only in the control but not in alcoholic cases, while only three genes were consistently observed only in the alcoholic but not control cases (Class I). A close examination of those 49 genes excluded the possibility that the difference in the detection resulted from factors other than a difference in intensity. PLS analysis was not appropriate for this set of genes due to undetectable expression in alcoholics or controls. Among the 15 762 genes consistently expressed in both groups, 484 genes were differentially expressed (Class II). Genes in both classes were combined (531 genes). Among the genes with altered expression levels, over half of them were downregulated in the alcoholics (304/531 genes). This reflects a possible suppression of many cellular functions, such as cell adhesion and lipid metabolism shown in Table 2. The complete list of genes is shown in Supplementary Table 6. A total of 232 of the 531 differentially expressed genes are well documented in the literature. Given that only a small



**Figure 2** Example of class prediction using the top 15 genes identified by PLS analysis. All cases except one (A14 in this example) were used as the training set. The left-out case (A14) was then classified based on the top 15 genes with highest PLS loadings in training data set. The arrow indicates the correct prediction of the left-out case (A14) on the plot of first two canonicals. See Supplementary Table 8 for the list of 15 genes. Filled circles indicate controls cases and filled triangles indicate alcoholic cases.

**Table 2** Functional Groups of Differentially Expressed Genes

Functional groups	Up	Down	Total
Apoptosis	3	8 (1)	11
Carbohydrate metabolism	1	5 (2)	6
Cell adhesion	2	18 (1)	20
Cytoskeleton	5	7 (1)	12
Immune/stress response	11 (1)	13 (2)	24
Ion channels	2	6	8
Lipid metabolism	2	11 (1)	13
Mitochondrial proteins	5	8	13
Myelination-related genes	1	10	11
Neural disease-related genes	3	11	14
Neurogenesis/development	15	20	35
Protein turnover/modification	20	30 (1)	50
Synaptic transmission	6	12 (2)	18
Trafficking proteins	10 (1)	18	28
Transcription-related genes	17	26	43
Ubiquitin-mediated protein degradation	9	12	21
Total	79	153	232

Up, number of up-regulated genes; down, number of down-regulated genes. Number of Class I genes in each functional group is indicated in the parenthesis.

number of Class I genes are annotated, we decided to analyze the functional grouping only with the combined data set based on their functional annotations in the literature (Table 2). The full list of genes in each group is shown in Supplementary Table 7. Consistent with previous studies, many of these genes are associated with myelination, immune/stress response, ion channels, lipid metabolism, synaptic transmission, trafficking, and ubiquitination. In addition, genes involved in cell adhesion and apoptosis are also significantly downregulated. Interestingly, five

genes in the neural disease-related group are known to be involved in Alzheimer's disease. Apoptosis, cell adhesion, and ubiquitin-mediated protein degradation are three groups with highest average PLS loading scores (data not shown).

In order to verify our current results, the list of candidate alcohol-responsive genes identified in this study was compared with published studies. Only alcohol studies using autopsy human brain were included in this comparison (Flatscher-Bader *et al*, 2005; Lavoie and Butterworth, 1995; Lewohl *et al*, 2000; Mayfield *et al*, 2002; Sokolov *et al*, 2003). Two of these five published reports are from our laboratory; however, different cases and methodologies were used in those studies. The first study used both cDNA arrays (UniGEMV; Genome Systems Inc.) and oligonucleotide arrays (HuGeneFL; Affymetrix Inc.) (Lewohl *et al*, 2000). The second study used cDNA arrays (UniGEM V2.0; InCyte Genomics Inc.) (Mayfield *et al*, 2002). In addition, both studies used pooled RNA from cases that did not overlap with the cases included in this study. The accession number for each gene used in different studies was converted to the current UniGene Cluster ID (Build #180) to reflect the fact that a same gene (UniGene Cluster ID) may be represented by different clones (accession numbers) in different studies. Out of the 232 named genes, 27 genes have been identified in previous studies (Table 3). Among those 27 genes, 21 were found to be regulated in the same direction, and in general, to a similar extent. In particular, the downregulation of peripheral myelin protein 22 has been observed in both microarray and real-time PCR studies (Flatscher-Bader *et al*, 2005). The other six genes showed opposite direction of regulation.

To determine the contribution of individual variability to gene expression patterns and whether changes in only a subset of those genes could discriminate alcoholics from non-alcoholic controls, a PCA was performed using the Class II genes in each functional group. The first two principal components are plotted to illustrate the clustering of individual cases (Figure 3). Four representative functional groups are shown (cell adhesion, neural disease-related genes, synaptic transmission, and ubiquitin-mediated protein degradation). As expected, control and alcoholic cases are largely separated in all functional groups. Interestingly, three alcoholic cases are consistently different from other alcoholic cases. Available clinical information indicates several unique features of these three cases. Case A4 was also dependent on morphine in addition to alcohol while none of the other cases were poly-drug abusers. Case A7 had showed strong alcohol withdrawal symptoms and was abstinent for 2 years prior to death, whereas case A8 had been treated for depression. It is possible that these factors can contribute to the gene expression patterns.

Aging and PMI can induce gene expression changes in human brain (Lu *et al*, 2004; Schramm *et al*, 1999). To determine if the gene expression changes identified in this study were correlated with these factors, a linear regression model was fitted to gene expression level and age or PMI. After correction for multiple comparisons, none of the 482 regulated genes was significantly correlated with age or PMI at a false discovery rate of less than 30% ( $q$ -value < 0.30) (Storey and Tibshirani, 2003). In addition, a hierarchical

**Table 3** Transcripts from Current Study which have been Reported in Previous Studies of Human Alcohol Abuse

UniGene cluster ID (Build #183)	Annotation	% change		Reference for published value
		Published	Present	
Hs.518267	Transferrin	-150	-49	Lewohl <i>et al</i> (2000)
Hs.143261	Calpain 3, (p94)	-75	-36	Lewohl <i>et al</i> (2000)
Hs.517582	MCM5 minichromosome maintenance deficient 5, cell division cycle 46	-54	-30	Mayfield <i>et al</i> (2002)
Hs.144197	UDP glycosyltransferase 8 (UDP-galactose ceramide galactosyltransferase)	-50	-46	Lewohl <i>et al</i> (2000)
Hs.19383	Angiotensinogen (serine (or cysteine) proteinase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 8)	-50	Turned off	Lewohl <i>et al</i> (2000)
Hs.534322	Major histocompatibility complex, class II, DR beta 4	-45	-48	Lewohl <i>et al</i> (2000)
Hs.496684	Lysosomal-associated membrane protein 2	-42	-40	Sokolov <i>et al</i> (2003); Flatscher-Bader <i>et al</i> (2005)
Hs.482730	EGF-like repeats and discoidin I-like domains 3	-40	-33	Lewohl <i>et al</i> (2000)
Hs.124611	Sparc/osteonectin, cwcv and kazal-like domains proteoglycan (testican)	-37	-21	Flatscher-Bader <i>et al</i> (2005)
Hs.104839	Tissue inhibitor of metalloproteinase 2	-33	-24	Flatscher-Bader <i>et al</i> (2005)
Hs.471441	Proteasome (prosome, macropain) subunit, beta type, 2	-32	-20	Flatscher-Bader <i>et al</i> (2005)
Hs.372031	Peripheral myelin protein 22	-32	-44	Lewohl <i>et al</i> (2000); Flatscher-Bader <i>et al</i> (2005)
Hs.135705	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 5	-31	-27	Sokolov <i>et al</i> (2003)
Hs.7879	Interferon-related developmental regulator 1	-31	-32	Sokolov <i>et al</i> (2003)
Hs.529890	Calnexin	-17	-30	Sokolov <i>et al</i> (2003)
Hs.168762	Unc-51-like kinase 2 ( <i>C. elegans</i> )	-20	-27	Sokolov <i>et al</i> (2003)
Hs.167017	Gamma-aminobutyric acid (GABA) B receptor, I	55	29	Flatscher-Bader <i>et al</i> (2005)
Hs.1787	Proteolipid protein 1 (Pelizaeus-Merzbacher disease, spastic paraplegia 2, uncomplicated)	-100	-46	Lewohl <i>et al</i> (2000)
Hs.273621	2',3'-cyclic nucleotide 3' phosphodiesterase	-100	-20	Lewohl <i>et al</i> (2000)
Hs.162585	Solute carrier family 12 (sodium/potassium/chloride transporters), member 2	-40	-21	Mayfield <i>et al</i> (2002)
Hs.89643	Transketolase (Wernicke-Korsakoff syndrome)	-19	-36	Lavoie and Butterworth, (1995)
Hs.99886	Complement component 4 binding protein, beta	-52	25	Flatscher-Bader <i>et al</i> (2005)
Hs.406062	NADH dehydrogenase (ubiquinone) I alpha subcomplex, I1, 14.7kDa	-32	45	Flatscher-Bader <i>et al</i> (2005)
Hs.368641	Stearoyl-CoA desaturase (delta-9-desaturase)	27	-37	Mayfield <i>et al</i> (2002)
Hs.385913	Acidic (leucine-rich) nuclear phosphoprotein 32 family, member E	54	-23	Sokolov <i>et al</i> (2003)
Hs.435570	Cyclin-dependent kinase-like 5	90	-26	Sokolov <i>et al</i> (2003)
Hs.528803	Ubiquinol-cytochrome c reductase core protein II	45	-24	Lewohl <i>et al</i> (2000)

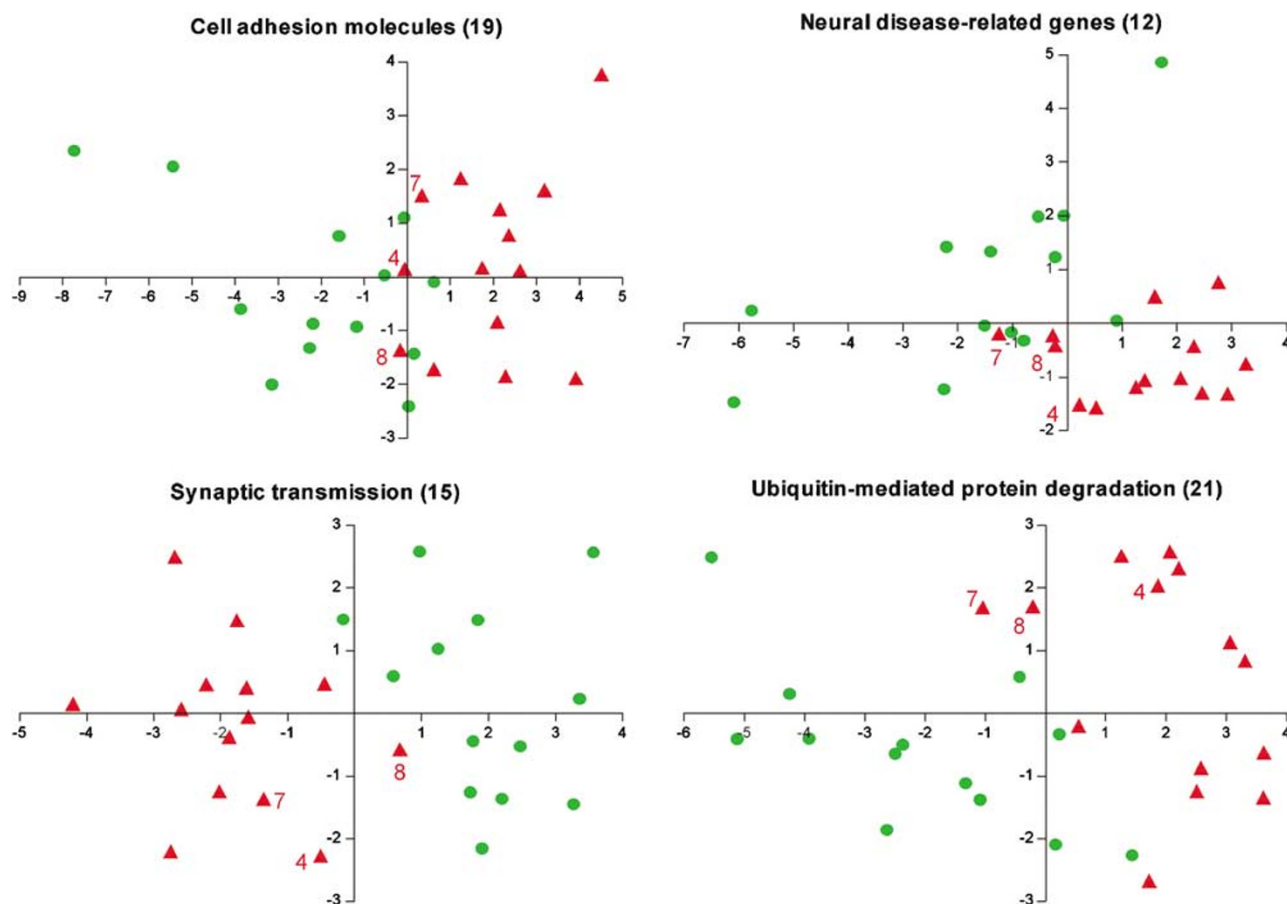
clustering of the 27 cases based on the 482 Class II genes revealed that cases formed clusters based only on the history of alcohol consumption, but not age or PMI (Figure 4). This is further supported by the fact that even cases within the same group (control or alcoholic group) did not form clusters based on age or PMI.

## DISCUSSION

The current data set provides the largest study to date of gene expression in human alcoholism and extends findings from earlier studies (Flatscher-Bader *et al*, 2005; Lewohl

*et al*, 2000; Liu *et al*, 2004; Mayfield *et al*, 2002; Sokolov *et al*, 2003). Importantly, this study revealed novel genes involved in cell adhesion, apoptosis, and other neural diseases that were altered by long-term alcohol abuse.

Array experiments characteristically have many measured variables (genes) and relatively few observations (experiments) and the statistical techniques utilized in this study are particularly well suited for these types of data (Datta, 2001; Datta, 2003). It should be noted that expression analysis using PLS techniques have been used to classify and discriminate among tumor types (Nguyen and Rocke, 2002b,c), and to predict clinical outcome and patient survival with a high degree of accuracy (Nguyen and Rocke,



**Figure 3** PCA analysis of gene expression levels for individual functional groups. The first two principal components are plotted. Case A4, A7, and A8 consistently differ from the other alcoholic cases in all functional groups. Green circles represent controls cases and red triangles represent alcoholic cases. Cases A4, A7, and A8 are identified by number. The number of genes included in the PCA analysis is shown in parentheses.

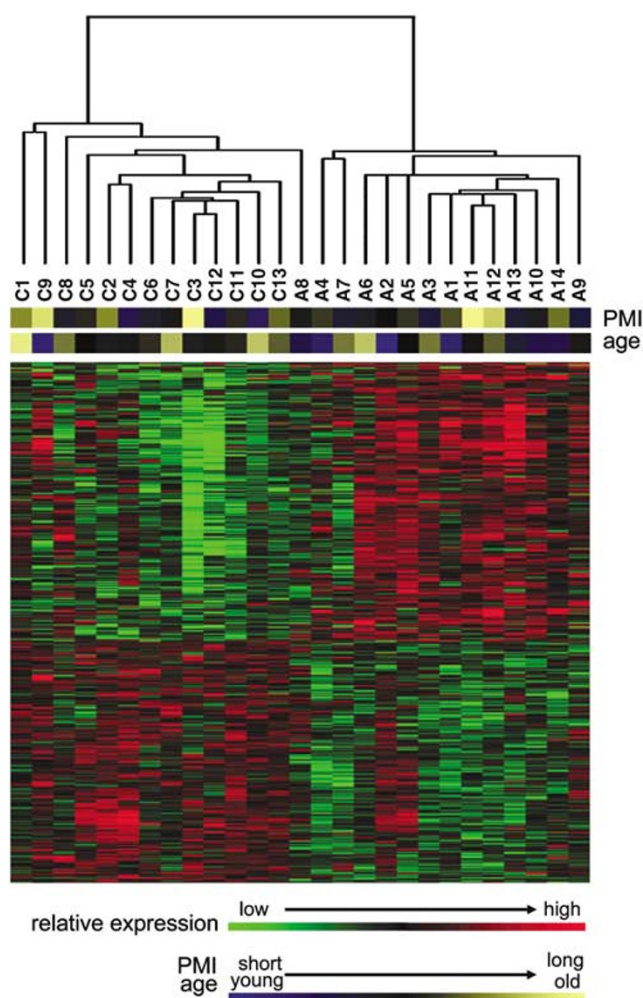
2002a; Perez-Enciso and Tenenhaus, 2003). However, cancer studies report relatively large changes in expression and this statistical technique has not been applied to data sets with modest expression changes such as those observed in alcoholism (typically 20-50%) (Lewohl *et al*, 2000; Liu *et al*, 2004; Mayfield *et al*, 2002). Our data suggest that chronic alcohol abuse affects multiple genes in multiple functional systems. Changes at the gene expression level reflect a possible pre-existing gene expression differences or an alteration as a result of alcohol consumption. An understanding of mechanisms of manifestations of alcoholism must encompass these small but abundant and widespread changes in cellular regulation.

Alcohol's effects on gene expression in the central nervous system have been widely studied in animals, using both array and nonarray techniques (reviewed by Worst and Vrana, 2005). In contrast, there are only limited studies with human subjects, largely due to the shortage of high-quality autopsy human brain available for alcohol research and other challenges in studies of autopsy human brain (Mirnics *et al*, 2004; Mirnics and Pevsner, 2004). To validate our array findings, we compared our results with those in the existing literature. Interestingly, most of the published results show a decrease in mRNA levels, consistent with our findings of downregulation of particular biological groups, for example, the myelination (four genes) and calpain/

calpastatin systems. The agreement between the current study and the literature is remarkably good considering the differences among the studies in methodologies, array platforms, individual cases, and other factors.

Myelin deficiency in the frontal cortex of alcoholic human brain was noted in a previous morphological study (Harper and Kril, 1989), and myelination-related genes have also been identified with altered expression levels in this and other studies (Lewohl *et al*, 2000; Liu *et al*, 2004; Mayfield *et al*, 2002; Sokolov *et al*, 2003). In the present study, among 11 myelination-related genes, 10 were downregulated and only one was upregulated. A major component of the myelin sheath, proteolipid protein 1, and several minor components (UDP glycosyltransferase 8, CNP, CD9 antigen, and claudin 11) were all downregulated. These results further support the conclusion that myelin structure is compromised by chronic alcohol abuse.

A striking finding of this study was that 18 of the 20 identified cell adhesion genes were downregulated. Cell adhesion molecules play critical roles in the development of the central nervous system, synapse formation, and immune responses (Hirano *et al*, 2003; Huntley *et al*, 2002; Lee and Benveniste, 1999; Milner and Campbell, 2002; Scheiffele, 2003). For instance, neurotrimin and activated leukocyte cell adhesion molecule are necessary for outgrowth of neurites and deficiency in scavenger receptor class B



**Figure 4** Age and PMI were not significantly correlated with gene expression. Cases were clustered based on the expression levels of 482 Class II genes. Control and alcoholic cases form two separate nodes. The age and PMI for each individual are color coded under the case IDs (yellow and blue scale). The expression level of each gene was median-centered; green indicates lower expression than the median level for that gene, and red indicates higher expression than median levels. Note that case A8 was clustered with the control cases.

(member 2) can cause neuropathies. Ethanol has been shown to change neural cell-cell adhesion *in vitro* (Charness *et al*, 1994). Downregulation of genes in this group may contribute to the compromise of neuronal functions in alcoholic human brain.

Several genes have been proposed to be risk factors for both alcoholism and Alzheimer's Disease, for example, transferrin (van Rensburg *et al*, 2004, 2000; Zambenedetti *et al*, 2003). Genes including calpain3, presenilin 1 and a glial specific gene (PRKC, apoptosis, WT1, regulator) are all related to Alzheimer's Disease and were regulated in the current alcohol study (Guo *et al*, 1998). These findings partly support the idea that different brain disorders may show similar changes in gene expression, and such changes reflect a 'nonspecific brain tissue malfunction' (Mirnics and Pevsner, 2004). However, this is purely speculative and further investigations will be necessary to make inferences.

In the PCA analysis of functional subgroups, case groups could be distinguished based on the expression of as few as

12 genes. This supports our previous conclusion that relatively small expression changes in the frontal cortex of alcoholic human brain are quite consistent across individuals and likely represent important alcohol-responsive genes. The hierarchical clustering of 27 cases identified two primary clusters, one comprised of control cases and the other alcoholic cases. One exception was A8 which was the only case that had been treated for another psychiatric disorder (depression). Depression alters gene expression (Sibille *et al*, 2004) and may explain the deviation from the alcoholic cluster. Overall, clustering of 27 cases only by history of alcohol consumption but not age or PMI suggests strongly the accuracy of our list of genes as candidate alcohol-responsive genes.

In summary, our present study identified a number of novel candidate alcohol-responsive genes in alcoholic frontal cortex. These genes are involved in cellular functions such as cell adhesion and other neural diseases. This study together with others suggest that the molecular pathology of alcoholism is manifest as changes in several key functional pathways that can be defined by gene expression profiles.

#### ACKNOWLEDGEMENTS

We are grateful to the Brisbane Node of the National Health and Medical Research Council (NHMRC) Brain Bank and the New South Wales Tissue Resource Centre for providing alcoholic and control brain tissue for analysis. We acknowledge Mrs Sonja Webb and Mrs Donna Sheedy for their assistance in providing detailed information for each of the cases used. We also thank the next of kin for providing informed written consent for the studies. We appreciate statistical assistance from Dr Igor Ponomarev. Financial support was provided by the National Institute of Alcoholism and Alcohol Abuse (USA, NIH AA12404) and the NHMRC.

#### REFERENCES

- Bueno R, Loughlin KR, Powell MH, Gordon GJ (2004). A diagnostic test for prostate cancer from gene expression profiling data. *J Urol* 171: 903–906.
- Bull JH, Ellison G, Patel A, Muir G, Walker M, Underwood M *et al* (2001). Identification of potential diagnostic markers of prostate cancer and prostatic intraepithelial neoplasia using cDNA microarray. *Br J Cancer* 84: 1512–1519.
- Charness ME, Safran RM, Perides G (1994). Ethanol inhibits neural cell-cell adhesion. *J Biol Chem* 269: 9304–9309.
- Chomczynski P, Sacchi N (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem* 162: 156–159.
- Datta S (2001). Exploring relationships in gene expressions: a partial least squares approach. *Gene Expr* 9: 249–255.
- Datta S (2003). Comparisons and validation of statistical clustering techniques for microarray gene expression data. *Bioinformatics* 19: 459–466.
- DeRisi JL, Iyer VR, Brown PO (1997). Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* 278: 680–686.
- Flatscher-Bader T, van der Brug M, Hwang JW, Gochee PA, Matsumoto I, Niwa S-I *et al* (2005). Alcohol-responsive genes in the frontal cortex and nucleus accumbens of human alcoholics. *J Neurochem* 93: 359–370.
- Geschwind DH (2003). DNA microarrays: translation of the genome from laboratory to clinic. *Lancet Neurol* 2: 275–282.



- Guo Q, Fu W, Xie J, Luo H, Sells SF, Geddes JW *et al* (1998). Par-4 is a mediator of neuronal degeneration associated with the pathogenesis of Alzheimer disease. *Nat Med* **4**: 957–962.
- Harper C, Kril J (1989). Patterns of neuronal loss in the cerebral cortex in chronic alcoholic patients. *J Neurol Sci* **92**: 81–89.
- Hirano S, Suzuki ST, Redies C (2003). The cadherin superfamily in neural development: diversity, function and interaction with other molecules. *Front Biosci* **8**: d306–d355.
- Huntley GW, Gil O, Bozdagi O (2002). The cadherin family of cell adhesion molecules: multiple roles in synaptic plasticity. *Neuroscientist* **8**: 221–233.
- Killion PJ, Sherlock G, Iyer VR (2003). The Longhorn Array Database (LAD): an open-source, MIAME compliant implementation of the Stanford Microarray Database (SMD). *BMC Bioinformatics* **4**: 32.
- Kim H, Golub GH, Park H (2005). Missing value estimation for DNA microarray gene expression data: local least squares imputation. *Bioinformatics* **21**: 187–198.
- Kim S, Dougherty ER, Shmulevich L, Hess KR, Hamilton SR, Trent JM *et al* (2002). Identification of combination gene sets for glioma classification. *Mol Cancer Ther* **1**: 1229–1236.
- Lavoie J, Butterworth RF (1995). Reduced activities of thiamine-dependent enzymes in brains of alcoholics in the absence of Wernicke's encephalopathy. *Alcohol Clin Exp Res* **19**: 1073–1077.
- Lee SJ, Benveniste EN (1999). Adhesion molecule expression and regulation on cells of the central nervous system. *J Neuroimmunol* **98**: 77–88.
- Lewohl JM, Dodd PR, Mayfield RD, Harris RA (2001). Application of DNA microarrays to study human alcoholism. *J Biomed Sci* **8**: 28–36.
- Lewohl JM, Wang L, Miles MF, Zhang L, Dodd PR, Harris RA (2000). Gene expression in human alcoholism: microarray analysis of frontal cortex. *Alcohol Clin Exp Res* **24**: 1873–1882.
- Liu J, Lewohl JM, Dodd PR, Randall PK, Harris RA, Mayfield RD (2004). Gene expression profiling of individual cases reveals consistent transcriptional changes in alcoholic human brain. *J Neurochem* **90**: 1050–1058.
- Lu T, Pan Y, Kao SY, Li C, Kohane I, Chan J *et al* (2004). Gene regulation and DNA damage in the ageing human brain. *Nature* **429**: 883–891.
- Marcotte ER, Srivastava LK, Quirion R (2003). cDNA microarray and proteomic approaches in the study of brain diseases: focus on schizophrenia and Alzheimer's disease. *Pharmacol Ther* **100**: 63–74.
- Mayfield RD, Lewohl JM, Dodd PR, Herlihy A, Liu J, Harris RA (2002). Patterns of gene expression are altered in the frontal and motor cortices of human alcoholics. *J Neurochem* **81**: 802–813.
- Mayfield RD, Liu J, Randall PK, Lewohl JM, Dodd PR, Harris RA (2003). Methods for the identification of differentially expressed genes in human post-mortem brain. *Methods* **31**: 301–305.
- Milner R, Campbell IL (2002). The integrin family of cell adhesion molecules has multiple functions within the CNS. *J Neurosci Res* **69**: 286–291.
- Mirnics K, Pevsner J (2004). Progress in the use of microarray technology to study the neurobiology of disease. *Nat Neurosci* **7**: 434–439.
- Mirnics K, Levitt P, Lewis DA (2004). DNA microarray analysis of postmortem brain tissue. *Int Rev Neurobiol* **60**: 153–181.
- Mirnics K, Middleton FA, Marquez A, Lewis DA, Levitt P (2000). Molecular characterization of schizophrenia viewed by microarray analysis of gene expression in prefrontal cortex. *Neuron* **28**: 53–67.
- Mor O, Nativ O, Stein A, Novak L, Lehavi D, Shibolet Y *et al* (2003). Molecular analysis of transitional cell carcinoma using cDNA microarray. *Oncogene* **22**: 7702–7710.
- Nestler EJ (2000). Genes and addiction. *Nat Genet* **26**: 277–281.
- Nguyen DV, Rocke DM (2002a). Multi-class cancer classification via partial least squares with gene expression profiles. *Bioinformatics* **18**: 1216–1226.
- Nguyen DV, Rocke DM (2002b). Partial least squares proportional hazard regression for application to DNA microarray survival data. *Bioinformatics* **18**: 1625–1632.
- Nguyen DV, Rocke DM (2002c). Tumor classification by partial least squares using microarray gene expression data. *Bioinformatics* **18**: 39–50.
- Okutsu J, Tsunoda T, Kaneta Y, Katagiri T, Kitahara O, Zembutsu H *et al* (2002). Prediction of chemosensitivity for patients with acute myeloid leukemia, according to expression levels of 28 genes selected by genome-wide complementary DNA microarray analysis. *Mol Cancer Ther* **1**: 1035–1042.
- Perez-Enciso M, Tenenhaus M (2003). Prediction of clinical outcome with microarray data: a partial least squares discriminant analysis (PLS-DA) approach. *Hum Genet* **112**: 581–592.
- Scheiffele P (2003). Cell-cell signaling during synapse formation in the CNS. *Annu Rev Neurosci* **26**: 485–508.
- Schramm M, Falkai P, Tepest R, Schneider-Axmann T, Przkora R, Waha A *et al* (1999). Stability of RNA transcripts in post-mortem psychiatric brains. *J Neural Transm* **106**: 329–335.
- Sibille E, Arango V, Galfalvy HC, Pavlidis P, Erraji-Benchekroun L, Ellis SP *et al* (2004). Gene expression profiling of depression and suicide in human prefrontal cortex. *Neuropsychopharmacology* **29**: 351–361.
- Smith DI (2002). Transcriptional profiling develops molecular signatures for ovarian tumors. *Cytometry* **47**: 60–62.
- Sokolov BP, Jiang L, Trivedi NS, Aston C (2003). Transcription profiling reveals mitochondrial, ubiquitin and signaling systems abnormalities in postmortem brains from subjects with a history of alcohol abuse or dependence. *J Neurosci Res* **72**: 756–767.
- Storey JD, Tibshirani R (2003). Statistical significance for genome-wide studies. *Proc Natl Acad Sci USA* **100**: 9440–9445.
- Taxman DJ, MacKeigan JP, Clements C, Bergstralh DT, Ting JP (2003). Transcriptional profiling of targets for combination therapy of lung carcinoma with paclitaxel and mitogen-activated protein/extracellular signal-regulated kinase kinase inhibitor. *Cancer Res* **63**: 5095–5104.
- Tkachev D, Mimmack ML, Ryan MM, Wayland M, Freeman T, Jones PB *et al* (2003). Oligodendrocyte dysfunction in schizophrenia and bipolar disorder. *Lancet* **362**: 798–805.
- Troyanskaya O, Cantor M, Sherlock G, Brown P, Hastie T, Tibshirani R *et al* (2001). Missing value estimation methods for DNA microarrays. *Bioinformatics* **17**: 520–525.
- van Rensburg SJ, Berman P, Potocnik F, MacGregor P, Hon D, de Villiers N (2004). 5- and 6-glycosylation of transferrin in patients with Alzheimer's disease. *Metab Brain Dis* **19**: 89–96.
- van Rensburg SJ, Berman PA, Potocnik FC, Taljaard JJ (2000). Glycosylation of transferrin in Alzheimer's disease and alcohol-induced dementia. *Metab Brain Dis* **15**: 243–247.
- Worst TJ, Vrana KE (2005). Alcohol and gene expression in the central nervous system. *Alcohol Alcohol* **40**: 63–75.
- Zambenedetti P, De Bellis G, Biunno I, Musicco M, Zatta P (2003). Transferrin C2 variant does confer a risk for Alzheimer's disease in caucasians. *J Alzheimers Dis* **5**: 423–427.
- Zembutsu H, Ohnishi Y, Tsunoda T, Furukawa Y, Katagiri T, Ueyama Y *et al* (2002). Genome-wide cDNA microarray screening to correlate gene expression profiles with sensitivity of 85 human cancer xenografts to anticancer drugs. *Cancer Res* **62**: 518–527.

Supplementary Information accompanies the paper on the Neuropsychopharmacology website (<http://www.nature.com/npp>).