American Journal of Medical Genetics Part B (Neuropsychiatric Genetics) 144B:683–690 (2007)

# Transcriptional Profiling of Subjects From the Iowa Adoption Studies

**Robert A. Philibert,[1,2]* Gi-Yung Ryu,[3] Jae-Geun Yoon,[3] Harinder Sandhu,[1] Nancy Hollenbeck,[1] Tracy Gunter,[1] Amanda Barkhurst,[1] William Adams,[1] and Anup Madan[3]**

[1]*Department of Psychiatry, The University of Iowa, Iowa City, Iowa*
[2]*Neuroscience Program, The University of Iowa, Iowa City, Iowa*
[3]*Department of Neurosurgery, The University of Iowa, Iowa City, Iowa*

**Transcriptional profiling has been used to identify gene expression patterns indicative of general medical illnesses such as atherosclerosis. However, whether these methods can identify common psychiatric disorders has not been established. To answer this question with respect to nicotine use, we used genome-wide expression profiling lymphoblast cell lines from six actively smoking Iowa Adoption Studies (IAS) subjects and nine "clean" control subjects, followed by real-time PCR (RT-PCR) of gene expression patterns in lymphoblast derived RNA from 94 subjects in the IAS. As compared to those from controls without a history of smoking (n = 9), the expression levels of 579 of 29,098 genes were significantly up-regulated and expression levels of 584 of 29,098 genes were significantly down-regulated in lymphoblast lines from currently smoking subjects (n = 6). RT-PCR confirmation of four select RNA levels confirmed the validity of the overall profile and revealed highly significant relationships between the expression of some of these transcripts and (1) major depression, (2) antisocial personality, (3) nicotine dependence, and (4) cannabis dependence. We conclude that the use of expression profiling may contribute significant insights into the biology of complex behavioral disorders.**


**KEY WORDS: substance use; major depression; SLC6A4; AUTS2; CAPN2; ELN; lymphoblast**

**Please cite this article as follows: Philibert RA, Ryu G-Y, Yoon J-G, Sandhu H, Hollenbeck N, Gunter T, Barkhurst A, Adams W, Madan A. 2007. Transcriptional Profiling of Subjects From the Iowa Adoption Studies. Am J Med Genet Part B 144B:683–690.**

## INTRODUCTION

The Iowa Adoption Studies (IAS) are a set longitudinal case and control studies, founded by Dr. Remi Cadoret, that use the adoption paradigm to examine the role of genetic (G), environmental (E), and gene–environment (GxE) interaction effects in the development and maintenance of depression and substance use in a group of 950 adoptees from the State of Iowa (for review see Yates et al. [1998]). Four prior waves of structured interviews, spaced approximately 5 years apart, have characterized the behavior and adoptive environment of each of the IAS subjects. The analysis of the data from these interviews has provided a great deal of insight in the separate, but vital role that G, E, and GxE factors play in depression and substance use disorders. In the current fifth wave, we are re-interviewing and phlebotomizing these subjects in an attempt to add a molecular component to these studies. In addition, we are preparing lymphoblast cell lines in order to provide DNA for genotyping and mRNA for expression analysis. The DNA from these cell lines is commonly used for genetic studies by a large number of investigators. However, though these lymphoblast cell lines are commonly used as models for serotonin transporter regulation [Hranilovic et al., 2004; Bradley et al., 2005], whether the mRNA derived from these cell lines is useful for other purposes is not known.

This situation may be changing. For example, Morello et al. [2004] have shown that lymphoblast cell lines from subjects with hyperlipidemia retain the transcriptional profile of the subjects from whom they were derived. Furthermore, a recent review has highlighted the utility of lymphoblasts in deciphering the biology of complex non-behavioral disorders [Liew et al., 2006]. Therefore, we were curious whether gene expression in our lymphoblast cell lines could provide insight into the biology of substance use disorders.

To examine this question, in an exploratory first step we performed genome-wide expression profiling in lymphoblast cell lines derived from a subset of six nicotine-dependent subjects and nine controls without a history of substance use or depression to identify transcripts differentially expressed in cases and controls. We then followed the expression of several of the more interesting candidate genes using real-time PCR (RT-PCR) extension in these 15 lines and an additional 79 lines, then analyzed the resulting data with respect to clinical information from waves 4 and 5 (1999–2003; 2004–current) of the IAS.

## MATERIALS AND METHODS

The overall design and methodology of the IAS and the complete demographic and clinical cohorts were gathered have been described in detail previously [Yates et al., 1998]. All clinical data used in this study was derived from structured interviews using the Structured Assessment for the Genetic Studies of Alcoholism (SSAGA) [Bucholz et al., 1994]. Symptom counts for major depression (maximum score of 9), antisocial personality disorder (ASPD) (maximum score of 7) and individual substance dependence (maximum score of 7) were derived from the SSAGA data using criteria from the DSM-IV [Association, 1994] as per our previous methods

Grant sponsor: NIH; Grant numbers: K08MH064714, R01DA015789, R01AI053264.

*Correspondence to: Robert A. Philibert, The University of Iowa, Rm 2-126 MEB, Iowa City, IA.
E-mail: robert-philibert@uiowa.edu

Received 11 July 2006; Accepted 18 January 2007
DOI 10.1002/ajmg.b.30512

[Yates et al., 1996]. Current smoking status was defined as smoking one or more cigarettes or cigars on a daily basis. Phlebotomy and interviewing was performed by a trained research assistant blind to clinical status. All procedures and methods were approved by the University of Iowa Institutional Review Board for Human Subjects.

The lymphoblast cell lines used for this study were derived using standard EBV transfection methods [Ginns et al., 1996] and grown using standard bovine serum based growth media supplemented with L-glutamine and penicillin/streptomycin. Media was changed for each of the cell lines 24 hr prior to extraction.

As a first step, RNA from these lines was prepared using an Invitrogen RNA purification kit (Invitrogen, Carlsbad, CA) according to the instructions provided by the manufacturer and stored in Ambion RNA storage solution at $-80^{\circ}$. All samples were examined spectrophotometrically to ensure quality with a subset of the samples being further analyzed using an Agilent Bioanalyzer.

From the first 94 subjects with complete biological and completed interviews in the recent wave, we selected six subjects (four male, two female) with active nicotine dependence and nine subjects (six male, three female) without a history of substance use or depression. We then performed microarray analysis on RNA extracted from their lymphoblast cell lines. To do this, total RNA was extracted using an Invitrogen Total RNA purification kit. Quality and integrity of the RNA samples was verified by Agilent Bioanalyzer. Total RNA (2 μg) was reverse-transcribed with the Chemiluminescent RT-IVT Labeling Kit (Applied Biosystems, Foster City, CA) and was hybridized to the ABI 1700 Human Genome Expression Microarray containing 29,098 gene-specific 60-mer oligonucleotide probes per manufacturer's protocol. Data were quantile normalized and a *t*-test was applied to each gene to determine statistical significance and provide *P*-values. Unsupervised hierarchical clustering was performed using Spotfire software to classify genes that have similar expression patterns. The Euclidean distance was used as a similarity method, based on gene expression levels, with no gene preselected. WPGMA (weighted average) was used to cluster genes. The Storey q-value method was then used to assess false discovery rate to determine the number of genes differentially expressed in 80% of the cell lines and to provide q-values [Storey and Tibshirani, 2003]. The significantly up-regulated ($n = 579$) and down-regulated ($n = 584$) gene lists represent those genes whose expression was arithmetically changed in the same direction (i.e., up or down) in each of the cases and significantly changed as compared to the controls as described above. All differentially expressed genes were selected at the false discovery rate of 5% per the method of Storey [Storey and Tibshirani, 2003].

From these differentially affected mRNAs, we chose two mRNAs [5HTT (or SLC6A4) and ELN] for further examination based on their biological plausibility and three mRNAs from the list of 30 most up-regulated or down-regulated transcripts (AUTS2, RPS19, and CAPN2; see Table IA,B) for further RT-PCR examination in all 94 lymphoblast cell lines available to us. As a first step in this process, we converted 5 μg aliquots of total RNA samples from each cell line to cDNA using an Applied Biosystems cDNA Archiving Kit (Applied Biosystems) according to manufacturer's directions. The resulting cDNA solutions were then diluted and aliquoted into master plates for robotic liquid handling.

Quantitative RT-PCR quantification of gene expression for these five transcripts was then performed using Taqman® Universal Master Mix Reagents (PE Biosystems, Branchburg, NJ) in combination with Assays-on-Demand® primer/probe sets specific for the ELN, AUTS2, RPS19, CAPN2 and 5HTT transcripts and an ABI 7900HT Sequence Detection System. Relative levels of these RNAs were determined by the comparative $C_T$ method using LDHA and GAPDH as normalizing controls [Dheda et al., 2004]. Z-scores for each RT-PCR run were calculated as per Fleiss [1981] and all runs were performed in duplicate. Although RT-PCR using the RPS19 primer probe set gave excellent signal, the correlation coefficient between duplicate samples was very low (<85%; normally this is >99%). Therefore, the results using this marker were excluded from further analyses. The resulting genetic (continuous variables of RNA expression Z-scores) and clinical data (e.g., ordinal symptom counts) was analyzed using JMP (version 5.1; SAS Institute, Cary, NC) using ordinal logistic regression using single markers or multiple RNA expression levels as indicated in the text. When testing whether more than one marker would better fit the model, the significance of the Effect Wald test from the all four markers together test (5HTT, ELN, AUTS2, and CAPN2; Table IV) was used to determine whether a marker was retained in the "Best Markers Only" model.

Two types of pathway analyses were conducted. The first was an unsupervised clustering analysis of the differentially expressed genes to known pathways using the Panther analysis subroutine and default settings (Applied Biosystems, http://www.pantherdb.org/) and the binomial test as described by Cho and Campbell [2000]. The second analysis was conducted using Gene Ontology subroutine (http://www.godatabase.org/dev) [Feng et al., 2003]. Briefly, using the recently developed GoMiner program, a GO category structure based on genes currently carrying human GO annotations was constructed and used as Query Gene File [Zeeberg et al., 2003]. The resulting file was loaded into the GoMiner program to examine the distribution of these genes in the GO category structure.

The microarray data have been submitted to the NCBI GEO database. The accession numbers are GSM144125 through GSM144136.

## RESULTS

As our first step, we performed microarray analysis on RNA extracted from lymphoblast cell lines derived from six subjects with active nicotine dependence and nine controls without a history of substance abuse or major depression selected from the first 94 lymphoblast cell lines derived from the IAS. Four of the currently smoking nicotine cases were completely medication free, a fifth was on risperidone, while the sixth nicotine case was on atorvastatin and used albuterol, monteleukast, and fluticasone inhalers. Four of the controls were also medication free, one was atorvastatin, atenolol, and lisinopril; one was on atenolol; one was on naproxen, clopidogrel, and metoprolol; and one was on azathioprine, atorvastatin, glyburide-metformin, rosiglitazone, amlodipine, and enalopril. Overall, only one subject, the nicotine case who was prescribed risperidone, was on medications for psychiatric reasons.

Figure 1 shows the unsupervised hierarchical clustering of normalized expression levels of genes in lymphoblast cell lines. The expression levels of 579 genes were significantly up-regulated while the expression levels of 584 genes were significantly down-regulated in the lymphoblast cell lines from the six nicotine-dependent subjects as compared to those from the nine "clean" controls. Table IA,B gives a list of the top 30 most up-regulated and the 30 most down-regulated transcripts.

We compared genes that were differentially expressed in those with active nicotine dependence with the genes present in 107 regulatory and metabolic pathways using the Panther pathway analysis program (http://www.pantherdb.org). These pathways represent about 12.0% of the known genes present in public databases. Of the 1,163 genes differentially regulated

TABLE I. The Most Down- and Up-Regulated Genes

| Probe ID | Gene symbol | Cytoband | Panther process | GO cellular component | Fold change | *P*-value | q-value |
|---|---|---|---|---|---|---|---|
| A: Thirty most down-regulated genes | | | | | | | |
| 212042 | RPS19 | 19q13.2 | Protein metabolism and modification | | −3.69 | <0.010 | <0.047 |
| 128153 | COX5B | 2cen-q13 | Electron transport | Mitochondrial membrane | −3.63 | <0.007 | <0.040 |
| 174637 | | | Protein metabolism and modification | | −3.48 | <0.010 | <0.049 |
| 214080 | | | | | −3.41 | <0.003 | <0.028 |
| 118887 | | | Protein metabolism and modification | | −3.33 | <0.003 | <0.028 |
| 114043 | | | Muscle contraction | | −3.28 | <0.006 | <0.039 |
| 153002 | RPS10 | 6p21.31 | | Small ribosomal subunit | −3.26 | <0.008 | <0.045 |
| 228451 | | | | | −3.21 | <0.010 | <0.048 |
| 206935 | MRPL23 | 11p15.5-p15.4 | Protein metabolism and modification | Ribosome | −3.15 | <0.001 | <0.023 |
| 130943 | RPL10A | 6p21.3-p21.2 | Protein metabolism and modification | | −3.07 | <0.005 | <0.037 |
| 226948 | HIST1H4I | 6p21.33 | Nucleoside, nucleotide and nucleic acid metabolism | | −3.05 | <0.010 | <0.048 |
| 171120 | | | | | −3.03 | <0.005 | <0.035 |
| 119116 | COMMD6 | | | | −2.95 | <0.003 | <0.027 |
| 201308 | | | Protein metabolism and modification | | −2.93 | <0.009 | <0.047 |
| 146877 | | | Lipid, fatty acid and steroid metabolism | | −2.85 | <0.004 | <0.032 |
| 106007 | | | | | −2.85 | <0.008 | <0.045 |
| 225790 | GAMT | 19p13.3 | Biological process unclassified | | −2.84 | <0.008 | <0.043 |
| 127866 | | | | | −2.75 | <0.010 | <0.048 |
| 130996 | | | | | −2.72 | <0.005 | <0.035 |
| 158641 | CKLFSF7 | 3p22.3 | Biological process unclassified | Integral to membrane | −2.71 | <0.002 | <0.025 |
| 234453 | | | Protein metabolism and modification | | −2.69 | <0.007 | <0.040 |
| 170400 | | | | | −2.69 | <0.006 | <0.039 |
| 176535 | | | | | −2.65 | <0.009 | <0.045 |
| 154419 | DULLARD | 17p13.2; | Developmental processes | Plasma membrane | −2.64 | <0.001 | <0.019 |
| 128466 | | | | | −2.59 | <0.002 | <0.027 |
| 165281 | | | | | −2.59 | <0.007 | <0.043 |
| 187132 | TRAPPC1 | 17p13.2 | Biological process unclassified | Endoplasmic reticulum | −2.57 | <0.003 | <0.027 |
| 208933 | | | Biological process unclassified | | −2.56 | <0.003 | <0.028 |
| 214094 | RPL23A | 17q11 | Protein metabolism and modification | Ribosome | −2.50 | <0.010 | <0.048 |
| 209685 | | | Biological process unclassified | | −2.48 | <0.002 | <0.026 |
| B: Thirty most up-regulated genes | | | | | | | |
| 103642 | | | | | 5.76 | 0.002 | 0.025 |
| 202763 | | | | | 4.63 | 0.005 | 0.035 |
| 182098 | | | | Sigl transduction | 4.08 | 0.006 | 0.040 |
| 210742 | CAPN2 | 1q41-q42 | Protein metabolism and modification intracellular | | 3.65 | 0.000 | 0.011 |
| 135433 | MARVELD3 | 16q22.2 | | Membrane | 3.64 | 0.004 | 0.031 |
| 149398 | CAPN2 | 1q41-q42 | Protein metabolism and modification | Intracellular | 3.42 | 0.000 | 0.013 |
| 236527 | | | | | 3.34 | 0.007 | 0.040 |
| 220463 | | | | | 3.33 | 0.009 | 0.046 |
| 103654 | | | | | 3.31 | 0.007 | 0.040 |
| 182347 | | | | | 3.09 | 0.006 | 0.038 |
| 175689 | | | Cell structure and motility | | 3.08 | 0.002 | 0.027 |
| 223790 | PCDHGB7 | 5q31 | Sigl transduction | Integral to membrane | 3.06 | 0.002 | 0.026 |
| 225245 | | | | | 3.03 | 0.009 | 0.047 |
| 152537 | LCN9 | 9q34.3 | Transport | | 2.95 | 0.001 | 0.023 |
| 113849 | | | | | 2.81 | 0.011 | 0.049 |
| 233070 | | | | | 2.63 | 0.003 | 0.030 |
| 191441 | RSAD2 | 2p25.2 | Coenzyme and prosthetic group metabolism | | 2.62 | 0.000 | 0.008 |
| 118637 | DCN | 12q13.2 | Developmental processes | Extracellular matrix | 2.61 | 0.004 | 0.032 |

(*Continued*)

TABLE I. (*Continued*)

| Probe ID | Gene symbol | Cytoband | Panther process | GO cellular component | Fold change | *P*-value | q-value |
|---|---|---|---|---|---|---|---|
| 105004 | KCTD12 | 13q22.1 | Biological process unclassified | Membrane | 2.60 | 0.008 | 0.045 |
| 179831 | FLJ90586 | 7q35 | | | 2.58 | 0.006 | 0.039 |
| 103193 | | | Biological process unclassified | | 2.54 | 0.011 | 0.050 |
| 234834 | | | | | 2.50 | 0.003 | 0.028 |
| 164423 | | | | | 2.49 | 0.003 | 0.030 |
| 101575 | | | | | 2.43 | 0.005 | 0.035 |
| 189694 | COCH | 14q12-q13 | Sensory perception | | 2.39 | 0.004 | 0.032 |
| 118644 | MLKL | 16q22.3 | Biological process unclassified | | 2.35 | 0.001 | 0.018 |
| 115055 | C2orf26 | 2q13 | Biological process unclassified | | 2.34 | 0.008 | 0.043 |
| 164834 | | | Immunity and defense | | 2.34 | 0.003 | 0.028 |
| 198965 | LILRB4 | 19q13.4 | Immunity and defense | Integral to membrane | 2.34 | 0.000 | 0.011 |
| 172527 | GRK4 | 4p16.3 | Protein metabolism and modification | | 2.34 | 0.008 | 0.045 |

Gene name is per GenBank. The extent of up-regulation is expressed as the 'fold' of the net change. Where known, the function of the gene according to Panther analytic or Gene Ontogeny subroutine is given as described in the Materials and Methods Section. *P*-values and q-values were calculated as described in the Materials and Methods Section.




Fig. 1. Hierarchical cluster of genes identified in lymphoblast cell lines from subjects without a history of depression or nicotine use (control, n = 9) versus relative to subjects with active nicotine dependence (n = 6). Each column represents data from single lymphoblast cell line, and each row represents expression levels of single gene across the 15 samples. Up-regulated transcripts are shown in red and transcripts down-regulated in green. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

those subjects with active nicotine dependence, only 97 of them mapped into identified Panther pathways. In particular, a large number of these genes mapped into apoptosis signaling (n = 9), FAS signaling (n = 3), p53 pathway (n = 9), TGF-beta (n = 7), de novo purine biosynthesis (n = 5),

G-protein signaling (n = 8), and Alzheimer disease-presenilin (n = 7) pathways.

As a second step in our analysis, we tested the validity of our genome-wide transcription profiling using RT-PCR and one marker from the list top 30 most up-regulated and the 30 most down-regulated transcripts. As Figure 2 demonstrates, the amount of AUTS2 expression was significantly decreased (higher Z-score; ANOVA $P < 0.008$) and the amount of CAPN2 was significantly increased (lower Z-score; ANOVA $P < 0.004$) in the six nicotine cases as compared to the nine controls.

From the total list of up-regulated or down-regulated genes, we chose two transcripts for further examination based on a number of factors including biological plausibility, our prior interest, physical clustering of differential expressed genes, and the degree of differential expression. 5HTT (SLC6A4) the serotonin reuptake transporter was chosen because of its obvious plausibility and our past research interest with the gene [Cadoret et al., 2003; Bradley et al., 2005]. Elastin (ELN) was chosen because our prior interest in this gene [Philibert et al., 2003] and its localization to 7q11, a linkage peak for panic disorder [Crowe et al., 2001], a disorder whose frequency is increased in subjects with nicotine dependence [Isensee et al., 2003].

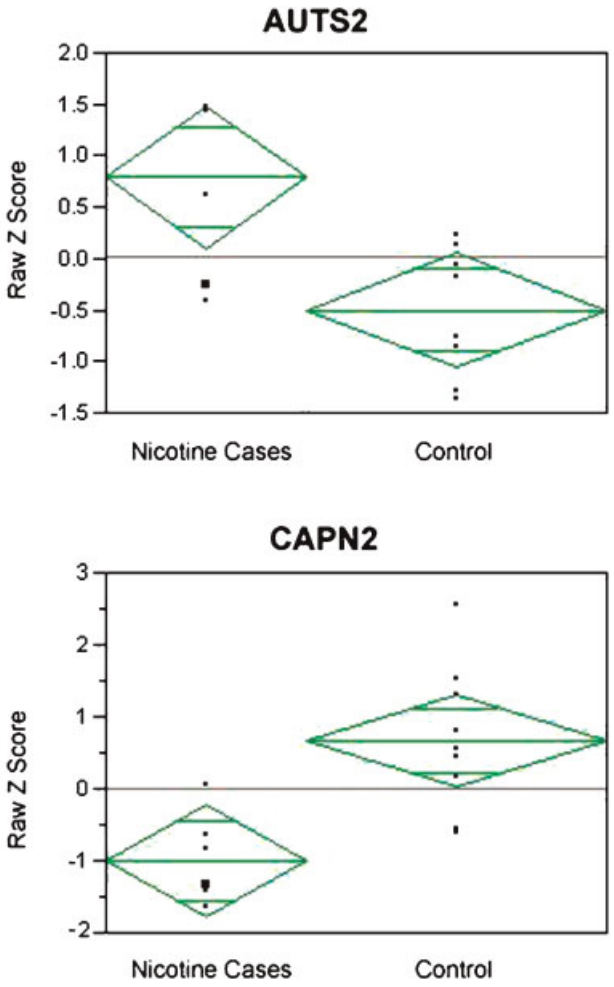


Fig. 2. Validation of the AUTS2 and CAPN2 expression profiling using RT-PCR. In order to validate the results of the microarray analyses, one marker from each of the 30 most highly up-regulated and down-regulated markers from Table 1A,B were evaluated using RT-PCR. AUTS2 expression was significantly decreased (higher Z-score; ANOVA $P < 0.008$) and the amount of CAPN2 was significantly increased (lower Z-score; ANOVA $P < 0.004$) in the six nicotine cases as compared to the nine controls. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

In order to extend and replicate this profile, we prepared RNA from the 79 additional lymphoblast lines derived from subjects from the IAS. Table II gives the distribution of the DSM-IV symptom counts for all 94 of the corresponding subjects for each of the most common behavioral disorders, major depressive disorder (MDD), ASPD, nicotine dependence (Nicotine Dep.), alcohol dependence (Alcohol Dep.), and Cannabis dependence (Cannabis Dep.) found in the IAS. Consistent with the design of this adoption study in which half of the participants are intentionally enrolled because of the strong history of MDD, ASPD and substance use in their biological parents [Yates et al., 1998], high symptom counts for MDD, ASPD and substance use are relatively frequent in these 94 adoptee subjects.

Since prior studies have shown that these common behavioral disorders have partially overlapping genetic and environmental diatheses, we analyzed the relationship of these variables using Spearman's Rho for ordinal variables. As Table III illustrates, each of the substance dependence count variables were significantly correlated with one another and ASPD symptom counts were significantly correlated with both alcohol and cannabis dependence symptom counts. Surprisingly, lifetime symptom counts for MDD were not significantly related to any of the other symptoms.

We then analyzed the relationship between these symptom counts and individual RNA expression levels using RT-PCR (Table IV). Contrary to our initial expectations given the results of the microarray analysis, the relationship between both AUTS2 and CAPN2 expression levels and nicotine dependence was not significant, but the relationships between AUTS2 and CAPN2 expression levels and cannabis dependence were highly significant. Retrospective analysis of the six nicotine "case" samples shed light on this matter. Four of the six nicotine cases had significant histories of cannabis dependence (4 of the 13 [31%] with cannabis dependence

TABLE II. DSM-IV Symptom Count for the Common Disorders

| | N | Number of Sx | MDD | ASPD | Nicotine Dep. | Alcohol Dep. | Cannabis Dep. |
|---|---|---|---|---|---|---|---|
| Men | 38 | | | | | | |
| | | 0 | 0 | 6 | 24 | 15 | 30 |
| | | 1 | 27 | 10 | 4 | 9 | 2 |
| | | 2 | 2 | 9 | 1 | 8 | 0 |
| | | 3 | 2 | 6 | 3 | 0 | 3 |
| | | 4 | 0 | 3 | 3 | 2 | 1 |
| | | 5 | 1 | 0 | 2 | 3 | 1 |
| | | 6 | 3 | 4 | 1 | 1 | 1 |
| | | 7 | 1 | 0 | 0 | 0 | 0 |
| | | 8 | 2 | | | | |
| | | 9 | 0 | | | | |
| Women | 56 | | | | | | |
| | | 0 | 0 | 25 | 29 | 37 | 51 |
| | | 1 | 25 | 14 | 3 | 9 | 0 |
| | | 2 | 0 | 8 | 4 | 7 | 1 |
| | | 3 | 2 | 5 | 5 | 1 | 0 |
| | | 4 | 2 | 2 | 11 | 0 | 0 |
| | | 5 | 2 | 0 | 2 | 1 | 2 |
| | | 6 | 7 | 1 | 2 | 0 | 2 |
| | | 7 | 9 | 1 | 0 | 1 | 0 |
| | | 8 | 4 | | | | |
| | | 9 | 5 | | | | |

Symptom counts for major depression (MDD, maximum score of 9), antisocial personality disorder (ASPD, maximum score of 7) and individual substance dependence (maximum score of 7) were derived from the SSAGA data using criteria from the DSM-IV [Association, 1994] as per our previous methods [Yates et al., 1996].

TABLE III. Spearman's Rho Correlation of Symptom Counts

| Variable | By variable | Spearman Rho | $P >$ Rho |
|---|---|---|---|
| Nicotine Dep. symptom | MDD symptom | 0.0897 | 0.3902 |
| Alcohol Dep. symptom | MDD symptom | 0.1285 | 0.2172 |
| Alcohol Dep. symptom | Nicotine Dep. symptom | **0.3521** | **0.0005** |
| Cannabis Dep. symptom | MDD symptom | 0.0236 | 0.8212 |
| Cannabis Dep. symptom | Nicotine Dep. symptom | **0.2596** | **0.0115** |
| Cannabis Dep. symptom | Alcohol Dep. symptom | **0.5453** | **<0.000** |
| ASPD symptom Ct | MDD symptom | 0.0862 | 0.4089 |
| ASPD symptom Ct | Nicotine Dep. symptom | 0.1998 | 0.0535 |
| ASPD symptom Ct | Alcohol Dep. symptom | **0.6087** | **<0.0001** |
| ASPD symptom Ct | Cannabis Dep. symptom | **0.5059** | **<0.0001** |

Values for which $P < 0.05$ are given in bold.

symptoms in the sample). In contrast, these six samples only represented 15% of those 41 individuals in the sample who had some symptoms of nicotine dependence, which suggest that in some respects these six nicotine "cases" were actually better cases for cannabis dependence. AUTS2 expression was also highly correlated with ASPD symptom count as well. ELN expression was significantly correlated with MDD and nicotine symptom counts, as well as the current smoker status. By itself, 5HTT expression was not correlated with any syndrome.

We next tested to see if using information from all the markers or markers with evidence of association, as evidenced by significant individual Effect Wald tests on the whole model test, would increase the significance of the relationship of gene expression and symptom counts using multiple ordinal regression (Table IV). Highly significant relationships could be found between two marker panel sets for current smoking status and both nicotine and cannabis symptom counts. But the addition of other markers did not improve the significance of the relationships between ELN and CAPN2 expression and MDD and ASPD symptom counts, respectively.

## DISCUSSION

In summary, we report that the expression of select genes in RNA obtained from lymphoblast cell lines is correlated with lifetime symptom counts of major depression, alcohol dependence, cannabis dependence, and nicotine dependence.

Before discussing the data, several caveats should be noted. First, since data for each of these subjects from the most recent interview are not fully available, all clinical information with the exception of current nicotine use was derived from structured interviews performed 2–5 years before the lymphoblast immortalization. However, our hand check of these interviews demonstrates that the information is highly accurate and that no significant changes in substance use or vulnerability to depression have occurred. Second, the clinical information is not independently verified (i.e., urine drug tests). Therefore, it is quite possible that some of clinical information may be inaccurate. Third, we have not fully accounted for the possible effects of medications on gene expression. But we note that (a) most of our subjects are not on psychiatric medications at all including only one of the six nicotine cases and none of the nine "clean" controls, (b) these are cell cultures considerably removed from the direct presence of these medications, and (c) our case and control profiling of lymphoblast cultures derived from 16 panic disorder probands (unpublished work), some of whom were not on medications, shows similar results. Fourth, these results are not corrected for multiple comparisons. However, we will note that many of these syndromes are inter-correlated (Table II) and the significance of many of these findings would survive any correction. Hence, we present these results uncorrected and advise the reader to use caution. Fifth and finally, it should be noted that the profiled cells are lymphoblasts, not lymphocytes. However, we note that others have demonstrated that they have similar gene expression properties [Morello et al., 2004].

The microarray and RT-PCR data with respect to ELN expression and nicotine dependence was the most interesting to us for several reasons. First, panic disorder is threefold more common in those with nicotine dependence and linkage for panic disorder has been linked to the 7q21 locus containing elastin [Crowe et al., 2001; Isensee et al., 2003]. Second, in previous work, we have investigated the role of this gene in panic disorder and found mildly suggestive evidence of a role for in panic disorder [Philibert et al., 2003]. Third, ELN and AUTS2 (which was associated with current smoking status in this study), both of which were down-regulated in the nicotine microarray expression data, each map to 7q21. This suggests the possibility that altered gene expression across the gene region may be differentially affected in nicotine dependence and that role of genetic variability and epigenetic modifications at these linked loci should be explored.

TABLE IV. Ordinal Regression Analysis of RNA Expression Levels and Symptom Counts

| Condition | Individual marker tests | | | | All markers together | Best markers only |
|---|---|---|---|---|---|---|
| Marker name | 5HTT | AUTS2 | ELN | CAPN2 | | |
| MDD symptom count | $P < 0.96$ | $P < 0.58$ | $P < 0.02$ | $P < 0.91$ | $P < 0.13$ | $P < 0.02$ (ELN only) |
| Nicotine Dep. symptom count | $P < 0.21$ | $P < 0.44$ | $P < 0.003$ | $P < 0.39$ | $P < 0.003$ | $P < 0.0005$ (ELN, 5HTT) |
| Alcohol Dep. symptom count | $P < 0.32$ | $P < 0.06$ | $P < 0.89$ | $P < 0.19$ | $P < 0.16$ | $P < 0.04$ (ELN, AUTS2) |
| Cannabis Dep. symptom count | $P < 0.51$ | $P < 0.002$ | $P < 0.57$ | $P < 0.002$ | $P < 0.0004$ | $P < 0.0001$ (AUTS2, CAPN2) |
| ASPD symptom count | $P < 0.25$ | $P < 0.0001$ | $P < 0.97$ | $P < 0.97$ | $P < 0.0003$ | $P < 0.0001$ (CAPN2) |
| Current smoker | $P < 0.36$ | $P < 0.10$ | $P < 0.001$ | $P < 0.93$ | $P < 0.0006$ | $P < 0.0003$ (ELN, AUTS2) |

Symptom counts and current smoker status are derived from the SSAGA using DSM-IV criteria as described in the Materials and Methods Section. All *P*-values are uncorrected for multiple comparisons. A detailed description of the analytic techniques is given in the Materials and Methods Section.

In contrast to our initial expectations, 5HTT expression by itself was not associated with any syndrome. There was a modest effect of 5HTT expression levels when combined with information from ELN expression with respect to nicotine dependence. This is not surprising since nicotine dependence and major depression have a common overlapping genetic substrate [Levinson et al., 2003] and the role for altered serotonergic activity in both major depression and nicotine dependence is strong. Still, it is mildly surprising that 5HTT levels did not significantly correlate with lifetime symptom counts for depression. However, since some of the subjects who were depressed probably were being treated with drugs that affect 5HTT function, it is possible that the presence of these antidepressants confounded the subsequent expression in the lymphoblast cell lines.

The distinctive association patterns of AUTS2 and CAPN2 gene expression in both substance use and ASPD is particularly intriguing to us. ASPD and substance dependence in general have strongly overlapping genetic and environmental diatheses [Yates et al., 1998; Fu et al., 2002]. If replicated, the strong association of AUTS2 expression with both cannabis use and ASPD, but CAPN2 only with cannabis dependence, may allow the beginning of the dissection of these intertwined etiologies.

A critical question is why gene expression in cell lines derived from the periphery is correlated with processes more associated with CNS function. We cannot be certain, but one reason may be that their expression may be affected by the same genetic and environmental factors as their counterparts in the CNS. On the face, this proposition may seem ill-founded. But it may well be that physiological triggers of abnormal behavior such as low thyroid hormone or high cortisol levels that can be produced by somatic illness or environmental stress may leave the same indelible signature on gene transcription in the periphery as it does on behavior emanating from CNS processing [Liew et al., 2006]. Furthermore, we are noting similar results from our case and control profiling of subjects with panic disorder (unpublished work). Hence, if these data can be replicated, a new window into our understanding of the effect of stressors on behavior may be open for further investigation.

This caveat is particularly important because at least two groups have attempted to develop expression profile signa tures of schizophrenia [Bowden et al., 2006; Glatt et al., 2005]. Unfortunately, the profiles developed by the two groups largely do not overlap. However, since Glatt and colleagues used whole blood RNA while Bowden and colleagues used whole blood lymphocyte RNA, both of which contain RNA from mixed cell populations, it is quite possible that the failure to generate similar profiles has resulted from methodological differences. Since the current study uses RNA derived from a single precursor cell type, B-lymphocytes, it may well generate a more robust signature for certain illnesses.

It is quite conceivable that the results of the current study could lead to diagnostic testing for certain behavioral illnesses. If so, the current findings have significant clinical impact. But before that becomes a reality, the development of marker panels that not only detect but also distinguish partially related disorders such as panic disorder and major depression from one another need to be developed. We are in the process of systematically profiling "pure cases" (e.g., no co-morbid illnesses) of various common disorders such as panic disorder, to develop such a panel. Once this is done, we plan on revisiting this subject more intensively using large batteries (>50) of markers in the cell lines that are currently being prepared. However, our experience with six nicotine case profiled in this article, four of whom also have substantial co-morbid cannabis use, suggests that defining what constitutes a "pure" case of any given disorder may not be a simple task and that complete specification of the clinical information may be essential for the accurate replication of findings. Hence, the result described in this communication can be described as exploratory. We fully expect that other efforts will more finely delineate and differentiate the changes attributable to nicotine use from those from the co-morbid disorders.

In summary, we report that expression profiling of lymphoblast cell lines demonstrates significant relationships of gene expression with lifetime symptom counts of major depression and antisocial personality, as well as nicotine, cannabis, and alcohol dependence. Since these cell lines strongly mimic the expression of their corresponding native B-lymphocytes, we suggest that expression profiling of mRNA expression may provide a mechanism for the development of laboratory diagnostic algorithms for the common mental disorders. However, it should be noted that the described results are best described as a hypothesis generating experiments and that additional efforts are necessary to differentiate the effects of nicotine use on gene expression from those caused by the co-morbid disorders.

## ACKNOWLEDGMENTS

The University of Iowa has filed for intellectual protection rights on behalf of Dr. Philibert and Dr. Madan for some of the ideas described in this communication. The authors do not have any other conflicts with respect to this work.

## REFERENCES

Association AP. 1994. Diagnostic and statistical manual of mental disorder, fourth edition. Washington, DC: American Psychiatric Association.

Bowden NA, Weidenhofer J, et al. 2006. Preliminary investigation of gene expression profiles in peripheral blood lymphocytes in schizophrenia. Schizophr Res 82(2–3):175–183.

Bradley S, Dodelzon K, et al. 2005. The relationship of serotonin transporter gene polymorphisms and haplotypes to mRNA transcription. Am J Med Genet 136(1):58–61.

Bucholz KK, Cadoret R, et al. 1994. A new, semi-structured psychiatric interview for use in genetic linkage studies: A report on the reliability of the SSAGA. J Stud Alcohol 55(2):149–158.

Cadoret RJ, Langbehn D, et al. 2003. Associations of the serotonin transporter promoter polymorphism with aggressivity, attention deficit, and conduct disorder in an adoptee population. Compr Psychiatry 44(2):88–101.

Cho RJ, Campbell MJ. 2000. Transcription, genomes, function. Trends Genet 16(9):409–415.

Crowe RR, Goedken R, et al. 2001. Genomewide survey of panic disorder. Am J Med Genet 105(1):105–109.

Dheda K, Huggett JF, et al. 2004. Validation of housekeeping genes for normalizing RNA expression in real-time PCR. Biotechniques 37(1): 112–114. 116. 118–119.

Feng W, Wang G, et al. 2003. Development of gene ontology tool for biological interpretation of genomic and proteomic data. AMIA Annu Symp Proc 2003:839.

Fleiss JL. 1981. Statistical Methods for Rates and Proportions, second edition. New York, NY: John Wiley & Sons Inc.

Fu Q, Heath AC, et al. 2002. Shared genetic risk of major depression, alcohol dependence, and marijuana dependence: Contribution of antisocial personality disorder in men. Arch Gen Psychiatry 59(12):1125–1132.

Ginns EI, Ott J, et al. 1996. A genome-wide search for chromosomal loci linked to bipolar affective disorder in the Old Order Amish. Nat Genet 12(4):431–435.

Glatt SJ, Everall IP, et al. 2005. Comparative gene expression analysis of blood and brain provides concurrent validation of SELENBP1 up-regulation in schizophrenia. Proc Natl Acad Sci USA 102(43):15533–15538.

Hranilovic D, Stefulj J, et al. 2004. Serotonin transporter promoter and intron 2 polymorphisms: Relationship between allelic variants and gene expression. Biol Psychiatry 55(11):1090–1094.

Isensee B, Wittchen HU, et al. 2003. Smoking increases the risk of panic: Findings from a prospective community study. Arch Gen Psychiatry 60(7):692–700.

Levinson DF, Zubenko GS, et al. 2003. Genetics of recurrent early-onset depression (GenRED): Design and preliminary clinical characteristics of a repository sample for genetic linkage studies. Am J Med Genet Part B 119(1):118–130.

Liew CC, Ma J, et al. 2006. The peripheral blood transcriptome dynamically reflects system wide biology: A potential diagnostic tool. J Lab Clin Med 147(3):126–132.

Morello F, de Bruin TWA, et al. 2004. Differential gene expression of blood-derived cell lines in familial combined hyperlipidemia. Arterioscler Thromb Vasc Biol 24(11):2149–2154.

Philibert RA, Nelson JJ, et al. 2003. Role of elastin polymorphisms in panic disorder. Am J Med Genet Part B 117B(1):7–10.

Storey JD, Tibshirani R. 2003. Statistical methods for identifying differentially expressed genes in DNA microarrays. Methods Mol Biol 224:149–157.

Yates WR, Cadoret RJ, et al. 1996. An adoption study of DSM-IIIR alcohol and drug dependence severity. Drug Alcohol Depend 41(1):9.

Yates W, Cadoret R, et al. 1998. The Iowa Adoption Studies Methods and Results. In: LaBuda M, Grigorenko E, editors. On the way to individuality: Methodological Issues in Behavioral Genetics. Hauppauge NY: Nova Science Publishers. pp 95–125.

Zeeberg BR, Feng W, et al. 2003. GoMiner: A resource for biological interpretation of genomic and proteomic data. Genome Biol 4(4):R28.