THE JOURNAL OF CHILD PSYCHOLOGY AND PSYCHIATRY

ACAMH THE ASSOCIATION FOR CHILD AND ADOLESCENT MENTAL HEALTH

Journal of Child Psychology and Psychiatry 54:10 (2013), pp 1109–1119 doi:10.1111/jcpp.12119

Gene × smoking interactions on human brain gene expression: finding common mechanisms in adolescents and adults

Samuel L. Wolock,[1]* Andrew Yates,[2]* Stephen A. Petrill,[3] Jason W. Bohland,[4] Clancy Blair,[5] Ning Li,[1] Raghu Machiraju,[2,6] Kun Huang,[2,6,7] and Christopher W. Bartlett[1,8]

[1]Battelle Center for Mathematical Medicine, Nationwide Children's Hospital, Columbus, OH, USA; [2]Department of Biomedical Informatics, The Ohio State University College of Medicine, Columbus, OH, USA; [3]Department of Psychology, The Ohio State University, Columbus, OH, USA; [4]Department of Health Sciences, Boston University, Boston, MA, USA; [5]Department of Applied Psychology, New York University, New York, NY, USA; [6]Department of Computer Science and Engineering, The Ohio State University, Columbus, OH, USA; [7]The CCC Biomedical Informatics Shared Resource, The Ohio State University, Columbus, OH, USA; [8]Department of Pediatrics, The Ohio State University College of Medicine, Columbus, OH, USA

Background: Numerous studies have examined gene × environment interactions (G × E) in cognitive and behavioral domains. However, these studies have been limited in that they have not been able to directly assess differential patterns of gene expression in the human brain. Here, we assessed G × E interactions using two publically available datasets to assess if DNA variation is associated with post-mortem brain gene expression changes based on smoking behavior, a biobehavioral construct that is part of a complex system of genetic and environmental influences. **Methods:** We conducted an expression quantitative trait locus (eQTL) study on two independent human brain gene expression datasets assessing G × E for selected psychiatric genes and smoking status. We employed linear regression to model the significance of the Gene × Smoking interaction term, followed by meta-analysis across datasets. **Results:** Overall, we observed that the effect of DNA variation on gene expression is moderated by smoking status. Expression of 16 genes was significantly associated with single nucleotide polymorphisms that demonstrated G × E effects. The strongest finding ($p = 1.9 \times 10^{-11}$) was neurexin 3-alpha (*NRXN3*), a synaptic cell–cell adhesion molecule involved in maintenance of neural connections (such as the maintenance of smoking behavior). Other significant G × E associations include four glutamate genes. **Conclusions:** This is one of the first studies to demonstrate G × E effects within the human brain. In particular, this study implicated NRXN3 in the maintenance of smoking. The effect of smoking on *NRXN3* expression and downstream behavior is different based upon SNP genotype, indicating that DNA profiles based on SNPs could be useful in understanding the effects of smoking behaviors. These results suggest that better measurement of psychiatric conditions, and the environment in post-mortem brain studies may yield an important avenue for understanding the biological mechanisms of G × E interactions in psychiatry. **Keywords:** Genetics, environment, brain development, developmental psychopathology.

Introduction

The effects of gene by environment (G × E) interaction on behavior are presumably mediated by differences in gene expression in the brain. There is a large and growing literature assessing measured G × E interactions to behaviors (Duncan & Keller, 2011), and less frequently with methylation or gene expression in peripheral tissues such as blood (Bagot & Meaney, 2010; Boulle et al., 2012; Mehta et al., 2013). That said, there is an emerging literature examining gene expression in the prefrontal cortex to several important psychiatric and health related outcomes. Early-life adversity is associated with widespread variation in gene expression in prefrontal cortex of rhesus macaques (Provencal et al., 2012). As well, evidence for G × E variation in gene expression in the brain from these monkeys indicates that serotonin transporter genotype interacts with early-life stress to predict an epigenetic mark, H3K4me3, binding at the SLC6A4 promoter in the hippocampus (Lindell et al., 2012). Generalization of brain research findings to human has been accomplished with smaller hypothesis-driven replication study designs (Mcgowan et al., 2009). However, there has been no previous large sample study to directly examine G × E in brain-based gene expression for complex human behaviors. This is, in part, due to the impossibility of acquiring gene expression data in vivo and difficulty in acquiring appropriate behavioral and environmental data retrospectively from a postmortem cohort. Thus, although child and adolescent psychiatry has made progress defining DNA variants and their relationship with behavior, the gene expression mechanisms that mediate these relationships have only recently begun to be characterized (Hebebrand, Scherag,

*These authors contributed equally to the work.

Conflicts of interest statement: No conflicts declared.

Schimmelmann, & Hinney, 2010; Naumova, Lee, Rychkov, Vlasova, & Grigorenko, 2013). Understanding the neural underpinnings of behavior would improve if direct links from DNA variation to brain gene expression were to be studied in the context of behavior (Kleinman et al., 2011; Richards et al., 2012).

This study is the first to examine gene expression in a gene × environment interaction study of smoking. Smoking is a leading cause of preventable death in the United States at about one in five deaths each year (US Surgeon General, 2012). And despite this fact, 88% of smokers begin by age 18 (Center for Behavioral Health Statistics & Quality, 2011). However, the neurobiological and epigenetic basis of smoking addiction is not well known. In this introduction, we first discuss the intersection of behavior and molecular genetics and discuss the advantage of gene expression studies on power to detect gene × environment interactions. Next, we define the eQTL study paradigm and design while explaining relevant concepts. We then establish the basis of this present study, examining brain gene expression and smoking behaviors. Next, we establish relevance of the molecular underpinning of smoking maintenance effects on the brain as it relates to preventing smoking initiation in youth and possible avenues for intervention. In the discussion section, we further argue how the methods applied here may be generalized to other behavioral and psychiatric traits in children and adolescents.

Molecular genetic studies of gene × environment interactions require an assay of DNA variants, a defined and measured behavior of interest, and an environment that can be quantified or used for stratification/grouping (Caspi & Moffitt, 2006). Assaying DNA variation has become robust and routine. The choice of environment and selected metrics pose another barrier as these choices come from an unbounded pool of candidates, many of which may be related to behavior, but not via interactions with genes; power is difficult to establish when designing a study and generally assumed to be low (Duncan & Keller, 2011). Finally, in psychiatric studies, the behavioral or clinical outcome is separated from processes involving DNA by many intervening biological steps.

By focusing on gene expression changes in human brain as the outcome, it is possible to reduce the number of biological steps between G × E and the outcome. Moreover, power is also enhanced by employing molecular variables that have strong a priori relationships (Johnson, Wang, & Sadee, 2005). By doing so, it is possible to model the principles of a G × E expression study in a setting that has greater signal to noise. However, we have chosen our variable set to provide more than a demonstration of a G × E study. It is also to understand smoking behaviors in a way that is relevant to prevention and intervention of smoking that could ideally be applied to adolescents before the cumulative and gradual psychological and medical consequences manifest (Audrain-Mcgovern, Rodriguez, Rodgers, Cuevas, & Sass, 2012; Baler & Volkow, 2011). This is the first molecular genetic study of smoking behavior, using gene × environment interactions directly associated with brain biology.

Mapping traits to DNA variation and gene expression

In animals and lower order model systems, it is possible to reverse engineer the role of genes by modifying or even removing a gene and observing the consequences called the reverse genetics paradigm (Lehner, 2013). However, the forward genetics paradigm applied to humans maps observed characteristics to specific chromosomal regions, with the ultimate goal of identifying specific genes and specific changes within those genes that influence the observed characteristics. In psychiatry, forward genetics is an observational design where observed traits are measured and DNA variation is assessed for an association with the observed trait through one of several different statistical genetic paradigms (Bailey-Wilson & Wilson, 2011; Do, Kathiresan, & Abecasis, 2012; Dube & Hegele, 2013). When the observed trait is quantitative in nature, mapping the molecular genetic basis can be called quantitative trait locus (QTL) mapping (Falconer & Mackay, 1996). The chromosomal location, or locus, is the quantity of interest in the calculations. One quantitative trait and perhaps the most fundamental in genetics is measurement of gene expression. This special type of QTL is called an expression quantitative trait locus or eQTL (Damerval, Maurice, Josse, & De Vienne, 1994; De Vienne, Leonardi, & Damerval, 1988). Therefore, eQTL mapping is the paradigm for finding DNA variations that underlie variation in gene expression. As DNA (i.e. the chromosomal location) is the direct template for the synthesis of RNA (i.e. gene expression), they represent a close biological relationship, and therefore, eQTL studies have much greater effect sizes than observed in behavioral studies.

Gene expression may vary by cell and tissue type, environmental conditions covering everything from moment-to-moment changes in metabolic activities to fully extrinsic variables like temperature, and past experience (Maranville, Luca, Stephens, & Di Rienzo, 2012). Some genes are stably expressed regardless of other factors, while others may be expressed so quickly or in such a complicated pattern that they cannot be studied using current technology. In between these two boundary conditions lie genes that show variation in expression to some degree. When eQTL studies are conducted on gene expression from the human brain, they will clearly not be sensitive to genes with rapid responses to the environment as the logistics of tissue donation preclude

the necessary speed in sample processing. However, many of the genes expressed in the brain do seem to have variability from person to person that is stable enough and can be reliably measured for successful eQTL mapping (Colantuoni et al., 2011; Gibbs et al., 2010; Heinzen et al., 2008; Kang et al., 2011; Liu et al., 2010; Myers et al., 2007; Webster et al., 2009).

To accommodate eQTL studies, datasets must include genotyping and expression profiling *in the same subjects*, although only two such datasets have meaningful measures of behavior (e.g. smoking status) at the time of death (Colantuoni et al., 2011; Liu et al., 2010). Here, we interrogated those two studies within a meta-analysis framework, to increase power, as the use of different microarray platforms prevented a single unified reanalysis. Our study is able to find novel eQTLs using SNP, gene expression, and environmental measures, in this case smoking status, and it quantifies the interaction between SNPs and the environment that may affect gene expression in the brain.

Gene expression as an outcome of smoking

Aside from demonstrating the veracity of human brain eQTL studies in understanding the brain basis of G × E, smoking behavior represents an important outcome in its own right from genetic, environmental, and biological perspectives. Smoking behavior and nicotine dependence have strong genetic and environmental components with genetics probably playing a more significant role in dependence (Vink, Willemsen, & Boomsma, 2005). A number of genetic loci have been previously associated with smoking behaviors. For example, a recent meta-analysis (Tobacco & Genetics Consortium, 2010) identified SNPs associated with the quantitative phenotype number of cigarettes per day within the nicotinic receptor gene *CHRNA3* (chromosome 15q25.1), in LOC100188947 (10q15) and in *EGLN2* (19q13). In addition, eight SNPs in the gene encoding brain-derived neurotrophic factor (BDNF) were associated with smoking initiation. There is evidence for the involvement of a variety of other genes in smoking-related behaviors, including those encoding the class of cell adhesion proteins known as neurexins, specifically neurexin 3 (Bierut et al., 2007; Docampo et al., 2012; Nussbaum et al., 2008). Neurexins are important developmentally and are directly associated with dopamine neurons in experimental data in cell culture (Noelker et al., 2012) and indirectly associated through protein–protein interactions with CASK (calcium/calmodulin-dependent serine protein kinase), a protein where mutations in a binding partner cause Parkinson's disease (Houlden & Singleton, 2012) – a disorder to dopamine depletion.

Twin studies in humans also provide evidence that the genetics of smoking behavior in adulthood is informative about the genetics in adolescence. Several meta-analyses of the heritability of tobacco use report 37%–60% of variation in smoking initiation behavior and 46%–59% in persistence of smoking to be accounted by additive genetics (Hall, Madden, & Lynskey, 2002; Li, Cheng, Ma, & Swan, 2003; Sullivan & Kendler, 1999). When considering smoking initiation longitudinally, the heritability was not found to change significantly from ages 13 to 46 in a meta-analysis consisting of only longitudinal datasets (Bergen, Gardner, & Kendler, 2007). More recent data on the heritability of nicotine dependence (defined using *DSM-III-R* criteria), used here as an indicator of smoking persistence, increased modestly from age 15 to 18, and thereafter stabilized from age 18 to 21. However, the additive genetic correlation between ages 15 and 18, and between ages 18 and 21 was quite stable at 0.22 and 0.27, respectively (Tully, Iacono, & Mcgue, 2010). Using cigarettes per day as a quantitative metric of smoking behavior assessed at ages 14, 17, 20, and 24 showed that the majority of smoking persistence across those ages was due to additive genetics (Vrieze, Mcgue, & Iacono, 2012).

The genetics of smoking from adolescence into adulthood also shows a longitudinally stable common risk factor with illicit drug use and alcohol use (Baker, Maes, Larsson, Lichtenstein, & Kendler, 2011). The heritability of this common genetic factor was consistent, estimated between 52% and 54% at each age range 13–14, 16–17, and 19–20. Importantly, the heritability of the common variance across ages was >40% for all time points, indicating that the majority of additive genetic effects common to smoking, alcohol and illicit drug use is longitudinally stable.

Smoking is an environmental factor that impacts the brain, resulting in, for example, a greater risk of dementia and cognitive impairments (Anstey, Von Sanden, Salim, & O'kearney, 2007). Animal studies have demonstrated that exposure to nicotine can drive changes in gene expression within the cerebral cortex and other brain structures (e.g. Kenny, File, & Rattray, 2001; Konu et al., 2001; Trauth, Seidler, Mccook, & Slotkin, 1999; Trauth, Seidler, & Slotkin, 2000), and particularly genes associated with the mesocorticolimbic dopaminergic pathway known to play a role in addiction (Flatscher-Bader & Wilce, 2009). Thus, a natural question that we address in the present analysis is whether the expression levels of certain genes in the human frontal cortex, treated as quantitative phenotypes in the eQTL approach, are driven by interactions between genotype and smoking history.

Gene-by-smoking interactions in child and adolescent psychology

In this study, we examined whether smoking behavior changes the expression of genes in the frontal lobe and whether these changes can be explained by

interactions between smoking status and SNPs. It is important to note that as the environmental variable measured here is smoking at the time of death, this study design may not be sensitive to genetic factors related to the *initiation* of smoking (Thorgeirsson et al., 2010; Tobacco & Genetics Consortium, 2010; Yoon et al., 2012), but will be important for understanding the effect of smoking on the genetic regulation of the maintenance of smoking behavior. For this reason, we are interested in the extent to which the reward pathways typically studied in the addiction literature are similar to or distinct from neural pathways related to learning and stabilization of synaptic connections to maintain a behavior.

To accomplish the goal of the study, to understand which genes are involved in smoking maintenance in a gene × environment analysis, smoking status at time of death was used here as a measure of the environment. Regular doses of nicotine to the brain alter the environment for neurons and drive changes in gene expression to adapt to that environment (Flatscher-Bader & Wilce, 2009). The outcome of interest was the statistical significance of the interaction term from a linear regression of gene expression values on SNP × Smoking. A significant interaction term indicates a nonadditive effect of smoking and SNP, whereby the effects of smoking on gene expression are not the same across all levels of SNP genotype. We focused our study to genes most commonly studied in psychiatric genetics as these genes have the greatest a priori likelihood for association.

Table 1 Genes of interest from text mining psychiatric literature*

AANAT	***ABAT***	*ACHE*	*ADRA1D*	*ADRA2A*
ADRA2B	*ADRB1*	*ADRB2*	*ADRB3*	*AKT1*
APOE	***APP***	*BDNF*	***CHAT***	*CHRFAM7A*
CHRM1	*CHRM2*	*CHRM3*	*CHRM4*	*CHRM5*
CHRNA1	*CHRNA10*	*CHRNA2*	*CHRNA3*	*CHRNA4*
CHRNA5	*CHRNA6*	*CHRNA7*	*CHRNA9*	*CHRNB1*
CHRNB2	*CHRNB3*	*CHRNB4*	*CHRND*	*CHRNE*
CHRNG	*CLOCK*	*COMT*	*CRELD2*	*CRHR1*
DAG1	*DAOA*	*DBH*	***DBI***	*DDC*
DISC1	*DISC2*	*DRD1*	*DRD2*	*DRD3*
DRD4	*DRD5*	*DTNBP1*	*DUSP27*	*DUT*
EFNB3	*FOXP2*	*GABBR1*	***GABBR2***	***GABRA1***
GABRA2	*GABRA3*	*GABRA4*	***GABRA5***	*GABRA6*
GABRB1	***GABRB2***	*GABRB3*	*GABRD*	*GABRE*
GABRG1	***GABRG2***	*GABRG3*	*GABRP*	*GABRQ*
GABRR1	*GABRR2*	*GABRR3*	***GAD1***	*GALR3*
GRIA1	***GRIA2***	*GRIA3*	*GRIA4*	*GRID1*
GRID2	*GRID2IP*	***GRIK1***	***GRIK2***	*GRIK3*
GRIK4	***GRIK5***	*GRIN1*	*GRIN2A*	*GRIN2B*
GRIN2C	*GRIN2D*	*GRIN3A*	*GRIN3B*	*GRINA*
GRM1	*GRM2*	***GRM3***	*GRM4*	***GRM5***
GRM6	***GRM7***	*GRM8*	*HTR1A*	*HTR1B*
HTR1D	*HTR1E*	*HTR1F*	***HTR2A***	*HTR2B*
HTR2C	*HTR3A*	*HTR3B*	*HTR3C*	*HTR3D*
HTR3E	*HTR4*	*HTR5A*	*HTR6*	*HTR7*
HTR7P1	***ISOC2***	*KIAA0125*	***MAOA***	*NOTCH4*
NRXN1	***NRXN3***	***NTRK2***	*OPRM1*	*PPP1R1B*
PRODH	***RGS4***	*RIC3*	*SLC18A2*	*SLC18A3*
SLC22A2	*SLC29A4*	*SLC32A1*	***SLC36A1***	*SLC5A7*
SLC6A1	*SLC6A11*	*SLC6A12*	*SLC6A2*	*SLC6A3*
SLC6A4	*SLC6A7*	*SNAI3*	***SNAP25***	*SOD2*
TPH1	*TPH2*	*YIF1B*		

*Bolded genes passed probe filtration and were included in the analysis.

Methods

In this section, we include a general outline of the methods, leaving more specific details, mostly from genomic and bioinformatics aspects of the study, for the online appendix (Online supporting information: Methodology). The greater detail included in the online supplementary information provides additional data and citations to allow for replication of our work. Note that we used publicly available de-identified datasets where informed consent was obtained by the primary dataset investigators.

We began with a text mining of the psychiatric genetics literature to determine the most commonly studied genes (see Online supplementary information: Methodology). The top 100 genes from this text mining approach were initially selected for study inclusion. We also expanded the list to include additional neurotransmitter family genes including receptors and metabolism genes for the major neurotransmitter systems: acetylcholine, dopamine, GABA, glutamate, and serotonin. Table 1 shows the 158 genes selected based on these criteria. However, after filtering expression data (see below) for data quality, suitable variability for analysis, and number of subjects with expression values above background detection levels, as well as requiring that gene expression data were available in both datasets, only 29 genes remained (bolded in Table 1).

The data consisted of two publicly available human brain eQTL datasets (Colantuoni et al., 2011; Liu et al., 2010), COLANTUONI and LIU, respectively, that include smoking status at time of death, genome-wide SNP microarray genotypes that assay genetic variants distributed across the entire human genome, and microarray-based gene expression profiles from frontal lobe samples. Each dataset, COLANTUONI and LIU, was pre-processed independently. For gene expression data, we performed covariate correction, including correction for multiple batches run at different times, tissue source, sex, ancestry, age, postmortem interval (PMI), RNA integrity number (RIN), brain pH and psychiatric status for LIU subjects (see online supplementary information: Methodology for statistical details on covariate correction). A sudden cause of death was reported in 86% of the subjects and no agonal states were reported to be neurologically based. These datasets then underwent quality control checks to ensure that the processed data were appropriate for downstream eQTL analysis (see sections *COLANTUONI gene expression* and *LIU gene expression* in the online supporting information:

Methodology). Genotypes were filtered for quality before use in (a) estimating ancestry principal component covariates, which are critical for controlling false positives related to ethnic stratification, and (b) input for genotype imputation, a method that makes meta-analysis possible by ensuring that both datasets are genotyped using the same set of SNPs.

After the above pre-processing workflow was carried out, G × E analysis included $N = 144$ subjects from COLANTUONI (38 smokers and 106 nonsmokers; ages 18–77, avg = 43.1; 68% male) and $N = 84$ subjects from LIU (56 smokers and 28 nonsmokers; ages 19–65, avg = 44.6; 70% male). These datasets were used for the primary analysis in the study. Secondary analysis of the effects of SNPs without regard to smoking was carried out using larger sample sizes, as information on smoking status was not a requirement in the filtering process ($N = 186$ for COLANTUNI and $N = 127$ for LIU). Within or nearby the genes selected for analysis, there were 405,875 SNPs in common across both datasets.

A meta-analysis of the two eQTL analyses (see *Statistical Analysis* in online supporting information: Methodology) to test for G × E effects was conducted by assessing the significance of the linear regression interaction term (SNP genotype × smoking status) as a predictor of gene expression. After correcting for multiple testing (see online supporting information: Methodology), the threshold for significance was 8.1×10^{-8}. Additional follow-up analyses assessed if smoking alone or SNP genotype alone (i.e. main effects) accounted for additional variance in gene expression. Meta-analysis R^2 values were obtained by pooling sums of squares for residuals and sums of squares for error across studies and applying the standard adjusted R^2 formula with pooled *df* to account for both studies. Lastly, properties of the SNPs and the loci were examined to interpret the main findings (see *SNP annotations* in the online supporting information: Methodology).

Results

Analysis of SNPs (G) × smoking status (E) on gene expression

G × E analysis was conducted on all SNP-gene pairs (405,875 SNPs and 29 genes; with multiple gene measurements present in some COLANTUONI genes). The meta-analysis yielded 312 SNPs (328 SNP-gene pairs) meeting the typical cut-off for significance in a genome-wide association study, $p = 5 \times 10^{-8}$, with the most significant result found for expression of the neurexin 3 (*NRXN3*) gene at SNP rs12411798 ($p = 1.9 \times 10^{-11}$). There were 16 genes associated with significant G × E interactions (see Table 2), including four glutamate-related genes (*GRIK1*, *GRIK2*, *GRIK5*, are kainite-sensitive ionotropic receptor subunits and *GRM3*, a metabotropic glutamate receptor) and three gamma-aminobutyric acid (GABA)-related genes (*ABAT*, *GAD1*, involved in synthesis and *GABBR2,* a GABA receptor subunit), in addition to another member of the neurexin family, *NRXN1,* implicated in stabilizing synaptic proteins and possibly has a special role in dopamine reward pathways. Most SNPs were more than 1 Mb distant from the associated gene, indicating an indirect effect on transcription is most likely (i.e. the effect is mediated by at least one other gene). Many associated SNPs are close together and represent the same underlying eQTL. We examined this by assuming that correlation patterns among nearby SNPs (called linkage disequilibrium) represent the same locus (i.e. the same cause in gene expression change). Table 2 lists the number of independent G × E loci for the top 16 genes. The proportion of variance accounted for by SNPs and G × E in this study was determined by forward selection of regression terms with R^2 values presented in Table 2. While the average R^2 was .16, comparison of the two studies indicated an average shrinkage of 60%, indicating that additional studies will be needed to more accurately estimate the true effect sizes,

Table 2 List of top 16 genes and their characteristics

Gene	Significant eQTLs	Significant SNPs	Peak SNP	Peak *p*-value	Peak adjusted R^2
NRXN3	17	150	rs12411798	1.87×10^{-11}	.25
ABAT	4	23	rs6558338	9.30×10^{-10}	.14
GRIK1	3	5	rs2268132	9.50×10^{-10}	.12
GRIK5	2	22	rs118031442	1.09×10^{-9}	.12
NRXN1	4	20	rs1646239	1.94×10^{-9}	.17
GRIK2	4	21	rs2585459	1.98×10^{-9}	.13
NTRK2	1	5	rs11135168	3.95×10^{-9}	.08
RGS4	1	1	rs35472486	8.42×10^{-9}	.17
APP	3	22	rs4608331	8.74×10^{-9}	.29
ISOC2	7	18	rs1046695	9.77×10^{-9}	.15
GAD1	1	4	rs149541349	1.08×10^{-8}	.16
PRODH	6	14	rs73735512	1.27×10^{-8}	.15
GABBR2	3	7	rs216137	1.30×10^{-8}	.13
SNAP25	4	14	rs8073177	1.54×10^{-8}	.22
HTR2A	1	1	rs35472486	1.95×10^{-8}	.13
GRM3	1	1	rs74005241	4.08×10^{-8}	.11

a regression of eQTL number on gene length indicates a relationship (R^2 = .50, p = .002), whereby longer genes have more eQTLs. When trans SNPs are included. This indicates that when a gene is important for G × E, length offers more opportunities for trans regulation of gene expression.

We examined if the SNPs in our study had previously described biological functions (i.e. database annotations) to determine if the eQTLs had a common underlying functional characteristic. No such trends were observed. Only 0.3% of SNPs were located in a transcription factor-binding site where the SNP could possibly disrupt binding as a mechanism to alter transcription. Only one SNP was predicted to have functional consequences at the protein level using standard prediction tools SIFT and PolyPhen 2 (Kumar, Henikoff, & Ng, 2009; Ramensky, Bork, & Sunyaev, 2002). We examined CHiP-seq data from all neutrally derived cell lines from the whole genome survey known as the Encyclopedia of DNA Elements (ENCODE) database to see if any of our SNPs fell within experimentally determined DNA-binding proteins (presumed regulatory regions) or open markers of chromatin (presumed transcriptionally active regions). A total of 69% of the genome fell within ENCODE regions, while 75% of the significant SNPs fell within those regions. This enrichment was significant (binomial test, p = .009), although the mechanism for how our eQTL SNPs affect transcription requires additional experimental work.

The effect of smoking alone

We examined if smoking is associated with changes in gene expression independent of SNP genotype. Analysis of each dataset yielded no significant genes with differential expression by smoking status. Top genes in COLANTUONI included two genes with prior evidence for a relationship with smoking, *DUSP27* (Nielsen et al., 2010) and *YIF1B* (Carrel et al., 2008), although the best false discovery rate (FDR) was 0.57, far from the necessary cut-off of 0.05. The fact that two genes previously identified as having expression changes associated with smoking were highlighted by the analysis, but not at an appropriate significance level, may indicate low power. Top hits within the LIU dataset had an FDR of 0.37, with no overlap of the top 100 genes from COLANTUONI.

NRXN3

SNPs associated with *NRXN3* expression comprised 6 of the top 10 SNPs based on p-value from the eQTL analysis, including the top SNP, and thus we sought to further characterize results related to this gene. Upon visual inspection of the data, the G × E loci display the classical 'X' shape (Figure 1), which can cause assessment of the main effect of genotype and the main effect of smoking to be absent, as averaging over either factor minimizes the remaining effect. In agreement with this observation, regression of gene expression on these SNPs does not yield significant SNP effects. To assess the possibility that the additive linear model was only approximating a more genetically complex model, we performed Bayesian genetic modeling to estimate parameters that can differentiate between the additive model and models that have dominance (i.e. where one copy of an allele has the same effect as two copies of that allele, thus introducing a nonlinearity). Parameter estimates were fully consistent with the additive linear model for the top 100 G × E SNPs, indicating that the linear regression model was adequate for these loci.

Discussion

This study represents the first attempt at examining G × E effects on gene expression in the human brain. In this study, we have demonstrated a signif-

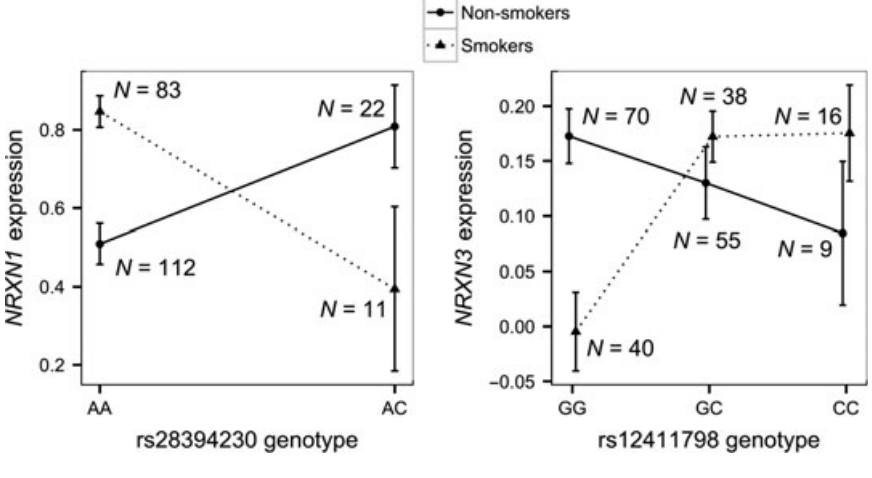

Figure 1 Examples of the observed SNP × Smoking interactions. Mean gene expression from the SCAN normalization method (*y*-axis) with standard error bars is plotted by SNP genotype (*x*-axis) and smoking status (symbols). A gene expression value of 0 indicates background levels (no expression). The pattern of results indicates an interaction effect that implies little to no main effects for either SNP or Smoking as averaging over either predictor variable will largely eliminate the effect of the other

icant interaction between genotype and smoking status on the expression of a set of 16 genes that were of previous interest in psychiatric genetics, although none had been shown to be involved in smoking persistence through G × E effects. These novel findings for psychiatrically important genes suggest that those genes may predispose a variety of behaviors specifically through gene expression changes as demonstrated by our results. The most significant results were for six SNPs in the neurexin 3 gene (*NRXN3*), one of three genes in the neurexin family. Neurexins are presynaptic cell adhesion proteins that play an important role in the formation and function of synapses. The three neurexin genes are each transcribed from two promoter regions, resulting in alpha and beta neurexin proteins, which are alternatively spliced, yielding more than 1000 distinct isoforms that are differentially expressed in the brain (Ullrich, Ushkaryov, & Sudhof, 1995). As a whole, neurexins are widely expressed in the mammalian brain, including in the prefrontal cortical areas studied here. Neurexin genes have been previously associated with smoking (Bierut et al., 2007; Docampo et al., 2012) and also with alcohol (Hishimoto et al., 2007) and opioid (Lachman et al., 2007) dependence. SNPs in *NRXN3* specifically have been previously associated with susceptibility to smoking (Docampo et al., 2012), and with degree of smoking in schizophrenic patients (Novak, Boukhadra, Shaikh, Kennedy, & Le Foll, 2009). The present results add new evidence, indicating that *NRXN3* expression in the human prefrontal cortex is differentially affected by smoking behavior depending on an individual SNP profile. And like previous studies, the association is exclusively with the membrane-bound *NRXN3*-alpha transcript and not in the soluble *NRXN3*-beta isoform (Docampo et al., 2012; Novak et al., 2009). The association of both glutamatergic and GABAergic genes may imply that the balance between excitation and inhibition neural circuitry of prefrontal corticolimbic circuitry was associated with addiction, specifically prefrontal cortex-striatal circuitry. In addition to neurotransmitter findings, neurexins, involved in stabilizing synaptic proteins to perhaps modulate transmission efficiency, might affect the balance of excitation and inhibition to alter smoking persistence.

The lack of a main effect of smoking on expression of *NRXN3* in this study is accounted for by the nature of the observed cross-over interactions (Dick & Kendler, 2012) between a large set of SNPs and the categorical smoking variable on *NRXN3* expression. In particular, interactions such as those depicted in Figure 1 (X-shaped) result in the effect of either factor (genotype or smoking status) being largely eliminated when averaging across levels of the other. Such a result indicates, for instance, that even the direction of the effect of smoking on *NRXN3* expression in these brain areas depends on specific genomic variations. The lack of any observable main effects of smoking (i.e. genes that are differentially expressed in smoking group vs. nonsmoking group) may reflect heterogeneity within the two groups of subjects or lack of statistical power.

Our analyses used only a single categorical variable, which was coded in the existing eQTL datasets, to determine smoking behavior. Such a binary variable cannot account for the full complexity of smoking-related phenotypes, and quantitative measures (e.g. number of cigarettes per day) might provide additional power if they were available. Indeed, it has been suggested that smoking initiation, number of cigarettes per day, and nicotine dependence each reflects distinct stages of smoking behavior, and may involve distinct biological and genetic mechanisms (Breitling, Yang, Korn, Burwinkel, & Brenner, 2011; Mayhew, Flay, & Mott, 2000). However, the lack of more detailed smoking behavioral indicators would only reduce study power and would not induce false-positive results. Therefore, our findings are likely to be an underestimate of the number of specific genetic loci that are involved in smoking and gene expression in the human brain.

The overall results of this study reflect the utility of analyzing eQTLs, in which associations are tested between individual SNP genotypes and gene expression levels in brain tissue samples, and of G × E analyses, in which the effects of genotype and environmental variables on the phenotypic measure (gene expression in this case) are simultaneously modeled. In particular, we have shown that the effects of smoking to alter expression of psychiatric genes in the brain can depend on particular genotypes, based on two independent studies. These data imply that G × E for smoking persistence is driven by differential expression of genes based upon the genotype at nearby SNP loci. Or, to reframe the conclusion, expression of psychiatric genes relevant to smoking depends upon the environment and genotypes of the individual.

An issue when interpreting the lack of main effects on gene expression for G × E SNPs is the possibility that smoking causes epigenetic changes, here referring to modifications of DNA that change the regulatory potential of the DNA without changing the sequence (The Encode Project Consortium, 2012). Therefore, a SNP that is normally silent can induce changes in gene expression when one of the two alleles is more likely to be epigenetically modified than the other. Without assessing the known markers for DNA modification, such as methylation or histone binding, it is not possible here to assert positive evidence for this hypothesis. Regardless of the specific mechanism, however, understanding how the environment changes gene expression will shed light on which genes to study in more detail in conjunction with other behavioral designs. Because no main effect of smoking was found on *NRXN3* expression and also because there was no significant main effect of genotype, it is clear that many

standard statistical models, which do not explicitly account for G × E effects, will miss such results. Thus, we suggest that the combination of the eQTL approach with modeling of G × E interactions is a promising direction for establishing the relationships between genes, brain, and behavior in complex psychiatric disorders. These results also suggest the importance of – where feasible – including environmental and behavioral measures from individual subjects as future eQTL datasets are gathered.

Brain gene expression profiling has only recently been used in the domain of psychiatric disease etiology (Geschwind & Konopka, 2009; Luykx et al., 2013; Mexal et al., 2005; Richards et al., 2012), but this study extends that approach to G × E and indicates that future studies using this approach may be powerful for addressing other types of questions. This study suggests that human brain gene expression datasets, when coupled with behavioral and genetic data, provide useful phenotypes that can be incorporated into psychiatric genetics studies, and provides the first demonstration of G × E effects on gene expression in the human brain. As brain banks for child and adolescent subjects continue to grow, it would be invaluable to the field to collect environmental data wherever possible to facilitate additional studies on issues more widely appreciated in earlier developmental stages.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Data S1: Online supporting information

Acknowledgements

We gratefully acknowledge NIH funding from the National Institute of Deafness and Other Communication Disorders R01 DC009453 (to C.W.B.) and the National Institute of General Medical Sciences U01 GM092655 (to K.H.). This work was also supported in part by an allocation of computing time from the Ohio Supercomputer Center Grant PCCR0001 (C.W.B.).

Correspondence

Christopher W. Bartlett, 575 Children's Crossroad, Columbus, OH 43215, USA; Email: Christopher.Bartlett@NationwideChildrens.org

Key points

- Gene-by-environment interaction effects on human behavior are assumed to be mediated by changes in gene expression patterns in the brain, although no empirical demonstrations exist.
- An expression quantitative trait locus study assesses the relationship between DNA loci and gene expression values and can be extended to include a linear regression interaction term for DNA locus by environment (e.g. smoking).
- We observed that the effect of DNA variation on gene expression changes depends on smoking status, demonstrating that behavioral gene-by-environment interactions can affect brain gene expression.
- Several psychiatrically important genes involving neural connectivity and glutamate transmission associate with DNA variants that interact with smoking, suggesting an important role for these genes in the maintenance of smoking.

References

Anstey, K.J., Von Sanden, C., Salim, A., & O'kearney, R. (2007). Smoking as a risk factor for dementia and cognitive decline: a meta-analysis of prospective studies. *American Journal of Epidemiology*, *166*, 367–378.

Audrain-Mcgovern, J., Rodriguez, D., Rodgers, K., Cuevas, J., & Sass, J. (2012). Longitudinal variation in adolescent physical activity patterns and the emergence of tobacco use. *Journal of Pediatric Psychology*, *37*, 622–633.

Bagot, R.C., & Meaney, M.J. (2010). Epigenetics and the biological basis of gene × environment interactions. *Journal of the American Academy of Child and Adolescent Psychiatry*, *49*, 752–771.

Bailey-Wilson, J.E., & Wilson, A.F. (2011). Linkage analysis in the next-generation sequencing era. *Human Heredity*, *72*, 228–236.

Baker, J.H., Maes, H.H., Larsson, H., Lichtenstein, P., & Kendler, K.S. (2011). Sex differences and developmental stability in genetic and environmental influences on psychoactive substance consumption from early adolescence to young adulthood. *Psychological Medicine*, *41*, 1907–1916.

Baler, R.D., & Volkow, N.D. (2011). Addiction as a systems failure: focus on adolescence and smoking. *Journal of the American Academy of Child and Adolescent Psychiatry*, *50*, 329–339.

Bartlett, C.W., & Vieland, V.J. (2005). Two novel quantitative trait linkage analysis statistics based on the posterior probability of linkage: application to the COGA families. *BMC Genetics*, *6* (Suppl 1), S121.

Bartlett, C.W., Vieland, V.J., Bartlett, J., Bell, J.T., Bhattacharjee, S., Clerget-Darpoux, F., … & Zhou, J.Y. (2007). Discussing gene-gene interaction: warning–translating equations to English may result in jabberwocky. *Genetic Epidemiology*, *31* (Suppl 1), S61–S67.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *57*, 289–300.

Bergen, S.E., Gardner, C.O., & Kendler, K.S. (2007). Age-related changes in heritability of behavioral phenotypes over adolescence and young adulthood: a meta-analysis. *Twin Res Hum Genet*, *10*, 423–433.

Bierut, L.J., Madden, P.A., Breslau, N., Johnson, E.O., Hatsukami, D., Pomerleau, O.F., … & Ballinger, D.G. (2007). Novel genes identified in a high-density genome wide association study for nicotine dependence. *Human Molecular Genetics*, *16*, 24–35.

Boulle, F., Van Den Hove, D.L., Jakob, S.B., Rutten, B.P., Hamon, M., Van Os, J., … & Kenis, G. (2012). Epigenetic regulation of the BDNF gene: implications for psychiatric disorders. *Mol Psychiatry*, *17*, 584–596.

Breitling, L.P., Yang, R., Korn, B., Burwinkel, B., & Brenner, H. (2011). Tobacco-smoking-related differential DNA methylation: 27K discovery and replication. *American Journal of Human Genetics*, *88*, 450–457.

Carrel, D., Masson, J., Al Awabdh, S., Capra, C.B., Lenkei, Z., Hamon, M., … & Darmon, M. (2008). Targeting of the 5-HT1A serotonin receptor to neuronal dendrites is mediated by Yif1B. *The Journal of Neuroscience*, *28*, 8063–8073.

Caspi, A., & Moffitt, T.E. (2006). Gene-environment interactions in psychiatry: joining forces with neuroscience. *Nature Reviews Neuroscience*, *7*, 583–590.

Center for Behavioral Health Statistics and Quality (2011). *Results from the 2010 National Survey on Drug Use and Health: summary of National Findings*. Rockville, MD: US Department of Health and Human Services.

Chanda, P., Yuhki, N., Li, M., Bader, J.S., Hartz, A., Boerwinkle, E., … & Arking, D.E. (2012). Comprehensive evaluation of imputation performance in African Americans. *Journal of Human Genetics*, *57*, 411–421.

Chelala, C., Khan, A., & Lemoine, N.R. (2009). SNPnexus: a web database for functional annotation of newly discovered and public domain single nucleotide polymorphisms. *Bioinformatics*, *25*, 655–661.

Colantuoni, C., Lipska, B.K., Ye, T., Hyde, T.M., Tao, R., Leek, J.T., … & Kleinman, J.E. (2011). Temporal dynamics and genetic control of transcription in the human prefrontal cortex. *Nature*, *478*, 519–523.

Dai, M., Wang, P., Boyd, A.D., Kostov, G., Athey, B., Jones, E.G., … & Meng, F. (2005). Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. *Nucleic Acids Research*, *33*, e175.

Damerval, C., Maurice, A., Josse, J.M., & De Vienne, D. (1994). Quantitative trait loci underlying gene product variation: a novel perspective for analyzing regulation of genome expression. *Genetics*, *137*, 289–301.

Dayem Ullah, A.Z., Lemoine, N.R., & Chelala, C. (2012). SNPnexus: a web server for functional annotation of novel and publicly known genetic variants (2012 update). *Nucleic Acids Research*, *40*, W65–W70.

Dayem Ullah, A. Z., Lemoine, N. R., & Chelala, C. (in press). A practical guide for the functional annotation of genetic variations using SNPnexus. *Briefings in Bioinformatics*.

De Vienne, D., Leonardi, A., & Damerval, C. (1988). Genetic aspects of variation of protein amounts in maize and pea. *Electrophoresis*, *9*, 742–750.

Dick, D.M., & Kendler, K.S. (2012). The impact of gene-environment interaction on alcohol use disorders. *Alcohol Res*, *34*, 318–324.

Do, R., Kathiresan, S., & Abecasis, G.R. (2012). Exome sequencing and complex disease: practical aspects of rare variant association studies. *Human Molecular Genetics*, *21*, R1–R9.

Docampo, E., Ribases, M., Gratacos, M., Bruguera, E., Cabezas, C., Sanchez-Mora, C., … & Estivill, X. (2012). Association of Neurexin 3 polymorphisms with smoking behavior. *Genes, Brain and Behavior*, *11*, 704–711.

Dube, J.B., & Hegele, R.A. (2013). Genetics 100 for cardiologists: basics of genome-wide association studies. *Canadian Journal of Cardiology*, *29*, 10–17.

Duncan, L.E., & Keller, M.C. (2011). A critical review of the first 10 years of candidate gene-by-environment interaction research in psychiatry. *American Journal of Psychiatry*, *168*, 1041–1049.

Falconer, D.S., & Mackay, T.F.C. (1996). *Introduction to Quantitative Genetics*. London, England: Benjamin Cummings.

Flatscher-Bader, T., & Wilce, P.A. (2009). The effect of alcohol and nicotine abuse on gene expression in the brain. *Nutrition Research Reviews*, *22*, 148–162.

Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., … & Zhang, J. (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology*, *5*, R80.

Geschwind, D.H., & Konopka, G. (2009). Neuroscience in the era of functional genomics and systems biology. *Nature*, *461*, 908–915.

Gibbs, J.R., Van Der Brug, M.P., Hernandez, D.G., Traynor, B.J., Nalls, M.A., Lai, S.L., … & Singleton, A.B. (2010). Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS Genetics*, *6*, e1000952.

Haibe-Kains, B., Desmedt, C., Sotiriou, C., & Bontempi, G. (2008). A comparative study of survival models for breast cancer prognostication based on microarray data: does a single gene beat them all? *Bioinformatics*, *24*, 2200–2208.

Hall, W., Madden, P., & Lynskey, M. (2002). The genetics of tobacco use: methods, findings and policy implications. *Tob Control*, *11*, 119–124.

Hebebrand, J., Scherag, A., Schimmelmann, B.G., & Hinney, A. (2010). Child and adolescent psychiatric genetics. *European Child & Adolescent Psychiatry*, *19*, 259–279.

Heinzen, E.L., Ge, D., Cronin, K.D., Maia, J.M., Shianna, K.V., Gabriel, W.N., … & Goldstein, D.B. (2008). Tissue-specific genetic control of splicing: implications for the study of complex traits. *PLOS Biology*, *6*, e1.

Hishimoto, A., Liu, Q.R., Drgon, T., Pletnikova, O., Walther, D., Zhu, X.G., … & Uhl, G.R. (2007). Neurexin 3 polymorphisms are associated with alcohol dependence and altered expression of specific isoforms. *Human Molecular Genetics*, *16*, 2880–2891.

Hou, L., Phillips, C., Azaro, M., Brzustowicz, L.M., & Bartlett, C.W. (2011). Validation of a cost-efficient multi-purpose SNP panel for disease based research. *PLoS ONE*, *6*, e19699.

Hou, L., Wang, K., & Bartlett, C.W. (2012). Evaluation of a bayesian model integration-based method for censored data. *Human Heredity*, *74*, 1–11.

Houlden, H., & Singleton, A.B. (2012). The genetics and neuropathology of Parkinson's disease. *Acta Neuropathologica*, *124*, 325–338.

Howie, B., Fuchsberger, C., Stephens, M., Marchini, J., & Abecasis, G.R. (2012). Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nature Genetics*, *44*, 955–959.

Huang, Y., Bartlett, C.W., Segre, A.M., O'connell, J.R., Mangin, L., & Vieland, V.J. (2007). Exploiting gene × gene interaction in linkage analysis. *BMC Proceedings*, *1* (Suppl 1), S64.

Huang, Y., & Vieland, V.J. (2010). Association statistics under the PPL framework. *Genetic Epidemiology*, *34*, 835–845.

Johnson, W.E., Li, C., & Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*, *8*, 118–127.

Johnson, A.D., Wang, D., & Sadee, W. (2005). Polymorphisms affecting gene regulation and mRNA processing: broad implications for pharmacogenetics. *Pharmacology & Therapeutics*, *106*, 19–38.

Kang, H.J., Kawasawa, Y.I., Cheng, F., Zhu, Y., Xu, X., Li, M., … & Sestan, N. (2011). Spatio-temporal transcriptome of the human brain. *Nature*, *478*, 483–489.

Kenny, P.J., File, S.E., & Rattray, M. (2001). Nicotine regulates 5-HT(1A) receptor gene expression in the cerebral cortex and dorsal hippocampus. *European Journal of Neuroscience*, *13*, 1267–1271.

Kleinman, J.E., Law, A.J., Lipska, B.K., Hyde, T.M., Ellis, J.K., Harrison, P.J., & Weinberger, D.R. (2011). Genetic neuropathology of schizophrenia: new approaches to an old question and new uses for postmortem human brains. *Biological Psychiatry*, *69*, 140–145.

Konu, O., Kane, J.K., Barrett, T., Vawter, M.P., Chang, R., Ma, J.Z., … & Li, M.D. (2001). Region-specific transcriptional response to chronic nicotine in rat brain. *Brain Research*, *909*, 194–203.

Kumar, P., Henikoff, S., & Ng, P.C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols*, *4*, 1073–1081.

Lachman, H.M., Fann, C.S., Bartzis, M., Evgrafov, O.V., Rosenthal, R.N., Nunes, E.V., … & Knowles, J.A. (2007). Genomewide suggestive linkage of opioid dependence to chromosome 14q. *Human Molecular Genetics*, *16*, 1327–1334.

Lehner, B. (2013). Genotype to phenotype: lessons from model organisms for human genetics. *Nature Reviews Genetics*, *14*, 168–178.

Li, M.D., Cheng, R., Ma, J.Z., & Swan, G.E. (2003). A meta-analysis of estimated genetic and environmental effects on smoking behavior in male and female adult twins. *Addiction*, *98*, 23–31.

Lindell, S.G., Yuan, Q., Zhou, Z., Goldman, D., Thompson, R.C., Lopez, J.F., … & Barr, C.S. (2012). The serotonin transporter gene is a substrate for age and stress dependent epigenetic regulation in rhesus macaque brain: potential roles in genetic selection and gene × environment interactions. *Development and Psychopathology*, *24*, 1391–1400.

Liu, C., Cheng, L., Badner, J.A., Zhang, D., Craig, D.W., Redman, M., & Gershon, E.S. (2010). Whole-genome association mapping of gene expression in the human prefrontal cortex. *Mol Psychiatry*, *15*, 779–784.

Luykx, J.J., Bakker, S.C., Lentjes, E., Neeleman, M., Strengman, E., Mentink, L., … & Ophoff, R.A. (2013). *Genome-wide association study of monoamine metabolite levels in human cerebrospinal fluid*. Mol Psychiatry.

Maranville, J.C., Luca, F., Stephens, M., & Di Rienzo, A. (2012). Mapping gene-environment interactions at regulatory polymorphisms: insights into mechanisms of phenotypic variation. *Transcription*, *3*, 56–62.

Mayhew, K.P., Flay, B.R., & Mott, J.A. (2000). Stages in the development of adolescent smoking. *Drug and Alcohol Dependence*, *59* (Suppl 1), S61–S81.

Mcgowan, P.O., Sasaki, A., D'alessio, A.C., Dymov, S., Labonte, B., Szyf, M., … & Meaney, M.J. (2009). Epigenetic regulation of the glucocorticoid receptor in human brain associates with childhood abuse. *Nature Neuroscience*, *12*, 342–348.

Mehta, D., Klengel, T., Conneely, K. N., Smith, A. K., Altmann, A., Pace, T. W., … & Binder, E. B. (2013). Childhood maltreatment is associated with distinct genomic and epigenetic profiles in posttraumatic stress disorder. *Proceedings of the National Academy of Sciences of the United States of America*, *110*, 8302–332.

Mexal, S., Frank, M., Berger, R., Adams, C.E., Ross, R.G., Freedman, R., & Leonard, S. (2005). Differential modulation of gene expression in the NMDA postsynaptic density of schizophrenic and control smokers. *Molecular Brain Research*, *139*, 317–332.

Myers, A.J., Gibbs, J.R., Webster, J.A., Rohrer, K., Zhao, A., Marlowe, L., … & Hardy, J. (2007). A survey of genetic human cortical gene expression. *Nature Genetics*, *39*, 1494–1499.

Naumova, O.Y., Lee, M., Rychkov, S.Y., Vlasova, N.V., & Grigorenko, E.L. (2013). Gene expression in the human brain: the current state of the study of specificity and spatiotemporal dynamics. *Child Development*, *84*, 76–88.

Nielsen, D.A., Ji, F., Yuferov, V., Ho, A., He, C., Ott, J., & Kreek, M.J. (2010). Genome-wide association study identifies genes that may contribute to risk for developing heroin addiction. *Psychiatric Genetics*, *20*, 207–214.

Noelker, C., Schwake, M., Balzer-Geldsetzer, M., Bacher, M., Popp, J., Schlegel, J., … & Dodel, R.C. (2012). Differentially expressed gene profile in the 6-hydroxy-dopamine-induced cell culture model of Parkinson's disease. *Neuroscience Letters*, *507*, 10–15.

Novak, G., Boukhadra, J., Shaikh, S.A., Kennedy, J.L., & Le Foll, B. (2009). Association of a polymorphism in the NRXN3 gene with the degree of smoking in schizophrenia: a preliminary study. *The World Journal of Biological Psychiatry*, *10*, 929–935.

Nussbaum, J., Xu, Q., Payne, T.J., Ma, J.Z., Huang, W., Gelernter, J., & Li, M.D. (2008). Significant association of the neurexin-1 gene (NRXN1) with nicotine dependence in European- and African-American smokers. *Human Molecular Genetics*, *17*, 1569–1577.

Piccolo, S.R., Sun, Y., Campbell, J.D., Lenburg, M.E., Bild, A.H., & Johnson, W.E. (2012). A single-sample microarray normalization method to facilitate personalized-medicine workflows. *Genomics*, *100*, 337–344.

Provencal, N., Suderman, M.J., Guillemin, C., Massart, R., Ruggiero, A., Wang, D., … & Szyf, M. (2012). The signature of maternal rearing in the methylome in rhesus macaque prefrontal cortex and T cells. *Journal of Neuroscience*, *32*, 15626–15642.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., … & Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, *81*, 559–575.

Ramensky, V., Bork, P., & Sunyaev, S. (2002). Human non-synonymous SNPs: server and survey. *Nucleic Acids Research*, *30*, 3894–3900.

Richards, A.L., Jones, L., Moskvina, V., Kirov, G., Gejman, P.V., Levinson, D.F., … & O'donovan, M.C. (2012). Schizophrenia susceptibility alleles are enriched for alleles that affect gene expression in adult human brain. *Mol Psychiatry*, *17*, 193–201.

Scott, L.J., Mohlke, K.L., Bonnycastle, L.L., Willer, C.J., Li, Y., Duren, W.L., … & Boehnke, M. (2007). A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science*, *316*, 1341–1345.

Shabalin, A.A. (2012). Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics*, *28*, 1353–1358.

Simmons, T.R., Flax, J.F., Azaro, M.A., Hayter, J.E., Justice, L.M., Petrill, S.A., … & Bartlett, C.W. (2010). Increasing genotype-phenotype model determinism: application to bivariate reading/language traits and epistatic interactions in language-impaired families. *Human Heredity*, *70*, 232–244.

Smyth, G.K. (2005). Limma: linear models for microarray data. In R. Gentleman, V. Carey, S. Dudoit, R. Irizarry & W. Huber (Eds.), *Bioinformatics and Computational Biology Solutions using R and Bioconductor* (pp. 397–420). New York: Springer.

Sullivan, P. F., & Kendler, K. S. (1999). The genetic epidemiology of smoking. *Nicotine & Tobacco Research*, *1* (Suppl 2), S51–S57; discussion S69-70.

The Encode Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, *489*, 57–74.

Thorgeirsson, T.E., Gudbjartsson, D.F., Surakka, I., Vink, J.M., Amin, N., Geller, F., … & Stefansson, K. (2010). Sequence variants at CHRNB3-CHRNA6 and CYP2A6 affect smoking behavior. *Nature Genetics*, *42*, 448–453.

Tobacco and Genetics Consortium (2010). Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nature Genetics*, *42*, 441–447.

Trauth, J.A., Seidler, F.J., Mccook, E.C., & Slotkin, T.A. (1999). Adolescent nicotine exposure causes persistent upregulation of nicotinic cholinergic receptors in rat brain regions. *Brain Research*, *851*, 9–19.

Trauth, J.A., Seidler, F.J., & Slotkin, T.A. (2000). An animal model of adolescent nicotine exposure: effects on gene expression and macromolecular constituents in rat brain regions. *Brain Research*, *867*, 29–39.

Tsuruoka, Y., Tsujii, J., & Ananiadou, S. (2008). FACTA: a text search engine for finding associated biomedical concepts. *Bioinformatics*, *24*, 2559–2560.

Tully, E.C., Iacono, W.G., & Mcgue, M. (2010). Changes in genetic and environmental influences on the development of nicotine dependence and major depressive disorder from middle adolescence to early adulthood. *Development and Psychopathology*, *22*, 831–848.

Ullrich, B., Ushkaryov, Y.A., & Sudhof, T.C. (1995). Cartography of neurexins: more than 1000 isoforms generated by alternative splicing and expressed in distinct subsets of neurons. *Neuron*, *14*, 497–507.

Us Surgeon General (2012). *Preventing Tobacco Use Among Youth and Young Adults*. Rockville, MD: US Department of Health and Human Services.

Vieland, V.J., Huang, Y., Seok, S.C., Burian, J., Catalyurek, U., O'connell, J., … & Valentine-Cooper, W. (2011). KELVIN: a software package for rigorous measurement of statistical evidence in human genetics. *Human Heredity*, *72*, 276–288.

Vink, J.M., Willemsen, G., & Boomsma, D.I. (2005). Heritability of smoking initiation and nicotine dependence. *Behavior Genetics*, *35*, 397–406.

Vrieze, S.I., Mcgue, M., & Iacono, W.G. (2012). The interplay of genes and adolescent development in substance use disorders: leveraging findings from GWAS meta-analyses to test developmental hypotheses about nicotine consumption. *Human Genetics*, *131*, 791–801.

Webster, J.A., Gibbs, J.R., Clarke, J., Ray, M., Zhang, W., Holmans, P., … & Myers, A.J. (2009). Genetic control of human brain transcript expression in Alzheimer disease. *The American Journal of Human Genetics*, *84*, 445–458.

Whitlock, M.C. (2005). Combining probability from independent tests: the weighted Z-method is superior to Fisher's approach. *Journal of Evolutionary Biology*, *18*, 1368–1373.

Yoon, D., Kim, Y.J., Cui, W.Y., Van Der Vaart, A., Cho, Y.S., Lee, J.Y., … & Park, T. (2012). Large-scale genome-wide association study of Asian population reveals genetic factors in FRMD4A and other loci influencing smoking initiation and nicotine dependence. *Human Genetics*, *131*, 1009–1021.

Accepted for publication: 4 June 2013
Published online: 27 August 2013