European Journal of Human Genetics (2011) 19, 76–83

www.nature.com/ejhg

ARTICLE

Nicotinic α5 receptor subunit mRNA expression is associated with distant 5′ upstream polymorphisms

Ryan M Smith[1], Houda Alachkar[2], Audrey C Papp[3], Danxin Wang[3], Deborah C Mash[4], Jen-Chyong Wang[5], Laura J Bierut[5] and Wolfgang Sadee*[,3]

CHRNA5*, encoding the nicotinic α5 subunit, is implicated in multiple disorders, including nicotine addiction and lung cancer. Previous studies demonstrate significant associations between promoter polymorphisms and *CHRNA5* mRNA expression, but the responsible sequence variants remain uncertain. To search for *cis*-regulatory variants, we measured allele-specific mRNA expression of *CHRNA5* in human prefrontal cortex autopsy tissues and scanned the *CHRNA5* locus for regulatory variants. A cluster of six frequent single-nucleotide polymorphisms (rs1979905, rs1979906, rs1979907, rs880395, rs905740, and rs7164030), in complete linkage disequilibrium (LD), fully account for a >2.5-fold allelic expression difference and a fourfold increase in overall *CHRNA5* mRNA expression. This proposed enhancer region resides more than 13 kilobases upstream of the *CHRNA5* transcription start site. The same upstream variants failed to affect *CHRNA5* mRNA expression in peripheral blood lymphocytes, indicating tissue-specific gene regulation. Other promoter polymorphisms were also correlated with overall *CHRNA5* mRNA expression in the brain, but were inconsistent with allelic mRNA expression ratios, a robust and proximate measure of *cis*-regulatory variants. The enhancer region and the nonsynonymous polymorphism rs16969968 generate three main haplotypes that alter the risk of developing nicotine dependence. Ethnic differences in LD across the *CHRNA5* locus require consideration of upstream enhancer variants when testing clinical associations.
European Journal of Human Genetics (2011) **19,** 76–83; doi:10.1038/ejhg.2010.120; published online 11 August 2010

Keywords: nicotinic receptor; α5 subunit; gene expression; nicotine dependence; lung cancer; enhancer

INTRODUCTION

Genetic surveys link the nicotinic receptor subunit α5 gene (*CHRNA5*) to smoking,[1–11] lung cancer,[10,12–14] chronic obstructive pulmonary disease (COPD),[15] cocaine addiction,[6,16] and alcohol dependence.[6,17,18] In Europeans, *CHRNA5* harbors only one frequent nonsynonymous single-nucleotide polymorphism (SNP), rs16969968 in exon 5 (minor allele frequency ~40%), resulting in an aspartic acid to asparagine residue change (D398N). This variant has been associated with risk for nicotine dependence, lung cancer, and COPD.[2–15] *In vitro*, rs16969968 reduces the maximal response of the receptor to epibatidine,[2] whereas response to nicotine remains to be determined.

Expression of α5 mRNA appears to be modulated by genetic factors. To uncover the expression quantitative trait locus (eQTL), two previous studies correlated mRNA expression in postmortem frontal cortex tissue to *CHRNA5* polymorphisms, yielding significant associations with a 22-bp deletion (rs3841324) located less than 100 bases upstream of the *CHRNA5* transcription start[10,18] as a potential promoter variant. Homozygous carriers of the variant short allele expressed 2- to 2.9-fold more *CHRNA5* mRNA *versus* the wild-type long allele in the human prefrontal cortex.[10,18] However, the association between rs3841324 and mRNA expression levels in the brain is less than definitive, because numerous other SNPs are in partial to strong linkage disequilibrium (LD) with rs3841324, and total mRNA levels used to determine eQTLs are affected by multiple *trans*-acting factors and variants in *cis*-regulatory regions.

In this study we measured allelic mRNA expression of *CHRNA5* in human brain autopsy tissues to mitigate *trans*-acting factors and distinguish between functional and marker SNPs embedded in a haplotype block with high but incomplete LD. Different mRNA expression between two alleles in the same tissue, termed allelic expression imbalance (AEI), is the most immediate and quantitative phenotype that one can use to scan a gene locus for the presence of *cis*-regulatory variants[19,20] or *cis*-eQTLs,[21] with greater confidence than is possible with the use of total mRNA levels. This approach revealed a distal enhancer region more than 13 kb upstream of *CHRNA5*, harboring six SNPs in complete LD with each other in European Americans and African Americans, but in lower LD with other SNPs previously suspected of affecting expression. The minor alleles of these proposed enhancer SNPs correspond to a 4.1-fold increase in overall *CHRNA5* mRNA expression. In light of these findings, we discuss genetically variable *CHRNA5* expression in disease associations and assess risk for nicotine addiction through the use of previous large-scale association studies.[3–5]

MATERIALS AND METHODS

Nucleic acid isolation

Genomic DNA (gDNA) was isolated from 91 prefrontal cortex tissues (provided by Dr Deborah Mash) originating from Brodmann Area 46, using a 'salting out' method,[22] supplemented with additional sodium dodecyl sulfate for lipid-rich brain tissue. Total RNA was isolated with RNeasy Mini Kit spin

[1]Neuroscience Graduate Studies Program, Department of Pharmacology, The Ohio State University, Columbus, OH, USA; [2]Integrated Biomedical Science Graduate Program, Comprehensive Cancer Center, The Ohio State University, Columbus, OH, USA; [3]Department of Pharmacology, Program in Pharmacogenomics, The Ohio State University, Columbus, OH, USA; [4]Department of Neurology, Miller School of Medicine, University of Miami, Miami, FL, USA; [5]Department of Psychiatry, Washington University, St Louis, MO, USA
*Correspondence: Dr W Sadee, Department of Pharmacology, Program in Pharmacogenomics, The Ohio State University, 5078 Graves Hall, 333 W. 10th Ave., Columbus, OH 43210, USA. Tel: +1 614 292 6673; Fax: +1 614 292 1957; E-mail: wolfgang.sadee@osumc.edu
Received 23 February 2010; revised 1 June 2010; accepted 4 June 2010; published online 11 August 2010

columns (Qiagen, Germantown, MD, USA) from brain homogenates dissolved in TRIzol (Life Technologies, Carlsbad, CA, USA). RNA preparations were treated with DNase to remove any remaining gDNA fragments. Two separate complementary DNA (cDNA) preparations were obtained using 0.5 μg total RNA per sample preparation. Oligo-dT was used to prime mRNA in the reverse transcription reaction. Similarly, gDNA and RNA from 87 patients were isolated from lymphocytes originating from the CEPH Coriell cell collection (Coriell Institute, Camden, NJ, USA).

Brain and lymphocyte sample genotyping

SNPs rs16969968 and rs615740 used as markers for measuring AEI were genotyped by restriction fragment length polymorphism methods, using primers tagged with a fluorophore (6-FAM or HEX). Taqα1 (rs16969968) or CviQI (rs615470) (from New England Biolabs, Ipswich, MA, USA) were used to cut the amplified gDNA wild-type alleles and were analyzed on an ABI 3730 DNA Analyzer (Life Technologies). restriction fragment length polymorphism was also used to genotype rs17486195 (NsiI), rs680244 (BsmAI), rs17408276 (CviQI), and rs578776 (NlaIII). The deletion polymorphism rs3841324 was genotyped using PCR with fluorescent primers and analyzed on the ABI 3730. A SNaPshot Multiplex System (Life Technologies) was used to genotype additional SNPs rs1979907, rs880395, rs905740, rs7164030, rs2036527, rs667282, rs621849, rs692780, and rs514743, using previously described methods.[21]

Allelic mRNA expression imbalance in human brain tissue

Allele-specific mRNA expression was quantified in samples heterozygous for either or both marker SNPs (rs16969968 and rs615470) using SNaPshot (a fluorescent primer extension method), as described elsewhere.[23] Primer sequences for quantitative measures (AEI and quantitative PCR (qPCR)) are listed in Supplementary Table S1. Because of low *CHRNA5* expression, 12.5 ng cDNA/25 ng gDNA was first preamplified over 10 PCR cycles using tailed primers targeting the exon containing the marker SNP, enabling equal application to both cDNA and gDNA. Subsequently, 1 μl of preamplified cDNA/gDNA was PCR amplified over 30 cycles using primers targeting the tails attached to the preamplification primers. Following the SNaPshot primer extension step and separation on an ABI 3730, area under the curve was determined using GeneMapper 4.0 software (Life Technologies) to calculate relative allelic expression ratios. For each SNP, samples were measured in triplicate for each of two independent cDNA syntheses, whereas gDNA was measured in duplicate. Allelic ratios for cDNA were normalized against the overall average ratio obtained for gDNA for each marker SNP (ancestral/variant allele). The percentage change from gDNA ratio was $\log_{10}$ transformed and used for data analysis.

Sequencing

gDNA (25 ng) from two tissues was amplified with Phusion High-Fidelity DNA Polymerase (New England Biolabs) using primers targeting 13 loci within and around the *CHRNA5* gene (Table 2). The PCR product was gel-purified using a Qiaquick Gel Extraction Kit (Qiagen) or a GenElute Gel Extraction Kit (Sigma-Aldrich, St Louis, MO, USA) and then sequenced by the Ohio State University Plant-Microbe Genomics Facility using an ABI 3730 DNA Analyzer (Life Technologies). Sequence data were aligned using the Basic Local Alignment Search Tool – bl2seq (http://blast.ncbi.nlm.nih.gov/Blast.cgi) – and compared against the reference sequence (NCBI Build 36.1 on the UCSC Genome Browser:[24] http://genome.ucsc.edu).

Quantitative real-time PCR (qRT–PCR)

CHRNA5 mRNA expression was measured in triplicate by qPCR on an ABI 7000 Sequence Detection System (Life Technologies). cDNA of 12.5 ng was amplified in a 15 μl reaction using Power SYBR Green PCR Master Mix (Life Technologies). Primers for *CHRNA5* measurements targeted the same region as used for SNaPshot analysis of rs16969968. The relative quantity of *CHRNA5* mRNA, normalized within each sample to *ACTB* expression (measured in duplicate), was used for data analysis. *CHRNB4* mRNA was measured by a single qPCR reaction performed for each sample, also normalized to *ACTB*.

Clinical association design and sample genotyping

Nicotine-dependent and nondependent smoking subjects were recruited by the multisite Collaborative Genetic Study of Nicotine Dependence project. For this study, only rs880395 and rs16969968 were considered for association analysis, and a total of 2050 subjects of European descent had complete genotyping data at both SNPs (1055 nicotine-dependent cases and 995 nondependent smoking controls). This cohort is described in previous genetic association studies on nicotine dependence, which also include more detailed genotyping methods.[3–5]

Data analysis

Samples with allelic expression ratios >2 were considered as potentially harboring functional *cis*-acting polymorphisms, with a 2.5-fold allelic difference chosen as a stringent cutoff. HelixTree software (Golden Helix, Bozeman, MT, USA) was used for testing genotype associations with allelic expression (available for tissues heterozygous for the two marker SNPs), and for performing linear regression analysis of genotype on overall *CHRNA5* mRNA expression (available for all tissues). Univariate analysis of variance comparing *CHRNA5* mRNA expression across demographic categories, with age and postmortem interval as covariates, and subsequent correlations and one-way analysis of varience comparing *CHRNA5* and *CHRNB4* expression across rs905740 genotype for brain and lymphocyte samples, was performed with SPSS *v*17.0 (SPSS, Chicago, IL, USA). SAS *v*9.1 (SAS Institute, Cary, NC, USA) was used for logistic regression analyses and to calculate odds ratios for nicotine-dependence risk according to genotype.

RESULTS

Allelic expression imbalance analysis

Table 1 shows sample demographics for postmortem prefrontal cortex (Brodmann area 46) samples used in this study, obtained from a population of cocaine abusers and controls. Drug group, race, sex, smoking history, age, or postmortem interval had no effect on overall *CHRNA5* mRNA expression (analysis of varience; Table 1). A total of 43 of 91 postmortem brain samples were heterozygous for one or both marker SNPs, with locations shown in Figure 1, yielding reproducible allelic mRNA measurements for analysis. Within each assay, standard deviations for gDNA ratios, normalized to 1 for subsequent comparisons with allelic mRNA ratios, varied by 8.9–9.8% of the mean for both marker SNPs. One sample showing an allelic gDNA ratio greater than two SD below the mean across multiple replicates was excluded from further analysis because of potential copy number variation. Because of the low abundance of *CHRNA5* transcripts, the within-sample standard deviation of allelic mRNA ratios was higher than that for gDNA ratios (SD=26% within-subject variation for rs16969968, and 47% for rs615470). Despite the relatively high error observed for mRNA measurements, allelic mRNA ratios were consistent between two separate cDNA syntheses (correlation coefficient, r=0.62) and between the two marker SNPs in subjects heterozygous for both (Figure 2; r=0.73). Given the large allelic mRNA expression imbalance (AEI) in a number of subjects, with ratios exceeding a fourfold difference between alleles in many samples, the S.D. of the allelic mRNA ratio assay was acceptable for detecting large AEI. We selected a conservative cutoff of allelic mRNA ratios >2.5 to indicate the presence of significant AEI, detectable with high confidence, and therefore cannot exclude the possibility that smaller allelic differences (ratios <2.5) are also present that escape reliable detection using this approach.

Considering AEI ratios with >2.5-fold difference, 27 of the 42 analyzable samples showed significant AEI (Figure 3). Of the 22 samples heterozygous for rs16969968, 11 showed significant AEI, whereas 25 of 31 samples heterozygous for rs615470 were AEI positive. All significant AEI ratios showed greater expression for the ancestral allele versus the variant allele of the respective marker SNPs, suggesting the presence of a functional polymorphism that is in high LD with the

Table 1 Demographics for prefrontal cortex samples

Demographic	*Group* (n)	*CHRNA5 Expression* (C_T *Avg.* ± *St. Dev.*)	*ANOVA* (P)	*No. samples in AEI analysis*
Sex	Male (62)	13.4 ± 1.5	0.597	29 (18 AEI positive)
	Female (29)	13.8 ± 1.4		13 (9)
Race	Caucasian (59)	13.6 ± 1.3	0.904	28 (17)
	African American (14)	13.2 ± 1.6		4 (2)
	Mixed/other (18)	13.7 ± 2.0		10 (8)
Drug Group	Cocaine (39)	13.5 ± 1.6	0.809	21 (15)
	Control (52)	13.6 ± 1.4		21 (12)
Smoking History	Smoker (35)	13.7 ± 1.6	0.187	12 (6)
	Non-smoker (56)	13.4 ± 1.4		30 (21)
Age[a]			0.224	
Postmortem interval[a]			0.828	

Abboreviation: AEI, allelic expression imbalance.
[a]Considered as a covariate variable.

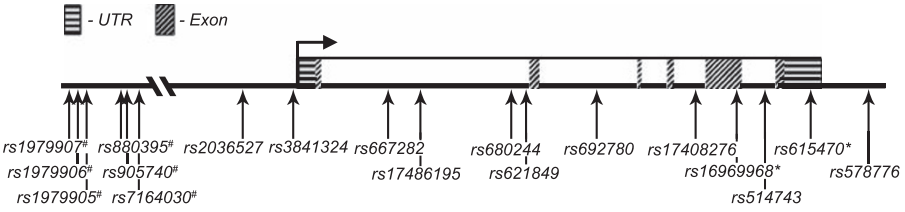

Figure 1 Polymorphisms genotyped in the *CHRNA5* locus, including SNPs used to establish haplotype structure, marker SNPs (*), and putative functional enhancer SNPs (#).

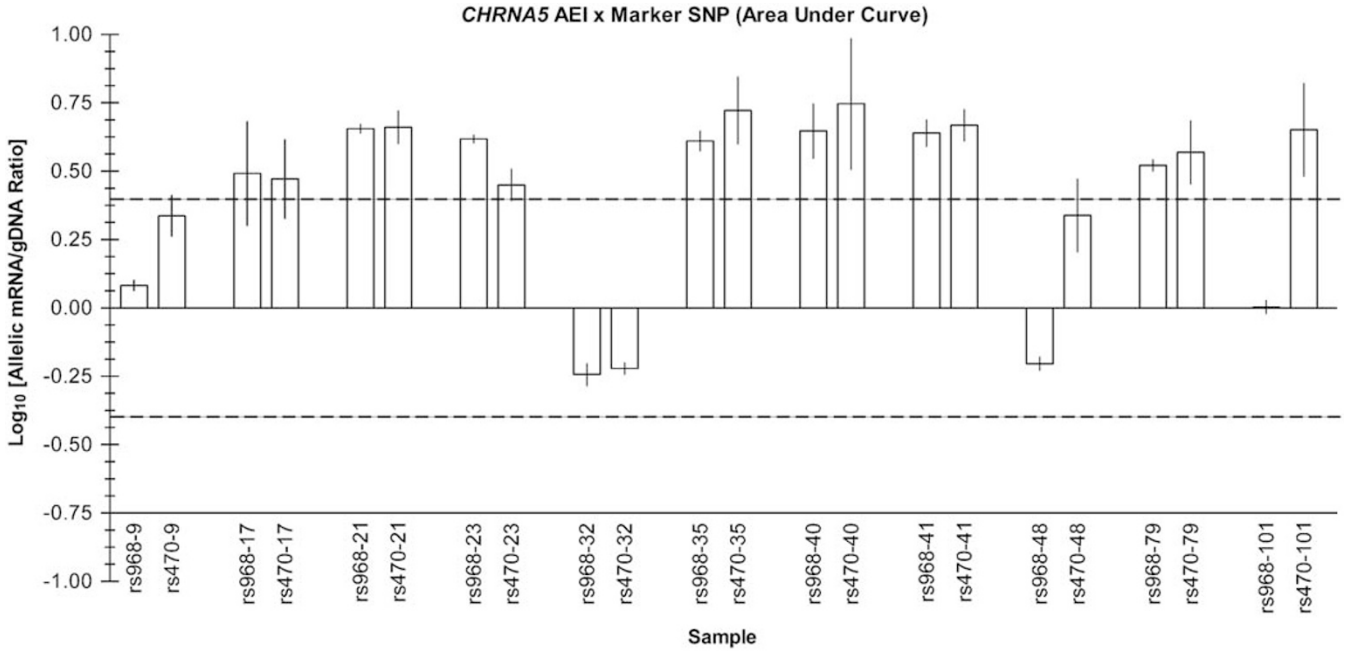

Figure 2 AEI ratios in brain tissues heterozygous for both marker SNPs rs16969968 (rs968) and rs615470 (rs470). In all, 9 of 10 samples display consistent AEI across the two marker SNPs, whereas one sample (101) shows differences in AEI, a possible indicator of RNA processing leading to different mRNA isoforms (eg, alternative splicing), or analysis error in this sample. *Note:* Solid line at 0 indicates no difference between gDNA and mRNA ratios; dotted lines indicate a 2.5-fold change in allelic mRNA expression (normalized to gDNA), taken here as strong evidence for AEI; error bars=SEM.

two marker SNPs, as equilibrated SNPs affecting allelic expression would cause AEI ratios above and below unity. The marker SNPs themselves are unlikely to account for the AEI ratios, as several samples were heterozygous at these SNP locations but lacked AEI. In a separate study of prefrontal cortex autopsy tissues obtained from the Buckeye Brain Bank at The Ohio State University, a similar frequency of large AEI ratios was observed (data not shown), supporting the accuracy of these results.

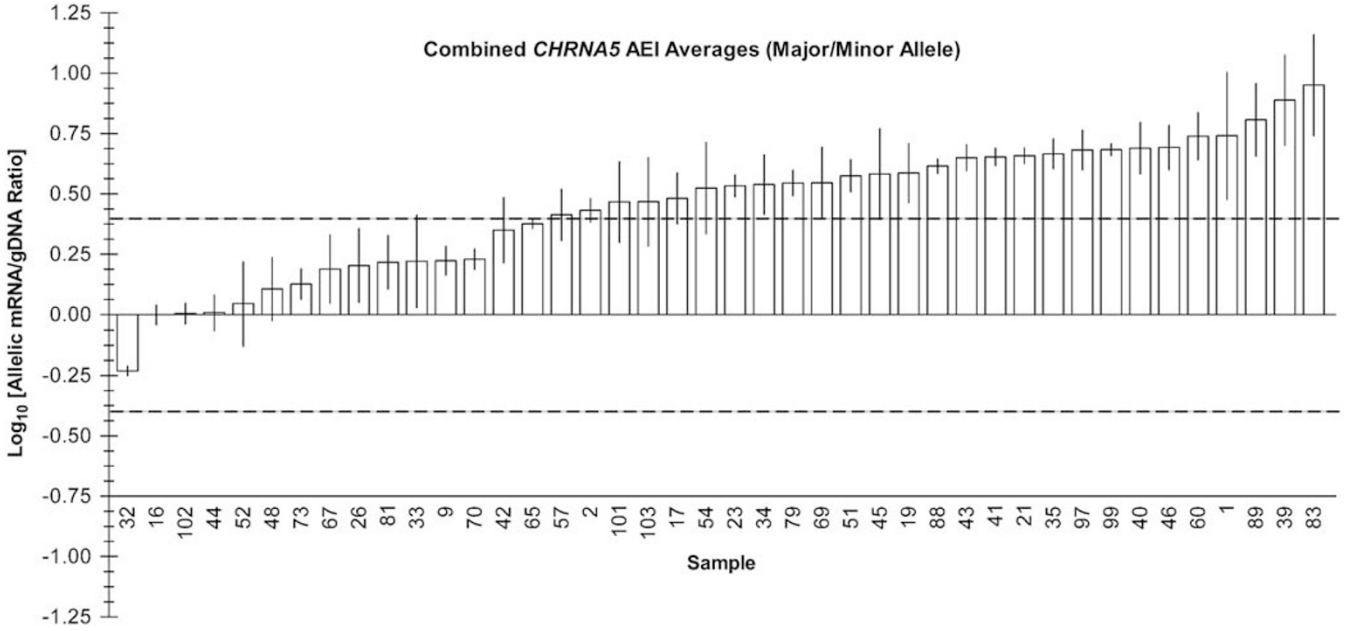

Figure 3 Combined rs16969968 and rs615470 allelic mRNA expression ratios for 42 prefrontal cortex tissues. A total of 27 of 42 samples display >2.5-fold AEI (represented by dotted lines). Bars represent mean AEI, with error bars indicating SEM.

Table 2 SNP association with overall *CHRNA5* expression and AEI, and linkage disequilibrium (LD) with rs905740

SNP rs#	*Chromosome position*[a]	*Samples genotyped (n)*	*Minor allele (Freq)*[b]	*Regression for mRNA expression (P)*	*% Samples meeting AEI*[c]	*AEI genotype test (P)*	*LD with rs905740 D′ (r^2)*
rs1979907[d]	78842239	86	A (0.34)	1.13E-04	100	5.37E-12	1.00 (1.00)
rs880395[d]	78844356	87	A (0.33)	2.50E-05	100	5.37E-12	1.00 (1.00)
rs905740[d]	78844386	88	T (0.33)	2.42E-05	100	5.37E-12	—
rs7164030[d]	78844661	88	G (0.33)	2.42E-05	100	5.37E-12	1.00 (1.00)
rs2036527	78851615	90	A (0.35)	5.75E-04	55	6.09E-01	1.00 (0.27)
rs3841324[e]	78857813:78857834	89	S (0.35)	1.81E-03	86	4.48E-03	0.83 (0.64)
rs667282	78863472	87	C (0.29)	8.95E-01	30	7.13E-02	0.54 (0.06)
rs17486195	78865197	88	G (0.30)	5.92E-03	55	8.29E-01	0.89 (0.17)
rs680244	78871288	91	T (0.38)	2.71E-05	79	2.73E-03	0.81 (0.52)
rs621849	78872861	90	G (0.37)	3.45E-06	83	7.16E-04	0.84 (0.57)
rs692780	78876505	91	G (0.30)	6.75E-05	77	6.77E-03	0.83 (0.61)
rs17408276	78881618	88	C (0.27)	1.38E-02	77	2.69E-02	0.83 (0.53)
rs16969968[f]	78882925	87	A (0.30)	1.46E-02	58	5.83E-01	0.89 (0.17)
rs514743	78884227	91	A (0.34)	1.21E-03	79	1.11E-02	0.71 (0.49)
rs615470	78885988	89	T (0.37)	4.76E-03	76	5.65E-02	0.71 (0.42)
rs578776	78888400	91	A (0.35)	9.40E-01	29	3.55E-03	0.66 (0.11)

Abbreviations: AEI, allelic expression imbalance; LD, linkage disequilibrium.
[a]Position according to Genome Reference Consortium build 37.1 (GRCh37).
[b]Allele frequency calculated in our sample population.
[c]To meet criterion sample must be heterozygous in AEI positive samples, homozygous in AEI negative samples.
[d]proposed enhancer SNPs.
[e]22-bp deletion variant.
[f]nonsynonymous SNP.

CHRNA5 SNP scanning and sequencing

We performed SNP scanning across the gene locus to identify *cis*-eQTL regions associated with AEI, as an indication of regions harboring functional regulatory variants. Guided by strong blocks of LD in *CHRNA5*, samples were first genotyped for 10 variants along the length of the gene (Figure 1) in addition to the two marker SNPs. Several SNPs were significantly associated with AEI ratios and overall expression, as shown in Table 2, but no single variant fully accounted for the observed AEI, as these variants were either homozygous when AEI was present or heterozygous when it was not, or both, in at least some subjects. To identify functional polymorphisms, two samples sharing a similar haplotype (concordant genotypes at 11 of 12 polymorphic sites) but differing in AEI phenotype (one AEI positive, one negative) were sequenced at 13 loci within and surrounding

Table 3 Loci sequenced in *CHRNA5* region

Chromosome position[a]	*39 HapMap SNPs in region (minor allele frequency in CEU Samples)*
78844131–78845361	rs880395 (0.367), rs905740 (0.367), rs7164030 (0.367)
78855228–78856482	
78857582–78859073	rs3841324, rs684513 (0.217)
78861175–78861676	rs7165657 (0.000), rs7166003 (0.000)
78862083–78863528	rs7178897 (0.000), rs667282 (0.208)
78864617–78865964	rs479385 (0.000), rs16969948 (0.000), rs17486195 (0.341), rs588765 (0.339), rs6495306 (0.375)
78866874–78868280	rs16969949 (0.000), rs12903839 (0.034), rs17486278 (0.409), rs1875869 (0.009)
78869168–78870991	rs601079 (0.367), rs495956 (0.333)
78872740–78874145	rs621849 (0.375), rs569207 (0.216), rs637137 (0.202), rs7180002 (0.417)
78877363–78878900	rs481134 (0.373), rs951266 (0.424), rs10519205 (0.000)
78880355–78881722	rs647041 (0.373), rs12898919 (0.133), rs12903575 (0.042), rs17408276 (0.325)
78882772–78884589	rs16969968 (0.425), rs518425 (0.250), rs514743 (0.331)
78887407–78888524	rs12899226 (0.042), rs660652 (0.325), rs472054 (0.328), rs8029939 (0.000), rs578776 (0.242)

Abbreviation: SNP, single-nucleotide polymorphism; CEU, CEPH (Utah Residents with Northern and Western European Ancestry).
[a]Position according to Genome Reference Consortium build 37.1 (GRCh37).

CHRNA5 (Table 3). Polymorphisms heterozygous in the AEI-positive sample but homozygous in the AEI-negative sample were considered candidates for further study. Of 39 polymorphic sites with appreciable frequency in CEU (CEPH (Utah Residents with Northern and Western Ancestry)) or YRI HapMap populations (Table 3), only three SNPs matched these criteria. These three SNPs, rs880395, rs905740, and rs7164030, reside in a span of 300 nucleotides ~13.5 kb upstream of the *CHRNA5* transcription start site (Figure 1), and their minor allele frequency varies by race (CEU=0.37, YRI=0.18). An additional 12 samples were sequenced at the upstream locus (Table 3; chromosomal position 78844131–78845361), confirming heterozygosity in AEI-positive samples and homozygosity in AEI-negative samples for all three SNPs. Finally, the three SNPs were genotyped in all brain tissues, showing them to be in complete LD with each other and accounting fully for all AEI data. This finding is valid across racial groups. As for all other measured SNP genotypes, at least one tissue was homozygous in AEI-positive or heterozygous in AEI-negative samples in each group. As predicted, the putative functional SNPs were in high LD with the marker SNPs used to measure AEI in our samples (D'=0.89 for rs16969968, 0.70 for rs615470; Supplementary Figure S1). The SNP Annotation and Proxy Search website (http://www.broadinstitute.org/mpg/snap/)[25] was used to find additional SNPs predicted to be in complete LD with rs880395, rs905740, or rs7164030 in both CEU and YRI HapMap populations, yielding three additional SNPs (rs1979905, rs1979906, and rs1979907) located 2 kb farther upstream of *CHRNA5*. Genotyping rs1979907 as a surrogate for all three additional SNPs confirmed complete LD between all of the upstream SNPs in our samples (Table 2). Furthermore, all samples with a >2.5-fold change in AEI were heterozygous at these four measured SNP locations, whereas all samples with <2.5-fold AEI were homozygous, designating one or more of these six SNPs as the likely functional variant(s); however, we cannot exclude the possibility of additional SNPs in complete LD in more distal regions. Genotype association with AEI ratios for these SNPs reached strong significance ($P=5.37\times10^{-12}$, Table 2), whereas other SNPs previously suspected of being responsible for altered expression, in particular rs3841324, failed to fully account for high AEI ratios and displayed lower association with AEI (P=0.00448, Table 2). Of the 42 subjects analyzed, three showed significant AEI but were homozygous for rs3841324, whereas an additional three were heterozygous and lacked AEI.

CHRNA5 mRNA expression in brain and lymphocytes

Linear regression was performed for the 16 polymorphisms genotyped in all 91 samples to uncover possible eQTLs, which are also subject to *trans*-acting factors, in contrast to the *cis*-eQTLs detected by AEI analysis. The results support a strong association between putative functional variants rs1979907, rs880395, rs905740, and rs7164030 and total brain *CHRNA5* mRNA expression ($P=2.42\times10^{-5}$, Table 2). These upstream SNPs scored better than the 22-bp deletion rs3841324 that was previously associated with mRNA expression ($P=1.81\times10^{-3}$), consistent with the AEI results. An intronic SNP, rs621849, was also highly correlated with total mRNA levels ($P=3.45\times10^{-6}$) but to a lesser extent with AEI ($P=7.16\times10^{-4}$ *versus* $P=5.37\times10^{-12}$). The high eQTL value for rs621849 could have been a fortuitous result of *trans*-acting factors in the tissues analyzed, or rs621849 could have been in partial LD with other regulatory variants with a relatively low impact on mRNA expression. Also significantly associated with expression, rs3841324 and rs621849 are in substantial LD with the six upstream SNPs in our samples (Table 2, Supplementary Figure S1), thereby at least partially accounting for their significant associations with expression. Although we cannot discount the possibility that they have smaller effects (<2.5-fold) on mRNA expression, examining regression and LD within each racial cohort argues against a functional role for rs3841324. In African Americans, there is no significant relationship between mRNA expression and rs3841324 genotype (P=0.672; Supplementary Table S2), whereas LD with the upstream 'enhancer' SNPs decreases accordingly (r^2=0.12; Supplementary Table S2). Similarly, race was not a significant factor when included as a covariate in the regression analysis (Supplementary Table S2), suggesting the fact that upstream SNPs dominate the genetic effect on mRNA expression.

A comparison of expression in brain samples across the rs905740 genotype (Figure 4a) revealed a significant 4.1-fold increase in *CHRNA5* expression for homozygous carriers of the minor SNP 'T' allele compared with homozygous carriers of the major wild-type 'C' allele (overall analysis of varience F=9.97, P=0.0001), with heterozygotes in between. As rs3841324 is in substantial LD with the upstream enhancer SNPs identified here, we find a lesser but still significant 2.9-fold increase in expression for carriers of the deletion, identical to previously reported changes in expression that correlated with this variant.[8] Nevertheless, these results indicate that rs3841324

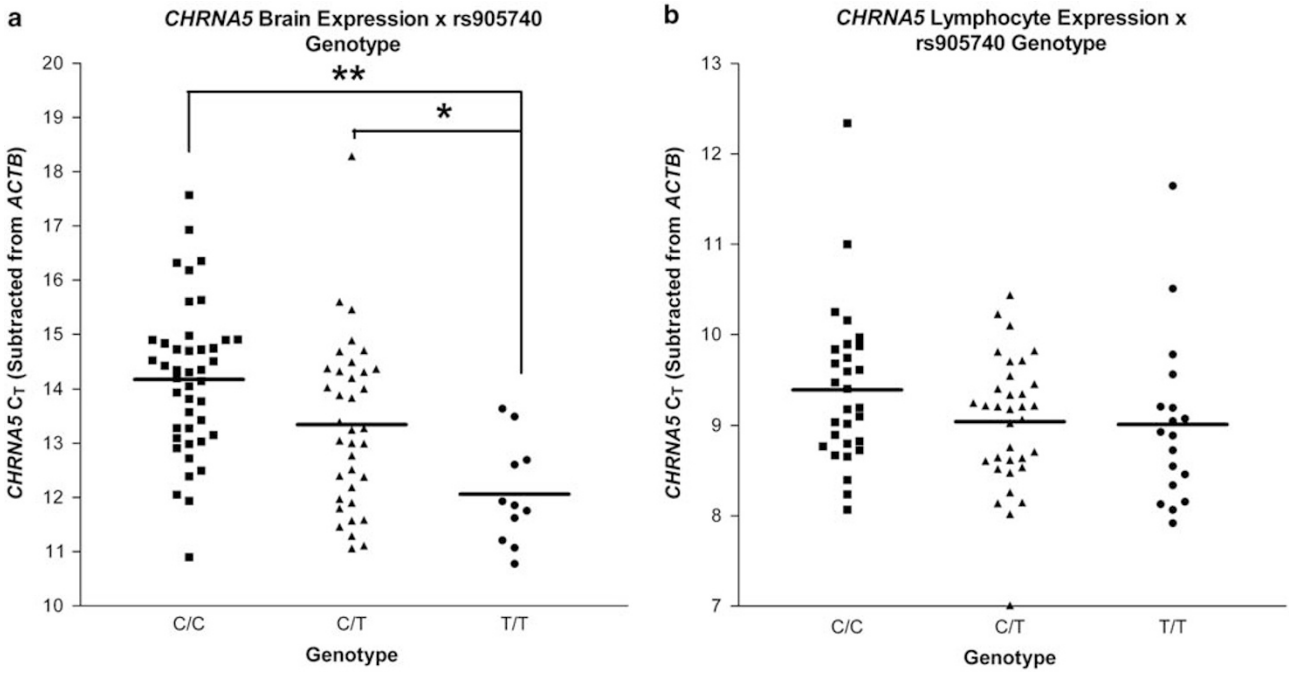

Figure 4 (**a**) One-way analysis of varience shows a significant effect of rs905740 genotype on total*CHRNA5* mRNA expression in prefrontal cortex tissue ($P<0.001$). Bonferroni-corrected *post hoc* analysis shows a significant increase in T/T expression (indicated by lower CT values) compared with either C/C or C/T genotype. $^{**}P<0.001$; $^{*}P<0.05$. CT: RT-PCR cycle threshold, with each unit corresponding to a twofold change. (**b**) *CHRNA5* mRNA expression in lymphocytes when compared across the rs905740 genotype. No significant difference is observed. CT: RT-PCR cycle threshold.

can only be considered as a surrogate marker for upstream enhancer SNPs, and only in European Americans.

To test for tissue-specific regulation, *CHRNA5* mRNA expression was measured in human peripheral lymphocytes, showing that the rs905740 genotype had no detectable effect ($F=1.83$, $P=0.17$; Figure 4b). We further considered whether the distal enhancer SNPs affect expression of the adjacent nicotine receptor subunit *CHRNB4*, because tissue-specific expression of *CHRNA5* and *CHRNB4* is correlated according to microarray data from the Genomics Institute of the Novartis Research Foundation[26] (http://symatlas.gnf.org) ($r=0.76$ for *CHRNA5* reporter 206533_at and *CHRNB4* reporter 207516_at). Using RT-PCR, we confirmed a correlation in mRNA expression in our brain samples ($r=0.58$); however, *CHRNB4* mRNA expression did not correlate with the rs905740 genotype ($F=0.55$, $P=0.58$).

Enhancer SNP association with nicotine dependence

Genotyping results generated by previous studies were used to test associations of the upstream enhancer region to nicotine dependence in European Americans recruited by the Collaborative Genetic Study of Nicotine Dependence.[3–5] Among the upstream enhancer SNPs, only rs880395 was included in the genotyping arrays, and, when analyzed alone, failed to show any significant associations with nicotine dependence, whereas the nonsynonymous rs16969968 variant conferred significant risk[2,3,5] (Table 4). A joint SNP analysis of rs880395 and rs16969968 finds a modest but significant contribution of rs880395 to nicotine dependence in European Americans (odds ratio (O.R.)=1.22, confidence interval (C.I.): 1.04–1.44, $P=0.01$, Table 4), whereas the nonsynonymous SNP rs16969968 remains robustly associated (O.R.=1.57, C.I.: 1.33–1.86, $P=1.2\times10^{-7}$, Table 4). The distribution of rs16969968 and rs880395 is nonrandom. The risk allele for rs16969968 (A genotype) occurs almost exclusively on the low expression main (nonrisk) G allele of rs880395 (Supplementary Table S3).

Table 4 Results from single-SNP and Joint-Effect Association Models testing for Nicotine Dependence in COGEND

	Risk allele	N	*OR (95% CI)*	P
Single SNP analysis[a]				
rs880395	A	2050	0.93 (0.82–1.06)	2.71E-01
rs16969968	A	2050	1.39 (1.21–1.58)	1.27E-06
Joint SNP analysis[b]				
rs880395	A	2050	1.22 (1.04–1.44)	1.58E-02
rs16969968	A	2050	1.57 (1.33–1.86)	1.24E-07

Abbreviations: COGEND, Collaborative Genetic Study of Nicotine Dependence; SNP, single-nucleotide polymorphism.
[a]Logistic regression model per SNP in which case=SNP+sex+age.
[b]Logistic regression model for both SNPs in which case=SNP_1+SNP_2+sex+age.

DISCUSSION

We have identified six *CHRNA5* SNPs (rs1979905, rs1979906, rs1979907, rs880395, rs905740, and rs7164030) in complete LD located in a region more than 13 kb upstream of the transcription site that are robustly associated with mRNA expression. Quantitative analysis in prefrontal cortex tissues revealed a fourfold increase in *CHRNA5* mRNA expression for homozygous carriers of the minor allele enhancer SNPs, supporting the hypothesis that the six upstream SNPs are located in a highly penetrant enhancer region. Given the level of precision attainable for measuring allelic mRNA ratios of transcripts with low expression, we cannot address here whether additional regulatory polymorphisms exist that have lesser effects on expression.

Evolutionary changes in gene expression across species commonly result from polymorphisms in *cis*-regulatory elements that create or abolish *cis*-acting transcription/enhancer/suppressor factor binding

sites.[27] The accumulation of the proposed enhancer SNPs to very high allele frequencies (37% in populations of European descent), resulting in gain of function compared with ancestral alleles, together with their position within a large haplotype block, suggests that these enhancer SNPs have undergone positive selection in the human lineage. Analysis of transcription factor binding sites created or abolished by the enhancer SNPs in this region using JASPAR[28,29] (http://jaspar.cgb.ki.se/;) suggests that rs905740 is a leading functional candidate: the variant 'T' allele increases the likelihood of Ets-1 binding, which has been shown to interact with the transcriptional coactivator EA1 binding protein p300 to regulate transcription at distal enhancer sites.[30,31] Significant upregulation of both Ets-1 and *CHRNA5* in lung cancer[32–35] provides a disease model in which the relationship between enhancer SNP rs905740 genotype and Ets-1 expression can be tested.

Approximately 50% of tissue-specific distal enhancers are found within 20 kb of the nearest transcription start site, but can also occur >90 kb up- or downstream of the genes they regulate.[31] The pattern of multiple transcription factor binding sites around a given gene is proposed to determine tissue-selective enhancers.[36] Therefore, we tested the effect of rs905740 genotype on *CHRNA5* mRNA expression in human lymphocytes, in which it is also expressed. Lack of association between the rs905740 genotype and expression in lymphocytes is consistent with this region serving as a tissue-specific distal enhancer. The activity of the enhancer in other tissues remains to be determined, but cortical and subcortical structures of the forebrain seem to show variable allelic expression (unpublished observations). It is important to determine whether the proposed enhancer region regulates *CHRNA5* expression in brain regions crucial for addiction, or whether *CHRNA5* expression in lung tissues associates with lung cancer independent of nicotine addiction.

Association of the enhancer region with clinical traits

Previous clinical association studies on nicotine dependence have failed to reveal a role for the enhancer region alone on phenotype when using marker SNP rs880395.[2,4,5] However, the results of this study strongly implicate rs880395, and those adjacent SNPs in complete LD, as strong drivers of *CHRNA5* expression. This prompted a reevaluation of rs880395 in nicotine dependence in conjunction with other functional SNPs. A joint analysis of rs880395 and rs16969968 revealed a modest but significant association of rs880395 with nicotine dependence. This is similar to a previous analysis examining the promoter variant rs3841324 (22 bp deletion) with rs16969968.[10] Both studies confirm that the minor allele of the nonsynonymous variant rs16969968 (*A/A* genotype) is a risk factor when occurring on the low-expressing background, that is, the ancestral allele for the proposed enhancer SNPs, whereas rs880395 and SNPs in high LD located in this proposed upstream enhancer region now appears to emerge as another independent risk factor.

It may, at first glance, be surprising that the joint rs880395/rs16969968 analysis did not yield stronger associations with nicotine addiction compared with the rs3841324/rs16969968 analysis,[10] when considering the fact that rs3841324 does not account for allelic expression differences. However, rs880395 and rs3841324 genotypes are highly correlated in populations of European descent (Table 2). Examining ethnic groups in which the correlation between these two variants is not as strong might magnify the difference in risk potential for nicotine dependence. Given the large effect of the enhancer region on mRNA expression, future studies should consider the haplotype on which the risk polymorphism resides as an important factor in penetrance.

Our data cannot rule out the possibility that alternative splicing contributes to allelic differences in specific transcript expression, as indicated by differences in AEI ratios between the two marker SNPs, even when the ratios correlate (r=0.62) (Figure 2). Human mRNA clones and expressed sequence tags display evidence of alternative splicing of *CHRNA5* in exon 5, in which rs16969968 resides. Presumably, any genetic variant driving a decrease in inclusion of rs16969968 in the mature mRNA could also decrease risk for dependence.

Conclusions

CHRNA5 exists in three similarly abundant main haplotypes (Supplementary Table S3) that have distinct biological functions, defined by the distal enhancer SNPs determining expression and the nonsynonymous rs16969968 SNP determining ligand-mediated signaling: the ancestral haplotype with low expression, a high-expressing haplotype, and a low-expressing haplotype with altered protein and channel activity. The effects of high *versus* low *CHRNA5* expression on channel activity *in vivo* remain to be determined. Because LD at the *CHRNA5* locus varies by ethnicity, the proposed enhancer variants should be considered when evaluating the penetrance of *CHRNA5* in nicotinic receptor-related disorders. The presence of additional functional variants cannot be excluded, possibly resulting in a more complex haplotype repertoire with distinct functions.

CONFLICT OF INTEREST

Drs LJ Bierut and JC Wang are listed as inventors on a patent (US 20070258898) covering the use of SNPs in determining diagnosis, prognosis, and treatment of addiction. Dr Bierut was a consultant for Pfizer, Inc. in 2008.

ACKNOWLEDGEMENTS

Funding: This work was supported in part by NIH grant DA022199 to W. Sadee and K02DA021237 to L.J. Bierut from the National Institute of Drug Abuse. The Collaborative Genetic Study on Nicotine Dependence (P01 CA89392 PI to L.J. Bierut) from the National Cancer Institute supported subject collection and genetic analysis to study nicotine dependence.

1 Berrettini W, Yuan X, Tozzi F *et al*: Alpha-5/alpha-3 nicotinic receptor subunit alleles increase risk for heavy smoking. *Mol Psychiatry* 2008; **13**: 368–373.
2 Bierut LJ, Stitzel JA, Wang JC *et al*: Variants in nicotinic receptors and risk for nicotine dependence. *Am J Psychiatry* 2008; **165**: 1163–1171.
3 Saccone SF, Hinrichs AL, Saccone NL *et al*: Cholinergic nicotinic receptor genes implicated in a nicotine dependence association study targeting 348 candidate genes with 3713 SNPs. *Hum Mol Genet* 2007; **16**: 36–49.
4 Saccone NL, Saccone SF, Hinrichs AL *et al*: Multiple distinct risk loci for nicotine dependence identified by dense coverage of the complete family of nicotinic receptor subunit (CHRN) genes. *Am J Med Genet B Neuropsychiatr Genet* 2009; **150B**: 453–466.
5 Saccone NL, Wang JC, Breslau N *et al*: The *CHRNA5-CHRNA3-CHRNB4* nicotinic receptor subunit gene cluster affects risk for nicotine dependence in African-Americans and in European-Americans. *Cancer Res* 2009; **69**: 6848–6856.
6 Sherva R, Kranzler HR, Yu Y *et al*: Variation in nicotinic acetylcholine receptor genes is associated with multiple substance dependence phenotypes. *Neuropsychopharmacology* 2010, e-pub ahead of print 19 May 2010. doi: 10.1038/npp.2010.64.
7 Sherva R, Wilhelmsen K, Pomerleau CS *et al*: Association of a single nucleotide polymorphism in neuronal acetylcholine receptor subunit alpha 5 (*CHRNA5*) with smoking status and with 'pleasurable buzz' during early experimentation with smoking. *Addiction* 2008; **103**: 1544–1552.
8 Stevens VL, Bierut LJ, Talbot JT *et al*: Nicotinic receptor gene variants influence susceptibility to heavy smoking. *Cancer Epidemiol Biomarkers Prev* 2008; **17**: 3517–3525.
9 The Tobacco and Genetics Consortium: Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nature Genet* 2010; **42**: 441–447.
10 Wang JC, Cruchaga C, Saccone NL *et al*: Risk for nicotine dependence and lung cancer is conferred by mRNA expression levels and amino acid change in *CHRNA5*. *Hum Mol Genet* 2009; **18**: 3125–3135.

11 Weiss RB, Baker TB, Cannon DS *et al*: A candidate gene approach identifies the *CHRNA5-A3-B4* region as a risk factor for age-dependent nicotine addiction. *PLoS Genet* 2008; **4**: e1000125.
12 Amos CI, Wu X, Broderick P *et al*: Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. *Nat Genet* 2008; **40**: 616–622.
13 Hung RJ, McKay JD, Gaborieau V *et al*: A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. *Nature* 2008; **452**: 633–637.
14 Liu P, Yang P, Wu X *et al*: A second genetic variant on chromosome 15q24-25.1 associates with lung cancer. *Cancer Res* 2010; **70**: 3128–3135.
15 Pillai SG, Ge D, Zhu G *et al*: A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. *PLoS Genet* 2009; **5**: e1000421.
16 Grucza RA, Wang JC, Stitzel JA *et al*: A risk allele for nicotine dependence in *CHRNA5* is a protective allele for cocaine dependence. *Biol Psychiatry* 2008; **64**: 922–929.
17 Schlaepfer IR, Hoft NR, Collins AC *et al*: The *CHRNA5/A3/B4* gene cluster variability as an important determinant of early alcohol and tobacco initiation in young adults. *Biol Psychiatry* 2008; **63**: 1039–1046.
18 Wang JC, Grucza R, Cruchaga C *et al*: Genetic variation in the *CHRNA5* gene affects mRNA levels and is associated with risk for alcohol dependence. *Mol Psychiatry* 2009; **14**: 501–510.
19 Johnson AD, Zhang Y, Papp AC *et al*: Polymorphisms affecting gene transcription and mRNA processing in pharmacogenetic candidate genes: detection through allelic expression imbalance in human target tissues. *Pharmacogenet Genomics* 2008; **18**: 781–791.
20 Wang D, Sadee W: Searching for polymorphisms that affect gene expression and mRNA processing: example *ABCB1* (MDR1). *AAPS J* 2006; **8**: E515–E520.
21 Ge B, Pokholok DK, Kwan T *et al*: Global patterns of cis variation in human cells revealed by high-density allelic expression analysis. *Nat Genet* 2009; **41**: 1216–1222.
22 Miller SA, Dykes DD, Polesky HF: A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* 1988; **16**: 1215.
23 Lim JE, Pinsonneault J, Sadee W, Saffen D: Tryptophan hydroxylase 2 (*TPH2*) haplotypes predict levels of *TPH2* mRNA expression in human pons. *Mol Psychiatry* 2007; **12**: 491–501.
24 Kent WJ, Sugnet CW, Furey TS *et al*: The human genome browser at UCSC. *Genome Res* 2002; **12**: 996–1006.
25 Johnson AD, Handsaker RE, Pulit S, Nizzari MM, O'Donnell CJ, de Bakker PIW: SNAP: A web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics* 2008; **24**: 2938–2939.
26 Su AI, Cooke MP, Ching KA *et al*: Large-scale analysis of the human and mouse transcriptomes. *Proc Natl Acad Sci USA* 2002; **99**: 4465–4470.
27 Wittkopp PJ, Haerum BK, Clark AG: Evolutionary changes in cis and trans gene regulation. *Nature* 2004; **430**: 85–88.
28 Sandelin A, Alkema W, Engstrom P, Wasserman WW, Lenhard B: JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Res* 2004; **32**: D91–D94.
29 Vlieghe D, Sandelin A, De Bleser PJ *et al*: A new generation of JASPAR, the open-access repository for transcription factor binding site profiles. *Nucleic Acids Res* 2006; **34**: D95–D97.
30 Kumar P, Pandey KN: Cooperative activation of *Npr1* gene transcription and expression by interaction of Ets-1 and p300. *Hypertension* 2009; **54**: 172–178.
31 Yang C, Shapiro LH, Rivera M, Kumar A, Brindle PK: A role for CREB binding protein and p300 transcriptional coactivators in Ets-1 transactivation functions. *Mol Cell Biol* 1998; **18**: 2218–2229.
32 Bolon I, Gouyer V, Devouassoux M *et al*: Expression of c-ets-1, collagenase 1, and urokinase-type plasminogen activator genes in lung carcinomas. *Am J Pathol* 1995; **147**: 1298–1310.
33 Yamaguchi E, Nakayama T, Nanashima A *et al*: Ets-1 proto-oncogene as a potential predictor for poor prognosis of lung adenocarcinoma. *Tohoku J Exp Med* 2007; **213**: 41–50.
34 Falvella FS, Galvan A, Frullanti E *et al*: Transcription deregulation at the 15q25 locus in association with lung adenocarcinoma risk. *Clin Cancer Res* 2009; **15**: 1837–1842.
35 Improgo MR, Schlichting NA, Cortes RY, Zhao-Shea R, Tapper AR, Gardner PD: *ASCL1* regulates the expression of the *CHRNA5/A3/B4* lung cancer susceptibility locus. *Mol Cancer Res* 2010; **8**: 194–203.
36 Pennacchio LA, Loots GG, Nobrega MA, Ovcharenko I: Predicting tissue-specific enhancers in the human genome. *Genome Res* 2007; **17**: 201–211.

Supplementary Information accompanies the paper on European Journal of Human Genetics website (http://www.nature.com/ejhg)