Supporting Information

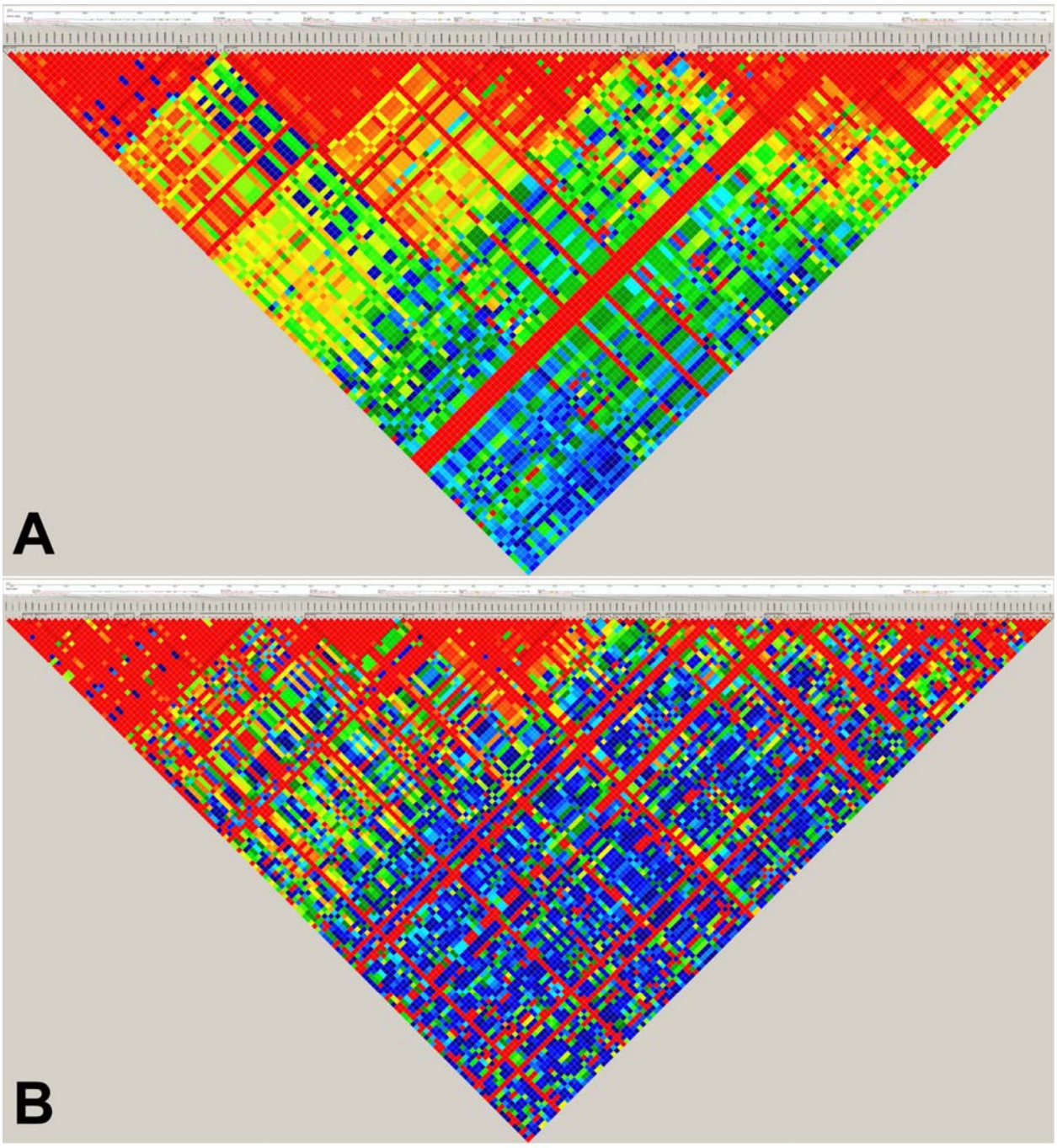
Funding Sources for External Databases

The GTEx Project was supported by the Common Fund of the Office of the Director of the NIH. Additional funds were provided by the NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. Donors were enrolled at Biospecimen Source Sites funded by NCI\SAIC-Frederick, Inc. (SAIC-F) subcontracts to the National Disease Research Interchange (10XS170), Roswell Park Cancer Institute (10XS171), and Science Care, Inc. (X10S172). The Laboratory, Data Analysis, and Coordinating Center (LDACC) was funded through a contract (HHSN268201000029C) to The Broad Institute, Inc. Biorepository operations were funded through an SAIC-F subcontract to Van Andel Institute (10ST1035). Additional data repository and project management were provided by SAIC-F (HHSN261200800001E). The Brain Bank was supported by a supplement to University of Miami grants DA006227 & DA033684 and to contract N01MH000028. Statistical Methods development grants were made to the University of Geneva (MH090941 & MH101814), the University of Chicago (MH090951, MH090937, MH101820, MH101825), the University of North Carolina - Chapel Hill (MH090936 & MH101819), Harvard University (MH090948), Stanford University (MH101782), Washington University St Louis (MH101810), and the University of Pennsylvania (MH101822).

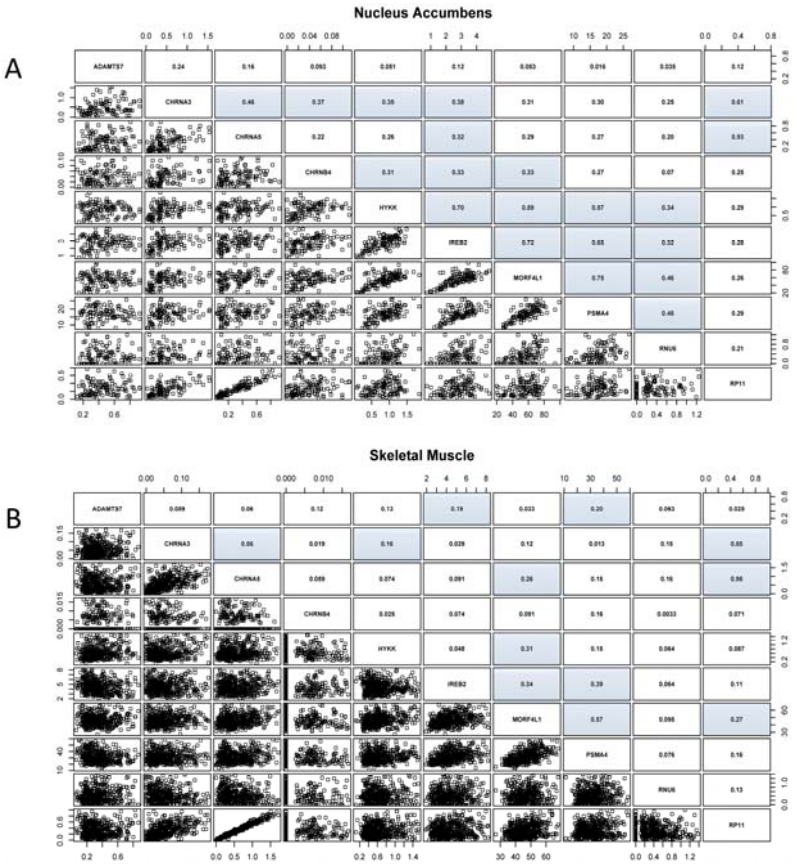
Funding for SAGE was provided through the NIH Genes, Environment and Health Initiative [GEI] (U01 HG004422). SAGE is a GWAS funded as part of the Gene Environment Association Studies (GENEVA) under GEI. Assistance with phenotype harmonization and genotype cleaning, as well as with general study coordination, was provided by the GENEVA Coordinating Center (U01 HG004446). Assistance with data cleaning was provided by the National Center for Biotechnology Information. Support for collection of datasets and samples

was provided by COGA; U10 AA008401, COGEND; P01 CA089392, and FSCD; R01 DA013423. Genotyping at the Johns Hopkins University Center for Inherited Disease Research (CIDR) was funded by NIH GEI (U01HG004438), the National Institute on Alcohol Abuse and Alcoholism, NIDA, and the NIH contract High throughput genotyping for studying the genetic contributions to human disease (HHSN268200782096C).

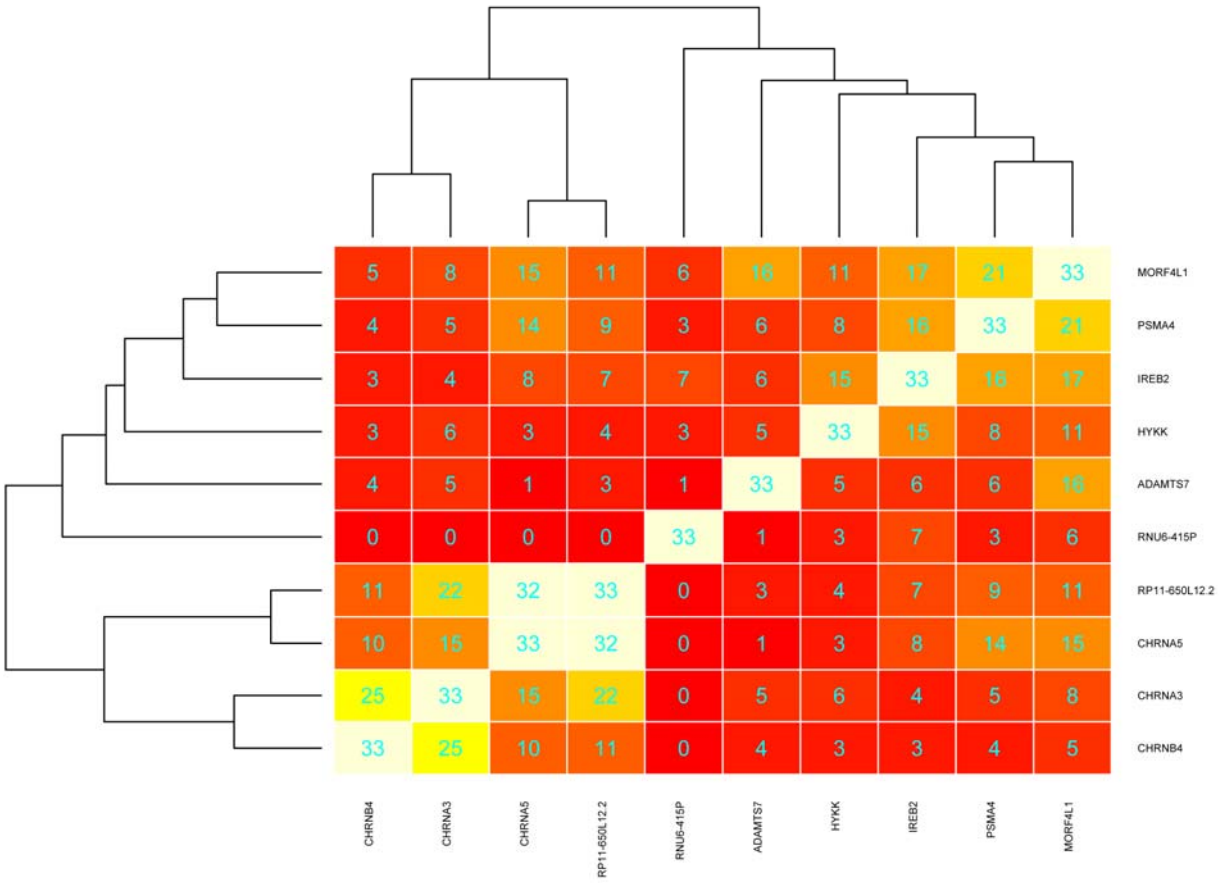
The Genetic Architecture of Smoking and Smoking Cessation data was accessed through dbGAP. Genotyping, performed at CIDR, was funded by 1 X01HG005274-01. CIDR is fully funded through a federal contract from the National Institutes of Health to The Johns Hopkins University, contract number HHSN268200782096C. Assistance with genotype cleaning, as well as with general study coordination, was provided by the GENEVA Coordinating Center (U01HG004446). Funding support for collection of datasets and samples was provided by COGEND (P01 CA089392) and UW-TTURC (P50 DA019706, P50 CA084724).



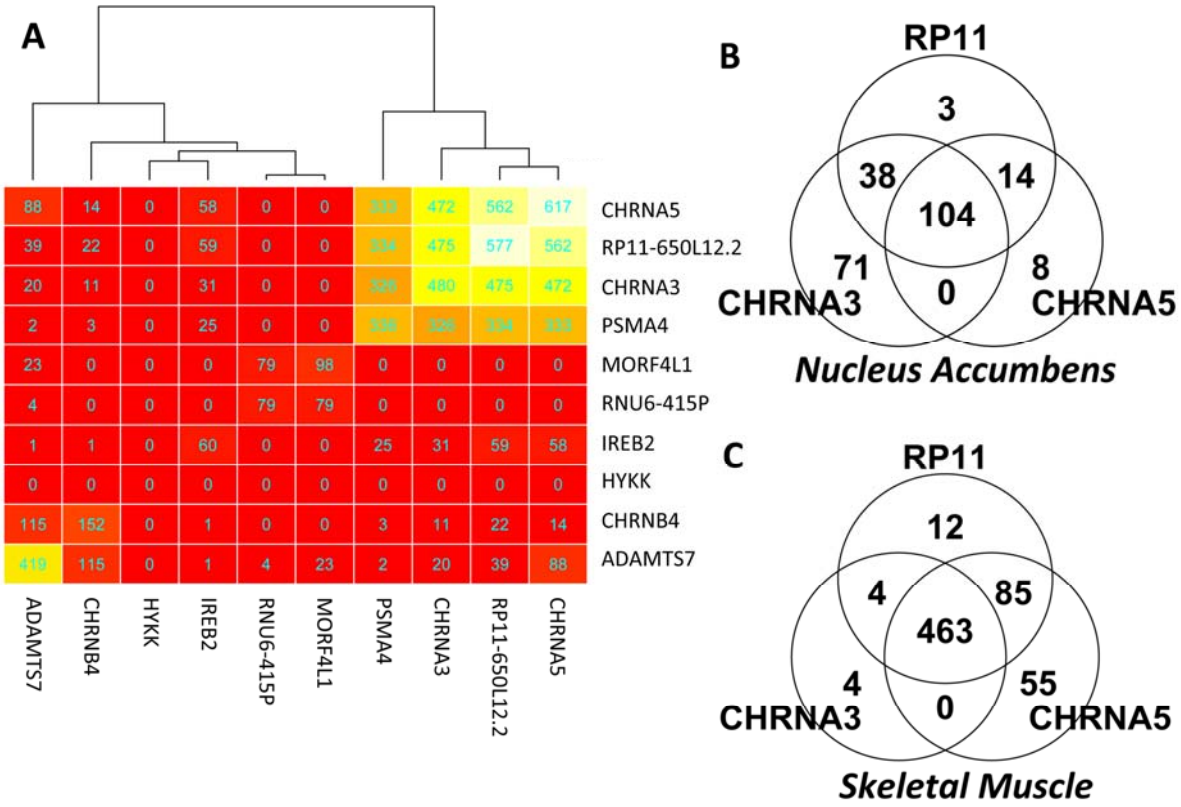
Suppl. Figure S1. LD of nicotinic receptor locus in YRI and CEU. LD structure for all SNPs in the 500kb (chr15:76507-76985 kb) region (available from HapMap v3 in Haploview (v4.2)) for people of European (1000 Genomes population CEU, A) and African descent (YRI, B). Boxes are shaded according to the standardized LD coefficient, D, with red representing D′=1, and haplotype blocks defined by 'confidence interval' parameter (Gabriel, 2002).



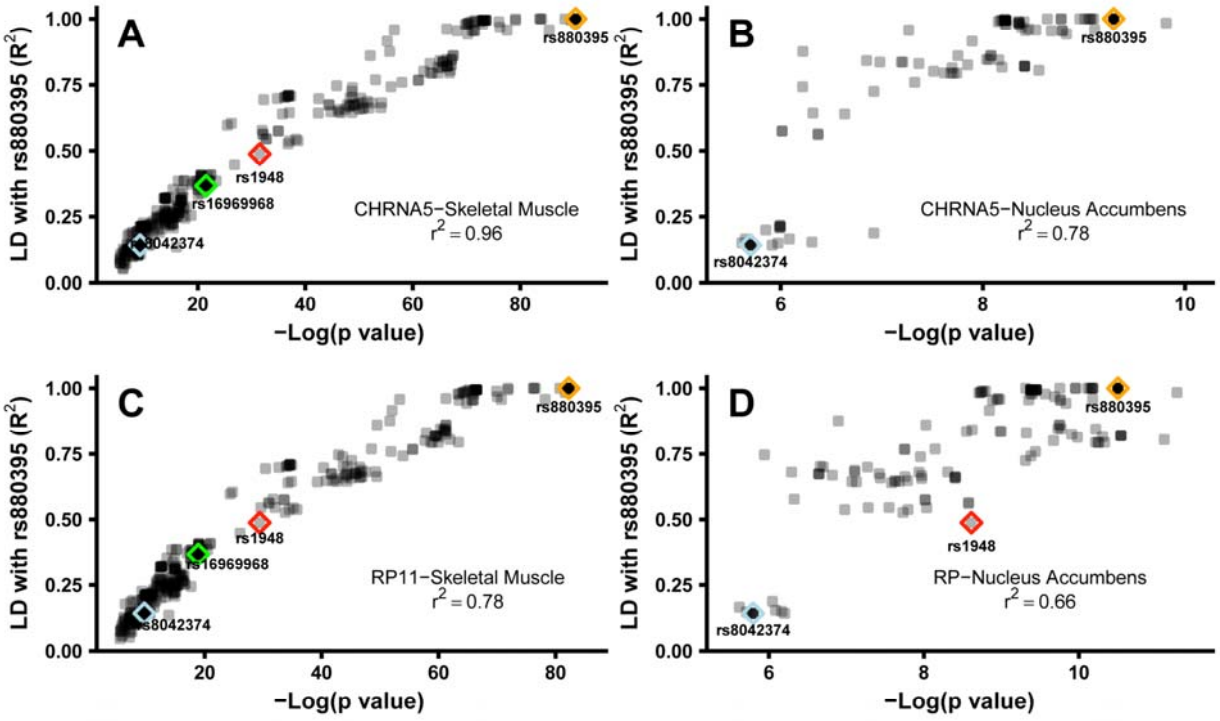
Suppl. Figure S2. Tissue-specific mRNA co-expression. Correlation between mRNA levels (RPKM in GTEx) between genes in the locus locus in nucleus accumbens (top) and skeletal muscle (bottom). The estimated correlation is show in the upper matrix with significant values (multiple hypothesis/Bonferroni corrected p value < 0.05) highlighted in blue. Scatter plots are shown in the lower matrix. Each dot represents an individual sample. Outliers were removed using quantiles.



Suppl. Figure S3. Number of tissues co-expressed. For each pair of genes (x/y intersection) the number of tissues where the two genes were significantly co-expressed (multiple hypothesis/Bonferroni corrected p value < 0.05) is displayed. Colors on a scale from no co-expression (red) to complete co-expression (white).

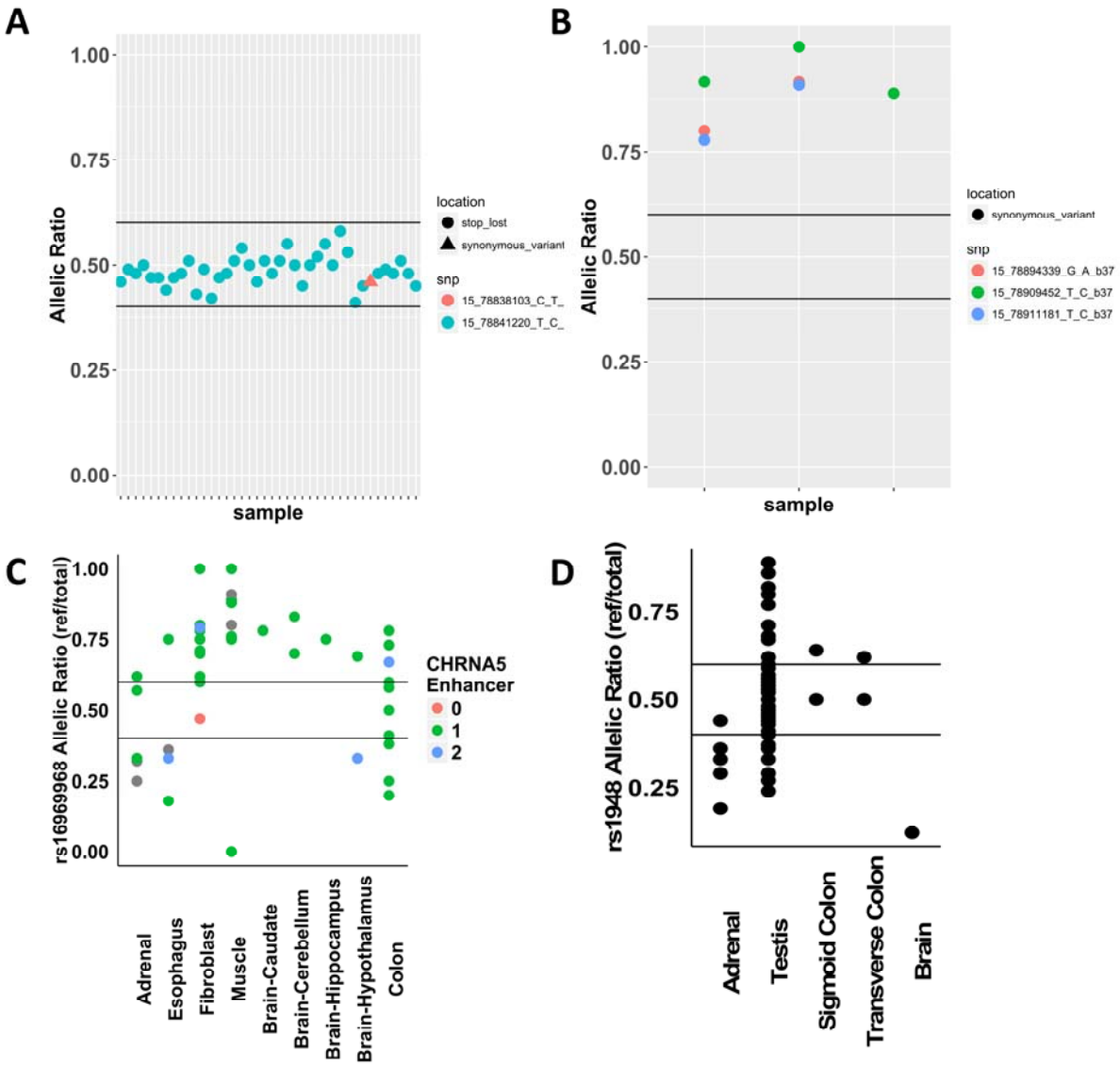


Suppl. Figure S4. Overlap between eQTLs. A) For each pair of genes (x/y intersection) the number of SNPs that are identified as eQTLs by GTEx in any tissue for both genes in the pair is displayed. For CHRNA3, CHRNA5 and RP11-650L12.2 the number of overlapping eQTLs in nucleus accumbens (B) and skeletal muscle (C) again as reported by GTEx. Colors on a scale from no overlap (red) to complete overlap (white).

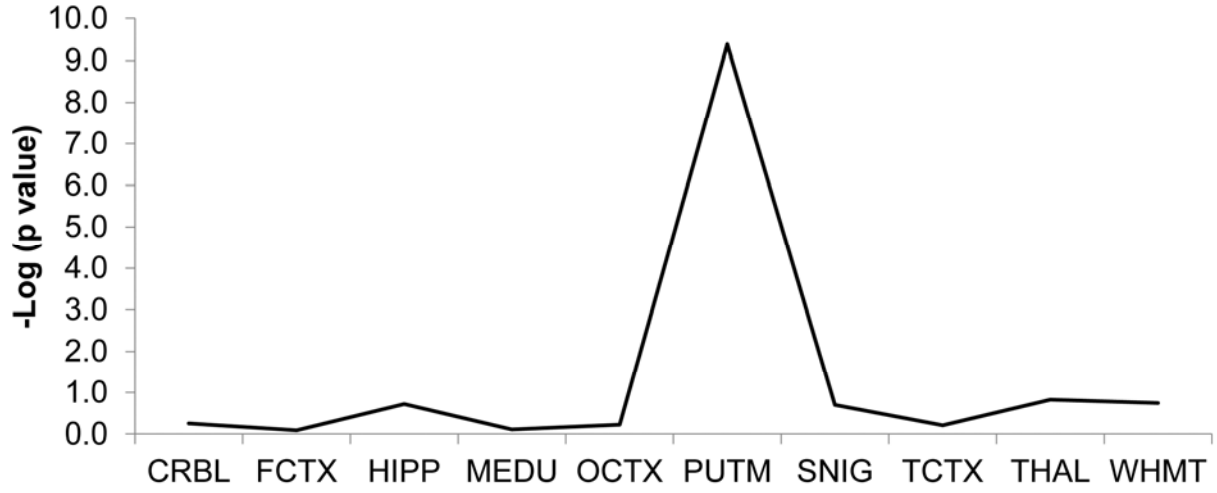


Suppl. Figure S5. LD association with eQTL strength. Correlation between -log (p value) and LD (R^2) with the top eQTL haplotype: for *CHRNA5* rs880395 in skeletal muscle (A. r^2=0.96) and nucleus accumbens (B. r^2=0.78); for *RP11-650L12.2* rs880395 in skeletal muscle (C. r^2=0.78) and nucleus accumbens (D. r^2 = 0.66.. Diamonds highlight rs880395 (orange), rs16969968 (green), - rs1948 (red), and a potential additional regulatory haplotype for *CHRNA5*, and *RP11-650L12.2* in brain tagged by rs8042374 (blue).

.



Suppl. Figure S6. Allelic expression ratios. More sensitive than eQTL analysis, measurement of allelic ratios can reveal the presence of *cis*-acting regulatory variants. SNPs with equal representation of both alleles have a ratio of 0.5 (reference/total), with deviations demonstrating an imbalance. Allelic ratios are plotted for all marker SNPs in *PSMA4* (A) and *CHRNA3* (B) in nucleus accumbens. *CHRNA3* shows robust allelic ratios in nucleus accumbens measured at 3 different SNPs (B), indicating the presence of a strong *cis*-acting regulatory element. (Fig. 4A). Allelic ratios are also measured at rs16969968 (C) and rs1948 (D). Each dot represents the allelic ratio for an individual in the given tissue. In C, the color indicates the genotype for the *CHRNA5*-enhancer haplotype: red, green and blue are 0, 1 or 2 minor alleles respectively. In most tissues, significant AEI is accounted for by the *CHRNA5*-enhancer haplotype, but not in every tissue (colon and LCLs). In D, CHRNB4 mRNA exhibits evidence of AEI, at rs1948, in several tissues. Allelic ratios consistently below 0.4 in adrenal tissue suggest this SNP is in strong LD with a functional variant affecting CHRNB4 expression in adrenal tissue.



Suppl. Figure S7. CHRNA3 eQTLs replicate in Braineac. Microarray data from braineac.org provide gene expression across ten brain regions. Plot shows the association between CHRNA3 mRNA expression and rs1948 in different brain regions (CRBL = Cerebellar cortex, FCTX = Frontal cortex, HIPP = hippocampus, MEDU = medulla, OCTX = Occipital cortex, PUTM = putamen, SNIG = substantia nigra, TCTX = temporal cortex, THAL = thalamus, WHMT = intralobular white matter).

Supplementary Table 1. Number of nicotine dependent (cases) and nicotine non-dependent (controls) by study/sub-study.

Study/Sub-study	Nicotine Non-dependent (controls)	Nicotine Dependent (cases)
COGA	37	363
COGEND	763	959
DEPEND	1	229
ED_SR	2	122
FSCD	28	221
TTURC2	15	862

Supplementary Table 2. Tissue specific expression profile of genes in nicotinic receptor locus.

Tissues with RNA expression	Gene
not expressed in any tissue	*RP11-160C18.4, RP11-650L12.1, RNU6-415P*
not in GTEx	*RP11-160C18.5, RP11-16K12.2, RP11-335K5.3, RP11-650L12.4*
universally well-expressed	*ADAMTS7, HYKK, MORF4L1, PSMA4, IREB2*
tissue specific expression	*CHRNA3, CHRNA5, CHRNB4, RP11-650L12.2, RPL21P116*
testes only	*RP11-160C18.2, RPL18P11, RP11-335K5.2*

Supplementary Table 3. Number of GTEx reported eQTLs in locus by tissue.

Tissue	Number of eQTLs
Muscle Skeletal	2100
Nerve Tibial	1095
Artery Tibial	1047
Cells Transformed fibroblasts	949
Heart Left Ventricle	851
Adipose Subcutaneous	779
Lung	552
Brain Caudate basal ganglia	539
Breast Mammary Tissue	509
Brain Nucleus accumbens basal ganglia	498
Brain Putamen basal ganglia	498
Esophagus Mucosa	477
Skin Sun Exposed Lower leg	468
Adipose Visceral Omentum	442
Testis	422
Artery Aorta	395
Heart Atrial Appendage	361
Whole Blood	356
Skin Not Sun Exposed Suprapubic	354
Brain Frontal Cortex BA9	329
Thyroid	299
Brain Cerebellar Hemisphere	284
Artery Coronary	267
Brain Anterior cingulate cortex BA24	238

Brain Cerebellum	234
Brain Cortex	233
Brain Hippocampus	229
Brain Hypothalamus	209
Pancreas	138
Cells EBV-transformed lymphocytes	134
Esophagus Muscularis	16
Colon Sigmoid	1
Adrenal Gland	0
Colon Transverse	0
Esophagus Gastroesophageal Junction	0
Liver	0
Ovary	0
Pituitary	0
Prostate	0
Small Intestine Terminal Ileum	0
Spleen	0
Stomach	0
Uterus	0
Vagina	0

Supplemental Table S4. eQTLs reported by GTEx for *CHRNA5*-enhancer locus (rs880395) in each gene x tissue combination available. Bold if it is the source of the top eQTL signal for that tissue. The effect of this SNP is negative for all genes except *PSMA4*.

Gene	**Tissue**	**p value**	**Effect Size**
CHRNA3	**Muscle - Skeletal**	**8.30E-40**	**-0.79**
	Brain - Caudate (basal ganglia)	1.10E-11	-0.72
	Brain - Nucleus accumbens (basal ganglia)	4.80E-08	-0.74
	Brain - Putamen (basal ganglia)	6.60E-08	-0.60
	Testis	9.7E-06	-0.50
CHRNA5	**Adipose**	**5.40E-38**	**-0.85**
	Artery - Aorta	**1.40E-07**	**-0.37**
	Artery - Coronary	**1.60E-09**	**-0.62**

	Artery - Tibial	**3.50E-21**	**-0.63**
	Brain - Anterior cingulate cortex (BA24)	**3.00E-12**	**-0.91**
	Brain - Caudate (basal ganglia)	**3.40E-12**	**-0.79**
	Brain - Cerebellar Hemisphere	**1.40E-16**	**-0.84**
	Brain - Cerebellum	**5.10E-14**	**-0.87**
	Brain - Cortex	**4.00E-14**	**-0.97**
	Brain - Frontal Cortex (BA9)	**6.50E-14**	**-0.82**
	Brain - Hippocampus	**7.60E-13**	**-0.82**
	Brain - Hypothalamus	**1.50E-09**	**-0.67**
	Brain - Nucleus accumbens (basal ganglia)	**5.10E-10**	**-0.75**
	Brain - Putamen (basal ganglia)	**4.80E-12**	**-0.82**
	Breast - Mammary Tissue	**1.70E-14**	**-0.62**
	Cells - Transformed fibroblasts	**7.50E-51**	**-0.69**
	Esophagus - Mucosa	**5.10E-15**	**-0.36**
	Heart - Atrial Appendage	**1.10E-19**	**-0.85**
	Heart - Left Ventricle	**1.50E-30**	**-0.81**
	Lung	**1.80E-18**	**-0.6**
	Muscle - Skeletal	**4.90E-91**	**-1.1**
	Nerve - Tibial	**2.30E-44**	**-0.92**
	Skin	**4.00E-16**	**-0.51**

	Thyroid	**1.70E-07**	**-0.39**
	Whole Blood	**1.40E-07**	**-0.37**
PSMA4	Adipose	1.70E-06	0.17
	Artery - Aorta	7.80E-06	0.16
	Artery - Tibial	**6E-06**	**0.15**
	Heart - Left Ventricle	**1.10E-11**	**0.41**
	Muscle - Skeletal	**2.30E-15**	**0.23**
	Nerve - Tibial	6.40E-08	0.19
	Skin	**2.60E-06**	**0.17**
RP11-650L12.2	**Adipose**	**3.30E-35**	**-0.82**
	Artery - Aorta	**2.70E-06**	**-0.32**
	Artery - Coronary	**3.90E-08**	**-0.56**
	Artery - Tibial	**1.50E-18**	**-0.59**
	Brain - Anterior cingulate cortex (BA24)	**8.60E-09**	**-0.8**
	Brain - Caudate (basal ganglia)	1.70E-12	-0.85
	Brain - Cerebellar Hemisphere	**8.50E-09**	**-0.55**
	Brain - Cortex	**8.40E-13**	**-0.95**
	Brain - Frontal Cortex (BA9)	**5.00E-13**	**-0.8**
	Brain - Hippocampus	**1.20E-10**	**-0.79**
	Brain - Hypothalamus	**2.60E-07**	**-0.56**
	Brain - Nucleus accumbens (basal ganglia)	**3.10E-11**	**-0.79**

	Brain - Putamen (basal ganglia)	**9.90E-12**	**-0.88**
	Breast - Mammary Tissue	**6.50E-14**	**-0.6**
	Cells - Transformed fibroblasts	**4.70E-52**	**-0.69**
	Esophagus - Mucosa	**2.20E-12**	**-0.35**
	Heart - Atrial Appendage	**4.90E-17**	**-0.83**
	Heart - Left Ventricle	**6.60E-27**	**-0.81**
	Lung	**2.00E-20**	**-0.64**
	Muscle - Skeletal	**5.60E-83**	**-1**
	Nerve - Tibial	**1.70E-40**	**-0.9**
	Skin	2.40E-14	-0.45
	Thyroid	**1.50E-08**	**-0.42**
	Whole Blood	**6.10E-07**	**-0.35**

Supplemental Table S5. Evidence sources for candidate SNPs.

Source		SNPs
(1) Top eQTLs	Most significant eQTLs (lowest p values) reported by GTEx for CHRNA5, CHRNA3, and CHRNB4 and evaluated in preceding sections of manuscript	rs4886580, rs4887074, rs880395, rs7164030, rs1979905, rs1979906, rs1979907, rs905740, rs4887064, rs12907966, rs1948
(2) Multi-gene eQTLs	SNPs identified as genome-wide significant eQTLs by GTEx for multiple genes in the locus. There are 835 SNPs that are eQTLs for more than one gene in the locus. These 5 SNPs that are eQTLs for all nicotinic genes and RP11-650L12.2 were included as SNPs of interest.	rs12915669: CHRNA3, CHRNA5, CHRNB4, RP11-335K5.2, RP11-650L12.2 rs28661610: CHRNA3, CHRNA5, CHRNB4, RP11-160C18.2, RP11-160C18.4, RP11-650L12.2 rs2869552: ADAMTS7,

		CHRNA3, CHRNA5, CHRNB4, PSMA4, RP11-160C18.2, RP11-650L12.2 rs2869553: CHRNA3, CHRNA5, CHRNB4, RP11-160C18.2, RP11-160C18.4, RP11-650L12.2 rs62010552: ADAMTS7, CHRNA3, CHRNA5, CHRNB4, PSMA4, RP11-160C18.2, RP11-650L12.2
(3) eQTL/GWAS overlap	To integrate molecular and clinical associations, we determined which SNPs in the 500 kb region surrounding the nicotinic cluster were both eQTLs for the *CHRNA5/CHRNA3/CHRNB4* cluster (in any tissue as reported by GTEx) and GWAS hits (with any clinical phenotype in GRASP). This query resulted in 11 SNPs (rs1051730, rs11858836, rs12914385, rs13180, rs17486278, rs1994016, rs2036527, rs2219939, rs3825807, rs4380028, rs55958997, rs667282, rs7173743, rs8034191, rs8040868, rs8042238, rs8042374, rs899997) assigned to our genes of interest (*CHRNA3*, *CHRNA5*, *CHRNB4*, and *RP11-160C18.2*). Including all SNPs in perfect LD (R^2 = 1, D′ = 1) with these 11 variants, we then filtered with biomaRt in R for those falling in regulatory regions defined by ENSEMBL (ENSRs) resulting in 17 SNPs.	rs8040868, rs1814880, rs11852909, rs2017091, rs61390267, rs7176070, rs12899940, rs12907511, rs11632672, rs7182993, rs7182694, rs7181486, rs17405217, rs17483548, rs55781567, rs55853698, rs8040868, rs1051730, rs2036527, rs6495309, rs660652, rs55781567, rs55853698, rs55958997, rs11858836, rs8034191, rs17486278, rs1051730, rs12914385, rs8042374, rs8040868, rs11858836, rs8034191, rs2036527, rs17486278, rs1051730, rs12914385, rs55958997, rs8042374, rs667282, rs899997
(4) Allelic ratios	Allelic ratios reported by GTEx were evaluated for each gene in the locus. Interpretation of these ratios	rs11633223, rs8040868, rs55781567, rs55853698, rs1948, rs16969968, rs1051730,

	revealed several SNPs of interest.	rs3743075, rs8040868, rs8192475, rs386616281
(5) Popular in the literature	A thorough review of the literature revealed these SNPs as potentially functional and/or relevant.	rs16969968, rs6495309, rs660652, rs55781567, rs55853698, rs8040868, rs1051730, rs2036527, rs16969968, rs880395, rs7164030, rs1979905, rs1979906, rs1979907, rs905740, rs588765, rs578776, rs12914385

Supplementary Table 6. Genotypes of rs16969968 and rs880395 from 1000 genomes.
Phased data reveals presence of two variants on opposite alleles (i.e. rarely occur together).

		rs16969968			
		0|0	0|1	1|0	1|1
rs880395	0|0	187	1	1	0
	0|1	283	123	2	2
	1|0	324	5	115	0
	1|1	1077	142	128	114