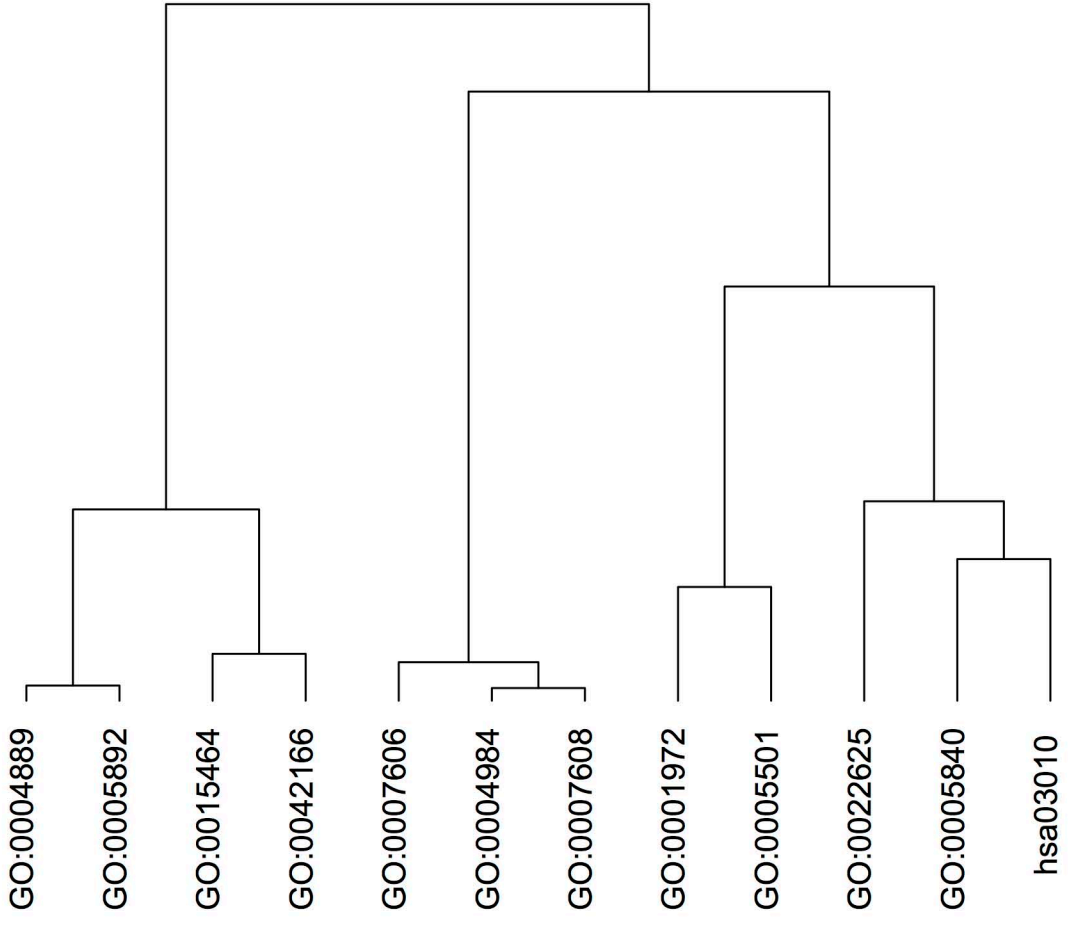
GO:0004889
GO:0005892
GO:0015464
GO:0042166
GO:0007606
GO:0004984
GO:0007608
GO:0001972
GO:0005501
GO:0022625
GO:0005840
hsa03010

Figure S1.

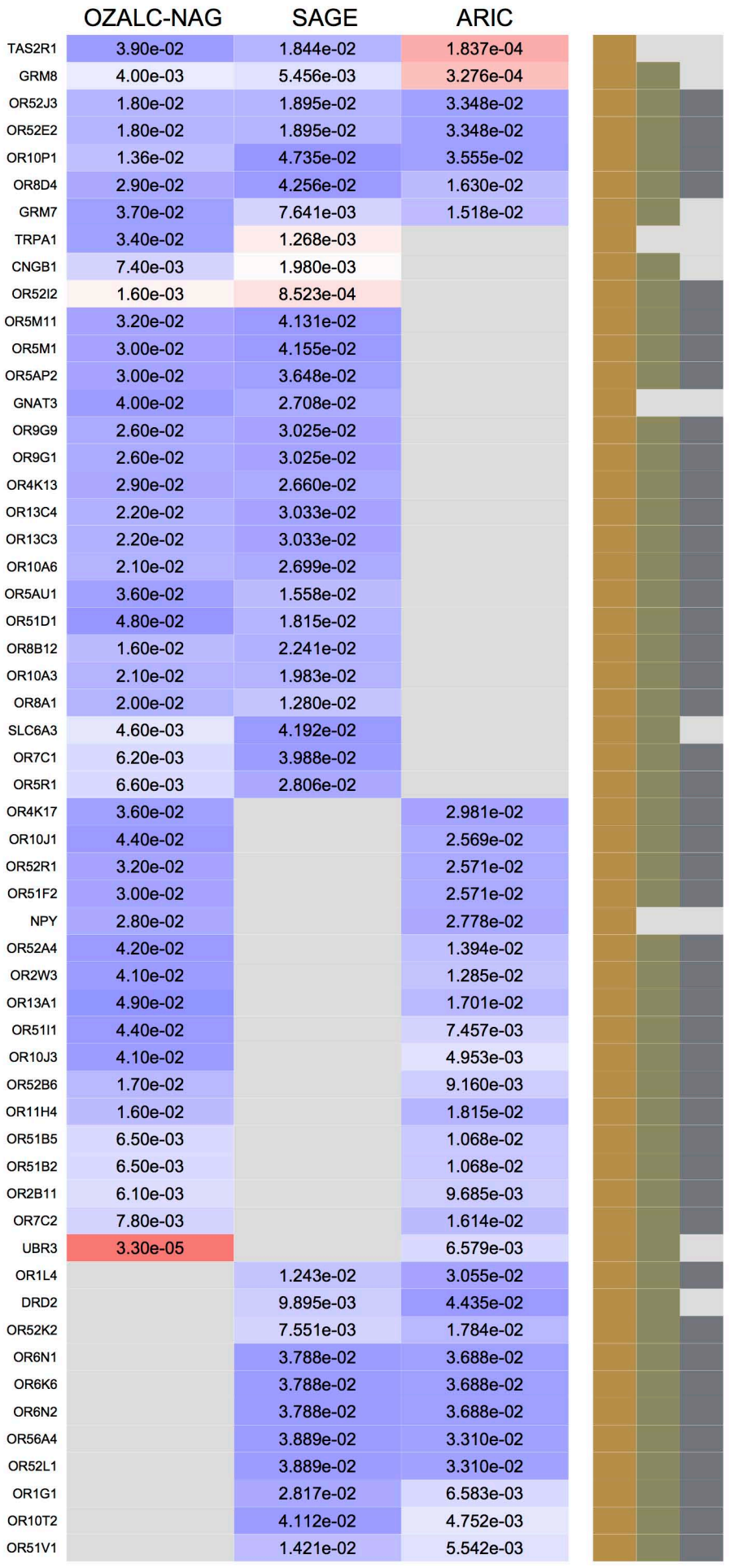
OZALC-NAG
SAGE
ARIC
TAS2R1 3.90e-02 1.844e-02 1.837e-04
GRM8 4.00e-03 5.456e-03 3.276e-04
OR52J3 1.80e-02 1.895e-02 3.348e-02
OR52E2 1.80e-02 1.895e-02 3.348e-02
OR10P1 1.36e-02 4.735e-02 3.555e-02
OR8D4 2.90e-02 4.256e-02 1.630e-02
GRM7 3.70e-02 7.641e-03 1.518e-02
TRPA1 3.40e-02 1.268e-03
CNGB1 7.40e-03 1.980e-03
OR52I2 1.60e-03 8.523e-04
OR5M11 3.20e-02 4.131e-02
OR5M1 3.00e-02 4.155e-02
OR5AP2 3.00e-02 3.648e-02
GNAT3 4.00e-02 2.708e-02
OR9G9 2.60e-02 3.025e-02
OR9G1 2.60e-02 3.025e-02
OR4K13 2.90e-02 2.660e-02
OR13C4 2.20e-02 3.033e-02
OR13C3 2.20e-02 3.033e-02
OR10A6 2.10e-02 2.699e-02
OR5AU1 3.60e-02 1.558e-02
OR51D1 4.80e-02 1.815e-02
OR8B12 1.60e-02 2.241e-02
OR10A3 2.10e-02 1.983e-02
OR8A1 2.00e-02 1.280e-02
SLC6A3 4.60e-03 4.192e-02
OR7C1 6.20e-03 3.988e-02
OR5R1 6.60e-03 2.806e-02
OR4K17 3.60e-02 2.981e-02
OR10J1 4.40e-02 2.569e-02
OR52R1 3.20e-02 2.571e-02
OR51F2 3.00e-02 2.571e-02
NPY 2.80e-02 2.778e-02
OR52A4 4.20e-02 1.394e-02
OR2W3 4.10e-02 1.285e-02
OR13A1 4.90e-02 1.701e-02
OR51I1 4.40e-02 7.457e-03
OR10J3 4.10e-02 4.953e-03
OR52B6 1.70e-02 9.160e-03
OR11H4 1.60e-02 1.815e-02
OR51B5 6.50e-03 1.068e-02
OR51B2 6.50e-03 1.068e-02
OR2B11 6.10e-03 9.685e-03
OR7C2 7.80e-03 1.614e-02
UBR3 3.30e-05 6.579e-03
OR1L4 1.243e-02 3.055e-02
DRD2 9.895e-03 4.435e-02
OR52K2 7.551e-03 1.784e-02
OR6N1 3.788e-02 3.688e-02
OR6K6 3.788e-02 3.688e-02
OR6N2 3.788e-02 3.688e-02
OR56A4 3.889e-02 3.310e-02
OR52L1 3.889e-02 3.310e-02
OR1G1 2.817e-02 6.583e-03
OR10T2 4.112e-02 4.752e-03
OR51V1 1.421e-02 5.542e-03
GO:0007606 “sensory perception of chemical stimulus”
GO:0007608 “sensory perception of smell”
GO:0004984 “olfactory receptor activity”
0.05
1.0E-05

Figure S2.

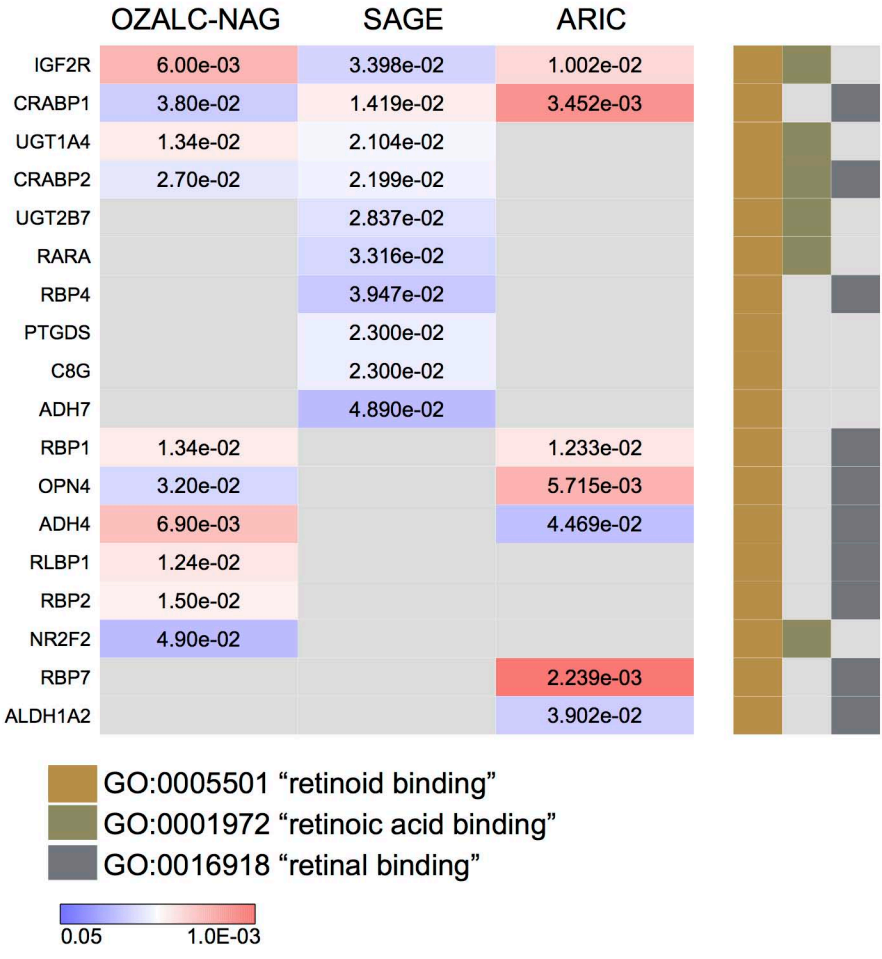
OZALC-NAG
SAGE
ARIC
IGF2R
CRABP1
UGT1A4
CRABP2
UGT2B7
RARA
RBP4
PTGDS
C8G
ADH7
RBP1
OPN4
ADH4
RLBP1
RBP2
NR2F2
RBP7
ALDH1A2
6.00e-03
3.398e-02
1.002e-02
3.80e-02
1.419e-02
3.452e-03
1.34e-02
2.104e-02
2.70e-02
2.199e-02
2.837e-02
3.316e-02
3.947e-02
2.300e-02
2.300e-02
4.890e-02
1.34e-02
1.233e-02
3.20e-02
5.715e-03
6.90e-03
4.469e-02
1.24e-02
1.50e-02
4.90e-02
2.239e-03
3.902e-02
GO:0005501 “retinoid binding”
GO:0001972 “retinoic acid binding”
GO:0016918 “retinal binding”
0.05
1.0E-03

Figure S3.

Table S1.

A)

SNP Threshold	Study	GO term enrichment threshold 0.005 #cat.	0.005 p-value	0.01 #cat.	0.01 p-value	0.05 #cat.	0.05 p-value
0.001	OZALC-NAG	23	1.20E-03	35	4.40E-03	101	1.02E-02
	SAGE	15	6.40E-03	34	2.20E-03	119	4.60E-03
0.005	OZALC-NAG	18	3.18E-02	26	9.02E-02	139	9.40E-02
	SAGE	11	1.37E-01	34	1.20E-01	179	2.44E-02
0.01	OZALC-NAG	19	8.84E-02	37	9.12E-02	163	1.62E-01
	SAGE	21	7.46E-02	41	8.74E-02	192	1.01E-01
0.05	OZALC-NAG	33	3.44E-02	62	4.60E-02	262	2.33E-01
	SAGE	42	1.80E-02	81	3.10E-02	313	5.78E-02

B)

SNP Threshold	Study	KEGG pathways enrichment threshold 0.005 #cat.	0.005 p-value	0.01 #cat.	0.01 p-value	0.05 #cat.	0.05 p-value
0.001	OZALC-NAG	1	2.80E-03	1	1.16E-02	1	1.76E-01
	SAGE	0	1.00E+00	1	7.80E-03	3	1.80E-03
0.005	OZALC-NAG	0	1.00E+00	1	7.00E-02	8	7.70E-02
	SAGE	0	1.00E+00	1	7.44E-02	2	5.54E-01
0.01	OZALC-NAG	0	1.00E+00	0	1.00E+00	12	4.62E-02
	SAGE	0	1.00E+00	0	1.00E+00	5	3.48E-01
0.05	OZALC-NAG	6	8.00E-04	10	1.20E-03	19	1.54E-02
	SAGE	4	3.20E-03	4	2.60E-02	13	8.40E-02

Table S2

			OZALC-NAG			SAGE			ARIC			Combined
acc	**name**	**# Genes Cat.**	**# Genes**	***p-value***	**Expected**	**# Genes**	***p-value***	**Expected**	**# Genes**	***p-value***	**Expected**	***p-value****
GO:0035095	behavioral response to nicotine	7	3	2.00E-05	0.034	2	2.60E-04	0.028	3	2.00E-05	0.032	2.97E-10
GO:0004889	nicotinic acetylcholine-activated cation-selective channel activity	17	4	2.00E-05	0.114	2	4.20E-03	0.095	4	2.00E-05	0.000	1.33E-09
GO:0005892	nicotinic acetylcholine-gated receptor-channel complex	16	4	2.00E-05	0.114	2	4.20E-03	0.095	4	2.00E-05	0.000	1.33E-09
GO:0042166	acetylcholine binding	22	4	4.00E-05	0.125	2	5.56E-03	0.105	4	4.00E-05	0.000	5.99E-09
GO:0015464	acetylcholine receptor activity	18	4	4.00E-05	0.110	2	4.16E-03	0.093	3	2.80E-04	0.000	4.74E-08
GO:0035094	response to nicotine	20	3	2.16E-03	0.263	2	2.13E-02	0.225	3	1.82E-03	0.002	2.11E-05
GO:0007274	neuromuscular synaptic transmission	15	2	2.75E-02	0.256	2	2.10E-02	0.218	2	2.50E-02	0.025	2.56E-03

Table S3

	OZALC-NAG	SAGE	ARIC
CHRNA5	rs16969968	rs16969968	rs569207
CHRNA3	rs16969968	rs16969968	rs569207
CHRNB4	rs6495309	rs6495309	rs8040868
CHRNB3	rs4950	rs1530848	rs4950
CHRNA6	rs16891604		rs7012713
CHRNA9	rs4861079	rs4469115	rs10021263
CHRNA7	rs904951	rs4779969	rs11858834
CHRNA2	rs6557999	rs9773817	
CHRM5	rs8035805		
CHRM3	rs2355228		
CHRNB2	rs1127313		
CHRM2	rs17506733		
CHRM4			rs12574668
CHRM1		rs12418496	rs12418496
CHRNA4		rs3787138	
CHRNG		rs733603	
CHRND		rs733603	

Table S4

		OZALC-NAG	SAGE	ARIC
Acc	**Name**	***q-value***	***q-value***	***q-value***
GO:0035095	behavioral response to nicotine	1.90E-01	9.89E-02	5.34E-02
GO:0060084	synaptic transmission involved in micturition	4.53E-01	8.14E-02	6.34E-02
GO:0006942	regulation of striated muscle contraction	5.15E-01	2.25E-01	1.00E+00
GO:0004889	nicotinic acetylcholine-activated cation-selective channel activity	1.00E+00	3.90E-01	6.03E-02
GO:0005892	nicotinic acetylcholine-gated receptor-channel complex	5.42E-01	3.51E-01	4.86E-02
GO:0015464	acetylcholine receptor activity	1.70E-01	7.23E-01	8.54E-02
GO:0005230	extracellular ligand-gated ion channel activity	5.09E-01	4.55E-01	3.86E-01
GO:0007271	synaptic transmission, cholinergic	4.77E-01	5.83E-01	3.23E-01
GO:0042060	wound healing	4.93E-01	4.53E-01	8.11E-01
GO:0006940	regulation of smooth muscle contraction	5.88E-01	4.90E-01	5.77E-01
GO:0005216	ion channel activity	6.65E-01	4.07E-01	6.97E-01
GO:0042552	myelination	5.96E-01	4.75E-01	9.04E-01
GO:0007257	activation of JUN kinase activity	6.30E-01	3.59E-01	8.35E-01
GO:0006548	histidine catabolic process	6.33E-01	4.00E-01	7.10E-01
GO:0034185	apolipoprotein binding	6.30E-01	4.21E-01	9.97E-01
GO:0007265	Ras protein signal transduction	7.52E-01	4.46E-01	1.00E+00

Table S5

acc	Category of gene	OZALC-NAG	SAGE	ARIC
GO:0042166	acetylcholine binding	1.00E-03	1.00E-03	1.00E-02
GO:0015464	acetylcholine receptor activity	5.00E-03	2.60E-03	6.20E-03
GO:0004889	nicotinic acetylcholine-activated cation-selective channel activity	4.00E-03	1.00E-03	2.90E-02
GO:0005892	nicotinic acetylcholine-gated receptor-channel complex	4.00E-03	1.00E-03	2.90E-02
GO:0004984	olfactory receptor activity	2.54E-02	1.34E-02	2.80E-03
GO:0007608	sensory perception of smell	1.14E-01	1.91E-01	2.58E-01
GO:0007606	sensory perception of chemical stimulus	9.20E-02	1.87E-01	1.33E-01

Pathway analysis of smoking quantity in multiple GWAS identifies cholinergic and sensory pathways

Oscar Harari PhD[1], Jen-Chyong Wang PhD[1], Kathleen Bucholz PhD[1], Howard J. Edenberg PhD[2], Andrew Heath DPhil[1], Nicholas G. Martin PhD[3], Michele L. Pergadia PhD[1], Grant Montgomery PhD[4], Andrew Schrage MS[1], Laura J. Bierut MD[1], Pamela F. Madden PhD[1],and Alison M. Goate DPhil[1,*]

[1]Department of Psychiatry, Washington University School of Medicine, St. Louis, Missouri, USA

[2] Department of Biochemistry and Molecular Biology, School of Medicine, Indiana University, Indianapolis, Indiana

[3] Genetic Epidemiology, Queensland Institute of Medical Research, Brisbane, Australia

[4] Molecular Epidemology, Queensland Institute of Medical Research, Brisbane, Australia

Supporting Text S1 - Samples and study design

Study of Addiction: Genetics and Environment (SAGE). This dataset includes unrelated individuals of European American (69%) and African American (31%) descent. We analyzed the EA subjects, with reported CPD values, selected from three studies: the Collaborative Genetic Study of Nicotine Dependence (N=1063), the Collaborative Study in Genetics of Alcoholism (N=626), and the Family Study of Cocaine Dependence (N=325). The assessment was based on the Semi-Structured Assessment for the Genetics of Alcoholism harmonized across all three studies. The study design is described elsewhere: [1]; and can be accessed in dbGaP (study

accession phs000092.v1.p1). We analyzed the quantity smoked per day (question 4 of the Fagerstrom test for nicotine dependence (FTND) [2]), for the subjects with European ancestry (Table 1). The Institutional Review Board at each contributing institution reviewed and approved the protocols for genetic studies under which all subjects were recruited. Subjects provided written informed consent for genetic studies and agreed to have their DNA and phenotypic information available to qualified investigators [1]

Nicotine Addiction Genetics (OZALC-NAG) [3]. This study includes participants enrolled at the Queensland Institute of Medical Research (QIMR) in Australia (Table 1) from a pool of families identified through diagnostic interview surveys of two cohorts of the Australian twin panel, which included spouses of the older of these two cohorts, for a total of about 12,500 families with information about smoking. The ancestry of the Australian samples is predominantly Anglo-Celtic or northern European (>90%). Index cases from these families, their full siblings, and parents were recruited for three coordinated studies: 1) the Nicotine Addiction Genetics (OZALC-NAG) Study, which ascertained heavy smoking index cases [4]; 2) the Australian Alcohol Extreme Discordant and Concordant Sibship (OZALCX-EDAC) study, which ascertained index cases with a history of alcohol dependence or heavy drinking (operationalized as in [5]); and 3) the Australian Alcohol Large Sibship (OZALC-BIGSIB) study, which ascertained large sibships, regardless of sibling phenotypic values [4, 3] Telephone interview survey data included FTND and DSM-IV-based assessments of nicotine dependence, as well as measures of quantity and

frequency of cigarette use. In this study, we use a measure of quantity of cigarettes smoked on average per day during the heaviest period of smoking. All data collection procedures were approved by institutional review boards at Washington University and the Queensland Institute of Medical Research. If the subject was an index case, permission was requested to contact other family members.

Atherosclerosis Risk in Communities study (ARIC). This is a multicenter population-based study designed to investigate the etiology of atherosclerosis in middle-aged adults [6], recruited from four U.S. communities. We used the average estimated number of cigarettes smoked per day across the entire time the subject smoked; using the values from the fourth question of the FTND (≤10; 11-20; 21-30; and >30). We selected only those subjects with European ancestry (Table 1). This study is part of the Gene Environment Association Studies initiative funded by the trans-NIH Genes, Environment, and Health Initiative (GEI). We analyzed the limited set of the data, accessed in dbGaP (Study Accession: phs000090.v1.p1). The study was approved by institutional review boards at each center, and all participants gave informed consent.

References

1. Bierut LJ, Agrawal A, Bucholz KK, Doheny KF, Laurie C, Pugh E, *et al.* A genome-wide association study of alcohol dependence. *Proc Natl Acad Sci USA* 2010 Mar. 16; **107**: 5082–5087.

2. Heatherton TF, Kozlowski LT, Frecker RC, Fagerstrom K-O. The Fagerstrom Test for Nicotine Dependence: a revision of the Fagerstrom Tolerance Questionnaire. *Addiction* 1991 Sep.; **86**: 1119–1127.

3. Heath AC, Whitfield JB, Martin NG, Pergadia ML, Goate AM, Lind PA, *et al.* A

Quantitative-Trait Genome-Wide Association Study of Alcoholism Risk in the Community: Findings and Implications. *Biol Psychiatry* 2011 Apr. 27;

4. Saccone S. Genetic Linkage to Chromosome 22q12 for a Heavy-Smoking Quantitative Trait in Two Independent Samples. *The American Journal of Human Genetics* 2007 May; **80**: 856–866.

5. Grant JD, Agrawal A, Bucholz KK, Madden PAF, Pergadia ML, Nelson EC, *et al.* Alcohol consumption indices of genetic risk for alcohol dependence. *Biol Psychiatry* 2009 Oct. 15; **66**: 795–800.

6. Sharrett AR. The Atherosclerosis Risk in Communities (ARIC) Study. Introduction and objectives of the hemostasis component. *Ann Epidemiol* 1992 Jul.; **2**: 467–469.

Pathway analysis of smoking quantity in multiple GWAS identifies cholinergic and sensory pathways

Oscar Harari PhD[1], Jen-Chyong Wang PhD[1], Kathleen Bucholz PhD[1], Howard J. Edenberg PhD[2], Andrew Heath DPhil[1], Nicholas G. Martin PhD[3], Michele L. Pergadia PhD[1], Grant Montgomery PhD[4], Andrew Schrage MS[1], Laura J. Bierut MD[1], Pamela F. Madden PhD[1],and Alison M. Goate DPhil[1,*]

[1]Department of Psychiatry, Washington University School of Medicine, St. Louis, Missouri, USA
[2] Department of Biochemistry and Molecular Biology, School of Medicine, Indiana University, Indianapolis, Indiana
[3] Genetic Epidemiology, Queensland Institute of Medical Research, Brisbane, Australia
[4] Molecular Epidemology, Queensland Institute of Medical Research, Brisbane, Australia

Supporting Text S2 - Genotypes

SAGE. Genotyping was performed using Illumina Human1Mv1_C BeadChips. Data cleaning and quality-control standards are described in detail in [1]. We analyzed 391,003 SNPs with a minor allele frequency (MAF) > 5% in populations of European descent, according to the human genome assembly build 37.1.

OZALC-NAG. Genotyping was conducted on Illumina platforms, including the Human 317K, the Human CNV370-Quadv3, and the Human 610-Quad. A detailed description of the genotypic data, data cleaning and quality-control measures is

available in[2]. We analyzed the SNPs genotyped in at least 2000 subjects, and with a MAF > 5%. Employing the human genome assembly build 37.1, a total of 154,477 SNPs met these criteria. All of these SNPs were also present on the Illumina Human 1M platform. We extended the set of SNPs analyzed to match the ones evaluated on the Illumina Human 1M by including imputing SNPs. Imputation was performed using MaCH[3] and HapMap samples of European ancestry (CEU; build 36, release 22) as the reference population[2].

ARIC. Genotyping was performed using the Affymetrix Genome-Wide Human SNP Array 6.0. Data cleaning and quality-control standards are described elsewhere: [4]. We analyzed 96,902 SNPs with a MAF > 5% in European descent populations that were also included in the Illumina Human1M. We extended this set of SNP including imputed SNPs to match the ones genotyped in the Illumina Human 1M chip. Imputation was performed at the GENEVA coordination center[5], executing BEAGLE version 3.3[6] on the full set of SNPs genotyped and employing HapMap Phase 3 samples of European ancestry (CEU + TSI) as the reference population[7] .

References

1. Bierut LJ, Agrawal A, Bucholz KK, Doheny KF, Laurie C, Pugh E, *et al.* A genome-wide association study of alcohol dependence. *Proc Natl Acad Sci USA* 2010 Mar. 16; **107**: 5082–5087.

2. Medland SE, Nyholt DR, Painter JN, McEvoy BP, McRae AF, Zhu G, *et al.* Common variants in the trichohyalin gene are associated with straight hair in Europeans. *Am J Hum Genet* 2009 Nov.; **85**: 750–755.

3. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet*

Epidemiol 2010 Dec.; **34**: 816–834.

4. Rasmussen-Torvik LJ, Alonso A, Li M, Kao W, Köttgen A, Yan Y, *et al.* Impact of repeated measures and sample selection on genome-wide association studies of fasting glucose. *Genet Epidemiol* 2010 Nov.; **34**: 665–673.

5. Cornelis MC, Agrawal A, Cole JW, Hansel NN, Barnes KC, Beaty TH, *et al.* The gene, environment association studies consortium (GENEVA): maximizing the knowledge obtained from GWAS by collaboration across studies of multiple conditions. *Genet Epidemiol* 2010 Jan. 20; **34**: 364–372.

6. Browning BL, Browning SR. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am J Hum Genet* 2009 Feb.; **84**: 210–223.

7. GENEVA Coordinating Center. GENEVA ARIC project imputation report [Internet]. 2010. Available from: https://www.genevastudy.org/sites/www/content/files/datacleaning/imputation/ARIC_Imputation_Report.pdf

Pathway analysis of smoking quantity in multiple GWAS identifies cholinergic and sensory pathways

Oscar Harari PhD[1], Jen-Chyong Wang PhD[1], Kathleen Bucholz PhD[1], Howard J. Edenberg PhD[2], Andrew Heath DPhil[1], Nicholas G. Martin PhD[3], Michele L. Pergadia PhD[1], Grant Montgomery PhD[4], Andrew Schrage MS[1], Laura J. Bierut MD[1], Pamela F. Madden PhD[1],and Alison M. Goate DPhil[1,*]

[1]Department of Psychiatry, Washington University School of Medicine, St. Louis, Missouri, USA

[2] Department of Biochemistry and Molecular Biology, School of Medicine, Indiana University, Indianapolis, Indiana

[3] Genetic Epidemiology, Queensland Institute of Medical Research, Brisbane, Australia

[4] Molecular Epidemology, Queensland Institute of Medical Research, Brisbane, Australia

Supporting Text S3 - Statistical analysis

Association *p-values* were calculated for unrelated samples datasets (SAGE and ARIC) employing linear regression models implemented in PLINK v1.07 [1], including gender as a covariate. In the case of the SAGE dataset, we also adjusted for alcohol and cocaine dependence and age, which was represented as quartiles (<35; 35-40; 40-45; and ≥45) [2]. We employed MERLIN v1.1.2 [3] to perform the association test for the family-based OZALC-NAG dataset (fastAssoc option), adjusting for age and gender.

References

1. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007 Sep.; **81**: 559–575.

2. Bierut LJ, Agrawal A, Bucholz KK, Doheny KF, Laurie C, Pugh E, *et al.* A genome-wide association study of alcohol dependence. *Proc Natl Acad Sci USA* 2010 Mar. 16; **107**: 5082–5087.

3. Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet* 2002 Jan.; **30**: 97–101.