Supplementary Information for:

Genome-wide meta-analyses identify three loci associated with cigarettes smoked per day.

Tobacco and Genetics (TAG) Consortium

Table of Contents

Nature Genetics: doi:10.1038/ng.571

Supplementary Table 1. Summary of prior GWAS of smoking behavior.

Study	Phenotype	N	SNP	Chr	Position	Gene	P-value
Liu et al, 2009 Molecular Psychiatry	Ever vs never smoking (any)	840	rs4956302	4	142,684,172	*IL15*	$1.19X10^{-6}$
			rs17354547	4	142,610,319	*IL15*	$5.61X10^{-6}$
			rs1402812	4	142,661,116	*IL15*	$6.67X10^{-6}$
			rs2036627	1	152,936,201	*KCNN3*	$7.24X10^{-6}$
			rs6009041	22	45,674,541	*TBC1D22*	$7.71X10^{-6}$
Caporaso et al, 2009 PLOS One	Ever vs never smoking	4,342	rs1402279	12	76,231,853		$5.20X10^{-6}$
			rs933688	5	90,798,504	*LOC133789*	$5.70X10^{-6}$
			rs1889899	9	26,779,940		$5.70X10^{-6}$
			rs6452953	5	90,758,790		$7.00X10^{-6}$
			rs6862125	5	90,760,354		$8.50X10^{-6}$
Vink et al, 2009 AJHG	Ever vs never smoking	3,497	rs10794595	10	124,871,492		$4.33X10^{-6}$
			rs16860281	3	149,478,045		$7.34X10^{-6}$
			rs8019291	14	92,000,917	*SLC24A4*	$8.08X10^{-6}$
			rs6997956	8	136,553,572	*KHDRBS3*	$8.47X10^{-6}$
			rs10865016	2	98,447,822	*CNGA3*	$9.67X10^{-6}$
Thorgeirsson et al, 2008 Nature	CPD (categorical)	10,995	rs1051730	15	76,681,394	*CHRNA3*	$5.00X10^{-16}$
			rs8034191	15	76,593,078	*LOC123688*	$1.00X10^{-15}$
			rs4887077	15	76,765,419		$8.00X10^{-10}$
			rs11638372	15	76,770,614		$1.00X10^{-9}$
			rs1996371	15	76,743,861		$3.00X10^{-9}$
Caporaso et al, 2009 PLOS One	CPD (quantitative)	2,617	rs6437740	3	108,948,507	*BBX*	$2.40X10^{-7}$
			rs910696	1	30,236,689		$3.00X10^{-6}$
			rs10411195	19	19,897,176	*ZNF505*	$5.80X10^{-6}$
			rs7050529	X	110,961,378	*TRPC5*	$6.20X10^{-6}$
			rs758642	17	3,733,656	*CAMKK1*	$7.30X10^{-6}$
Berrettini et al, 2008 Molecular Psychiatry	CPD (quantitative)	7,500	rs5522	4	149,576,925	*NR3C2*	$1.52X10^{-5}$
			rs10869409	9	76,313,209	*RORB*	$5.53X10^{-5}$
			rs6495308	15	76,694,711	*CHRNA3*	$6.9X10^{-5}$
			rs7804771	7	136,783,133	*DGK1*	$9.81X10^{-5}$
Caporaso et al, 2009 PLOS One	CPD (binary <10 CPD vs 10+ CPD, smokers only)	2,617	rs3112740	16	7,776,298		$6.00X10^{-6}$
			rs2268983	14	68,478,450	*ACTN1*	$6.70X10^{-6}$
			rs3027409	X	43,363,287	*MAOA*	$6.70X10^{-6}$
			rs886716	7	26,330,858		$7.70X10^{-6}$
			rs4722613	7	26,385,573	*LOC441205*	$9.30X10^{-6}$
Caporaso et al, 2009 PLOS One	Former vs current smoking	2,617	rs10989661	9	101,702,423		$6.30X10^{-6}$
			rs1847461	12	89,579,976		$8.20X10^{-6}$
			rs10859032	12	89,599,693		$8.40X10^{-6}$
Vink et al, 2009 AJHG	Current vs former + non- smokers	3,497	rs6484176	11	25,744,207		$8.83X10^{-7}$
			rs4741746	9	2,608,969	*VLDLR*	$2.98X10^{-6}$
			rs6661946	1	233,049,149	*HEATR1*	$4.11X10^{-6}$
			rs2404646	4	136,443,886		$6.48X10^{-6}$
			rs4758405	11	6,281,210		$8.41X10^{-6}$
Caporaso et al, 2009 PLOS One	Age at Initiation	2,617	rs11082304	18	18,974,971	*CABLES1*	$6.00X10^{-6}$
			rs17050782	4	140,780,739	*SET7*	$8.40X10^{-6}$
Bierut et al, 2006 Human Molecular Genetics	Nicotine dependence among smokers (FTND = 0 vs FTND = 4+)	1,929	rs2836823	21	39,302,119		$1.53X10^{-6}$
			rs4142041	10	68,310,957	*CTNNA3*	$5.64X10^{-6}$
			rs999	6	32,261,864	*PBX2*	$1.42X10^{-5}$
			rs12623467	2	51,136,740	*NRXN1*	$1.48X10^{-5}$
Caporaso et al, 2009 PLOS One	Pack years	2,617	rs800082	3	145,822,903		$3.30X10^{-6}$
			rs9289678	3	145,789,794		$3.70X10^{-6}$

Nature Genetics: doi:10.1038/ng.571

Caporaso et al, 2009 PLOS One	Duration of smoking	2,617	rs7553864	1	87,325,379	*AK002179*	$2.70X10^{-6}$
			rs719015	1	87,323,731	*AK002179*	$3.80X10^{-6}$
			rs912969	13	102,665,105		$7.80X10^{-6}$
			rs950063	4	126,789,524		$9.00X10^{-6}$

Nature Genetics: doi:10.1038/ng.571

Supplementary Table 2. Genotyping, imputation, and statistical analysis for TAG studies.

Study	Genotyping					Imputation			Association analyses		
	Platform	Inclusion criteria			SNPs met QC criteria	Imputation software	Inclusion criteria		SNPs in meta-analysis	λ_{GC}	Analyses software
		MAF	Call rate	P HWE			MAF	Imputation quality			
Population-based cohort studies											
ARIC	Affymetrix 6.0	≥1%	≥95%	$\geq 10^{-5}$	708,116	MACH	≥1%	Rsq≥0.7	2,337,252 (f) 2,337,252 (m)	1.034 (f) 1.007 (m)	ProbABEL
BLSA	Illumina HumanHap550	≥1%	>99%	$\geq 10^{-4}$	531,689	MACH	≥1%	Rsq≥0.7	2,368,916 (f) 2,369,031 (m)	1.025 (f) 1.029 (m)	MERLIN
CHS	Illumina HumanHap-370K	-	≥97%	$\geq 10^{-5}$	306,655	BimBam	-	oevar_imp≥0.7	1,714,818 (f) 1,714,504 (m)	1.019 (f) 1.014 (m)	R
InCHIANTI	Illumina HumanHap550 V1 & V3	≥1%	>97%	$\geq 10^{-4}$	496,032	IMPUTE	≥1%	Proper_info≥0.7	2,375,251 (f) 2,375,963 (m)	1.017 (f) 1.012 (m)	SNPTEST
Rotterdam	Illumina HumanHap 550	≥1%	≥90%	$\geq 10^{-5}$	530,683	MACH	≥1%	Rsq≥0.7	2,358,314 (f) 2,358,311 (m)	1.029 (f) 1.003 (m)	ProbABEL
FHS	Affymetrix 500K and MIPS 50K combined	-	≥97%	$\geq 10^{-6}$	503,551	MACH	-	Rsq≥0.7	2,543,273 (f) 2,543,200 (m)	1.011 (f) 1.047 (m)	LMEKIN
WGHS	Illumina HumanHap300 Duo+	≥1%	≥90%	$\geq 10^{-6}$	339,913	MACH	≥1%	Rsq≥0.7	2,160,712 (f)	1.102 (f)	ProbABEL
Case-control studies											
ADVANCE	Illumina 550K	-	≥95%	$\geq 10^{-3}$	557,724(?)	BimBam	≥1%	Proper_info≥0.7	1,899,305 (f) 1,895,997 (m)	1.010(f) 1.001 (m)	SNPTEST
ATVB	Affymetrix 6.0	≥1%	≥95%	$\geq 10^{-5}$	614,175	MACH	≥1%	Rsq≥0.7	2,282,340 (f) 2,282,478 (m)	1.006 (f) 1.017 (m)	ProbABEL
DGI	Affymetrix 500K	≥1%	≥95%	$\geq 10^{-6}$	389,057	MACH	≥1%	Rsq≥0.7	2,022,668 (f) 2,022,494 (m)	1.08 (f) 0.998 (m)	ProbABEL
FUSION	Illumina HumanHap300	-	≥90%	$\geq 10^{-6}$	315,635	MACH	≥1%	Rsq≥0.7	2,179,752 (f) 2,179,752 (m)	1.021 (f) 1.024 (m)	R glm function
IARC	Illumina HumanHap 300	-	≥95%	$\geq 10^{-7}$	310,023	MACH	≥1%	Rsq≥0.7	2,240,416 (f) 2,242,083 (m)	1.021 (f) 1.02 (m)	ProbABEL
MIGen	Affymetrix 6.0	≥1%	≥95%	$\geq 10^{-6}$	727,496	MACH	≥1%	Rsq≥0.7	2,316,220 (f) 2,316,220 (m)	1.088 (f) 1.088 (m)	SNPTEST
NHS	Illumina 550k	≥1%	≥89%	-	528,173	MACH	≥1%	Rsq≥0.7	2,372,155 (f)	1.011 (f)	ProbABEL
NTR/NESDA	Perlegen 600K	≥1%	≥ 95%	-	427,037	IMPUTE	≥1%	Proper_info≥0.7	2,349,830 (f) 2,349,711 (m)	1.008 (f) 1.011 (m)	SNPTEST
GAIN controls	Affymetrix 6.0	≥1%	≥95%	$\geq 10^{-5}$	704,216	MACH	≥1%	Rsq≥0.7	2,100,033 (f) 2,100,032 (m)	1.001 (f) 1.024 (m)	ProbABEL

Lambda is shown for smoking initiation (ever versus never smokers) stratified by sex. MAF=minor allele frequency, HWE=Hardy Weinberg Equilibrium, f=female, m=male.

Nature Genetics: doi:10.1038/ng.571

Supplementary Table 3. Sample quality control for studies in the TAG Consortium.

Study	Sample QC		Sample size included
	Call rate	Other exclusions	
Population-based cohort studies			
ARIC	>90%	1) sex discrepancy with genetic data from X-linked markers 2) first and second degree relatives and ancestry outliers 3) missing phenotype information	8,330
BLSA	>97%	1) sex discrepancy with genetic data from X-linked markers 2) ancestry outliers 3) duplicates and related individuals 4) missing phenotype information	856
CHS	>95%	1) sex discrepancy with genetic data from X-linked markers 2) missing phenotype information	3,236
InCHIANTI	>97%	1) sex discrepancy with genetic data from X-linked markers 2) heterozygosity <30% 3) missing phenotype information	1,200
Rotterdam	>97.5%	1) sex discrepancy with genetic data from X-linked markers 2) duplicates and first and second degree relatives with IBS 3) ancestry outliers 4) missing phenotype information	5,610
FHS	>97%	1) missing phenotype information	7,257
WGHS	>98%	1) ancestry outliers 2) missing phenotype information	22,037
Case-control studies			
ADVANCE	>98.5%	1) sex discrepancy with genetic data from X-linked markers 2) ancestry outliers 3) duplicates and related individuals 4) missing phenotype information	585
ATVB	>90%	1) unusual autosomal homozygosity 2) first and second degree relatives and ancestry outliers 3) missing phenotype information	3,260
DGI	>95%	1) sex discrepancy with genetic data from X-linked markers 2) duplicates and related individuals 3) missing phenotype information	2,504
FUSION	≥97.5%	1) missing phenotype information	1,055
IARC	>95%	1) sex discrepancy with genetic data from X-linked markers 2) ethnic outliers 3) duplicates and related individuals 4) missing phenotype information	8,381
MIGen	>95%	1) heterozygosity check 2) duplicates and related individuals 3) ethnic outliers 4) missing phenotype information	2,647
NHS	>90%	1) duplicates and ethnic outliers 2) missing phenotype information	2,249
NTR/NESDA	>95%	1) sex discrepancy with genetic data from X-linked markers 2) ethnic outliers 3) duplicates and first and second degree relatives with IBS 4) missing phenotype information	3,438
GAIN controls	>90%	1) first and second degree relatives and ethnic outliers 2) unusual autosomal homozygosity 3) missing phenotype information	1,390

Nature Genetics: doi:10.1038/ng.571

Supplementary Table 4. Per study findings for the most significant SNP for each smoking phenotype.

Phenotype	Top SNP	Chromomsome (bp position)	Nearby genes	Alleles	Study	Sample size (*N* case/*N* control)	λ_{GC}	λ_{1000}	Beta (SE)	*P*-value
Ever vs. never smokers	rs16941640 (I^2=0%)	17 (42,619,520)	*CDC27* *MYL4*	A/T	ARIC	8330 (5033/3297)	1.021	1.005	0.0684 (0.066)	0.3023
					BLSA	856 (462/394)	1.027	1.063	0.2839 (0.222)	0.2014
					CHS	NA	NA	NA	NA	NA
					InCHIANTI	1200 (527/673)	1.015	1.025	0.2634 (0.236)	0.2648
					Rotterdam	5610 (3323 /2287)	1.015	1.006	0.0841 (0.088)	0.3396
					FHS	7257 (3932/3325)	1.029	1.008	0.2989 (0.081)	0.0002
					WGHS	22037 (10832/11205)	1.102	1.009	0.1098 (0.040)	0.0059
					ADVANCE	585 (279/306)	1.006	1.021	-0.1933 (0.242)	0.4235
					ATVB	3260 (2220/1040)	1.012	1.008	0.2153 (0.139)	0.1217
					DGI	2504 (946/1558)	1.037	1.031	0.2605 (0.127)	0.0403
					FUSION	1055 (494/561)	1.023	1.044	0.2256 (0.183)	0.2182
					IARC	8381 (6305/2076)	1.021	1.007	-0.0129 (0.089)	0.8858
					MIGen	2647 (1701/946)	1.088	1.072	0.2632 (0.116)	0.0236
					NHS	2249 (1211/1038)	1.011	1.010	0.0762 (0.126)	0.5454
					NESDA/NTR	3438 (2231/1207)	1.009	1.006	0.1613 (0.109)	0.1407
					GAIN	NA	NA	NA	NA	NA
					Overall	**69409 (39022/30387)**	**1.092**	**1.003**	**0.1298 (0.025)** ***0.1298 (0.025)***	**$2.21X10^{-7}$** ***$2.21X10^{-7}$***
CPD	rs12914385 (I^2=60%)	15 (76,685,778)	*CHRNA3* *CHRNA5* *CHRNB4*	C/T	ARIC	5002	0.991		-1.4733 (0.226)	$6.77X10^{-11}$
					BLSA	NA	NA		NA	NA
					CHS	1642	1.04		-0.7448 (0.431)	0.0837
					InCHIANTI	527	1.013		-0.4702 (0.498)	0.3454
					Rotterdam	3298	1.009		-0.6275 (0.279)	0.0243
					FHS	3932	1.005		-1.3328 (0.248)	$7.62X10^{-8}$
					WGHS	10668	1.017		-1.2981 (0.158)	$2.22X10^{-16}$
					ADVANCE	76	1.041		1.6565 (1.840)	0.3680
					ATVB	2190	0.996		-1.5730 (0.402)	$9.20X10^{-5}$
					DGI	NA	NA		NA	NA
					FUSION	443	0.996		0.4372 (0.814)	0.5910
					IARC	6280	1.003		-0.1098(0.202)	0.5861
					MIGen	NA	NA		NA	NA
					NHS	1211	1.014		-1.4378 (0.430)	$8.34X10^{-4}$
					NESDA/NTR	1928	1.007		-1.5452 (0.327)	$2.33X10^{-6}$

Nature Genetics: doi:10.1038/ng.571

					GAIN	984	1.015		-0.4524 (0.748)	0.5453
					Overall	**38181**	**1.055**		**-1.0235 (0.083)** ***-1.0497 (0.153)***	**$4.23X10^{-35}$** *$7.40x10^{-12}$*
Former vs. Current smokers	rs7872903 (I^2=0%)	9 (135,474,113)	*DBH* *ADAMTS L2* *FAM163B* *SARDH*	C/T	ARIC	4905 (2815/2090)	1.002	1.001	-1.5452 (0.327)	$2.33X10^{-6}$
					BLSA	NA	NA	NA	NA	NA
					CHS	1651 (1285/366)	1.011	1.019	-0.1333 (0.108)	0.2158
					InCHIANTI	527 (300/227)	1.014	1.054	0.0487 (0.163)	0.7646
					Rotterdam	3256 (2040/1216)	0.998	0.999	-0.0786 (0.062)	0.2016
					FHS	NA	NA	NA	NA	NA
					WGHS	10410 (7831/2579)	1.055	1.013	-0.1523 (0.038)	$5.46X10^{-5}$
					ADVANCE	279 (182/97)	1.034	1.269	-0.3274 (0.211)	0.1216
					ATVB	2207 (470/1737)	0.99	0.986	-0.0790 (0.098)	0.4196
					DGI	NA	NA	NA	NA	NA
					FUSION	455 (321/134)	1.014	1.068	-0.1594 (0.202)	0.4307
					IARC	6296 (1974 /4322)	1.018	1.007	-0.0968 (0.049)	0.0483
					MIGen	1701 (699/1002)	1.071	1.086	-0.1022 (0.085)	0.2300
					NHS	1211 (1074/137)	0.997	0.988	0.0299 (0.159)	0.8511
					NESDA/NTR	2187 (1150/1037)	1.024	1.022	-0.1038 (0.075)	0.1658
					GAIN	760 (478/282)	1.018	1.051	-0.0164 (0.130)	0.90
					Overall	**35845 (20619/15226)**	**1.051**	**1.003**	**-0.1111 (0.020)** ***-0.1111 (0.020)***	**$5.55X10^{-8}$** ***$5.55X10^{-8}$***
Age at initiation	rs2806464 (I^2=48%)	1 (230,269,046)	*DISC1*	C/T	ARIC	5016	0.999		0.0196 (0.005)	$2.33X10^{-4}$
					BLSA	426	1.015		0.0494 (0.019)	0.0119
					CHS	NA	NA		NA	NA
					InCHIANTI	527	1.023		0.0684 (0.036)	0.0567
					Rotterdam	3297	1.006		-0.0039 (0.007)	0.6006
					FHS	NA	NA		NA	NA
					WGHS	NA	NA		NA	NA
					ADVANCE	203	0.976		NA	NA
					ATVB	2167	1.007		-0.0039 (0.007)	0.5778
					DGI	679	0.983		0.0257 (0.016)	0.0982
					FUSION	457	1.005		-0.0007 (0.021)	0.9724
					IARC	6274	1.007		0.0155 (0.005)	0.0022
					MIGen	NA	NA		NA	NA
					NHS	1205	1.02		0.0206 (0.007)	0.0053
					NESDA/NTR	2187	1.01		0.0062 (0.009)	0.4937
					GAIN	NA	NA		NA	NA

Nature Genetics: doi:10.1038/ng.571

					Overall	**22438**	**1.05**		**0.0123 (0.003)** ***0.0131 (0.004)***	**$1.60X10^{-6}$** *$1.9x10^{-3}$*

Supplementary Table 4. The most significant SNP associated with each smoking phenotype are presented in this table. In each study, λ_{GC}, λ_{1000}, regression beta and its standard error, and *P*-value are based on sex combined, fixed effect meta-analyses. The overall beta, standard error, and *P*-value are obtained by multiplying the λ_{GC} factors specific to each study calculated separately for each sex by the variance of the beta estimates. Nearby genes are within 50kb of the index SNP. Betas, standard errors and *p*-values from random effects meta-analyses are shown in italics.

NA=Not available (not included in meta-analysis).

Nature Genetics: doi:10.1038/ng.571

Supplementary Table 5. Regions selected for follow-up for each smoking phenotype.

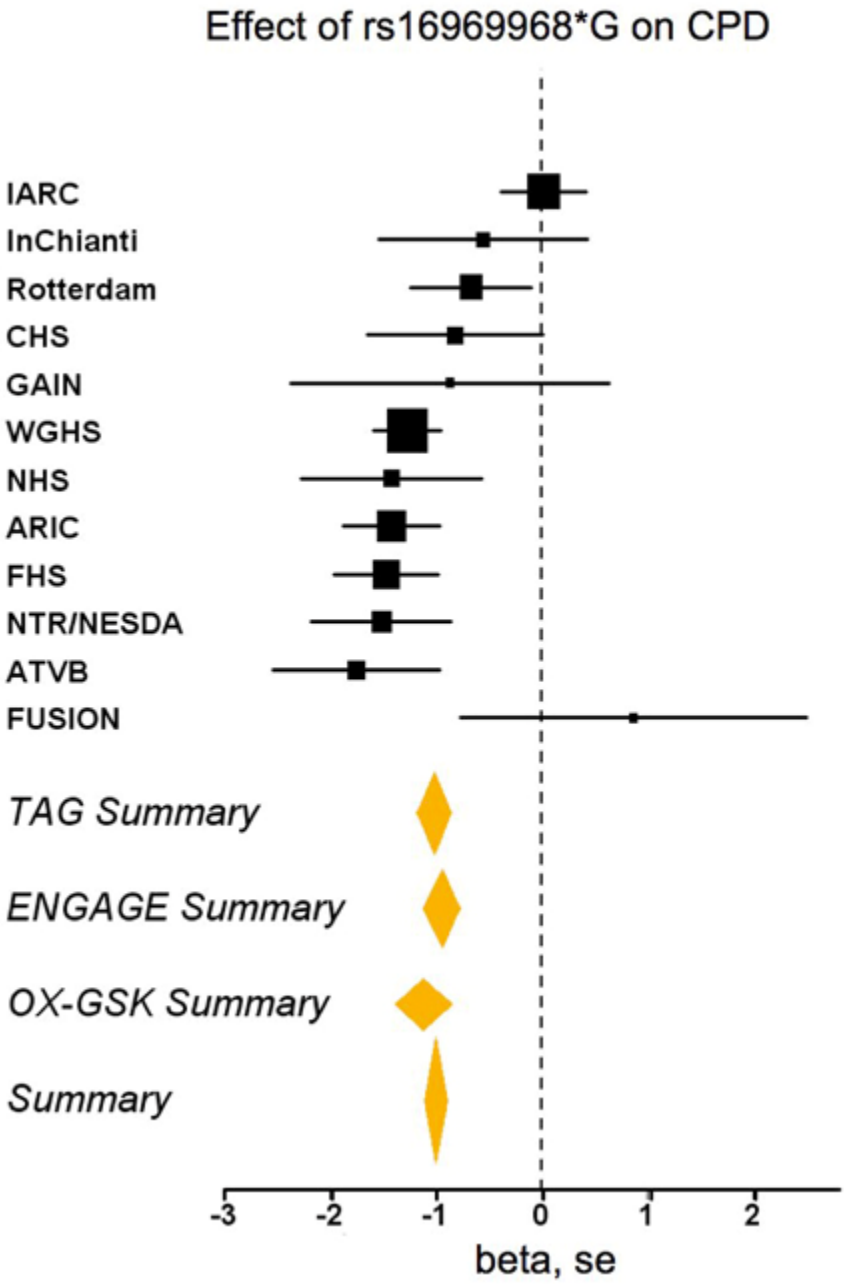
Phenotype	Chr	Start	End	Genes in region (±50 kb)	SNPs with $p<1X10^{-4}$	Most significant SNP
Smoking initiation	1	41,528,143	41,535,605	*SCMH1*	10	rs7548367
	4	11,935,089	12,027,840	< intergenic >	73	rs4377605
	4	63,211,114	63,222,973	< intergenic >	2	rs10013579
	4	159,069,274	159,265,293	*GRIA2, C4orf18, TMEM144*	24	rs1839129
	5	167,274,172	167,640,026	*ODZ2, WWC1*	10	rs725695
	8	3,465,461	4,160,033	*CSMD1*	20	rs2449222
	8	8,759,788	8,763,956	*MFHAS1*	3	rs10108954
	8	59,471,286	59,998,555	*UBXN2B, TOX*	30	rs7000535
	11	27,508,555	27,685,115	*LIN7C, BDNF*	31	rs11030084
	12	36,528,296	37,411,158	*ALG10B, CPNE8*	46	rs1817648
	12	113,653,862	113,668,436	*TBX3, CR591392*	7	rs11067275
	12	130,772,542	130,843,178	*SFRS8, MMP17*	9	rs3782288
	16	59,086,821	59,138,599	< intergenic >	29	rs11076393
	17	42,327,809	42,643,770	*GOSR2, WNT9B, CDC27, MYL4, ITGB3*	14	rs16941640
CPD	1	77,740,944	78,319,686	*AK5, NEXN, FUBP1, DNAJB4, GIPC2*	9	rs11162413
	1	200,058,929	200,154,344	*NAV1, IPO9, SHISA4, hsa-mir-1231, LMOD1, TIMM17A*	5	rs10920270
	3	26,541,958	26,612,410	*LRRC3B*	15	rs9856956
	4	57,519,897	57,525,738	*C4orf14, POLR2B, REST*	3	rs781668
	4	157,393,006	157,396,195	< intergenic >	3	rs4691319
	6	79,958,973	79,991,150	*HMGN3*	13	rs989191
	8	20,775,925	20,832,225	< intergenic >	13	rs11204140
	8	95,021,903	95,022,756	*PPM2C*	4	rs6983462
	8	128,844,326	128,868,185	*MYC*	6	rs4733560
	8	134,870,418	134,875,728	< intergenic >	7	rs7812376
	9	79,738,617	80,023,993	*GNAQ, CEP78*	19	rs7037778
	10	93,328,580	93,350,311	*PPP1R3C, LOC100188947*	6	rs1028936
	11	122,113,056	122,117,002	*UBASH3B*	9	rs4421757
	15	76,685,778	76,960,060	*IREB2, LOC123688*, PSMA4, CHRNA5*, CHRNA3, CHRNB4, ADAMTS7*, MORF4L1, CTSH*	174	rs12914385
	19	46,002,411	46,100,586	*CYP2A6, EGLN2, MIA, RAB4B, SNRPA, CYP2A7*	7	rs4802091
Smoking cessation	1	237,738,574	237,813,355	*CHRM3*	10	rs1782352
	2	29,548,434	29,653,986	*ALK*	30	rs10779970
	2	44,500,047	44,602,979	*C2orf34*	11	rs698782

Nature Genetics: doi:10.1038/ng.571

	2	182,683,202	182,703,718	*PDE1A, PPP1R1C*	22	rs10205020
	6	145,078,043	145,292,919	*UTRN*	41	rs6570644
	7	39,064,851	39,147,068	*POU6F2*	15	rs10234489
	8	34,911,658	35,003,823	< intergenic >	58	rs17250591
	8	92,427,905	92,445,365	*SLC26A7*	8	rs6471277
	9	135,468,176	135,475,419	*DBH, ADAMTSL2, FAM163B, SARDH*	10	rs3025360
	11	28,547,744	28,666,010	< intergenic >	69	rs2582895
	11	101,117,645	101,261,411	*ANGPTL5, KIAA1377*	46	rs10895192
	14	36,577,540	36,713,381	*SLC25A21, MIPOL1**	53	rs17178639
	15	45,623,301	45,758,743	*SEMA6D*	34	rs1797227
	15	76,499,754	76,715,319	*TBC1D2B, IREB2, PSMA4, LOC123688, CHRNA5, CHRNA3, CHRNB4*	36	rs4362358
	20	11,095,726	11,114,172	< intergenic >	8	rs6134154

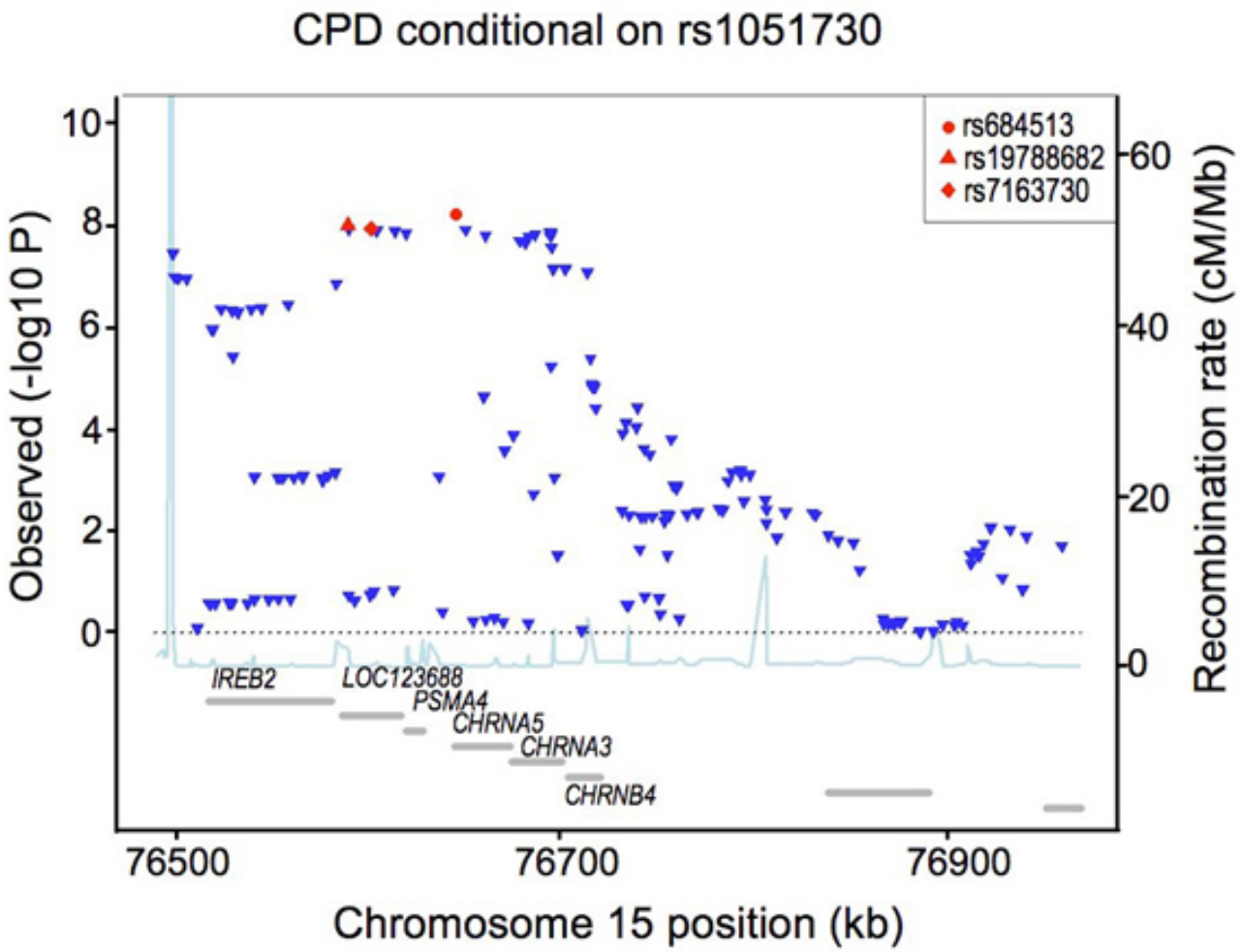
Nature Genetics: doi:10.1038/ng.571

Supplementary Figure 1. Forest plot: Effect of rs16969968*G on CPD. Forest plot of the association between rs16969968 and CPD in each study of the TAG Consortium, each consortium, and meta-analysis across the three consortia.



Nature Genetics: doi:10.1038/ng.571

Supplementary Figure 2. Regional plot: CPD conditional on rs1051730. Regional association plots show SNPs plotted by position on chromosome against –log10 P-value with CPD, from the TAG Consortium. Estimated recombination rates (from HapMap-CEU) are plotted in light blue to reflect the local LD structure on a secondary y-axis. The three most significant SNPs are indicated as red diamonds. The gray bars at the bottom of the plot represent the relative size and location of genes in the region.



Nature Genetics: doi:10.1038/ng.571

Supplementary Note

TAG Consortium Overview.

The organizational structure of the TAG Consortium was comprised of an Executive Committee (EC), a Phenotype Working Group (PWG) chaired by Drs. Caryn Lerman and Jaakko Kaprio, and an Analytic Working Group (AWG) chaired by Drs. Peter Kraft, John Ioannidis, and Danyu Lin. The EC was responsible for developing general guidelines for collaboration, authorship, sharing of results, publication, and timely participation. The PWG harmonized the smoking phenotypes across studies by reviewing questionnaire information and smoking variable distributions from all studies and determining how the smoking phenotypes would be constructed. The AWG developed the uniform analytic plan for each of the smoking traits and specified protocols for assessing genotype quality control, testing for main effects, conducting sex-specific analyses, considering the effects of smoking-related phenotypes in the analyses and conducting and interpreting the meta-analyses. At least one representative from each study participated in the monthly teleconferences, and up to five co-authors from each study were named in the manuscript. All of the studies participating in the TAG Consortium were approved by their respective institutional review committee, and the subjects from all the studies provided written informed consent.

Summary of the studies participating in TAG.

1. Atherosclerosis Risk Communities Study (ARIC): The ARIC study is a population-based, prospective cohort study of cardiovascular disease and its risk factors sponsored by National Heart, Lung and Blood Institute (NHLBI) [1]. ARIC included 15,792 individuals aged 45-64 years at baseline (1987-89), chosen by probability sampling from four US communities [2]. Cohort members completed four clinic examinations, conducted three years apart between 1987 and 1998. Follow-up for clinical events was annual. The current analysis included 8330 males and females of European ancestry on whom baseline smoking information was available.
2. Baltimore Longitudinal Study of Aging GWAS (BLSA): BLSA is a population-based study aimed to evaluate contributors of healthy aging in the older population residing predominantly in the Baltimore-Washington DC area [3,4]. Starting in 1958, participants are examined every one to four years depending on their age. Currently there are approximately 1100 active participants enrolled in the study. Blood samples were collected for DNA extraction, and genome-wide genotyping was completed for 1231

Nature Genetics: doi:10.1038/ng.571

subjects using Illumina 550K. The analysis was limited to subjects of European decent with data on smoking behavior (N=856).

3. Cardiovascular Health Study (CHS): CHS is a population-based National Heart, Lung, and Blood Institute-funded cohort study of risk factors for cardiovascular disease in adults 65 years of age or older conducted at 4 field centers [5]. The original predominantly white cohort of 5,201 persons was recruited in 1989–1990 from random samples of Medicare lists. An additional 687 African-Americans were enrolled in 1992–1993. CHS participants completed standardized clinical examinations and questionnaires at study baseline and at 9 annual follow-up visits. Follow-up for clinical events occurs every 6 months. The current analysis included 3,236 white males and females without clinical cardiovascular disease (coronary heart disease, congestive heart failure, peripheral vascular disease, valvular heart disease, stroke or transient ischemic attack) at baseline on whom baseline smoking information was available [6].
4. Invecchiare in Chianti (InCHIANTI): InCHIANTI is a population-based epidemiological study aimed at evaluating factors that influence mobility in the older population living in the Chianti region of Tuscany, Italy[7]. Briefly, 1616 residents were selected from the population registry of Greve in Chianti (a rural area: 11,709 residents with 19.3% of the population greater than 65 years of age) and Bagno a Ripoli (Antella village near Florence; 4,704 inhabitants, with 20.3% greater than 65 years of age). The participation rate was 90% (n= 1,453) and participants ranged between 21–102 years of age. The study protocol was approved by the Italian National Institute of Research and Care of Aging Institutional Review. There were 85 parent-offspring pairs, 6 sib-pairs and 2 half-sibling pairs documented. Further familial relationships were investigated using IBD of 10,000 random SNPs with RELPAIR and uncovered 1 parent-offspring, 79 siblings and 13 half-sibling. The correct family structure inferred from genetic data was utilized for all analyses.
5. The Rotterdam Study (RS): RS was planned and designed in the early 1990s as a longitudinal study investigating the incidence and progression of diseases in the elderly [8,9]. From 1991 to 1995 all inhabitants of Ommoord, a district of Rotterdam in the Netherlands, who were 55 years or older, were invited to participate in this study (37). Of 10,275 eligible individuals, 7,983 agreed to participate (78%). Participants completed home interviews and underwent an extensive set of examinations in a research facility. Follow-up research center visits were repeated every 3-4 years and incident diseases were assessed through general practitioner (GP) medical records as well as

Nature Genetics: doi:10.1038/ng.571

hospitalization and prescription medication history. The RS has been approved by the institutional review board (Medical Ethics Committee) of the Erasmus Medical Center and by the review board of the Netherlands Ministry of Health, Welfare and Sports. All participants provided written informed consent. The current analysis included 5,610 participants on whom genotyping and baseline smoking information was available.

6. Framingham Heart Study (FHS): Details of study designs for the FHS cohorts have been previously published[10-12]. Briefly, the FHS Original Cohort began in 1948 with the enrollment of 5,209 men and women who were 28 to 62 years of age (mean age 44 years; 55% women) at their baseline visit. In 1971, 5,124 children and spouses of the children of the Original Cohort participants were enrolled into the Offspring cohort (mean age 37 years; 52% women). The Third Generation cohort began in 2002 with the enrollment of 4,095 children and their spouses (mean age 40 years; 53% women) of the Offspring cohort participants. The current analysis included 7,257 participants from the Offspring and Third Generation cohorts for whom smoking status and genotyping information was available.
7. Women's Genome Health Study (WGHS): The WGHS cohort is a study of 28,345 American female health professionals who are participants in the ongoing observational follow-up of the Women's Health Study (WHS), a randomized trial of aspirin and vitamin E in the prevention of cardiovascular disease and cancer completed in 2004. Eligible women were at least 45 years of age at baseline and free of cardiovascular disease, cancer and other major chronic illness at the time of enrollment (1993). Details of the rationale, design, and methodology of the WGHS and the WHS are described elsewhere [13,14]. The current analysis included 22,037 women of verified European ancestry for whom baseline smoking information was available.
8. Atherosclerotic Disease Vascular Function and Genetic Epidemiology (ADVANCE): ADVANCE is a large epidemiological study of genetic and non-genetic determinants of coronary artery disease (CAD) that started in 2000 as a collaborative effort between researchers at Stanford University and Kaiser Permanente of Northern California [15-18]. The overarching goal of the study is to improve our ability to prevent, diagnose and treat CAD. The ADVANCE study is a case-control study among adults (age ≥18 years) receiving medical care within KPNC. Briefly, between October 28, 2001 and December 31, 2003, a total of 3,179 case and control subjects were recruited. Cases consisted of subjects presenting with clinically significant CAD (MI or angina with angiogram showing

Nature Genetics: doi:10.1038/ng.571

at least one coronary artery stenosis of > 50%) at a young age (≤ 45 years for males, ≤ 55 years for females) or subjects presenting with incident stable angina or incident acute myocardial infarction (AMI) at an older age. Controls consisted of young subjects with no history of CAD (30 to 45 years for males, 30 to 55 for females) or subjects aged 60 to 72 with no history of CAD, cerebrovascular accident (CVA), or peripheral arterial disease (PAD). Young controls recruited de novo from Kaiser were complemented with a subset of 479 subjects from the Coronary Artery Risk Development in Young Adults (CARDIA) Study originally recruited at the Oakland field center and attending the study's year 15 examination in 2000-2001. During recruitment, some race/ethnic and gender strata were over sampled to maximize the probability that case and control groups were balanced in this respect. As part of the NHLBI sponsored STAMPEED initiative, GWA genotyping was performed using the Illumina 550K platform on a subset of ADVANCE subjects with early onset CAD (n = 513) as well as a similar number of controls (n = 527). As part of this manuscript, we supplied GWA data on 585 ADVANCE white/European subjects with data on smoking history.

9. Atherosclerosis, Thrombosis and Vascular Biology Italian Study Group (ATVB) – Premature MI cohort: ATVB was designed as a nationwide case-control study involving 125 Italian coronary care units[19]. Cases with premature MI were recruited whilst hospitalised for a first MI under the age of 45 years. All cases underwent coronary angiography. Controls were healthy subjects matched for sex, age and parental geographic origin and had no history of thromboembolic disease. Original enrolment involved 1,210 cases and 1,201 controls between 1998 and 2001, and this was extended to 1958 case and 1976 control samples in the period to 2005.
10. The Diabetes Genetics Initiative (DGI): The DGI sample consists of 1,462 Type 2 diabetes (T2D) cases and 1,467 normoglycemic controls from Sweden and Finland. Of these, 2,097 are population-based T2D cases and controls matched for BMI, gender and geographic origin, and 834 are T2D cases and controls in 326 sibships discordant for T2D[20]. The current analysis included all individuals from the population-based study component and one individual from each sibship, in total 2,504 males and females, on whom smoking information at inclusion was available.
11. The Finland-United States Investigation of NIDDM (FUSION) Study: FUSION aims to identify genetic variants that predispose to T2D or are responsible for variability in diabetes-related quantitative traits [21-23]. The study began as an affected sib pair family

Nature Genetics: doi:10.1038/ng.571

study and later expanded to include large numbers of T2D cases and controls for association analysis. The FUSION stage 1 GWAS was performed on a set of 1,161 Finnish T2D cases and 1,174 Finnish controls, approximately frequency matched on 5-year age category, sex, and birth province within Finland. The current analysis included 1,055 of these individuals on whom smoking data was available.

12. International Agency for Research on Cancer (IARC) contributed two cancer studies conducted in Europe to the TAG Consortium which were analyzed as one study (the IARC study in central Europe, and the Alcohol-Related Cancers and Genetic Susceptibility in Europe (ARCAGE) Study). The IARC study of lung cancer in central Europe was conducted with cancer institutes in 6 countries including Czech Republic (Prague, Olomouc, Brno), Hungary (Borsod, Heves, Szabolcs, Szolnok, Budapest), Poland (Warsaw, Lodz), Romania (Bucharest), Russia (Moscow) and Slovakia (Banska Bystrica, Bratislava, Nitra) between 1998 and 2002. Briefly, each centre followed an identical protocol and was responsible for recruiting a consecutive group of newly diagnosed cases of lung cancer and a comparable group of hospital or population controls. All subjects were interviewed based on a standard questionnaire which included information on lifestyle risk factors, occupational history, medical and family history. Written consent for participation was obtained from all study subjects and ethical approval has been obtained for all study centers as well as at IARC, the coordinating center. Further details on the questionnaire, as well as case and control recruitment, have been reported elsewhere[24]. Controls in all centers except Warsaw were chosen among subjects admitted as in-patients or outpatients in the same hospital as the cases with conditions unrelated to tobacco including minor surgical conditions, benign disorders, common infections, eye conditions (except cataract or diabetic retinopathy) and common orthopeadic diseases (except osteoporosis). In Warsaw, population controls were selected by random sampling from the Polish Electronic List of Residents. This resulted in a total of 3,052 potential controls. Cases and controls were frequency matched by sex, age (+/- 3 years), center, referral (or of residence) area and period of recruitment (+/- 6 months). The participation rates for both cases and controls were over 80% in all centres. Blood samples were collected from all subjects and DNA has been extracted. Candidate gene studies based on Taqman genotyping have been completed on over 100 candidate gene variants[25]. 1,989 lung cancer cases (434 adenocarcinoma, 815 squamous cell, 298 small cell and 379 of mixed cell or other histology) and 2,625 controls had sufficient DNA amount and quality and was genotyped with the Illumina

Nature Genetics: doi:10.1038/ng.571

HumanHap300 panel at Centre National Genotypage (CNG) in Paris [26,27]. We conducted a parallel study of head and neck cancer in 5 of the countries. For this study DNA was available for 771 cases recruited in Russia, Czech Republic, Romania, Hungary and Poland, and comprised of 249 oral cavity/pharynx, 328 larynx, 167 oesophagus, and 278 cases of cancers from overlapping sites. These were compared to an age-sex matched group of controls from the lung cancer study. 726 cases (238 oral cavity/pharynx, 312 larynx, 156 oesophagus and 20 with an overlapping site) and 694 controls were successfully genotyped. All upper aero-digestive cancer cases were restricted to squamous cell carcinoma, the predominant histological type. Finally a parallel study of kidney cancer was performed, consisting of 1,097 cases and 1,476 matched controls [28]. ARCAGE is a multicenter case-control study of head and neck cancer conducted by IARC from 2002 to 2005, in 12 centers from 8 European countries: Prague (Czech Republic), Athens (Greece), Aviano, Padova, Turin (Italy), Oslo (Norway), Glasgow, Manchester, Newcastle (United Kingdom), Barcelona (Spain), Zagreb (Croatia) and Paris (France). Head and neck cancer cases diagnosed at designated hospitals or cancer clinics and confirmed histologically or cytologically were recruited into the study within 3 months of diagnosis. Controls were recruited from the same hospitals as cases in all of the centers, except the UK centers where population-based controls were recruited. Only controls with a recent diagnosis of diseases that were not related to tobacco and alcohol were included. Cases and controls were interviewed with a structured questionnaire that included information on lifestyle and dietary habits as well as a series of questions relating to tobacco addiction based on the Fagerstrom test of nicotine addiction. 1,599 cases and 1,491 controls had sufficient amount of DNA for this study, of which 1,536 cases (749 oral cavity/pharynx, 574 larynx, 159 oesophagus and 54 cases with an overlapping site) and 1,443 controls were successfully genotyped [26,27].

13. Myocardial Infarction Genetics Consortium (MIGen): MIGen Consortium is a case-control study aimed at identifying genetic variants associated with early-onset myocardial infarction [2,29-31]. Most of the samples are selected from population based cross-sectional or cohort studies and population-based myocardial infarction registers from five different studies: Heart Attack Risk in Puget Sound (Seattle, USA), REGICOR (Girona, Spain), MGH Premature Coronary Artery Disease Study (Boston, USA), FINRISK (Finland); Malmö Diet and Cancer Study (Malmö, Sweden). All the participants were of European ancestry. The current analysis included 2,647 males and females from these studies for whom smoking information was available.

Nature Genetics: doi:10.1038/ng.571

14. Nurses' Health Study (NHS): Subjects of NHS were drawn from one previous genome-wide association study, performed as part of the Cancer Genetic Markers of Susceptibility (CGEMS) project[32]. Briefly, the NHS is a longitudinal study of 121,700 women enrolled in 1976. The CGEMS_NHS was a nested case-control study, including 1,145 postmenopausal breast cancer cases and 1,142 matched controls, derived from 32,826 NHS participants who provided a blood sample between 1989 and 1990 and were free of diagnosed breast cancer at blood collection and followed for incident disease until June 1, 2004. In this study we included 2,249 subjects (1,211 cases and 1,038 controls) with smoking behaviors data available. All subjects were of self-reported European ancestry, which was consistent with genetic analyses of population structure. All these samples were genotyped using the Illumina HumanHap 550k platform, Individual samples were removed if more than 10% of SNPs failed genotyping, and individual SNPs were removed if more than 10% of samples failed. The average call rate was 99.8%.
15. Netherlands Study of Depression and Anxiety (NESDA) and Netherlands Twin Registry (NTR): The two parent projects that supplied data for this sample, NESDA[33] and NTR[34], are large-scale longitudinal studies. The sample consisted of 1,777 participants from the NTR and 1763 participants from the NESDA. For participants of the NTR, longitudinal-survey data from seven waves of data collection (1991–2004) were used in determining the smoking behavior of participants. Current and past smoking behavior was assessed when DNA samples were collected (2004–2007) [35]. For participants from NESDA, DNA samples and data on smoking behavior were collected during a clinical interview between 2004 and 2007. The genotypic data were generated as part of one of the six initial Genetic Association Information Network (GAIN) studies sponsored by the Foundation for the National Institutes of Health (NIH). Sampling, data-collection characteristics, and genotype-cleaning procedures of the GAIN-major depressive disorder (MDD) study have been described in detail elsewhere[36,37].
16. Genetic Association Information Network (GAIN) controls: Controls were ascertained from a US national-sampling frame as part of the NIMH Genetics repository (MH059571, PI: Dr Pablo Gejman, release v4.0, June 2006). Controls were collected by Knowledge Networks, a survey and market research company whose panel contains approximately 60 000 households (>120,000 unrelated adults) [38]. Households were selected via random digit dialing and proportionally from 25 major US population areas and financial incentives were provided for participation. The Knowledge Networks panel is generally

Nature Genetics: doi:10.1038/ng.571

representative of the US population but with a slight bias toward higher income and education. The TAG meta-analyses included 1,390 individuals of European decent previously used as controls for the GAIN schizophrenia GWAS[39]. Smoking data was collected on these individuals as part of the Knowledge Networks questionnaire administered from 2003-2006.

Acknowledgements.

1. Atherosclerosis Risk Communities Study (ARIC): The Atherosclerosis Risk in Communities Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts N01-HC-55015, N01-HC-55016, N01-HC-55018, N01-HC-55019, N01-HC-55020, N01-HC-55021, N01-HC-55022, R01HL087641, R01HL59367 and R01HL086694; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health contract HHSN268200625226C. The authors thank the staff and participants of the ARIC study for their important contributions. Infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research. Nora Franceschini was supported by AHA 0675001N (NF).

2. Baltimore Longitudinal Study of Aging GWAS (BLSA): This research was supported in part by the Intramural Research Program of the NIH, National Institute of Aging. A portion of that support was through a R&D contract with MedStar Research Institute.

3. Cardiovascular Health Study (CHS): The CHS research reported in this article was supported by contract numbers N01-HC-85079 through N01-HC-85086, N01-HC-35129, N01 HC-15103, N01 HC-55222, N01-HC-75150, N01-HC-45133, grant numbers U01 HL080295 and R01 HL087652 from the National Heart, Lung, and Blood Institute, with additional contribution from the National Institute of Neurological Disorders and Stroke. A full list of principal CHS investigators and institutions can be found at http://www.chs-nhlbi.org/pi.htm. DNA handling and genotyping was supported in part by National Center for Research Resources grant M01-RR00425 to the Cedars-Sinai General Clinical Research Center Genotyping core and National Institute of Diabetes and Digestive and Kidney Diseases grant DK063491 to the Southern California Diabetes Endocrinology Research Center. Evan L. Thacker was supported by NHLBI training grant T32-HL007902.

Nature Genetics: doi:10.1038/ng.571

4. Invecchiare in Chianti (INCHIANTI): The InCHIANTI study baseline (1998-2000) was supported as a “targeted project” (ICS110.1/RF97.71) by the Italian Ministry of Health and in part by the U.S. National Institute on Aging (Contracts: 263 MD 9164 and 263 MD 821336); the InCHIANTI Follow-up 1 (2001-2003) was funded by the U.S. National Institute on Aging (Contracts: N.1-AG-1-1 and N.1-AG-1-2111); the InCHIANTI Follow-ups 2 and 3 studies (2004-2010) were financed by the U.S. National Institute on Aging (Contract: N01-AG-5-0002);supported in part by the Intramural research program of the National Institute on Aging, National Institutes of Health, Baltimore, Maryland.

5. The Rotterdam Study (RS): The Rotterdam Study is funded by Erasmus Medical Center and Erasmus University, Rotterdam, Netherlands Organization for the Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam. The generation and management of GWAS genotype data for the Rotterdam Study is supported by the Netherlands Organisation of Scientific Research NWO Investments (nr. 175.010.2005.011, 911-03-012). This study is funded by the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), and the Netherlands Genomics Initiative (NGI)/Netherlands Organisation for Scientific Research (NWO) project nr. 050-060-810.

6. Framingham Heart Study (FHS): This research was conducted in part using data and resources from the Framingham Heart Study of the National Heart, Lung, and Blood Institute of the National Institutes of Health and Boston University School of Medicine. The analyses reflect intellectual input and resource development from the Framingham Heart Study investigators participating in the SNP Health Association Resource (SHARe) project. This work was partially supported by the National Heart, Lung, and Blood Institute's Framingham Heart Study (contract no. N01-HC-25195) and its contract with Affymetrix for genotyping services (contract no. N02-HL-6-4278). A portion of this research used the Linux Cluster for Genetic Analysis (LinGA-II) funded by the Robert Dawson Evans Endowment of the Department of Medicine at Boston University School of Medicine and Boston Medical Center. Emelia Benjamin is supported in part by 1R01HL092577-01A1, 1R01 AG028321, and N01-HC 25195.

Nature Genetics: doi:10.1038/ng.571

7. Women's Genome Health Study (WGHS): This work was funded by the National Heart, Lung and Blood Institute (NHLBI; HL043851) and the National Cancer Institute (NCI; CA047988) and the Donald W Reynolds Foundation. Genotyping and collaborative scientific support was provided by Amgen, Inc. Dr. Everett is supported by the American Heart Association (AHA 0835304N). Dr. Paré is supported by a a Fonds de la echerche en Santé due Quebec scholarship.

8. Atherosclerotic Disease Vascular Function and Genetic Epidemiology (ADVANCE): The ADVANCE study is a collaboration between Stanford Medical School and Kaiser-Permanente of Northern California and was supported by a grant from the Donald W. Reynolds Foundation, Las Vegas, NV from 2000-2006. ADVANCE is also supported by a grant from the NIH through the National Heart, Lung, and Blood Institute, RO1 HL087647-01, to Dr. Quertermous. The ADVANCE study would also like to acknowledge the volunteer participants of the study as well as other members of the ADVANCE group that have contributed to the overall effort including: Drs. Alan G. at Kaiser and Mark Hlatky, Audrey Southwick, Neil Risch and Richard Myers. The ADVANCE study would also like to acknowledge the work of Drs. Stephen Sidney and Dale Williams for the continued relationship between ADVANCE and the CARDIA study. The CARDIA study is supported by contracts N01-HC-48047, N01-HC-48048, N01-HC-48049, N01-HC-48050, and N01-HC-95095 from the National Heart, Lung, and Blood Institute.

9. Atherosclerosis, Thrombosis and Vascular Biology Italian Study Group (ATVB) – Premature MI cohort: This study was supported by the Italian charitable Foundation "Associazione per lo Studio della Trombosi in Cardiologia" and the Atherosclerosis, Thrombosis, and Vascular Biology Italian Study Group.

10. The Diabetes Genetics Initiative (DGI): Novartis, Sigrid Juselius Foundation, and Swedish Research council

11. The Finland-United States Investigation of NIDDM Genetics Study (FUSION): Support was provided by NIH grants DK062370 (M.B.) and DK072193 (K.L.M.)

Nature Genetics: doi:10.1038/ng.571

12. International Agency for Research on Cancer (IARC): US National Cancer Institute (RO1 CA092039) & Institut National du Cancer, France.

13. Myocardial Infarction Genetics Consortium (MIGen): The MIGen study was funded by the U.S. National Institutes of Health (NIH) and National Heart, Lung, and Blood Institute's STAMPEED genomics research program through a grant to D.A. S.K. is supported by a Doris Duke Charitable Foundation Clinical Scientist Development Award, a charitable gift from the Fannie E. Rippel Foundation, the Donovan Family Foundation, a career development award from the NIH, and institutional support from the Department of Medicine and Cardiovascular Research Center at Massachusetts General Hospital. Genotyping was partially funded by The Broad Institute Center for Genotyping and Analysis, which is supported by grant U54 RR020278 from the National Center for Research Resources. V.S. was supported by the Sigrid Juselius Foundation and the Finnish Foundation for Cardiovascular Research. The REGICOR study was partially funded by the Ministerio de Sanidad y Consumo, Instituto de Salud Carlos III (Red HERACLES RD06/0009), Fundació Marató TV3, Fondos FEDER Unión Europea, the CIBER Epidemiología y Salud Pública, the FIS (CP05/00290, PI061254), and AGAUR (SGR 2005/00577); G.L. is supported by the Juan de la Cierva Program, Ministerio de Educación. The HARPS study was supported by grants and contracts from the US NIH (R01HL056931, P30ES007033, N01-HD-1-3107). The Myocardial Infarction Genetics (MIGen) Consortium Investigators include Stephen M. Schwartz[1,2], David S. Siscovick[1], Jean Yee[1,2], Yechiel Friedlander[3], Roberto Elosua[4,5], Jaume Marrugat[4], Gavin Lucas[4], Isaac Subirana[5,4], Joan Sala[6], Rafael Ramos[7], Christopher J O'Donnell[8,9], Sekar Kathiresan[10,11,12,8], Calum A. MacRae[10,8], Veikko Salomaa[13], Aki S. Havulinna[13], Leena Peltonen[12,14,15], Olle Melander[16], Goran Berglund[17], Benjamin F. Voight[11,12,18], Joel N. Hirschhorn[12,19], Kiran Musunuru[10,11,12,8], Mark Daly[11,12,8], Shaun Purcell[11,12,20], Aarti Surti[12], Candace Guiducci[12], Lauren Gianniny[12], Daniel Mirel[12], Melissa Parkin[12], Noël Burtt[12], Stacey Gabriel[12], David Altshuler[11,12,8,18,19]. Affiliations:[1] Cardiovascular Health Research Unit, Departments of Medicine and Epidemiology, University of Washington, Seattle, Washington, USA. [2] Department of Epidemiology, University of Washington, Seattle, Washington, USA. [3] Unit of Epidemiology, Hebrew University-Hadassah School of Public Health, Jerusalem, Israel. [4] Cardiovascular Epidemiology and Genetics, Institut Municipal d'Investigacio Medica, Barcelona, Spain. [5] CIBER Epidemiología y Salud Pública, Barcelona, Spain. [6] Servei de Cardiologia i Unitat Coronària, Hospital de Girona Josep Trueta and Institut de Investigació Biomedica de Girona, Spain. [7] Unitat de Recerca i Unitat Docent de Medicina de Familia de Girona, IDIAP Jordi Gol, Institut Català de la Salut, Spain. [8] Department

Nature Genetics: doi:10.1038/ng.571

of Medicine, Harvard Medical School, Boston, Massachusetts 02115, USA. [9] Framingham Heart Study and National, Heart, Lung, and Blood Institute, Framingham, Massachusetts, 01702, USA. [10] Cardiovascular Research Center and Cardiology Division, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. [11] Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. [12] Program in Medical and Population Genetics, Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA. [13] Department of Chronic Disease Prevention, THL-National Institute for Health and Welfare, Helsinki, Finland. [14] Wellcome Trust Sanger Institute, Cambridge CB10 1SA, UK. [15] Institute for Molecular Medicine, University of Helsinki, Helsinki 00029, Finland. [16] Department of Clinical Sciences, Hypertension and Cardiovascular Diseases, University Hospital, Malmö, Lund University, Malmö 20502, Sweden. [17] Department of Clinical Sciences, Internal Medicine, University Hospital Malmö, Lund University, Malmö 20502, Sweden. [18] Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts, 02114, USA. [19] Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. [20] Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, 02142, USA.

14. Nurses' Health Study (NHS): PK, FG, DH and SH were supported by NIH grant 5P01CA087969-10.

15. Netherlands Study of Depression and Anxiety (NESDA) and Netherlands Twin Registry (NTR): NESDA: Geestkracht program of ZonMW (10-000-1002); matching funds from universities and mental health care institutes involved in NESDA (GGZ Buitenamstel-Geestgronden, Rivierduinen, University Medical Center Groningen, GGZ Lentis, GGZ Friesland, GGZ Drenthe). Genotyping was funded by the Genetic Association Information Network (GAIN) of the Foundation for the US National Institutes of Health. NTR: NWO/ZonMW: 904-61-090; 904-61-193; 480-04-004; SPI 56-464-14192; NBIC-Bioassist; Center for Medical Systems Biology; Neuroscience Campus Amsterdam and NWO-VENI 451-06-004. Genotyping was funded by the Genetic Association Information Network (GAIN) of the Foundation for the US National Institutes of Health.

Nature Genetics: doi:10.1038/ng.571

17. Genetic Association Information Network (GAIN): Funding Source - MGS1 and MGS2 Sample Collections -- R01 MH67257; R01 MH59588; R01 MH59571; R01 MH59565; R01 MH59587; R01 MH60870; R01 MH60879; R01 MH59566; R01 MH59586; R01 MH61675; Funding Source - Genotyping and Analyses -- U01 MH79469; U01 MH79470. Institutions -- Atlanta Veterans Affairs Medical Center and Emory University, Atlanta, GA, USA.; NorthShore University HealthSystem (NUH), Evanston, IL, USA.; Louisiana State University (LSU) Health Sciences Center, New Orleans, LA, USA.; Mount Sinai School of Medicine, New York, NY, USA.; Queensland Ctr. for Mental Health Res. (QCMHR) and Univ. Queensland, Brisbane, Australia.; Stanford University, Palo Alto, CA, USA.; University of California, San Francisco, San Francisco, CA, USA.; University of Colorado Health Sciences Center, Denver, CO, USA.; University of Iowa, Iowa City, IA, USA.; Washington University, St. Louis, MO, USA. Lead Principal Investigator and Clinical Site PI -- Pablo V. Gejman, MD[1] ; Statistical Task Force Chair, Clinical Site PI, Clinical Database -- Douglas F. Levinson, MD[2] ; Clinical Site PI -- Farooq Amin, MD[3] ; Donald W. Black, MD[4] ; Nancy G. Buccola, APRN, CNS-BC[5] ; William F. Byerley, MD[6] ; C. Robert Cloninger, MD[7] ; Robert Freedman, MD[8] ; Bryan J. Mowry, MD[9] ; Jeremy M. Silverman, PhD[10] ; Statistician -- Frank Dudbridge, PhD[11] ; Peter A. Holmans, PhD[12] ; Maria Martinez, PhD[13] ; Itsik Pe'er, PhD[14] ; Jianxin Shi, PhD[2] ; Alice Whittemore, PhD[2] ; Co-Investigator -- Jubao Duan, PhD[1] ; Ayman H. Fanous, MD[15] ; Kenneth S. Kendler, MD[16] ; John P. Rice, PhD[17] ; Alan R. Sanders, MD[1] . Affiliations: [1] NorthShore University HealthSystem (NUH), Evanston, IL, USA; [2] Stanford University School of Medicine, Palo Alto, CA, USA; [3] Emory University, Atlanta, GA, USA; [4] University of Iowa Hospitals and Clinics, Iowa City, IA, USA; [5] Louisiana State University Health Services Center, New Orleans, LA, USA; [6] University of San Francisco, San Francisco, CA, USA; [7] Washington University School of Medicine, St. Louis, MO, USA; [8]University of Colorado Health Sciences Center, Denver, CO, USA; [9] Queensland Centre for Mental Health Research, Queensland, Australia; [10] Mount Sinai School of Medicine, New York, New York, USA; [11] University of Cambridge, Institute of Public Health, Cambridge, UK; [12] Wales College of Medicine, Cardiff University, Cardiff, Wales, UK; [13] Institut National de la Santé et de la Recherche Médicale (INSERM), Toulouse, France; [14] Columbia University, New York, NY, USA; [15] Washington VA Medical Center and Georgetown University Medical Center, Washington, DC, USA; [16] Virginia Commonwealth University, Richmond, VA, USA; [17] Washington University School of Medicine, St. Louis, MO, USA.

17. Jaakko Kaprio is funded by The Academy of Finland Centre of Excellence in Complex Disease Genetics.

Nature Genetics: doi:10.1038/ng.571

18. Janet Audrain-McGovern is funded by NIH/NCI grants R01 CA10925 and R01 CA126958

19. John P. A. Ioannidis: Scientific support for this project was provided through the Tufts Clinical and Translational Science Institute (Tufts CTSI) under funding from the National Institute of Health/National Center for Research Resources (UL1 RR025752). Points of view or opinions in this paper are those of the authors and do not necessarily represent the official position or policies of the Tufts CTSI.

Nature Genetics: doi:10.1038/ng.571

Literature cited.

1. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *Am J Epidemiol* **129**, 687-702 (1989).
2. Newton-Cheh, C. et al. Genome-wide association study identifies eight loci associated with blood pressure. *Nat Genet* (2009).
3. Normal Human Aging: The Baltimore Longitudinal Study of Aging. *Ann Intern Med %R 10.1059/0003-4819-104-2-294_4* **104**, 294-c- (1986).
4. Tanaka, T. et al. Genome-wide association study of vitamin B6, vitamin B12, folate, and homocysteine blood concentrations. *Am J Hum Genet* **84**, 477-82 (2009).
5. Fried, L.P. et al. The Cardiovascular Health Study: design and rationale. *Ann Epidemiol* **1**, 263-76 (1991).
6. Psaty, B.M. et al. Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: Design of Prospective Meta-Analyses of Genome-Wide Association Studies From 5 Cohorts. *Circ Cardiovasc Genet %R 10.1161/CIRCGENETICS.108.829747* **2**, 73-80 (2009).
7. Ferrucci, L. et al. Subsystems contributing to the decline in ability to walk: bridging the gap between epidemiology and geriatric practice in the InCHIANTI study. *J Am Geriatr Soc* **48**, 1618-25 (2000).
8. Hofman, A. et al. The Rotterdam Study: objectives and design update. *Eur J Epidemiol* **22**, 819-29 (2007).
9. Hofman, A., Grobbee, D.E., de Jong, P.T. & van den Ouweland, F.A. Determinants of disease and disability in the elderly: the Rotterdam Elderly Study. *Eur J Epidemiol* **7**, 403-22 (1991).
10. Dawber, T.R., Meadors, G.F. & Moore, F.E., Jr. Epidemiological approaches to heart disease: the Framingham Study. *Am J Public Health Nations Health* **41**, 279-81 (1951).
11. Splansky, G.L. et al. The Third Generation Cohort of the National Heart, Lung, and Blood Institute's Framingham Heart Study: design, recruitment, and initial examination. *Am J Epidemiol* **165**, 1328-35 (2007).
12. Feinleib, M., Kannel, W.B., Garrison, R.J., McNamara, P.M. & Castelli, W.P. The Framingham Offspring Study. Design and preliminary data. *Prev Med* **4**, 518-25 (1975).
13. Ridker, P.M. et al. Rationale, design, and methodology of the Women's Genome Health Study: a genome-wide association study of more than 25,000 initially healthy american women. *Clin Chem* **54**, 249-55 (2008).
14. Ridker, P.M. et al. A randomized trial of low-dose aspirin in the primary prevention of cardiovascular disease in women. *N Engl J Med* **352**, 1293-304 (2005).
15. Assimes, T.L. et al. Susceptibility locus for clinical and subclinical coronary artery disease at chromosome 9p21 in the multi-ethnic ADVANCE study. *Hum Mol Genet* **17**, 2320-8 (2008).
16. Assimes, T.L. et al. A near null variant of 12/15-LOX encoded by a novel SNP in ALOX15 and the risk of coronary artery disease. *Atherosclerosis* **198**, 136-44 (2008).
17. Assimes, T.L. et al. Common polymorphisms of ALOX5 and ALOX5AP and risk of coronary artery disease. *Hum Genet* **123**, 399-408 (2008).
18. Knowles, J.W. et al. Failure to replicate an association of SNPs in the oxidized LDL receptor gene (OLR1) with CAD. *BMC Med Genet* **9**, 23 (2008).

Nature Genetics: doi:10.1038/ng.571

19. No evidence of association between prothrombotic gene polymorphisms and the development of acute myocardial infarction at a young age. *Circulation* **107**, 1117-22 (2003).
20. Saxena, R. et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* **316**, 1331-6 (2007).
21. Valle, T. et al. Mapping genes for NIDDM. Design of the Finland-United States Investigation of NIDDM Genetics (FUSION) Study. *Diabetes Care* **21**, 949-58 (1998).
22. Silander, K. et al. A large set of Finnish affected sibling pair families with type 2 diabetes suggests susceptibility loci on chromosomes 6, 11, and 14. *Diabetes* **53**, 821-9 (2004).
23. Scott, L.J. et al. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* **316**, 1341-5 (2007).
24. Scelo, G. et al. Occupational exposure to vinyl chloride, acrylonitrile and styrene and lung cancer risk (europe). *Cancer Causes Control* **15**, 445-52 (2004).
25. Brennan, P. et al. Uncommon CHEK2 mis-sense variant and reduced risk of tobacco-related cancers: case control study. *Hum Mol Genet* **16**, 1794-801 (2007).
26. Hung, R.J. et al. A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. *Nature* **452**, 633-7 (2008).
27. McKay, J.D. et al. Lung cancer susceptibility locus at 5p15.33. *Nat Genet* **40**, 1404-6 (2008).
28. Brennan, P. et al. Tobacco smoking, body mass index, hypertension, and kidney cancer risk in central and eastern Europe. *Br J Cancer* **99**, 1912-5 (2008).
29. Erdmann, J. et al. New susceptibility locus for coronary artery disease on chromosome 3q22.3. *Nat Genet* **41**, 280-2 (2009).
30. Kathiresan, S. A PCSK9 missense variant associated with a reduced risk of early-onset myocardial infarction. *N Engl J Med* **358**, 2299-300 (2008).
31. Kathiresan, S. et al. Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nat Genet* **41**, 334-41 (2009).
32. Hunter, D.J. et al. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat Genet* **39**, 870-4 (2007).
33. Penninx, B.W. et al. The Netherlands Study of Depression and Anxiety (NESDA): rationale, objectives and methods. *Int J Methods Psychiatr Res* **17**, 121-40 (2008).
34. Boomsma, D.I. et al. Netherlands Twin Register: from twins to twin families. *Twin Res Hum Genet* **9**, 849-57 (2006).
35. Vink, J.M., Willemsen, G. & Boomsma, D.I. Heritability of smoking initiation and nicotine dependence. *Behav Genet* **35**, 397-406 (2005).
36. Boomsma, D.I. et al. Genome-wide association of major depression: description of samples for the GAIN Major Depressive Disorder Study: NTR and NESDA biobank projects. *Eur J Hum Genet* **16**, 335-42 (2008).
37. Sullivan, P.F. et al. Genome-wide association for major depressive disorder: a possible role for the presynaptic protein piccolo. *Mol Psychiatry* **14**, 359-75 (2009).
38. Sullivan, P.F. et al. Genomewide association for schizophrenia in the CATIE study: results of stage 1. *Mol Psychiatry* **13**, 570-84 (2008).
39. Purcell, S.M. et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748-52 (2009).

Nature Genetics: doi:10.1038/ng.571