A Genome-wide Association Study of Lung Cancer

Identifies a Region of Chromosome 5p15

Associated with Risk for Adenocarcinoma

Maria Teresa Landi, Nilanjan Chatterjee, Kai Yu, Lynn R. Goldin, Alisa M. Goldstein, Melissa Rotunno, Lisa Mirabello, Kevin Jacobs, William Wheeler, Meredith Yeager, Andrew W. Bergen, Qizhai Li, Dario Consonni, Angela C. Pesatori, Sholom Wacholder, Michael Thun, Ryan Diver, Martin Oken, Jarmo Virtamo, Demetrius Albanes, Zhaoming Wang, Laurie Burdette, Kimberly F. Doheny, Elizabeth W. Pugh, Cathy Laurie, Paul Brennan, Rayjean Hung, Valerie Gaborieau, James D. McKay, Mark Lathrop, John McLaughlin, Ying Wang, Ming-Sound Tsao, Margaret R. Spitz, Yufei Wang, Hans Krokan, Lars Vatten, Frank Skorpen, Egil Arnesen, Simone Benhamou, Christine Bouchard, Andres Metsapalu, Tonu Vooder, Mari Nelis, Kristian Välk, John K. Field, Chu Chen, Gary Goodman, Patrick Sulem, Gudmar Thorleifsson, Thorunn Rafnar, Timothy Eisen, Wiebke Sauter, Albert Rosenberger, Heike Bickeböller, Angela Risch, Jenny Chang-Claude, H. Erich Wichmann, Kari Stefansson, Richard Houlston, Christopher I. Amos, Joseph F. Fraumeni, Jr., Sharon A. Savage, Pier Alberto Bertazzi, Margaret A. Tucker, Stephen Chanock, and Neil E. Caporaso

Figure S1. Illumina Infinium Genotype Cluster Plots Showing the Calling for the Notable SNP, rs2736100
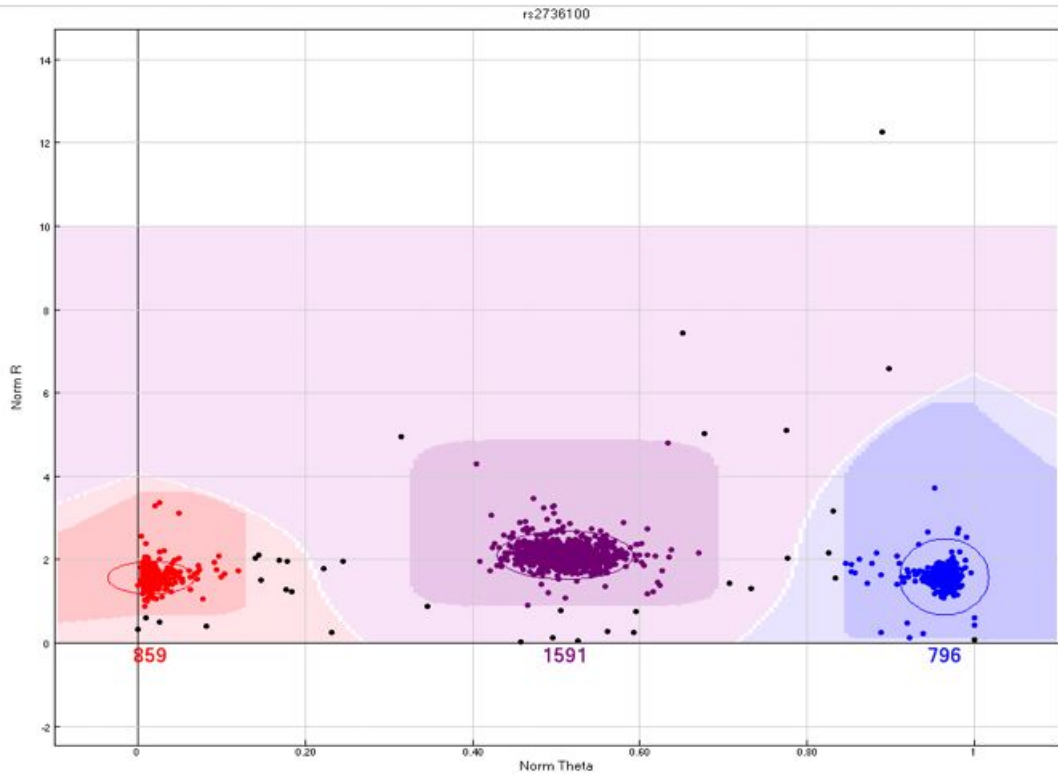
Figure S2. Lung GWAS QQ Plot of Hardy-Weinberg Proportion

ATBC=Alpha-Tocopherol, Beta-Carotene Cancer Prevention study
CPSII=Cancer Prevention Study II Nutrition Cohort
EAGLE=Environment and Genetics in Lung cancer Etiology study
PLCO=Prostate, Lung, Colon, Ovary Screening Trial

Hardy-Weinberg Proportion

Cohort	total	p<.001	p<.05
ATBC	503857	2621	27100
CPSII_1M	848021	1231	33749
CPSII_610K	505239	1647	23941
EAGLE	492692	2401	28479
PLCO	494865	1472	24076

Legend:
- ATBC
- CPSII_1M
- CPSII_610K
- EAGLE
- PLCO

X-axis: Expected P value
Y-axis: Observed P value
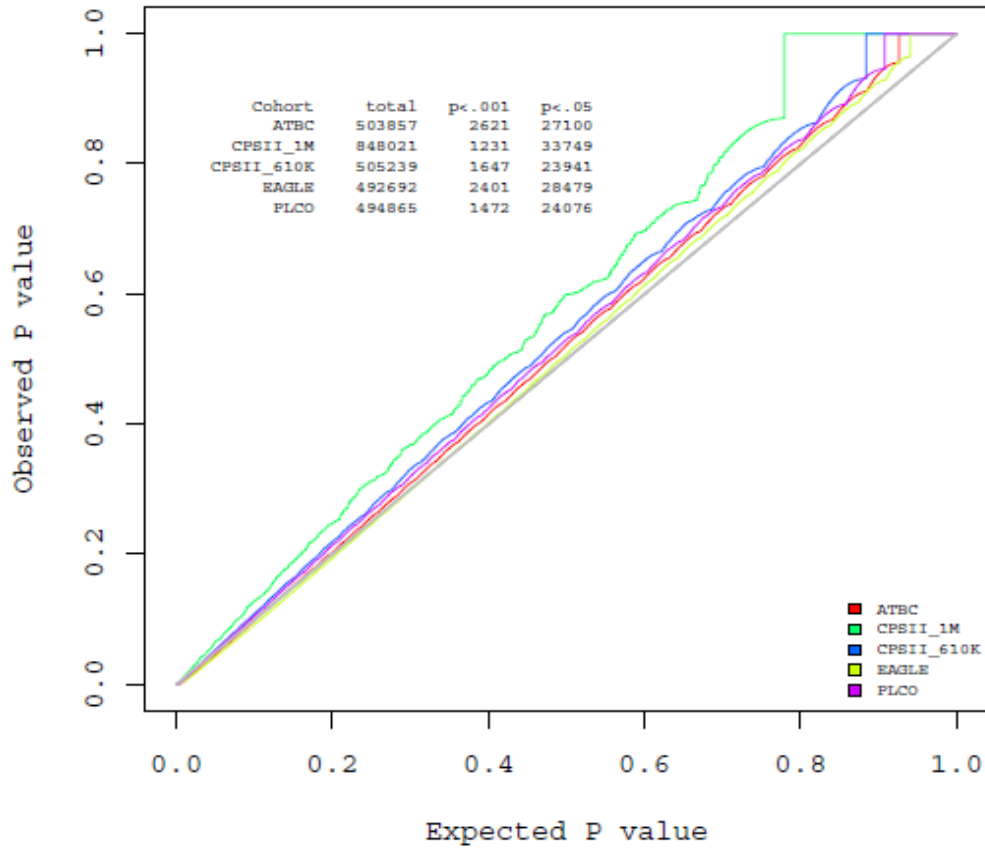
Figure S3. Lung GWAS Plot of Admixture Defined by Analysis with the STRUCTURE Program

ATBC=Alpha-Tocopherol, Beta-Carotene Cancer Prevention study
CPSII=Cancer Prevention Study II Nutrition Cohort
EAGLE=Environment and Genetics in Lung cancer Etiology study
PLCO=Prostate, Lung, Colon, Ovary Screening Trial
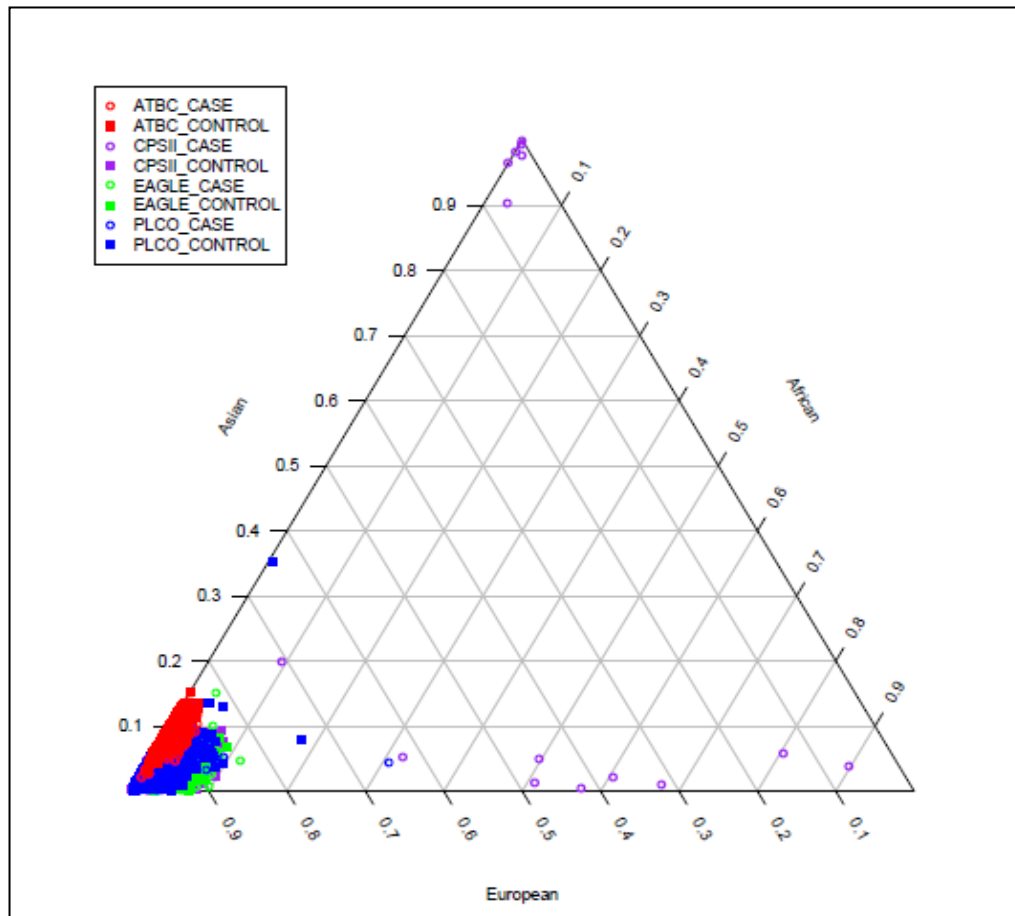
Figure S4A. CPSII_1M PCA Plot (EV1-EV6)
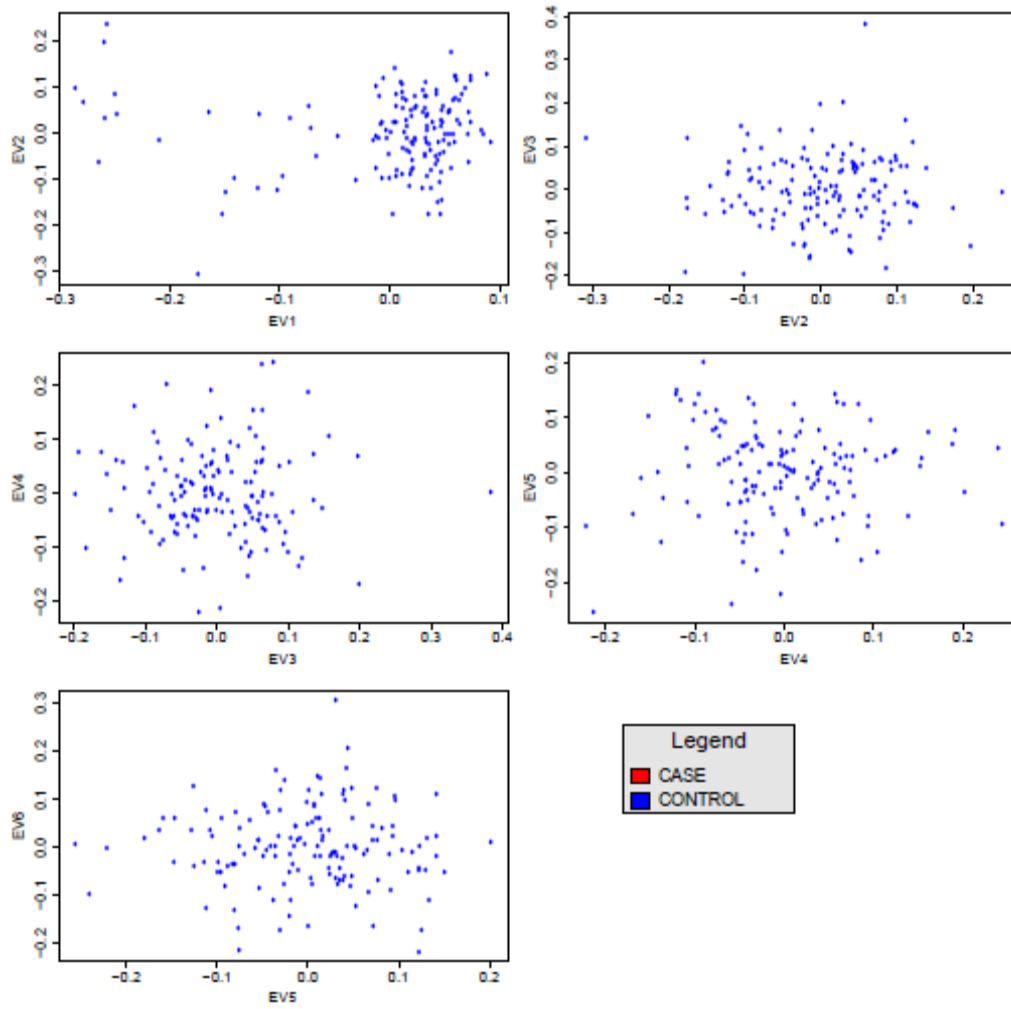
Figure S4B. CPSII_610K Set PCA Plot (EV1-EV6)
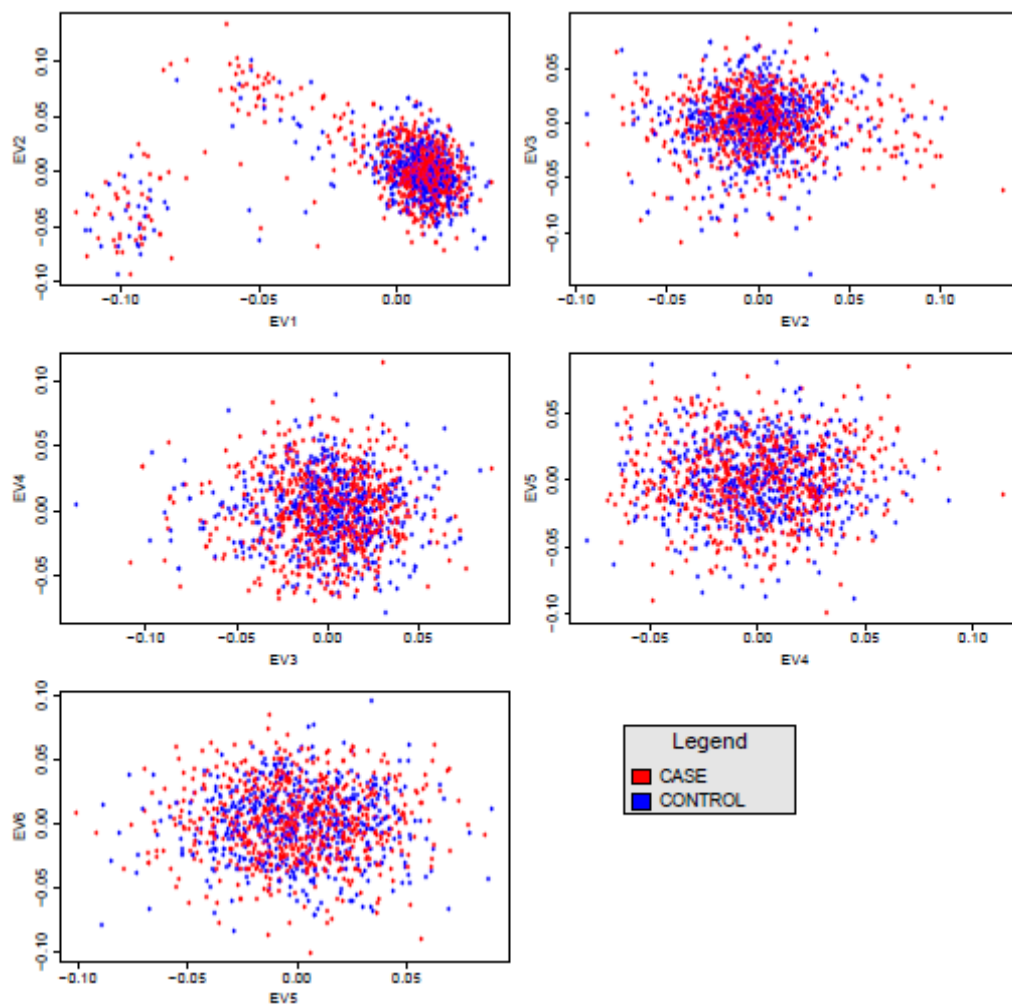
Figure S4C. EAGLE Study PCA Plot (EV1-EV6)
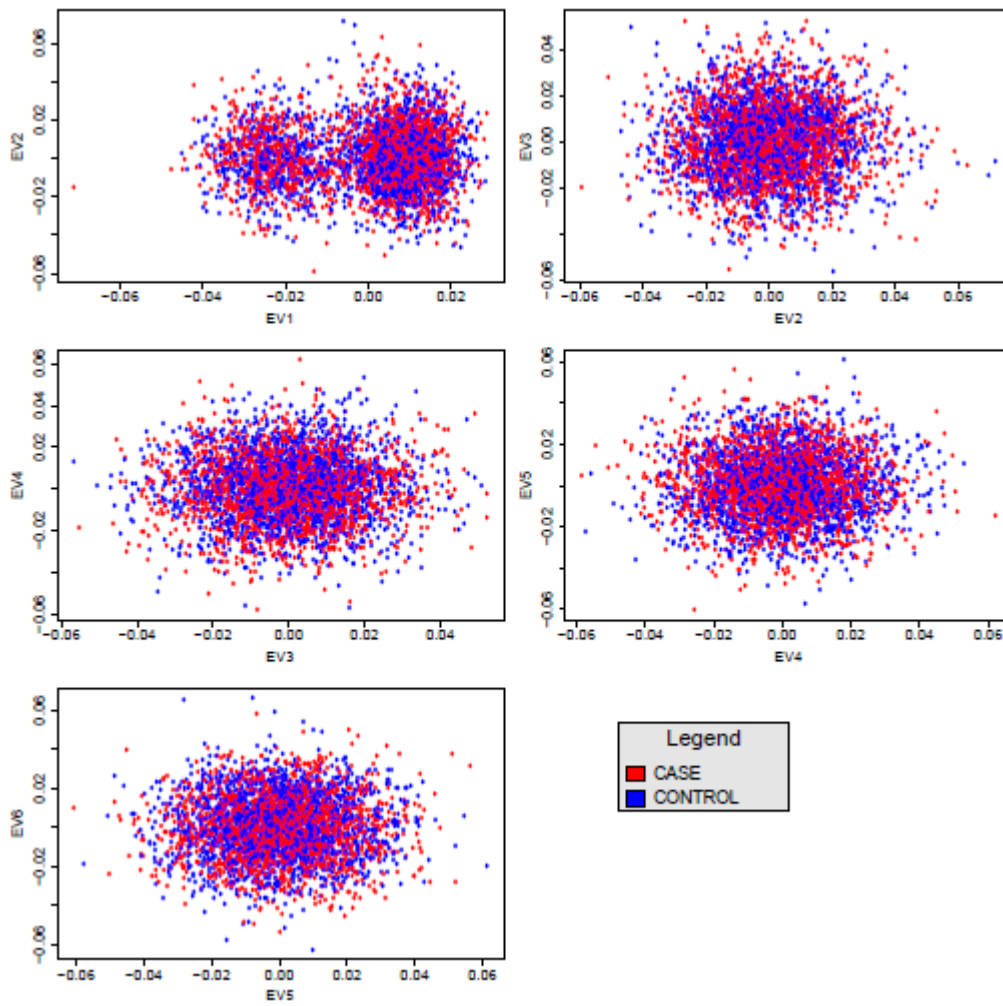
Figure S4D. PLCO PCA Plot (EV1-EV6)

Legend
- CASE
- CONTROL
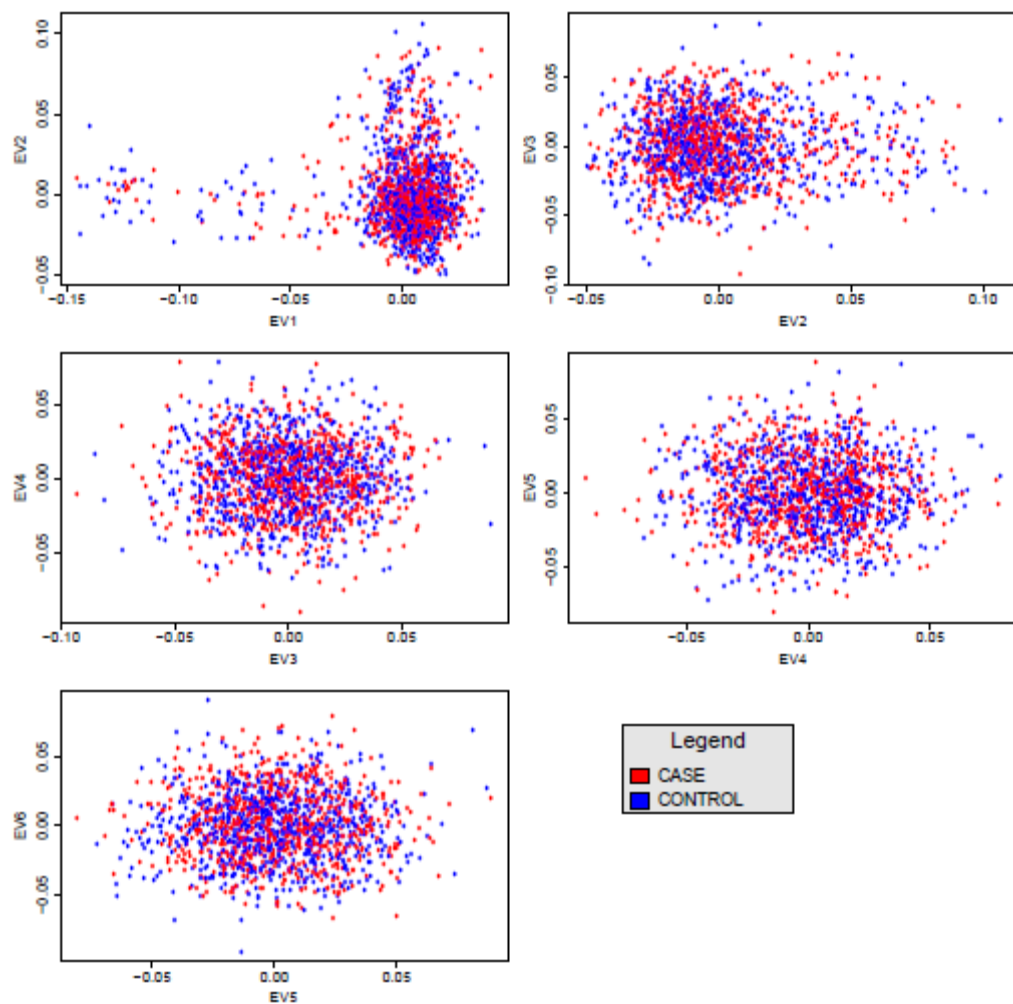
Figure S5A. Quantile-Quantile Plot of P Values for the Environment and Genetics in Lung Cancer Etiology (EAGLE) Study

EAGLE Unadjusted
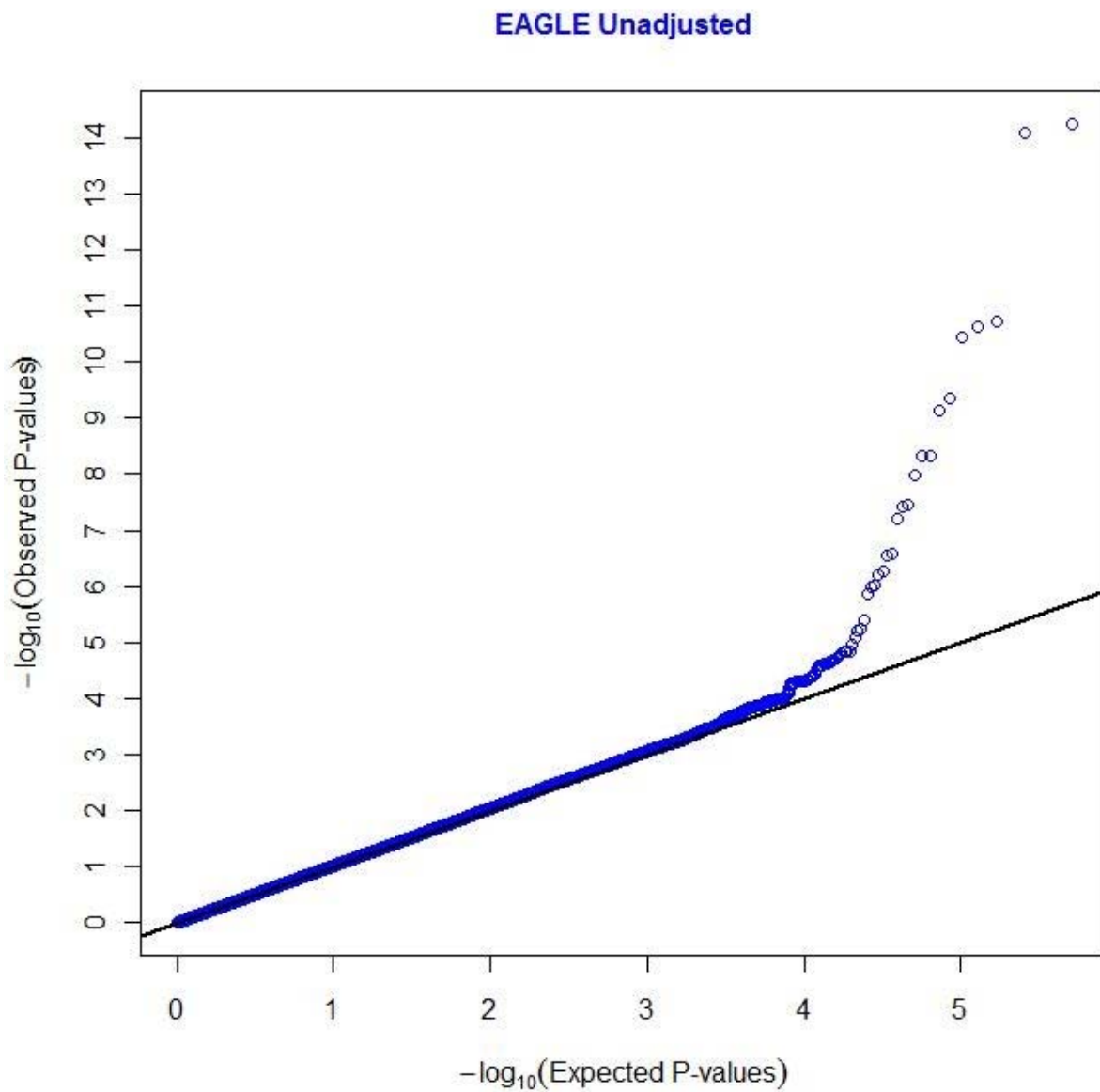
Figure S5B. Quantile-Quantile Plot of P Values for the Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study

ATBC Unadjusted
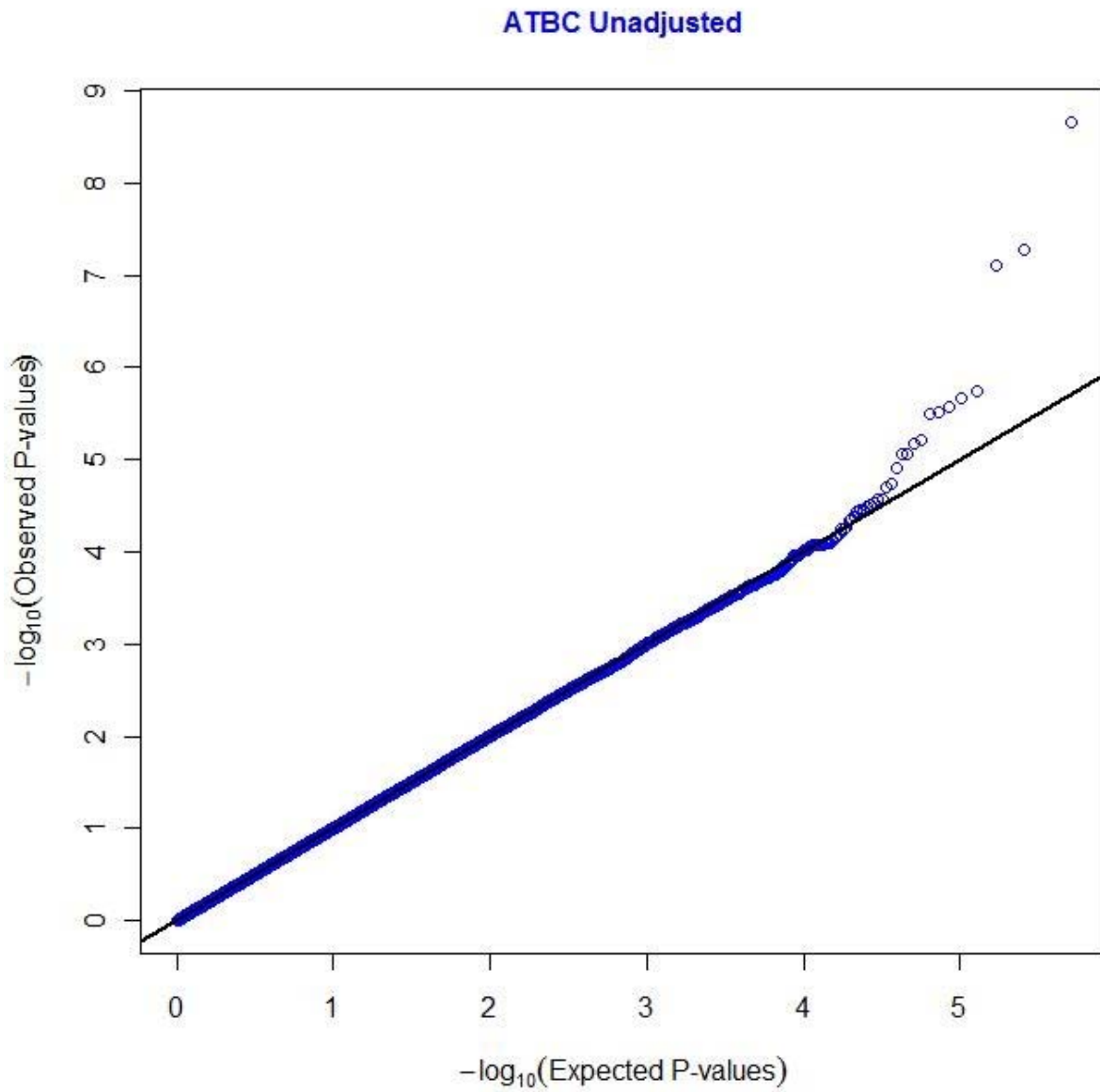
Figure S5C. Quantile-Quantile Plot of P Values for the Prostate, Lung, Colon, Ovary (PLCO) Screening Trial

PLCO Unadjusted
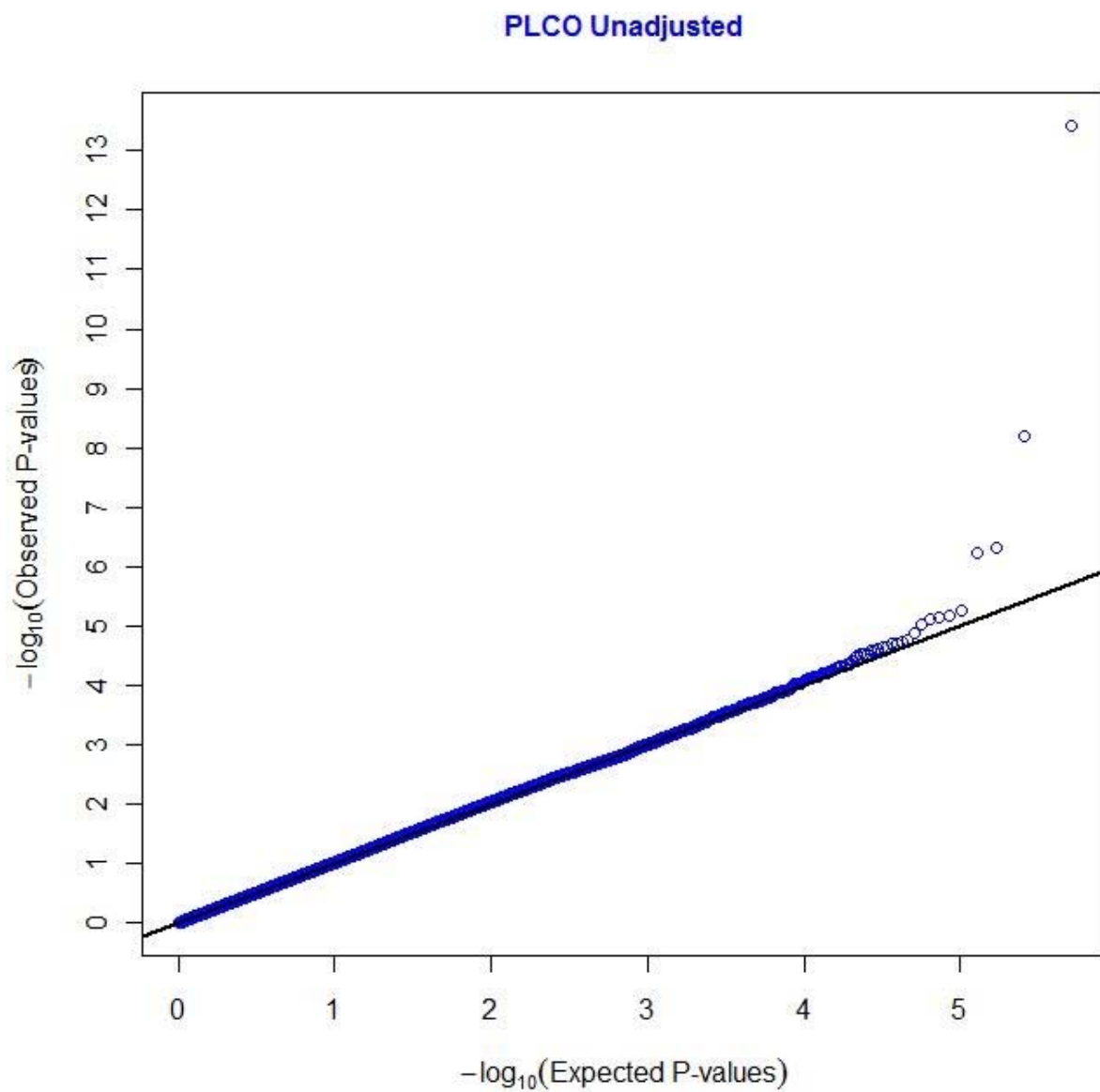
Figure S5D. Quantile-Quantile Plot of P Values for the Cancer Prevention Study II (CPSII) Nutrition Cohort

CPSII Unadjusted
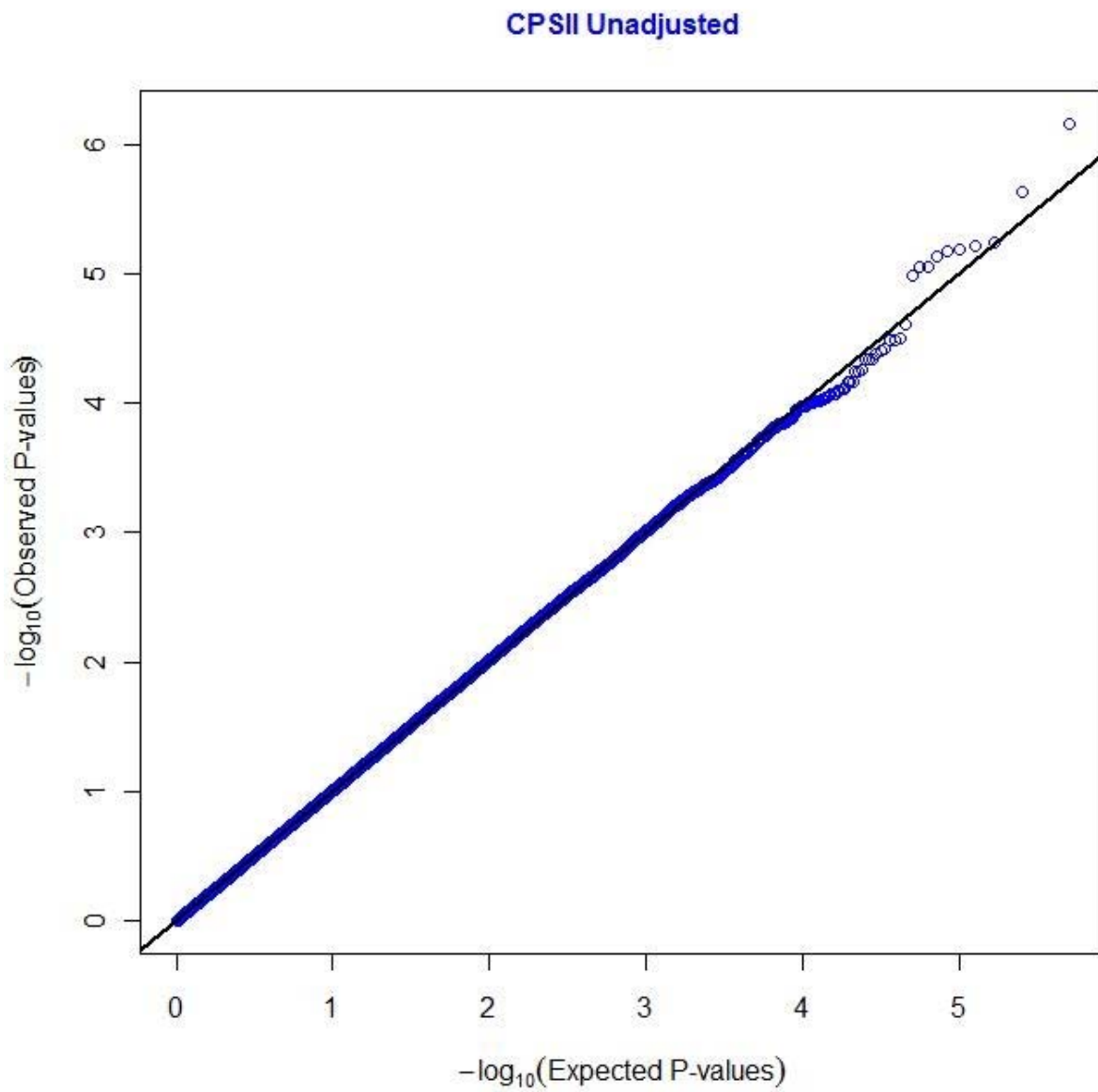
Table S1. Lung GWAS Subjects after Sample Quality Control

STUDY	PROJECT	Illumina HapMap	CASES	CONTROLS	Total
EAGLE[a]	GEI[b] Lung	550K	1955	1997	3952
ATBC[c]	Common Controls	610Quad		1286	1286
ATBC[c]	Lung study	610 Quad	1760		1760
ATBC[c]	PanScan[d]	550K		208	208
PLCO[e]	GEI[a] Lung	550K	806	859	1665
PLCO[e]	Prostate Controls	317K/240K		1105[f]	1105[f]
PLCO[e]	Lung study	610 Quad	615		615
CPSII[g]	Common Controls	1M Duo		148	148
CPSII[g]	Common Controls	610 Quad		533	533
CPSII[g]	Lung study	610 Quad	721		721
CPSII[g]	PanScan[d]	550K		118	118
Grand Total			**5857**	**6254**	**12111**

[a]Environment and Genetics in Lung Cancer Etiology study
[b]Gene and Environment Health Initiative
[c]Alpha-Tocopherol, Beta-Carotene Cancer Prevention study
[d]Pancreatic Cancer Cohort Consortium
[e]Prostate, Lung, Colon, Ovary Screening Trial
[f]Data from 1,105 PLCO control subjects from HumanHap317 chips after removing samples with low completion rates. Of these, 1,098 PLCO control subjects had also data from HumanHap240 chips.
[g]Cancer Prevention Study II Nutrition Cohort

Table S2. Subjects Excluded from the Genome-wide Association Analysis

Exclusion Criteria	Cutoff	N. Subjects
Overall missing rate[a]	> 5%	0[b]
X Chr. missing rate[a]	ATBC[c] (> 7%)	6
	CPSII[d]_1M (>5%)	0
	CPSII[d]_610K (>6%)	11
	EAGLE[e], PLCO[f] (>5%)	12
	Sub Total	29
Autosomal heterozygosity[a]	ATBC[c] (> 34%)	0
	CPSII[d]_1M (> 34%)	0
	CPSII[d]_610K (> 34%)	1
	EAGLE[e], PLCO[f] (> 34%)	1
	Sub Total	2
X Chr. heterozygosity for female[a]	CPSII[d]_1M (> 26% or < 18%)	0
	CPSII[d]_610K (> 35% or < 20%)	1
	EAGLE[e], PLCO[f] (> 35% or <20%)	4
	Sub Total	5
X Chr. Heterozygosity for male[a]	ATBC[c] (> 1.2%)	5
	CPSII[d]_1M (> 0.5%)	0
	CPSII[d]_610K (> 0.5%)	2
	EAGLE[e], PLCO[f] (> 0.5%)	11
	Sub Total	18
Female observed as male	CPSII[d]_1M (< 5%)	0
	CPSII[d]_610K (< 5%)	0
	EAGLE[e], PLCO[f] (< 5%, non-aneuploid)	13
	Sub Total	13
Male observed as female	ATBC[c] (> 5%)	5
	CPSII[d]_1M (> 5%)	0
	CPSII[d]_610K (> 5%)	1
	EAGLE[e], PLCO[f] (> 5%, non-aneuploid)	14
	Sub Total	20
Unexpected duplicates	Genotype concordance > 98%	49
Relative pairs	Exclude the one with lower completion	40
Estimated CEU admixture	< 80%	18
Total		183

[a]QC exclusions; the others are analytic exclusions
[b]Missing rate filtering was applied at the sample level instead than subject level
[c]Alpha-Tocopherol, Beta-Carotene Cancer Prevention study
[d]Cancer Prevention Study II Nutrition Cohort
[e]Environment and Genetics in Lung Cancer Etiology study
[f]Prostate, Lung, Colon, Ovary Screening Trial

Table S3_ Lung GWAS Summary of Locus Exclusions

Quality Control Group	Locus Count[a]
ATBC[b]	888
CPSII[c] (HapMap1M)	128
CPSII[c] (HapMap610QUAD)	485
EAGLE[d]	991
PLCO[e]	410
PanScan[f] ATBC[b]	111
PanScan[f] CPSII[c]	16
Prostate PLCO[e]	604
Total*	2011

[a]Based on loci with p-values< 10^{-7} for deviation of Hardy Weinberg proportions
[b]Alpha-Tocopherol, Beta-Carotene Cancer Prevention study
[c]Cancer Prevention Study II Nutrition Cohort
[d]Environment and Genetics in Lung Cancer Etiology study
[d]Gene and Environment Health Initiative
[e]Prostate, Lung, Colon, Ovary Screening Trial
[f]Pancreatic Cancer Cohort Consortium

Table S4. Association of Selected SNPs on 6p21-6p22 with Lung Cancer Risk Overall and by Histologic Types and Smoking Status

Marker[a], Allele[b], Location[c], Gene[d], MAF[e]	Study	Overall NCI data: 5848 controls (2384 current, 2041 former,1402 never smokers) 5739 cases (3410 current, 1946 former, 362 never smokers) Meta-analysis: 19666 controls, 13300 cases		Adenocarcinoma NCI data: 5848 controls (2384 current, 2041 former,1402 never smokers) 1730 cases (786 current, 731 former, 203 never smokers) Meta-analysis: 14933 controls, 3333 ca		Squamous cell carcinoma NCI data: 5848 controls (2384 current, 2041 former,1402 never smokers) 1400 cases (953 current, 422 former, 23 never smokers) Meta-analysis: 14933 controls, 2589 cases		Small cell carcinoma NCI data: 5848 controls (2384 current, 2041 former,1402 never smokers) 678 cases (516 current, 152 former, 9 never smokers)[f] Meta-analysis: 13796 controls, 1418 cases	
		P-value[f]	**OR**[g] **(95% CI)**	**P-value**[f]	**OR**[g] **(95% CI)**	**P-value**[f]	**OR**[g] **(95% CI)**	**P-value**[f]	**OR**[g] **(95% CI)**
rs2395471, (G,A) 31348671 HLA-C, 0.543/0.455	NCI_Smoking Unadjusted	0.002	0.92 (0.87, 0.97)	0.040	0.92 (0.85, 1.00)	0.25	0.95 (0.87, 1.04)	0.42	0.95 (0.85, 1.07)
	NCI_Smoking Adjusted	0.001	0.91 (0.85, 0.96)	0.023	0.90 (0.83, 0.99)	0.086	0.92 (0.84, 1.01)	0.18	0.92 (0.81, 1.04)
	NCI_Current smokers	0.06	0.93 (0.86, 1.00)	0.35	0.94 (0.83, 1.07)	0.41	0.95 (0.85, 1.07)	0.99	1.00 (0.87, 1.15)
	NCI_Former smokers	0.04	0.91 (0.83, 1.00)	0.18	0.92 (0.81, 1.04)	0.47	0.95 (0.81, 1.10)	0.23	0.86 (0.68, 1.10)
	NCI_Never smokers	0.54	0.95 (0.79, 1.13)	0.11	0.83 (0.67, 1.05)	0.54	1.21 (0.66, 2.20)	0.35	0.61 (0.22, 1.71)
	Meta-analysis Overall	6.87x10⁻⁴	0.93 (0.89, 0.97)	0.031	0.93 (0.88, 0.99)	0.17	0.95 (0.88, 1.02)	0.45	0.96 (0.86, 1.07)
	Meta-analysis 3 studies[h]	0.001	0.93 (0.89, 0.97)						
rs3117582, (A,C) 31728499 BAT3,APOM, 0.104/0.116	NCI_Smoking Unadjusted	0.001	1.16 (1.06, 1.27)	0.50	1.05 (0.91, 1.21)	0.015	1.20 (1.04, 1.38)	0.90	0.99 (0.79, 1.23)
	NCI_Smoking Adjusted	0.010	1.14 (1.03, 1.26)	0.43	1.06 (0.91, 1.23)	0.011	1.24 (1.05, 1.46)	0.27	1.13 (0.91, 1.39)
	NCI_Current smokers	0.03	1.16 (1.02, 1.32)	0.75	1.04 (0.83, 1.29)	0.13	1.16 (0.96, 1.39)	0.50	1.09 (0.86, 1.38)
	NCI_Former smokers	0.03	1.19 (1.02, 1.40)	0.50	1.08 (0.86, 1.35)	0.17	1.20 (0.92, 1.57)	0.94	1.02 (0.67, 1.55)
	NCI_Never smokers	0.72	0.94 (0.68, 1.30)	0.78	1.06 (0.70, 1.59)	0.75	1.18 (0.42, 3.32)	0.27	2.01 (0.58, 7.0)
	Meta-analysis Overall	4.79x10⁻¹²	1.22 (1.15, 1.29)	0.21	1.07 (0.96, 1.19)	0.017	1.16 (1.03, 1.31)	0.46	1.07 (0.90, 1.27)
	Meta-analysis 3 studies	0.001	1.13 (1.05, 1.22)						
rs4324798, (G,A) 28884096, TRNAA-UGC, 0.086/0.080	NCI_Smoking Unadjusted	0.057	1.11 (1.00-1.23)	0.68	0.97 (0.82, 1.14)	0.064	1.17 (0.99, 1.39)	0.31	1.12 (0.90, 1.41)
	NCI_Smoking Adjusted	0.040	1.13 (1.00-1.26)	0.96	1.00 (0.84, 1.19)	0.039	1.22 (1.01, 1.47)	0.20	1.17 (0.92, 1.49)

NCI_Current smokers	0.009	1.23 (1.05, 1.44)	0.26	1.16 (0.90, 1.50)	0.10	1.21 (0.96, 1.51)	0.35	1.15 (0.86, 1.52)
NCI_Former smokers	0.75	1.02 (0.86, 1.23)	0.42	0.90 (0.70, 1.16)	0.52	1.10 (0.82, 1.48)	0.26	1.27 (0.84, 1.94)
NCI_Never smokers	0.61	0.91 (0.64, 1.29)	0.31	0.78 (0.48, 1.26)	0.59	0.68 (0.16, 2.82)	0.58	1.52 (0.35, 6.65)
Meta-analysis Unadjusted	1.74×10^{-8}	1.16 (1.09-1.24)	0.96	1.00 (0.88, 1.12)	0.17	1.10 (0.96, 1.27)	0.26	1.12 (0.92, 1.37)
Meta-analysis 3 Studies	0.04	1.09 (1.00, 1.18)						

SNPs identified as relevant for lung cancer risk in previous studies are shown. The lung cancer cases in the Overall analyses include adenocarcinomas, squamous cell carcinomas, small cell carcinomas, and other groups of different histologic types or mixed histologies. The Overall meta-analyses include four NCI GWAS studies plus ten additional studies; the meta-analyses in the three major histologic types include four NCI GWAS studies plus three replication studies. Results for each SNP are shown following this order: Top row (NCI_Smoking Unadjusted), results from the NCI GWAS data adjusted for age in five-year intervals (defined as age-at-diagnosis/interview for case-control study and age at baseline for cohort study), gender, study (EAGLE, PLCO, ATBC, CPS-II) and four principal components for population stratification within studies (two for PLCO, and one each for EAGLE and ATBC). Adjustment for the same variables plus smoking status (current, former, never), cigarettes smoked per day (≤10,11-20,21-30,31-40,41+), duration in ten-year intervals and number of years since quitting (1-5,6-10,11-20,21-30,30+) for former smokers showed virtually identical results (data not shown). Second row (NCI_Curremt_smokers), results from the NCI GWAS data restricted to current smoker subjects, adjusted for the same variables of the top row. Third row (NCI_Former_smokers), results from the NCI GWAS data restricted to former smoker subjects who quit smoking >6 months prior to the study, adjusted for the same variables of the top row. Fourth row (NCI_Never_smokers), results from the NCI GWAS data restricted to never smoker subjects who smoked <100 cigarettes in their lifetime, adjusted for the same variables of the top row. Bottom row (Meta-analysis_Overall, results from the meta-analysis combining NCI GWAS data and 10 additional studies, unadjusted models. Bottom row (Meta-analysis 3 Studies, results from the meta-analysis combining NCI GWAS data and 3 additional studies included in the analyses stratified by histology, unadjusted models.

The first two SNPs are on chromosome 6p21, the last SNP is on chromosome 6p22.

[a]NCBI dbSNP identifier.
[b]Major allele, minor allele.
[c]Chromosome and NCBI Human genome Build 36 location.
[d]Gene neighborhood within 20 kb upstream and 10 kb downstream of SNP.
[e]Minor allele frequency in controls/cases.
[f] Subjects categorized by smoking status; sm=smokers
[g]1d.f. Wald test.
[h]OR, per-allele odds ratio; CI, 95% confidence interval.