## Supplementary Materials

## Supplementary Methods

**Study Subjects.** The LSC has been actively enrolling smokers from the Albuquerque, NM metropolitan area since 2001. Longitudinal studies to predict lung cancer and other chronic pulmonary diseases are conducted through assessing biomarkers present in sputum, blood, and urine samples (2–4). Enrollment was restricted to current and former smokers age 40 to 74 y with a minimum of 10 pack-years of smoking and no personal history of lung cancer. A detailed questionnaire written in English was used to collect information on demographics, medical, cigarette smoking, and exposure history, socioeconomic status, diet, and quality of life. Sputum samples were collected by induction and stored in Saccomanno's fixative. Pulmonary function testing was performed at each visit. All participants signed a consent form, and the Western Institutional Review Board approved this project. The GWAS discovery set was comprised of 1200 Caucasian (self-reported) smokers with methylation status of 12 tumor suppressor genes measured in sputum DNA samples.

The PLuSS Cohort was established in 2002 to support translational studies of the Pittsburgh Lung Cancer Specialized Programs of Research Excellence (5). Eligibility criteria for inclusion were 50–79 years old; smoke half pack cigarettes per day or more for at least 25 years; if quit, smoking cessation was no more than 10 years; and no personal history of lung cancer. Enrollment of 3,638 persons was completed in 2005 along with pulmonary function testing, blood collection, and two screening CTs separated by 1 year on 90% of participants. Home collection of sputum was initiated in 2006. The replication set comprised 718 Caucasian (self-identified) smokers with methylation status of 12 tumor suppressor genes measured in sputum DNA samples. Female smokers were over-sampled in the PLuSS to match the sex distribution in the LSC because sex is a major clinical risk factor for gene methylation. The prevalence for gene methylation in sputum from PLuSS participants was comparable, albeit slightly lower than seen in the LSC. Twenty-five percent of subjects in the PLuSS set had genetic ancestry analyzed previously with > 99.7% concordance to self-reported Caucasian ethnicity (6).

**Pathway Analysis.** A LD-based clumping approach in PLINK (version 1.06) was used to identify candidate genomic intervals that contained SNPs associated with risk for gene methylation. We started with 742 index SNPs with MAF $\geq 0.05$ and $P \leq 5 \times 10^{-4}$ for association with risk for gene methylation in the GWAS. Three hundred and seventy-six independent candidate genomic intervals were defined by identification of additional SNPs that located within 250kb upstream and downstream of the index SNPs, that were in LD with the index SNPs ($r^2 \geq 0.2$), and that were associated with risk for gene methylation at $P \leq 0.05$. Each individual genomic interval was indexed by the SNP most significantly associated with the risk for gene methylation within it and has a median size of 50 kb (mean = 83 kb). Then, a total of 389 genes located within these candidate genomic intervals were identified by expanding 20kb upstream and downstream of the coding area of the genes. The gene list was then applied in the pathway analysis using Ingenuity software.

**Measurement of DSBR capacity.** PHA-stimulated lymphocytes were treated with bleomycin to evaluate the generation of chromatid breaks as an index of DSBR capacity (4). Sixty-seven hours after PHA stimulation, the culture was split into two T25 flasks and treated with bleomycin or

vehicle for 5 h. The final concentration for bleomycin in culture medium was 3 units per liter, a concentration defined through dose-response studies using isolated lymphocytes from cohort subjects and two lymphoblastoid cells lines: GM02782 (mutant ATM) and GM00131 (wild-type ATM) (data not shown). The dose selected was within the linear dose-response range and caused obvious genotoxicity, but minimal cytotoxicity. One hour before harvest, colcemid was added to the cultures at a final concentration of 0.06 mg/ml. Slides were prepared according to conventional procedures and 100 well-spread metaphases were examined for chromatid breaks using the criteria of Hsu et al (7).

**Statistical Analysis.** Logistic regression models were used to evaluate the association between rs73371737 and risk for CMH with adjustment for age and sex or NSCLC with adjustment for age, sex, and smoking history. A meta-analysis was conducted to calculate the combined P values from multiple studies. Generalized linear models were used to assess the associations between SNPs and gene expression (delta Ct relative to β-actin) in primary HBECs and normal lung tissues with adjustment for smoking status; between rs73371737 and pack years (log transformed) in NSCLC cases from New Mexico and Pittsburgh with adjustment for age, sex, and cohort; and between composite risk score in double strand break repair by homologous recombination (DSBR-HR) pathway ($\geq$ 4 risk alleles versus $\leq$1 risk allele) and cells with chromatid breaks per 100 cells with adjustment for age, sex, smoking history, and methylation status.

**Supplementary Table 1.** Demographics for three non-small cell lung cancer (NSCLC) case-control studies

| Variable | New Mexico | | | Pittsburgh | | | MDACC | | |
|---|---|---|---|---|---|---|---|---|---|
| | Control | Case | P value | Control | Case | P value | Control | Case | P value |
| Design | Population-based case-control study[f] | | | Population-based case-control study[f] | | | Hospital-based case-control study | | |
| N | 330 | 335 | | 1216 | 751 | | 1018 | 1042 | |
| Age, yrs, mean (SD) | 65.7 (8.8) | 63.8 (9.5) | 0.0072 [a] | 59.4 ± 6.8 | 68.0 ± 8.2 | $< 1 \times 10^{-18}$ [a] | 61.3 ± 8.0 | 62.4 ± 9.6 | 0.0062 [a] |
| Male sex, % | 62.7 | 66.6 | 0.30 [b] | 52.6 | 51.7 | 0.68 [b] | 57.8 | 58.2 | 0.8552 [b] |
| NHW ethnicity, % | 100 | 100 | | 100 | 100 | | 100 | 100 | |
| Current smoker, % | 29.1 | 31.0 | 0.58 [b] | NA [e] | NA [e] | | 43.9 | 49.8 | 0.0073 [b] |
| Pack years (SD) | 44.7 (28.5) | 58.4 (33.4) | < 0.0001 [c] | 51.6 ± 21.9 | 55.2 ± 27.9 | 0.054 [c] | 48.5 ± 28.9 | 54.9 ± 29.9 | <0.0001 [c] |
| Tumor stage | | | | | | | | | |
| I and II, % | | 44.4 | | | 61.7 | | | 32.5 | |
| III and IV, % | | 55.6 | | | 38.3 | | | 61.8 | |
| Others, % | | | | | | | | 5.7 | |
| Histology | | | | | | | | | |
| Adeno, % | | 52.1 | | | 49.0 | | | 52.4 | |
| Squam, % | | 29.4 | | | 37.2 | | | 28.2 | |
| Others, % [d] | | 18.5 | | | 13.8 | | | 19.4 | |

[a] Two-sided two-sample *t* test between cases and controls.
[b] $\chi^2$ test for differences between cases and controls.
[c] Two-sided Wilcoxon rank sum test between cases and controls.
[d] Others included large cell lung cancer, poorly differentiated, and mixed NSCLC.
[e] Not available.
[f] Information for controls was from baseline enrollment. All controls remained cancer-free in the follow-ups.

**Supplementary Table 2.** Demographics for NCI NSCLC case-control study (PLCO and EAGLE)

| Variable | Control | Case | P value |
|---|---|---|---|
| N | 1410 | 1609 | |
| PLCO, % | 37.4 | 21.3 | $< 1 \times 10^{-18}$ [a] |
| Age, % | | | |
| 59 or less | 23.6 | 21.9 | 0.00027 [a] |
| 60 – 64 | 23.6 | 20.3 | |
| 65 – 69 | 26.6 | 24.9 | |
| 70 – 74 | 18.4 | 20.6 | |
| 75 – 79 | 7.9 | 12.3 | |
| Male sex, % | 80.1 | 81.9 | 0.23 [a] |
| NHW ethnicity, % | 100 | 100 | |
| Current smoker, % | 55.2 | 51.7 | 0.052 [a] |
| Pack years, % | | | |
| 15-30 | 33.5 | 17.9 | $< 1 \times 10^{-18}$ [a] |
| 30-40 | 25.7 | 21.1 | |
| 40-50 | 15.3 | 18.5 | |
| 50-60 | 9.7 | 16.3 | |
| 60-70 | 5.0 | 8.3 | |
| 70-80 | 3.9 | 5.4 | |
| > 80 | 6.9 | 12.5 | |
| Tumor stage | | | |
| I and II, % | | 40.7 | |
| III and IV, % | | 59.3 | |
| Histology | | | |
| Adenocarcinoma, % | | 49.4 | |
| Squamous cell carcinoma, % | | 34.2 | |
| Other NSCLC, % [b] | | 16.4 | |

[a] $\chi^2$ test for differences between cases and controls

[b] Others included large cell lung cancer and other NSCLC.

**Supplementary Table 3**. Gene list for pathway analysis

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| ABCA1 | C1orf101 | DNAJC15 | FUT4 | KCTD16 | NEBL | PDE8B | RNF144A | TSC22D2 | PCDHA11 |
| ABCB4 | C1orf87 | DPCR1 | FYCO1 | KIF26B | NEK11 | PGAP1 | RPL5 | TSHR | PCDHA13 |
| ABCC8 | C1QL3 | DST | GABRG3 | KREMEN1 | NEU2 | PHF20 | RPUSD2 | TSPAN18 | MAS1L |
| ABL1 | C22orf31 | DUOX2 | GALK2 | KTELC1 | NFATC1 | PID1 | RYR2 | TXNDC3 | LOC651503 |
| ACCN1 | C2orf48 | DUSP10 | GALNT10 | L3MBTL3 | NGEF | PIK3AP1 | SAMD3 | UBD | HCG27 |
| ACTG1 | C2orf66 | DYNC1I1 | GALNTL4 | LAMB1 | NOP5/NOP58 | PITPNA | SAPS2 | UNC13A | CDSN |
| ACVR1 | C3orf1 | DYTN | GDF10 | LIN7A | NPHP4 | PIWIL4 | SBF1 | UNC5C | CCHCR1 |
| ADAM12 | C7orf50 | ECE2 | GDF2 | LMO4 | NPS | PKNOX2 | SCAND1 | USH1C | C6orf15 |
| ADCYAP1R1 | C9orf4 | EDN1 | GEN1 | LOC200726 | NSUN2 | PLAA | SDCBP2 | USP6NL | PCDHA12 |
| ALAD | C9orf43 | EEPD1 | GHR | LOC553158 | NTN1 | PLD1 | SERP1 | VARS2 | PCDHA10 |
| ALDH1A3 | C9orf82 | EFNB3 | GLRX3 | LRRC19 | ODF3L2 | PLD5 | SERPINA11 | VSTM2A | PCDHA1 |
| ALDH1L1 | CACNA1A | EIF2A | GNG7 | LRRK1 | OR10G4 | PLXNA2 | SERPINA12 | WDR79 | |
| ALS2CR13 | CASC5 | EIF4G1 | GPER | LRRN2 | OR10G8 | PLXNB2 | SERPINA9 | WWC1 | |
| ANAPC1 | CCR6 | EMR2 | GPR141 | LRRTM3 | OR10G9 | PMP22 | SFTPG | WWOX | |
| ANKRD44 | CCR9 | EMR3 | GPR146 | LTBP1 | OR10S1 | PMP22CD | SH2D4B | XKR5 | |
| ANKRD49 | CDGAP | EPB41L4B | GPR31 | LZTFL1 | OR11A1 | PNLIPRP1 | SLC12A8 | YIPF5 | |
| ANLN | CDH13 | EPHB1 | GPR83 | MACROD2 | OR12D2 | PNLIPRP2 | SLC1A1 | YPEL1 | |
| ARHGAP15 | CDH4 | ERBB4 | GRIN2A | MADCAM1 | OR12D3 | POLE3 | SLC23A2 | ZBTB7C | |
| ARHGAP24 | CDKL1 | ERGIC2 | GRM5 | MAML3 | OR14J1 | POLR1A | SLC43A2 | ZDHHC21 | |
| ARHGAP8 | CER1 | ESRRG | GSX1 | MAPK1 | OR2AK2 | POU5F1 | SLC4A5 | ZFAND2A | |
| ARID1B | CHP2 | EVI5 | GTF2H4 | MAPK11 | OR2H1 | PPM1F | SLC6A20 | ZFPM2 | |
| ASTE1 | CHRNA2 | EXOC2 | GTF3C3 | MAPK12 | OR2H2 | PRDM11 | SLIT3 | ZFYVE28 | |
| ASTN2 | CNTN4 | EZH2 | HAS2 | MBL2 | OR2L13 | PRKAG2 | SNPH | ZNF143 | |
| ATP10A | CNTNAP2 | FAM116B | HDAC10 | MCM5 | OR2L8 | PRKCB1 | SNX29 | ZNF160 | |
| ATP10B | COPS2 | FAM155A | HDHD3 | ME3 | OR2W3 | PSMB1 | SORD | ZNF254 | |
| ATP2C1 | COX19 | FAM49A | HECW2 | MEGF6 | OR4D5 | PSMD2 | SRD5A1 | ZNF333 | |
| ATP5S | CPO | FAM69A | HEG1 | MERTK | OR5V1 | PSORS1C1 | SRGAP3 | ZNF385B | |
| ATXN1 | CROT | FAM82A2 | HLA-B | MIA2 | OR6T1 | PSORS1C2 | STK35 | ZNF415 | |
| BACH2 | CSF1 | FAR2 | HLA-C | MICA | OR8D4 | PTCD3 | SUMO1 | ZNF488 | |
| BAHCC1 | CSMD1 | FBXO33 | HMOX1 | MICB | PARP12 | PTER | SYT13 | ZNRF3 | |
| BAIAP2 | CTAGE5 | FECH | HOXB13 | MOBKL1B | PARVB | PTK2B | TBP | ZSCAN20 | |
| BAZ2B | CTNNA3 | FGF7 | ICA1L | MRE11A | PAX5 | PTPRC | TCF19 | PCDHA7 | |
| BMPR2 | CUL1 | FGFR1OP | IFT74 | MRPL35 | PCDHA3 | QRFP | TEK | OR10C1 | |
| BOLA3 | CXCR6 | FHOD3 | IMMT | MSGN1 | PCDHAC2 | RAD18 | TEKT3 | HCP5 | |
| BRWD2 | CYP2W1 | FIBCD1 | INHBA | MSR1 | PCDHB1 | RAD51 | TET3 | OR10G7 | |
| BSPRY | DAPK1 | FKBP1A | IPO7 | MTHFD2 | PCDHB2 | RAPGEF4 | THRB | PCDHA2 | |
| C14orf145 | DDEF1 | FLJ16165 | JHDM1D | MUC16 | PCDHB3 | RBP3 | TMEM39A | PCDHA9 | |
| C15orf33 | DDR1 | FLRT3 | JUN | MXD4 | PCDHB4 | REEP1 | TMSB10 | PCDHAC1 | |
| C15orf57 | DGKH | FREM1 | KCMF1 | MYH11 | PCDHB5 | RGS3 | TOP3B | PCDHA8 | |
| C16orf63 | DHRS2 | FRMD6 | KCNC2 | MYRIP | PCDHB6 | RGS6 | TP53I11 | PCDHA5 | |
| C17orf70 | DLG2 | FSCN2 | KCNH5 | NAV2 | PDCD2 | RICH2 | TRIM35 | PCDHA6 | |
| C17orf92 | DNAH2 | FST | KCNJ11 | NDUFS4 | PDE6C | RNASET2 | TRIM58 | PCDHA4 | |

**Supplementary Table 4.** Association between SNPs in DSBR-HR genes and risk for gene methylation in LSC

| Gene | SNP | Allele[a] | MAF | OR (95%CI) | P value[b] |
|---|---|---|---|---|---|
| GEN1 | rs12470034 | C/T | 0.49 | 1.17 (1.08 – 1.27) | 0.00011 |
| ABL1 | rs7025469 | A/G | 0.27 | 1.18 (1.08 – 1.30) | 0.00033 |
| MRE11A | rs10741490 | G/A | 0.37 | 1.18 (1.08 – 1.29) | 0.00022 |
| RAD51 | rs11070284 | C/T | 0.15 | 0.81 (0.72 – 0.91) | 0.00047 |
| Risk score[c] | | | | 1.20 (1.14 – 1.26) | $1.20 \times 10^{-13}$ |

[a] Allele after '/' is the minor allele and test allele.

[b] Age, sex, current smoking status, and pack years were included in GEE model for covariate adjustment.

[c] Risk score = rs12470034_T + rs7025469_G + rs10741490_A – rs11070284_T.

**Supplementary Table 5.** Haplotype-based analyses for the two top SNPs on chromosome 15q12 with the risk for gene methylation

| Haplotype[a] | LSC | | | PLuSS | | | Combined | |
|---|---|---|---|---|---|---|---|---|
| | Freq[b] | OR (95% CI) | P value | Freq[b] | OR (95% CI) | P value | OR (95% CI)[c] | P value |
| CC | 57.2 | 1.00 | REF | 56.5 | 1.00 | REF | 1.00 | REF |
| CT | 32.2 | 0.82 (0.74 – 0.91) | 0.00014 | 33.7 | 0.87 (0.72 – 1.04) | 0.13 | 0.83 (0.76 – 0.91) | $6.2 \times 10^{-5}$ |
| AC | 6.6 | 1.32 (1.10 – 1.59) | 0.0030 | 5.4 | 1.59 (1.06 – 2.39) | 0.025 | 1.36 (1.15 – 1.61) | $3.0 \times 10^{-4}$ |
| AT | 3.9 | 1.12 (0.90 – 1.40) | 0.31 | 4.4 | 1.03 (0.67 – 1.58) | 0.91 | 1.10 (0.90 – 1.34) | 0.34 |
| Not CC | 42.8 | 1.00 | REF | 43.5 | 1.00 | REF | 1.00 | REF |
| CC | 57.2 | 1.10 (1.006 – 1.20) | 0.037 | 56.5 | 1.05 (0.90 – 1.22) | 0.54 | 1.09 (1.007 – 1.17) | 0.032 |
| Not CT | 67.8 | 1.00 | REF | 66.3 | 1.00 | REF | 1.00 | REF |
| CT | 32.2 | 0.78 (0.71 – 0.86) | $4.86 \times 10^{-7}$ | 33.7 | 0.82 (0.70 – 0.96) | 0.015 | 0.79 (0.73 – 0.86) | $1.9 \times 10^{-8}$ |
| Not AC | 93.4 | 1.00 | REF | 94.6 | 1.00 | REF | 1.00 | REF |
| AC | 6.6 | 1.52 (1.29 – 1.80) | $1.07 \times 10^{-6}$ | 5.4 | 1.79 (1.26 – 2.54) | $1.16 \times 10^{-3}$ | 1.57 (1.35 – 1.82) | $2.5 \times 10^{-9}$ |
| Not AT | 96.1 | 1.00 | REF | 95.6 | 1.00 | REF | 1.00 | REF |
| AT | 3.9 | 1.23 (0.99 – 1.52) | 0.065 | 4.4 | 1.24 (0.86 – 1.78) | 0.24 | 1.23 (1.025 – 1.48) | 0.026 |

[a] Haplotype was constructed for rs73371737 and rs7179575 using PHASE. LD between two SNPs: D' = 0.02, $r^2 = 0.000098$.
[b] Population haplotype frequency.
[c] Fixed model.

**Supplementary Table 6.** Gene expression of GABRB3, GABRA5 and GABRG3 in HBECs (mean ± SD)

| Gene | No. tested | No. determined Ct[a] | Ct | Delta Ct (relative to β-actin) |
|---|---|---|---|---|
| GABRB3 | 48 | 48 | 33.8 ± 3.8 | 12.3 ± 3.5 |
| GABRA5 | 48 | 31 | 42.5 ± 4.6 | 21.2 ± 4.0 |
| GABRG3 | 48 | 16 | 45.3 ± 3.6 | 25.5 ± 3.5 |

[a] TaqMan quantitative real time PCR was ran for 50 cycles.

**Supplementary Table 7.** Association between rs7179575 and expression of GABRB3 in HBECs and normal lungs

| Genotype | HBECs | | | Normal lungs | | |
|---|---|---|---|---|---|---|
| | n | lsMean ± Stderr[a] | P value for trend | n | lsMean ± Stderr[a] | P value for trend |
| CC | 22 | 13.3 ± 0.7 | 0.095 | 16 | 10.7 ± 0.3 | 0.77 |
| CT | 22 | 11.6 ± 0.7 | | 13 | 10.3 ± 0.3 | |
| TT | 5 | 11.1 ± 1.5 | | 11 | 10.7 ± 0.3 | |

[a] Delta CT was calculated using β-actin as the endogenous control. Smoking status was adjusted in the generalized linear models. CC, CT, or TT genotypes were coded as 0, 1, or 2. lsMean, least square mean; Stderr, standard error.

**Supplementary Table 8.** Association between rs73371737 and risk for CMH in Caucasian former smokers[a]

| rs73371737 | LSC | | PLuSS | |
|---|---|---|---|---|
| | non-CMH | CMH | non-CMH | CMH |
| CC | 536 (84.3) | 102 (75.0) | 237 (84.6) | 23 (74.2) |
| CA | 93 (14.6) | 28 (20.6) | 42 (15.0) | 6 (19.5) |
| AA | 7 (1.1) | 6 (4.41) | 1 (0.36) | 2 (6.45) |
| Cochran-Armitage Trend Test | 0.0021 | | 0.0303 | |

[a] Analyses were conducted in 772 Caucasian former smokers from LSC and 311 Caucasian former smokers from PLuSS. Age and sex were included for covariate adjustment in the logistic regression models.

**Supplementary Table 9.** Meta-analysis for association between rs7179575 and risk for NSCLC

| Variable[a] | OR and 95%CI | P value | Weight | P for Q-test | $I^2$ | $Tau^2$ |
|---|---|---|---|---|---|---|
| rs7179575 (CT vs. CC) | | | | | | |
| Lovelace | 0.83 (0.59 – 1.17) | 0.29 | 32.78 | 0.708 | 0.000 | 0.000 |
| Pittsburgh | 0.98 (0.80 – 1.19) | 0.83 | 97.45 | | | |
| NCI | 0.92 (0.78 – 1.09) | 0.36 | 137.22 | | | |
| MDACC | 1.04 (0.82 – 1.32) | 0.75 | 67.79 | | | |
| Combined | 0.95 (0.85 – 1.06) | 0.36 | | | | |
| rs7179575 (TT vs. CC) | | | | | | |
| Lovelace | 0.71 (0.42 – 1.19) | 0.19 | 14.17 | 0.229 | 30.56 | 0.107 |
| Pittsburgh | 0.85 (0.64 – 1.13) | 0.27 | 47.54 | | | |
| NCI | 0.90 (0.71 – 1.13) | 0.38 | 71.15 | | | |
| MDACC | 1.24 (0.89 – 1.74) | 0.21 | 34.19 | | | |
| Combined | 0.93 (0.80 – 1.08) | 0.33 | | | | |

[a] CC is wild homozygote. TT is variant homozygote. Fixed model was selected for combing results from the four studies. Rs7179575 was imputed in MDACC with dosage $R^2 = 0.75$.

**Supplementary Table 10.** Association between rs73371737 and pack years in NSCLC cases[a]

| rs73371737 | n | Geometric mean (95% CI) | P value | P value |
|---|---|---|---|---|
| CC | 869 | 48.8 (47.0 – 50.7) | REF | 0.56 |
| CA | 201 | 47.6 (43.9 – 51.5) | 0.56 | REF |
| AA | 16 | 35.2 (26.6 – 46.6) | 0.023 | 0.043 |

[a] The P value for F test is 0.069 with adjustment for age, sex, and cohort (New Mexico vs Pittsburgh) in generalized linear model with log transformed pack years as the outcome.

**Supplementary Table 11.** List of genes containing SNPs associated with risk for gene methylation in pathway analysis

| Ingenuity Canonical Pathways | -log(p-value) | Ratio | Genes |
|---|---|---|---|
| DNA Double-Strand Break Repair by Homologous Recombination | 3.76 | 0.250 | RAD51,GEN1,ABL1,MRE11A |
| CCR5 Signaling in Macrophages | 3.25 | 0.086 | JUN,MAPK1,PTK2B,MAPK12,MAPK11,GNG7,PRKCB |
| BMP signaling pathway | 2.99 | 0.090 | JUN,FST,MAPK1,PRKAG2,BMPR2,MAPK12,MAPK11 |
| April Mediated Signaling | 2.91 | 0.119 | JUN,MAPK1,MAPK12,MAPK11,NFATC1 |
| Netrin Signaling | 2.86 | 0.111 | RYR2,PRKAG2,NTN1,NFATC1,UNC5C |
| B Cell Activating Factor Signaling | 2.81 | 0.114 | JUN,MAPK1,MAPK12,MAPK11,NFATC1 |
| ATM Signaling | 2.80 | 0.098 | RAD51,JUN,ABL1,MRE11A,MAPK12,MAPK11 |
| IL-17A Signaling in Gastric Cells | 2.74 | 0.160 | JUN,MAPK1,MAPK12,MAPK11 |
| Axonal Guidance Signaling | 2.71 | 0.042 | SLIT3,NGEF,RGS3,MAPK1,ABL1,PLXNA2,NTN1,GNG7, NFATC1,EPHB1,SRGAP3,ADAM12,ECE2,BAIAP2,PRKAG2, PLXNB2,EFNB3,PRKCB,UNC5C |
| UVC-Induced MAPK Signaling | 2.71 | 0.119 | JUN,MAPK1,MAPK12,MAPK11,PRKCB |

**Supplementary Table 12.** Association between six lung cancer risk loci and risk for gene methylation in LSC

| Locus | Lung cancer GWAS[a] | | Methylation GWAS | | | | | |
|---|---|---|---|---|---|---|---|---|
| | SNP | Allele[b] | SNP | Minor allele[c] | $R^{2d}$ | MAF | OR(95% CI) | P value |
| 15q25 | rs1051730 | C/T | Same | T | 1 | 0.36 | 0.95 (0.87 – 1.04) | 0.26 |
| | rs6495309 | T/C | rs6495308 | C | 0.85 | 0.22 | 1.05 (0.95 – 1.16) | 0.34 |
| | rs951266 | C/T | Same | T | 1 | 0.36 | 0.95 (0.87 – 1.03) | 0.23 |
| 5p15.33 | rs2736100 | T/G | Same | G | 1 | 0.49 | 0.98 (0.90 – 1.06) | 0.57 |
| | rs401681 | T/C | rs380286 | A | 1 | 0.44 | 0.96 (0.88 – 1.04) | 0.30 |
| **6p21- p22** | **rs3117582** | **A/C** | **Same** | **C** | **1** | **0.11** | **1.13 (1.00 – 1.28)** | **0.05** |
| | **rs3131379** | **G/A** | **Same** | **A** | **1** | **0.11** | **1.14 (1.01 – 1.29)** | **0.032** |
| 12p13.33 | rs10849605 | T/C | rs11571379 | G | 0.93 | 0.48 | 1.01 (0.93 – 1.09) | 0.80 |
| | rs3748522 | A/C | Same | A | 1 | 0.44 | 1.03 (0.94 – 1.12) | 0.55 |
| 9q21 | rs1333040 | T/C | Same | C | 1 | 0.43 | 0.95 (0.87 – 1.03) | 0.20 |
| 2q32.1 | rs11683501 | A/G | Same | G | 1 | 0.47 | 1.02 (0.94 – 1.11) | 0.62 |

[a] Information for SNPs and associations with risk for lung cancer was from reference 27.
[b] Allele after '/' is the risk allele for lung cancer.
[c] Minor allele is the test allele.
[d] $R^2$ is between the two SNPs listed for lung cancer GWAS (1) and methylation GWAS (this study).

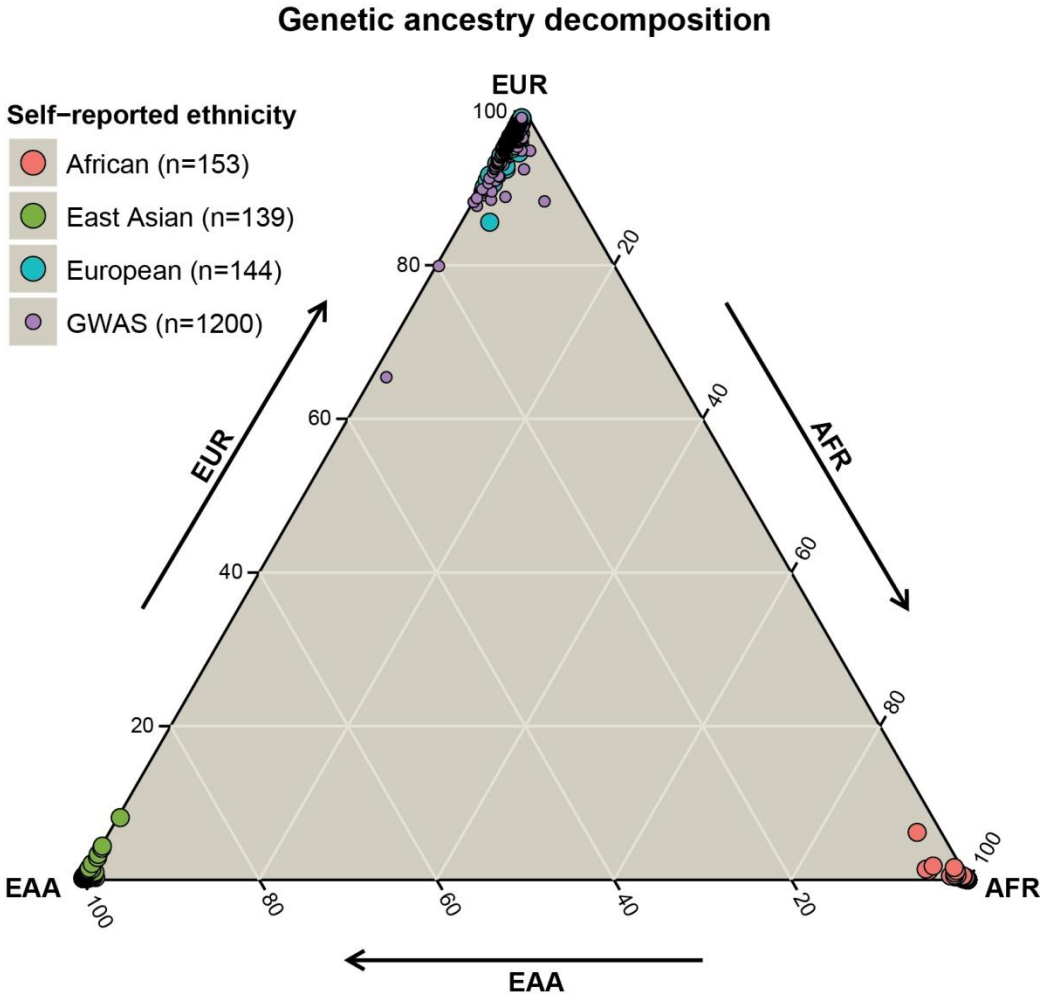
**Supplementary Table 13.** MAF of rs73371737 across four ethnic groups in 1000 Genomes project.

| Ethnic group[a] | Population[b] | Allele | |
|---|---|---|---|
| | | C | A |
| EUR (n = 379) | CEU, FIN, GBR, IBS, TSI | 0.898 | 0.102 |
| AFR (n = 246) | YRI, LWK, ASW | 0.711 | 0.289 |
| ASN (n = 286) | CHB, CHS, JPT | 1 | 0 |
| AMR (n = 181) | CLM, MXL, PUR | 0.914 | 0.086 |

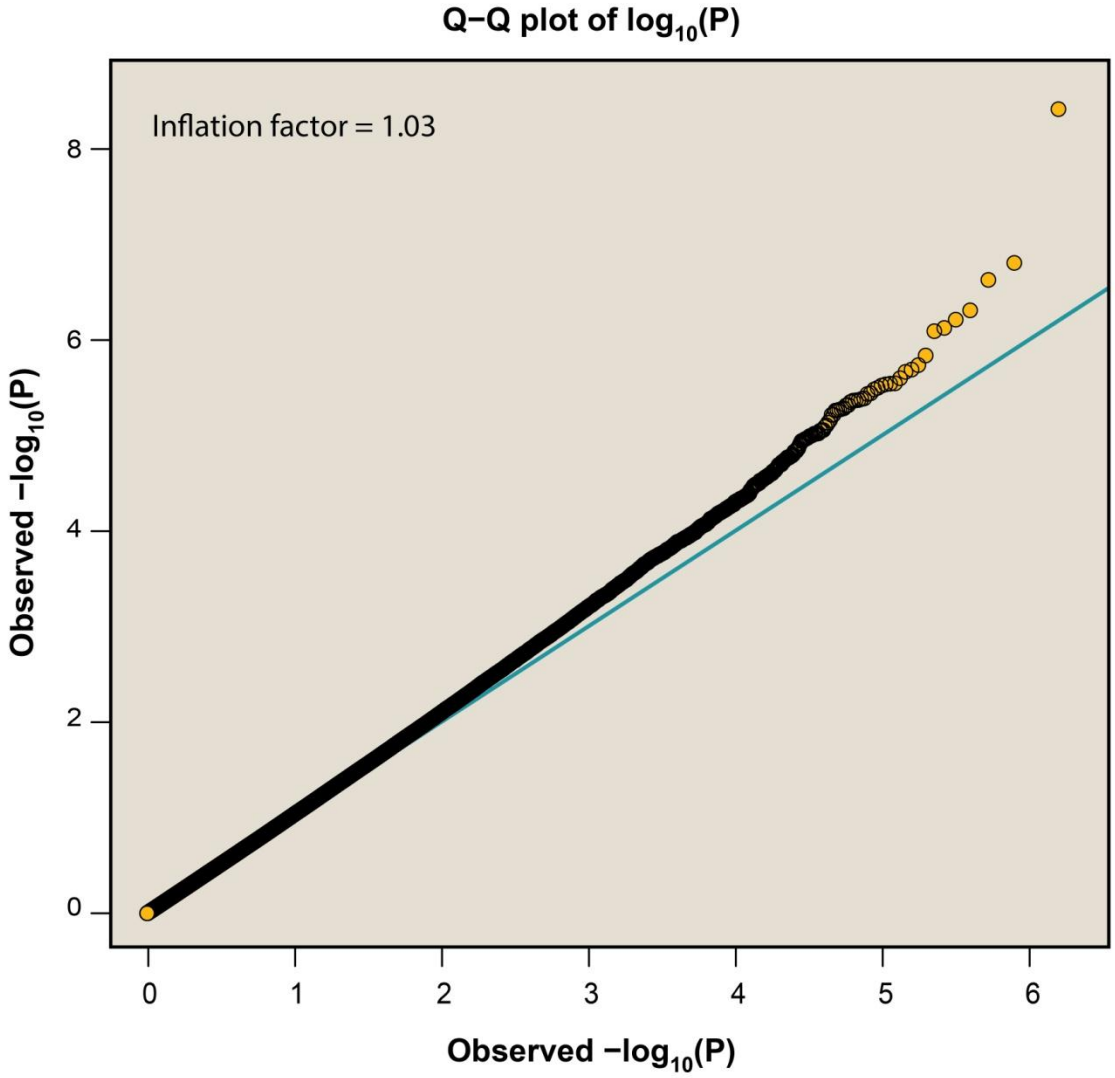
[a] AFR, African; AMR, Ad Mixed American; ASN, East Asian; EUR, European

[b] CEU, Utah Residents (CEPH) with Northern and Western European ancestry; FIN, Finnish in Finland; GBR, British in England and Scotland; IBS, Iberian population in Spain; TSI, Toscani in Italia; YRI, Yoruba in Ibadan, Nigera; LWK, Luhya in Webuye, Kenya; ASW, Americans of African Ancestry in SW USA; CHB, Han Chinese in Bejing, China; CHS, Southern Han Chinese; JPT, Japanese in Tokyo, Japan; CLM, Colombians from Medellin, Colombia; MXL, Mexican Ancestry from Los Angeles USA; PUR, Puerto Ricans from Puerto Rico.

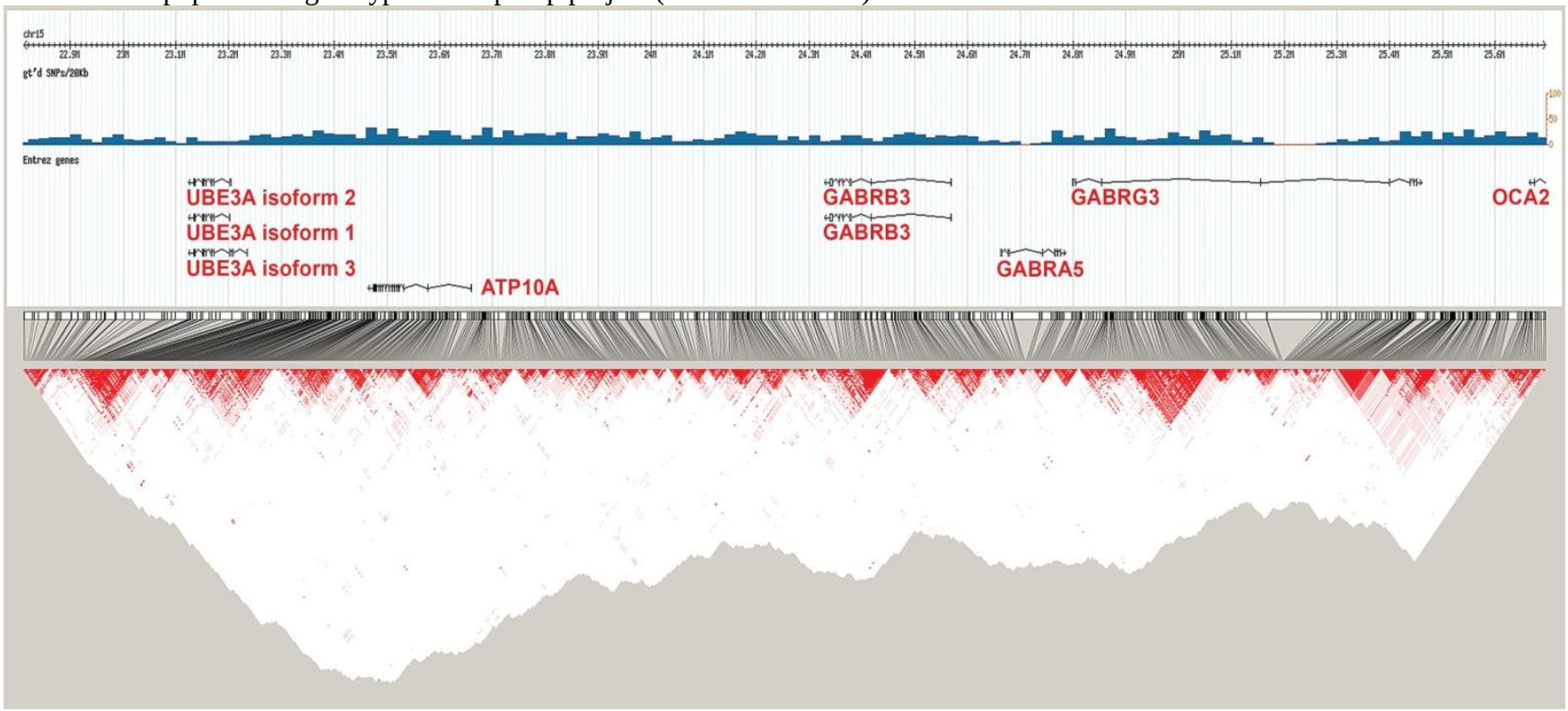
**Supplementary Figure 1.** Ternary plot for genetic ancestry of European, East Asian, and African in EUR (n = 144), ASN (n = 139), AFR (n = 153), and Caucasian (self-identified) smokers in LSC (n = 1200). Genetic ancestry was analyzed using the Bayesian Markov Chain Monte Carlo algorithm implemented in STRUCTURE 2.3.3 under the admixture model. The 1200 LSC smokers (GWAS) overlapped almost completely with European subjects except two LSC subjects that were removed from genetic association analysis due to having < 85% European ancestry.



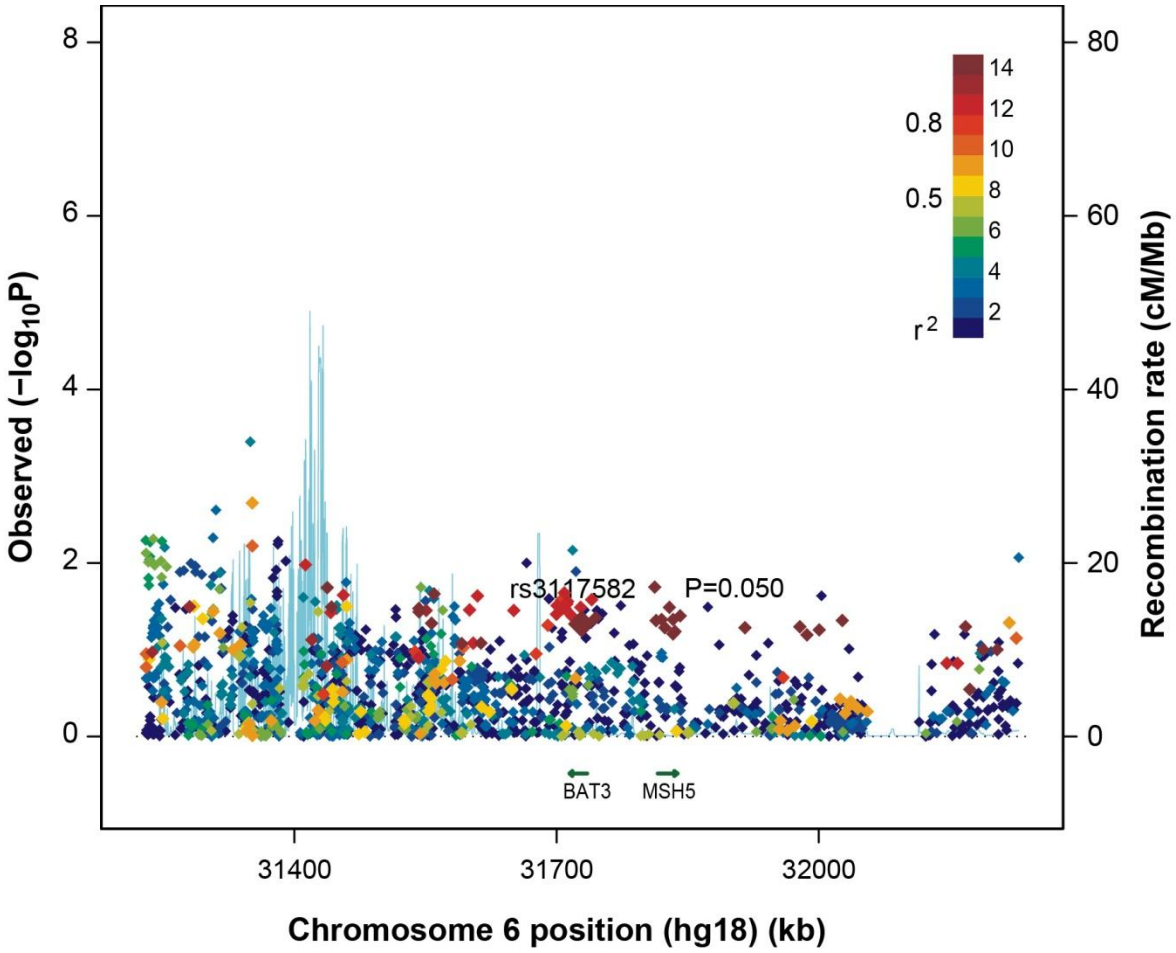
**Supplementary Figure 2.** Quantile-quantile plots for genetic associations with risk for gene methylation with observed P values plotted as a function of theoretical P values.



**Supplementary Figure 3.** Linkage disequlibirum mapping for common SNPs (MAF ≥ 0.05) within chr15: 22.8 – 25.7 Mb (B36) in CEU and TSI populations genotyped in HapMap project (Phase 3 release 2).



**Supplementary Figure 4.** Associations for variants within 500kb upstream and downstream of rs3117582 and risk for gene methylation in LSC. Multiple variants in high LD with rs3117582 ($R^2 > 0.8$) were significantly ($P \leq 0.05$) associated with risk for gene methylation. This area has 63 genes annotated in the University of California Santa Cruz Genome Browser March 2006 assembly (Genome Build 36). Due to space limitation, only two genes (BAT3 and MSH5) that have been suggested in literature (1) to be responsible for the genetic association signals observed in this locus for lung cancer are labeled. BAT3 is implicated in the control of apoptosis and regulating heat shock protein. MSH5 is involved in DNA mismatch repair and meiotic recombination.



**Supplementary References**

1. Timofeeva MN, Hung RJ, Rafnar T, et al.Influence of common genetic variation on lung cancer risk: meta-analysis of 14 900 cases and 29 485 controls. *Hum Mol Genet.* 2012;21(22):4980-4995.

2. Leng S, Stidley CA, Liu Y, et al. Genetic Determinants for promoter hypermethylation in the lungs of smokers: a candidate gene-based study. *Cancer Res.* 2012;72(3):707–715.3. Leng S, Liu Y, Thomas CL, et al. Native American ancestry affects the risk for gene methylation in the lungs of Hispanic smokers from New Mexico. *Am. J. Respir. Crit. Care Med.* 2013;188(9):1110–1116.

4.Leng S, Stidley CA, Willink R, et al. Double-strand break damage and associated DNA repair genes predispose smokers to gene methylation. *Cancer Res.* 2008;68(8):3049–3056.

5. Wilson DO, Weissfeld JL, Fuhrman CR, et al. The Pittsburgh Lung Screening Study (PLuSS): outcomes within 3 years of a first computed tomography scan. *Am. J. Respir. Crit. Care Med.* 2008;178(9):956–961.

6. Buch SC, Diergaarde B, Nukui T, et al. Genetic variability in DNA repair and cell cycle control pathway genes and risk of smoking-related lung cancer. *Mol. Carcinog.* 2012;51(suppl 1):E11–E20.

7. Hsu TC, Johnston DA, Cherry LM, et al, Sensitivity to genotoxic effects of bleomycin in humans: possible relationship to environmental carcinogenesis. Int. J. Cancer 1989;43(3):403-409.