|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Supplementary Table S1.** Comparison of characteristics between the PLCO trial and the HLCS study | | | | | | |
| **Characteristics** | **PLCO** | |  | **HLCS** | | ***P* \*** |
| **Frequency** | **Deaths (%)** |  | **Frequency** | **Deaths (%)** |
| Total | 1,185 | 798 (67.3) |  | 984 | 665 (67.5) |  |
| Median overall survival (months) | 23.8 |  |  | 39.9 |  |  |
| Age |  |  |  |  |  |  |
| ≤71 | 636 | 400 (62.9) |  | 654 | 428 (65.4) | <0.0001 |
| >71 | 549 | 398 (72.5) |  | 330 | 237 (71.8) |  |
| Sex |  |  |  |  |  |  |
| Male | 698 | 507 (72.6) |  | 507 | 379 (74.7) | 0.0006 |
| Female | 487 | 291 (59.8) |  | 477 | 286 (59.9) |  |
| Smoking status |  |  |  |  |  |  |
| Never | 115 | 63 (54.8) |  | 92 | 52 (56.5) | 0.166 |
| Current | 423 | 272 (64.3) |  | 390 | 266 (68.2) |  |
| Former | 647 | 463 (71.6) |  | 502 | 347 (69.1) |  |
| Histology |  |  |  |  |  |  |
| Adenocarcinoma | 577 | 348 (60.3) |  | 597 | 378 (63.3) | <0.0001 |
| Squamous cell carcinoma | 285 | 192 (67.4) |  | 216 | 156 (72.2) |  |
| Others | 323 | 258 (79.9) |  | 171 | 131 (76.6) |  |
| Stage |  |  |  |  |  |  |
| I - IIIA | 655 | 315 (48.1) |  | 606 | 352 (58.0) | 0.003 |
| IIIB - IV | 528 | 482 (91.3) |  | 377 | 313 (83.0) |  |
| *Missing* | 2 |  |  | -- |  |  |
| Abbreviations: PLCO, the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial; HLCS, Harvard Lung Cancer Susceptibility Study.  \* Chi-square test for the comparison of characteristics between the PLCO trial and the HLCS study for each clinical variable. | | | | | | |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Supplementary Table 2.** List of 220 selected immunity B cell pathway genes used in the discovery analysis | | | | |
| **Dataset** | **Name of pathway** | **Selected genes a** | **Number of genes** | | |
| GO | GOBP\_B\_CELL\_MEDIATED\_IMMUNITY | *AICDA, APLF, ATAD5, BATF, BCL10, BCL3, BCL6, BTK, C17orf99, C1QA, C1QB, C1QBP, C1QC, C1R, C1RL, C1S, C2, C3, C4A, C4B, C4BPA, C4BPB, C5, C6, C7, C8A, C8B, C8G, C9, CARD9, CCR6, CD19, CD226, CD27, CD28, CD40, CD40LG, CD46, CD55, CD70, CD74, CD81, CFI, CLCF1, CLU, CR1, CR1L, CR2, CSF2RB, ERCC1, EXO1, EXOSC3, EXOSC6, FCER1A, CER1G, FCER2, FCGR1A, FCGR2B, FCGR3A, FOXJ1, FOXP3, GAPT, HLA-E, HLA-G, HMCES, HPX, HSPD1, IGHA1, IGHA2, IGHD, IGHE, IGHG1, IGHG2, IGHG3, IGHG4, IGHM, IGHV1-18, IGHV1-24, IGHV1-3, IGHV1-45, IGHV1-58, IGHV1-69, IGHV1-69-2, IGHV1-69D, IGHV2-26, IGHV2-5, IGHV2-70, IGHV2-70D, IGHV3-11, IGHV3-13, IGHV3-15, IGHV3-16, IGHV3-20, IGHV3-21, IGHV3-23, IGHV3-30, IGHV3-33, IGHV3-35, IGHV3-38, IGHV3-43, IGHV3-48, IGHV3-49, IGHV3-53, IGHV3-64, IGHV3-64D, IGHV3-66, IGHV3-7, IGHV3-72, IGHV3-73, IGHV3-74, IGHV4-28, IGHV4-31, IGHV4-34, IGHV4-39, IGHV4-4, IGHV4-59, IGHV4-61, IGHV5-10-1, IGHV5-51, IGHV6-1, IGHV7-4-1, IGHV7-81, IGHV8-51-1, IGKC, IGLC1, IGLC2, IGLC3, IGLC6, IGLC7, IGLL1, IGLL5, IL10, IL13RA2, IL2, IL21R, IL27RA, IL2RB, IL4, IL4R, IL9, IL9R, INPP5D, IRF7, KMT5B, KMT5C, LIG4, LTA, MAD2L2, MASP2, MBL2, MLH1, MPL, MSH2, MSH6, MYD88, NBN, NCR3LG1, NDFIP1, NECTIN2, NFKBIZ, NOD2, NSD2, PARP3, PAXIP1, PMS2, PRKCD, PTPN6, PTPRC, RIF1, RNF168, RNF8, SANBR, SERPING1, SHLD1, SHLD2, SHLD3, SLA2, SLC15A4, STAT6, SUPT6H, SUSD4, SVEP1, SWAP70, TBX21, TCIRG1, TFRC, TGFB1, THOC1, TLR8, TNF, TNFSF13, TNFSF4, TP53BP1, TRAF3IP2, TREM2, TREX1, UNG, XCL1, ZP3* | 199 | | |
| GO | GOBP\_B\_CELL\_ACTIVATION\_INVOLVED\_IN\_IMMUNE\_RESPONSE | *ABL1, ADA, AICDA, APLF, ATAD5, BATF, BCL3, BCL6, C17orf99, CCR6, CD180, CD19, CD28, CD40, CD40LG, CDH17, CLCF1, CR1, DLL1, DOCK10, DOCK11, ERCC1, EXO1, EXOSC3, EXOSC6, FCGR2B, FOXP3, GAPT, GPR183, HMCES, HSPD1, IFNB1, IL10, IL2, IL21, IL27RA, IL4, IL6, IRF8, ITFG2, ITM2A, KMT5B, KMT5C, LFNG, LGALS1, LIG4, MAD2L2, MFNG, MLH1, MSH2, MSH6, NBN, NDFIP1, NFKBIZ, NKX2-3, NOTCH2, NSD2, PARP3, PAXIP1, PHF14, PLCG2, PLCL2, PMS2, POU2AF1, PTK2B, PTPRC, RAG2, RIF1, RNF168, RNF8, SANBR, SHLD1, SHLD2, SHLD3, SLC15A4, SPI1, ST3GAL1, STAT6, SUPT6H, SWAP70, TBX21, TFRC, TGFB1, THOC1, TLR4, TNFSF13, TNFSF4, TP53BP1, UNG, XBP1* | 90 | | |
| GO | GOBP\_MATURE\_B\_CELL\_DIFFERENTIATION\_INVOLVED\_IN\_IMMUNE\_RESPONSE | *ADA, BCL3, BCL6, C17orf99, CDH17, CR1, DLL1, DOCK10, DOCK11, FCGR2B, GPR183, IL10, IL2, IL21, IL6, IRF8, ITFG2, ITM2A, LFNG, LGALS1, MFNG, NFKBIZ, NKX2-3, NOTCH2, PHF14, PLCG2, POU2AF1, PTK2B, RAG2, SPI1, ST3GAL1, XBP1* | 32 | | |
| GO | GOBP\_B\_CELL\_PROLIFERATION\_INVOLVED\_IN\_IMMUNE\_RESPONSE | *ABL1, CD180, CD19, GAPT, PLCL2, TLR4* | 6 | | |
| Total |  |  | 220 b | | |
| a Genes were selected based on online datasets;  b 99 duplicated genes and eight genes on the X chromosome had been removed;  Keyword: “B cell” AND “immunity”;  Organism: Homo sapiens. | | | |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Supplementary Table S3.** Associations of the first 10 principal components and OS of NSCLC in the PLCO trial | | | | |
| **PC\*** | **Parameter Estimate** | **Standard Error** | **Chi-Square** | ***P*** |
| **PC1** | **4.821** | **1.353** | **12.697** | **<0.001** |
| **PC2** | **-0.681** | **1.228** | **0.308** | **0.579** |
| **PC3** | **-3.054** | **0.949** | **10.351** | **0.001** |
| **PC4** | **-2.837** | **1.246** | **5.184** | **0.023** |
| PC5 | -0.910 | 1.232 | 0.546 | 0.460 |
| PC6 | 1.355 | 1.252 | 1.172 | 0.279 |
| PC7 | -0.236 | 1.218 | 0.038 | 0.846 |
| PC8 | -1.684 | 1.322 | 1.622 | 0.203 |
| PC9 | -1.886 | 1.267 | 2.216 | 0.137 |
| PC10 | 0.347 | 1.240 | 0.078 | 0.180 |
| \* The first four PC were used for adjustment for population stratification in the multivariate analysis.  Abbreviations: OS, overall survival; NSCLC, non-small cell lung cancer; PLCO, the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial; PC, principal component. | | | | |

**Supplementary Table 4.** Function prediction for the three Validated SNPs

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **SNP** | **Gene** | **Chr** | **Genotyped** | **RegDB1** | **Haploreg v4.22** | | | | | |
| **Promoter histone marks** | **Enhancer histone marks** | **DNAse** | **Motifs changed** | **Selected eQTL hits** | **dbSNP func annot** |
| **rs13385922** | *INPP5D* | 2 | Yes | 5 | -- | -- | -- | ERalpha-a, HES1, HNF1 | 1 hit | -- |
| rs6836770 | *CFI* | 4 | Yes | 5 | -- | LIV | -- | 6 altered motifs | 1 hit | intronic |
| **rs3208406** | *EXOSC3* | 9 | No | 1f | -- | FAT | SKIN, SKIN | Rad21, SIX5 | 1 hit | missense |

Abbreviations: SNP, single nucleotide polymorphism; NSCLC, non-small cell lung cancer; Chr, chromosome; dbSNP func annot, dbSNP function annotation.

1 RegulomeDB: http://regulomedb.org/

2 Haploreg: http://archive.broadinstitute.org/mammals/haploreg/haploreg.php

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Supplementary Table S5.** Stratified analysis for associations between unfavorable genotypes and NSCLC survival in the PLCO trial | | | | | | | | | |
| **Characteristics** | **0-1 unfavorable genotype a** | **2-3 unfavorable genotype a** | **Multivariate Analysis b for OS** | | | | **Multivariate Analysis b for DSS** | | |
| **Frequency** | **Frequency** | **HR (95% CI)** | ***P*** | ***P* inter c** | **HR (95% CI)** | | ***P*** | ***P* inter c** |
| Age (years) |  |  |  |  | 0.181 |  | |  | 0.075 |
| ≤ 71 | 196 | 423 | 1.45 (1.16-1.82) | 0.001 |  | 1.64 (1.29-2.10) | | <0.0001 |  |
| > 71 | 154 | 374 | 1.26 (1.00-1.59) | 0.049 |  | 1.30 (1.02-1.67) | | 0.038 |  |
| Sex |  |  |  |  | 0.191 |  | |  | 0.489 |
| Male | 207 | 471 | 1.24 (1.01-1.51) | 0.038 |  | 1.38 (1.11-1.72) | | 0.004 |  |
| Female | 143 | 326 | 1.59 (1.21-2.11) | 0.001 |  | 1.64 (1.23-2.20) | | 0.001 |  |
| Smoking status |  |  |  |  | 0.231 |  | |  | 0.219 |
| Never | 35 | 76 | 0.71 (0.37-1.37) | 0.309 |  | 0.71 (0.37-1.37) | | 0.308 |  |
| Current | 126 | 282 | 1.37 (1.03-1.81) | 0.031 |  | 1.55 (1.14-2.11) | | 0.005 |  |
| Former | 189 | 439 | 1.44 (1.16-1.78) | 0.001 |  | 1.56 (1.24-1.96) | | 0.0002 |  |
| Histology |  |  |  |  | 0.701 |  | |  | 0.605 |
| Adeno | 168 | 393 | 1.19 (0.92-1.53) | 0.184 |  | 1.23 (0.94-1.61) | | 0.126 |  |
| Squamous | 79 | 199 | 1.44 (1.02-2.01) | 0.036 |  | 1.67 (1.14-2.46) | | 0.009 |  |
| Others | 103 | 205 | 1.40 (1.061.84) | 0.018 |  | 1.52 (1.13-2.04) | | 0.006 |  |
| Tumor stage |  |  |  |  | 0.710 |  | |  | 0.164 |
| I-IIIA | 195 | 440 | 1.71 (1.32-2.22) | <0.0001 |  | 2.10 (1.54-2.86) | | <0.0001 |  |
| IIIB-IV | 155 | 357 | 1.13 (0.92-1.39) | 0.259 |  | 1.18 (0.95-1.46) | | 0.393 |  |
| Chemotherapy |  |  |  |  | 0.381 |  | |  | 0.690 |
| No | 196 | 427 | 1.17 (0.92-1.48) | 0.195 |  | 1.28 (0.98-1.67) | | 0.075 |  |
| Yes | 154 | 370 | 1.36 (1.09-1.70) | 0.007 |  | 1.43 (1.14-1.80) | | 0.002 |  |
| Radiotherapy |  |  |  |  | 0.870 |  | |  | 0.617 |
| No | 241 | 502 | 1.37 (1.18-1.69) | 0.003 |  | 1.52 (1.21-1.91) | | 0.0003 |  |
| Yes | 109 | 295 | 1.31 (1.01-1.69) | 0.040 |  | 1.38 (1.06-1.81) | | 0.019 |  |
| Surgery |  |  |  |  | 0.730 |  | |  | 0.202 |
| No | 187 | 433 | 1.33 (1.10-1.62) | 0.003 |  | 1.39 (1.14-1.70) | | 0.001 |  |
| Yes | 163 | 364 | 1.33 (0.98-1.79) | 0.067 |  | 1.67 (1.16-2.40) | | 0.006 |  |
| Abbreviations: OS, overall survival; DSS, disease-specific survival; NSCLC, non-small cell lung cancer; PLCO, the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial; HR, hazards ratio; CI, confidence interval.  a 38 missing data were excluded, unfavorable genotypes were *INPP5D* rs13385922 CT+TT and *EXOSC3* rs3208406 AG+GG;  b Adjusted for age, sex, stage, histology, smoking status, chemotherapy, radiotherapy, surgery, PC1, PC2, PC3 and PC4;  c *P* inter: *P* value for interaction analysis between characteristics and unfavorable genotypes. | | | | | | | | | |

**Supplemental Acknowledgements**

We wish to thank all of the investigators and funding agencies that enabled the deposition of data in dbGaP and PLCO that we used in the present study:

The datasets used for the analyses described in thr present study were obtained from dbGaP at http://www.ncbi.nlm.nih.gov/gap through dbGaP accession number phs000336.v1.p1 and phs000093.v2.p2. Principal Investigators: Maria Teresa Landi. Genetic Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA. Neil E. Caporaso. Genetic Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA.

Funding support for the GWAS of Lung Cancer and Smoking was provided through the NIH Genes, Environment and Health Initiative [GEI] (Z01 CP 010200). The human subjects participating in the GWAS derive from The Environment and Genetics in Lung Cancer Etiology (EAGLE) case-control study and the Prostate, Lung Colon and Ovary Screening Trial and these studies are supported by intramural resources of the National Cancer Institute. Assistance with phenotype harmonization and genotype cleaning, as well as with general study coordination, was provided by the Gene Environment Association Studies, GENEVA Coordinating Center (U01HG004446). Assistance with data cleaning was provided by the National Center for Biotechnology Information. Funding support for genotyping, which was performed at the Johns Hopkins University Center for Inherited Disease Research, was provided by the NIH GEI (U01HG004438).

PLCO was also supported by the Intramural Research Program of the Division of Cancer Epidemiology and Genetics and by contracts from the Division of Cancer Prevention, National Cancer Institute, NIH, DHHS. The authors thank PLCO screening center investigators and staff, and the staff of Information Management Services Inc. and Westat Inc. Most importantly, we acknowledge trial participants for their contributions that made this study possible.