|  |  |  |  |
| --- | --- | --- | --- |
| **Supporting Information Table 1.** List of 111 selected genes in the complement-related gene-set | | | |
| **Dataset** | **Name of pathwaya** | **Selected genesb** | **Number of genes** |
| Genecards | Complement Pathway | *C1QA, C1QB, C1QBP, C1QC, C1R, C1S, C2, C3, C4A, C4B, C5, C5AR1, C6, C7, C8A, C8B, C8G, C9, CD46, CD55, CD59, CFB, CFD, CFH, CR1, CR2, MASP1, MASP2, MBL2, SERPING1* | 30 |
| Genecards | Complement and coagulation cascades | *A2M, BDKRB1 , BDKRB2 , C3AR1, C4BPA, C4BPB, CFI, CLTC, CLU, CPB2, F10, F11, F12, F13A1, F13B, F2, F2R, F2RL2, F2RL3, F3, F5, F7, FGA, FGB, FGG, ITGAM, ITGAX, ITGB2, KLKB1, KNG1, LMAN1, MIR6843, PLAT, PLAU, PLAUR, PLG, PROC, PROCR, PROS1, SERPINA1, SERPINA5, SERPINB2, SERPINC1, SERPIND1, SERPINE1, SERPINF2, TFPI, THBD, VTN, VWF* | 50 |
| Genecards | Immune response Lectin induced complement pathway | *C4B\_2, CD93, CFHR3, CRP, FCN2, FCN3, IGH, IGHG1, IGHM, IGHV3-23, IGK, IGKC, IGL* | 13 |
| Genecards | Creation of C4 and C2 activators | *C5AR2, COLEC10, COLEC11, FCN1, GZMM, IGHG2, IGHG3, IGHG4, IGKV3D-11, IGKV4-1, IGKV5-2, IGLC1, IGLC6* | 13 |
| HGNC | Complement system | *CFHR1, CFHR2, CFHR4, CFHR5, CR1L* | 5 |
| Total |  |  | 111 |
| a Genes were selected based on online datasets and literatures;  b Duplicated genes and genes in X chromosome had been removed  Keyword: complement  Organism: Homo sapiens | | | |

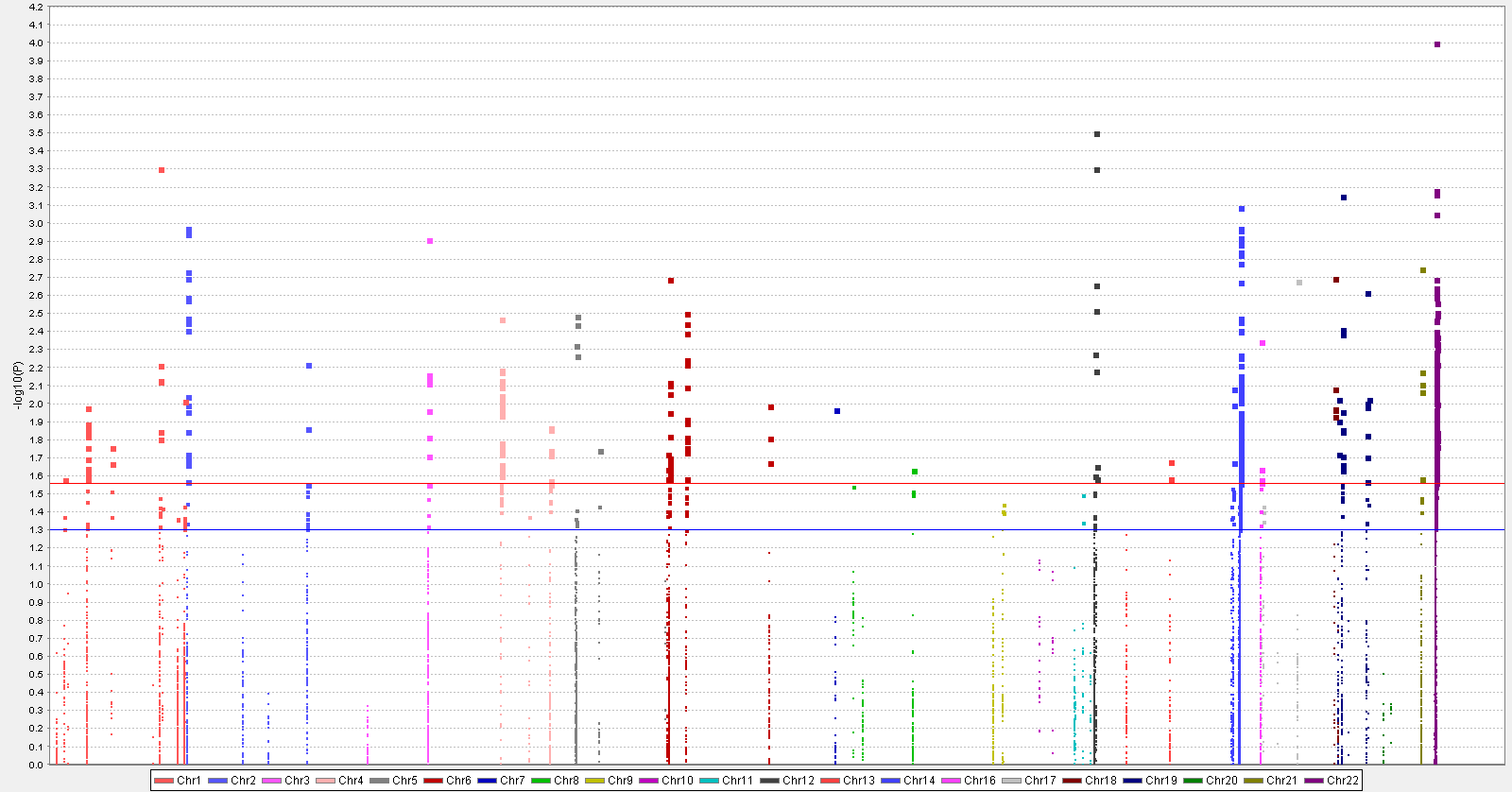
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Supporting Information Table 2.** Associations of the first 10 principal components and OS of NSCLC in the PLCO trial | | | | |
| **PC\*** | **Parameter Estimate** | **Standard Error** | **Chi-Square** | ***P*** |
| PC1 | 4.821 | 1.353 | 12.697 | <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 4 were used for the 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. | | | | |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Supporting Information Table 3.** Function prediction of the four validated SNPs associated with survival of NSCLC\* | | | | | | | | | | |
| **SNP** | **Gene** | **Chr** | **Genotyped** | **RegDB1** | **Haploreg v4.12** | | | | | |
| **Promoter histone marks** | **Enhancer histone marks** | **DNAse** | **Motifs changed** | **Selected eQTL hits** | **dbSNP func annot** |
| rs73049469 | *VWF* | 12 | no | 2b | ESC, IPSC | 10 tissues | ESC,IPSC | 5 altered motifs | -- | intronic |
| rs3788142 | *ITGB2* | 21 | yes | 3a | 12 tissues | 12 tissues | HRT,GI,BLD | AhR,Egr-1,Sp4 | 3hits | intronic |
| rs116750895 | *CFI* | 4 | no | -- | -- | -- | -- | HNF1,Homez,Ncx | -- | intronic |
| rs4251906 | *PLAUR* | 19 | no | -- | -- | BRST | BRST | -- | 1hit | intronic |
| 1 RegulomeDB: http://regulomedb.org/  2 Haploreg:http://archive.broadinstitute.org/mammals/haploreg/haploreg.php  \* All the four SNPs had no function in the SNPinfo: https://snpinfo.niehs.nih.gov/snpinfo/snpfunc.html  Abbreviations: SNP, single nucleotide polymorphism; NSCLC, non-small cell lung cancer; chr, chromosome; dbSNP fun cannot, dbSNP function annotation. | | | | | | | | | | |

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Supporting Information Table 4.** Stratified analysis for association between the risk genotypes and OS of NSCLC in the PLCO triala | | | | | | | | | | | |
| **Characteristics** | **0 risk Genotype** | |  | **1-2 risk genotypes** | |  |  | **Multivariate analysisb** | |  | ***P*interc** |
| **All** | **Death (%)** |  | **All** | **Death (%)** |  |  | **HR (95% CI)** | ***P*** |
| Age (years) |  |  |  |  |  |  |  |  |  |  |  |
| ≤ 71 | 293 | 176 (60.07) |  | 338 | 222 (65.68) |  |  | 1.37 (1.12-1.68) | 0.002 |  |  |
| > 71 | 248 | 175 (70.56) |  | 292 | 214 (73.29) |  |  | 1.15 (0.94-1.41) | 0.173 |  | 0.080 |
| Sex |  |  |  |  |  |  |  |  |  |  |  |
| Male | 333 | 239 (71.77) |  | 361 | 266 (73.68) |  |  | 1.26 (1.05-1.51) | 0.012 |  |  |
| Female | 208 | 112 (53.85) |  | 269 | 170 (63.20) |  |  | 1.17 (0.92-1.49) | 0.214 |  | 0.973 |
| Smoking status | |  |  |  |  |  |  |  |  |  |  |
| Never | 53 | 23 (43.40) |  | 61 | 39 (63.93) |  |  | 2.05 (1.16-3.61) | 0.013 |  |  |
| Current | 190 | 123 (64.74) |  | 227 | 143 (63.00) |  |  | 0.98 (0.76-1.26) | 0.860 |  |  |
| Former | 298 | 205 (68.79) |  | 342 | 254 (74.27) |  |  | 1.32 (1.10-1.60) | 0.004 |  | 0.860 |
| Histology |  |  |  |  |  |  |  |  |  |  |  |
| Adenocarcinoma | 270 | 149 (55.19) |  | 304 | 197 (64.80) |  |  | 1.23 (0.99-1.54) | 0.067 |  |  |
| Squamous cell carcinoma | 117 | 80 (68.38) |  | 167 | 111 (66.47) |  |  | 1.16 (0.85-1.58) | 0.360 |  |  |
| Others | 154 | 122 (79.22) |  | 159 | 128 (80.50) |  |  | 1.12 (0.85-1.46) | 0.429 |  | 0.840 |
| Tumor stage |  |  |  |  |  |  |  |  |  |  |  |
| I-IIIA | 307 | 137 (44.63) |  | 344 | 176 (51.16) |  |  | 1.40 (1.12-1.76) | 0.004 |  |  |
| IIIB-IV | 234 | 214 (91.45) |  | 286 | 260 (90.91) |  |  | 1.10 (0.92-1.32) | 0.306 |  | 0.188 |
| Chemotherapy |  |  |  |  |  |  |  |  |  |  |
| No | 301 | 162 (53.82) |  | 335 | 203 (60.60) |  |  | 1.29 (1.05-1.60) | 0.017 |  |  |
| Yes | 240 | 189 (78.75) |  | 295 | 233 (78.98) |  |  | 1.20 (0.98-1.46) | 0.076 |  | 0.406 |
| Radiotherapy |  |  |  |  |  |  |  |  |  |  |  |
| No | 355 | 197 (55.49) |  | 403 | 251 (62.28) |  |  | 1.27 (1.05-1.54) | 0.014 |  |  |
| Yes | 186 | 154 (82.80) |  | 227 | 185 (81.50) |  |  | 1.08 (0.86-1.35) | 0.509 |  | 0.593 |
| Surgery |  |  |  |  |  |  |  |  |  |  |  |
| No | 295 | 262 (88.81) |  | 339 | 302 (89.09) |  |  | 1.13 (0.95-1.34) | 0.157 |  |  |
| Yes | 246 | 89 (36.18) |  | 291 | 134 (46.05) |  |  | 1.41 (1.07-1.86) | 0.014 |  | 0.165 |
| a 14 missing date were excluded;  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 characteristic and risk genotypes.  Abbreviations: OS, overall survival; NSCLC, non-small cell lung cancer; PLCO, the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial; HR, hazards ratio; CI, confidence interval. | | | | | | | | | | | |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Supporting Information Table 5.** Correlation of the two functional SNPs with mRNA expression in whole blood and normal lung tissue in the GTEx project. Correlation between rs73049469, rs3788142, and *VWF, ITGB2* mRNA expression, respectively, in the GTEx project. | | | | | | | | |
| **SNP** | **Gene** | **Whole blood** | | |  | **Normal lung tissue** | | |
| **β** | **SE** | ***P*** |  | **β** | **SE** | ***P*** |
| rs73049469 | *VWF* | 0.04 | 0.07 | 0.558 |  | 0.08 | 0.06 | 0.195 |
| rs3788142 | *ITGB2* | 0.08 | 0.03 | 0.007 |  | 0.01 | 0.03 | 0.828 |
| Abbreviations: SE: standard error | | | | | | | | |

**Supporting Information Figure 1.** Manhattan plot of the discovery set in genotype data of the PLCO trial. The statistical values across the autosomes of associations between 14,699 SNPs and overall survival are plotted as −log10 *P* values. The blue horizontal line indicates *P* = 0.05 and the red line indicates FPRP = 0.2.



***PLAUR***

**rs4251906**

***ITGB2***

**rs3788142**

***VWF***

**rs73049469**

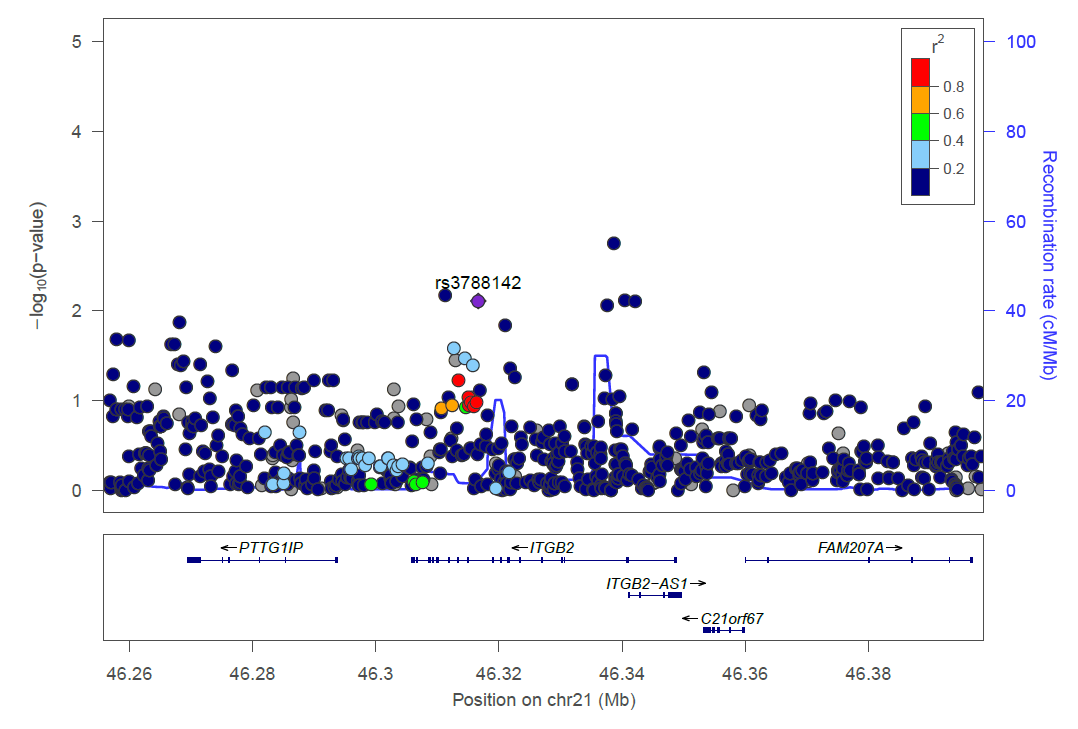
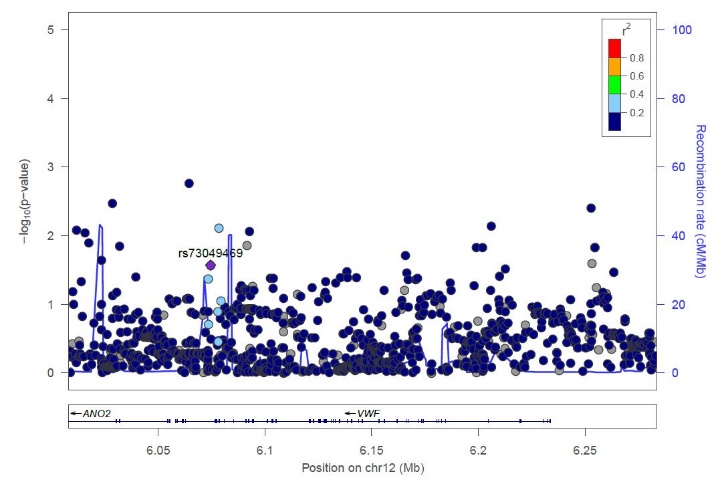
***CFI***

**rs116750895**

**FPRP=0.2**

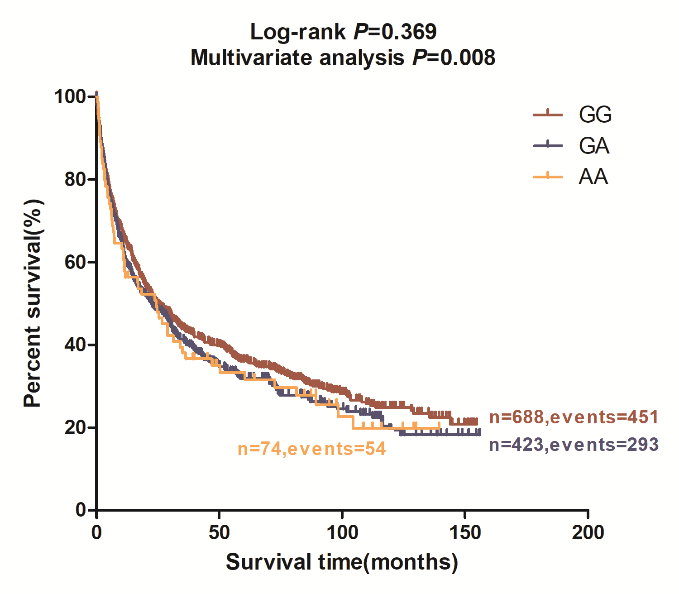
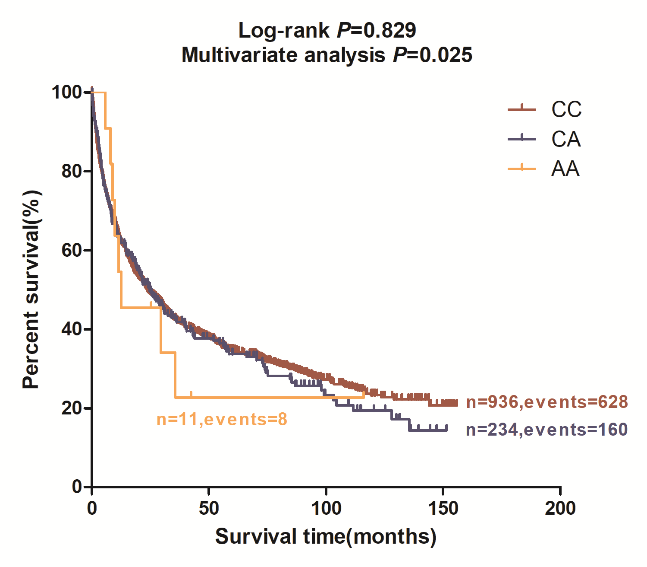
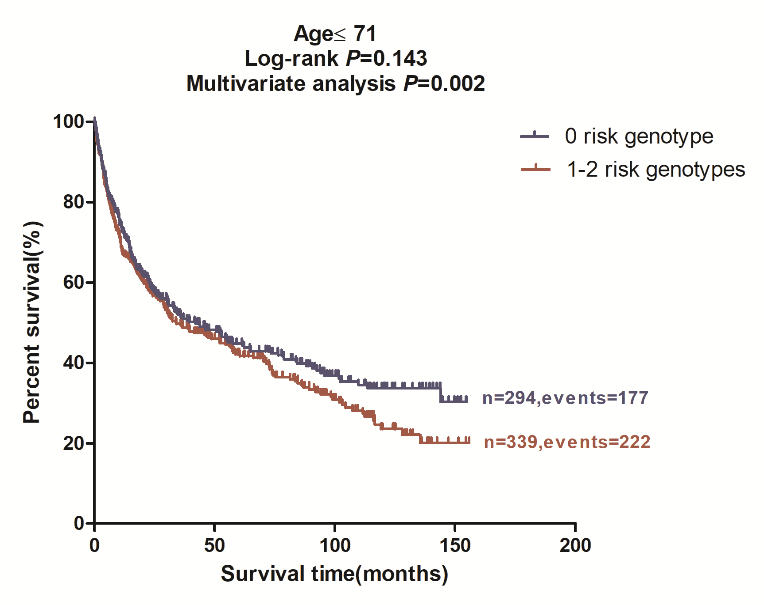
***P*=0.05**

**Supporting Information Figure 2.** Regional association plots for the four independent SNPs in the complement pathway genes. Single nucleotide polymorphisms (SNPs) in the region of 50 kilobases up or downstream of (a) *VWF* rs73049469 and (b) *ITGB2* rs3788142. Data points are colored according to the level of linkage disequilibrium of each pair of SNPs based on the hg19/1000 Genomes European population. The left-hand y-axis shows the association P-value of individual SNPs in the discovery dataset, which is plotted as -log10 (*P*) against chromosomal base-pair position. The right-hand y-axis shows the recombination rate estimated from HapMap Data Rel 22/phase II European population.

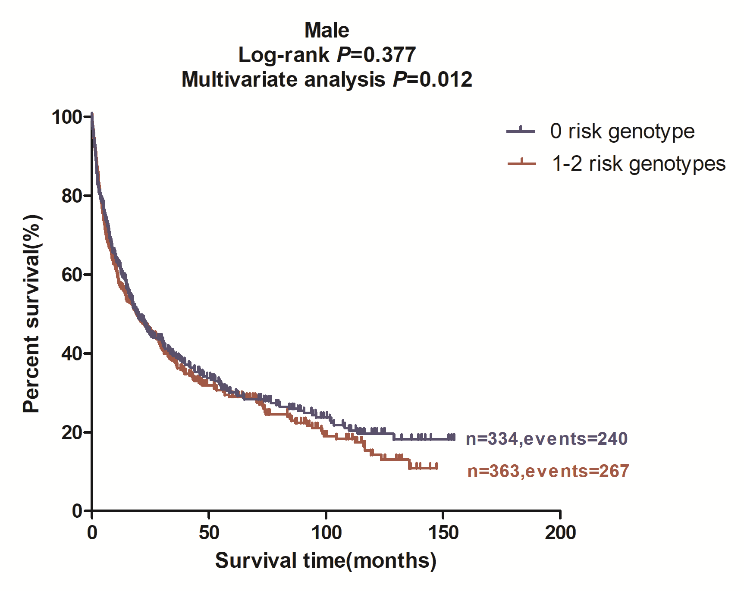
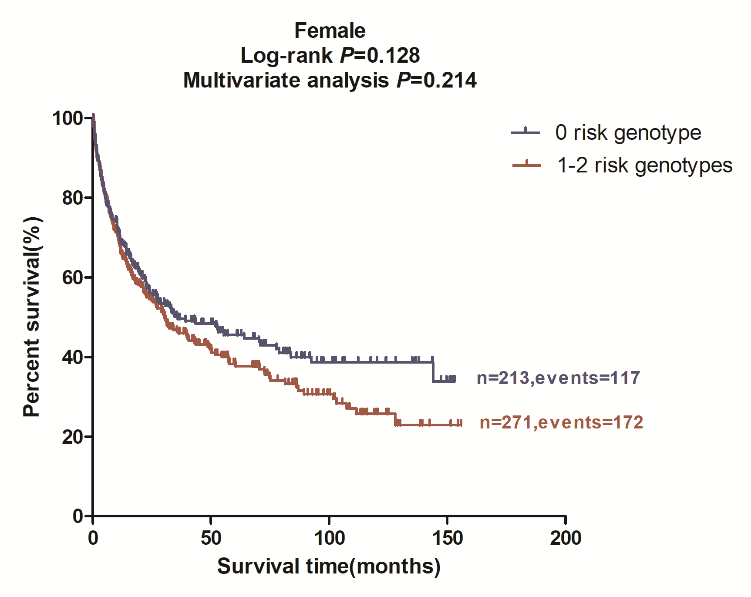
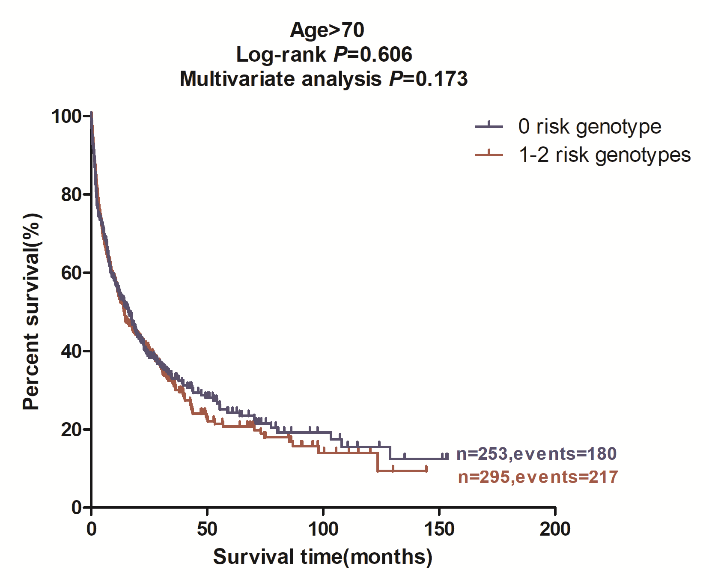
****a b

*P*=0.05

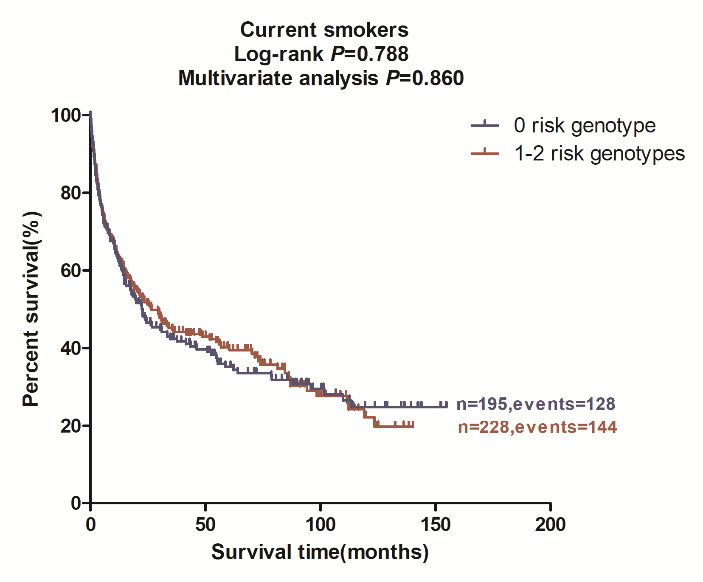
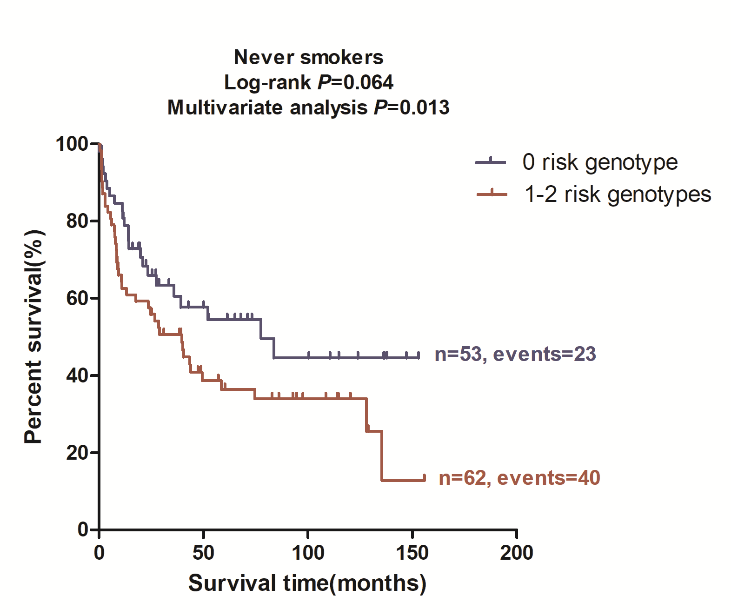
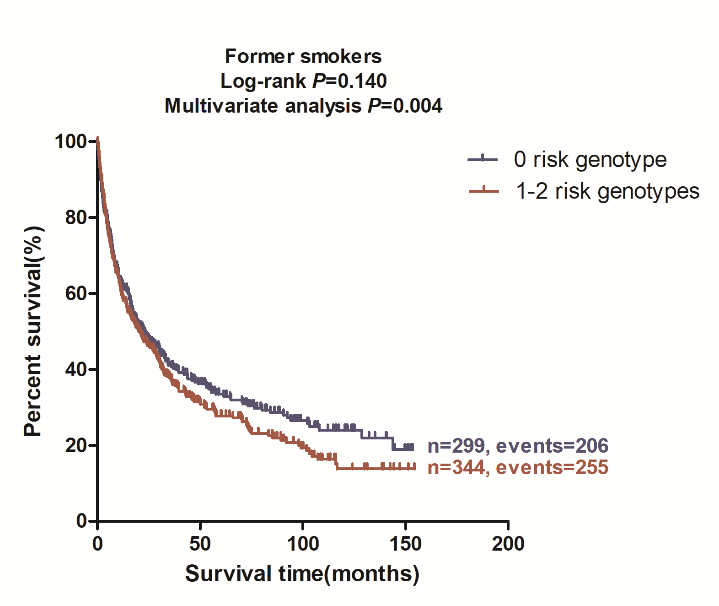
**Supporting Information Figure 3.** Kaplan–Meier analysis for patients with NSCLC of single SNP(a,b), and 0 and 1-2 risk genotypes stratified by clinicopathological characters. (c1, c2) By age; (d1, d2) by sex; (e1, e2 and e3) by smoking status; (f1, f2 and f3) by histology; (g1,g2) by tumor stage; (h1, h2) by chemotherapy; (i1, i2) by radiotherapy and (j1, j2) by surgery.

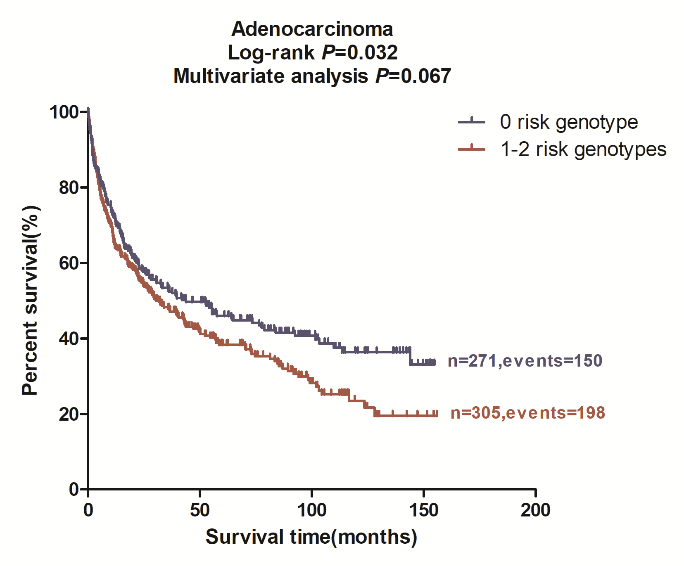
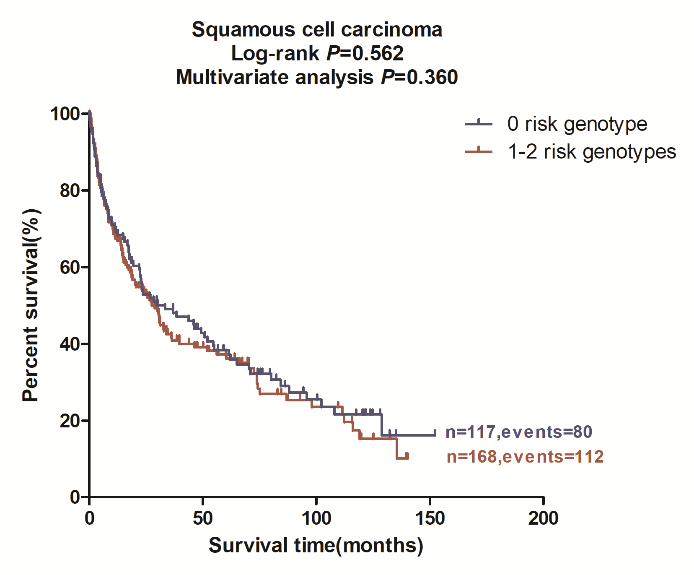
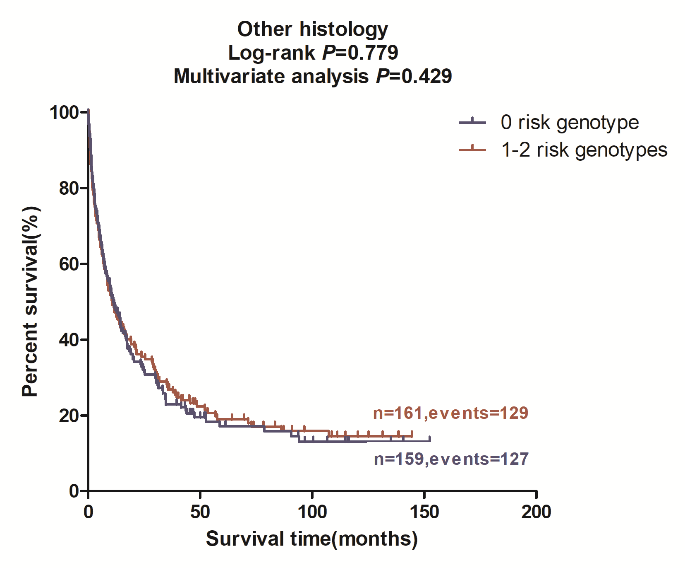


a b c1

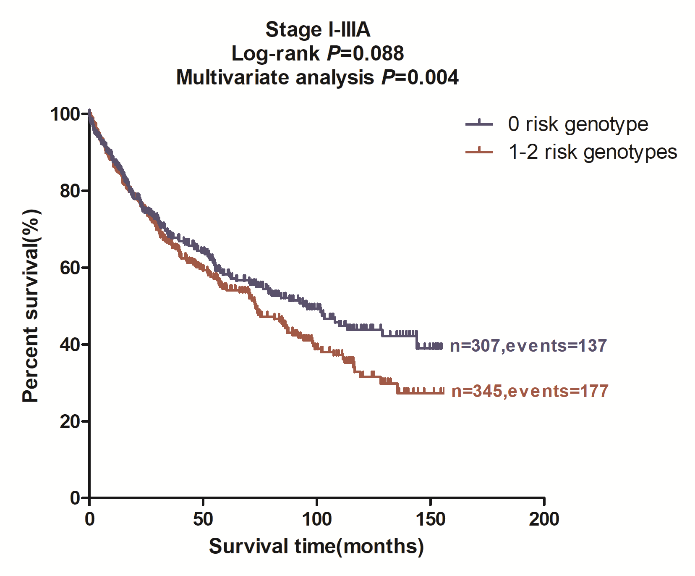
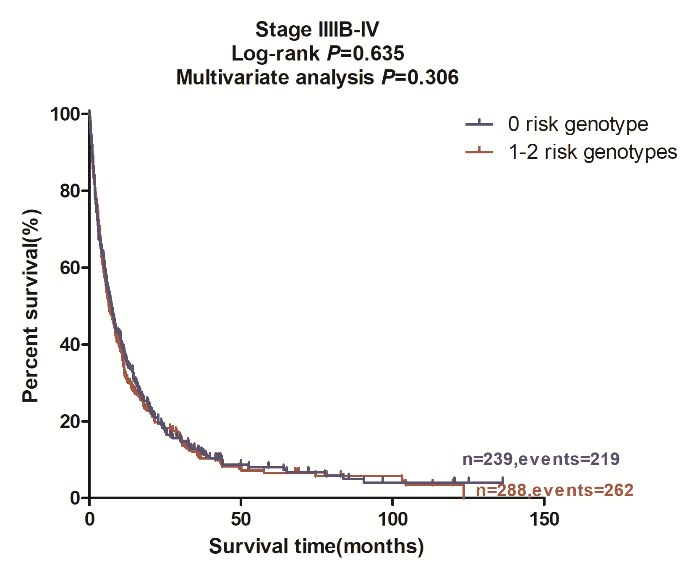
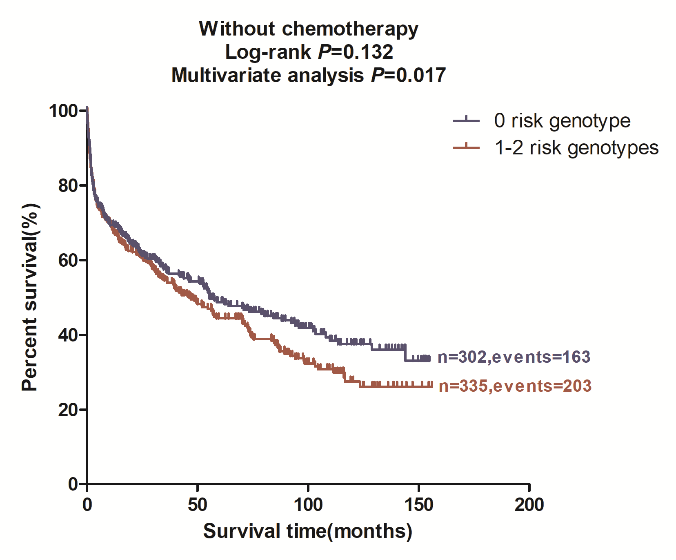


c2 d1 d2

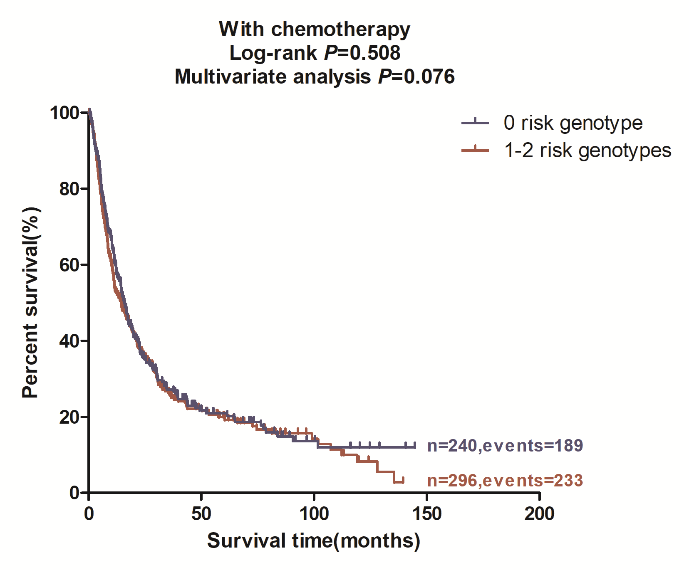
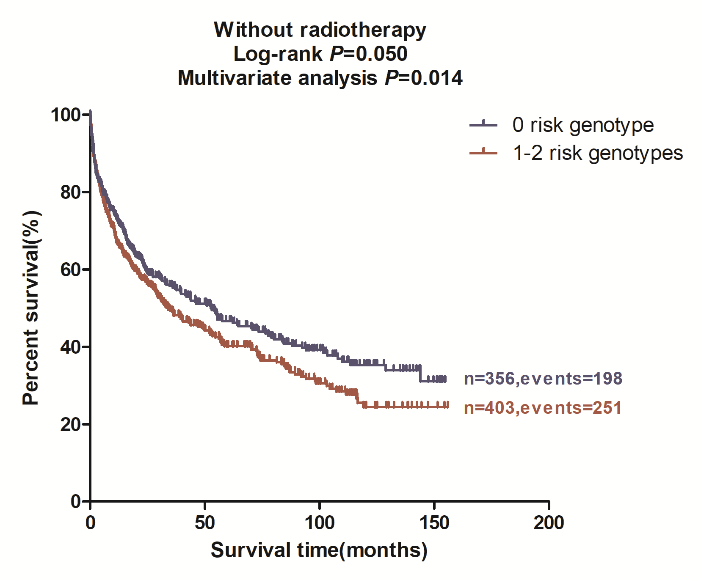
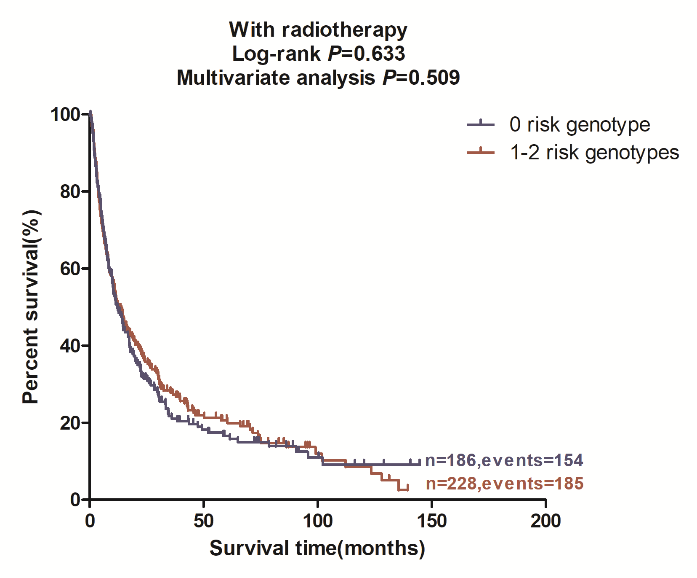
e1 e2 e3

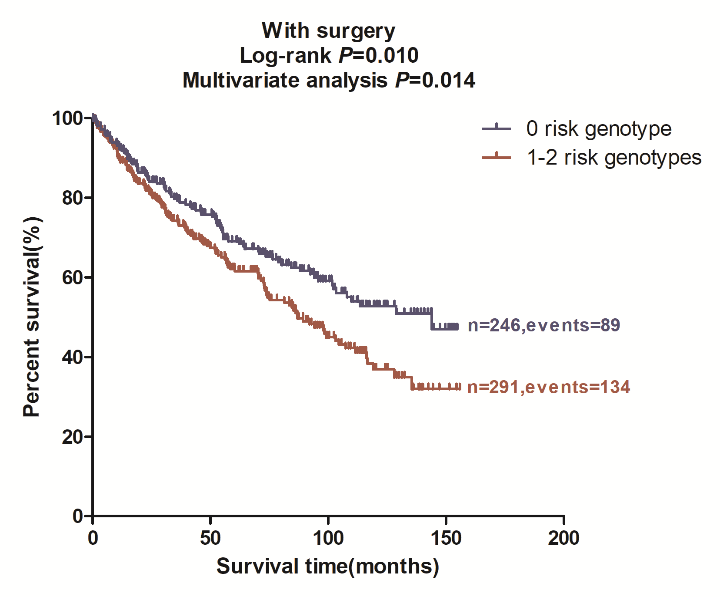


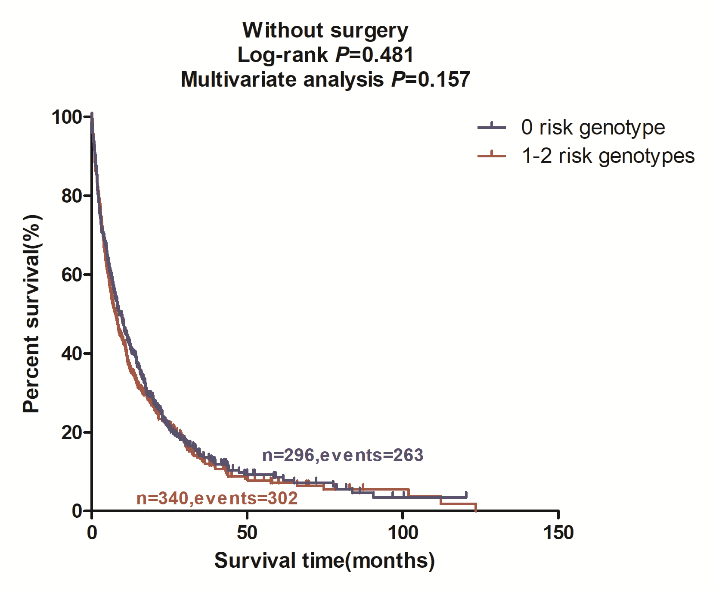
f1 f2 f3



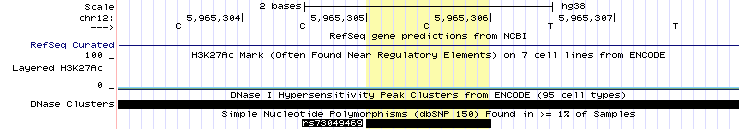
g1 g2 h1

h2 i1 i2

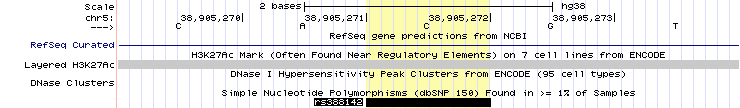


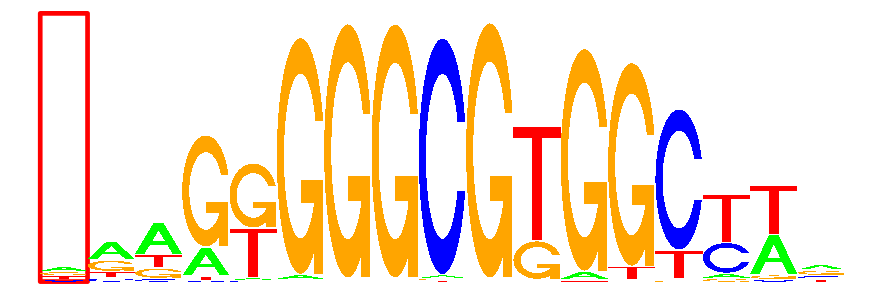
j1 j2

**Supporting Information Figure 4.** Functional prediction of SNPs in the ENCODE project. (a) Location and functional prediction of SNPs rs73049469, (b) Position weight matrix based on the consensus logo, which showed rs73049469 is located on the PLAG1 motif, (c) Location and functional prediction of SNPs rs3788142, (d) Position weight matrix based on the consensus logo, which shows rs3788142 is located on the SP4 motif.

a

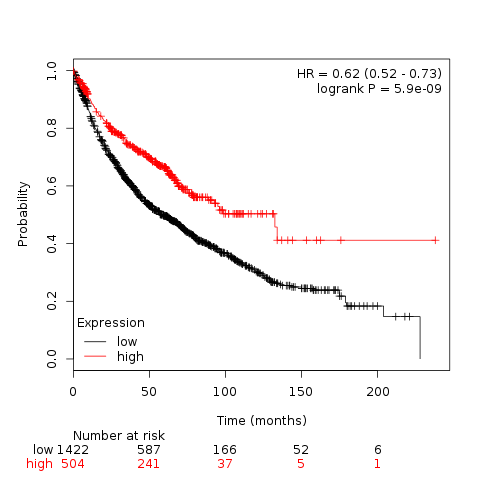
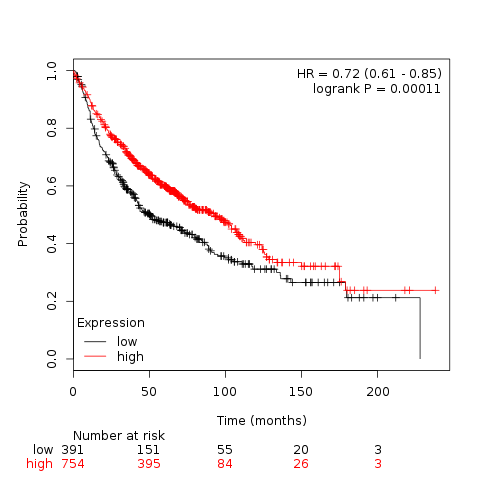
b

c

d

**Supporting Information Figure 5.** Differential mRNA expression analysis by using the data of the Cancer Genome Atlas (TCGA). Lower expression of (a) *VWF* and(b) *ITGB2* were found in the 109 paired tumor tissues compared to the adjacent normal tissues.

E:\ITGB2.tifE:\VWF.tifa b

** Supporting Information Figure 6.** Kaplan–Meier analysis for patients with NSCLC by the four genes. Based on online survival analysis software, (a) higher *VWF* expression and (b) higher *ITGB2* expression was associated with better survival of NSCLC.

a b