

Supplementary Table 1. Association of demographics and clinical characteristics with OS in NSCLC patients from the PLCO trial

Parameter		No.		Median survival time ^a	Univariate analysis		Multivariate analysis ^b	
		Patients	Deaths (%)		HR (95% CI)	P	HR (95% CI)	P
Total		1185	798 (60.3)	23.77				
Age ^c	≤71	636	400 (62.9)	36.53	1.00		1.00	
	>71	549	398 (72.5)	14.60	1.74 (1.51-2.00)	<0.001	1.26 (1.09-1.46)	0.002
Sex	Male	698	507 (72.6)	19.33	1.00		1.00	
	Female	487	291 (59.8)	31.77	0.73 (0.63-0.84)	<0.001	0.80 (0.69-0.94)	0.005
Smoking status	Never	115	63 (54.8)	43.50	1.00		1.00	
	Current	423	272 (64.3)	23.70	1.34 (1.02-1.76)	0.038	1.79 (1.32-2.44)	<0.001
	Former	647	463 (71.6)	21.30	1.51 (1.16-1.96)	0.002	1.69 (1.26-2.27)	<0.001
Pack years ^c	≤49	603	399 (66.2)	23.77	1.00		1.00	
	>49	581	398 (68.5)	23.97	1.05 (0.92-1.21)	0.458	1.01 (0.87-1.17)	0.936
	Missing	1						
Histology	Adenocarcinoma	577	348 (60.3)	36.37	1.00		1.00	
	Squamous cell	285	192 (67.4)	29.83	1.19 (1.00-1.42)	0.057	1.14 (0.94-1.37)	0.180
	Others	323	258 (79.9)	10.87	1.89 (1.60-2.22)	<0.001	1.33 (1.12-1.57)	0.003
Stage	I-III A	655	315 (48.1)	80.13	1.00		1.00	
	IIIB-IV	528	482 (91.3)	6.87	5.08 (4.37-5.91)	<0.001	2.70 (2.23-3.27)	<0.001
	Missing	2						
Chemotherapy	No	639	367 (57.4)	53.03	1.00		1.00	
	Yes	538	423 (78.6)	14.83	1.85 (1.60-2.14)	<0.001	0.42 (0.35-0.50)	<0.001
	Missing	8						
Radiotherapy	No	762	450 (59.1)	40.50	1.00		1.00	
	Yes	415	340 (81.9)	12.33	1.94 (1.68-2.24)	<0.001	0.90 (0.77-1.05)	0.160
	Missing	8						
Surgery	No	637	566 (88.9)	8.27	1.00		1.00	
	Yes	540	224 (41.5)	103.33	0.15 (0.13-0.18)	<0.001	0.26 (0.21-0.33)	<0.001
	Missing	8						

OS = Overall survival; NSCLC = Non-small cell lung cancer; PLCO = Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial; HR = Hazard ratio; 95% CI = 95% Confidence interval.

^a Median survival time (months);

^b Multivariate cox regression analyses were adjusted for all factors listed in this table;

^c The median age is 71 years; the median pack years is 49.

Supplementary Table 2. List of 84 selected genes in the c-Myb pathway from MSigDB

Name	Selected genes	Number of genes
PID_CMYB_PATHWAY	<i>ADA, ADORA2B, ANPEP, ATP2B1, BCL2, BIRC3, CA1, CASP6, CBX4, CCNA1, CCNB1, CCND1, CD34, CD4, CDK6, CDKN1A, CDKN1B, CDKN2A, CEBPA, CEBPB, CEBPD, CLTA, COL1A2, COPA, CREBBP, CSF1R, ELANE, EP300, ETS1, ETS2, GATA3, GSTM1, H2AFZ, HES1, HIPK2, HRAS, HSPA8, IQGAP1, KIT, KITLG, KRAS, LECT2, LEF1, LYZ, MAD1L1, MAF, MAP3K7, MAT2A, MCM4, MPO, MYB, MYC, MYF6, MYOD1, NCOR1, NLK, NRAS, PAX5, PIAS3, PIM1, PPID, PPP3CA, PRTN3, PTCRA, PTGS2, RAG2, SIN3A, SKI, SLC25A3, SMARCA2, SND1, SP1, SPI1, TAB1, TAB2, TFEC, TOM1, TRIM28, UBE2I, WNT1, YEATS4, ZFHX3, ZFPM1, GATA1</i>	84
Gene removed ^a	<i>GATA1</i>	1
Total genes		83

MSigDB = Molecular signatures database.

^aGene in X chromosome was removed;

Keyword for MSigDB: c-Myb

Organism: Homo sapiens

Supplementary Table 3. The selected 42 representative SNPs with $P < 0.05$ and FPRP ≤ 0.10 in the PLCO discovery dataset and HLCS GWAS validation dataset

SNP	Allele ^a	Gene	PLCO (n=1185)		HLCS (n=984)	
			HR (95% CI)	P^b	HR (95% CI)	P^b
rs10480889	T/C	<i>HIPK2</i>	0.78 (0.65-0.94)	0.009	0.94 (0.78-1.13)	0.532
rs10893879	G/C	<i>ETS1</i>	0.84 (0.75-0.94)	0.002	0.99 (0.88-1.11)	0.876
rs11429644	T/-	<i>ETS1</i>	1.20 (1.09-1.33)	3.7×10^{-4}	N/A	N/A
rs11606640	G/A	<i>ETS1</i>	0.82 (0.71-0.95)	0.006	0.73 (0.62-0.85)	7.2×10^{-5}
rs11641076	T/A	<i>ZFPM1</i>	1.26 (1.08-1.48)	0.004	1.05 (0.85-1.29)	0.674
rs11767177	G/T	<i>MAD1L1</i>	1.33 (1.12-1.59)	0.001	1.10 (0.92-1.32)	0.284
rs11773884	A/G	<i>CDK6</i>	0.85 (0.75-0.95)	0.004	1.08 (0.96-1.21)	0.199
rs1353357	T/G	<i>MYF6</i>	1.27 (1.07-1.51)	0.007	0.95 (0.79-1.14)	0.585
rs138754	C/T	<i>TOM1</i>	1.23 (1.06-1.42)	0.006	1.15 (0.96-1.37)	0.120
rs1474637	C/T	<i>MAP3K7</i>	0.87 (0.79-0.96)	0.006	1.03 (0.93-1.15)	0.567
rs1506799	G/A	<i>PPP3CA</i>	1.26 (1.06-1.48)	0.008	0.96 (0.79-1.16)	0.650
rs16824948	C/T	<i>SKI</i>	1.30 (1.06-1.58)	0.010	1.09 (0.85-1.38)	0.500
rs180989349	A/G	<i>IQGAP1</i>	1.21 (1.05-1.40)	0.009	1.10 (0.91-1.32)	0.325
rs200673885	T/-	<i>IQGAP1</i>	1.22 (1.05-1.42)	0.010	N/A	N/A
rs2079147	A/G	<i>CDK6</i>	0.85 (0.77-0.94)	0.002	1.06 (0.94-1.19)	0.318
rs2097795	C/T	<i>CREBBP</i>	0.74 (0.60-0.92)	0.008	0.82 (0.64-1.05)	0.113
rs31527	G/A	<i>LECT2</i>	1.21 (1.06-1.38)	0.004	1.02 (0.88-1.19)	0.785
rs34761089	T/-	<i>SP1</i>	1.27 (1.06-1.52)	0.011	N/A	N/A
rs34910888	T/-	<i>MYB</i>	0.77 (0.65-0.92)	0.004	N/A	N/A
rs3758166	C/T	<i>PAX5</i>	0.81 (0.68-0.95)	0.010	1.04 (0.88-1.23)	0.651
rs3765580	A/G	<i>SMARCA2</i>	1.18 (1.04-1.33)	0.012	0.94 (0.82-1.09)	0.423
rs3804361	G/A	<i>PPP3CA</i>	1.14 (1.03-1.25)	0.011	0.91 (0.82-1.02)	0.097
rs42523	T/C	<i>COL1A2</i>	1.17 (1.04-1.31)	0.007	1.09 (0.97-1.23)	0.144
rs4564353	A/T	<i>ETS1</i>	0.84 (0.76-0.93)	0.001	0.97 (0.86-1.08)	0.573
rs4935825	A/C	<i>HSPA8</i>	0.83 (0.75-0.92)	3.1×10^{-4}	0.95 (0.85-1.06)	0.321
rs5030520	C/G	<i>CASP6</i>	0.74 (0.60-0.90)	0.003	0.88 (0.69-1.13)	0.331
rs5030552	A/G	<i>CASP6</i>	1.25 (1.06-1.47)	0.009	1.02 (0.84-1.23)	0.876
rs56386066	TGG/-	<i>MYB</i>	0.86 (0.77-0.95)	0.004	N/A	N/A
rs61372618	-/T	<i>SMARCA2</i>	0.86 (0.78-0.95)	0.004	N/A	N/A
rs61761154	T/C	<i>KRAS</i>	1.26 (1.08-1.47)	0.003	1.02 (0.84-1.25)	0.812
rs61916272	C/G	<i>CD4</i>	1.15 (1.04-1.28)	0.007	0.95 (0.85-1.06)	0.387
rs62053220	C/G	<i>ZFHX3</i>	0.71 (0.58-0.86)	4.4×10^{-4}	0.73 (0.60-0.89)	0.002
rs6475529	C/T	<i>SMARCA2</i>	0.80 (0.69-0.93)	0.004	0.93 (0.78-1.10)	0.393
rs6590334	C/T	<i>ETS1</i>	0.82 (0.74-0.90)	7.0×10^{-5}	0.99 (0.89-1.11)	0.912
rs7024379	A/C	<i>SMARCA2</i>	1.29 (1.10-1.51)	0.001	0.95 (0.79-1.14)	0.567
rs7039339	C/T	<i>SMARCA2</i>	1.22 (1.06-1.39)	0.004	0.98 (0.85-1.13)	0.807
rs7204751	T/C	<i>ZFHX3</i>	1.20 (1.05-1.38)	0.009	1.04 (0.89-1.22)	0.614
rs72642130	C/T	<i>SKI</i>	1.38 (1.13-1.69)	0.002	1.09 (0.85-1.41)	0.487
rs73047102	G/A	<i>MAD1L1</i>	1.27 (1.06-1.52)	0.009	1.09 (0.86-1.39)	0.464
rs8062492	G/A	<i>ZFHX3</i>	0.73 (0.63-0.86)	1.2×10^{-4}	0.93 (0.79-1.08)	0.344
rs8181551	A/T	<i>ETS1</i>	1.15 (1.04-1.28)	0.008	1.01 (0.90-1.14)	0.844

rs9298820	A/G	SMARCA2	0.87 (0.78-0.96)	0.007	1.06 (0.95-1.18)	0.325
-----------	-----	---------	------------------	-------	------------------	-------

FPRP = False positive report probability; SNPs = Single nucleotide polymorphisms; PLCO = Prostate, Lung, Colorectal and Ovarian cancer trial; HLCS = Harvard Lung Cancer Susceptibility study; GWAS = Genome-wide association study; HR = Hazard ratio; 95% CI = 95% Confidence interval; N/A = Not applicable.

^a Reference/effect allele;

^b Adjusted for age, sex, smoking status, histology, tumor stage, chemotherapy, radiotherapy, surgery, and principal components; in an additive model.

Supplementary Table 4. Associations of the two validated SNPs in the c-Myb pathway with DSS of NSCLC in the PLCO dataset

Genotype	DSS Univariate analysis				DSS Multivariate analysis ^a			
	Frequency				Frequency			
	All	Death(%)	HR (95% CI)	P	All	Death(%)	HR (95% CI)	P
ETS1 rs11606640 G > A								
	n=1177				n=1167			
GG	842	511 (60.7)	1.00		835	506 (60.6)	1.00	
GA	307	180 (58.6)	0.93 (0.79-1.10)	0.415	304	179 (58.9)	0.79 (0.66-0.93)	0.006
AA	28	19 (67.9)	1.08 (0.68-1.70)	0.752	28	19 (67.9)	0.86 (0.54-1.37)	0.528
Trend				0.629				0.013
GA+AA	335	199 (59.4)	0.94 (0.80-1.11)	0.490	332	198 (59.6)	0.79 (0.67-0.94)	0.006
Genotype missing ^b	8				8			
Phenotype missing ^d	N/A				10			
ZFHX3 rs62053220 C > G								
	n=1181				n=1171			
CC	1002	610 (60.9)	1.00		992	604 (60.9)	1.00	
CG	173	101 (58.4)	0.87 (0.71-1.08)	0.206	173	101 (58.4)	0.78 (0.63-0.97)	0.023
GG	6	2 (33.3)	0.41 (0.10-1.66)	0.212	6	2 (33.3)	0.21 (0.05-0.86)	0.030
Trend				0.100				0.002
CG+GG	179	103 (57.5)	0.86 (0.69-1.05)	0.140	179	103 (57.5)	0.74 (0.60-0.92)	0.006
Genotype missing ^c	4				4			
Phenotype missing ^d	N/A				10			
NPG ^a								
	n=1173				n=1163			
0	717	439 (61.2)	1.00		710	434 (61.1)	1.00	
1	402	238 (59.2)	0.90 (0.77-1.05)	0.176	399	237 (59.4)	0.71 (0.60-0.83)	<0.001
2	54	31 (57.4)	0.87 (0.60-1.25)	0.449	54	31 (57.4)	0.73 (0.50-1.05)	0.092
Trend				0.156				<0.001
0	717	439 (61.2)	1.00		710	434 (61.1)	1.00	
1-2	456	269 (59.0)	0.89 (0.77-1.04)	0.146	453	268 (59.2)	0.71 (0.61-0.83)	<0.001
Genotype missing ^c	12							
Phenotype missing ^d	N/A							

SNPs = Single nucleotide polymorphisms; DSS = Disease specific survival; NSCLC = Non-small cell lung cancer; PLCO = Prostate, Lung, Colorectal and Ovarian cancer trial; NPG = Number of protective genotypes; HR = Hazards ratio; 95% CI = 95% Confidence interval; N/A = Not applicable.

^a Protective genotypes were *ETS1* rs11606640 GA+AA, *ZFHX3* rs62053220 CG+GG;

^b Multivariate Cox hazards regression analyses were adjusted for age, sex, smoking, stage, histology, chemotherapy, radiotherapy, surgery, and top four principal components;

^c Eight observations missing of *ETS1* rs11606640; Four observations missing of *ZFHX3* rs62053220;

^d Two observations missing of tumor stage and eight observations missing of chemotherapy/radiotherapy/surgery in the PLCO dataset.

Supplement Table 5. Stratified multivariate analyses for association between protective genotypes and OS in NSCLC patients in the PLCO trial

Characteristics	No. of 0 protective genotypes		No. of 1-2 protective genotypes		Multivariate analysis ^a		<i>P</i> _{inter} ^b
	All	Death (%)	All	Death (%)	HR (95% CI)	<i>P</i>	
Age (years)							
≤ 71	387	243 (62.8)	238	151 (63.5)	0.82 (0.67-1.01)	0.060	
> 71	323	242 (75.0)	215	145 (67.4)	0.60 (0.48-0.74)	<0.001	0.002
Sex							
Male	413	310 (75.1)	277	190 (68.6)	0.67 (0.55-0.80)	<0.001	
Female	297	175 (58.9)	176	106 (60.2)	0.72 (0.56-0.93)	0.012	0.186
Smoking status							
Never	72	37 (51.4)	41	25 (61.0)	0.93 (0.54-1.60)	0.802	
Current	253	169 (66.8)	160	94 (58.8)	0.70 (0.54-0.90)	0.006	
Former	385	279 (72.5)	252	177 (70.2)	0.69 (0.57-0.84)	<0.001	0.480
Histology							
Adenocarcinoma	357	220 (61.6)	211	123 (58.3)	0.72 (0.58-0.91)	0.005	
Squamous	166	113 (68.1)	115	75 (65.2)	0.89 (0.65-1.21)	0.449	
Others	187	152 (81.3)	127	98 (77.2)	0.59 (0.46-0.77)	<0.001	0.144
Tumor stage							
I-III A	405	202 (49.9)	245	110 (44.9)	0.80 (0.63-1.01)	0.056	
IIIB-IV	305	283 (92.8)	208	186 (89.4)	0.71 (0.59-0.85)	<0.001	0.394
Chemotherapy							
No	382	224 (58.6)	250	139 (55.6)	0.69 (0.55-0.86)	0.001	
Yes	328	261 (79.6)	203	157 (77.3)	0.72 (0.59-0.88)	0.002	0.888
Radiotherapy							
No	453	272 (60.0)	300	173 (57.7)	0.79 (0.65-0.96)	0.016	
Yes	257	213 (82.9)	153	123 (80.4)	0.61 (0.48-0.77)	<0.001	0.040
Surgery							
No	383	346 (90.3)	246	213 (86.6)	0.66 (0.56-0.79)	<0.001	
Yes	327	139 (42.5)	207	83 (40.1)	0.84 (0.64-1.11)	0.221	0.067
Genotype missing ^c	12						
Phenotype missing ^d	10						

OS = Overall survival; NSCLC = Non-small cell lung cancer; HR = Hazards ratio; 95% CI = 95% Confidence interval.

^a Adjusted for age, sex, stage, histology, smoking status, chemotherapy, radiotherapy, surgery, PC1, PC2, PC3 and PC4;

^b *P*_{inter}: *P*-value for interaction analysis between characteristic and number of protective genotypes;

^c Eight observations missing of *ETS1* rs11606640; Four observations missing of *ZFHX3* rs62053220;

^d Two observations missing of tumor stage and eight observations missing of chemotherapy/radiotherapy/surgery in PLCO dataset.

Supplementary Table 6. Associations of the two validated SNPs in the c-Myb pathway with OS of NSCLC in subgroup of age

Genotype	OS Multivariate analysis ^a (age ≤ 71 years)				OS Multivariate analysis ^a (age > 71 years)			
	Frequency				Frequency			
	All	Death(%)	HR (95% CI)	P	All	Death(%)	HR (95% CI)	P
ETS1 rs11606640 G > A	n=629				n=538			
GG	456	281 (61.6)	1.00		379	281 (74.1)	1.00	
GA	160	105 (65.6)	0.92 (0.73-1.16)	0.474	144	95 (66.0)	0.65 (0.51-0.83)	<0.001
AA	13	10 (76.9)	1.06 (0.56-2.01)	0.854	15	11 (73.3)	0.73 (0.39-1.35)	0.311
Trend				0.627				0.001
GA+AA	173	115 (66.5)	1.09 (0.88-1.36)	0.525	159	106 (66.7)	0.66 (0.52-0.83)	<0.001
Total Genotype missing ^b	8							
Total Phenotype missing ^c	10							
ZFHX3 rs62053220 C > G	n=630				n=541			
CC	537	343 (63.9)	1.00		455	333 (73.2)	1.00	
CG	89	53 (59.6)	0.81 (0.60-1.09)	0.163	84	56 (66.7)	0.71 (0.53-0.95)	0.020
GG	4	1 (25.0)	0.11 (0.02-0.80)	0.029	2	1 (50.0)	0.47 (0.07-3.37)	0.449
Trend				0.010				0.014
CG+GG	93	54 (58.1)	0.73 (0.55-0.99)	0.040	86	57 (66.3)	0.70 (0.52-0.93)	0.015
Total Genotype missing ^b	4							
Total Phenotype missing ^c	10							

SNPs = Single nucleotide polymorphisms; OS = Overall survival; NSCLC = Non-small cell lung cancer; PLCO = Prostate, Lung, Colorectal and Ovarian cancer trial; HR = Hazards ratio; 95% CI = 95% Confidence interval; N/A = Not applicable.

^a Adjusted for age, sex, smoking status, histology, tumor stage, chemotherapy, radiotherapy, surgery, and top four principal components;

^b Eight observations missing of *ETS1* rs11606640 removed in analyses;

^c Four observations missing of *ZFHX3* rs62053220 removed in analyses;

^d Two observations missing of tumor stage and eight observations missing of chemotherapy/radiotherapy/surgery in PLCO dataset removed in analyses.

Supplementary Table 7. Associations of the two validated SNPs in the c-Myb pathway with OS of NSCLC in subgroup of radiotherapy

Genotype	OS Multivariate analysis ^a (without radiotherapy)				OS Multivariate analysis ^a (with radiotherapy)			
	Frequency				Frequency			
	All	Death(%)	HR (95% CI)	P	All	Death(%)	HR (95% CI)	P
ETS1 rs11606640 G > A	n=757				n=410			
GG	531	312 (58.8)	1.00		304	250 (82.2)	1.00	
GA	208	123 (59.1)	0.86 (0.69-1.06)	0.163	96	77 (80.2)	0.66 (0.51-0.87)	0.002
AA	18	12 (66.7)	0.83 (0.46-1.48)	0.523	10	9 (90.0)	1.00 (0.51-1.97)	0.996
Trend				0.151				0.014
GA+AA	226	135 (59.7)	1.09 (0.88-1.36)	0.141	106	86 (81.1)	0.69 (0.53-0.89)	0.004
Total Genotype missing ^b	8							
Total Phenotype missing ^c	10							
ZFHX3 rs62053220 C > G	n=757				n=414			
CC	642	385 (60.0)	1.00		350	291 (83.1)	1.00	
CG	112	61 (54.5)	0.73 (0.55-0.96)	0.025	61	48 (78.7)	0.83 (0.61-1.14)	0.252
GG	3	1 (33.3)	0.48 (0.07-3.45)	0.464	3	1 (33.3)	0.13 (0.02-0.97)	0.046
Trend				0.019				0.024
CG+GG	115	62 (53.9)	0.72 (0.55-0.95)	0.020	64	49 (75.6)	0.75 (0.55-1.03)	0.072
Total Genotype missing ^b	4							
Total Phenotype missing ^c	10							

SNPs = Single nucleotide polymorphisms; OS = Overall survival; NSCLC = Non-small cell lung cancer; PLCO = Prostate, Lung, Colorectal and Ovarian cancer trial; HR = Hazards ratio; 95% CI = 95% Confidence interval; N/A = Not applicable.

^a Adjusted for age, sex, smoking status, histology, tumor stage, chemotherapy, radiotherapy, surgery, and top four principal components;

^b Eight observations missing of *ETS1* rs11606640 removed in analyses;

^c Four observations missing of *ZFHX3* rs62053220 removed in analyses;

^d Two observations missing of tumor stage and eight observations missing of chemotherapy/radiotherapy/surgery in PLCO dataset removed in analyses.

Supplementary Table 8. Functional prediction analyses of the two validated SNPs

SNP	Gene	Chr.	Allele	MAF	Regulome DB Score	HaploReg v4.1				
						Promoter histone marks	Enhancer histone marks	DNase	Motifs changed	dbSNP function annotation
rs11606640	<i>ETS1</i>	11	A/G	0.154	4	BLD ^a	13 tissues	BLD ^a , SKIN	Ets	intronic
rs62053220	<i>ZFHX3</i>	16	G/C	0.078	4	--	7 tissues	MUS ^b , GI ^c	5 altered motif	intronic

SNP = Single nucleotide polymorphisms; MAF = Minor allele frequency.

^a Primary T helper 17 cells PMA-I stimulated primary T CD8⁺ naive cells from peripheral blood;

^b Fetal muscle trunk;

^c Gastric.