

Genetic variants of *CHEK1*, *PRIM2* and *CDK6* in the mitotic phase-related pathway are associated with non-small cell lung cancer survival

List of Authors:

Rui Mu; Hongliang Liu; Sheng Luo; Edward F. Patz Jr.; Carolyn Glass; Li Su; Mulong Du; David C. Christiani; Lei Jin; Qingyi Wei

Contents		
Supplementary Table S1	Comparison of characteristics between the PLCO trial and the HLCS study	Page 1
Supplementary Table S2	List of 368 selected genes in mitotic phase-related gene-set used in discovery analysis	Page 2-3
Supplementary Table S3	Associations of the first 10 principal components and OS of NSCLC in the PLCO trial	Page 4
Supplementary Table S4	Three independent SNPs were adjusted for 41 previously published SNPs for NSCLC in the PLCO Trial	Page 5-6
Supplementary Table S5	Stratified analysis for associations between (0-1 vs. 2-3) unfavorable genotypes and NSCLC survival in the PLCO trial	Page 7
Supplementary Table S6	Stratified Cox proportional hazards model analysis for clinical covariates and three independent SNPs by the combination of stage and chemotherapy for NSCLC in the PLCO trial	Page 8
Supplementary Table S7	Function prediction for <i>CHEK1</i> rs76744140 and <i>PRIM2</i> rs6939623	Page 9
Supplementary Table S8	Function prediction for <i>CDK6</i> rs113181986 and its 24 high LD ($r^2 > 0.8$) SNPs in <i>CDK6</i>	Page 10-11
Supplementary Figure S1	Manhattan plot	Page 12
Supplementary Figure S2	Regional association plots for the three independent SNPs in the mitotic phase-related pathway genes	Page 13
Supplementary Figure S3	Receiver operating characteristic (ROC) curve and time dependent area under the ROC curve (AUC) estimation for NSCLC OS and DSS prediction using three newly independent SNPs in the PLCO trail	Page 14-15
Supplementary Figure S4	Testing proportional Hazards assumption by using Schoenfeld residuals plot and zph test	Page 16
Supplementary Figure S5	Functional prediction of three independent SNPs in mitotic phase-related pathway genes in the ENCODE data	Page 17
Supplementary Figure S6	Functional prediction of 10 high LD ($r^2 > 0.8$) SNPs of representative SNP rs113181986 in <i>CDK6</i> in the ENCODE data	Page 18
Supplementary Figure S7	The eQTL analysis for <i>CHEK1</i> rs76744140, <i>PRIM2</i> rs6939623, and <i>CDK6</i> rs113181986 from the GTEx (V8) database	Page 19
Supplementary Figure S8	The eQTL analysis for 19 SNPs had a high LD ($r^2 > 0.8$) with rs113181986 in <i>CDK6</i> from the GTEx (V8) database	Page 20
Supplementary Figure S9	The eQTL analysis for <i>CHEK1</i> rs76744140, <i>PRIM2</i> rs6939623, and <i>CDK6</i> rs113181986 from 1,000 Genomes Project	Page 21
Supplementary Figure S10	Kaplan-Meier analysis for patients with NSCLC by the three genes	Page 22
Supplementary Figure S11	Mutation frequency of <i>CHEK1</i> , <i>PRIM2</i> , and <i>CDK6</i> in non-small cell lung tumor tissues	Page 23

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)	
Tumor stage					
I-III A	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 S2. List of 368 selected genes in mitotic phase-related gene-set used in discovery analysis

Dataset	Name of pathway	Selected genes ^a	Number of genes
GO	GO_POSITIVE_REGULATION_OF_G1_S_TRANSITION_OF_MITOTIC_CELL_CYCLE	ADAM17, ADAMTS1, AIF1, AKT1, ANKRD17, ANXA1, APEX1, CCND1, CCND2, CCND3, CDC45, CDC6, CDK10, CUL4A, CUL4B, CYP1A1, DDX3X, EGFR, EIF4G1, FGF10, HYAL1, KCNA5, KMT2E, LSM10, LSM11, MDM2, MEPCE, MIR208A, MIR214, MIR221, MIR222, MIR29A, MIR372, MIR495, MIR515-1, MIR515-2, MIR519D, MIR520A, MIR520H, MIR590, MTBP, PLCB1, PLRG1, RDX, RGCC, RPTOR, TERT, TFDP1, UBE2E2	49
GO	GO_MITOTIC_G1_S_TRANSITION_CHECKPOINT	ARID3A, ATM, AURKA, BAX, BTG2, CARM1, CASP2, CCNB1, CCND1, CDC25C, CDK1, CDK2, CDKN1A, CDKN1B, CHEK2, CNOT1, CNOT10, CNOT11, CNOT2, CNOT3, CNOT4, CNOT6, CNOT6L, CNOT7, CNOT8, CNOT9, CRADD, E2F1, E2F4, E2F7, E2F8, EP300, FBXO31, GADD45A, GIGYF2, GML, GTSE1, MDM2, MDM4, MUC1, PCBP4, PCNA, PIDD1, PLAGL1, PLK2, PLK3, PLK5, PML, PRKDC, PRMT1, RBL2, RFWD3, RGCC, RPA2, RPL26, RPS27L, SFN, SOX4, TFDP1, TFDP2, TNKS1BP1, TP53, TRIAP1, ZNF385A	64
GO	GO_G0_TO_G1_TRANSITION	APAF1, BMI1, BRCA1, CBX3, CBX5, CCN2, CDC7, CDK3, CHEK1, DAB2IP, DUX4, E2F1, E2F6, EED, EHMT1, EHMT2, EPC1, EZH2, FOXO4, HLA-G, L3MBTL2, MAX, MED1, MGA, MIR424, MIR503, PCGF2, PCGF6, PHC1, PHC3, PPP2R5B, RAD51, RBBP4, RBBP7, RBBP8, RHNO1, RING1, RNF2, RRM2, RYBP, SOX15, SUZ12, TFDP1, TFDP2, UXT, YAF2	46
GO	GO_EXIT_FROM_MITOSIS	ANLN, CDC14A, CDC14B, CDC14C, CDC23, CDCA5, CDKN1B, CHMP2A, CHMP4B, CHMP7, CLASP1, CLASP2, CTDSP1, EPS8, KNTC1, MAD2L1BP, NEUROG1, NPM2, PHB2, PPP1R9B, PPP2R2D, RGCC, SIRT2, SIRT7, SPAST, UBE2C, UBE2S, ZW10	28
GO	GO_MITOTIC_G2_M_TRANSITION_CHECKPOINT	ATM, BLM, BRCA1, BRSK1, CDC6, CDK1, CDK5RAP3, CDKN1A, CLSPN, DONSON, FOXN3, FOXO4, HMGA2, HUS1, HUS1B, MIR21, MRE11, MRNIP, NAE1, NBN, NOP53, ORC1, RAD17, RINT1, SYF2, TAOK1, TAOK2, TAOK3, TICRR, TOPBP1, TRIM39, ZNF830	32
GO	GO_REGULATION_OF_TRANSCRIPTION_INVOLVED_IN_G1_S_TRANSITION_OF_MITOTIC_CELL_CYCLE	BACH1, BRD4, CCNA1, CCNE1, CDC45, CDC6, CDK14, CDK15, CDK16, CDK17, CDK18, CDK5, CDT1, DHFR, DHFRP1, E2F1, E2F4, E2F6, E2F7, ESRRB, FBXO5, GFI1, HINFP, ID2, KLF11, NPAT, ORC1, PCNA, POLA1, RB1, RRM2, TFDP1, TFDP2, TYMS, ZPR1	35
GO	GO_MITOTIC_SISTER_CHROMATID_COHESION	ATRX, BOD1, CDC20, CDCA5, CHTF8, DSCC1, HASPIN, MACROH2A1, MAU2, NAA10, NAA50, NIPBL, NSMCE2, PDS5A, PDS5B, POGZ, RAD21, RAD21L1, RB1, REC8, SGO1, SLF1, SLF2, SMC1A, SMC5, TNKS	26
REACTOME	REACTOME_MITOTIC_G1_PHASE_AND_G1_S_TRANSITION	ABL1, AKT1, AKT2, AKT3, CABLES1, CCNA1, CCNA2, CCNB1, CCND1, CCND2, CCND3, CCNE1, CCNE2, CCNH, CDC25A, CDC45, CDC6, CDC7, CDK1, CDK2, CDK4, CDK6, CDK7, CDKN1A, CDKN1B, CDKN1C, CDKN2A, CDKN2B, CDKN2C, CDKN2D, CDT1, CKS1B, CUL1, DBF4, DHFR, DYRK1A, E2F1, E2F2, E2F3, E2F4, E2F5, E2F6, FBXO5, GMNN, HDAC1, JAK2, LIN37, LIN52, LIN54, LIN9, LYN, MAX, MCM10, MCM2, MCM3, MCM4, MCM5, MCM6, MCM7, MCM8, MNAT1, MYBL2, MYC, ORC1, ORC2, ORC3, ORC4, ORC5, ORC6, PCNA, POLA1, POLA2, POLE, POLE2, POLE3, POLE4, PPP2CA, PPP2CB, PPP2R1A, PPP2R1B, PPP2R2A, PPP2R3B, PRIM1, PRIM2, PSMA1, PSMA2, PSMA3, PSMA4, PSMA5, PSMA6, PSMA7, PSMB1, PSMB10, PSMB2, PSMB3, PSMB4, PSMB5, PSMB6, PSMB7, PSMB8, PSMB9, PSMC1, PSMC2, PSMC3, PSMC4, PSMC5, PSMC6, PSMD1, PSMD10, PSMD11, PSMD12, PSMD13, PSMD14, PSMD2, PSMD3, PSMD4, PSMD5, PSMD6, PSMD7, PSMD8, PSMD9, PSME1, PSME2, PSME3,	149

		<i>PSMF1, PTK6, RB1, RBBP4, RBL1, RBL2, RPA1, RPA2, RPA3, RPA4, RPS27A, RRM2, SEM1, SKP1, SKP2, SRC, TFDP1, TFDP2, TK1, TOP2A, TYMS, UBA52, UBB, UBC, WEE1</i>	
KEGG	-	-	0
PID	-	-	0
BIOCARTA	-	-	0
Total		<i>ABL1, AKT1, AKT2, AKT3, CABLES1, CCNA1, CCNA2, CCNB1, CCND1, CCND2, CCND3, CCNE1, CCNE2, CCNH, CDC25A, CDC45, CDC6, CDC7, CDK1, CDK2, CDK4, CDK6, CDK7, CDKN1A, CDKN1B, CDKN1C, CDKN2A, CDKN2B, CDKN2C, CDKN2D, CDT1, CKS1B, CUL1, DBF4, DHFR, DYRK1A, E2F1, E2F2, E2F3, E2F4, E2F5, E2F6, FBXO5, GMNN, HDAC1, JAK2, LIN37, LIN52, LIN54, LIN9, LYN, MAX, MCM10, MCM2, MCM3, MCM4, MCM5, MCM6, MCM7, MCM8, MNAT1, MYBL2, MYC, ORC1, ORC2, ORC3, ORC4, ORC5, ORC6, PCNA, POLA1, POLA2, POLE, POLE2, POLE3, POLE4, PPP2CA, PPP2CB, PPP2R1A, PPP2R1B, PPP2R2A, PPP2R3B, PRIM1, PRIM2, PSMA1, PSMA2, PSMA3, PSMA4, PSMA5, PSMA6, PSMA7, PSMB1, PSMB10, PSMB2, PSMB3, PSMB4, PSMB5, PSMB6, PSMB7, PSMB8, PSMB9, PSMC1, PSMC2, PSMC3, PSMC4, PSMC5, PSMC6, PSMD1, PSMD10, PSMD11, PSMD12, PSMD13, PSMD14, PSMD2, PSMD3, PSMD4, PSMD5, PSMD6, PSMD7, PSMD8, PSMD9, PSME1, PSME2, PSME3, PSMF1, PTK6, RB1, RBBP4, RBL1, RBL2, RPA1, RPA2, RPA3, RPA4, RPS27A, RRM2, SEM1, SKP1, SKP2, SRC, TFDP1, TFDP2, TK1, TOP2A, TYMS, UBA52, UBB, UBC, WEE1, ADAM17, ADAMTS1, AIF1, ANKRD17, ANXA1, APEX1, CDK10, CUL4A, CUL4B, CYP1A1, DDX3X, EGFR, EIF4G1, FGF10, HYAL1, KCNA5, KMT2E, LSM10, LSM11, MDM2, MEPCE, MIR208A, MIR214, MIR221, MIR222, MIR29A, MIR372, MIR495, MIR515-1, MIR515-2, MIR519D, MIR520A, MIR520H, MIR590, MTBP, PLCB1, PLRG1, RDX, RGCC, RPTOR, TERT, UBE2E2, APAF1, BMI1, BRCA1, CBX3, CBX5, CCN2, CDK3, CHEK1, DAB2IP, DUX4, EED, EHMT1, EHMT2, EPC1, EZH2, FOXO4, HLA-G, L3MBTL2, MED1, MGA, MIR424, MIR503, PCGF2, PCGF6, PHC1, PHC3, PPP2R5B, RAD51, RBBP7, RBBP8, RHNO1, RING1, RNF2, RYBP, SOX15, SUZ12, UXT, YAF2, ANLN, CDC14A, CDC14B, CDC14C, CDC23, CDCA5, CHMP2A, CHMP4B, CHMP7, CLASP1, CLASP2, CTD1P1, EPS8, KNTC1, MAD2L1BP, NEUROG1, NPM2, PHB2, PPP1R9B, PPP2R2D, SIRT2, SIRT7, SPAST, UBE2C, UBE2S, ZW10, ATM, BLM, BRSK1, CDK5RAP3, CLSPN, DONSON, FOXN3, HMGA2, HUS1, HUS1B, MIR21, MRE11, MRNIP, NAE1, NBN, NOP53, RAD17, RINT1, SYF2, TAOK1, TAOK2, TAOK3, TICRR, TOPBP1, TRIM39, ZNF830, BACH1, BRD4, CDK14, CDK15, CDK16, CDK17, CDK18, CDK5, E2F7, ESRRB, GFI1, HINFP, ID2, KLF11, NPAT, ZPR1, ATRX, BOD1, CDC20, CHTF8, DSCC1, HASPIN, MACROH2A1, MAU2, NAA10, NAA50, NIPBL, NSMCE2, PDS5A, PDS5B, POGZ, RAD21, RAD21L1, REC8, SGO1, SLF1, SLF2, SMC1A, SMC5, TNKS, ARID3A, AURKA, BAX, BTG2, CARM1, CASP2, CDC25C, CHEK2, CNOT1, CNOT10, CNOT11, CNOT2, CNOT3, CNOT4, CNOT6, CNOT6L, CNOT7, CNOT8, CNOT9, CRADD, E2F8, EP300, FBXO31, GADD45A, GIGYF2, GML, GTSE1, MDM4, MUC1, PCBP4, PIDD1, PLAGL1, PLK2, PLK3, PLK5, PML, PRKDC, PRMT1, RFWD3, RPL26, RPS27L, SFN, SOX4, TNKS1BP1, TP53, TRIAP1, ZNF385A</i>	368

^a Genes were selected based on online datasets (<http://software.broadinstitute.org/gsea/msigdb/search.jsp>) and literatures;

^b 50 duplicated genes and 1 pseudogene had been excluded;

Keyword: Mitotic AND phase;

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

Abbreviations: OS, overall survival; NSCLC, non-small cell lung cancer; PLCO, the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial; PC, principal component.

* The first 4 PC were used for adjustment for population stratification in the multivariate analysis.

Supplementary Table S4. Three independent SNPs were adjusted for 41 previously published SNPs for NSCLC in the PLCO Trial

Variables	Allele^a	HR (95% CI)	P
rs779901	C/T	0.80 (0.68-0.94)	0.008
rs3806116	G/T	1.18 (1.06-1.32)	0.003
rs199731120	C/CA	1.15 (0.95-1.39)	0.153
rs10794069	A/G	1.15 (1.02-1.30)	0.025
rs1732793	G/A	1.20 (1.04-1.38)	0.013
rs225390	G/A	0.87 (0.75-1.02)	0.088
rs3788142	G/A	1.16 (1.02-1.31)	0.024
rs73049469	C/A	1.31 (1.10-1.56)	0.003
rs35970494	T/TC	1.03 (0.87-1.22)	0.733
rs225388	G/A	1.02 (0.88-1.18)	0.824
rs7553295	G/T	0.84 (0.74-0.95)	0.006
rs1279590	G/A	0.86 (0.73-1.00)	0.052
rs73534533	C/A	0.73 (0.60-0.89)	0.002
rs677844	T/C	0.85 (0.75-0.97)	0.013
rs4978754	C/T	0.83 (0.68-1.02)	0.077
rs1555195	C/T	0.85 (0.75-0.97)	0.012
rs11660748	A/G	1.27 (1.08-1.50)	0.005
rs73440898	A/G	1.14 (0.95-1.37)	0.148
rs13040574	C/A	0.86 (0.77-0.96)	0.007
rs469783	T/C	0.84 (0.75-0.94)	0.002
rs36071574	G/A	1.32 (1.04-1.67)	0.021
rs7242481	G/A	0.87 (0.78-0.98)	0.019
rs1049493	T/C	0.90 (0.81-1.01)	0.065
rs1801701	C/T	0.79 (0.65-0.96)	0.018
rs35859010	C/T	0.82 (0.71-0.96)	0.011

rs1833970	T/A	0.84 (0.75-0.94)	0.003
rs254315	T/C	1.26 (1.00-1.46)	0.002
rs425904	T/C	1.24 (1.09-1.41)	0.001
rs35385129	C/A	0.80 (0.69-0.94)	0.005
rs4487030	A/G	0.85 (0.77-0.95)	0.004
rs60571065	T/A	1.06 (0.89-1.26)	0.505
rs13213007	G/A	1.16 (1.02-1.31)	0.021
rs115613985	T/A	1.21 (1.08-1.36)	0.001
rs9673682	T/G	0.94 (0.84-1.05)	0.278
rs2011404	C/T	0.83 (0.71-0.96)	0.013
rs7867814	G/A	1.12 (0.95-1.32)	0.171
rs2547235	C/T	1.23 (1.06-1.42)	0.005
rs4733124	T/C	1.30 (1.12-1.50)	0.001
rs11225211	G/A	0.82 (0.70-0.95)	0.008
rs11787670	A/G	0.84 (0.69-1.04)	0.113
rs67715745	T/C	0.90 (0.77-1.04)	0.150
rs76744140	T/C	1.35 (1.09-1.66)	0.006
rs6939623	G/T	1.24 (1.03-1.50)	0.021
rs113181986	G/C	0.77 (0.60-1.00)	0.046

Abbreviations: SNP: single-nucleotide polymorphisms; NSCLC, non-small cell lung cancer; PLCO, the Prostate, Lung, Colorectal and Ovarian cancer screening trial; HR: hazards ratio; CI: confidence interval;

^a Reference/effect allele.

Supplementary Table S5. Stratified analysis for associations between (0-1 vs. 2-3) unfavorable genotypes and NSCLC survival in the PLCO trial

Characteristics	0-1 unfavorable genotype	2-3 unfavorable genotype	Multivariate Analysis ^b for OS			Multivariate Analysis ^b for DSS		
	Frequency ^a	Frequency ^a	HR (95% CI)	<i>P</i>	<i>P</i> _{inter} ^c	HR (95% CI)	<i>P</i>	<i>P</i> _{inter} ^c
Age (years)					0.628			0.666
≤ 71	472	162	1.31 (1.05-1.63)	0.016		1.30 (1.03-1.64)	0.026	
> 71	400	141	1.41 (1.12-1.77)	0.003		1.36 (1.06-1.72)	0.016	
Sex					0.236			0.227
Male	509	186	1.39 (1.15-1.69)	0.001		1.36 (1.11-1.68)	0.004	
Female	363	117	1.46 (1.12-1.89)	0.005		1.41 (1.07-1.86)	0.014	
Smoking status					0.441			0.746
Never	80	34	1.37 (0.77-2.44)	0.279		1.32 (0.74-2.36)	0.345	
Current	321	96	1.28 (0.96-1.72)	0.095		1.19 (0.87-1.63)	0.277	
Former	471	173	1.48 (1.21-1.81)	<0.001		1.47 (1.19-1.82)	<0.001	
Histology					0.005			0.016
Adeno	428	147	1.44 (1.14-1.81)	0.003		1.39 (1.09-1.78)	0.008	
Squamous	211	73	1.32 (0.95-1.84)	0.103		1.28 (0.89-1.84)	0.191	
Others	233	83	1.42 (1.07-1.89)	0.017		1.38 (1.02-1.88)	0.036	
Tumor stage					0.005			0.053
I-III A	499	155	1.65 (1.29-2.12)	<0.001		1.61 (1.22-2.12)	<0.001	
III B-IV	373	148	1.19 (0.98-1.46)	0.086		1.19 (0.97-1.46)	0.105	
Chemotherapy					0.002			0.022
No	473	165	1.38 (1.10-1.73)	0.006		1.28 (1.00-1.65)	0.051	
Yes	399	138	1.29 (1.04-1.61)	0.022		1.30 (1.04-1.62)	0.023	
Radiotherapy					0.018			0.028
No	557	204	1.37 (1.12-1.67)	0.002		1.37 (1.11-1.71)	0.004	
Yes	315	99	1.27 (0.99-1.64)	0.064		1.22 (0.93-1.59)	0.148	
Surgery					0.013			0.029
No	467	168	1.26 (1.04-1.52)	0.019		1.24 (1.02-1.51)	0.031	
Yes	405	135	1.55 (1.17-2.05)	0.003		1.58 (1.15-2.17)	0.005	

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 10 with missing data 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 characteristics and unfavorable genotypes.

Supplementary Table S6. Stratified Cox proportional hazards model analysis for clinical covariates and three independent SNPs by the combination of stage and chemotherapy for NSCLC in the PLCO trial

Characteristics	Category	zph <i>P</i> ^a	HR (95% CI) ^b	<i>P</i> ^b
Age (years)	Continuous	0.102	1.03 (1.02-1.05)	<0.0001
Sex	Male	--	1.00	--
	Female	0.405	0.79 (0.68-0.91)	0.002
Smoking status	Never	--	1.00	--
	Current	0.537	1.74 (1.30-2.33)	0.0002
	Former	0.964	1.64 (1.25-2.17)	0.0004
Histology	Adenocarcinoma	--	1.00	--
	Squamous cell carcinoma	0.211	1.12 (0.93-1.35)	0.231
	Others	0.586	1.23 (1.03-1.45)	0.020
Radiotherapy	No	--	1.00	--
	Yes	0.196	0.90 (0.76-1.06)	0.197
Surgery	No	--	1.00	--
	Yes	0.130	0.25 (0.19-0.32)	<0.0001
<i>CHEK1</i> rs76744140	TT/TC/CC	0.118	1.24 (1.03-1.50)	0.026
<i>PRIM2</i> rs6939623	GG/GT/TT	0.350	1.32 (1.11-1.56)	0.001
<i>CDK6</i> rs113181986	GG/GC/CC	0.492	0.81 (0.64-1.02)	0.077

Abbreviations: SNP: single-nucleotide polymorphisms; NSCLC, non-small cell lung cancer; PLCO, the Prostate, Lung, Colorectal and Ovarian cancer screening trial; HR: hazards ratio; CI: confidence interval;

^a zph test of stratified Cox proportional hazards model by the combination of stage and chemotherapy;

^b Adjusted for age, sex, smoking status, histology, radiotherapy, surgery, PC1, PC2, PC3 and PC4.

Supplementary Table S7. Function prediction for *CHEK1* rs76744140 and *PRIM2* rs6939623

SNP	Gene	Chr	RegDB ^a	Haploreg v4.1 ^b						
				Promoter histone marks	Enhancer histone marks	DNase	Proteins bound	Motifs changed	Selected eQTL hits	dbSNP func annot
rs76744140	<i>CHEK1</i>	11	7	--	LNG	--	--	BDP1,KAP1	2 hits	intronic
rs6939623	<i>PRIM2</i>	6	5	--	--	PANC	--	5 altered motifs	--	intronic

Abbreviations: SNP, single nucleotide polymorphism; Chr, chromosome; DNase, deoxyribonuclease; eQTL, expression quantitative trait loci; dbSNP func annot, dbSNP function annotation;

^a <http://regulomedb.org/>

^b Haploreg: <https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php>

Supplementary Table S8. Function prediction for *CDK6* rs113181986 and its 24 high LD ($r^2 > 0.8$) SNPs in *CDK6*.

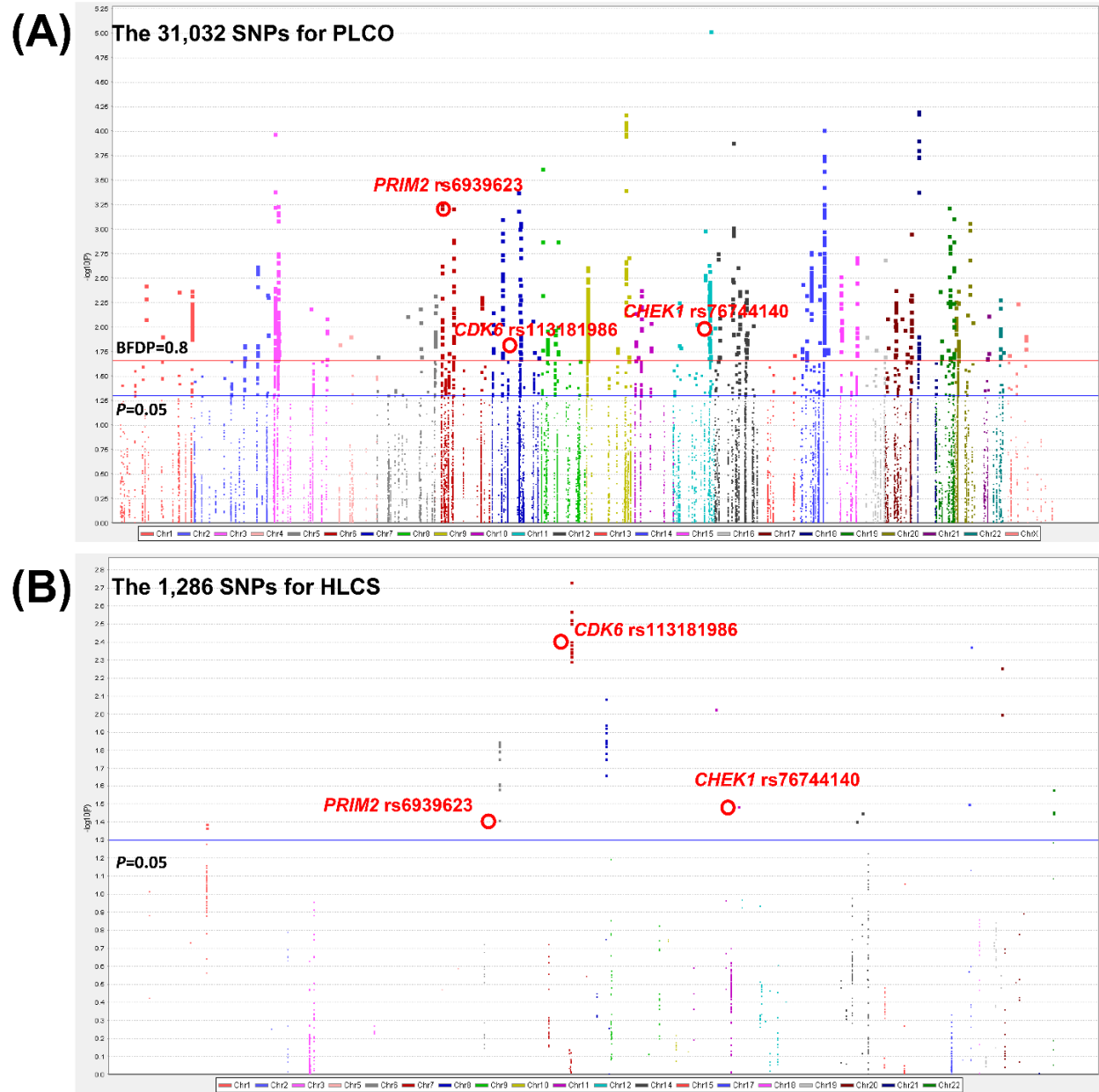
SNP	Gene	Chr	RegDB ^a	Haploreg v4.1 ^b							
				LD (r^2)	Promoter histone marks	Enhancer histone marks	DNase	Proteins bound	Motifs changed	Selected eQTL hits	dbSNP func annot
rs10277569	2.9kb 3' of <i>CDK6</i>	7	3a	0.94	--	--	BLD	--	12 altered motifs	--	--
rs10277576	2.9kb 3' of <i>CDK6</i>	7	5	0.94	--	--	BLD	--	LBP-1,STAT	--	--
rs76682494	1.4kb 3' of <i>CDK6</i>	7	4	0.94	--	4 tissues	BLD	--	AP-2,AP-4,p53	1 hit	--
rs36045255	1.4kb 3' of <i>CDK6</i>	7	5	0.94	--	4 tissues	LIV,SKIN	--	5 altered motifs	--	--
rs113665429	<i>CDK6</i>	7	6	0.92	IPSC, PANC	ESC, IPSC	--	--	11 altered motifs	--	3'-UTR
rs76908789	<i>CDK6</i>	7	3a	0.92	14 tissues	7 tissues	17 tissues	ERALPHA_A, TCF4,ELF1	13 altered motifs	--	3'-UTR
rs3731384	<i>CDK6</i>	7	2b	0.95	--	4 tissues	13 tissues	CTCF	31 altered motifs	--	intronic
rs3731380	<i>CDK6</i>	7	3a	0.95	--	HRT, MUS	MUS	--	4 altered motifs	--	intronic
rs3731373	<i>CDK6</i>	7	5	0.95	--	BLD, MUS	--	--	Mxi1,Myc,ZEB1	--	intronic
rs3731348	<i>CDK6</i>	7	5	1.00	--	5 tissues	--	--	Foxl1,HMG-IY,Mef2	--	intronic
rs75427382	<i>CDK6</i>	7	2a	1.00	--	5 tissues	HRT,KID,MUS	--	NF-I,TLX1::NFIC	--	intronic
rs61628822	<i>CDK6</i>	7	3a	1.00	--	5 tissues	ADRL	--	Dlx2,Dlx3,Sox	--	intronic
rs111450372	<i>CDK6</i>	7	3a	1.00	--	12 tissues	SKIN,VAS	--	15 altered motifs	--	intronic
rs76230393	<i>CDK6</i>	7	4	0.97	BLD	BLD, HRT, MUS	ADRL,HRT	--	--	--	intronic
rs79355747	<i>CDK6</i>	7	2a	0.90	11 tissues	15 tissues	41 tissues	4 bound proteins	4 altered motifs	--	intronic
rs113181986	<i>CDK6</i>	7	6	1.00	--	BLD, THYM	--	--	CTCF,Myf,RP58	--	intronic
rs75237922	<i>CDK6</i>	7	5	0.97	BLD	7 tissues	4 tissues	--	4 altered motifs	--	intronic
rs77183195	<i>CDK6</i>	7	3a	1.00	--	21 tissues	8 tissues	GATA1	7 altered motifs	--	intronic

rs3731334	CDK6	7	4	1.00	GI, LNG, CRVX	14 tissues	10 tissues	--	Myc	--	intronic
rs113861828	CDK6	7	4	0.97	--	10 tissues	HRT	EBF1	NRSF,Sin3Ak- 20	--	intronic
rs76527309	CDK6	7	6	1.00	--	BLD, BRN, GI	SKIN	--	13 altered motifs	--	intronic
rs79835522	CDK6	7	3a	1.00	--	7 tissues	9 tissues	--	ERalpha- a,TBX5,Tgif1	--	intronic
rs78716704	CDK6	7	7	1.00	--	BLD	SKIN	--	PRDM1	--	intronic
rs113793043	CDK6	7	5	0.89	--	BLD	--	--	--	--	intronic
rs111702206	CDK6	7	3a	0.85	--	4 tissues	6 tissues	--	Nrf1	--	intronic

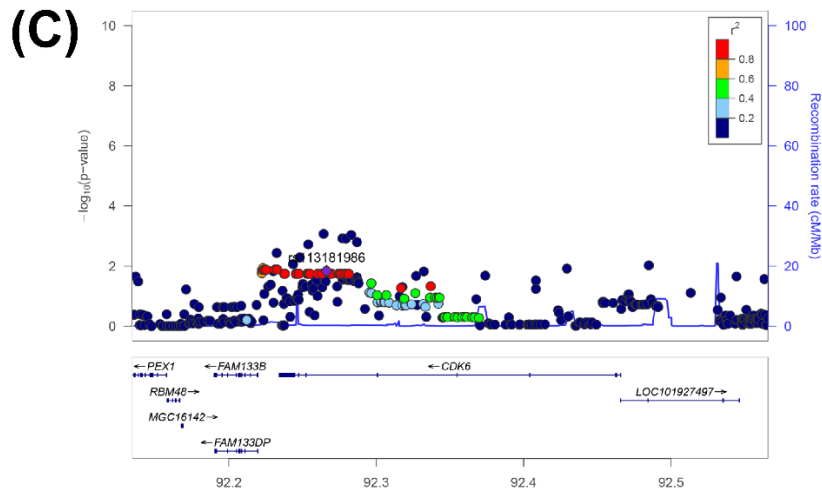
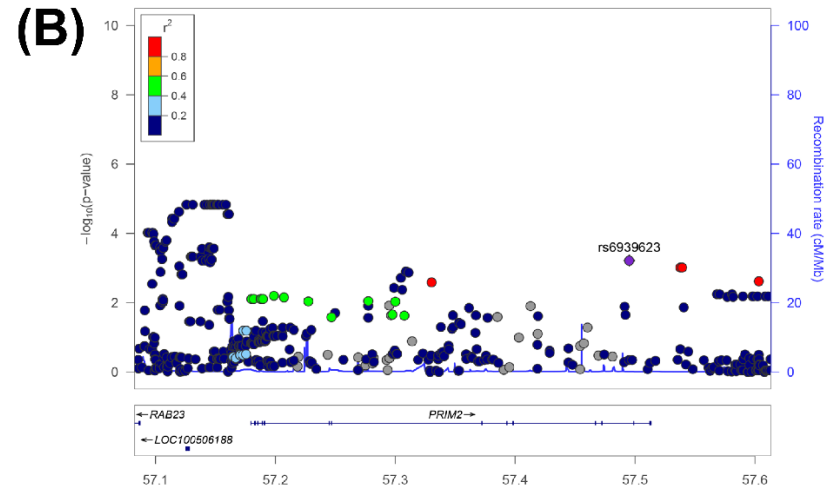
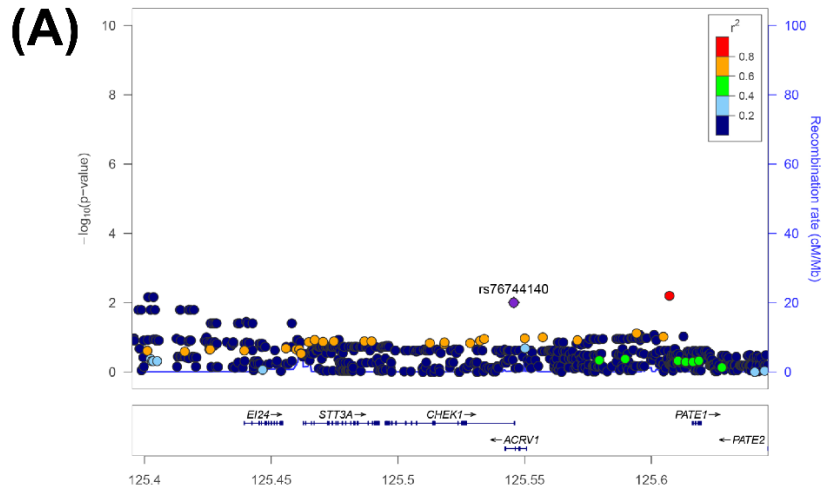
Abbreviations: SNP, single nucleotide polymorphism; Chr, chromosome; DNase, deoxyribonuclease; eQTL, expression quantitative trait loci; dbSNP func annot, dbSNP function annotation;

^a <http://regulomedb.org/>

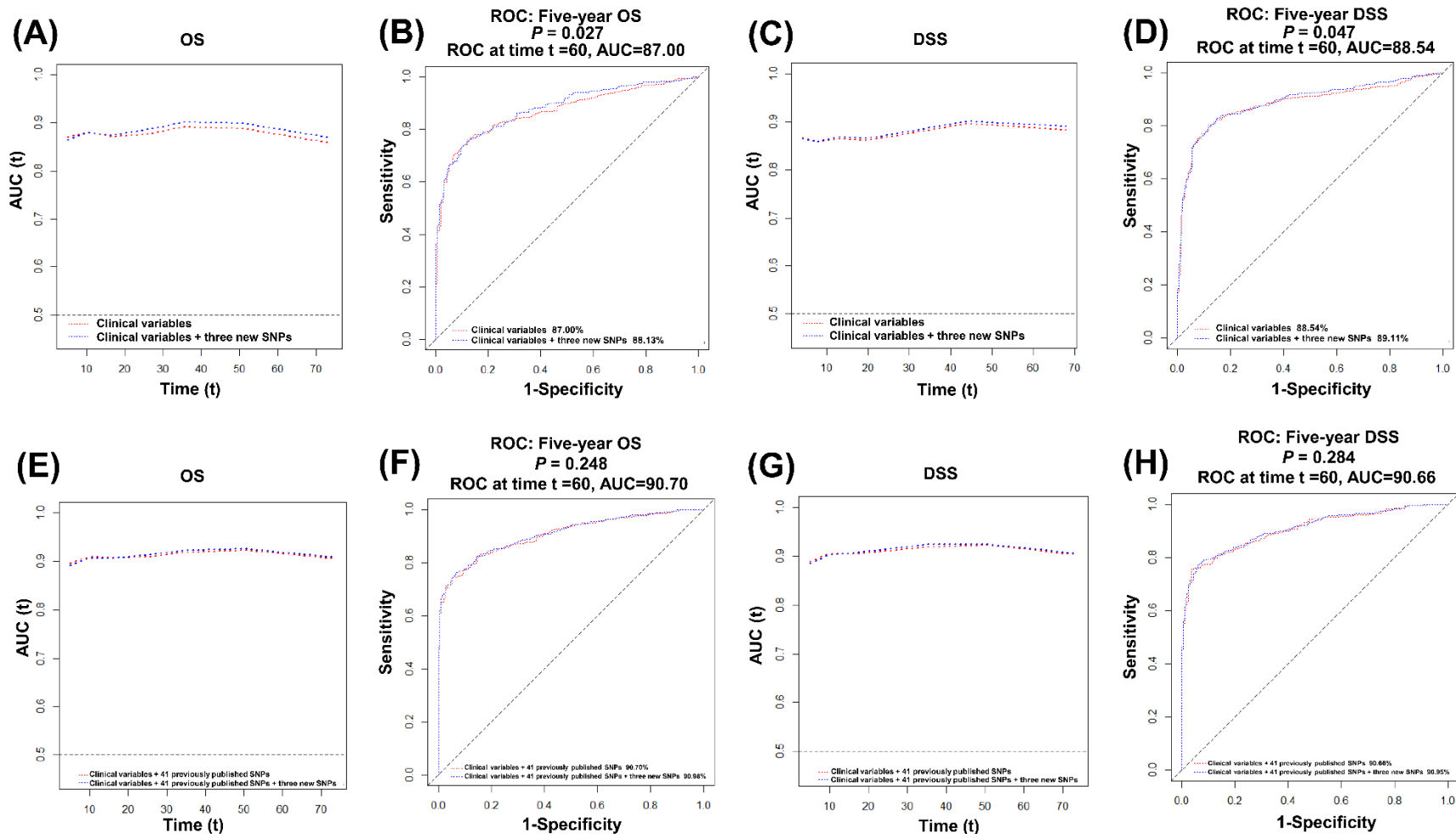
^b Haploreg: <https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php>



Supplementary Figure S1. Manhattan plot. **(A)** Manhattan plot for 31,032 SNPs of mitotic phase-related pathway genes in the PLCO trial; **(B)** Manhattan plot for 1,286 SNPs in the HLCS dataset. The blue horizontal line indicates $P = 0.05$ and the red line indicates $BFDP = 0.80$. Abbreviations: PLCO, Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial; HLCS, the Harvard Lung Cancer Susceptibility Study; BFDP, Bayesian false-discovery probability.

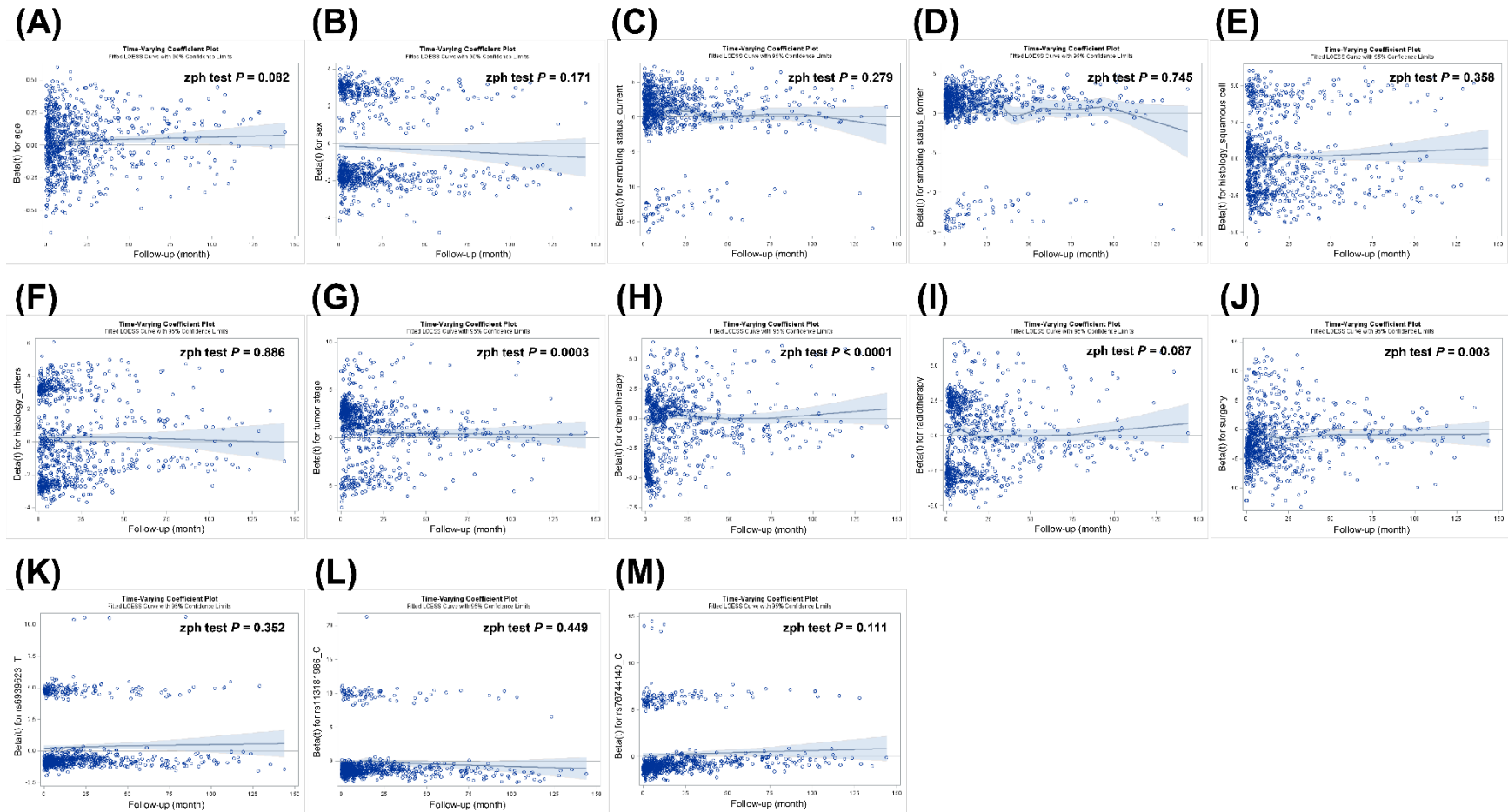


Supplementary Figure S2. Regional association plots for the three independent SNPs in the mitotic phase-related pathway genes. Regional association plots contained 100 kb up or downstream of *CHEK1* (A), *PRIM2* (B) and *CDK6* (C). Data points are colored according to the level of linkage disequilibrium 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 $-\log_{10}(P)$ against chromosomal base-pair position. The right-hand y-axis shows the recombination rate estimated from HapMap Data Rel 22/phase II European population.

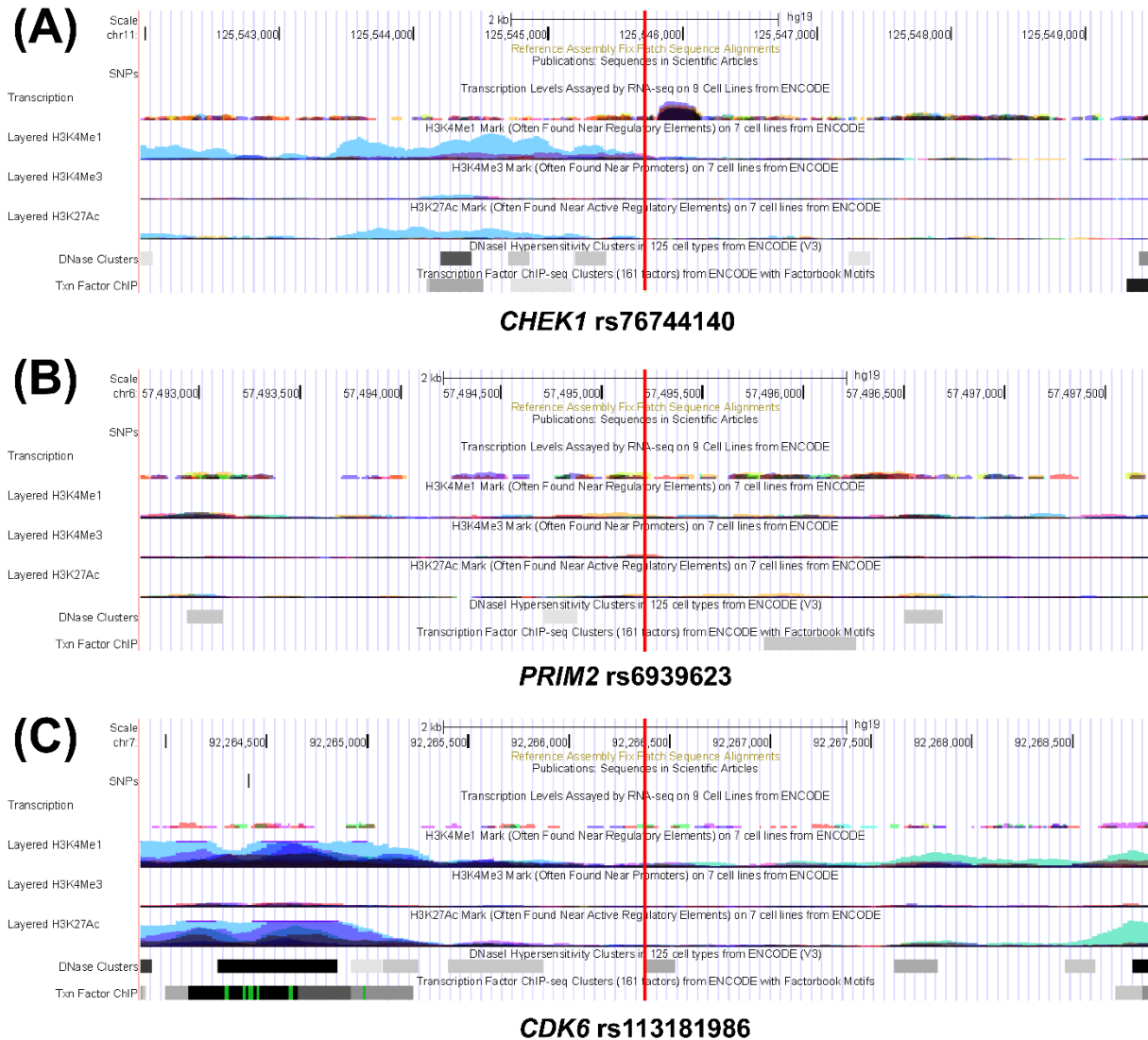


Supplementary Figure S3. Receiver operating characteristic (ROC) curve and time dependent area under the ROC curve (AUC) estimation for NSCLC OS and DSS prediction using three newly independent SNPs in the PLCO trial. The Five-year NSCLC OS prediction by time-dependent AUC (A) and ROC curve based on clinical variables plus three new SNPs (B), red line indicates clinical variables 87.00%; blue line indicates clinical variables plus three new SNPs 88.13%; time t = time (months). The Five-year NSCLC DSS prediction by time-dependent AUC (C) and ROC curve based on clinical variables plus three new SNPs (D), red line indicates clinical variables 88.54%; blue line indicates clinical variables plus new SNPs 89.11%; time t = time (months). The Five-year NSCLC OS prediction by time-dependent AUC (E) and ROC curve based on clinical variables and 41 previously published SNPs plus three new SNPs (F), red line indicates clinical variables plus 41 previously published

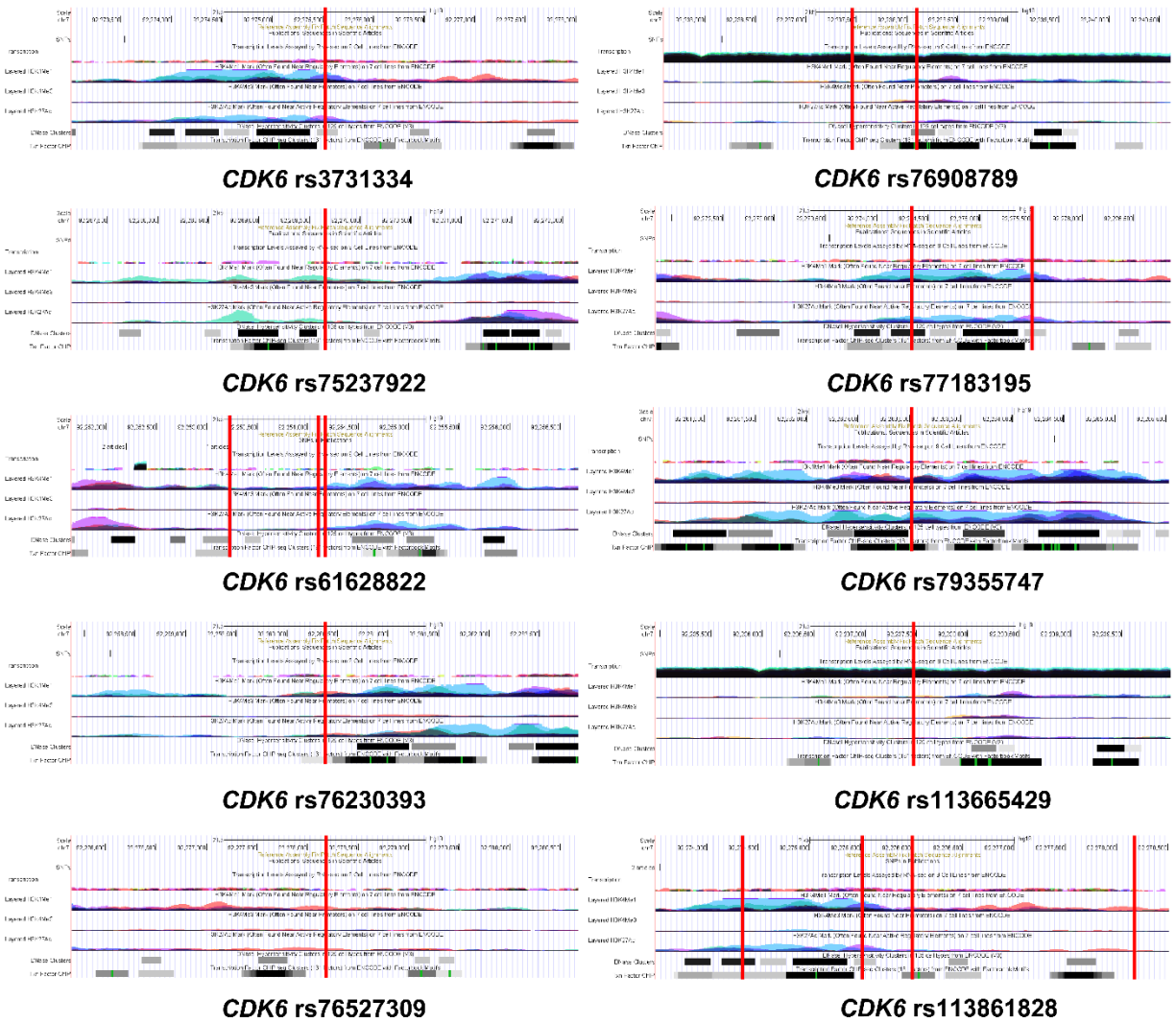
SNPs 90.70%; blue line indicates clinical variables and 41 previously published SNPs plus three new SNPs 90.98%; time t = time (months). The Five-year NSCLC DSS prediction by time-dependent AUC (**G**) and ROC curve based on clinical variables and 41 previously published SNPs plus three new SNPs (**H**), red line indicates clinical variables plus 41 previously published SNPs 90.66%; blue line indicates clinical variables and 41 previously published SNPs plus three new SNPs 90.95%. Abbreviations: OS, Overall survival; DSS, disease-specific survival; NSCLC, Non-small cell lung cancer; PLCO, The Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial.



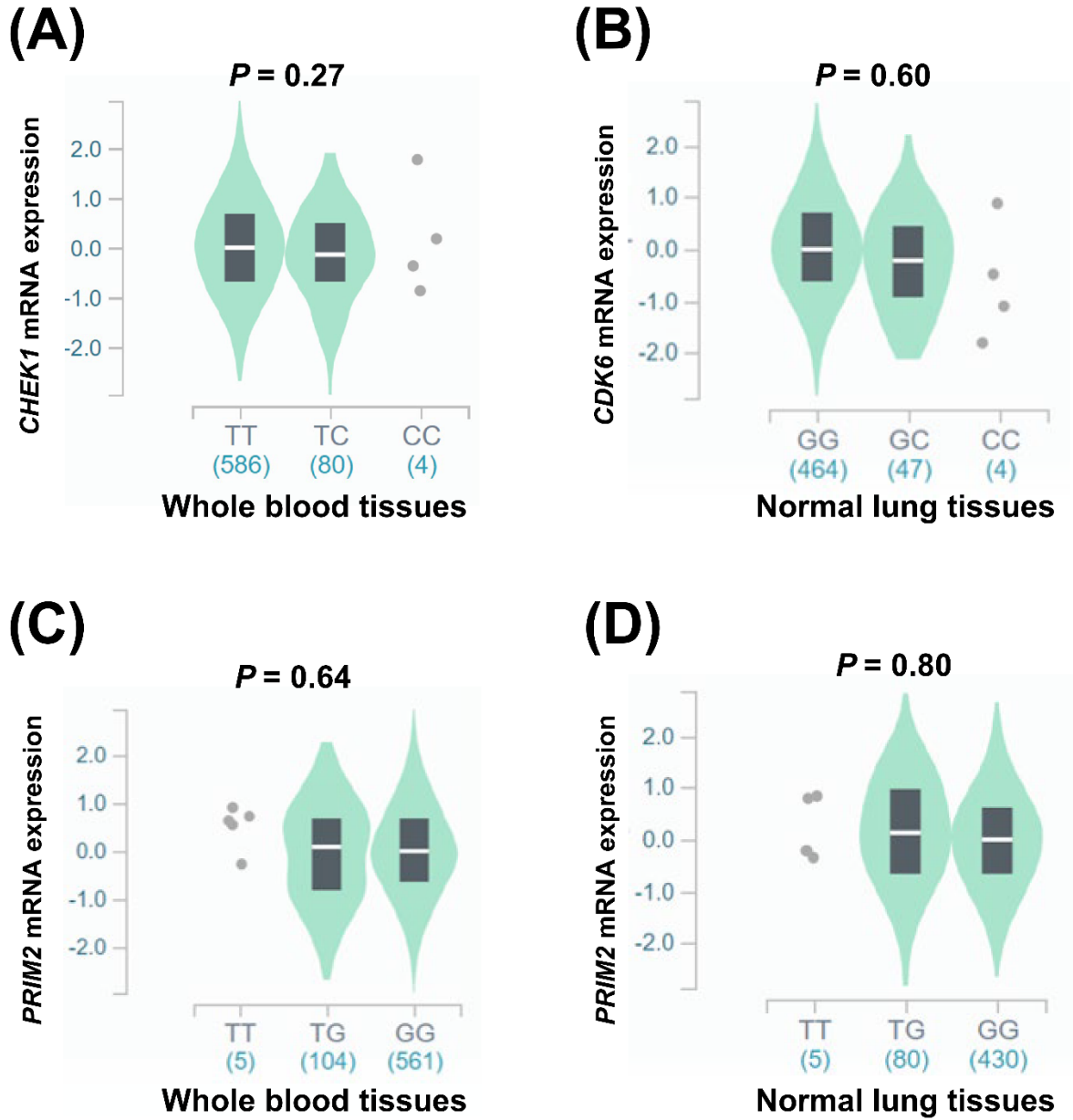
Supplementary Figure S4. Testing proportional Hazards assumption by using Schoenfeld residuals plot and zph test for (A) age; (B) sex; (C) smoking status (current); (D) smoking status (former); (E) histology (squamous cell); (F) histology (others); (G) stage; (H) chemotherapy; (I) radiotherapy; (J) surgery; (K) rs6939623; (L) rs113181986; (M) rs76744140.



Supplementary Figure S5. Functional prediction of three independent SNPs in mitotic phase-related pathway genes in the ENCODE data. Location and functional prediction of *CHEK1* rs76744140 **(A)**. Location and functional prediction of *PRIM2* rs6939623 **(B)**. Location and functional prediction of *CDK6* rs113181986 **(C)**. The H3K4Me3, H3K4Me1, and H3K27Ac tracks showed the genome-wide levels of enrichment of acetylation of lysine 27, the mono-methylation of lysine 4, and tri-methylation of lysine 4 of the H3 histone protein. DNase clusters track showed DNase hypersensitivity areas. Txn factor track showed regions of transcription factor binding of DNA. Abbreviations: ENCODE, Encyclopedia of DNA Elements; SNP, single nucleotide polymorphism.



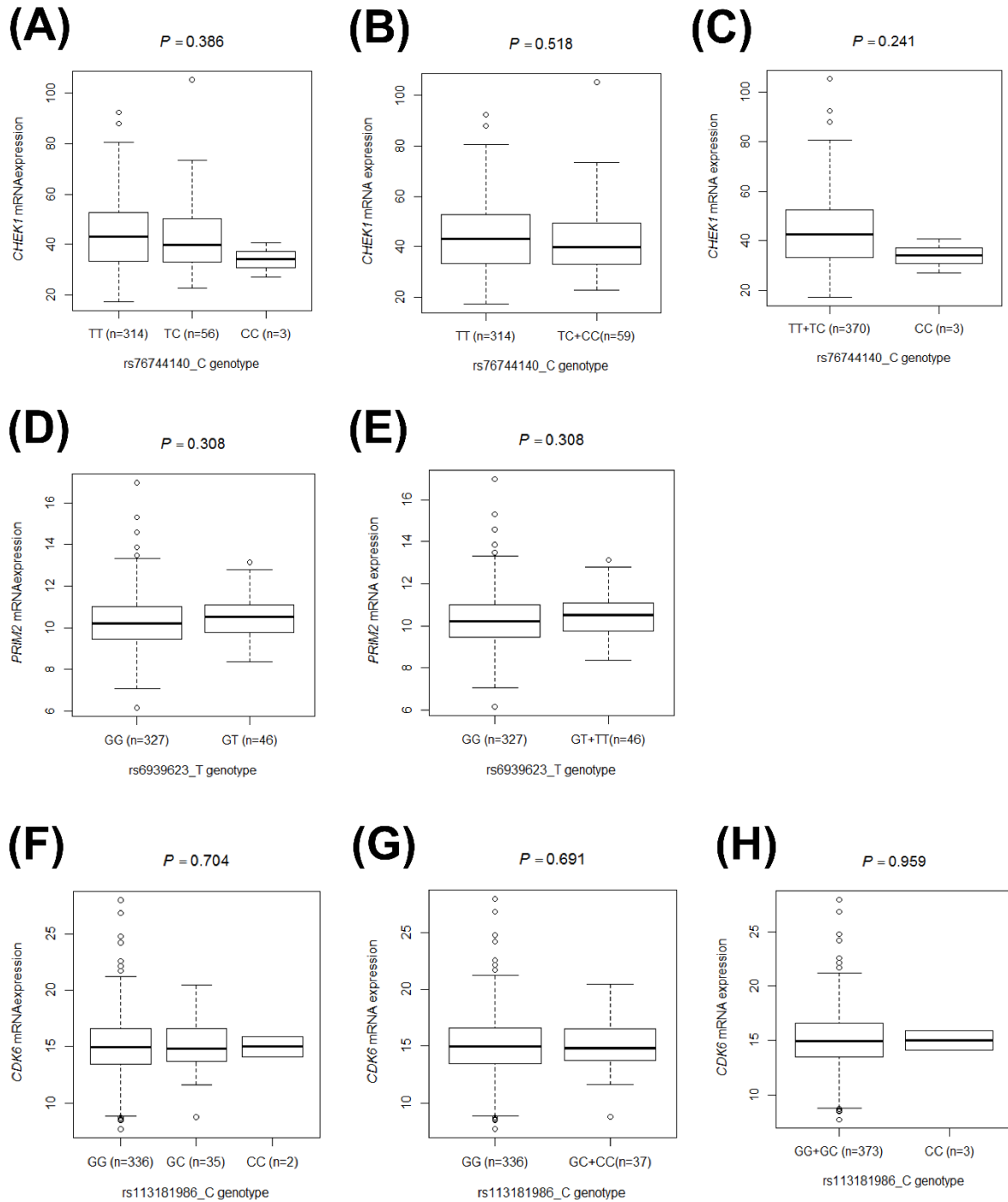
Supplementary Figure S6. Functional prediction of 10 high LD ($r^2 > 0.8$) SNPs of representative SNP rs113181986 in *CDK6* in the ENCODE data. The H3K4Me3, H3K4Me1, and H3K27Ac tracks showed the genome-wide levels of enrichment of acetylation of lysine 27, the mono-methylation of lysine 4, and trimethylation of lysine 4 of the H3 histone protein. DNase clusters track showed DNase hypersensitivity areas. Txn factor track showed regions of transcription factor binding of DNA. Abbreviations: ENCODE, Encyclopedia of DNA Elements; SNP, single nucleotide polymorphism; LD, linkage disequilibrium.



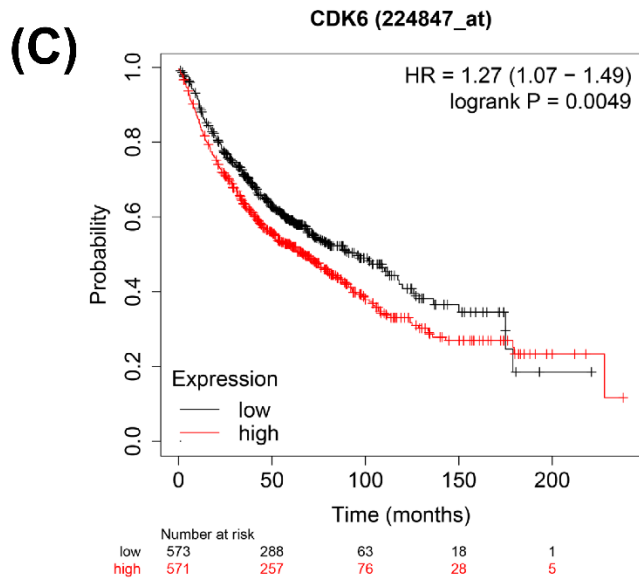
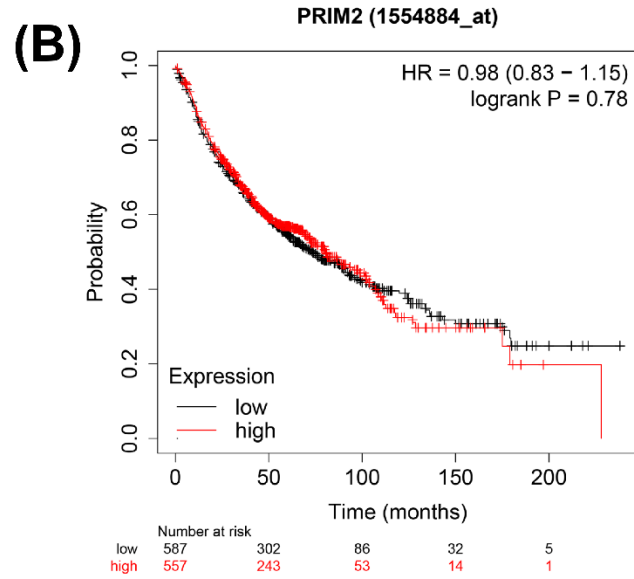
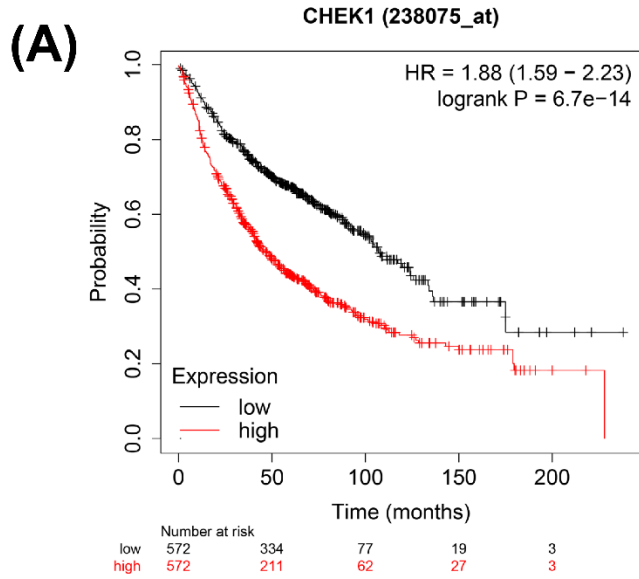
Supplementary Figure S7. The eQTL analysis for *CHEK1* rs76744140, *PRIM2* rs6939623, and *CDK6* rs113181986 from the GTEx (V8) database. **(A)** The correlation of rs76744140 genotypes and *CHEK1* mRNA expression in whole blood tissues; **(B)** The correlation of rs113181986 genotypes and *CDK6* mRNA expression in normal lung tissues; **(C)** The correlation of rs6939623 genotypes and *PRIM2* mRNA expression in whole blood tissues; **(D)** The correlation of rs6939623 genotypes and *PRIM2* mRNA expression in normal lung tissues. Abbreviations: eQTL, expression quantitative trait loci; GTEx, Genotype Tissue Expression project.



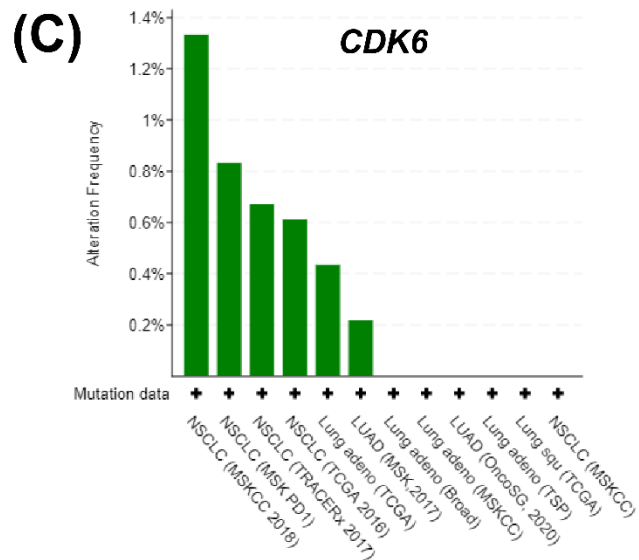
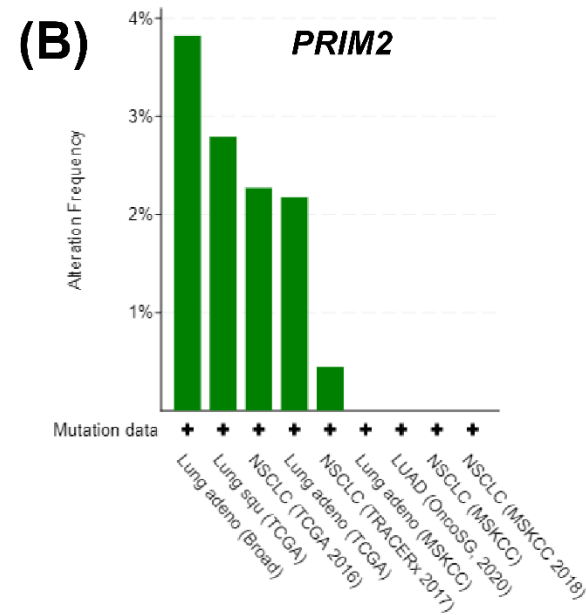
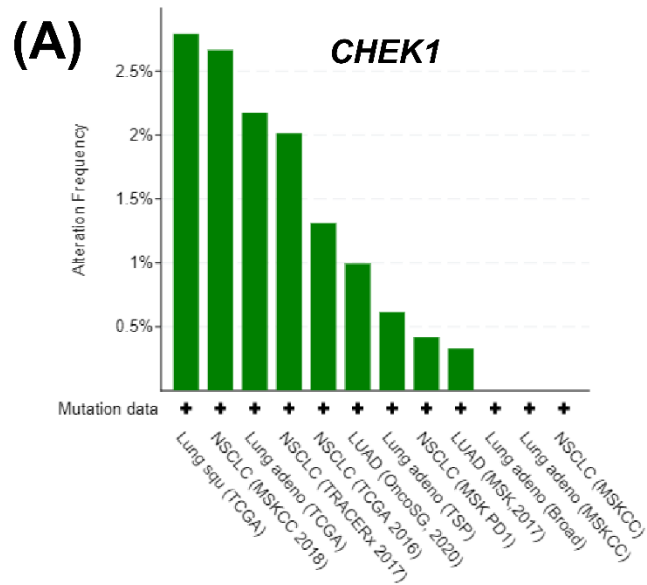
Supplementary Figure S8. The eQTL analysis for 19 SNPs had a high LD ($r^2 > 0.8$) with rs113181986 in *CDK6* from the GTEx (V8) database. The correlation of the 19 SNPs genotypes and *CDK6* mRNA expression in whole blood tissues.



Supplementary Figure S9. The eQTL analysis for *CHEK1* rs76744140, *PRIM2* rs6939623, and *CDK6* rs113181986 from 1,000 Genomes Project. The correlation of rs76744140 genotypes and *CHEK1* mRNA expression in the additive (A), dominant (B) and recessive (C) model, respectively. The correlation of rs6939623 genotypes and *PRIM2* mRNA expression in the additive (D) and dominant (E) model, respectively. The correlation of rs113181986 genotypes and *CDK6* mRNA expression in the additive (F), dominant (G) and recessive (H) model, respectively.



Supplementary Figure S10. Kaplan-Meier analysis for patients with NSCLC by the three genes. Based on online survival analysis software, **(A)** high *CHEK1* expression was associated with worse survival of patients with NSCLC; **(B)** *PRIM2* expression were not significantly associated with survival of patients with NSCLC; **(C)** high *CDK6* expression was associated with worse survival of patients with NSCLC. Abbreviations: NSCLC, non-small cell lung cancer. the online Kaplan-Meier plotter was obtained from <http://kmplot.com/analysis/>.



Supplementary Figure S11. Mutation frequency of *CHEK1*, *PRIM2*, and *CDK6* in non-small cell lung tumor tissues. **(A)** Mutation frequency of *CHEK1* in NSCLC, LUAD and LUSC using the online database of the cBioPortal for Cancer Genomics; **(B)** Mutation frequency of *PRIM2* in NSCLC, LUAD and LUSC using the online database of the cBioPortal for Cancer Genomics; **(C)** Mutation frequency of *CDK6* in NSCLC, LUAD and LUSC using the online database of the cBioPortal for Cancer Genomics. Abbreviations: NSCLC, non-small cell lung cancer; LUAD, Lung adenocarcinoma; LUSC, Lung squamous cell carcinoma. The online cBioPortal for Cancer Genomics database was obtained from <http://www.cbioportal.org>.