Chinese Medical Journal Pulmonary and Critical Care Medicine 2 (2024) 88–94

Contents lists available at ScienceDirect

Chinese Medical Journal Pulmonary and Critical Care Medicine

journal homepage: www.elsevier.com/locate/pccm

Review Article

Single nucleotide variants in lung cancer

Xiaoling Tian, Zhe Liu*

Zhejiang Key Laboratory of Medical Epigenetics, Department of Cell Biology, School of Basic Medical Sciences, Hangzhou Normal University, Hangzhou, Zhejiang 311121, China

ARTICLE INFO

Edited by: Peifang Wei

Keywords:
Lung cancer
Germline mutation
Single nucleotide variants
Susceptibility
Outcome
Precision medicine

ABSTRACT

Germline genetic variants, including single-nucleotide variants (SNVs) and copy number variants (CNVs), account for interpatient heterogeneity. In the past several decades, genome-wide association studies (GWAS) have identified multiple lung cancer-associated SNVs in Caucasian and Chinese populations. These variants either reside within coding regions and change the structure and function of cancer-related proteins or reside within non-coding regions and alter the expression level of cancer-related proteins. The variants can be used not only for cancer risk assessment and prevention but also for the development of new therapies. In this review, we discuss the lung cancer-associated SNVs identified to date, their contributions to lung tumorigenesis and prognosis, and their potential use in predicting prognosis and implementing therapeutic strategies.

Introduction

Lung cancer is the most frequent cause of cancer-related death worldwide, with 5-year survival rates varying from 4–17%.[1] Small cell lung carcinoma (SCLC) and non-small cell lung carcinoma (NSCLC) are the two histopathologic types of lung cancer. NSCLC consists of adenocarcinoma (ADC), squamous cell carcinoma (SCC), adenosquamous carcinoma (ASC), and large cell carcinoma (LCC). The alteration in oncogenic driver genes is one of the important causes of lung cancer. Tumor-suppressor gene *TP53* is the most frequently mutated gene (approximately 40%) in all types of lung cancer.[2] Lung cancer subtypes also harbor specific mutations and variants in other genes, such as *EGFR* and *KRAS* in ADC,[3] *CDKN2A* and *RB1* in SCC,[4] and *RB1*, *PTEN*, and *MYC* in SCLC.[2]

Pathogenic variants in the germline of an individual, which are heritable, are called germline variants. A growing amount of evidence has shown the important role of germline variations in lung cancer initiation and progression. Genetic variations include single nucleotide variants (SNVs), insertions, deletions, structural variants, and repeat variations.[5] If the minor allele frequency (MAF) of an SNV is more than 1%, a variant is called a polymorphism. Single nucleotide polymorphisms (SNPs) are the most common genetic variation type among all germline genetic variations, accounting for 90% of all polymorphisms in the genome.

SNVs in different regions

SNVs can reside in coding regions and non-coding regions of the genome. Approximately 10% of disease-associated SNVs are located in coding regions and 90% are located in non-coding regions.[6] SNVs in different regions modulate diseases through different mechanisms.

SNVs in coding regions

Genetic variants in a coding region change the sequence of amino acids and influence protein function. Germline variants in lung cancer driver genes such as *EGFR*, *KRAS*, and *P53* contribute greatly to tumorigenesis and progression. The epidermal growth factor receptor (EGFR) tyrosine kinase domain is encoded by exons 18–24. Over 90% of the known *EGFR* mutations in lung cancer are in exons 19–21. *EGFR*-K757R is the most common *EGFR* germline mutation in Chinese lung cancer patients. The mutation of K757R is associated with the response of lung cancer to chemotherapy. K757R and exon 19del + K757R show similar sensitivity to icotinib and osimertinib, whereas exon 19del + K757R is more sensitive to afatinib and gefitinib than K757R.[7] The less commonly observed *EGFR* germline variants in coding exons have been associated with susceptibility or treatment response of lung cancer. For example, *EGFR* V834L and V843I are associated with susceptibility to lung adenocarcinoma (LUAD), and R776H, V843I, L858R, and P848L are associated with squamous cell lung cancer.[8] Synonymous mutations also affect lung cancer in some cases. The common *EGFR* Q787Q polymorphism showed significant protective effects on the overall survival of patients with *EGFR*-mutant stage IV LUAD treated with EGFR tyrosine kinase inhibitors (TKIs).[9]

RAS proteins are a family of small GTPases that play critical roles in multiple cellular signaling pathways, such as the RAS–mitogen-activated protein kinase (MAPK) pathway. Approximately 83% of

* Corresponding author at: Department of Cell Biology, School of Basic Medical Sciences, Hangzhou Normal University, Hangzhou, Zhejiang 311121, China
E-mail address: Zheliu@hznu.edu.cn (Z. Liu)

https://doi.org/10.1016/j.pccm.2024.04.004
Received 8 May 2023; Available online 14 June 2024

KRAS somatic mutations are found at G12, followed by G13 (14%) and Q61 (2%). Germline *KRAS* mutations have been observed at numerous locations, including K5N, V14I, Q22R, Q22E, P34R, P34L, T58I, G60R, E153V, and F156L, and are associated with lung cancer risk.[10,11] Germline variants in codons 12, 13, or 61 in *KRAS* are rarely found.

The *P53* tumor suppressor gene plays a central role in many cellular processes, such as DNA repair and apoptosis. The *P53* V157D mutation was identified in a family with hereditary lung cancer syndrome. Further mechanistic study showed that the *P53* V157D variant promotes lung cancer cell proliferation.[12] The *P53* P72A germline variant in the *P53* coding region is also associated with lung cancer.[13]

In addition to hotpot oncogene mutation, germline variants in other candidate susceptibility genes have also been widely studied. Eleven lung cancer-associated germline variants are located in genes encoding components of the growth hormone/insulin-like growth factor (GH-IGF) pathway, including rs7214723 in *CAMKK1* (E375G), rs6964587 in *AKAP9* (M463I), and rs6183 in *GHR* (P495T), and genes in the DNA damage-response pathway, including rs11571833 in *BRCA2* (K3326X) and rs28360135 in *XRCC4* (I137T).[14] Rare deleterious variants in DNA repair pathway genes, such as rs77187983 in *EHBP1* (D590V), rs11571833 in *BRCA2* (K326X), and rs752672077 in *MPZL2* (I24M), are also associated with lung cancer risk.[15] Germline mutations in *PARK2* correlate with impaired mitophagy and increase the death of lung cancer cells. Mechanically, A46T Parkin results in inability to translocate to mitochondria and recruit downstream mitophagic regulators, such as optineurin (OPTN) and transcription factor EB (TFEB). Besides, N254S and R275W Parkin display slower mitochondrial translocation than WT Parkin.[16]

Whole-exome sequencing revealed that SCLC frequently harbors germline pathogenic variants in *RAD51D* (Q62X), *CHEK1* (Q346X, R379X), *BRCA2* (R1699W, Y2215X), and *MUTYH* (G396D, V493F, Y179C), which are significantly associated with recurrence-free survival after platinum-based chemotherapy.[17]

SNVs in non-coding regions

A significant number of lung cancer-associated SNVs have been identified in gene promoters, introns, and intergenic regions. Promoter SNVs frequently alter the binding affinity of transcription factors, subsequently influencing the transcriptional regulation of key genes involved in lung tumorigenesis. Intron or intergenic SNVs either modulate the activity of cis-regulatory elements (CREs), such as enhancers and silencers, usually by affecting transcription factor binding or changing the function of non-coding RNAs (ncRNAs).

SNVs in cis-regulatory elements

A non-coding SNV can exert its functional effects through various mechanisms: regulating the transcription of neighboring genes, regulating the transcription of distant genes, or regulating genes located on other chromosomes. This suggests the complexity of eukaryotic genomic assembly. rs3769201 and rs722864, which are located in the introns of *ZAK*, are associated with decreased ZAK mRNA expression and reduced lung cancer risk.[18] rs3117582, located in intron 1 of *BAT3*, is involved in the regulation of p53 acetylation in response to DNA damage; and rs3131379, located in intron 10 of *MSH5*, is involved in DNA mismatch repair and regulates lung cancer susceptibility.[19] Whether these two SNP loci regulate *BAT3* or *MSH5* is not clear. Growing evidence has shown that a non-coding SNP may bypass nearby genes to regulate a gene that is located far away. For example, rs402710, which is located in the last intron of *CLPTM1L*, physically interacts with the *TERT* promoter by looping out the intervening sequences, regulating *TERT* gene expression and lung tumorigenesis[20] (Fig. 1).

Although intrachromosomal regulation is a frequently reported mechanism used by non-coding SNPs to regulate disease, some lung cancer-associated SNP loci exhibit trans-effects. For example, rs1663689, which is located in the intergenic region in chromosome 10p1.4, regulates lung cancer susceptibility and outcome through regulation of adhesion G protein-coupled receptor G6 (*ADGRG6*), which is located chromosome 6, through interchromosomal interaction[21] (Fig. 2).

A non-coding SNP usually exerts its function by changing the binding affinity of transcription factors. For example, rs2853677 is located within the Snail1 binding site in a *TERT* enhancer. The enhancer increases *TERT* transcription when juxtaposed to the *TERT* promoter. rs2853677-T results in the binding of Snail1 to the enhancer and disrupting enhancer-promoter colocalization, which subsequently silences *TERT* transcription (Fig. 3).[22] SNP rs17079281-C, located in the *DCBLD1* promoter, creates a YY1-binding site, resulting in decreased DCBLD1 expression and subsequent decreased cell proliferation.[23] rs9399451 and rs9390123 reside within an enhancer region and influence the binding of POU2F1, which subsequently affects the promoter activity of *PHACTR2-AS1* and *PEX3* in lung cancer cell lines.[24] rs4142441 is located in a MYC binding site in the *OSER1-AS1* promoter region. The G allele of rs4142441 results a higher binding affinity of MYC. MYC binding suppresses the transcription of *OSER1-AS1*, and promotes tumor progression.[25] A lung-specific p53-responsive enhancer of *TNFRSF19* harbors three highly linked common SNPs (rs17336602, rs4770489, and rs34354770) and six p53 binding sequences either close to or located between the variations. The enhancer effectively protects normal lung cell lines against pulmonary carcinogen nicotine-derived nitrosamine ketone (NNK)-induced DNA damage and malignant transformation by upregulating *TNFRSF19* through chromatin looping. These variations significantly weaken the enhancer activity by affecting the p53 response, especially when cells are exposed to NNK.[26]

SNVs in ncRNAs

Most of the human genome is transcribed RNAs that do not encode proteins. These ncRNAs, including microRNAs (miRNAs) and long ncRNAs (lncRNAs), play crucial roles in regulating the initiation and progression of various cancers.[27] Lung cancer-associated SNPs in ncRNAs have been identified. rs11614913 in hsa-mir-196a2 regulates its binding to the *LSP1* 3′UTR, which changes LSP1 expression and survival in individuals with NSCLC.[28] rs10505477 in the lncRNA *CASC8* is highly related to ADC risk in males and highly relevant to severe hematologic toxicity in NSCLC and gastrointestinal toxicity in SCLC after platinum-based chemotherapy.[29] rs140618127, located in the lncRNA *LOC146880*, regulates binding between miR-539-5p and *LOC146880*, which modulates phosphorylation of enolase 1 (ENO1) and subsequent phosphorylation of phosphoinositide 3-kinase (PI3K) and Akt, and is associated with NSCLC susceptibility in the Chinese population.[30] The risk T allele of rs12740674, located in the enhancer of miR-1262, reduces the expression level of miR-1262 in lung tissue through chromosomal looping and increases the expression levels of UNC-51-like kinase 1 (ULK1) and RAB3D, member RAS oncogene family, promoting lung cancer cell proliferation (Table 1).[31]

SNVs and lung cancer susceptibility

Over the past decade, several genome-wide association studies (GWAS) focused on cancer susceptibility have been performed. To date, 51 lung cancer-associated SNP loci have been identified,[32] and a substantial proportion of these loci are specific to different subgroups in terms of histological subtype, smoking status, and ancestry.[33]

SNVs in different lung cancer subgroups

Genomic heterogeneity is associated with different histopathological types of lung cancer. A study by Dai et al[34] identified 19 SNP loci that were significantly associated with NSCLC risk in a Chinese population. Among these variants, rs17038564 ($P = 1.87 \times 10^{-8}$), rs35201538 ($P = 1.99 \times 10^{-8}$), and rs77468143 ($P = 7.48 \times 10^{-12}$) were significant in

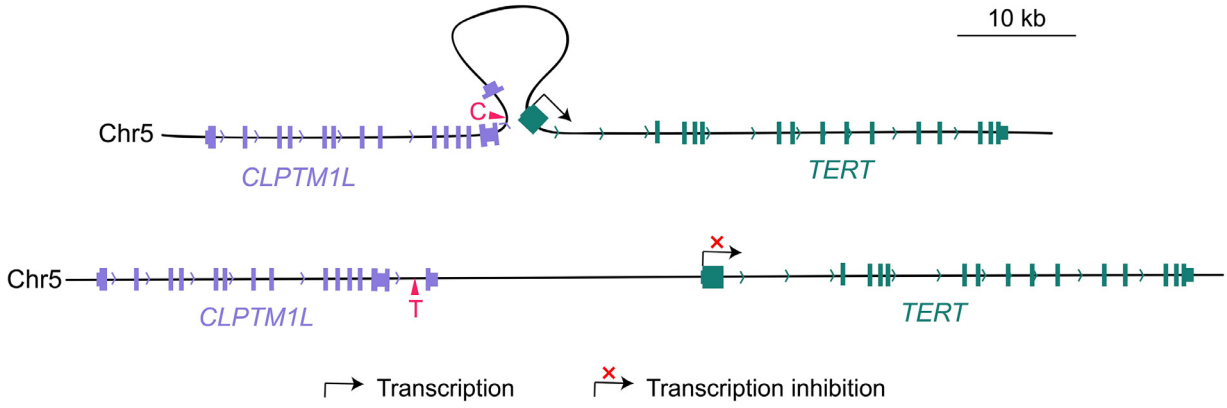

Fig. 1. Schematic representation of the mechanism by non-coding SNV rs402710, which is located in the last intron of *CLPTM1L*, and regulates *TERT* via physical interactions with the *TERT* promoter. SNV: Single nucleotide variant.

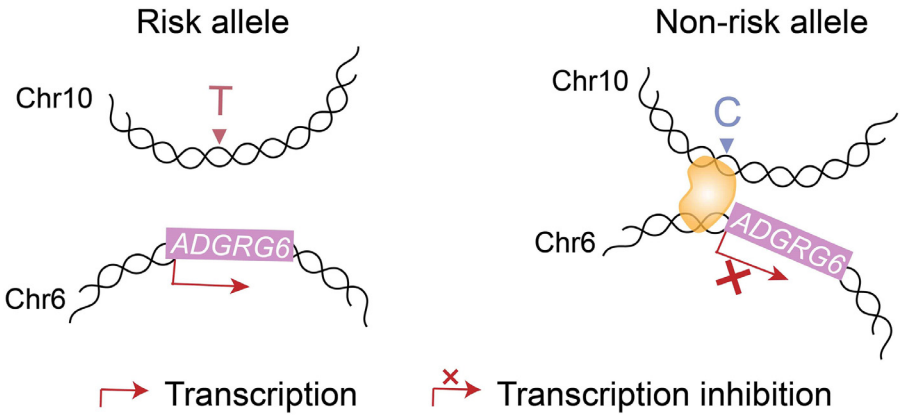

Fig. 2. Schematic representation of the mechanism of the intergenic SNV rs1663689 in regulating *ADGRG6* by interchromosomal interaction. SNV: Single nucleotidevariant.

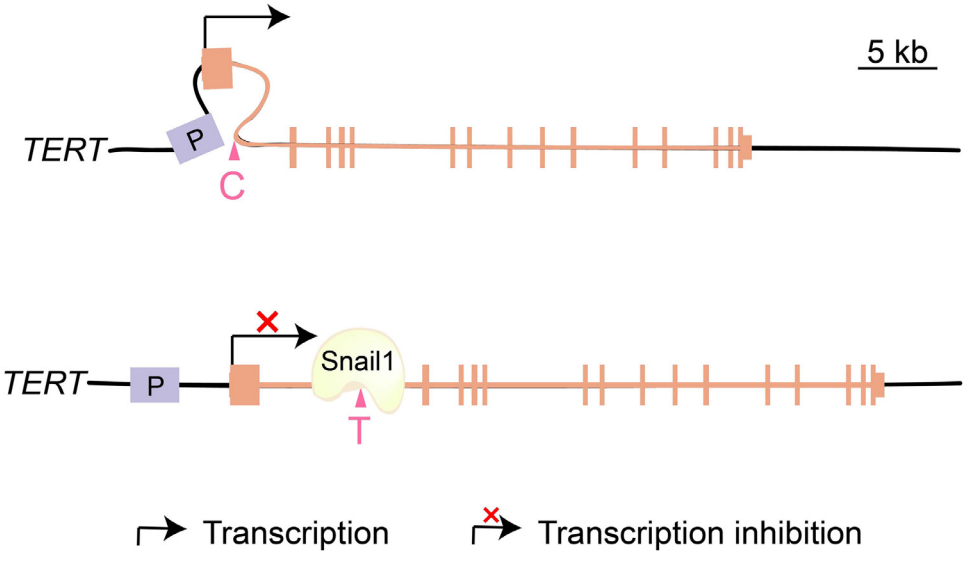

Fig. 3. Schematic representation of rs2853677 changing the binding affinity of the Snail1 transcription factor.

Table 1
Mechanisms of non-coding SNVs.

Mechanism	SNVs
Alter transcriptional activity of promoters by cis- or trans-effects.	rs402710[20]; rs1663689[21]
Influence the binding affinity of transcription factors.	rs2853677[22]; rs17079281[23]; rs9399451[24]
Influence the binding affinity of ncRNAs.	rs11614913[28]; rs140618127[30]

ncRNAs: Non-coding RNAs; SNVs: Single nucleotide variants.

Table 2
SNVs and lung cancer susceptibility in different subgroups.

SNVs	Region	Gene	Disease	Effect allele	Reference allele	OR (95% CI)	*P* value	References
rs17038564	2p14	*AFTPH*	LUAD	G	A	1.15 (1.10–1.21)	1.87×10^{-8}	34
rs35201538	9p13.3	*AQP3*	LUAD	C	T	1.10 (1.06–1.13)	1.99×10^{-8}	
rs77468143	15q21.1	*SECISBP2L*	LUAD	T	G	1.14 (1.10–1.18)	7.48×10^{-12}	
rs4573350	9q33.2	*DAB21P*	LUSC	T	C	1.13 (1.09–1.18)	3.23×10^{-9}	
rs77468143	15q21.1	*SECISBP2L*	LUAD	G	T	0.90 (0.83–0.89)	1.7×10^{-16}	35
rs13080835	3q28	*TP63*	LUAD	T	G	0.90 (0.87–0.92)	7.5×10^{-12}	
rs7705526	5p15.33	*TERT*	LUAD	A	C	1.30 (1.21–1.29)	3.8×10^{-35}	
rs4236709	8p12	*NRG1*	LUAD	G	A	1.10 (1.09–1.18)	1.3×10^{-10}	
rs885518	9p21.3	*CDNK2A*	LUAD	G	A	1.20 (1.11–1.23)	9.96×10^{-10}	
rs11591710	10q24.3	*OBFC1*	LUAD	C	A	1.20 (1.11–1.22)	6.3×10^{-11}	
rs1056562	11q23.3	*AMICA1*	LUAD	T	C	1.10 (1.07–1.14)	2.8×10^{-10}	
rs41309931	20q13.33	*RTEL1*	LUAD	T	G	1.20 (1.11–1.23)	1.3×10^{-9}	
rs116822326	6p21.33	*MHC*	LUSC	G	A	1.30 (1.19–1.32)	3.8×10^{-9}	
rs7953330	12p13.33	*RAD52*	LUSC	C	G	0.90 (0.83–0.90)	7.3×10^{-13}	
rs17879961	22q12.1	*CHEK2*	LUSC	G	A	0.40 (0.32–0.52)	5.7×10^{-13}	
rs3134615	1p34.2	*MYCL1*	SCLC	T	G	2.08 (1.39–3.12)	0.0004	36

CI: Confidence interval; LUAD: Lung adenocarcinoma; LUSC: Lung squamous cell carcinoma; OR: Odds (log-additive) ratio; SCLC: Small cell lung carcinoma; SNVs: Single nucleotide variants.

the LUAD subgroup, whereas rs4573350 was specific for SCC. SNPs can be verified through different cohorts. Amos et al[35] confirmed the association of rs77468143 with LUADs in a European population. The authors also identified additional seven SNPs (rs13080835, rs7705526, rs4236709, rs885518, rs11591710, rs1056562, and rs41309931) that were associated with LUAD and three SNPs (rs116822326, rs7953330, and rs17879961) that were associated with SCC.[35] rs3134615, located in the 3′UTR of *MYCL1*, is associated with an increased risk of SCLC.[36] The identification of subtype-specific associations of genetic variants indicates that the genetic architecture of lung cancer varies markedly among LUAD, SCC, and SCLC (Table 2).

SNVs and smoking status

Smoking is the main cause of lung cancer, especially LUAD and SCLC, and nicotine is the most addictive component in tobacco. An analysis by Gabriel et al[37] demonstrated that genetic variants influence the smoking behavior of individuals, which in turn influences their carcinogenic exposure and, consequently, their somatic mutation burden. The 15q25 susceptibility region, which contains six coding genes, including three cholinergic nicotine receptor genes (*CHRNA3*, *CHRNA5*, and *CHRNB4*) that exhibit independent effects on smoking behavior, contains multiple SNVs that are strongly associated with lung cancer.[38–42] The most robust lung cancer-associated SNP in 15q25 is rs16969968, which results in an amino acid change from aspartate to asparagine at position 398 of the nicotinic receptor α5 subunit protein sequence. rs16969968 predicts delayed smoking cessation and earlier age of lung cancer diagnosis.[43–45] rs9439519 and rs4809957 are associated with cigarette smoking. rs4809957 interacts with smoking dose to contribute to lung cancer risk.[46] rs6441286, rs17723637, and rs4751674 stratify lung cancer risk by smoking behavior. rs6441286 and rs17723637 variants increase the risk for lung cancer in ever-smokers, whereas the rs4751674 variant has a protective effect in ever-smokers compared with never-smokers.[47] rs910083, located in the intron of *DNMT3B*, is associated with an increased risk of nicotine dependence. In International Lung Cancer Consortium data, the C allele at rs910083 was found to increase the risk of squamous cell lung carcinoma.[48] rs34211819, located in an intron region of tensin-3 (*TNS3*), and rs1143149, located in an intron region of septin 7 (*SEPT7*), are significantly associated with the survival of NSCLC patients who are long-term former smokers. Both SNPs have significant interaction effects with years of smoking cessation. As the years of smoking cessation increase in long-term former smokers, the protective effect of rs34211819 and the detrimental effect of rs1143149 on survival are both enhanced.[49]

While the primary cause of lung cancer is smoking, approximately 25% of lung cancers worldwide occur in never-smokers.[50] These cancers are caused by environmental risk factors, such as passive smoking, hormonal factors, and, most importantly, genetic factors.[51–53] A study using WES and RNA-sequencing data for never-smoker LUADs sequenced by The Cancer Genome Atlas (TCGA) and Clinical Proteomic Tumor Analysis Consortium (CPTAC) found that pathogenic germline variants in cancer predisposition genes such as *BRCA1*, *BRCA2*, *FANCG*, *FANCM*, *HMBS*, *MSH6*, *NF1*, *POLD1*, *TMEM127*, and *WRN* are exclusively associated with lung cancer in never-smokers.[54] A GWAS of lung cancer in never-smoking females in Asia identified three lung cancer susceptibility loci, at 10q25.2 (rs7086803), 6q22.2 (rs9387478), and 6p21.32 (rs2395185), with no evidence of an association of 15q25 with lung cancer.[55] The *TERT* and *CLPTM1L* genes are located in 5p15. *TERT* is an established telomere maintenance locus. rs10936599 variants, which are located in the *TERT* coding region and associated with telomere length, were robustly associated with increased lung cancer risk among never-smoking women in Asia.[56] *CLPTM1L* was identified by screening for cisplatin (CDDP) resistance-related genes and was found to induce apoptosis in CDDP-sensitive cells. rs402710 is located in intron 4 of the *CLPTM1L* gene and is associated with lung cancer susceptibility.[41,57]

Estrogen receptor (*ER*) gene SNPs such as rs7753153 and rs985192, located in *ESR1*, and rs3020450, located in *ESR2*, are associated with LUAD risk in never-smoking women.[58] Moreover, rs12233719, a sex hormone regulation-related SNP in *UGT2B7*, is associated with NSCLC risk among never-smoking Chinese women.[59] These studies suggest the important role of sex hormones in regulating lung tumorigenesis in never-smoking populations. rs11080466 and rs11663246, located in the intron of *PIEZO2*, show a significant association with NSCLC susceptibility of never smokers in Korean populations.[60] rs4648127, located in the *NFKB1* gene, was associated with lung cancer in the screening arm of the Prostate, Lung, Colon, and Ovarian (PLCO) Cancer Screening Trial.[61]

The genetic susceptibility differences in histopathological types and in accordance with smoking status may implicate distinct biological mechanisms of lung cancer and therapeutic strategies (Table 3).

Interplay between SNVs and somatic mutations in regulating lung cancer susceptibility

Most of the mutations detected clinically are somatic mutations. Germline mutations can co-occur or be mutually exclusive with somatic cancer gene alterations.[62] For example, the susceptibility variant rs36600 is associated with somatic mutations within *ARID1A*; the susceptibility variants rs2395185 and rs3817963 are associated with somatic alterations in the cell cycle pathway; and rs3817963 is associated with somatic alterations in the MAPK signaling pathway.[63] rs2395185

Table 3
SNVs and lung cancer susceptibility with different smoking status.

SNVs	Region	Gene	Effect allele	Reference allele	OR (95% CI)	*P* value	References
Smoking							
rs16969968	15q25	*CHRNA5*	A	G	1.40	5.47×10^{-40}	44
rs9439519	1p36.32	*AJAP1-NPHP4*	G	A	1.11 (1.06–1.16)	3.65×10^{-6}	46
rs4809957	20q13.2	*CYP24A1*	T	C	1.13 (1.08–1.18)	1.20×10^{-8}	
rs6441286	3q25.33	*IL12A-AS1*	C	A	1.24 (1.13–1.37)	1.16×10^{-5}	47
rs17723637	9q31.2	*ZNF462*	G	A	1.36 (1.19–1.56)	1.06×10^{-5}	
rs4751674	10q25.3	*AFAP1L2*	C	T	0.58 (0.46–0.74)	1.07×10^{-5}	
rs910083	20q11	*DNMT3B*	C	A	1.06 (1.04–1.07)	3.70×10^{-8}	48
rs34211819	7p12.3	*TNS3*	C	A	0.73 (0.66–0.81)	3.90×10^{-9}	49
rs1143149	7p14.2	*SEPT7*	C	G	1.36 (1.22–1.51)	9.75×10^{-9}	
Never-smoking							
rs7086803	10q25.2	*VTI1A*	A	G	1.28 (1.21–1.35)	3.54×10^{-18}	55
rs9387478	6q22.2	*ROS1-DCBLD1*	A	C	0.85 (0.81–0.90)	4.14×10^{-10}	
rs2395185	6p21.32	*HLA*	T	G	1.17 (1.11–1.23)	9.51×10^{-9}	
rs10936599	3q26.2	*TERC*	C	T	0.10 (0.08–0.11)	2.54×10^{-31}	56
rs402710	5p15.33	*CLPTM1L*	T	C	1.14 (1.09–1.19)	5.00×10^{-8}	41
rs7753153	6q25	*ESR1*	A	G	1.51 (1.17–1.95)	0.0008	58
rs985192	6q25	*ESR1*	C	A	1.31 (1.00–1.71)	0.0233	
rs3020450	14q23	*ESR2*	A	C	2.11 (1.01–4.42)	0.0098	
rs12233719	4q13.2	*UGT2B7*	T	G	1.54 (1.21–1.96)	0.006	59
rs11080466	18p11.22	*PIEZO2*	C	T	0.66 (0.56–0.78)	1.19×10^{-6}	60
rs11663246	18p11.22	*PIEZO2*	T	C	0.67 (0.57–0.79)	2.40×10^{-6}	
rs4648127	4q24	*NFKB1*	T	C	0.56 (0.37–0.86)	0.02	61

CI: Confidence interval; OR: Odds (log-additive) ratio; SNVs: Single nucleotide variants.

is associated with elevated *APOBEC* mutagenesis.[64] Some lung cancer risk-related SNPs were shown to influence genetic damage in coke oven workers exposed to polycyclic aromatic hydrocarbons (PAHs). Some SNP loci (rs1333040, rs1663689, and rs3813572) are associated with decreased micronuclei frequency, which is a biomarker of chromosomal damage, genome instability, and cancer risk that associates acquired mutations with genetic susceptibility.[65] Recently, Peng et al[66] identified 111 pathogenic or likely pathogenic (P/LP) germline mutations in 35 cancer genes in 106 of 1794 Chinese patients (5.91%). Chinese patients with germline mutations show different prevalence rates of somatic *KRAS* mutation, *MET* exon 14 skipping, and *TP53* mutations compared with those without germline mutations.

SNVs associated with lung cancer outcome and drug sensitivity

In recent years, significant efforts have been dedicated to investigate the biological mechanisms underlying the association of SNVs with lung cancer outcomes and the clinical implications. The main goal is to translate these discoveries into clinical application. rs3743073-G in the *CHRNA3* gene is significantly associated with short survival among patients with advanced stage NSCLC.[67] rs942190-G and rs2401863-A located in *TDP1* are associated with relatively poor survival among SCLC patients.[68] Four SNPs (rs2107561, rs6882451, rs1826692, and rs6595026) modulate overall survival of lung cancer, and rs2107561, an intron variant of *PTPRG*, exhibits the strongest association.[69] rs5030740 in *RPA1* and rs1776148 and rs1047840 in *EXO1* are associated with disease-free survival and overall survival in lung cancer patients receiving platinum-based chemotherapy. Patients with the C allele of rs5030740 are regarded as protective allele of the prolonged progression-free survival. Patients with the A/A or A/G genotype of rs1776148 and the A/A genotype of rs1047840 have longer overall survival than G/G genotype of rs1776148 and A/G or G/G genotype of rs1047840.[70]

In addition to predicting prognosis, SNPs can be used in determining therapeutic strategies. The therapeutic efficacy suffers from large patient variability. Genetic variants often alter the sensitivity to the treatments in clinical practice. For example, rs712829 (216G/T) and rs4644 (191C/A) in *EGFR* are predictive of sensitivity to gefitinib. Mechanically, rs712829 is located in the binding site for the transcription factor Sp1. The T allele promotes Sp1 binding, enhances *EGFR* transcription, and increases sensitivity to gefitinib. rs4644 is located in the transcriptional start site of the *EGFR* promoter. The A allele increases promoter activity and protein expression and therefore increases sensitivity to gefitinib.[71] rs2231142 (421C/A) in *ABCG2* has been correlated with drug transport. The A allele reduces TKI transport and increases the accumulation of gefitinib, which results in adverse effects.[71] Patients with the *H19* rs2839698 A allele have a smaller chance of response to platinum-based chemotherapy.[72] rs1052566 (A273V) in *BRMS1v2* is associated with aggressive tumors in LUAD. The A allele of rs1052566 increases c-fos, thereby upregulating *CEACAM6*, which drives metastasis. T5224, a c-fos pharmacologic inhibitor, suppresses metastases in mice bearing A/A tumors.[73] rs1663689 A enhances ADGRG6 expression, which elevates the downstream cyclic adenosine monophosphate (cAMP)-protein kinase A (PKA) signaling. rs1663689 A/A tumors are more sensitive to the PKA inhibitor H89 than the rs1663689 C/C tumors.[21] The A allele of rs16906252 in *MGMT* is associated with increased *MGMT* methylation and lower MGMT expression. MGMT can reduce the tumor response to temozolomide. Thus, lung cancer patients with rs16906252-A may benefit from temozolomide treatment.[74] Although the clinical treatment strategies based on germline variants have not been implemented at present, these studies demonstrate the utility for SNVs in predicting drug sensitivity of tumors, highlighting their important role in precision medicine.

Outlook

SNVs regulate cellular behavior and subsequent disease phenotypes; therefore, SNVs can be used to select the appropriate therapeutic strategies. To achieve this goal, understanding the causal function of SNVs is needed. SNVs reside in different regions and function through different mechanisms. SNVs in coding regions change the sequence of amino acids, and subsequently change the structure and biological function of relative proteins. However, over 90% of disease-associated SNVs are located in non-coding regions of the genome, often at considerable genomic distances from annotated genes. These non-coding SNVs either change the sequences of noncoding RNAs or change the binding affinity of transcription factors and thereby posttranscriptionally regulate or transcriptionally regulate their downstream target genes (Fig. 4). Although some well-characterized non-coding SNVs regulate their neighboring genes, assignment based on linear proximity is error prone, as many cis-regulatory elements map large distances away from their targets, bypassing the nearest gene, which makes identifying

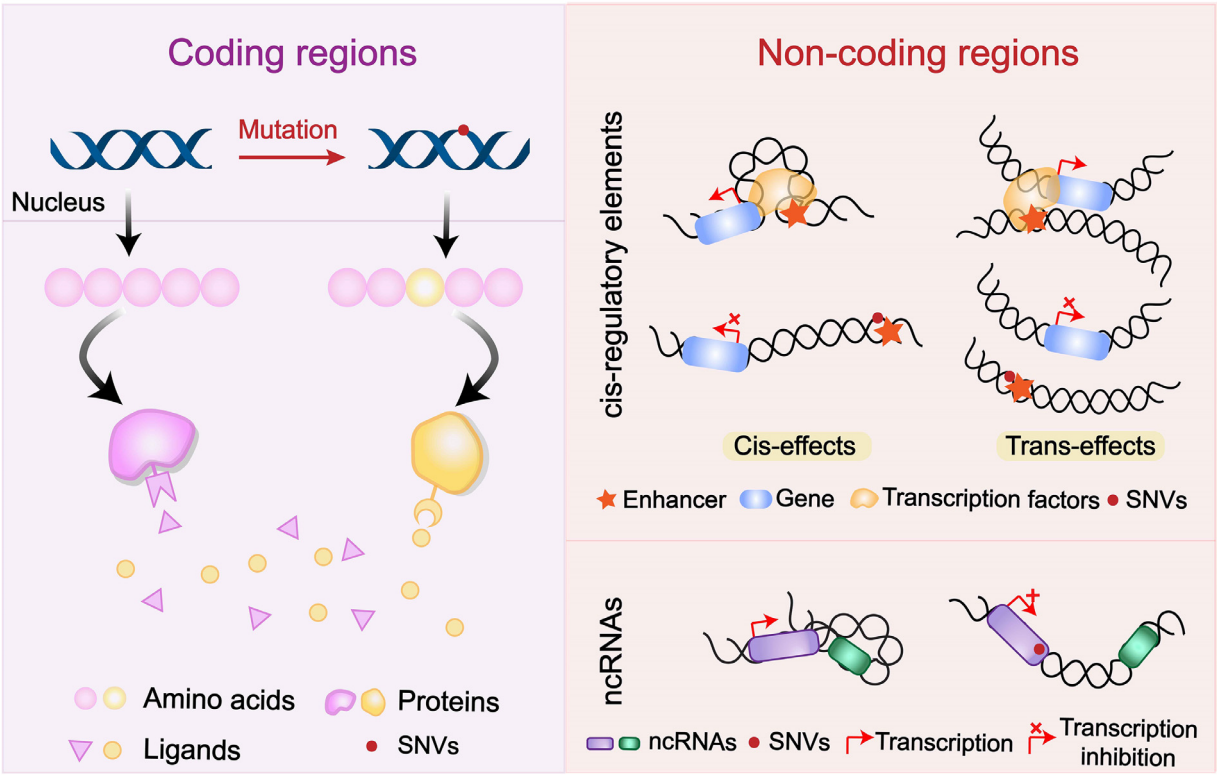

Fig. 4. Schematic representation of the mechanism of the SNVs located in different regions. ncRNAs: Non-coding RNAs; SNVs: Single nucleotide variants.

their downstream genes problematic. Therefore, the functions of most lung cancer-associated non-coding SNVs remain unknown. Developing efficient strategies to decipher the regulatory pathways for non-coding SNVs is needed. Given that long-range regulation requires direct physical interactions in eukaryotes, genomic screening for non-coding SNV-interacting genes will serve as a strategy to identify target genes.

Funding

This work was supported by grants from the National Natural Science Foundation of China (Nos. 81825017 and 81773034 to Z.L.), and the Interdisciplinary Research Project of Hangzhou Normal University (No. 2024JCXK03 to Z. L.).

Declaration of competing interest

The authors declare there is no conflict of interests.

References

1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. *CA Cancer J Clin*. 2022;72:7–33. doi:10.3322/caac.21708.
2. Zhang Y, Wang DC, Shi L, Zhu B, Min Z, Jin J. Genome analyses identify the genetic modification of lung cancer subtypes. *Semin Cancer Biol*. 2017;42:20–30. doi:10.1016/j.semcancer.2016.11.005.
3. Shigematsu H, Lin L, Takahashi T, et al. Clinical and biological features associated with epidermal growth factor receptor gene mutations in lung cancers. *J Natl Cancer Inst*. 2005;97:339–346. doi:10.1093/jnci/dji055.
4. Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. *Nature*. 2012;489:519–525. doi:10.1038/nature11404.
5. Jackson M, Marks L, May GHW, Wilson JB. The genetic basis of disease. *Essays Biochem*. 2018;62:643–723. doi:10.1042/EBC20170053.
6. Maurano MT, Humbert R, Rynes E, et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science*. 2012;337:1190–1195. doi:10.1126/science.1222794.
7. Lu S, Yu Y, Li Z, et al. EGFR and ERBB2 germline mutations in Chinese lung cancer patients and their roles in genetic susceptibility to cancer. *J Thorac Oncol*. 2019;14:732–736. doi:10.1016/j.jtho.2018.12.006.
8. Shukuya T, Takahashi K. Germline mutations in lung cancer. *Respir Investig*. 2019;57:201–206. doi:10.1016/j.resinv.2018.12.005.
9. Wu WJ, Yang SH, Chung HP, et al. EGFR Q787Q polymorphism is a germline variant and a prognostic factor for lung cancer treated with TKIs. *Front Oncol*. 2022;12:816801. doi:10.3389/fonc.2022.816801.
10. Dunnett-Kane V, Burkitt-Wright E, Blackhall FH, Malliri A, Evans DG, Lindsay CR. Germline and sporadic cancers driven by the RAS pathway: parallels and contrasts. *Ann Oncol*. 2020;31:873–883. doi:10.1016/j.annonc.2020.03.291.
11. Gremer L, Merbitz-Zahradnik T, Dvorsky R, et al. Germline KRAS mutations cause aberrant biochemical and physical properties leading to developmental disorders. *Hum Mutat*. 2011;32:33–43. doi:10.1002/humu.21377.
12. Wang Z, Sun Y, Gao B, et al. Two co-existing germline mutations P53 V157D and PMS2 R20Q promote tumorigenesis in a familial cancer syndrome. *Cancer Lett*. 2014;342:36–42. doi:10.1016/j.canlet.2013.08.032.
13. Papadakis ED, Soulitzis N, Spandidos DA. Association of p53 codon 72 polymorphism with advanced lung cancer: The Arg allele is preferentially retained in tumours arising in Arg/Pro germline heterozygotes. *Br J Cancer*. 2002;87:1013–1018. doi:10.1038/sj.bjc.6600595.
14. Rudd MF, Webb EL, Matakidou A, et al. Variants in the GH-IGF axis confer susceptibility to lung cancer. *Genome Res*. 2006;16:693–701. doi:10.1101/gr.5120106.
15. Liu Y, Xia J, McKay J, et al. Rare deleterious germline variants and risk of lung cancer. *NPJ Precis Oncol*. 2021;5:12. doi:10.1038/s41698-021-00146-7.
16. Zhang ZL, Wang NN, Ma QL, et al. Somatic and germline mutations in the tumor suppressor gene PARK2 impair PINK1/Parkin-mediated mitophagy in lung cancer cells. *Acta Pharmacol Sin*. 2020;41:93–100. doi:10.1038/s41401-019-0260-6.
17. Tlemsani C, Takahashi N, Pongor L, et al. Whole-exome sequencing reveals germline-mutated small cell lung cancer subtype with favorable response to DNA repair-targeted therapies. *Sci Transl Med*. 2021;13:eabc7488. doi:10.1126/scitranslmed.abc7488.
18. Feng Y, Wang Y, Liu H, et al. Novel genetic variants in the P38MAPK pathway gene ZAK and susceptibility to lung cancer. *Mol Carcinog*. 2018;57:216–224. doi:10.1002/mc.22748.
19. Wang Y, Broderick P, Webb E, et al. Common 5p15.33 and 6p21.33 variants influence lung cancer risk. *Nat Genet*. 2008;40:1407–1409. doi:10.1038/ng.273.
20. Yang YC, Fu WP, Zhang J, Zhong L, Cai SX, Sun C. rs401681 and rs402710 confer lung cancer susceptibility by regulating TERT expression instead of CLPTM1L in East Asian populations. *Carcinogenesis*. 2018;39:1216–1221. doi:10.1093/carcin/bgy084.
21. Lei X, Tian X, Wang H, et al. Noncoding SNP at rs1663689 represses ADGRG6 via interchromosomal interaction and reduces lung cancer progression. *EMBO Rep*. 2023;24:e56212. doi:10.15252/embr.202256212.
22. Li X, Xu X, Fang J, et al. Rs2853677 modulates Snail1 binding to the TERT enhancer and affects lung adenocarcinoma susceptibility. *Oncotarget*. 2016;7:37825–37838. doi:10.18632/oncotarget.9339.
23. Wang Y, Ma R, Liu B, et al. SNP rs17079281 decreases lung cancer risk through creating an YY1-binding site to suppress DCBLD1 expression. *Oncogene*. 2020;39:4092–4102. doi:10.1038/s41388-020-1278-4.

24. Shi Q, Shi QN, Xu JW, et al. rs9390123 and rs9399451 influence the DNA repair capacity of lung cancer by regulating PEX3 and PHACTR2-AS1 expression instead of PHACTR2. *Oncol Rep*. 2022;47:59. doi:10.3892/or.2022.8270.
25. Xie W, Wang Y, Zhang Y, et al. Single-nucleotide polymorphism rs4142441 and MYC co-modulated long non-coding RNA OSER1-AS1 suppresses non-small cell lung cancer by sequestering ELAVL1. *Cancer Sci*. 2021;112:2272–2286. doi:10.1111/cas.14713.
26. Shao L, Zuo X, Yang Y, et al. The inherited variations of a p53-responsive enhancer in 13q12.12 confer lung cancer risk by attenuating TNFRSF19 expression. *Genome Biol*. 2019;20:103. doi:10.1186/s13059-019-1696-1.
27. Yan H, Bu P. Non-coding RNA in cancer. *Essays Biochem*. 2021;65:625–639. doi:10.1042/EBC20200032.
28. Hu Z, Chen J, Tian T, et al. Genetic variants of miRNA sequences and non-small cell lung cancer survival. *J Clin Invest*. 2008;118:2600–2608. doi:10.1172/JCI34934.
29. Hu L, Chen SH, Lv QL, et al. Clinical significance of long non-coding RNA CASC8 rs10505477 polymorphism in lung cancer susceptibility, platinum-based chemotherapy response, and toxicity. *Int J Environ Res Public Health*. 2016;13:545. doi:10.3390/ijerph13060545.
30. Feng T, Feng N, Zhu T, et al. A SNP-mediated lncRNA (LOC146880) and microRNA (miR-539-5p) interaction and its potential impact on the NSCLC risk. *J Exp Clin Cancer Res*. 2020;39:157. doi:10.1186/s13046-020-01652-5.
31. Xie K, Chen M, Zhu M, et al. A polymorphism in miR-1262 regulatory region confers the risk of lung cancer in Chinese population. *Int J Cancer*. 2017;141:958–966. doi:10.1002/ijc.30788.
32. Wang C, Dai J, Qin N, et al. Analyses of rare predisposing variants of lung cancer in 6004 whole genomes in Chinese. *Cancer Cell*. 2022;40:1223.e–1239.e. doi:10.1016/j.ccell.2022.08.013.
33. Long E, Patel H, Byun J, Amos CI, Choi J. Functional studies of lung cancer GWAS beyond association. *Hum Mol Genet*. 2022;31:R22–R36. doi:10.1093/hmg/ddac140.
34. Dai J, Lv J, Zhu M, et al. Identification of risk loci and a polygenic risk score for lung cancer: a large-scale prospective cohort study in Chinese populations. *Lancet Respir Med*. 2019;7:881–891. doi:10.1016/S2213-2600(19)30144-4.
35. McKay JD, Hung RJ, Han Y, et al. Large-scale association analysis identifies new lung cancer susceptibility loci and heterogeneity in genetic susceptibility across histological subtypes. *Nat Genet*. 2017;49:1126–1132. doi:10.1038/ng.3892.
36. Xiong F, Wu C, Chang J, et al. Genetic variation in an miRNA-1827 binding site in MYCL1 alters susceptibility to small-cell lung cancer. *Cancer Res*. 2011;71:5175–5181. doi:10.1158/0008-5472.CAN-10-4407.
37. Gabriel AAG, Atkins JR, Penha RCC, et al. Genetic analysis of lung cancer and the germline impact on somatic mutation burden. *J Natl Cancer Inst*. 2022;114:1159–1166. doi:10.1093/jnci/djac087.
38. David SP, Wang A, Kapphahn K, et al. Gene by environment investigation of incident lung cancer risk in African-Americans. *EBioMedicine*. 2016;4:153–161. doi:10.1016/j.ebiom.2016.01.002.
39. Bae EY, Lee SY, Kang BK, et al. Replication of results of genome-wide association studies on lung cancer susceptibility loci in a Korean population. *Respirology*. 2012;17:699–706. doi:10.1111/j.1440-1843.2012.02165.x.
40. Broderick P, Wang Y, Vijayakrishnan J, et al. Deciphering the impact of common genetic variation on lung cancer risk: a genome-wide association study. *Cancer Res*. 2009;69:6633–6641. doi:10.1158/0008-5472.CAN-09-0680.
41. Truong T, Hung RJ, Amos CI, et al. Replication of lung cancer susceptibility loci at chromosomes 15q25, 5p15, and 6p21: A pooled analysis from the International lung cancer consortium. *J Natl Cancer Inst*. 2010;102:959–971. doi:10.1093/jnci/djq178.
42. Wang Y, Broderick P, Matakidou A, Eisen T, Houlston RS. Role of 5p15.33 (TERT-CLPTM1L), 6p21.33 and 15q25.1 (CHRNA5-CHRNA3) variation and lung cancer risk in never-smokers. *Carcinogenesis*. 2010;31:234–238. doi:10.1093/carcin/bgp287.
43. Hung RJ, McKay JD, Gaborieau V, et al. A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. *Nature*. 2008;452:633–637. doi:10.1038/nature06885.
44. Saccone NL, Culverhouse RC, Schwantes-An TH, et al. Multiple independent loci at chromosome 15q25.1 affect smoking quantity: a meta-analysis and comparison with lung cancer and COPD. *PLoS Genet*. 2010;6:e1001053. doi:10.1371/journal.pgen.1001053.
45. Chen LS, Hung RJ, Baker T, et al. CHRNA5 risk variant predicts delayed smoking cessation and earlier lung cancer diagnosis – a meta-analysis. *J Natl Cancer Inst*. 2015;107:djv100. doi:10.1093/jnci/djv100.
46. Dong J, Hu Z, Wu C, et al. Association analyses identify multiple new lung cancer susceptibility loci and their interactions with smoking in the Chinese population. *Nat Genet*. 2012;44:895–899. doi:10.1038/ng.2351.
47. Li Y, Xiao X, Han Y, et al. Genome-wide interaction study of smoking behavior and non-small cell lung cancer risk in Caucasian population. *Carcinogenesis*. 2018;39:336–346. doi:10.1093/carcin/bgx113.
48. Hancock DB, Guo Y, Reginsson GW, et al. Genome-wide association study across European and African American ancestries identifies a SNP in DNMT3B contributing to nicotine dependence. *Mol Psychiatry*. 2018;23:1911–1919. doi:10.1038/mp.2017.193.
49. Shen S, Wei Y, Li Y, et al. A multi-omics study links TNS3 and SEPT7 to long-term former smoking NSCLC survival. *NPJ Precis Oncol*. 2021;5:39. doi:10.1038/s41698-021-00182-3.
50. Sun S, Schiller JH, Gazdar AF. Lung cancer in never smokers – a different disease. *Nat Rev Cancer*. 2007;7:778–790. doi:10.1038/nrc2190.
51. Corrales L, Rosell R, Cardona AF, Martin C, Zatarain-Barron ZL, Arrieta O. Lung cancer in never smokers: the role of different risk factors other than tobacco smoking. *Crit Rev Oncol Hematol*. 2020;148:102895. doi:10.1016/j.critrevonc.2020.102895.
52. Benusiglio PR, Fallet V, Sanchis-Borja M, Coulet F, Cadranel J. Lung cancer is also a hereditary disease. *Eur Respir Rev*. 2021;30:210045. doi:10.1183/16000617.0045-2021.
53. Malhotra J, Malvezzi M, Negri E, La Vecchia C, Boffetta P. Risk factors for lung cancer worldwide. *Eur Respir J*. 2016;48:889–902. doi:10.1183/13993003.00359-2016.
54. Devarakonda S, Li Y, Martins Rodrigues F, et al. Genomic profiling of lung adenocarcinoma in never-smokers. *J Clin Oncol*. 2021;39:3747–3758. doi:10.1200/jco.21.01691.
55. Lan Q, Hsiung CA, Matsuo K, et al. Genome-wide association analysis identifies new lung cancer susceptibility loci in never-smoking women in Asia. *Nat Genet*. 2012;44:1330–1335. doi:10.1038/ng.2456.
56. Machiela MJ, Hsiung CA, Shu XO, et al. Genetic variants associated with longer telomere length are associated with increased lung cancer risk among never-smoking women in Asia: a report from the female lung cancer consortium in Asia. *Int J Cancer*. 2015;137:311–319. doi:10.1002/ijc.29393.
57. Lu X, Ke J, Luo X, et al. The SNP rs402710 in 5p15.33 is associated with lung cancer risk: a replication study in Chinese population and a meta-analysis. *PLoS ONE*. 2013;8:e76252. doi:10.1371/journal.pone.0076252.
58. Chen KY, Hsiao CF, Chang GC, et al. Estrogen receptor gene polymorphisms and lung adenocarcinoma risk in never-smoking Women. *J Thorac Oncol*. 2015;10:1413–1420. doi:10.1097/JTO.0000000000000646.
59. Qian Y, Xie L, Li L, et al. Association between sex hormones regulation-related SNP rs12233719 and lung cancer risk among never-smoking Chinese women. *Cancer Med*. 2021;10:1880–1888. doi:10.1002/cam4.3772.
60. Ahn MJ, Won HH, Lee J, et al. The 18p11.22 locus is associated with never smoker non-small cell lung cancer susceptibility in Korean populations. *Hum Genet*. 2012;131:365–372. doi:10.1007/s00439-011-1080-z.
61. Shiels MS, Engels EA, Shi J, et al. Genetic variation in innate immunity and inflammation pathways associated with lung cancer risk. *Cancer*. 2012;118:5630–5636. doi:10.1002/cncr.27605.
62. Lu C, Xie M, Wendl MC, et al. Patterns and functional implications of rare germline variants across 12 cancer types. *Nat Commun*. 2015;6:10086. doi:10.1038/ncomms10086.
63. Shen H, Zhu M, Wang C. Precision oncology of lung cancer: genetic and genomic differences in Chinese population. *NPJ Precis Oncol*. 2019;3:14. doi:10.1038/s41698-019-0086-1.
64. Wang Y, Wang C, Zhang J, et al. Interaction analysis between germline susceptibility loci and somatic alterations in lung cancer. *Int J Cancer*. 2018;143:878–885. doi:10.1002/ijc.31351.
65. Dai X, Deng S, Wang T, et al. Associations between 25 lung cancer risk-related SNPs and polycyclic aromatic hydrocarbon-induced genetic damage in coke oven workers. *Cancer Epidemiol Biomarkers Prev*. 2014;23:986–996. doi:10.1158/1055-9965.EPI-13-1251.
66. Peng W, Li B, Li J, et al. Clinical and genomic features of Chinese lung cancer patients with germline mutations. *Nat Commun*. 2022;13:1268. doi:10.1038/s41467-022-28840-5.
67. Niu X, Chen Z, Shen S, et al. Association of the CHRNA3 locus with lung cancer risk and prognosis in Chinese Han population. *J Thorac Oncol*. 2010;5:658–666. doi:10.1097/JTO.0b013e3181d5e447.
68. Lohavanichbutr P, Sakoda LC, Amos CI, et al. Common TDP1 polymorphisms in relation to survival among small cell lung cancer patients: a multicenter study from the international lung cancer consortium. *Clin Cancer Res*. 2017;23:7550–7557. doi:10.1158/1078-0432.CCR-17-1401.
69. Galvan A, Colombo F, Frullanti E, et al. Germline polymorphisms and survival of lung adenocarcinoma patients: a genome-wide study in two European patient series. *Int J Cancer*. 2015;136:E262–E271. doi:10.1002/ijc.29195.
70. He J, Wang Z, Wang Y, Zou T, Li XP, Chen J. The roles of EXO1 and RPA1 polymorphisms in prognosis of lung cancer patients treated with platinum-based chemotherapy. *Dis Markers*. 2022;2022:3306189. doi:10.1155/2022/3306189.
71. Galvani E, Peters GJ, Giovannetti E. EGF receptor-targeted therapy in non-small-cell lung cancer: role of germline polymorphisms in outcome and toxicity. *Future Oncol*. 2012;8:1015–1029. doi:10.2217/fon.12.89.
72. Gong WJ, Yin JY, Li XP, et al. Association of well-characterized lung cancer lncRNA polymorphisms with lung cancer susceptibility and platinum-based chemotherapy response. *Tumour Biol*. 2016;37:8349–8358. doi:10.1007/s13277-015-4497-5.
73. Liu Y, Chudgar N, Mastrogiacomo B, et al. A germline SNP in BRMS1 predisposes patients with lung adenocarcinoma to metastasis and can be ameliorated by targeting c-fos. *Sci Transl Med*. 2022;14:eabo1050. doi:10.1126/scitranslmed.abo1050.
74. Leng S, Bernauer AM, Hong C, et al. The A/G allele of Rs16906252 predicts for MGMT methylation and is selectively silenced in premalignant lesions from smokers and in lung adenocarcinomas. *Clin Cancer Res*. 2011;17:2014–2023. doi:10.1158/1078-0432.Ccr-10-3026.