


# Functional studies of lung cancer GWAS beyond association

Erping Long<sup>1</sup>, Harsh Patel<sup>1</sup>, Jinyoung Byun<sup>2,3</sup>, Christopher I Amos<sup>2,3,4,\*</sup> and Jiyeon Choi <sup>1,\*</sup>

<sup>1</sup>Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, 20892, USA

<sup>2</sup>Institute for Clinical and Translational Research, Baylor College of Medicine, Houston, TX, 77030, USA

<sup>3</sup>Section of Epidemiology and Population Sciences, Department of Medicine, Baylor College of Medicine, Houston, TX, 77030, USA

<sup>4</sup>Dan L Duncan Comprehensive Cancer Center, Baylor College of Medicine, Houston, TX, 77030, USA

\*To whom correspondence should be addressed at: Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, National Cancer Institute, 9615 Medical center drive, Rockville, MD 20850, USA. Tel: +1 2407606474; Fax: +1 2405414442; Email: [jiyeon.choi2@nih.gov](mailto:jiyeon.choi2@nih.gov)

The authors wish it to be known that, in their opinion, the last two authors should be regarded as joint Corresponding authors.

## Abstract

Fourteen years after the first genome-wide association study (GWAS) of lung cancer was published, approximately 45 genomic loci have now been significantly associated with lung cancer risk. While functional characterization was performed for several of these loci, a comprehensive summary of the current molecular understanding of lung cancer risk has been lacking. Further, many novel computational and experimental tools now became available to accelerate the functional assessment of disease-associated variants, moving beyond locus-by-locus approaches. In this review, we first highlight the heterogeneity of lung cancer GWAS findings across histological subtypes, ancestries and smoking status, which poses unique challenges to follow-up studies. We then summarize the published lung cancer post-GWAS studies for each risk-associated locus to assess the current understanding of biological mechanisms beyond the initial statistical association. We further summarize strategies for GWAS functional follow-up studies considering cutting-edge functional genomics tools and providing a catalog of available resources relevant to lung cancer. Overall, we aim to highlight the importance of integrating computational and experimental approaches to draw biological insights from the lung cancer GWAS results beyond association.

## Introduction

Lung cancer is the second most common cancer and the leading cause of cancer death worldwide, accounting for 18% of the total cancer deaths (1). Although cigarette smoking is the major risk factor for lung cancer, up to 25% of lung cancer cases also arise in never-smokers (2,3). Further, the heritability of lung cancer has been estimated as ~8–20% by twin (4) and array-based studies (5–8), and rare high-penetrance genetic variants were identified for familial risk of lung cancer (e.g. EGFR T790M, TP53 mutations in Li–Fraumeni syndrome) (9–11), pointing to the genetic component of lung cancer susceptibility. For over a decade, lung cancer genome-wide association study (GWAS) successfully identified ~45 genomic loci (12–23). A substantial proportion of these loci were private to different subgroups in terms of histological subtypes, smoking status and ancestries, suggesting complex biological processes underlying diverse lung cancer subgroups. However, mechanistic understanding of the GWAS results in terms of what genes and functional variants explain the significant statistical association is often not straightforward for several reasons. First, the genetic variant with the strongest association *P*-value in a given locus (i.e. lead SNP) is not necessarily causal because

multiple co-inherited variants tied by linkage disequilibrium (LD) often present similar *P*-values that may not be distinguished from the margin of error in the current GWAS sample sizes. Second, the risk-associated variants primarily reside in nonprotein-coding regions and therefore the target genes of these variants may not be apparent by the location and proximity. Third, the effect of these variants on target genes might be detectable only in specific biological contexts. Despite these challenges, functional follow-up studies of individual lung cancer risk loci provided insights into underlying biology (24–31). Tapping into recent advances in functional genomics tools and resources, post-GWAS strategies are starting the transition to multi-loci approaches. Here we summarize the notable heterogeneity of lung cancer GWAS results, discuss published lung cancer post-GWAS functional studies, and provide available functional genomics tools and strategies relevant to lung cancer GWAS functional characterization.

## Heterogeneity of Lung Cancer GWAS Findings

### Histological subtypes

Lung cancer has two main histological subtypes—non-small cell lung cancer (NSCLC) and small cell lung cancer

(SCLC) (32). NSCLC accounts for ~80% of lung cancer cases and is further divided into multiple subtypes. Lung adenocarcinoma (LUAD) and squamous cell carcinoma (SQC) are the two most common subtypes of NSCLC. LUAD, SQC and SCLC (the third most common subtype) present significant differences in histological features and the cell types and locations from which they arise. LUAD arises from alveolar type II cells of the lung, while SQC arises from flat squamous cells of central airways. SCLC is characterized as a neuroendocrine carcinoma, which starts in the bronchi or central airways (32).

Based on the GWAS catalog (accessed July 15, 2021, searched by the term 'lung cancer' and filtered with genome-wide significance  $P \leq 5 \times 10^{-8}$ , [Supplementary Material, Table S1](#)) and a newly published cross-ancestry GWAS ([Supplementary Material, Table S2](#)), a total of 46 loci (by cytoband) were identified. Half of them (23 of 46) are either only significant ( $P < 5 \times 10^{-8}$  in one subtype) or mainly significant (i.e.  $P < 5 \times 10^{-8}$  in overall lung cancer or NSCLC as well but not in the other subtype) in LUAD (17 loci), SQC (5 loci) or SCLC (1 locus; [Supplementary Material, Table S3](#)). A comprehensive summary of the GWAS findings until 2018 was described before (29). Notably, three recent large-scale GWASs included subtype-stratified analyses mainly investigating LUAD and SQC and less frequently for SCLC. A study including 11 273 LUAD, 7426 SQC and 2664 SCLC cases in European populations identified 8 loci mainly significant in LUAD (1p31, 3q28, 8p12, 9p21, 10q24, 11q23, 15q21 and 20q13) and 3 in SQC (12p13, 13q13 and 22q12) (16). Another GWAS of 15 581 LUAD and 8350 SQC in Chinese and European populations identified 16 loci significant in LUAD, but only those at 5p15.33, 6p21–22 and 15q25.1 were also significant in SQC (23). One locus at 9q33.2 was unique to SQC. The third study including 27 301 LUAD, 13 389 SQC and 3779 SCLC in European, Asian and African populations identified 11 LUAD-, 4 SQC- and 1 SCLC-specific loci with heterogeneous effects (33). These findings together suggested that the genetic factors contributing to SCLC, LUAD and SQC risk are distinct and may indicate different underlying mechanisms.

## Ancestry

While a larger proportion of GWASs was performed in European populations, substantial discoveries were also made in Asian populations. Based on the GWAS catalog, a substantial proportion of the loci (25 of 46) are only significant in either European (17 loci) or East Asian (8 loci) populations ([Supplementary Material, Table S3](#)). Further, considerable heterogeneity was observed for commonly detected loci in both European and East Asian populations. For example, a recent study of 26 655 Chinese and 27 820 European individuals indicated that the effect sizes (i.e. odds ratio) of the lead SNPs in the loci at 3q28 and 5p15 were greater in Chinese populations than in European ones. In contrast, the opposite was true for the locus at 15q25.1 (23). Such heterogeneity can potentially be attributed to differences

in the allele frequency: the risk allele G (rs55781567) in the locus at 15q25.1 is ~38% in EUR (European populations in the 1000 Genomes Project) while it is ~3% in EAS (East Asian). Other factors (e.g. genetic interaction with environmental exposures and sample size differences across studies) could also result in discrepancies in GWAS results. Based on the largest-to-date cross-ancestry GWAS, including 911 288 individuals of European descent, 39 074 Asian descent and 6640 African descent (33), 44% (18 of 41 loci) of the identified loci showed significance in all three ancestries. Ancestry-specific signals were observed mainly in Europeans, likely due to imbalanced sample sizes, but EAS or AFR (African)-specific signals were also observed, including those for the rare variants.

It should be noted that lung cancer GWAS in non-European/East-Asian populations (e.g. African) is still rare. In a GWAS of 5339 African American individuals, two known loci originally discovered in European and Asian populations (15q25.1 and 5p15.33) reached genome-wide significance (34), but no new loci were identified. Notably, there is no GWAS reported among Hispanic/Latinx populations to date. Further studies with larger sample sizes in underrepresented populations are needed to dissect the heterogeneity across different ancestries.

## Smoking status

Approximately 25% of lung cancer diagnoses worldwide occur in lifelong never-smokers (2,3). Studies have shown considerable differences in characteristics between never- and ever-smokers. Never-smokers with lung cancer have been mainly female (35), mostly presented LUAD (36) and constituted a higher proportion in Asian populations compared with European or African populations (37). Besides, never-smokers developing lung cancer tend to have more frequent somatic EGFR mutations and show a better outcome after EGFR-inhibitor treatment than ever-smokers (36). These observations suggest that lung cancer among never-smokers presents distinct features in the etiology and carcinogenic pathways.

Multiple GWAS for lung cancer in never-smokers has been conducted. Lan and colleagues conducted a GWAS in never-smoking Asian women (5510 cases and 4544 controls), which confirmed the loci reported in the studies of mainly smokers (5p15, 3q28 and 17q24.3) and identified new loci (10q25.2, 6q22.1 and 6p21.32) (22). Notably, the smoking behavior-associated locus at 15q25.1 (the strongest signal in European, mainly smoker populations) was not detected in this population. The signals at 10q25.2 and 6q22.1 were later replicated in another study, including smokers and nonsmokers from Chinese and European populations (23). A GWAS with a larger sample size (6877 cases and 6277 controls) identified a locus at 12q13.13 that is specific to never-smoking Asian women (38) and two independent signals (6p21.1 and 9p21.3). A follow-up study stratifying the cases based on EGFR mutation status identified a locus only significant

in EGFR mutation carriers (6p21.3) and two loci displaying preferential association in EGFR mutation carriers versus non-carriers (6p21.32 and 6q22.2) (39). A recent never-smoker lung cancer GWAS in European populations (3636 cases and 6295 controls) replicated the association in a region at 5p15 and did not show any genome-wide significant signal at the smoking-associated locus in 15q25.1 (40).

Despite these efforts in identifying unique loci associated with lung cancer in never-smokers, it is hard to tease apart the effect of other factors such as ancestry, sex and main histology from the smoking effect with the current sample sizes.

### Post-GWA studies of lung cancer risk loci

Beyond significant association from the initial GWAS, statistical fine-mapping (to identify independent signals and prioritize the most plausible variants among multiple variants with similar *P*-values) and functional analyses provided clues to the biological implications of several lung cancer-associated loci. These loci can be potentially linked to pathways involving smoking behavior and nicotine metabolism, telomere biology, immune response, DNA damage response and repair, cellular stress response and cell cycle regulation. Notably, for many of these loci, a robust connection among GWAS signals, functional variants and target genes still needs to be made for a comprehensive understanding of the biological mechanisms.

### Smoking behavior and nicotine metabolism 15q25.1

The locus at 15q25.1 is the first and the strongest locus associated with lung cancer by GWAS in European populations (41). GWAS in African American (42) and East Asian (43) populations also reported this locus but with a weaker association or tagged by independent variants with higher effect allele frequencies than the European lead SNP, rs55781567. There was no indication of an association between lung cancer and 15q25.1 among never-smokers in both East Asian (22) and European populations (40).

This locus includes three nicotinic acetylcholine receptor (nAChR) subunit genes (*CHRNA5*, *CHRNA3* and *CHRNA4*), and variants on or near these genes are also associated with smoking status (44,45) and nicotine addiction (46). Although these variants are in high LD among Europeans (47), fine-mapping in African-Americans refined the association signal to the variants on or near *CHRNA5* including a missense variant, rs16969968 (D398N,  $R^2 = 0.9089$  with rs55781567; Fig. 1A) (42). Experimental data in human embryonic kidney (HEK293) cells (24) suggested that the smoking behavior-associated A allele (398N) decreases the response of nAChR to agonists (e.g. acetylcholine and nicotine) in the presence of high external calcium, which may lead to distinct downstream cellular signaling (Fig. 1B). An fMRI study showed that reduced connectivity in the

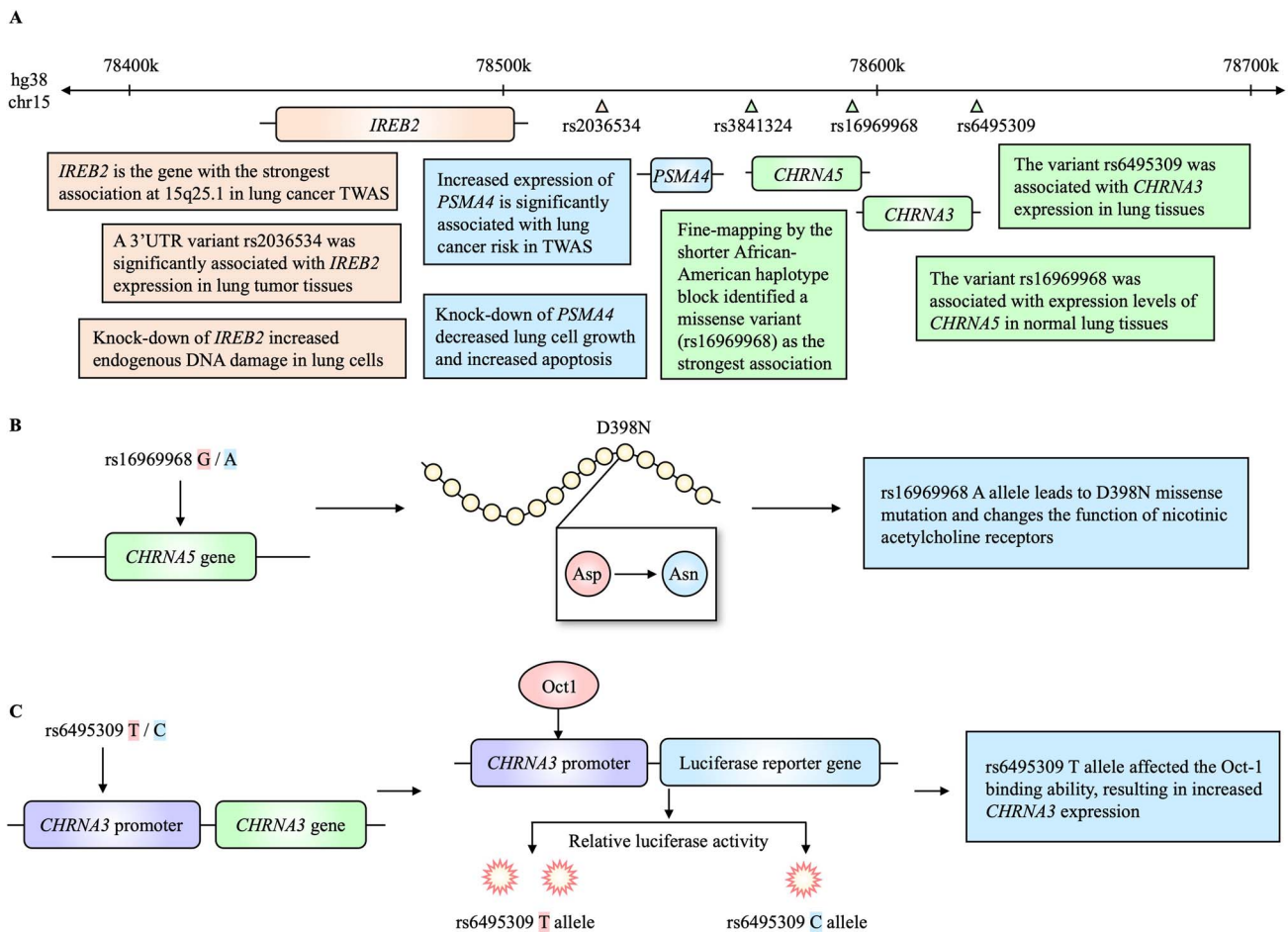
anterior cingulate cortex to ventral striatum circuits was associated with smoking and addiction severity among individuals with D398N substitution (48). In mice, the D398N substitution will cause a partial loss of nAChR function, along with increased nicotine intake (24,49). rs16969968 and other variants were expression quantitative trait loci (eQTL) for *CHRNA5* levels in tumor-adjacent normal lung tissues in European populations, which was replicated in two additional lung tissue datasets (27). Beyond simple overlap of GWAS and eQTL *P*-values for a single variant, a transcriptome-wide association study (TWAS) using GTEx lung tissues formally found a significant association between imputed *CHRNA5* expression levels based on genotypes of multiple local variants and lung cancer risk (31) (Fig. 1A). Lung cancer risk-associated alleles of these SNPs were correlated with lower *CHRNA5* levels.

*CHRNA3* eQTLs were also significant in GTEx lung tissues with lower levels being correlated with the risk (31). In an experimental assessment for *CHRNA3*, rs6495309 (reported in the Chinese population, weak LD  $R^2 = 0.0175$  with rs55781567 in EUR) (41), a variant in the promoter region was shown to affect the Oct-1 binding affinity, and risk allele C resulted in an increased expression of *CHRNA3* in lung tissues (opposite from TWAS; Fig. 1C). Moreover, epigenetic silencing of *CHRNA3* (25) leads to inhibition of apoptosis and  $Ca^{2+}$  influx in the presence of nicotine in lung cancer cells. These findings suggested that variants at 15q25.1 may influence lung cancer risk via both nAChR protein function and the expression of these genes.

In addition to nAChR subunit genes, a few other genes in the region have gained functional support in lung cancer development. Proliferation and apoptosis assays demonstrated that knock-down of *PSMA4* (Proteasome  $\alpha$ -4 Subunit Isoform 1) significantly decreased the growth and increased the apoptosis of lung cancer cell lines (50). Consistently, higher levels of *PSMA4* in normal lung tissues were associated with lung cancer risk in a TWAS study (31). In the same study, *IREB2* (Iron Responsive Element Binding Protein 2) was the gene most strongly associated with lung cancer at 15q25.1, where lower levels are associated with the risk. Further, knock-down of *IREB2* increased DNA damage in the immortalized lung fibroblasts, suggesting a potential mechanism through DNA damage and repair pathways (31). Another study found that the risk allele C of rs2036534 ( $R^2 = 0.1803$  with rs55781567 in EUR) was associated with increased *IREB2* expression in lung tumors (Fig. 1A) (51). Together, these findings provide strong support for nAChR genes underlying this locus and suggest a more complex mechanism involving other genes in tumorigenesis beyond smoking behavior.

### Telomere biology 5p15.33

The locus at 5p15.33 is a well-known multi-cancer locus (52,53) and was associated with lung cancer across different populations (12,14,16). The most robust



**Figure 1.** Summary of the functional findings from lung cancer GWAS locus at 15q25.1. (A) Summarized findings around genes *IREB2*, *PSMA4*, *CHRNA5* and *CHRNA3*. (B) Experimental findings of rs16969968 in *CHRNA5*. (C) Regulatory activity of rs6495309 on *CHRNA3* identified by luciferase reporter assay. Lung cancer risk-associated alleles are shaded in red and protective alleles in blue.

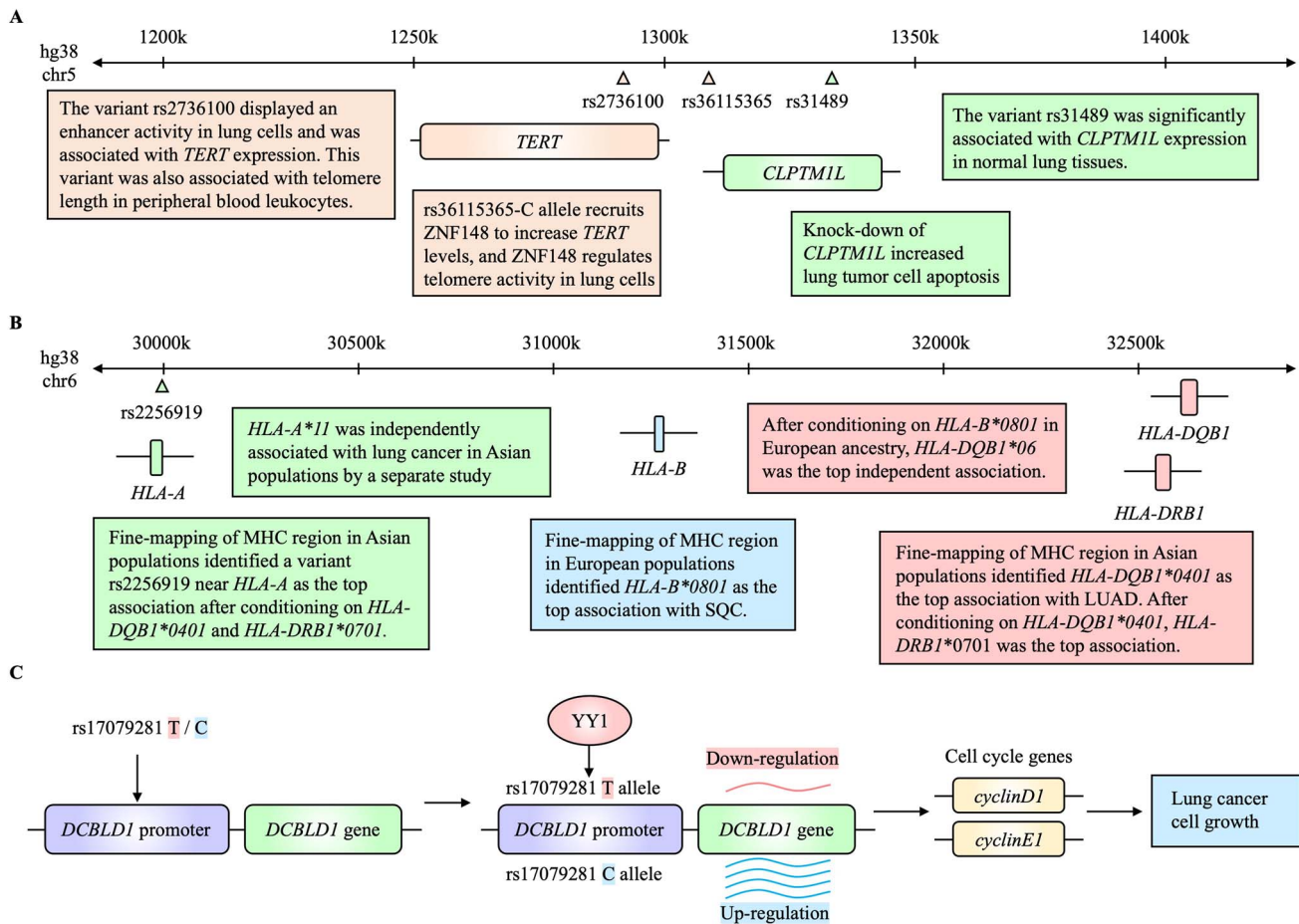
signal is tagged by rs2736100, which was consistently observed in European and East Asian populations among smokers and never-smokers. A functional study found that rs2736100 displayed a potential enhancer activity by luciferase reporter assays in lung cells, and the lung cancer risk-associated C allele was correlated with higher expression of *TERT* (Telomerase Reverse Transcriptase) in normal and lung tumor tissues (Fig. 2A) (54). *TERT* encodes a catalytic subunit of the telomerase complex, which enables the cells to circumvent senescence once reactivated and promotes cancer development by sustaining telomere length and chromosomal integrity (55). Notably, rs2736100-C allele is correlated with longer telomere length in peripheral white blood cells (Fig. 2A) (56). Moreover, genetically predicted longer telomere length (57) and longer telomere length in leukocytes are associated with increased lung cancer risk (58).

GWAS in diverse populations and stepwise conditional analyses have identified additional independent signals (still significant when conditioning on the lead SNP as a covariate in the regression) in this locus (40,54,59–61), suggesting that there are multiple functional variants contributing to lung cancer risk through *TERT* or other genes. A multi-cancer meta-analysis including lung can-

cer in this locus identified six independent regions (52). Functional characterization of one of them identified rs36115365 as a functional variant (tagged by rs37004) which is located in an intergenic region between *TERT* (~18 kb telomeric from the SNP) and *CLPTM1L* (Cleft Lip and Palate Transmembrane Protein 1-Like; ~5 kb centromeric from the SNP) (53). The lung cancer-protective C allele of rs36115365 preferentially recruits a transcription factor, ZNF148, which increases *TERT* expression, but not *CLPTM1L*, in lung cancer cell lines. Depletion of ZNF148 in a lung cancer cell line also reduced telomerase activity and increased telomere length. Notably, the same rs36115365-C allele increases the risk of pancreatic and testicular cancer but decreases the risk of lung cancer and melanoma, although the variant functionality on *TERT* was consistently shown in different cancer cell types.

Functional annotation of some fine-mapped variants (62) also suggested the involvement of *CLPTM1L* in this locus. rs31489 was significantly associated with *CLPTM1L* levels in normal lung tissues (14), with risk allele C associated with higher expression (Fig. 2A). Consistent with this observation, higher expression levels of *CLPTM1L* were reported in LUAD tissues compared to





**Figure 2.** Summary of the functional findings from lung cancer GWAS loci at 5p15.33, 6p21 (MHC) and 6p22.1. (A) Summarized findings around genes *TERT* and *CLPTM1L*. (B) Conditional fine-mapping analyses on the MHC region at 6p21. (C) Regulatory mechanisms of rs17079281 on *DCBLD1* expression and its functional consequences. Lung cancer risk-associated alleles are shaded in red and protective alleles in blue.

normal lung tissues (63). Functional studies demonstrated that *CLPTM1L* is required for K-Ras induced lung tumorigenesis in a mouse model (64), and knock-down of *CLPTM1L* sensitized lung tumor cells for genotoxic stress-induced apoptosis (63). These studies suggested that knock-down of *CLPTM1L* could serve as a potential treatment to impede the growth of KRAS-mutated tumors (65). These data support *TERT* as the most plausible lung cancer susceptibility gene in this locus but also support the involvement of *CLPTM1L*. Given the complexity of genetic signals in this locus, comprehensive identification of multiple functional variants and their respective target genes using high-throughput approaches will be needed.

## Immune response

### Major histocompatibility complex region

In the genomic loci encompassing major histocompatibility complex (MHC) on chromosome 6, heterogeneous lung cancer association signals have been identified. Namely, significant associations were independently found for the loci at 6p21.33 for SQC in Europeans (16) and at 6p21.32 for lung cancer of never-smoking women in East Asian populations (22). MHC is a region

of high gene density shaped by demographic and selective dynamics, resulting in high density of genetic variants, high polymorphisms in protein products of MHC genes (66), LDs spanning longer physical distances than elsewhere in the genome (67), and clustered genes with related functions (68,69). To fine-map the GWAS signals based on genetic variants in the MHC region to amino acid and serotype levels, Ferreiro-Iglesias and colleagues (70) utilized high-density SNP genotyping data and imputed allelic and amino acid polymorphisms of classical human leukocyte antigen (HLA) genes both in European and Asian populations (Fig. 2B). From the European dataset, significant associations were limited to SQC, and *HLA-B*\*0801 was the top risk-associated allele, which is in high-LD with the previous GWAS lead SNP in this region, rs3117582 ( $R^2=0.76$ ). This signal was refined to the 163rd amino acid of *HLA-B*\*0801 within the peptide-binding groove that is important for T cell receptor recognition. *HLA-B*\*0801 is a part of the ancestral haplotype 8.1 which is commonly associated with a higher risk of immune-related diseases in European populations. After conditioning on *HLA-B*\*0801, the *HLA-DQB1*\*06 allele group exhibited an independent protective effect, and *DQB1*\*0603, showing the strongest

association in the group, has been associated with multiple cancers in the same direction (71,72). In Asian populations, significant associations were found in LUAD, including two independent HLA alleles (located in *HLA-DQB1\*0401* and *HLA-DRB1\*0701*) and one variant near *HLA-A* (rs2256919; Fig. 2B) (70). A separate fine-mapping study performed HLA amino acid and serotype imputation in Asian populations (73). This study identified independent associations of rs12333226 (an eQTL locus with *HLA-A* and *HLA-H* in monocytes) and *HLA-A\*11*, as well as bi-allelic coding variants of *HLA-DRB1* that are in LD with a previously reported SNP, rs2395185 ( $R^2 = 0.51$ ). Notably, no overlap of signals was observed between European and Asian populations in the MHC region, which suggests a hypothesis that exposures to different antigens in two populations could have shaped the roles of different susceptibility alleles contributing to two lung cancer histological types. Studying potential tumor antigen and HLA interaction and their implication in immunotherapy might be beneficial in the treatment strategy for different types of lung cancer.

## Cell cycle regulation

### 6q22.1

The locus at 6q22.1 was initially identified in lung cancer of Asian never-smoking women (22) and later replicated in two trans-ancestry GWAS, including smokers and never-smokers, displaying a stronger association in LUAD (23,33). The two nearest genes in this locus are *DCBLD1* (Discoidin, CUB and LCCL domain containing 1) and *ROS1* (ROS proto-oncogene 1, receptor tyrosine kinase). Recurrent somatic fusion of *ROS1* is one of the known driver events of NSCLC tumorigenesis, especially in Asian never-smokers. While there have not been findings linking lung cancer-associated variants to *ROS1* function, TWAS of both European (31) and Chinese (74) populations showed that lower levels of *DCBLD1* in lung tissues are correlated with lung cancer risk. Wang and colleagues (26) further prioritized rs17079281 ( $R^2 = 0.3784$ ,  $D' = 0.9928$  with the Asian GWAS lead SNP rs9387478) in the *DCBLD1* promotor, based on a high functionality score from RegulomeDB (75). They found that rs17079281-T allele exhibited a higher binding affinity to transcription factor YY1 than the risk-associated C allele. T > C CRISPR editing in lung cells reduced YY1 binding and increased *DCBLD1* expression. Knockdown and over-expression of *DCBLD1* in lung cancer cell lines indicated that *DCBLD1* promotes cell proliferation and cell cycle progression. *DCBLD1* knockdown also reduced tumor growth in mouse xenografts, which is consistent with the tumor-promoting role of *DCBLD1* (Fig. 2C). Given that there are multiple high-LD variants with the GWAS lead SNP spanning the regulatory regions of *DCBLD1* and *ROS1*, functional assessment of these variants and their target genes will help rule out the role of *ROS1* and establish a robust link between the genetic signal and *DCBLD1*.

## DNA damage repair and cellular stress response

### 13q13.1

*BRCA2* (Breast Cancer gene 2), in the locus at 13q13.1, encodes a key regulator of homologous recombination (HR) that protects the genome from double-strand DNA damage during DNA replication. Pathogenic mutations in *BRCA2* were first associated with breast and ovarian cancer risk, but not with lung cancer risk (76). Wang et al. reported a rare variant, rs11571833 (minor allele frequency, MAF 1% in EUR; OR = 2.47) associated with lung cancer risk in this locus with a relatively large effect size, which is mainly driven by the association with SQC rather than LUAD (15). The association of rs11571833 was confirmed by several other studies in European populations for SQC and SCLC (16,77,78) and by the recent trans-ancestry GWAS, where two intronic variants, rs11571818 and rs11571815 in LD ( $R^2 = 1$ , EUR), were significant in lung cancer and SQC, respectively (33).

rs11571833 transverses A > T in the last exon of *BRCA2*, resulting in a premature stop codon (K3326X). Given that K3326X carriers were reported with normal levels of *BRCA2* mRNA in blood, this genotype might generate a truncated *BRCA2* protein by escaping nonsense-mediated decay (78). Although the stop codon is ~20 amino acids downstream of a RAD51-binding domain (79,80), which is critical to *BRCA2* functions in DNA damage repair, the functional impact of the truncated protein has not been assessed in experimental systems (e.g. effects on nuclear localization of *BRCA2* (81)). Given that rs11571833 is associated with lung cancer and other cancers linked with strong genotoxic stresses (e.g. smoking, ultraviolet radiation) (78,82), investigating the role of this variant in impaired DNA repair function interacting with genotoxic stress conditions will provide a deeper biological understanding to the genetic findings in this locus.

### 22q12.1

In this locus, a rare coding variant in *CHEK2* (cell cycle checkpoint kinase 2), rs17879961, was significantly associated with SQC (MAF < 3% in Finish Europeans; OR = 0.38) (15). The association was replicated in later studies of SQC and overall lung cancer in European (16) and multi-ancestry populations (33) as well as in a meta-analysis of aerodigestive squamous cancers in Europeans (83). rs17879961 is a missense variant (p.Ile157Thr, I157T) within *CHEK2* which encodes a cell cycle check point kinase that triggers growth arrest or apoptosis when DNA damage occurs. Ile157 is located in a critical position of the forkhead-associated (FHA) domain that is required for *CHEK2* dimerization and subsequent kinase activation, and I157T substitution impairs this process (84). Although the impaired *CHEK2* activation is considered to compromise downstream phosphorylation of *CHEK2* substrate proteins, the I157T substitution is correlated with reduced lung cancer and aerodigestive squamous cancer risk. In contrast, the I157T substitution is associated with increased breast cancer risk (85,86) and

is also found in Li–Fraumeni cancer syndrome families not caused by *TP53* mutations (87). One speculative explanation for the protective effect of rs17879961 in smoking-related cancers is that CHEK2-mediated stem cell apoptosis results in frequent renewal/replacement of stem cells in a seemingly mutagenic environment (e.g. constant exposure to tobacco smoke), which might increase the mutation rates in stem cell populations, eventually leading to cancer development. Therefore, compromised CHEK2 activity might be linked to a reduced risk of smoking-related cancers (88). However, this hypothesis needs to be tested in experimental systems incorporating exposures and in tumor-evolution analysis.

### 12p13.33

The locus at 12p13.33 was initially associated with SQC in European populations (16,89,90) tagged by SNPs on or near *RAD52* (recombination repair gene 52), which encodes a key regulator of HR-related genomic instability. In a follow-up study, two SQC-associated variants (rs6413436 and rs10744729) were *cis*-eQTLs of *RAD52* expression in lung tissues, where the risk alleles are correlated with higher *RAD52* levels (91). These two SNPs are in weak LD ( $R^2 = 0.215$  and  $0.2873$ ,  $D' = 0.9278$  and  $0.9456$ , respectively) with the lead SNP (rs7953330) that later reached genome-wide significance in an SQC GWAS (16). Consistent with these findings, higher *RAD52* levels in normal lung tissues were associated with lung cancer risk by TWAS (31). *RAD52* levels were increased in SQC tumors compared to normal tissues, and somatic amplification of *RAD52* region was reported in SQC tumors from TCGA (91). Experimental data suggested that depletion of *RAD52* in mouse immortalized bronchial epithelial cells attenuated the cell proliferation rate and induced senescence (91). *RAD52* plays an essential role in DNA repair pathways (92) and maintaining tumor genome integrity (91). *RAD52*-dependent synthetic lethality of HR-deficient cancer cells (e.g. *BRCA2*, *BRCA1* or *PALB2*) was also reported (93,94). Further studies are warranted to identify functional variants in this locus and characterize their roles in increasing *RAD52* expression in the context of tobacco smoking and SQC-specific somatic alterations.

### 11q22.3

A rare high-effect size coding variant, rs56009889, in *ATM* (Ataxia-Telangiectasia Mutated) was initially associated with LUAD in a rare-variant analysis (MAF < 1%) of European case–control sets (95) and was replicated in a recent cross-ancestry GWAS (33). The variant frequency was higher in LUAD, females, light-smokers and somatic *EGFR* mutation carriers in a separate dataset (95). rs56009889 is extremely rare in all populations except Jewish-descent individuals, where it has a MAF of 2%. Because of its higher prevalence in this population, several homozygotes, all with lung cancer, were observed. This variant results in L2307F substitution in the conserved FRAP-ATM-TRRAP (FAT) domain of *ATM* preceding

the kinase domain (96). Given that *ATM* is a key kinase regulating DNA damage response signaling, the variant effect on *ATM* function in the context of lung cancer development should be investigated. Additional rare deleterious mutations of *ATM* (that cause Ataxia Telangiectasia in homozygotes) identified by exome sequencing are also associated with increased LUAD or lung cancer risk (97–99). Findings of low-frequency medium-penetrance variants in *ATM* further support the contribution of DNA damage response and repair pathways in lung cancer susceptibility.

### 3q28

Multiple variants in the locus at 3q28 were associated with lung cancer risk, especially in LUAD, including rs4488809 (18) and rs11375254 (23) identified in Asian GWAS. Association of the same signal was observed in Asian female never-smokers (22) and European populations (15). rs4488809 (intron 1 of *TP63*, tumor protein p63) is associated with *TP63* expression in normal lung tissues (100) and lung tissue TWAS showed that lower *TP63* levels are correlated with the risk (31). *TP63* encodes a *TP53* family member protein which plays a key role in cellular differentiation, stress response and cancer development (101). The transactivation (TA) isoforms of *TP63* (TAp63) are induced when the cells are exposed to DNA damage, transactivating the target genes of *TP53* tumor suppressor and thus affecting DNA damage response (102). On the other hand, chromosomal regions encompassing *TP63* are amplified in squamous cell carcinomas of different organs including lung SQC, which is correlated with increased transcripts of  $\Delta N$  isoform of *TP63* ( $\Delta Np63$ ) (103,104).  $\Delta Np63$  is localized to basal cells of squamous epithelium and has roles in stem cell maintenance (104). Although the known functions of *TP63* isoforms have relevance in lung cancer, statistical fine-mapping and subsequent functional assessment of the candidate variants will be needed to identify functional variants linking genetic signals to *TP63* expression and/or function.

## Future Directions of Lung Cancer Post-GWA Studies

While functional studies of several loci improved our understanding of lung cancer susceptibility, most reported GWAS loci remain unexplored, and numerous new signals are being discovered with emerging GWAS in larger and more diverse populations. We summarize key considerations in lung cancer GWAS follow-up studies, including emerging high-throughput functional approaches and available lung cancer-relevant resources.

### Identification of functional variants

Prior to experimental testing of individual variants from GWAS loci, statistical fine-mapping can prioritize variants that are most likely to be responsible for the association (i.e. credible causal variants, CCV) from genetically linked variants and further identify



independent signals that might be hindered by the primary signal (reviewed in (105)). For example, stepwise conditional analyses of complex lung cancer loci in diverse populations identified multiple independent signals (e.g. 5p15.33 locus). Bayesian fine-mapping, on the other hand, provides posterior inclusion probability of each variant for capturing likely causal variants. An advantage of Bayesian methods is that functional annotations can be used as a prior to weight the variants to narrow down CCV based on lung cancer-relevant molecular features. Because GWAS variants are mostly non-protein-coding, cis-regulation is considered the main mechanism. We summarize the available functional annotation resources in [Supplementary Material, Table S4](#), which profiled cis-regulatory genomic regions in lung-relevant tissue/cell types (e.g. gene promoters, enhancers and transcription factor binding sites). Algorithms are also available to estimate the impact of protein-coding variants on protein structure and function (summarized in (106)).

One challenge in fine-mapping is its general requirement of (preferably in-sample) LD matrix drawn from a large number of individuals to accurately represent the GWAS population's LD structure. This issue becomes even more challenging with emerging multi-ancestry GWAS, including those in lung cancer (23,33) and warrants more advanced algorithms. At the same time, multi-ancestry GWAS provides an opportunity for cross-ancestry fine-mapping (107,108). For example, given that African ancestry populations have significantly narrower LD, fine-mapping could achieve a substantial reduction in CCV size assuming that the causal variants are shared across populations. This strategy was applied to resolving the large LD blocks in the 15q25.1 region (42).

Beyond fine-mapping and functional annotation, reporter assays based on the measurement of enzymatic activity have been widely utilized to investigate the regulatory potential of DNA sequences for gene transcription one variant at a time (e.g. luciferase assay) (109). For lung cancer loci, heuristic selection of variants (e.g. using  $R^2$  or  $P$ -value cutoff) followed by luciferase assays have been mainly used, in part due to the experimental burden of testing multiple variants. Recent advancements such as Massively Parallel Reporter Assays (MPRA) scale-up conventional reporter assays by adopting massively parallel sequencing and thus can test transcriptional regulatory activities of thousands of variants from multiple loci (110–112). Given that fine-mapping provides limited prioritization power, especially in high-LD regions, MPRA and similar high-throughput methods could break the LD tie and efficiently identify functional variants.

## Identification and characterization of susceptibility genes

### Molecular QTL-based approaches

Molecular QTL identifies genetic variants that contribute to inter-individual differences in molecular traits and

thus can identify candidate susceptibility genes from GWAS loci. Various molecular traits could be assessed for QTL analyses, including expression (eQTL), splicing (sQTL), protein (pQTL), DNA methylation (meQTL), microRNA (miQTL), chromatin accessibility (caQTL), histone modification (hQTL), RNA stability QTL and so on. We summarize available lung and related tissues QTL datasets in [Table 1](#).

To formally assess the overlap between QTL and GWAS signals, colocalization approaches compare the association patterns of GWAS and QTLs and investigate whether the same variants drive both signals (e.g. COLOC (113), eCAVIAR (114)). TWAS, on the other hand, utilizes the local genetic variants that contribute to gene expression in eQTL datasets to impute gene expression levels for the case and control individuals from the GWAS dataset. Association test between genetically predicted gene expression levels and the trait (e.g. PrediXcan (115), FUSION (116)) could find transcriptome-wide significant associations, including the loci that were underpowered in the GWAS itself based on the individual SNP-based testing. TWAS and colocalization were performed for two lung cancer GWAS datasets in recent studies. Bossé and colleagues (31) performed TWAS integrating lung eQTL ( $n = 1038$ ) with lung cancer GWAS summary statistics (16) (both datasets in EUR) and identified candidate susceptibility genes for 37% of GWAS loci and further identified novel loci. Zhu and colleagues (74) performed cross-tissue TWAS integrating eQTL of 44 GTEx tissue types (~85% EUR) with lung cancer GWAS data from Chinese populations using UTMOST (117). They identified multiple candidate genes from both known and novel loci in lung and other tissue types, and a subset of them was also supported by colocalization. Given that TWAS and colocalization models are typically based on the assumption that the allele frequencies and LD structures are the same between GWAS and QTL datasets, lung QTL resources in non-European populations are critically needed.

A critical consideration in QTL-based and functional annotation approaches in selecting the most relevant tissue type(s), where lung cancer susceptibility variants and genes are functional. There are multiple agnostic approaches that combine GWAS statistics with tissue-specific transcriptomic and epigenomic annotations to identify the most relevant tissues or cell types. Stratified LD score regression (S-LDSC) considers LD structure and heritability of all variants regardless of association cutoff to estimate the heritability attributable to different functional annotations. A recent study using S-LDSC in multiple solid cancers did not find significantly enriched tissue types for lung cancer using European GWAS statistics (6) perhaps due to a low heritability captured in the current size of GWAS and lack of cell-type-specific annotation datasets. Indeed, lung tissue consists of over 50 cell types (118), and multiple epithelial cell types are considered the cells of lung cancer origin (LUAD: alveolar type II, bronchiolar epithelial club and bronchioalveolar



**Table 1.** Available QTL-based datasets for lung cancer post-GWAS studies

Resources	Brief information	Data storage/links
GTEEx v8	Samples were collected from 54 tissues across 838 donors, mainly for eQTL and sQTL datasets, including 515 lung samples	<a href="https://gtexportal.org/home/">https://gtexportal.org/home/</a>
Hao et al.	Lung eQTL datasets from 1111 European ancestry individuals who underwent lung resectional surgery (non-tumor lung tissue samples). The majority were smokers or former smokers	<a href="https://pubmed.ncbi.nlm.nih.gov/23209423/">https://pubmed.ncbi.nlm.nih.gov/23209423/</a>
Luo et al.	Human small ( $n = 112$ ) and large airway ( $n = 40$ ) epithelial cells eQTL datasets	Data stored in NCBI GSE5057 and GSE40364
Shi et al.	Histologically normal human lung tissues ( $n = 210$ ) meQTL datasets of European ancestry individuals. Mostly smokers ( $n = 206$ )	Data stored in dbGAP phs000093.v2.p2 and GSE52401
Morrow et al.	Lung meQTL datasets from 116 European ancestry individuals (90 COPD cases and 36 controls). All were smokers	<a href="https://pubmed.ncbi.nlm.nih.gov/29313708/">https://pubmed.ncbi.nlm.nih.gov/29313708/</a>
PancanQTL	Comprehensive database for cis-eQTLs and trans-eQTLs in 33 cancer types from TCGA, including LUAD ( $n = 514$ ) and LUSC ( $n = 500$ )	<a href="http://gong_lab.hzau.edu.cn/PancanQTL/">http://gong_lab.hzau.edu.cn/PancanQTL/</a>

**Table 2.** Available single-cell RNA-seq datasets for human lung

Resources	Sample information	Data storage/links
Reyfman et al.	Donor lung biopsies ( $n = 8$ ) and lung explants from patients with pulmonary fibrosis ( $n = 4$ ), systemic sclerosis ( $n = 1$ ), polymyositis ( $n = 1$ ) and chronic hypersensitivity pneumonitis ( $n = 1$ )	<a href="https://doi.org/10.1164/rccm.201712-2410OC">https://doi.org/10.1164/rccm.201712-2410OC</a>
Braga et al.	Asthmatic lungs ( $n = 9$ , aged from 49 to 65 years old) and normal lungs from deceased organ donors ( $n = 11$ , aged from 44 to 65 years old)	<a href="https://asthma.cellgeni.sanger.ac.uk/">https://asthma.cellgeni.sanger.ac.uk/</a>
Travaglini et al.	Normal lung tissues from 3 patients with focal lung tumors (aged from 46 to 75 years old)	<a href="https://hlca.ds.czbiohub.org/">https://hlca.ds.czbiohub.org/</a>
Maynard et al.	Metastatic lung tumor samples from biopsies ( $n = 30$ ) (aged from 39 to 77 years old, all with LUAD except 1 SQC)	<a href="https://doi.org/10.1016/j.cell.2020.07.017">https://doi.org/10.1016/j.cell.2020.07.017</a>
Guo et al.	NSCLC tumors and adjacent non-tumor lung samples from 14 patients (LUAD, $n = 11$ ; SQC, $n = 3$ ; aged from 45 to 78 years old)	<a href="http://lung.cancer-pku.cn">http://lung.cancer-pku.cn</a>
Kim et al.	Normal lung ( $n = 11$ ), early-stage lung tumor ( $n = 11$ ), advance stage lung tumor ( $n = 4$ ), lymph node metastases ( $n = 7$ ), normal lymph node ( $n = 10$ ), pleural effusion ( $n = 5$ ), brain metastases ( $n = 10$ )	<a href="https://pubmed.ncbi.nlm.nih.gov/32385277/">https://pubmed.ncbi.nlm.nih.gov/32385277/</a>
Habermann et al.	Explanted lung samples from individuals with pulmonary fibrosis ( $n = 20$ ) and nonfibrotic controls ( $n = 10$ )	<a href="https://pubmed.ncbi.nlm.nih.gov/32832598/">https://pubmed.ncbi.nlm.nih.gov/32832598/</a>

stem cells (119,120), SQC: tracheal basal cell progenitors, SCLC: neuroendocrine cells (121)). Current QTL and other annotation data are mainly using bulk lung tissues and hence may not capture cell-type and context (e.g. cell state, cell-cell interaction)-specific gene regulation, which could result in limited utility in explaining GWAS loci. For example, colocalization analysis of the most recent lung cancer GWAS and eQTL data from eight types of bulk tissue in GTEEx identified susceptibility genes for ~50% of the GWAS loci (33). To address the limitations of bulk-tissue approaches, emerging single-cell-based human lung datasets can be explored (Table 2). These datasets profiled cell-type-specific lung transcriptomes and identified previously unknown lung cell types (e.g. transitional states from stem cells to differentiated cells) (122,123). S-LDSC and similar agnostic approaches (124) could be applied to these single-cell human lung datasets to prioritize relevant cell types.

### Other functional assays

While QTL-based approaches require a large collection of samples, other experimental approaches using cell-based systems could identify lung cancer susceptibility

genes and characterize their functions. Chromatin interaction analyses can identify susceptibility genes from GWAS loci by detecting physical interactions between GWAS variants and target gene promoters (e.g. Hi-C, HiChIP (125), Capture-Hi-C (126)) based on long-range cis-regulation via chromatin looping (127). This cell-based approach is particularly valuable in assessing low-frequency variants that might have insufficient statistical power in typical QTL datasets and can also incorporate diverse cellular contexts (e.g. stimulation, cell states) that might not be captured in QTL datasets. These approaches have successfully identified hundreds of candidate susceptibility genes from breast cancer GWAS loci (128,129).

Genome-editing technologies could test the variant effect on its target gene expression/function or assess candidate gene function in cell-based systems (130). Unlike the reporter assays, genome-editing investigates the impact of variants on their endogenous target genes in a physiologically relevant context (131). Wang and colleagues (26) used CRISPR editing to switch the genotype of a lung cancer-associated variant in a lung cell line and assess the allelic effect on transcription factor

binding and target gene expression. CRISPR-mediated high throughput screens adopting knock-out, activation (CRISPRa) or inhibition (CRISPRi) can test thousands of candidate genes or variants using a guide-RNA panel followed by massively parallel sequencing (132). An autoimmune disease locus was investigated using CRISPRi- and CRISPRa-tiling to efficiently narrow down functional variants affecting the target gene expression (133). For CRISPR-based approaches, compatibility of the cell type to lentiviral transduction and drug selection process (132) and p53-dependent cellular toxicity caused by Cas9-induced DNA double-strand breaks (134) should be considered.

Other functional assays relevant to lung cancer could be performed to characterize target gene function in tumorigenesis. Cellular phenotypes related to tumorigenesis (cell proliferation, migration, invasion, altered metabolism, DNA damage repair and apoptosis (135)) can be investigated *in vitro* or *in vivo* models (136). For example, endogenous DNA damage levels of lung cells upon knockdown or overexpression of candidate genes were used to screen multiple genes from lung cancer GWAS loci in recent studies (31,33). Importantly, experimental systems should reflect the lung cancer-relevant biological contexts, including frequent somatic alterations (e.g. EGFR, KRAS), cell types (primary or cancer) and exposures.

## Concluding Remarks

Post-GWAS functional studies will help improve etiological insights and biological understanding of lung tumorigenesis in general. Insights from investigating the interaction among susceptibility genes and exposures (e.g. smoking) and somatic driver events will further provide clues to better prevention and potential therapy. For example, understanding how large-effect-size lung cancer-associated coding variants (e.g. BRCA2, CHEK2 and ATM) contribute to early stages of lung tumorigenesis could shed light on potential therapeutic approaches tapping into synthetic lethality of DNA damage repair pathways in a subset of patients.

It is possible that functional studies could contribute to better predicting lung cancer risk via incorporation into a polygenic risk score (PRS). Previous studies proved that, beyond age and smoking pack years, PRS is an independent and effective predictor of lung cancer risk (23). Although cumulative effects of PRS and modifiable risk factors improved the performance (137), the discriminative power of the current risk-prediction model is still far from the requirement in clinical practice. Functional annotations can further improve PRS by assigning priors to effect sizes and reducing the inflation of association statistics in the high-LD region. For example, incorporating genomic and epigenomic features into PRS (138) or prioritizing cell-type-specific regulatory variants (139) improved prediction accuracy in multiple diseases. In lung cancer, restricting the GWAS variants based

on biological pathways improved the performance of prediction models (140). It is expected that PRS could be improved by post-GWAS findings, which could help construct clinical risk prediction models that benefit smokers and never-smokers via targeted lung cancer screening programs.

## Supplementary Material

Supplementary Material is available at HMGJ online.

## Acknowledgements

This work was supported by the Intramural Research Program (IRP) of the Division of Cancer Epidemiology and Genetics, National Cancer Institute, US National Institutes of Health.

*Conflict of Interest statement.* The authors declare no conflict of interest. The funder had no role in the writing of this review or the decision to submit it for publication.

## Funding

This research was partially supported by the National Institutes of Health (NIH) for Integrative Analysis of Lung Cancer Etiology and Risk (U19CA203654), Sequencing Familial Lung Cancer (R01CA243483) and Genetic Analysis of Lung Cancer Etiology and Risk (R03CA25622). Dr Amos is a Research Scholar of the Cancer Prevention Research Interest of Texas (CPRIT) award (RR170048).

## References

1. Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A. and Bray, F. (2021) Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.*, **71**, 209–249.
2. Ferlay, J., Shin, H.R., Bray, F., Forman, D., Mathers, C. and Parkin, D.M. (2010) Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int. J. Cancer*, **127**, 2893–2917.
3. Parkin, D.M., Bray, F., Ferlay, J. and Pisani, P. (2005) Global cancer statistics, 2002. *CA Cancer J. Clin.*, **55**, 74–108.
4. Mucci, L.A., Hjelmborg, J.B., Harris, J.R., Czene, K., Havelick, D.J., Scheike, T., Graff, R.E., Holst, K., Moller, S., Unger, R.H. et al. (2016) Familial risk and heritability of cancer among twins in Nordic countries. *JAMA*, **315**, 68–76.
5. Dai, J., Shen, W., Wen, W., Chang, J., Wang, T., Chen, H., Jin, G., Ma, H., Wu, C., Li, L. et al. (2017) Estimation of heritability for nine common cancers using data from genome-wide association studies in Chinese population. *Int. J. Cancer*, **140**, 329–336.
6. Jiang, X., Finucane, H.K., Schumacher, F.R., Schmit, S.L., Tyrer, J.P., Han, Y., Michailidou, K., Lesueur, C., Kuchenbaecker, K.B., Dennis, J. et al. (2019) Shared heritability and functional enrichment across six solid cancers. *Nat. Commun.*, **10**, 431.
7. Byun, J., Han, Y., Ostrom, Q.T., Edelson, J., Walsh, K.M., Pettit, R.W., Bondy, M.L., Hung, R.J., McKay, J.D., Amos, C.I. et al. (2021) The shared genetic architectures between lung cancer and multiple polygenic phenotypes in genome-wide association studies. *Cancer Epidemiol. Biomark. Prev.*, **30**, 1156–1164.

8. Sampson, J.N., Wheeler, W.A., Yeager, M., Panagiotou, O., Wang, Z., Berndt, S.I., Lan, Q., Abnet, C.C., Amundadottir, L.T., Figueroa, J.D. et al. (2015) Analysis of heritability and shared heritability based on genome-wide association studies for thirteen cancer types. *J. Natl. Cancer Inst.*, **107**, djv279.
9. Gazdar, A., Robinson, L., Oliver, D., Xing, C., Travis, W.D., Soh, J., Toyooka, S., Watumull, L., Xie, Y., Kernstine, K. et al. (2014) Hereditary lung cancer syndrome targets never smokers with germline EGFR gene T790M mutations. *J. Thorac. Oncol.*, **9**, 456–463.
10. Benusiglio, P.R., Fallet, V., Sanchis-Borja, M., Coulet, F. and Cadranet, J. (2021) Lung cancer is also a hereditary disease. *Eur. Respir. Rev.*, **30**, 210045.
11. Kanwal, M., Ding, X.J. and Cao, Y. (2017) Familial risk for lung cancer. *Oncol. Lett.*, **13**, 535–542.
12. Wang, Y., Broderick, P., Webb, E., Wu, X., Vijayakrishnan, J., Matakidou, A., Qureshi, M., Dong, Q., Gu, X., Chen, W.V. et al. (2008) Common 5p15.33 and 6p21.33 variants influence lung cancer risk. *Nat. Genet.*, **40**, 1407–1409.
13. McKay, J.D., Hung, R.J., Gaborieau, V., Boffetta, P., Chabrier, A., Byrnes, G., Zaridze, D., Mukeria, A., Szeszenia-Dabrowska, N., Lissowska, J. et al. (2008) Lung cancer susceptibility locus at 5p15.33. *Nat. Genet.*, **40**, 1404–1406.
14. Landi, M.T., Chatterjee, N., Yu, K., Goldin, L.R., Goldstein, A.M., Rotunno, M., Mirabello, L., Jacobs, K., Wheeler, W., Yeager, M. et al. (2009) A genome-wide association study of lung cancer identifies a region of chromosome 5p15 associated with risk for adenocarcinoma. *Am. J. Hum. Genet.*, **85**, 679–691.
15. Wang, Y., McKay, J.D., Rafnar, T., Wang, Z., Timofeeva, M.N., Broderick, P., Zong, X., Laplana, M., Wei, Y., Han, Y. et al. (2014) Rare variants of large effect in BRCA2 and CHEK2 affect risk of lung cancer. *Nat. Genet.*, **46**, 736–741.
16. McKay, J.D., Hung, R.J., Han, Y., Zong, X., Carreras-Torres, R., Christiani, D.C., Caporaso, N.E., Johansson, M., Xiao, X., Li, Y. et al. (2017) Large-scale association analysis identifies new lung cancer susceptibility loci and heterogeneity in genetic susceptibility across histological subtypes. *Nat. Genet.*, **49**, 1126–1132.
17. Wang, Y., Broderick, P., Matakidou, A., Vijayakrishnan, J., Eisen, T. and Houlston, R.S. (2011) Variation in TP63 is associated with lung adenocarcinoma in the UK population. *Cancer Epidemiol. Biomark. Prev.*, **20**, 1453–1462.
18. Hu, Z., Wu, C., Shi, Y., Guo, H., Zhao, X., Yin, Z., Yang, L., Dai, J., Hu, L., Tan, W. et al. (2011) A genome-wide association study identifies two new lung cancer susceptibility loci at 13q12.12 and 22q12.2 in Han Chinese. *Nat. Genet.*, **43**, 792–796.
19. Dong, J., Hu, Z., Wu, C., Guo, H., Zhou, B., Lv, J., Lu, D., Chen, K., Shi, Y., Chu, M. et al. (2012) Association analyses identify multiple new lung cancer susceptibility loci and their interactions with smoking in the Chinese population. *Nat. Genet.*, **44**, 895–899.
20. Shiraishi, K., Kunitoh, H., Daigo, Y., Takahashi, A., Goto, K., Sakamoto, H., Ohnami, S., Shimada, Y., Ashikawa, K., Saito, A. et al. (2012) A genome-wide association study identifies two new susceptibility loci for lung adenocarcinoma in the Japanese population. *Nat. Genet.*, **44**, 900–903.
21. Li, Y., Sheu, C.C., Ye, Y., de Andrade, M., Wang, L., Chang, S.C., Aubry, M.C., Aakre, J.A., Allen, M.S., Chen, F. et al. (2010) Genetic variants and risk of lung cancer in never smokers: a genome-wide association study. *Lancet Oncol.*, **11**, 321–330.
22. Lan, Q., Hsiung, C.A., Matsuo, K., Hong, Y.C., Seow, A., Wang, Z., Hosgood, H.D., 3rd, Chen, K., Wang, J.C., Chatterjee, N. et al. (2012) Genome-wide association analysis identifies new lung cancer susceptibility loci in never-smoking women in Asia. *Nat. Genet.*, **44**, 1330–1335.
23. Dai, J., Lv, J., Zhu, M., Wang, Y., Qin, N., Ma, H., He, Y.Q., Zhang, R., Tan, W., Fan, J. et al. (2019) Identification of risk loci and a polygenic risk score for lung cancer: a large-scale prospective cohort study in Chinese populations. *Lancet Respir. Med.*, **7**, 881–891.
24. Tammimaki, A., Herder, P., Li, P., Esch, C., Laughlin, J.R., Akk, G. and Stitzel, J.A. (2012) Impact of human D398N single nucleotide polymorphism on intracellular calcium response mediated by alpha3beta4alpha5 nicotinic acetylcholine receptors. *Neuropharmacology*, **63**, 1002–1011.
25. Paliwal, A., Vaissiere, T., Krais, A., Cuenin, C., Cros, M.P., Zaridze, D., Moukeria, A., Boffetta, P., Hainaut, P., Brennan, P. et al. (2010) Aberrant DNA methylation links cancer susceptibility locus 15q25.1 to apoptotic regulation and lung cancer. *Cancer Res.*, **70**, 2779–2788.
26. Wang, Y., Ma, R., Liu, B., Kong, J., Lin, H., Yu, X., Wang, R., Li, L., Gao, M., Zhou, B. et al. (2020) SNP rs17079281 decreases lung cancer risk through creating an YY1-binding site to suppress DCBLD1 expression. *Oncogene*, **39**, 4092–4102.
27. Nguyen, J.D., Lamontagne, M., Couture, C., Conti, M., Pare, P.D., Sin, D.D., Hogg, J.C., Nickle, D., Postma, D.S., Timens, W. et al. (2014) Susceptibility loci for lung cancer are associated with mRNA levels of nearby genes in the lung. *Carcinogenesis*, **35**, 2653–2659.
28. Ji, X., Bosse, Y., Landi, M.T., Gui, J., Xiao, X., Qian, D., Joubert, P., Lamontagne, M., Li, Y., Gorlov, I. et al. (2018) Identification of susceptibility pathways for the role of chromosome 15q25.1 in modifying lung cancer risk. *Nat. Commun.*, **9**, 3221.
29. Qin, N., Li, Y., Wang, C., Zhu, M., Dai, J., Hong, T., Albanes, D., Lam, S., Tardon, A., Chen, C. et al. (2021) Comprehensive functional annotation of susceptibility variants identifies genetic heterogeneity between lung adenocarcinoma and squamous cell carcinoma. *Front. Med.*, **15**, 275–291.
30. Chen, Z., Wen, W., Beeghly-Fadiel, A., Shu, X.O., Diez-Obrero, V., Long, J., Bao, J., Wang, J., Liu, Q., Cai, Q. et al. (2019) Identifying putative susceptibility genes and evaluating their associations with somatic mutations in human cancers. *Am. J. Hum. Genet.*, **105**, 477–492.
31. Bosse, Y., Li, Z., Xia, J., Manem, V., Carreras-Torres, R., Gabriel, A., Gaudreault, N., Albanes, D., Aldrich, M.C., Andrew, A. et al. (2020) Transcriptome-wide association study reveals candidate causal genes for lung cancer. *Int. J. Cancer*, **146**, 1862–1878.
32. Herbst, R.S., Heymach, J.V. and Lippman, S.M. (2008) Lung cancer. *N. Engl. J. Med.*, **359**, 1367–1380.
33. Byun, J., Han, Y., Li, Y., Xia, J., Xiao, X., Sun, R., Walsh, K.M., Gorlov, I., Gorlova, O., Zhou, W. et al. (2020) Trans-ethnic genome-wide meta-analysis of 35,732 cases and 34,424 controls identifies novel genomic cross-ancestry loci contributing to lung cancer susceptibility. *medRxiv*. <https://doi.org/10.1101/2020.10.06.20207753>.
34. Zanetti, K.A., Wang, Z., Aldrich, M., Amos, C.I., Blot, W.J., Bowman, E.D., Burdette, L., Cai, Q., Caporaso, N., Chung, C.C. et al. (2016) Genome-wide association study confirms lung cancer susceptibility loci on chromosomes 5p15 and 15q25 in an African-American population. *Lung Cancer*, **98**, 33–42.
35. Sun, S., Schiller, J.H. and Gazdar, A.F. (2007) Lung cancer in never smokers – a different disease. *Nat. Rev. Cancer*, **7**, 778–790.
36. Dias, M., Linhas, R., Campaignha, S., Conde, S. and Barroso, A. (2017) Lung cancer in never-smokers – what are the differences? *Acta Oncol.*, **56**, 931–935.
37. Thun, M.J., Hannan, L.M., Adams-Campbell, L.L., Boffetta, P., Buring, J.E., Feskanich, D., Flanders, W.D., Jee, S.H., Katanoda, K., Kolonel, L.N. et al. (2008) Lung cancer occurrence in never-



- smokers: an analysis of 13 cohorts and 22 cancer registry studies. *PLoS Med.*, **5**, e185.
38. Wang, Z., Seow, W.J., Shiraishi, K., Hsiung, C.A., Matsuo, K., Liu, J., Chen, K., Yamji, T., Yang, Y., Chang, I.S. et al. (2016) Meta-analysis of genome-wide association studies identifies multiple lung cancer susceptibility loci in never-smoking Asian women. *Hum. Mol. Genet.*, **25**, 620–629.
  39. Seow, W.J., Matsuo, K., Hsiung, C.A., Shiraishi, K., Song, M., Kim, H.N., Wong, M.P., Hong, Y.C., Hosgood, H.D., 3rd, Wang, Z. et al. (2017) Association between GWAS-identified lung adenocarcinoma susceptibility loci and EGFR mutations in never-smoking Asian women, and comparison with findings from western populations. *Hum. Mol. Genet.*, **26**, 454–465.
  40. Hung, R.J., Spitz, M.R., Houlston, R.S., Schwartz, A.G., Field, J.K., Ying, J., Li, Y., Han, Y., Ji, X., Chen, W. et al. (2019) Lung cancer risk in never-smokers of European descent is associated with genetic variation in the 5p15.33 TERT-CLPTM1L region. *J. Thorac. Oncol.*, **14**, 1360–1369.
  41. Amos, C.I., Wu, X., Broderick, P., Gorlov, I.P., Gu, J., Eisen, T., Dong, Q., Zhang, Q., Gu, X., Vijayakrishnan, J. et al. (2008) Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. *Nat. Genet.*, **40**, 616–622.
  42. Walsh, K.M., Gorlov, I.P., Hansen, H.M., Wu, X., Spitz, M.R., Zhang, H., Lu, E.Y., Wenzlaff, A.S., Sison, J.D., Wei, C. et al. (2013) Fine-mapping of the 5p15.33, 6p22.1-p21.31, and 15q25.1 regions identifies functional and histology-specific lung cancer susceptibility loci in African-Americans. *Cancer Epidemiol. Biomark. Prev.*, **22**, 251–260.
  43. Wu, C., Hu, Z., Yu, D., Huang, L., Jin, G., Liang, J., Guo, H., Tan, W., Zhang, M., Qian, J. et al. (2009) Genetic variants on chromosome 15q25 associated with lung cancer risk in Chinese populations. *Cancer Res.*, **69**, 5065–5072.
  44. Liu, J.Z., Tozzi, F., Waterworth, D.M., Pillai, S.G., Muglia, P., Middleton, L., Berrettini, W., Knouff, C.W., Yuan, X., Waeber, G. et al. (2010) Meta-analysis and imputation refines the association of 15q25 with smoking quantity. *Nat. Genet.*, **42**, 436–440.
  45. Tobacco and Genetics, C (2010) Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat. Genet.*, **42**, 441–447.
  46. Thorgeirsson, T.E., Geller, F., Sulem, P., Rafnar, T., Wiste, A., Magnusson, K.P., Manolescu, A., Thorleifsson, G., Stefansson, H., Ingason, A. et al. (2008) A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature*, **452**, 638–642.
  47. Saccone, N.L., Wang, J.C., Breslau, N., Johnson, E.O., Hatsukami, D., Saccone, S.F., Grucza, R.A., Sun, L., Duan, W., Budde, J. et al. (2009) The CHRNA5-CHRNA3-CHRNA4 nicotinic receptor subunit gene cluster affects risk for nicotine dependence in African-Americans and in European-Americans. *Cancer Res.*, **69**, 6848–6856.
  48. Hong, L.E., Hodgkinson, C.A., Yang, Y., Sampath, H., Ross, T.J., Buchholz, B., Salmeron, B.J., Srivastava, V., Thaker, G.K., Goldman, D. et al. (2010) A genetically modulated, intrinsic circulate circuit supports human nicotine addiction. *Proc. Natl. Acad. Sci. U. S. A.*, **107**, 13509–13514.
  49. Morel, C., Fattore, L., Pons, S., Hay, Y.A., Marti, F., Lambolez, B., De Biasi, M., Lathrop, M., Fratta, W., Maskos, U. et al. (2014) Nicotine consumption is regulated by a human polymorphism in dopamine neurons. *Mol. Psychiatry*, **19**, 930–936.
  50. Liu, Y., Liu, P., Wen, W., James, M.A., Wang, Y., Bailey-Wilson, J.E., Amos, C.I., Pinney, S.M., Yang, P., de Andrade, M. et al. (2009) Haplotype and cell proliferation analyses of candidate lung cancer susceptibility genes on chromosome 15q24-25.1. *Cancer Res.*, **69**, 7844–7850.
  51. Fehrer, G., Liu, G., Pintilie, M., Sykes, J., Cheng, D., Liu, N., Chen, Z., Seymour, L., Der, S.D., Shepherd, F.A. et al. (2012) Association of the 15q25 and 5p15 lung cancer susceptibility regions with gene expression in lung tumor tissue. *Cancer Epidemiol. Biomark. Prev.*, **21**, 1097–1104.
  52. Wang, Z., Zhu, B., Zhang, M., Parikh, H., Jia, J., Chung, C.C., Sampson, J.N., Hoskins, J.W., Hutchinson, A., Burdette, L. et al. (2014) Imputation and subset-based association analysis across different cancer types identifies multiple independent risk loci in the TERT-CLPTM1L region on chromosome 5p15.33. *Hum. Mol. Genet.*, **23**, 6616–6633.
  53. Fang, J., Jia, J., Makowski, M., Xu, M., Wang, Z., Zhang, T., Hoskins, J.W., Choi, J., Han, Y., Zhang, M. et al. (2017) Functional characterization of a multi-cancer risk locus on chr5p15.33 reveals regulation of TERT by ZNF148. *Nat. Commun.*, **8**, 15034.
  54. Wei, R., Cao, L., Pu, H., Wang, H., Zheng, Y., Niu, X., Weng, X., Zhang, H., Favus, M., Zhang, L. et al. (2015) TERT polymorphism rs2736100-C is associated with EGFR mutation-positive non-small cell lung cancer. *Clin. Cancer Res.*, **21**, 5173–5180.
  55. Dratwa, M., Wysoczanska, B., Lacina, P., Kubik, T. and Bogunia-Kubik, K. (2020) TERT-regulation and roles in cancer formation. *Front. Immunol.*, **11**, 589929.
  56. Codd, V., Nelson, C.P., Albrecht, E., Mangino, M., Deelen, J., Buxton, J.L., Hottenga, J.J., Fischer, K., Esko, T., Surakka, I. et al. (2013) Identification of seven loci affecting mean telomere length and their association with disease. *Nat. Genet.*, **45**, 422–427 427e1–2.
  57. Machiela, M.J., Hsiung, C.A., Shu, X.O., Seow, W.J., Wang, Z., Matsuo, K., Hong, Y.C., Seow, A., Wu, C., Hosgood, H.D., 3rd et al. (2015) 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*, **137**, 311–319.
  58. Seow, W.J., Cawthon, R.M., Purdue, M.P., Hu, W., Gao, Y.T., Huang, W.Y., Weinstein, S.J., Ji, B.T., Virtamo, J., Hosgood, H.D., 3rd et al. (2014) Telomere length in white blood cell DNA and lung cancer: a pooled analysis of three prospective cohorts. *Cancer Res.*, **74**, 4090–4098.
  59. Dong, J., Cheng, Y., Zhu, M., Wen, Y., Wang, C., Wang, Y., Geng, L., Shen, W., Liu, J., Li, Z. et al. (2017) Fine mapping of chromosome 5p15.33 identifies novel lung cancer susceptibility loci in Han Chinese. *Int. J. Cancer*, **141**, 447–456.
  60. Kachuri, L., Amos, C.I., McKay, J.D., Johansson, M., Vineis, P., Bueno-de-Mesquita, H.B., Boutron-Ruault, M.C., Johansson, M., Quiros, J.R., Sieri, S. et al. (2016) Fine mapping of chromosome 5p15.33 based on a targeted deep sequencing and high density genotyping identifies novel lung cancer susceptibility loci. *Carcinogenesis*, **37**, 96–105.
  61. Zhao, D.P., Yang, C.L., Zhou, X., Ding, J.A. and Jiang, G.N. (2014) Association between CLPTM1L polymorphisms (rs402710 and rs401681) and lung cancer susceptibility: evidence from 27 case-control studies. *Mol. Genet. Genom.*, **289**, 1001–1012.
  62. Li, Z., Pu, Z., Fan, J., Li, N., Zhu, M., Zhang, J., Wang, Y., Geng, L., Cheng, Y., Ma, H. et al. (2018) Fine mapping in TERT-CLPTM1L region identified three independent lung cancer susceptibility signals: a large-scale multi-ethnic population study. *Mol. Carcinog.*, **57**, 1289–1299.
  63. James, M.A., Wen, W., Wang, Y., Byers, L.A., Heymach, J.V., Coombes, K.R., Girard, L., Minna, J. and You, M. (2012) Functional characterization of CLPTM1L as a lung cancer risk candidate gene in the 5p15.33 locus. *PLoS One*, **7**, e36116.

64. James, M.A., Vikis, H.G., Tate, E., Rymaszewski, A.L. and You, M. (2014) CRR9/CLPTM1L regulates cell survival signaling and is required for Ras transformation and lung tumorigenesis. *Cancer Res.*, **74**, 1116–1127.
65. James, M.A., Vikis, H.G., You, M. (2016) Targeting CLPTM1L by RNA interference for treatment and prevention of cancer. *Patent US9265789B2*.
66. Janeway, C.A., Jr., Travers, P., Walport, M. et al. (2001) *Immunobiology: The Immune System in Health and Disease*. Garland Science, New York.
67. Walsh, E.C., Mather, K.A., Schaffner, S.F., Farwell, L., Daly, M.J., Patterson, N., Cullen, M., Carrington, M., Bugawan, T.L., Erlich, H. et al. (2003) An integrated haplotype map of the human major histocompatibility complex. *Am. J. Hum. Genet.*, **73**, 580–590.
68. de Bakker, P.I., McVean, G., Sabeti, P.C., Miretti, M.M., Green, T., Marchini, J., Ke, X., Monsuur, A.J., Whittaker, P., Delgado, M. et al. (2006) A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. *Nat. Genet.*, **38**, 1166–1172.
69. Sanchez-Mazas, A., Buhler, S. and Nunes, J.M. (2013) A new HLA map of Europe: regional genetic variation and its implication for peopling history, disease-association studies and tissue transplantation. *Hum. Hered.*, **76**, 162–177.
70. Ferreira-Iglesias, A., Lesseur, C., McKay, J., Hung, R.J., Han, Y., Zong, X., Christiani, D., Johansson, M., Xiao, X., Li, Y. et al. (2018) Fine mapping of MHC region in lung cancer highlights independent susceptibility loci by ethnicity. *Nat. Commun.*, **9**, 3927.
71. Lesseur, C., Diergaard, B., Olshan, A.F., Wunsch-Filho, V., Ness, A.R., Liu, G., Lacko, M., Eluf-Neto, J., Franceschi, S., Lagiou, P. et al. (2016) Genome-wide association analyses identify new susceptibility loci for oral cavity and pharyngeal cancer. *Nat. Genet.*, **48**, 1544–1550.
72. Chen, D., Juko-Pecirep, I., Hammer, J., Ivansson, E., Enroth, S., Gustavsson, I., Feuk, L., Magnusson, P.K., McKay, J.D., Wilander, E. et al. (2013) Genome-wide association study of susceptibility loci for cervical cancer. *J. Natl. Cancer Inst.*, **105**, 624–633.
73. Qin, N., Wang, C., Zhu, M., Lu, Q., Ma, Z., Huang, M., Dai, J., Ma, H., Jin, G., Hu, Z. et al. (2017) Fine-mapping the MHC region in Asian populations identified novel variants modifying susceptibility to lung cancer. *Lung Cancer*, **112**, 169–175.
74. Zhu, M., Fan, J., Zhang, C., Xu, J., Yin, R., Zhang, E., Wang, Y., Ji, M., Sun, Q., Dai, J. et al. (2021) A cross-tissue transcriptome-wide association study identifies novel susceptibility genes for lung cancer in Chinese populations. *Hum. Mol. Genet.*, **30**, 1666–1676.
75. Boyle, A.P., Hong, E.L., Hariharan, M., Cheng, Y., Schaub, M.A., Kasowski, M., Karczewski, K.J., Park, J., Hitz, B.C., Weng, S. et al. (2012) Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.*, **22**, 1790–1797.
76. Lynch, H.T., Snyder, C. and Casey, M.J. (2013) Hereditary ovarian and breast cancer: what have we learned? *Ann. Oncol.*, **24**, viii83–viii95.
77. Liu, Y., Lusk, C.M., Cho, M.H., Silverman, E.K., Qiao, D., Zhang, R., Scheurer, M.E., Kheradmand, F., Wheeler, D.A., Tsavachidis, S. et al. (2018) Rare variants in known susceptibility loci and their contribution to risk of lung cancer. *J. Thorac. Oncol.*, **13**, 1483–1495.
78. Rafnar, T., Sigurjonsdottir, G.R., Stacey, S.N., Halldorsson, G., Sulem, P., Pardo, L.M., Helgason, H., Sigurdsson, S.T., Gudjonsson, T., Tryggvadottir, L. et al. (2018) Association of BRCA2 K3326\* with small cell lung cancer and squamous cell cancer of the skin. *J. Natl. Cancer Inst.*, **110**, 967–974.
79. Davies, O.R. and Pellegrini, L. (2007) Interaction with the BRCA2 C terminus protects RAD51-DNA filaments from disassembly by BRC repeats. *Nat. Struct. Mol. Biol.*, **14**, 475–483.
80. McAllister, K.A., Bennett, L.M., Houle, C.D., Ward, T., Malphurs, J., Collins, N.K., Cachafeiro, C., Haseman, J., Goulding, E.H., Bunch, D. et al. (2002) Cancer susceptibility of mice with a homozygous deletion in the COOH-terminal domain of the Brca2 gene. *Cancer Res.*, **62**, 990–994.
81. Spain, B.H., Larson, C.J., Shihabuddin, L.S., Gage, F.H. and Verma, I.M. (1999) Truncated BRCA2 is cytoplasmic: implications for cancer-linked mutations. *Proc. Natl. Acad. Sci. U. S. A.*, **96**, 13920–13925.
82. Delahaye-Sourdeix, M., Anantharaman, D., Timofeeva, M.N., Gaborieau, V., Chabrier, A., Vallee, M.P., Lagiou, P., Holcatova, I., Richiardi, L., Kjaerheim, K. et al. (2015) A rare truncating BRCA2 variant and genetic susceptibility to upper aerodigestive tract cancer. *J. Natl. Cancer Inst.*, **107**, djv037.
83. Lesseur, C., Ferreira-Iglesias, A., McKay, J.D., Bosse, Y., Johansson, M., Gaborieau, V., Landi, M.T., Christiani, D.C., Caporaso, N.C., Bojesen, S.E. et al. (2021) Genome-wide association meta-analysis identifies pleiotropic risk loci for aerodigestive squamous cell cancers. *PLoS Genet.*, **17**, e1009254.
84. Cai, Z., Chehab, N.H. and Pavletich, N.P. (2009) Structure and activation mechanism of the CHK2 DNA damage checkpoint kinase. *Mol. Cell*, **35**, 818–829.
85. Michailidou, K., Lindstrom, S., Dennis, J., Beesley, J., Hui, S., Kar, S., Lemacon, A., Soucy, P., Glubb, D., Rostamianfar, A. et al. (2017) Association analysis identifies 65 new breast cancer risk loci. *Nature*, **551**, 92–94.
86. Han, F.F., Guo, C.L. and Liu, L.H. (2013) The effect of CHEK2 variant I157T on cancer susceptibility: evidence from a meta-analysis. *DNA Cell Biol.*, **32**, 329–335.
87. Bell, D.W., Varley, J.M., Szydlowski, T.E., Kang, D.H., Wahrer, D.C., Shannon, K.E., Lubratovich, M., Verselis, S.J., Isselbacher, K.J., Fraumeni, J.F. et al. (1999) Heterozygous germ line hCHK2 mutations in Li-Fraumeni syndrome. *Science*, **286**, 2528–2531.
88. Brennan, P., McKay, J., Moore, L., Zaridze, D., Mukeria, A., Szeszenia-Dabrowska, N., Lissowska, J., Rudnai, P., Fabianova, E., Mates, D. et al. (2007) Uncommon CHEK2 mis-sense variant and reduced risk of tobacco-related cancers: case control study. *Hum. Mol. Genet.*, **16**, 1794–1801.
89. Timofeeva, M.N., Hung, R.J., Rafnar, T., Christiani, D.C., Field, J.K., Bickeboller, H., Risch, A., McKay, J.D., Wang, Y., Dai, J. et al. (2012) Influence of common genetic variation on lung cancer risk: meta-analysis of 14 900 cases and 29 485 controls. *Hum. Mol. Genet.*, **21**, 4980–4995.
90. Shi, J., Chatterjee, N., Rotunno, M., Wang, Y., Pesatori, A.C., Consonni, D., Li, P., Wheeler, W., Broderick, P., Henrion, M. et al. (2012) Inherited variation at chromosome 12p13.33, including RAD52, influences the risk of squamous cell lung carcinoma. *Cancer Discov.*, **2**, 131–139.
91. Lieberman, R., Xiong, D., James, M., Han, Y., Amos, C.I., Wang, L. and You, M. (2016) Functional characterization of RAD52 as a lung cancer susceptibility gene in the 12p13.33 locus. *Mol. Carcinog.*, **55**, 953–963.
92. Rossi, M.J., DiDomenico, S.F., Patel, M. and Mazin, A.V. (2021) RAD52: paradigm of synthetic lethality and new developments. *Front. Genet.*, **12**, 780293.
93. Feng, Z., Scott, S.P., Bussen, W., Sharma, G.G., Guo, G., Pandita, T.K. and Powell, S.N. (2011) Rad52 inactivation is synthetically

- lethal with BRCA2 deficiency. *Proc. Natl. Acad. Sci. U. S. A.*, **108**, 686–691.
94. Lok, B.H., Carley, A.C., Tchang, B. and Powell, S.N. (2013) RAD52 inactivation is synthetically lethal with deficiencies in BRCA1 and PALB2 in addition to BRCA2 through RAD51-mediated homologous recombination. *Oncogene*, **32**, 3552–3558.
  95. Ji, X., Mukherjee, S., Landi, M.T., Bosse, Y., Joubert, P., Zhu, D., Gorlov, I., Xiao, X., Han, Y., Gorlova, O. et al. (2020) Protein-altering germline mutations implicate novel genes related to lung cancer development. *Nat. Commun.*, **11**, 2220.
  96. Baretic, D. and Williams, R.L. (2014) PIKKs—the solenoid nest where partners and kinases meet. *Curr. Opin. Struct. Biol.*, **29**, 134–142.
  97. Yan, Z., Tong, X., Ma, Y., Liu, S., Yang, L., Yang, X., Yang, X., Bai, M. and Fan, H. (2017) Association between ATM gene polymorphisms, lung cancer susceptibility and radiation-induced pneumonitis: a meta-analysis. *BMC Pulm. Med.*, **17**, 205.
  98. Liu, Y., Xia, J., McKay, J., Tsavachidis, S., Xiao, X., Spitz, M.R., Cheng, C., Byun, J., Hong, W., Li, Y. et al. (2021) Rare deleterious germline variants and risk of lung cancer. *NPJ Precis. Oncol.*, **5**, 12.
  99. Esai Selvan, M., Zauderer, M.G., Rudin, C.M., Jones, S., Mukherjee, S., Offit, K., Onel, K., Rennert, G., Velculescu, V.E., Lipkin, S.M. et al. (2020) Inherited rare, deleterious variants in ATM increase lung adenocarcinoma risk. *J. Thorac. Oncol.*, **15**, 1871–1879.
  100. Hao, K., Bosse, Y., Nickle, D.C., Pare, P.D., Postma, D.S., Laviolette, M., Sandford, A., Hackett, T.L., Daley, D., Hogg, J.C. et al. (2012) Lung eQTLs to help reveal the molecular underpinnings of asthma. *PLoS Genet.*, **8**, e1003029.
  101. Flores, E.R. (2007) The roles of p63 in cancer. *Cell Cycle*, **6**, 300–304.
  102. Petitjean, A., Ruptier, C., Tribollet, V., Hautefeuille, A., Chardon, F., Cavard, C., Puisieux, A., Hainaut, P. and Caron de Fromental, C. (2008) Properties of the six isoforms of p63: p53-like regulation in response to genotoxic stress and cross talk with DeltaNp73. *Carcinogenesis*, **29**, 273–281.
  103. Campbell, J.D., Yau, C., Bowly, R., Liu, Y., Brennan, K., Fan, H., Taylor, A.M., Wang, C., Walter, V., Akbani, R. et al. (2018) Genomic, pathway network, and immunologic features distinguishing squamous carcinomas. *Cell Rep.*, **23**, 194–212 e196.
  104. Moses, M.A., George, A.L., Sakakibara, N., Mahmood, K., Ponnamperuma, R.M., King, K.E. and Weinberg, W.C. (2019) Molecular mechanisms of p63-mediated squamous cancer pathogenesis. *Int. J. Mol. Sci.*, **20**, 3590.
  105. Schaid, D.J., Chen, W. and Larson, N.B. (2018) From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat. Rev. Genet.*, **19**, 491–504.
  106. Wong, A.K., Sealfon, R.S.G., Theesfeld, C.L. and Troyanskaya, O.G. (2021) Decoding disease: from genomes to networks to phenotypes. *Nat. Rev. Genet.*, **22**, 774–790.
  107. Lam, M., Chen, C.Y., Li, Z., Martin, A.R., Bryois, J., Ma, X., Gaspar, H., Ikeda, M., Benyamin, B., Brown, B.C. et al. (2019) Comparative genetic architectures of schizophrenia in east Asian and European populations. *Nat. Genet.*, **51**, 1670–1678.
  108. Chen, J., Spracklen, C.N., Marenne, G., Varshney, A., Corbin, L.J., Luan, J., Willems, S.M., Wu, Y., Zhang, X., Horikoshi, M. et al. (2021) The trans-ancestral genomic architecture of glycemic traits. *Nat. Genet.*, **53**, 840–860.
  109. Melnikov, A., Zhang, X., Rogov, P., Wang, L. and Mikkelsen, T.S. (2014) Massively parallel reporter assays in cultured mammalian cells. *J. Vis. Exp.*, **90**, 51719.
  110. Ulirsch, J.C., Nandakumar, S.K., Wang, L., Giani, F.C., Zhang, X., Rogov, P., Melnikov, A., McDonel, P., Do, R., Mikkelsen, T.S. et al. (2016) Systematic functional dissection of common genetic variation affecting red blood cell traits. *Cell*, **165**, 1530–1545.
  111. Kheradpour, P., Ernst, J., Melnikov, A., Rogov, P., Wang, L., Zhang, X., Alston, J., Mikkelsen, T.S. and Kellis, M. (2013) Systematic dissection of regulatory motifs in 2000 predicted human enhancers using a massively parallel reporter assay. *Genome Res.*, **23**, 800–811.
  112. Melnikov, A., Murugan, A., Zhang, X., Tesileanu, T., Wang, L., Rogov, P., Feizi, S., Gnirke, A., Callan, C.G., Jr., Kinney, J.B. et al. (2012) Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nat. Biotechnol.*, **30**, 271–277.
  113. Giambartolomei, C., Vukcevic, D., Schadt, E.E., Franke, L., Hingorani, A.D., Wallace, C. and Plagnol, V. (2014) Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.*, **10**, e1004383.
  114. Hormozdiari, F., van de Bunt, M., Segre, A.V., Li, X., Joo, J.W.J., Bilow, M., Sul, J.H., Sankaraman, S., Pasaniuc, B. and Eskin, E. (2016) Colocalization of GWAS and eQTL signals detects target genes. *Am. J. Hum. Genet.*, **99**, 1245–1260.
  115. Gamazon, E.R., Wheeler, H.E., Shah, K.P., Mozaffari, S.V., Aquino-Michaels, K., Carroll, R.J., Eyler, A.E., Denny, J.C., Consortium, G.T., Nicolae, D.L. et al. (2015) A gene-based association method for mapping traits using reference transcriptome data. *Nat. Genet.*, **47**, 1091–1098.
  116. Gusev, A., Ko, A., Shi, H., Bhatia, G., Chung, W., Penninx, B.W., Jansen, R., de Geus, E.J., Boomsma, D.I., Wright, F.A. et al. (2016) Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.*, **48**, 245–252.
  117. Hu, Y., Li, M., Lu, Q., Weng, H., Wang, J., Zekavat, S.M., Yu, Z., Li, B., Gu, J., Muchnik, S. et al. (2019) A statistical framework for cross-tissue transcriptome-wide association analysis. *Nat. Genet.*, **51**, 568–576.
  118. Travaglini, K.J., Nabhan, A.N., Penland, L., Sinha, R., Gillich, A., Sit, R.V., Chang, S., Conley, S.D., Mori, Y., Seita, J. et al. (2020) A molecular cell atlas of the human lung from single-cell RNA sequencing. *Nature*, **587**, 619–625.
  119. Mainardi, S., Mijimolle, N., Francoz, S., Vicente-Duenas, C., Sanchez-Garcia, I. and Barbacid, M. (2014) Identification of cancer initiating cells in K-Ras driven lung adenocarcinoma. *Proc. Natl. Acad. Sci. U. S. A.*, **111**, 255–260.
  120. Sutherland, K.D., Song, J.Y., Kwon, M.C., Proost, N., Zevenhoven, J. and Berns, A. (2014) Multiple cells-of-origin of mutant K-Ras-induced mouse lung adenocarcinoma. *Proc. Natl. Acad. Sci. U. S. A.*, **111**, 4952–4957.
  121. Hanna, J.M. and Onaitis, M.W. (2013) Cell of origin of lung cancer. *J. Carcinog.*, **12**, 6.
  122. Lavin, Y., Kobayashi, S., Leader, A., Amir, E.D., Elefant, N., Bigenwald, C., Remark, R., Sweeney, R., Becker, C.D., Levine, J.H. et al. (2017) Innate immune landscape in early lung adenocarcinoma by paired single-cell analyses. *Cell*, **169**, 750–765 e717.
  123. Vieira Braga, F.A., Kar, G., Berg, M., Carpaij, O.A., Polanski, K., Simon, L.M., Brouwer, S., Gomes, T., Hesse, L., Jiang, J. et al. (2019) A cellular census of human lungs identifies novel cell states in health and in asthma. *Nat. Med.*, **25**, 1153–1163.
  124. Iotchkova, V., Ritchie, G.R.S., Geihs, M., Morganello, S., Min, J.L., Walter, K., Timpson, N.J., Consortium, U.K., Dunham, I., Birney, E. et al. (2019) GARFIELD classifies disease-relevant genomic features through integration of functional annotations with association signals. *Nat. Genet.*, **51**, 343–353.
  125. Mumbach, M.R., Satpathy, A.T., Boyle, E.A., Dai, C., Gowen, B.G., Cho, S.W., Nguyen, M.L., Rubin, A.J., Granja, J.M., Kazane, K.R.



- et al. (2017) Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. *Nat. Genet.*, **49**, 1602–1612.
126. Hughes, J.R., Roberts, N., McGowan, S., Hay, D., Giannoulatou, E., Lynch, M., De Gobbi, M., Taylor, S., Gibbons, R. and Higgs, D.R. (2014) Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nat. Genet.*, **46**, 205–212.
127. Smemo, S., Tena, J.J., Kim, K.H., Gamazon, E.R., Sakabe, N.J., Gomez-Marin, C., Aneas, I., Credidio, F.L., Sobreira, D.R., Wasserman, N.F. et al. (2014) Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature*, **507**, 371–375.
128. Baxter, J.S., Leavy, O.C., Dryden, N.H., Maguire, S., Johnson, N., Fedele, V., Simigdala, N., Martin, L.A., Andrews, S., Wingett, S.W. et al. (2018) Capture hi-C identifies putative target genes at 33 breast cancer risk loci. *Nat. Commun.*, **9**, 1028.
129. Beesley, J., Sivakumaran, H., Moradi Marjaneh, M., Lima, L.G., Hillman, K.M., Kaufmann, S., Tuano, N., Hussein, N., Ham, S., Mukhopadhyay, P. et al. (2020) Chromatin interactome mapping at 139 independent breast cancer risk signals. *Genome Biol.*, **21**, 8.
130. Cano-Gamez, E. and Trynka, G. (2020) From GWAS to function: using functional genomics to identify the mechanisms underlying complex diseases. *Front. Genet.*, **11**, 424.
131. Rao, S., Yao, Y. and Bauer, D.E. (2021) Editing GWAS: experimental approaches to dissect and exploit disease-associated genetic variation. *Genome Med.*, **13**, 41.
132. Lichou, F. and Trynka, G. (2020) Functional studies of GWAS variants are gaining momentum. *Nat. Commun.*, **11**, 6283.
133. Ray, J.P., de Boer, C.G., Fulco, C.P., Lareau, C.A., Kanai, M., Ulirsch, J.C., Tewhey, R., Ludwig, L.S., Reilly, S.K., Bergman, D.T. et al. (2020) Prioritizing disease and trait causal variants at the TNFAIP3 locus using functional and genomic features. *Nat. Commun.*, **11**, 1237.
134. Ihry, R.J., Worringer, K.A., Salick, M.R., Frias, E., Ho, D., Theriault, K., Kommineni, S., Chen, J., Sondey, M., Ye, C. et al. (2018) p53 inhibits CRISPR-Cas9 engineering in human pluripotent stem cells. *Nat. Med.*, **24**, 939–946.
135. Hanahan, D. and Weinberg, R.A. (2011) Hallmarks of cancer: the next generation. *Cell*, **144**, 646–674.
136. Freedman, M.L., Monteiro, A.N., Gayther, S.A., Coetzee, G.A., Risch, A., Plass, C., Casey, G., De Biasi, M., Carlson, C., Duggan, D. et al. (2011) Principles for the post-GWAS functional characterization of cancer risk loci. *Nat. Genet.*, **43**, 513–518.
137. Kachuri, L., Graff, R.E., Smith-Byrne, K., Meyers, T.J., Rashkin, S.R., Ziv, E., Witte, J.S. and Johansson, M. (2020) Pan-cancer analysis demonstrates that integrating polygenic risk scores with modifiable risk factors improves risk prediction. *Nat. Commun.*, **11**, 6084.
138. Hu, Y., Lu, Q., Powles, R., Yao, X., Yang, C., Fang, F., Xu, X. and Zhao, H. (2017) Leveraging functional annotations in genetic risk prediction for human complex diseases. *PLoS Comput. Biol.*, **13**, e1005589.
139. Amariuta, T., Ishigaki, K., Sugishita, H., Ohta, T., Koido, M., Dey, K.K., Matsuda, K., Murakami, Y., Price, A.L., Kawakami, E. et al. (2020) Improving the trans-ancestry portability of polygenic risk scores by prioritizing variants in predicted cell-type-specific regulatory elements. *Nat. Genet.*, **52**, 1346–1354.
140. Qian, D.C., Han, Y., Byun, J., Shin, H.R., Hung, R.J., McLaughlin, J.R., Landi, M.T., Seminara, D. and Amos, C.I. (2016) A novel pathway-based approach improves lung cancer risk prediction using germline genetic variations. *Cancer Epidemiol. Biomark. Prev.*, **25**, 1208–1215.