



OPEN

The shared genetic architecture between epidemiological and behavioral traits with lung cancer

Rowland W. Pettit¹, Jinyoung Byun^{1,2}, Younghun Han^{1,2}, Quinn T. Ostrom³, Jacob Edelson⁴, Kyle M. Walsh³, Melissa L. Bondy⁵, Rayjean J. Hung^{6,7}, James D. McKay⁸ & Christopher I. Amos^{1,2,9}✉

The complex polygenic nature of lung cancer is not fully characterized. Our study seeks to identify novel phenotypes associated with lung cancer using cross-trait linkage disequilibrium score regression (LDSR). We measured pairwise genetic correlation (r_g) and SNP heritability (h^2) between 347 traits and lung cancer risk using genome-wide association study summary statistics from the UKBB and OncoArray consortium. Further, we conducted analysis after removing genomic regions previously associated with smoking behaviors to mitigate potential confounding effects. We found significant negative genetic correlations between lung cancer risk and dietary behaviors, fitness metrics, educational attainment, and other psychosocial traits. Alcohol taken with meals ($r_g = -0.41$, $h^2 = 0.10$, $p = 1.33 \times 10^{-16}$), increased fluid intelligence scores ($r_g = -0.25$, $h^2 = 0.22$, $p = 4.54 \times 10^{-8}$), and the age at which full time education was completed ($r_g = -0.45$, $h^2 = 0.11$, $p = 1.24 \times 10^{-20}$) demonstrated negative genetic correlation with lung cancer susceptibility. The body mass index was positively correlated with lung cancer risk ($r_g = 0.20$, $h^2 = 0.25$, $p = 2.61 \times 10^{-9}$). This analysis reveals shared genetic architecture between several traits and lung cancer predisposition. Future work should test for causal relationships and investigate common underlying genetic mechanisms across these genetically correlated traits.

In 2020 ~230,000 lung cancer cases will be diagnosed in the US, and ~140,000 people will die from their disease¹. In total, this morbidity ranks lung cancer as the leading cause of cancer-related deaths in the United States. Our current understanding of lung cancer is that it is a multi-factorial disease in which tumorigenesis results from inherited genetic variants^{2,3}, sustained environmental exposures⁴, and stochastic somatic mutations⁵. Environmental exposures associated with an increased risk of developing lung cancer are numerous and include cigarette smoke⁶, radon⁷, individual diet⁸, pollution in the atmosphere⁹, metallurgy¹⁰, and indoor pollution from cooking or heating with solid fuels¹¹. The most significant contributor to lung cancer development is due to tobacco smoking¹². However, clustering of lung cancer cases in families beyond a level that could be explained by shared environmental exposures to tobacco smoke or pollution supports a role of genetic factors contributing to disease risk^{13–17}. Investigations into the precise tumorigenic mechanisms behind the familial aggregation of lung cancer are complicated by genetic polygenicity^{18,19}, whereby a combination of multiple genes contributes to risk.

Genome-wide association studies (GWAS), which examine millions of single nucleotide polymorphisms (SNPs) for association with a trait of interest, are helpful for deciphering the genetic architecture of complex diseases². GWAS is not without limitations, and behavioral traits that are genetically influenced can mediate observed associations between SNPs and lung cancer risk^{20–23}. GWAS analysis can be further confounded when

¹Institute for Clinical and Translational Research, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA. ²Section of Epidemiology and Population Sciences, Department of Medicine, Baylor College of Medicine, Houston, TX, USA. ³Duke Cancer Institute, Duke University Medical Center, Durham, NC, USA. ⁴Department of Biomedical Data Science, School of Medicine, Stanford University, Stanford, CA, USA. ⁵Department of Epidemiology and Population Health, School of Medicine, Stanford University, Stanford, CA, USA. ⁶Lunenfeld-Tanenbaum Research Institute, Sinai Health System, Toronto, Canada. ⁷Division of Epidemiology, Dalla Lana School of Public Health, University of Toronto, Toronto, Canada. ⁸Section of Genetics, International Agency for Research on Cancer, World Health Organization, Lyon, France. ⁹Dan L Duncan Comprehensive Cancer Center, Baylor College of Medicine, Houston, TX, USA. ✉email: Chris.Amos@bcm.edu

unknown population stratification or cryptic relatedness exists in the underlying data²⁴. Prior GWAS investigations in lung cancer have revealed unique loci with strong statistical significance, yet, these regional associations vary across histological subtypes of lung cancer². On top of heterogeneity between histological subtypes, known lung cancer risk loci only account for a minor proportion of the total estimated heritability of lung cancer, indicating a substantial proportion of the heritable causes²⁵ of lung cancer remains unidentified.

A more comprehensive approach to understanding tumorigenic mechanisms may be fruitful. Focused work into understanding the genetic architecture behind disease co-development may be more informative than studying individual phenotypes^{26,27}. A knowledge gap exists today to quantify the extent that other diseases, environmental exposures, and phenotypic traits correlate with a predisposition to lung cancer. A novel regression statistical framework, known as cross-trait linkage disequilibrium score regression (LDSR), may be employed to fill this gap in knowledge. LDSR uses GWAS summary statistics to identify genome-wide genetic correlations between phenotypes of interest²⁸. The similarity of measured SNP effect estimates reported by GWAS summary statistics are compared between traits. LDSR allows for accurate calculations of genetic co-correlation (r_g) between phenotypes while minimizing effects from selection biases in the recruitment of comparable controls from the same source population²⁴. Use of this method can identify correlations in the genetic architecture between traits, allowing etiological insights to be gleaned.

Here, we quantify the association between genetically influenced epidemiological and behavioral traits and the risk of lung cancer. We use summary statistics generated by prior lung cancer GWAS and use LDSR to estimate cross-trait genetic correlations with lung cancer. We additionally evaluate how these traits correlate with each of the major histological subtypes of lung cancer—adenocarcinoma, squamous cell carcinoma, and small cell carcinoma, and further evaluate associations in ever- and never-smokers. We aimed to confirm prior associations with lung cancer and to identify novel phenotypic associations from GWAS datasets.

Methods

Summary statistics for lung cancer. This work is a continuation of efforts conducted by the Transdisciplinary Research of Cancer in Lung of the International Lung Cancer Consortium (TRICL-ILCCO)²⁹ and the OncoArray Consortium³⁰. The TRICL-OncoArray Consortium has previously published GWAS summary statistics results after a meta-analysis of lung cancer GWAS. The complete methods have been published previously^{29,30}, but are presented here in brief. Lung cancer patients and healthy controls with no personal lung cancer history were recruited after individual institutional IRB approval and informed consent for genotyping. Genotyping occurred using the Illumina OncoArray-500K BeadChip of 533,631 SNPs. Standard quality control measures were implemented to exclude underperforming samples and SNPs²⁹. Individuals and SNPs with genotyping call-rates < 95% were removed. Genotype imputation was conducted using the reference dataset of the 1000 Genomes Project Phase 3 (October 2014). The more common variant was included during the imputation process for positions with > 2 alleles. After imputation and quality control processes, 502,933 SNPs from 29,266 lung cancer patients and 56,450 healthy controls of European ancestry were incorporated into a meta-analysis²⁹. Amongst the lung cancer cases, 11,273 cases of adenocarcinoma, 2,664 cases of small cell carcinoma, and 7,426 cases of squamous cell carcinoma were represented as histological subtypes (Supplementary Table 1). We obtained and utilized the summary statistics from the TRICL-ILCCO GWAS meta-analysis²⁹ regarding lung cancer, the histological subtypes of lung cancer, and summary statistics for 'ever' vs. 'never' smoking status subcohorts.

Phenotype and exposure accession with United Kingdom Biobank genome-wide association studies. GWAS summary statistics for cross-trait LDSR analyses were obtained from the United Kingdom Biobank (UKBB). The UKBB is a national and international health repository³¹. Since its inception in 2006, the UKBB has collected clinical and genotypic data for 500,000 adult participants across 22 sites in the United Kingdom³¹. Participants in this longitudinal project were age 40–69 at enrollment. Initial relevant information is gathered by clinical exam, questionnaire, and biospecimen sampling. Participants will be followed for 30+ years. Periodically, follow-up health data are obtained by a linked unique encrypted identifier with electronic health records from the UK National Health Service (NHS). Each of the > 500,000 participants in the UKBB has been genotyped, 90% of which were genotyped using a custom Affymetrix UKBB Axiom array. This array assayed ~ 850,000 variants across the genome, which were used to impute 9.1 million SNPs with satisfactory quality control measures in place. These imputation procedures are conducted by the Wellcome Trust Center for Human Genetics and are conducted internally at the UKBB before the data release. GWAS was conducted from these imputed data, and summary level statistics were made publicly available (https://nealelab.github.io/UKBB_ldsc/downloads.html#reference_files). We obtained all of our GWAS summary statistics from the second batch of UKBB GWAS results published online and updated in August 2018.

Harmonization and quality control with SNP filtering. We harmonized the obtained publicly available GWAS summary statistics. Our final dataset included summary statistics for selected epidemiological and individual lifestyle traits, including alcohol use and fitness activity levels and routines. The final dataset also included biometric measurements, including BMI and body fat percentage measurements. Reported educational attainment, employment status, workplace environment, and psychological experiences were also included. These obtained UKBB summary statistics contained SNP-level effect sizes (beta) for each trait, with Z-scores calculated by dividing SNP effect sizes by their standard error. To harmonize these datasets, and as an additional quality control measure, we filtered the imputed SNPs from the UKBB to include only those autosomal SNPs with a minor allele frequency greater than 0.01 and imputation quality INFO score greater than 0.90. We further

removed SNPs from our harmonized data set that were not in HapMap3 with a minor allele frequency less than 5% in European populations, in line with previously published methods^{24,32}.

Estimating pairwise genetic correlations and heritability. With this information, we estimated genome-wide SNP heritability using LDSR. Additionally, we used LDSR to compute the pairwise genetic correlation between each of the UKBB traits with lung risk from the TRICL-OncoArray consortium. LDSR calculates genetic correlation by regressing the product of SNP z scores ($Z_{UKBB} * Z_{TRICL-OncoArray}$) against the SNP's calculated linkage disequilibrium score²⁴. The slope of this regression accurately estimates the genetic covariance between two traits. Genetic covariance is converted to a genetic correlation between traits by normalizing genetic covariance by the calculated heritability of each of the two compared traits. The heritability of a trait can be thought of as the genetic covariance of a trait with itself and ultimately represents the proportion of a trait that genetic effects can explain²⁸.

LDSR mitigates potential biases from population stratification¹⁹ and cryptic relatedness²⁴ by modeling an intercept term that accounts for any genomic inflation. We applied a cross-trait LDSR model that included an intercept in these analyses to account for hidden biases that may exist between reference and target populations, especially those that may arise due to the instability of linkage disequilibrium scores in European populations and sub-populations^{24,33}.

We used LDSR to calculate the genetic correlations between lung cancer risk and traits of interest. We additionally performed LDSR for each of the histological subtypes of lung cancer, including small cell carcinoma, squamous cell carcinoma, and adenocarcinoma. Further, we performed LDSR between traits of interest and lung cancer risk in ever- and never-smoker subgroups. Individuals who reported having smoked fewer than 100 cigarettes throughout their lives were defined as "never smokers," and those who had smoked more than 100 cigarettes in their life as "ever smokers"²⁹. We stratified both lung cancer cases and controls by smoking status for these analyses.

Removal of known regions related to smoking behaviors. If a trait shows a genetic correlation with lung cancer in LDSR analyses, this does not necessarily imply a causal relationship. Indeed, both the trait and lung cancer risk may be jointly influenced by a third, unmodeled trait that independently influences each. Notably, smoking status has the potential to confound our associations (e.g., the genetic correlation between lung cancer risk and emphysema risk would likely be attributable to the effect of smoking on both diseases)^{34–38}. In addition to stratifying our LDSR analyses by 'never' and 'ever' smoking status as available from the TRICL-OncoArray Consortium, we also excluded genomic loci previously associated with smoking behaviors. A recent meta-analysis quantified the effect of SNPs on several smoking behaviors, including "age of initiation of smoking", "cigarettes per day", "smoking cessation", and "smoking initiation"³⁹. These authors used a conditional analysis method⁴⁰ to identify SNPs independently associated with at least one of these smoking related traits. Applying a predetermined genome-wide significance threshold of $p < 5 \times 10^{-8}$, 467 SNPs were found to be associated with smoking related traits³⁹. We repeated our LDSR analyses after removing each of these 467 smoking-related SNPs from our summary statistics. Specifically, we identified the sentinel variant from the meta-analysis and removed all SNPs within ± 500 kb. SNPs that were filtered at this step appear in Supplementary Table 2, which also annotates the upper and lower bounds of the genomic regions removed. Changes in the number of SNPs included and excluded from this analysis, per histological subgroup and lifetime smoking status, appear in Supplementary Table 3. Quantile–Quantile plots of the p-values observed from the TRICL-OncoArray meta-analysis before and after removing smoking-related SNPs may be appreciated in Supplementary Figure 1.

We summarized and presented these methods graphically in Fig. 1. Multiple comparisons are conducted in executing these methods. We tested 347 traits and associated them to determine their genetic predispositions to develop overall lung cancers, adenocarcinomas, squamous cell carcinomas, and small cell carcinomas. Additionally, we tested these traits for associations in 'never' or 'ever' smoking populations. These 2082 independent tests were conducted twice, before and after the removal of smoking-related SNPs. In total, 4164 comparisons were performed. Using a stringent Bonferroni correction, we set our adjusted P value significance cutoff threshold to be less than 1.2×10^{-5} , or $-\log_{10}(P) > 4.92$. Here we report the trait associations with significance metrics less than the Bonferroni adjustment. In Supplementary Table 5, we present the heritability, genetic correlations, significance values for each comparison conducted. In this table, we further provide LDSR confidence thresholds and heritability thresholds for each UKBB trait. We finally offer a direct uniform resource locator link for each UKBB trait, allowing for ease of inquiry into trait type counts, inclusion criteria, distribution histograms, and other relevant metrics.

Results

Heritability of lung cancer and its histological subtypes. Overall, we found the heritability of lung cancer to be $8.3\% \pm 1.3\%$, which persisted even after smoking-related regions were removed ($6.9 \pm 0.8\%$). Stratifying by 'ever' and 'never' smoking status, we estimate the overall lung cancer heritability to be $10.0 \pm 2.1\%$ in 'ever' smokers and $3.0 \pm 4.8\%$ in 'never' smokers. After removing smoking-related SNPs, the estimated heritability in 'ever' smokers was $7.7 \pm 1.4\%$, and in 'never' smokers was $2.9 \pm 4.7\%$. Stratifying amongst the histological subtypes of lung cancer, and including all SNPs, adenocarcinoma heritability was $6.7 \pm 1.0\%$, small cell lung cancer heritability was $10.5 \pm 1.9\%$, and squamous cell carcinoma of the lung had an estimated heritability of $5.2 \pm 1.1\%$. After removing smoking-related SNPs, heritability estimates fell to $6.2 \pm 0.9\%$ (adenocarcinoma), $9.4 \pm 2.0\%$ (small cell), and $4.4 \pm 0.9\%$ (squamous cell carcinoma of the lung) (Supplementary Table 4). The heritability of each of 347 traits modeled using LDSR appears in Supplementary Table 5.

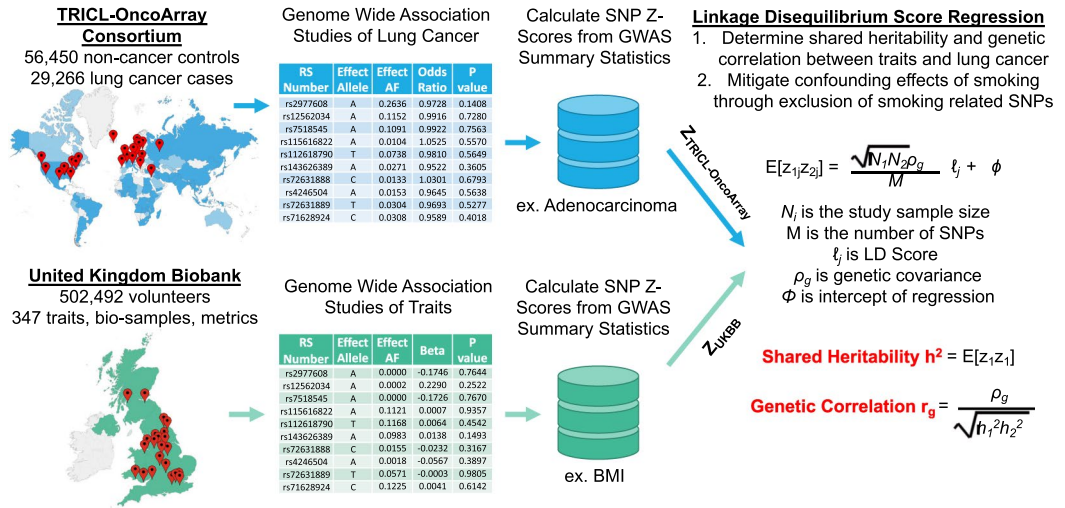


Figure 1. Graphical representation of the analytical workflow, including datasets utilized and analyses performed. Maps rendered with Tableau Desktop, 2021 Tableau Software, LLC, <https://www.tableau.com/>.

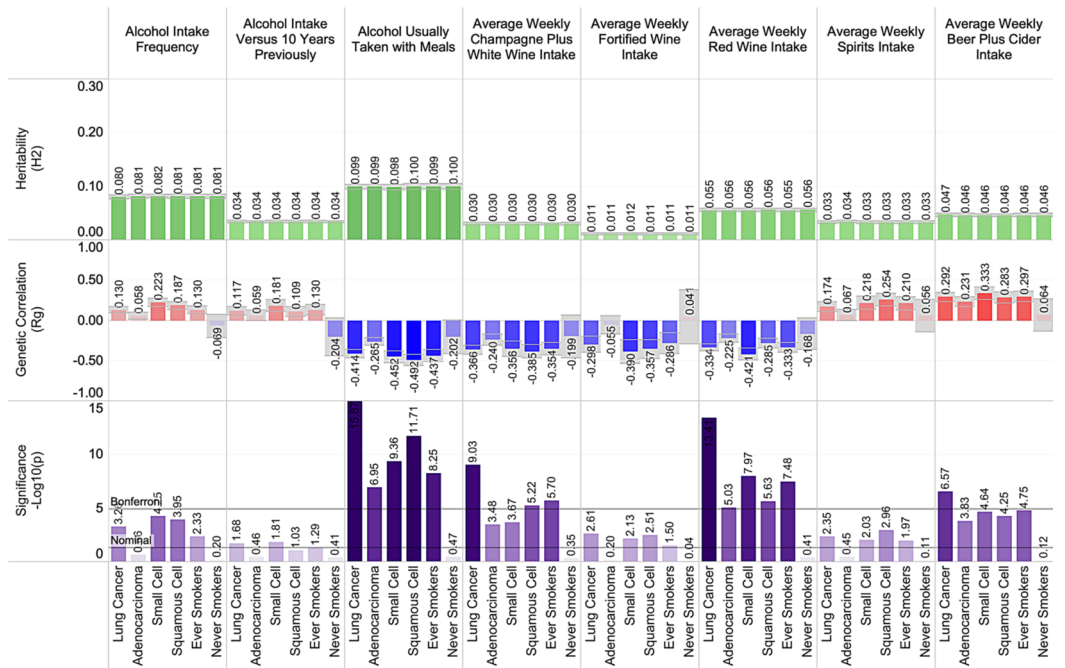


Figure 2. The shared heritability and genetic correlation between alcohol use and lung cancer.

Heritability and genetic correlations between lung cancer and alcohol use. Using cross-trait LDSR, we found that "alcohol usually taken with meals" had an estimated heritability of 0.1 and demonstrated a negative genetic correlation with lung cancer risk across histological subtypes and smoking status. Specifically, "alcohol usually taken with meals" demonstrated a -0.41 genetic correlation (r_g) with all lung cancer ($p = 1.33 \times 10^{-16}$). These findings remained consistent after excluding regions associated with smoking behaviors ($r_{g \dagger} = -0.37$, $p_{\dagger} = 4.46 \times 10^{-13}$). Further investigation revealed that average weekly beer plus cider intake demonstrated positive genetic correlation with lung cancer susceptibility (pre-removal: $r_g = 0.29$, $p = 2.68 \times 10^{-7}$; post-removal: $r_{g \dagger} = 0.29$, $p_{\dagger} = 9.87 \times 10^{-7}$), whereas average weekly red wine intake demonstrated negative genetic correlation with overall lung cancer susceptibility ($r_g = -0.33$, $p = 3.90 \times 10^{-14}$; $r_{g \dagger} = -0.31$, $p_{\dagger} = 3.08 \times 10^{-9}$). These findings were consistent across histological subtypes (Supplementary Table 5). A summary of the significant alcohol-related associations is presented in Fig. 2, and the results from association testing for all traits are included in Supplementary Table 5. All results from our LDSR analyses are publicly hosted and available for interactive viewing at <https://public.tableau.com/profile/rowland.pettit>.

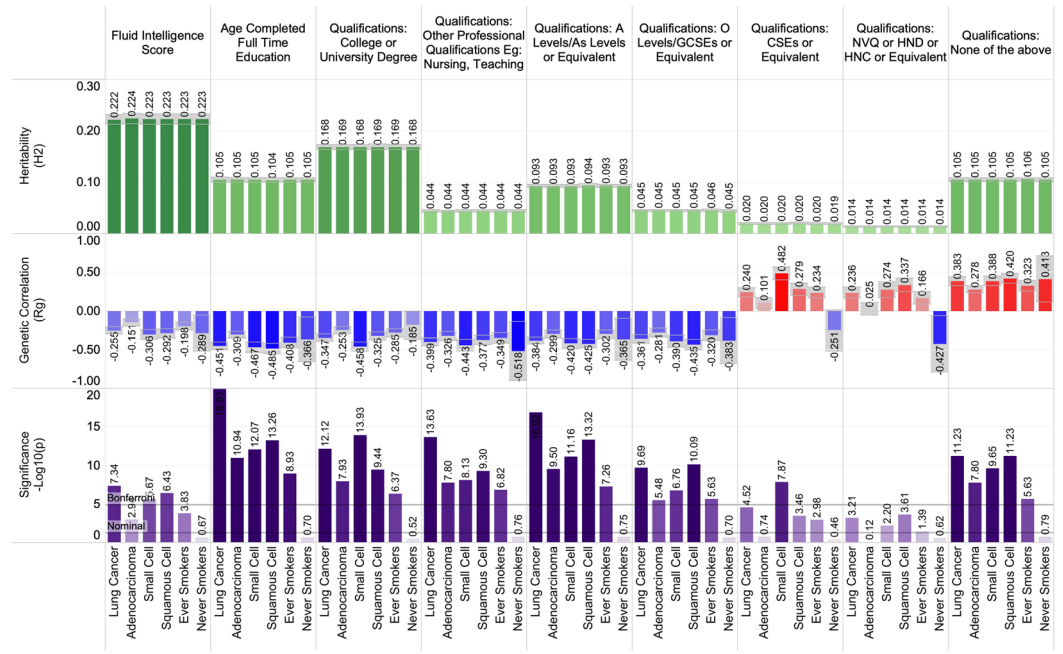


Figure 3. The shared heritability and genetic correlation between education and employment with lung cancer.

Heritability and genetic correlations between lung cancer and education/employment. Education and employment statuses were genetically correlated with lung cancer susceptibility. As this self-reported personal characteristic information comes from the UK biobank, educational ascertainment metrics follow the United Kingdom advanced learning schemas. These analyses found that total years of education, obtaining a college or university degree, earning other advanced professional qualifications such as nursing or teaching roles, gaining “A” level qualification, or earning general certificates of secondary education all demonstrated significant negative genetic correlation with lung cancer susceptibility (Fig. 3). These trends persisted across histological subtypes, but associations were not statistically significant among ‘never’ smokers. Here we highlight reported correlations for ‘age completed full time education’ with overall lung cancer, before and after removal of smoking-associated genomic regions ($r_g = -0.45$, $p = 1.24 \times 10^{-20}$; $r_{g \dagger} = -0.43$, $p_{\dagger} = 1.06 \times 10^{-19}$), small cell lung cancer ($r_g = -0.47$, $p = 8.55 \times 10^{-13}$; $r_{g \dagger} = -0.45$, $p_{\dagger} = 6.20 \times 10^{-9}$), squamous cell lung cancer ($r_g = -0.49$, $p = 5.46 \times 10^{-14}$; $r_{g \dagger} = -0.46$, $p_{\dagger} = 8.40 \times 10^{-10}$), adenocarcinoma ($r_g = -0.31$, $p = 1.15 \times 10^{-11}$; $r_{g \dagger} = -0.27$, $p_{\dagger} = 3.04 \times 10^{-7}$), ‘ever’ smokers ($r_g = -0.41$, $p = 1.17 \times 10^{-9}$; $r_{g \dagger} = -0.41$, $p_{\dagger} = 4.51 \times 10^{-10}$), and ‘never’ smokers ($r_g = -0.37$, $p = 0.20$; $r_{g \dagger} = -0.33$, $p_{\dagger} = 0.11$).

In contrast, obtaining none of the previously mentioned academic qualifications demonstrated a positive genetic correlation with lung cancer susceptibility, which was strongest in overall lung cancer ($r_g = 0.38$, $p = 5.91 \times 10^{-12}$; $r_{g \dagger} = 0.38$, $p_{\dagger} = 3.78 \times 10^{-16}$), and the trend held across histological subtypes and in ‘ever’ smokers. Fluid intelligence scores were genetically correlated with decreased lung cancer susceptibility across all histological and smoking status sub-classifications (overall $r_g = -0.25$, $p = 4.54 \times 10^{-8}$) but did not reach statistical significance in ‘never’ smokers. The calculated Townsend deprivation index⁴¹, which is a metric combining the census demographics of car ownership, household overcrowding, household employment status, and house ownership, demonstrated significant increased genetic predisposition with lung cancers (overall lung cancer $r_g = 0.35$, $p = 1.03 \times 10^{-10}$; $r_{g \dagger} = 0.28$, $p_{\dagger} = 9.61 \times 10^{-6}$). A summary of the significant education and employment-related associations is presented in Fig. 3.

Heritability and genetic correlations between lung cancer and fitness metrics. Measured and reported fitness metrics were genetically correlated with lung cancer susceptibility. Increased body fat percentage, impedance of the whole body, waist circumference, and increased body mass index (BMI) correlated positively with lung cancer susceptibility. Highlighting BMI, positive genetic correlations were observed for overall lung cancer ($r_g = 0.20$, $p = 2.61 \times 10^{-9}$; $r_{g \dagger} = 0.19$, $p_{\dagger} = 3.23 \times 10^{-8}$) as well as across small cell lung carcinoma ($r_g = 0.24$, $p = 3.54 \times 10^{-7}$; $r_{g \dagger} = 0.24$, $p_{\dagger} = 5.27 \times 10^{-5}$), and squamous cell carcinoma ($r_g = 0.27$, $p = 9.91 \times 10^{-10}$; $r_{g \dagger} = 0.26$, $p_{\dagger} = 1.01 \times 10^{-6}$). Similarly, positive genetic correlations were observed between body fat percentage and overall lung cancer ($r_g = 0.17$, $p = 6.11 \times 10^{-7}$; $r_{g \dagger} = 0.17$, $p_{\dagger} = 1.23 \times 10^{-6}$) and squamous cell carcinomas ($r_g = 0.23$, $p = 1.85 \times 10^{-7}$; $r_{g \dagger} = 0.23$, $p_{\dagger} = 9.81 \times 10^{-6}$). Participant-reported activity level traits demonstrated negative genetic correlation with lung cancer susceptibility. Physical activity traits include DIY physical activity in last 4 weeks, exercise such as swimming or cycling in the last 4 weeks, as well as cycling or walking as methods of transport when going to work. Contrarily, having ‘no physical activity in the last 4 weeks’ demonstrated increased genetic correlation with lung cancer susceptibility. We highlight “swimming, cycling, and keeping fit in the last 4 weeks” which demonstrated significant negative genetic correlations with lung cancer susceptibil-

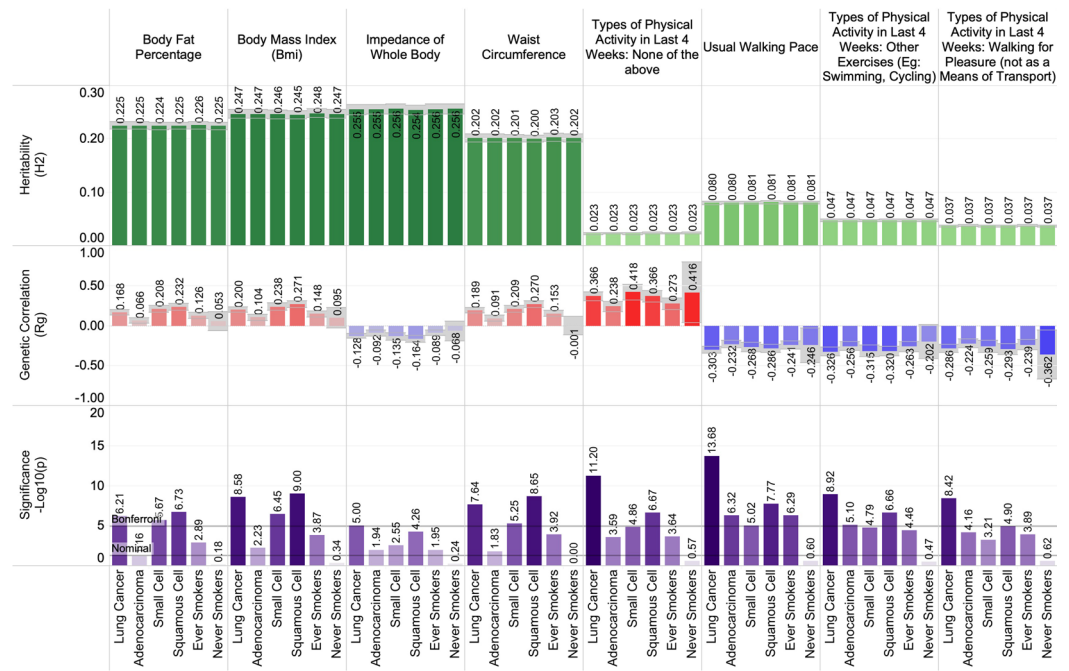


Figure 4. The shared heritability and genetic correlation between fitness with lung cancer.

ity: overall lung cancer ($r_g = -0.33$, $p = 1.20 \times 10^{-9}$; $r_{g\ddagger} = -0.33$, $p_{\ddagger} = 4.02 \times 10^{-10}$), adenocarcinoma ($r_g = -0.26$, $p = 7.92 \times 10^{-6}$; $r_{g\ddagger} = -0.25$, $p_{\ddagger} = 2.05 \times 10^{-5}$), squamous cell carcinoma ($r_g = -0.32$, $p = 2.20 \times 10^{-7}$; $r_{g\ddagger} = -0.33$, $p_{\ddagger} = 2.44 \times 10^{-6}$), and ‘ever’ smokers ($r_g = -0.26$, $p = 3.46 \times 10^{-5}$; $r_{g\ddagger} = -0.29$, $p_{\ddagger} = 1.72 \times 10^{-5}$). A summary of the significant fitness-related associations is presented in Fig. 4.

Heritability and genetic correlations between lung cancer and other specific traits. Significant genetic correlation and heritability estimates were observed for select psychosocial traits. A participant’s reported ‘frequency of depressed mood in the last 2 weeks’ demonstrated a positive genetic correlation with lung cancer susceptibility for overall lung cancer ($r_g = 0.23$, $p = 3.09 \times 10^{-6}$; $r_{g\ddagger} = 0.21$, $p_{\ddagger} = 9.44 \times 10^{-6}$). Specific depressive-related symptoms also demonstrated positive genetic correlation, including the frequency of un-enthusiasm/disinterest in the last 2 weeks: overall lung cancer ($r_g = 0.35$, $p = 1.11 \times 10^{-11}$; $r_{g\ddagger} = 0.32$, $p_{\ddagger} = 1.40 \times 10^{-10}$), adenocarcinoma ($r_g = 0.28$, $p = 8.70 \times 10^{-7}$; $r_{g\ddagger} = 0.26$, $p_{\ddagger} = 7.10 \times 10^{-6}$), and squamous cell carcinoma ($r_g = 0.39$, $p = 2.54 \times 10^{-9}$; $r_{g\ddagger} = 0.33$, $p_{\ddagger} = 1.67 \times 10^{-5}$). Being breastfed as a baby demonstrated a negative genetic correlation with lung cancer susceptibility. The genetic correlations for being breastfed as a baby were significant in the overall lung cancer ($r_g = -0.30$, $p = 3.46 \times 10^{-6}$; $r_{g\ddagger} = -0.30$, $p_{\ddagger} = 3.50 \times 10^{-5}$). In female only traits, both age at first live birth and age started oral contraceptive demonstrated negative genetic susceptibility with lung cancer. For age at first live birth the genetic predispositions for lung cancer are significant in overall lung cancer ($r_g = -0.45$, $p = 2.60 \times 10^{-14}$; $r_{g\ddagger} = 0.45$, $p_{\ddagger} = 4.98 \times 10^{-20}$), adenocarcinoma ($r_g = -0.29$, $p = 1.97 \times 10^{-8}$; $r_{g\ddagger} = -0.27$, $p_{\ddagger} = 1.78 \times 10^{-7}$), small cell carcinoma ($r_g = -0.53$, $p = 1.20 \times 10^{-13}$; $r_{g\ddagger} = -0.54$, $p_{\ddagger} = 1.78 \times 10^{-8}$), squamous cell carcinoma ($r_g = -0.53$, $p = 2.62 \times 10^{-14}$; $r_{g\ddagger} = -0.54$, $p_{\ddagger} = 1.68 \times 10^{-11}$), and ‘ever’ smokers ($r_g = -0.40$, $p = 2.16 \times 10^{-8}$; $r_{g\ddagger} = -0.43$, $p_{\ddagger} = 2.85 \times 10^{-11}$). Similarly, the age of last live birth demonstrated also demonstrated a significant decrease in lung cancer susceptibility overall, and in the small cell, squamous cell and ever smoking cohorts. The trait ‘age started oral contraceptive’ bore significant genetic predispositions with overall lung cancer ($r_g = -0.28$, $p = 1.30 \times 10^{-5}$; $r_{g\ddagger} = -0.27$, $p_{\ddagger} = 5.93 \times 10^{-5}$). These findings are further detailed in Fig. 5. A full correlation plot of all highly correlated traits is presented as Fig. 6, which includes all UKBB traits with significant genetic correlation with lung cancer after a Bonferroni correction for statistical significance. Figures 7 and 8 presents all nominally associated UKBB traits ($p < 0.05$) including their r_g and standard errors in cohort clustered forest plots.

Discussion

We sought to determine the shared genetic architecture between environmental and behavioral factors and lung cancer predisposition. LDSR has previously demonstrated efficacy and accuracy in determining the shared heritability and genetic correlation between phenotypes and disease states of interest^{42,43}. To date, the TRICL-OncoArray Lung consortium comprises the largest lung cancer GWAS conducted in European-ancestry populations³⁰. We leveraged these lung cancer GWAS meta-analysis data with GWAS summary statistics of traits from the UKBB to comprehensively assess shared genetic architectures between specific traits and lung cancer risk, observing numerous significant associations that were consistent across strata of lung cancer histology.

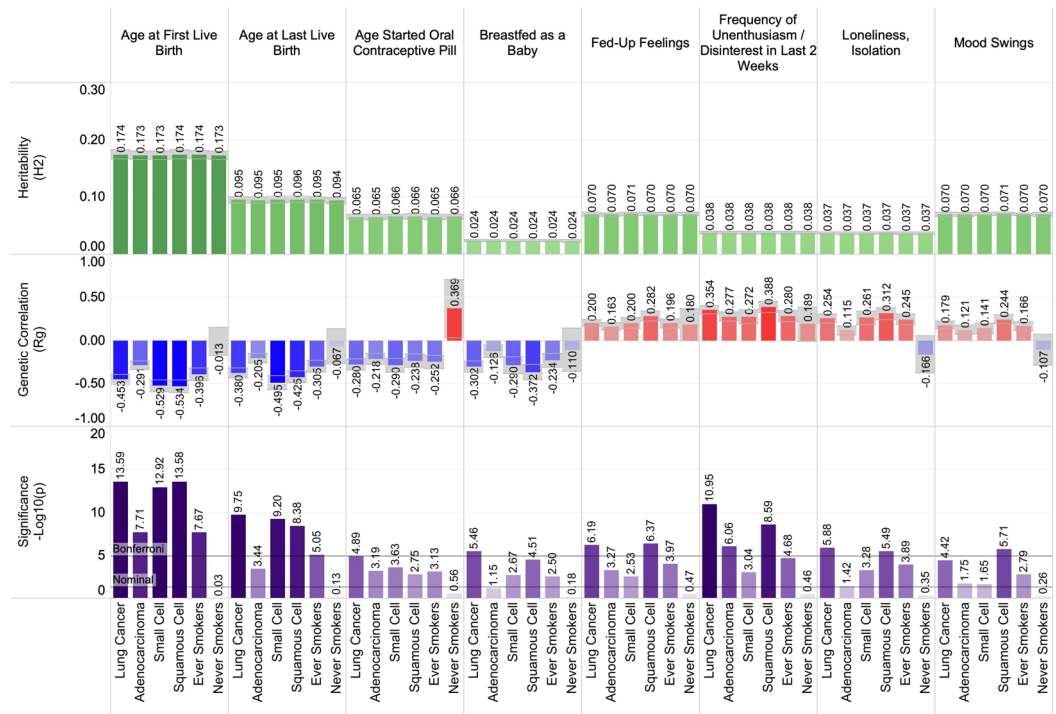


Figure 5. The shared heritability and genetic correlation between psychosocial and other specific traits with lung cancer.

We observed significant positive and negative (*i.e.*, protective) genetic correlations between lung cancer risk and individual behavioral characteristics and other environmental factors. We acknowledge that the strength of the LDSR method relies on the assumption that the genetic architectures between populations are similar. To ensure this, our analyses were conducted on European-ancestry populations in all studies, and SNPs included are those imputed using standard methods developed for application to the 1000 genomes project.

We provide further evidence that lung cancer is a heritable disease. Overall, our analysis estimated the heritability of lung cancer to be $8.3 \pm 1.3\%$, with comparable heritability in adenocarcinoma ($6.8 \pm 1.0\%$), higher heritability in small cell lung carcinoma ($10.5 \pm 1.9\%$), and lower heritability in squamous cell carcinoma of the lung ($5.2 \pm 1.1\%$). These findings are similar to previous reports²⁹. The heritability of lung cancer among never smokers was considerably lower than among smokers, which might indicate heterogeneity in etiology of lung cancer in never smokers obscures its heritable nature. It is noteworthy that we found no significant associations in LDSR analyses among the 'never' smoker's subgroup, but the observed genetic correlations in this cohort consistently mirroring the direction observed in 'ever' smokers and across histological subgroups. The never-smoker subgroup was a considerably smaller sample (2355 lung cancer cases, 7504 non-cancer controls) and had the lowest heritability of any of our lung cancer sub-strata, indicating that we may have been underpowered to detect cross-trait associations with this group.

The frequency and circumstance of alcohol consumption demonstrated a significant and mixed correlation with the genetic architecture of lung cancer. We found that "alcohol taken with meals" was negatively correlated with overall lung cancer. However, when analyzing this trend by type of alcohol consumed, higher average weekly beer and cider intake and higher weekly spirits intake were positively genetically correlated with lung cancer risk. In contrast, higher average weekly champagne, white wine, or red wine intake had a negative correlation. This effect has previously been observed through non-genetic epidemiological meta-analysis⁴⁴, and, notably, we observe concordant findings through LDSR. One possible explanation is that concurrent smoking consumption is more likely in those who drink beer or partake in spirits and less likely in wine drinkers, possibly due to socioeconomic differences⁴⁵. Evidence against this hypothesis includes that the genetic correlations with lung cancer and alcohol intake were consistent across histological subtypes and when contrasted against 'never' versus 'ever' smoking status, although non-significant in 'never' smokers.

Educational attainment traits demonstrated a consistent genetic correlation with lung cancer risk in LDSR analyses. Certifications of educational attainment were consistently negatively correlated with lung cancer susceptibility. The corollary is also true, with 'no educational qualifications' (*i.e.*, no college or university degree), no professional qualifications in nursing or teaching, no "A" levels, and no general certificate of secondary education, demonstrating a positive correlation with lung cancer risk. These findings retained significance across histological subtypes. Removal of smoking-related SNPs as a method to mitigate residual confounding effects did not change the identified correlations or significance of these findings. Complementing these findings, it was independently found that fluid intelligence score, which had a consistent $h^2 \sim 0.22 \pm 0.01$, demonstrated a consistent negative genetic correlation with lung cancer across histological subtypes and smoking statuses.

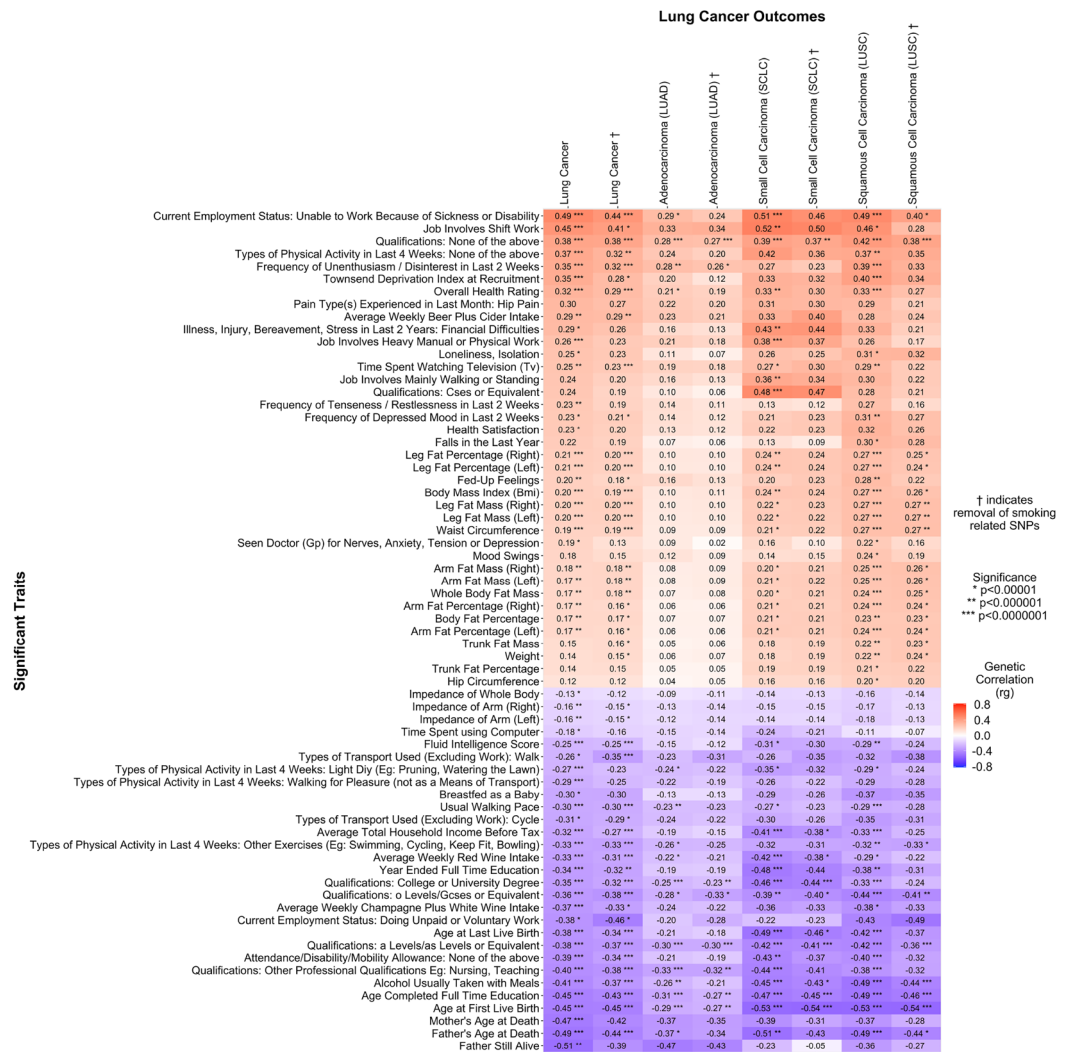


Figure 6. Genetic correlation plot of highly significant trait associations with lung cancer outcomes.

Summary statistics for several quantitative as well as binary fitness-related traits demonstrated consistent associations. However, consistency in statistical significance was not achieved among each of the three histological subgroups. Indicators of BMI demonstrated relatively consistent findings. We highlight BMI and body fat percentage. These traits demonstrate significant heritability ($h^2 \sim 0.22 \pm 0.01$) and have consistent positive genetic correlations with lung cancer. A general trend of negative genetic correlation between increased physical activity and lung cancer risk was observed; however, these findings had marginal estimated heritability at around ~ 0.03 . BMI's causal role in lung cancer oncogenesis was recently validated using Mendelian randomization⁶, however the strength of association measured in this prior study varied by lung cancer histology.

Several *specific* traits stood out from these analyses. A modest correlation was observed for depression and depression-related psychosocial traits, including 'frequency of fed-up feelings,' 'frequency of unenthusiasm/disinterest,' and 'loneliness, isolation,' and 'mood swings.' These captured symptoms are part of the diagnostic criteria for mental illnesses, and it is worth noting that the incidence of smoking behavior in populations who suffer from mental illness is higher than those without mental illnesses⁴⁶. Other specific standout traits genetically correlated with lung cancer risk included the participant-reported status of being breastfed as a baby. The heritability of this trait was 0.023 ± 0.002 , however, a consistent negative genetic correlation with lung cancer was observed. While interesting, these findings were only significant after correction for multiple comparisons testing for the overall and squamous cell lung cancer histological subgroups. The age at which a woman undergoes her 1st and last live birth and the age she started oral contraceptives were other specific traits that demonstrated a genetic correlation with lung cancer risk. These traits each revealed appreciable trait heritability and consistent, highly significant negative genetic correlations. It is well known that the ages of first live birth⁴⁷, last live birth⁴⁸, and initiation of oral contraceptive pills⁴⁹ are associated with androgen modulation and modified cancer risk. It is logical that these traits are annotating such a reality in lung cancer predisposition^{50,51}. We note that these results should be viewed as revealing only genetic associations, not for causal effect estimation.



Figure 7. Overlapping forest plot of nominally significant trait associations with lung cancer outcomes clustered by alcohol use, educational ascertainment and fitness metrics.

Our use of LDSR, with an intercept, allowed for acceptable mitigation of population stratification and cryptic relatedness confounders that could exist between the UKBB population and our TRICL-OncoArray lung cancer dataset. Further, we used individuals of European descent in these cohorts to mitigate this risk. Additional confounding, predominantly through smoking, have the potential to limit the strength of these analyses. To appreciate any hidden effects of smoking, we sub-stratified our analysis by those who had and had not smoked roughly 100 cigarettes in their lifetime. In addition to this 'never' versus 'ever' smoking comparison, we re-ran LDSR analyses after excluding genomic regions previously associated with smoking-related behaviors. Although GWAS meta-analyses of smoking behaviors have included upwards of 500,000 individuals, it is likely that additional genetic loci of small effect influence smoking behaviors and remain undetected by GWAS. Therefore, our analyses excluding known smoking-associated regions may not fully account for the contribution of smoking-associated genomic variation to our traits in our LDSR analyses. We present all our results, including these smoking sub-analyses, in the supplemental material.

Using cross trait LDSR, we have identified positively and negatively correlated traits with lung cancer. These findings indicate that shared genetic backgrounds exist between these traits, including alcohol use, educational attainment, fitness, and several other specific traits with lung cancer development. Our work should be viewed as a considerable step towards understanding the shared genetic architecture between these traits and lung cancer. A potential next step in future investigations is to perform causal analyses on strongly correlated traits we have described. Mendelian randomization studies may help determine causal versus mere association between these traits and the development of lung cancer. Ultimately identifying causal relationships may help to understand the shared genetic architecture of these traits with lung cancer, as well as to accurately create predictive risk models for lung cancer development. While causal modeling has an important role, it requires identifying and specifying sets of markers that can reliably represent intermediate traits. The LD Score regression approach evaluates the entire genome and so should be a more powerful filter for future causal modeling, once adequate genetic predictors for each of the traits that have been identified in our analysis are available.



Figure 8. Overlapping forest plot of nominally significant trait associations with lung cancer outcomes clustered by specific traits.

Received: 18 March 2021; Accepted: 6 August 2021
 Published online: 02 September 2021

References

- Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2019. *CA Cancer J. Clin.* **69**(1), 7–34. <https://doi.org/10.3322/caac.21551> (2019).
- Bosse, Y. & Amos, C. I. A decade of GWAS results in lung cancer. *Cancer Epidemiol. Biomarkers Prev.* **27**(4), 363–379. <https://doi.org/10.1158/1055-9965.EPI-16-0794> (2018).
- Bailey-Wilson, J. E. et al. A major lung cancer susceptibility locus maps to chromosome 6q23–25. *Am. J. Hum. Genet.* **75**(3), 460–474. <https://doi.org/10.1086/423857> (2004).
- Marant Micallef, C. et al. Occupational exposures and cancer: A review of agents and relative risk estimates. *Occup. Environ. Med.* **75**(8), 604–614. <https://doi.org/10.1136/oemed-2017-104858> (2018).
- Rosell, R. & Karachaliou, N. Large-scale screening for somatic mutations in lung cancer. *Lancet* **387**(10026), 1354–1356. [https://doi.org/10.1016/S0140-6736\(15\)01125-3](https://doi.org/10.1016/S0140-6736(15)01125-3) (2016).
- Zhou, W. et al. Causal relationships between body mass index, smoking and lung cancer: Univariable and multivariable Mendelian randomization. *Int. J. Cancer.* <https://doi.org/10.1002/ijc.33292> (2020).
- Pershagen, G. et al. Residential radon exposure and lung cancer in Sweden. *N. Engl. J. Med.* **330**(3), 159–164. <https://doi.org/10.1056/NEJM199401203300302> (1994).
- Hodge, A. M. et al. Dietary inflammatory index, Mediterranean diet score, and lung cancer: A prospective study. *Cancer Causes Control.* **27**(7), 907–917. <https://doi.org/10.1007/s10552-016-0770-1> (2016).
- Doll, R. Atmospheric pollution and lung cancer. *Environ. Health Perspect.* **22**, 23–31. <https://doi.org/10.1289/ehp.782223> (1978).
- Pershagen, G. Lung cancer mortality among men living near an arsenic-emitting smelter. *Am. J. Epidemiol.* **122**(4), 684–694. <https://doi.org/10.1093/oxfordjournals.aje.a114147> (1985).
- Lissowska, J. et al. Lung cancer and indoor pollution from heating and cooking with solid fuels: The IARC international multicentre case-control study in Eastern/Central Europe and the United Kingdom. *Am. J. Epidemiol.* **162**(4), 326–333. <https://doi.org/10.1093/aje/kwi204> (2005).

12. Dela Cruz, C. S., Tanoue, L. T. & Matthay, R. A. Lung cancer: Epidemiology, etiology, and prevention. *Clin. Chest Med.* **32**(4), 605–644. <https://doi.org/10.1016/j.ccm.2011.09.001> (2011).
13. Tokuhata, G. K. & Lilienfeld, A. M. Familial aggregation of lung cancer in humans. *J. Natl. Cancer Inst.* **30**(2), 289–312. <https://doi.org/10.1093/jnci/30.2.289> (1963).
14. Tokuhata, G. K. & Lilienfeld, A. M. Familial aggregation of lung cancer among hospital patients. *Public Health Rep.* **78**(4), 277–283. <https://doi.org/10.2307/4591778> (1963).
15. Sellers, T. A. *et al.* Evidence for mendelian inheritance in the pathogenesis of lung cancer. *J. Natl. Cancer Inst.* **82**(15), 1272–1279. <https://doi.org/10.1093/jnci/82.15.1272> (1990).
16. Sellers, T. A. *et al.* Segregation analysis of smoking-associated malignancies: Evidence for Mendelian inheritance. *Am. J. Med. Genet.* **52**(3), 308–314. <https://doi.org/10.1002/ajmg.1320520311> (1994).
17. Sellers, T. A. *et al.* Increased familial risk for non-lung cancer among relatives of lung cancer patients. *Am. J. Epidemiol.* **126**(2), 237–246. <https://doi.org/10.1093/aje/126.2.237> (1987).
18. Dragani, T. A., Manenti, G. & Pierotti, M. A. Polygenic inheritance of predisposition to lung cancer. *Ann. Ist. Super Sanita.* **32**(1), 145–150 (1996).
19. Yang, J. *et al.* Genomic inflation factors under polygenic inheritance. *Eur. J. Hum. Genet.* **19**(7), 807–812. <https://doi.org/10.1038/ejhg.2011.39> (2011).
20. Truong, T. *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.* **102**(13), 959–971. <https://doi.org/10.1093/jnci/djq178> (2010).
21. Hung, R. J. *et al.* A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. *Nature* **452**(7187), 633–637. <https://doi.org/10.1038/nature06885> (2008).
22. Thorgeirsson, T. E. *et al.* A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature* **452**(7187), 638–642. <https://doi.org/10.1038/nature06846> (2008).
23. Timofeeva, M. N. *et al.* Influence of common genetic variation on lung cancer risk: Meta-analysis of 14 900 cases and 29 485 controls. *Hum. Mol. Genet.* **21**(22), 4980–4995. <https://doi.org/10.1093/hmg/dds334> (2012).
24. Bulik-Sullivan, B. *et al.* LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**(3), 291–295. <https://doi.org/10.1038/ng.3211> (2015).
25. Maher, B. The case of the missing heritability. *Nature* **456**(7218), 18–22 (2008).
26. Tradigo G, Vacca R, Manini T, *et al.* A new approach to disentangle genetic and epigenetic components on disease comorbidities: Studying correlation between genotypic and phenotypic disease networks. In *Procedia Computer Science*, Vol. 110, 453–458 <https://doi.org/10.1016/j.procs.2017.06.119> (Elsevier B.V., 2017).
27. Rubio-Perez, C. *et al.* Genetic and functional characterization of disease associations explains comorbidity. *Sci. Rep.* **7**(1), 1–14. <https://doi.org/10.1038/s41598-017-04939-4> (2017).
28. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**(11), 1236–1241. <https://doi.org/10.1038/ng.3406> (2015).
29. McKay, J. D. *et al.* Large-scale association analysis identifies new lung cancer susceptibility loci and heterogeneity in genetic susceptibility across histological subtypes. *Nat. Genet.* **49**(7), 1126–1132. <https://doi.org/10.1038/ng.3892> (2017).
30. Amos, C. I. *et al.* The OncoArray Consortium: A network for understanding the genetic architecture of common cancers. *Cancer Epidemiol. Biomarkers Prev.* **26**(1), 126–135. <https://doi.org/10.1158/1055-9965.EPI-16-0106> (2017).
31. Sudlow, C. *et al.* UK Biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**(3), e1001779. <https://doi.org/10.1371/journal.pmed.1001779> (2015).
32. Altshuler, D. M. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**(7422), 56–65. <https://doi.org/10.1038/nature11632> (2012).
33. Byun, J. *et al.* Ancestry inference using principal component analysis and spatial analysis: A distance-based analysis to account for population substructure. *BMC Genomics* **18**(1), 1–12. <https://doi.org/10.1186/s12864-017-4166-8> (2017).
34. Zhang, L. R. *et al.* Cannabis smoking and lung cancer risk: Pooled analysis in the International Lung Cancer Consortium. *Int. J. Cancer.* **136**(4), 894–903. <https://doi.org/10.1002/ijc.29036> (2015).
35. Schuller, H. M. The neuro-psychological axis of smoking-associated cancer. *J. Immunol. Sci.* **3**(2), 1–5. <https://doi.org/10.29245/2578-3009/2019/2.1166> (2019).
36. Hecht SS. Tobacco and cancer: Approaches using carcinogen biomarkers and chemoprevention. In *Annals of the New York Academy of Sciences*, Vol. 833, 91–111. <https://doi.org/10.1111/j.1749-6632.1997.tb48596.x> (Blackwell Publishing Inc., 1997).
37. Leon, M. E. *et al.* European code against cancer, 4th edition: Tobacco and cancer. *Cancer Epidemiol.* **39**, S20–S33. <https://doi.org/10.1016/j.canep.2015.06.001> (2015).
38. Amos, C. I. *et al.* A susceptibility locus on chromosome 6q greatly increases lung cancer risk among light and never smokers. *Cancer Res.* **70**(6), 2359–2367. <https://doi.org/10.1158/0008-5472.CAN-09-3096> (2010).
39. Liu, M. *et al.* Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat. Genet.* **51**(2), 237–244. <https://doi.org/10.1038/s41588-018-0307-5> (2019).
40. Jiang, Y. *et al.* Proper conditional analysis in the presence of missing data: Application to large scale meta-analysis of tobacco use phenotypes. *PLoS Genet.* **14**(7), e1007452. <https://doi.org/10.1371/JOURNAL.PGEN.1007452> (2018).
41. Townsend, P. Deprivation. *Health Visit.* **45**(8), 223–224. <https://doi.org/10.1017/s0047279400020341> (1972).
42. Fuller, T. & Reus, V. Shared genetics of psychiatric disorders [version 1; peer review: 2 approved]. *F1000Research.* <https://doi.org/10.12688/f1000research.18130.1> (2019).
43. Hartz, S. M. *et al.* Genetic correlation between smoking behaviors and schizophrenia. *Schizophr. Res.* **194**, 86–90. <https://doi.org/10.1016/j.schres.2017.02.022> (2018).
44. Chao, C. Associations between beer, wine, and liquor consumption and lung cancer risk: A meta-analysis. *Cancer Epidemiol. Biomarkers Prev.* **16**(11), 2436–2447. <https://doi.org/10.1158/1055-9965.EPI-07-0386> (2007).
45. Friedman, G. D., Tekawa, I., Klatsky, A. L., Sidney, S. & Armstrong, M. A. Alcohol drinking and cigarette smoking: An exploration of the association in middle-aged men and women. *Drug Alcohol Depend.* **27**(3), 283–290. [https://doi.org/10.1016/0376-8716\(91\)90011-M](https://doi.org/10.1016/0376-8716(91)90011-M) (1991).
46. Lasser, K. *et al.* Smoking and mental illness: A population-based prevalence study. *J. Am. Med. Assoc.* **284**(20), 2606–2610. <https://doi.org/10.1001/jama.284.20.2606> (2000).
47. MacMahon, B. *et al.* Age at first birth and breast cancer risk. *Bull World Health Organ.* **43**(2), 209–221 (1970).
48. Setiawan, V. W. *et al.* Age at last birth in relation to risk of endometrial cancer: Pooled analysis in the epidemiology of endometrial cancer consortium. *Am. J. Epidemiol.* **176**(4), 269–278. <https://doi.org/10.1093/aje/kws129> (2012).
49. Marchbanks, P. A. *et al.* Oral contraceptives and the risk of breast cancer. *N. Engl. J. Med.* **346**(26), 2025–2032. <https://doi.org/10.1056/NEJMoa013202> (2002).
50. Siegfried, J. M., Hershberger, P. A. & Stabile, L. P. Estrogen receptor signaling in lung cancer. *Semin. Oncol.* **36**(6), 524–531. <https://doi.org/10.1053/j.seminoncol.2009.10.004> (2009).
51. Kawai, H. *et al.* Estrogen receptor α and β are prognostic factors in non-small cell lung cancer. *Clin. Cancer Res.* **11**(14), 5084–5089. <https://doi.org/10.1158/1078-0432.CCR-05-0200> (2005).

Acknowledgements

The authors would like to thank all members of the Integrative Analysis of Lung Cancer Etiology and Risk (INTEGRAL) Team of the International Lung Cancer Consortium (ILCCO) for providing summary results data for lung cancer. INTEGRAL-ILCCO acknowledges the following contributing investigators: Demetrius Albanes, Stephan Lam, Adonina Tardon, Chu Chen, Gary Goodman, Stig E. Bojesen, Maria Teresa Landi, Mattias Johansson, Angela Risch, H-Erich Wichmann, Heike Bickeboller, David C. Christiani, Gadi Rennert, Susanne Arnold, Paul Brennan, John K. Field, Sanjay Shete, Loic Le Marchand, Olle Melander, Hans Brunnström, Geoffrey Liu, Angeline Andrew, Lambertus A. Kiemeny, Hongbing Shen, Shan Zienolddiny, Kjell Grankvist, Mikael Johansson, Neil Caporaso, Penella Woll, Richard Houlston, Ying Wang, M. Dawn Teare, Yun-Chul Hong, Jian-Min Yuan, Philip Lazarus, Matthew B. Schabath, Melinda C. Aldrich. The authors would like to thank BRASS: Baylor Research Advocates for Student Scientists for their support (RWP).

Author contributions

Conception and design: J.B., Y.H., Q.T.O., J.E., K.M.W., M.L.B., J.D.M., and C.I.A.; Acquisition of data: The Integrative Analysis of Lung Cancer Etiology and Risk (INTEGRAL) Team of the International Lung Cancer Consortium (ILCCO); Analysis and interpretation of data: J.B., Y.H., Q.T.O., J.E., K.M.W., R.W.P., M.L.B., C.I.A.; R.W.P. and J.B. wrote the paper, and all authors provided feedback and contributed to the final format of this manuscript.

Funding

Cancer Prevention Research Interest of Texas (CPRIT) award: RR170048 (CIA, JB); National Institutes of Health (NIH) for INTEGRAL consortium: U19CA203654 (CIA, JB, YH, JE, RJH, JDM); Distinguished Scientist award from the Sontag Foundation (KMW); Research Training Grant from the Cancer Prevention and Research Institute of Texas: RP160097T (QTO); National Institutes of Health (NIH): R01CA139020 (MLB); NIH T32ES027801 (RWP).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-96685-x>.

Correspondence and requests for materials should be addressed to C.I.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021