Genome-wide gene-air pollution interaction analysis of lung function in 300,000 individuals

Carl A. Melbourne [a], A. Mesut Erzurumluoglu [b], Nick Shrine [a], Jing Chen [a], Martin D. Tobin [a,c], Anna L. Hansell [d,e,*], Louise V. Wain [a,c]

[a] Department of Health Sciences, University of Leicester, Leicester, UK
[b] MRC Epidemiology Unit, University of Cambridge, Cambridge, UK
[c] National Institute for Health Research, Leicester Respiratory Biomedical Research Centre, Glenfield Hospital, Leicester, UK
[d] Centre for Environmental Health and Sustainability, University of Leicester, Leicester, UK
[e] National Institute for Health Research Health Protection Research Unit in Environmental Exposures and Health at the University of Leicester, Leicester, UK

ARTICLE INFO

ABSTRACT

Background: Impaired lung function is predictive of mortality and is a key component of chronic obstructive pulmonary disease. Lung function has a strong genetic component but is also affected by environmental factors such as increased exposure to air pollution, but the effect of their interactions is not well understood.
Objectives: To identify interactions between genetic variants and air pollution measures which affect COPD risk and lung function. Additionally, to determine whether previously identified lung function genetic association signals showed evidence of interaction with air pollution, considering both individual effects and combined effects using a genetic risk score (GRS).
Methods: We conducted a genome-wide gene-air pollution interaction analysis of spirometry measures with three measures of air pollution at home address: particulate matter ($PM_{2.5}$ & PM_{10}) and nitrogen dioxide (NO_2), in approximately 300,000 unrelated European individuals from UK Biobank. We explored air pollution interactions with previously identified lung function signals and determined their combined interaction effect using a GRS.
Results: We identified seven new genome-wide interaction signals ($P < 5 \times 10^{-8}$), and a further ten suggestive interaction signals ($P < 5 \times 10^{-7}$). Additionally, we found statistical evidence of interaction for FEV_1/FVC between $PM_{2.5}$ and previously identified lung function signal, rs10841302, near *AEBP2*, suggesting increased susceptibility as copies of the G allele increased (but size of the impact was small - interaction beta: -0.363 percentage points, 95% CI: -0.523, -0.203 per 5 $\mu g/m^3$). There was no observed interaction between air pollutants and the weighted GRS.
Discussion: We carried out the largest genome-wide gene-air pollution interaction study of lung function and identified potential effects of clinically relevant size and significance. We observed up to 440 ml lower lung function for certain genotypes when exposed to mean levels of outdoor air pollution, which is approximately equivalent to nine years of average normal loss of lung function in adults.

1. Introduction

Impaired lung function is predictive of mortality and is a key component in the diagnosis of chronic obstructive pulmonary disease (COPD). Smoking is the biggest risk factor for COPD, which is thought to have caused as many as 2.9 million deaths worldwide in 2016 (GBD 2016 Causes of Death Collaborators, 2017) although other sources of indoor air pollution are also associated with COPD risk (Rabe & Watz, 2017; Agustí & Hogg, 2019). Furthermore, increased exposure to air pollution is associated with lower lung function (Doiron, et al., 2019).

Lung function and COPD risk is also influenced by genetic factors and we and others have discovered over 300 genetic association signals for COPD risk and/or lung function measures (Sakornsakolpat et al., 2019; Shrine et al., 2019). Combining these signals into a single genetic risk score, we have previously shown that individuals in the highest decile of genetic risk have an almost 5-fold increased risk of COPD compared to those in the lowest decile (Shrine et al., 2019). However, collectively, these variants only explain up to around 13% of the heritability of lung

function.

We hypothesised that there could be interactions between genetic variants and air pollution measures which affect COPD risk and lung function. Detection of such effects could enable identification of high-risk subgroups of the population and provide new biological insight into the mechanisms whereby air pollution affects respiratory health.

To test this hypothesis, we carried out the largest genome-wide gene-air pollution interaction study of lung function in $\sim$ 300,000 individuals from UK Biobank, using particulate matter (PM) and nitrogen dioxide (NO_2) concentrations as measures of air pollution exposure.

2. Methods

2.1. Study participants

We used spirometry, anthropometric, questionnaire and genetic data for individuals in UK Biobank, collected at baseline (upon recruitment) between 2006 and 2010. UK Biobank is a large-scale research database, containing both genetic and health information for a national cohort of over 500,000 individuals aged between 40 and 69 years.

2.2. Selection of individuals with lung function data

We selected unrelated European individuals from UK Biobank as previously described (Shrine et al., 2019). In summary, we selected individuals that had complete lung function data and passed our previously outlined quality control filters (N = 348,936) for forced expiratory volume in 1 second (FEV_1), forced vital capacity (FVC) and the ratio (FEV_1/FVC). From this we then selected a subsample of unrelated individuals (N = 303,320) of genetically determined European ancestry (KING kinship coefficient < 0.0884 corresponding to below 2nd degree kinship (Manichaikul et al., 2010)). All individuals had complete data for sex, age, height and ever smoking status (ever vs never).

2.3. Air pollution data

Air pollution concentrations at place of residence of UK Biobank participants at recruitment (at time of pulmonary function testing) were estimated using European Study of Cohorts and Air Pollution Effects (ESCAPE) land use regression models (Eeftens et al., 2012; Beelen et al., 2013). In these analyses, we explored associations with fine particles with average diameter < 2.5 μm ($PM_{2.5}$), particulate matter with average aerodynamic diameter < 10 μm (PM_{10}) and annual average concentrations of nitrogen dioxide (NO_2).

ESCAPE model predictions were compared to the UK's Automative Urban and Rural Network (AURN) data (Gulliver & Hoogh, 2015) to evaluate air pollution model estimates. NO_2 concentrations were predicted reasonably well throughout the country (R^2 = 0.67). PM_{10} concentrations were moderately well estimated for central and southern UK areas (R^2 = 0.53) but less so for nothern England or Scotland (R^2 < 0.5) as models were not robust > 400 km from Greater London. The PM analyses therefore did not include participants from northern England and Scotland [see (Doiron et al., 2017) for more details on air pollution concentration modelling].

2.4. Genome-wide interaction analysis

FEV_1, FVC and FEV_1/FVC were adjusted for sex, age, age^2, height and ever smoking. Residuals were then inverse normal transformed.

Individuals were genotyped using the Affymetrix Axiom UK BiLEVE and Affymetrix Axiom UK Biobank arrays (Bycroft et al., 2018) with imputation undertaken using the Haplotype Reference Consortium (HRC) (McCarthy et al., 2016) and combined UK10K + 1000 genomes (Huang et al., 2015) reference panels. Multiallelic variants were removed and variants imputed with low confidence were excluded (imputation quality r^2 < 0.5 for all SNPs and r^2 < 0.8 for rare SNPs with

minor allele frequency (MAF) < 1%). Variants with MAF < 0.5% were removed.

Each transformed lung function trait was used as the outcome in a multiple regression model which included the first 15 principal component terms for ancestry, genotyping array, SNP term (using an additive genetic model), air pollution variable and an interaction term for the interaction between SNP and air pollution:

$$Phenotype_i = \beta_0 + \beta_1 G_i + \beta_2 A_i + \beta_3 G_i A_i + PC_{1i} \cdots PC_{15i} + Array_i + \varepsilon_i$$

where G_i is the genotype for individual i, A_i is the air pollution value, $PC_{1i} \cdots PC_{15i}$ represent principal component values and $Array_i$ is the genotype array value (coded 0 and 1 for UK Biobank array and UK BiLEVE array respectively). The p-value returned for the β_3 estimate corresponds to the interaction effect between SNP and air pollution value ($G_i A_i$). Multiple regression was performed using PLINK2 (Chang et al., 2015).

Air pollution measures $PM_{2.5}$ and PM_{10} were transformed into standard z-scores due to observed collinearity issues as a result of strong correlation between the air pollution variable (A_i) and interaction ($G_i A_i$) in the regression model (observed due to small variances for air pollution measures $PM_{2.5}$ and PM_{10}, **Supplementary Fig. 1**). Air pollution measure NO_2 was analysed untransformed.

2.5. Signal selection and signal refinement

To define association signals and their sentinel variants, all variants were ranked by p-value and the SNP with the lowest p-value was selected as the first signal sentinel. All SNPs +/−1 megabase (Mb) either side of this first sentinel were then excluded and the process repeated for the next most significant SNP until all 2 Mb regions containing a sentinel SNP with $P < 5 \times 10^{-8}$ had been identified (genome-wide signals). The process was repeated to define a set of signals with sentinel SNPs at threshold of $P < 5 \times 10^{-7}$ (suggestive signals). Conditional analysis was used to identify additional independent genome-wide and suggestive signals by including the sentinel interaction term in the model, re-analysing all SNPs within each 2 Mb region and determining whether any SNPs remained below the pre-specified threshold. Region plots for each signal were created using LocusZoom (Pruim et al., 2010).

To aid the interpretation of interaction effects for genome-wide significant interaction signals, we presented the association between lung function trait and air pollution variable stratified by genotype group. To do this, dosages were converted to direct genotype calls by rounding to the nearest genotype group.

Using a Bayesian method (Wakefield, 2007) we refined each signal to a credible set of SNPs (the set of SNPs 95% likely to contain the causal SNP, under the assumption that the causal SNP was analysed).

2.6. Identification of putative causal genes

Credible set SNPs including the sentinel SNP were annotated using Annovar (Wang, et al., 2010) to identify coding variants with a putative functional effect (for example, missense). To identify whether any of the signals were independently associated with gene expression, we searched the GTEx (GTEx Consortium, 2013) and blood eQTLgen (Westra et al., 2013) eQTL catalogues. To identify a potential shared causal variant between the SNP-air pollution interaction signals and the eQTL gene expression signals, colocalisation was undertaken using COLOC (Giambartolomei et al., 2014) where full summary data was available in GTEx and eQTLgen databases (Võsa et al., 2021). An observed probability > 0.8 for a shared causal variant was used as the threshold to conclude colocalisation of SNP-air pollution and gene expression signals. We queried the sentinel SNPs in Open Target Genetics (Ghoussaini et al., 2021) for eQTL associations (which in addition to GTEx includes a further 14 consortia with eQTL expression association results) and to identify associations with protein expression (pQTL) and overlap with regions known to interact with gene promoters

(promotor capture HiC).

2.7. Association with other phenotypes

The SNP with the highest posterior probability for causality in each credible set was queried in PhenoScanner (Staley et al., 2016) and Open Targets Genetics (Ghoussaini et al., 2021) resources to identify shared associations with other phenotypes at a threshold of $P < 1 \times 10^{-3}$.

2.8. Tissue-specificity of interaction signals

To identify whether there was enrichment of SNP-air pollution interaction signals within regulatory regions of the genome (for example, DNase I Hypersensitive Sites (DHS)) in specific cell or tissue types we used GARFIELD (Iotchkova et al., 2019). The software determines whether signals are enriched for DHS across 55 tissues (with an adjusted significant enrichment threshold for 540 effective annotations of $P < 9.26 \times 10^{-5}$). We investigated the functional impact of SNPs (potential chromatin effects) which were highly probable to be the drivers of each signal (i.e. SNPs with posterior probability > 0.9 in credible sets) using DeepSEA (Zhou & Troyanskaya, 2015). To define a significant functional impact we used an E-value < 0.05 (the proportion of 1000 Genomes SNPs predicted to have a higher magnitude for chromatin effect compared to the chosen SNP being investigated) and an absolute probability difference > 0.1 between alternative and reference allele (the threshold defined for 'high confidence').

2.9. Sensitivity analyses

2.9.1. Effect of Socio-Economic status
Socio-economic status (SES) of an individual is a plausible moderator of lung function, with observed modification of air pollution effects (Doiron et al., 2019), however adjusting for SES in our analyses would have led to a reduction of approximately 13% in the discovery sample size due to missing data. We accounted for any effects of SES on genome-wide interaction signals in two ways. Firstly, we undertook a sensitivity analysis for the top signals adjusting for educational status and income status using a complete-case analysis (after inverse normalisation of lung function traits). Secondly, we present interaction effects for genome-wide signals across categorised groups for income and educational status to visualise any difference in effect (akin to a three-way interaction between SNP, air pollutant and education/income). Income status was categorised using the definition in UK Biobank of "less than £18,000", "£18,000 to £30,999", "£31,000 to £51,999", "£52,000 to 100,000" and "> 100,000". Educational status was dichotomised as "lower vocational qualification or less" vs "higher vocational qualification or more", grouping A-level (2), O-level (3), CSEs (4), and "None of the above" (-7) under "low education", and College/University (1), NVQ (5) and Other professional qualifications (6) under "high education". Individuals who selected "Do not know" (-1), "Prefer not to answer" (-3) or have missing data were excluded from subsequent analyses.

2.9.2. Exposure misspecification
Misspecification of a continuous exposure in statistical models, such as incorrectly modelling non-linear effects as linear, has been shown to inflate type I error rates when studying gene-environment interactions leading to identification of false-positives (Tchetgen Tchetgen & Kraft, 2011; Sun, et al., 2018). To determine whether this affected our conclusions (estimates and statistical significance), we re-calculated interaction effects for genome-wide gene-air pollution interaction signals using the same statistical model as before with inclusion of non-linear terms (quadratic and cubic to model the air pollution effect).

2.10. Previously reported lung function and COPD association signals

We performed a look-up in the genome-wide gene-air pollution interaction analyses (for all three air pollution measures and all three lung function measures), for the 304 signals previously reported for association with lung function and COPD (279 lung function signals from Shrine et al. 2019 (Shrine et al., 2019) and 25 signals from Sakornsakolpat et al. 2019 (Sakornsakolpat et al., 2019)). As these independent signals have *a priori* evidence for association with lung function or COPD, we applied a Bonferroni corrected threshold for 304 tests to define a significant air pollution interaction effect ($P < 1.6 \times 10^{-4}$). As before, to aid interpretation of the interaction effect for any statistically significant signal, we present the association between lung function trait and air pollution stratified by genotype group.

2.11. Weighted genetic risk score interaction analysis

We used a weighted genetic risk score (GRS) to explore whether the combined effect of previously reported lung function signals showed an interaction with air pollution measures (i.e. whether the phenotypic effects of the SNPs were modified by exposure to air pollution). Each individual's trait specific risk score was calculated using the effect sizes of the 279 SNPs reported in Shrine et al. 2019 (Shrine et al., 2019) on FEV_1, FVC and FEV_1/FVC (using the lung function reducing allele as the coded allele). Multiple regression was performed using the same model above, using the weighted GRS for each lung function trait in place of the genotype. As all three lung function traits are correlated, interaction terms (i.e. GRS × Air pollution measure) with $P < 0.05$ were defined as statistically significant.

2.12. Antioxidant genes and their interaction with air pollution

Genetic variation within antioxidant genes may contribute to susceptibility of adverse effects of air pollution on respiratory health (Fuertes et al., 2020). We have provided look-ups for the most commonly evaluated antioxidant genes (for which a SNP was reported) and for SNPs evaluated in previous antioxidant-gene-air pollution interaction studies, both of which are reviewed in Fuertes et al. (Fuertes et al., 2020). A Bonferroni adjusted threshold of $P < 3.85 \times 10^{-3}$ (for 13 variants) was used to determine statistical significance.

3. Results

The association between lung function and air pollutants PM_{10}, $PM_{2.5}$ and NO_2 in UK Biobank has previously been published in (Doiron et al., 2019), and we provide those associations in **supplementary Table 1**.

3.1. Genome-wide interaction analysis

Genome-wide interaction analysis was undertaken in 277,597 European individuals from UK Biobank for air pollution variables PM_{10}/$PM_{2.5}$, (**Supplementary Table 2**) and a total of 10,848,082 SNPs (**Supplementary Fig. 2**). For the NO_2 analysis, there were 299,015 European individuals and 10,846,777 SNPs. Manhattan plots are presented in Fig. 1 and QQ plots in **supplementary Fig. 3**.

We identified seven signals with an interaction effect reaching genome-wide statistical significance ($P < 5 \times 10^{-8}$) for at least one lung function trait and air pollution variable (Table 1, **Supplementary Table 3 and Supplementary Fig. 4**). Four signals were identified for an interaction with PM_{10}. There were two for FEV_1 (in 4q35.2 [near *LINC02374*] and in 19q12 [near *LOC100420587*]), one for FVC (in 1p36.33 [near *LINC01342*]) and one for FEV_1/FVC (in 6p25.1 [in *LY86-AS1*]). Two signals were identified for an interaction with $PM_{2.5}$; one for FEV_1 (in 7q31.33 [near *GRM8*]) and one for FVC (in 5q31.2 [in *KDM3B*].
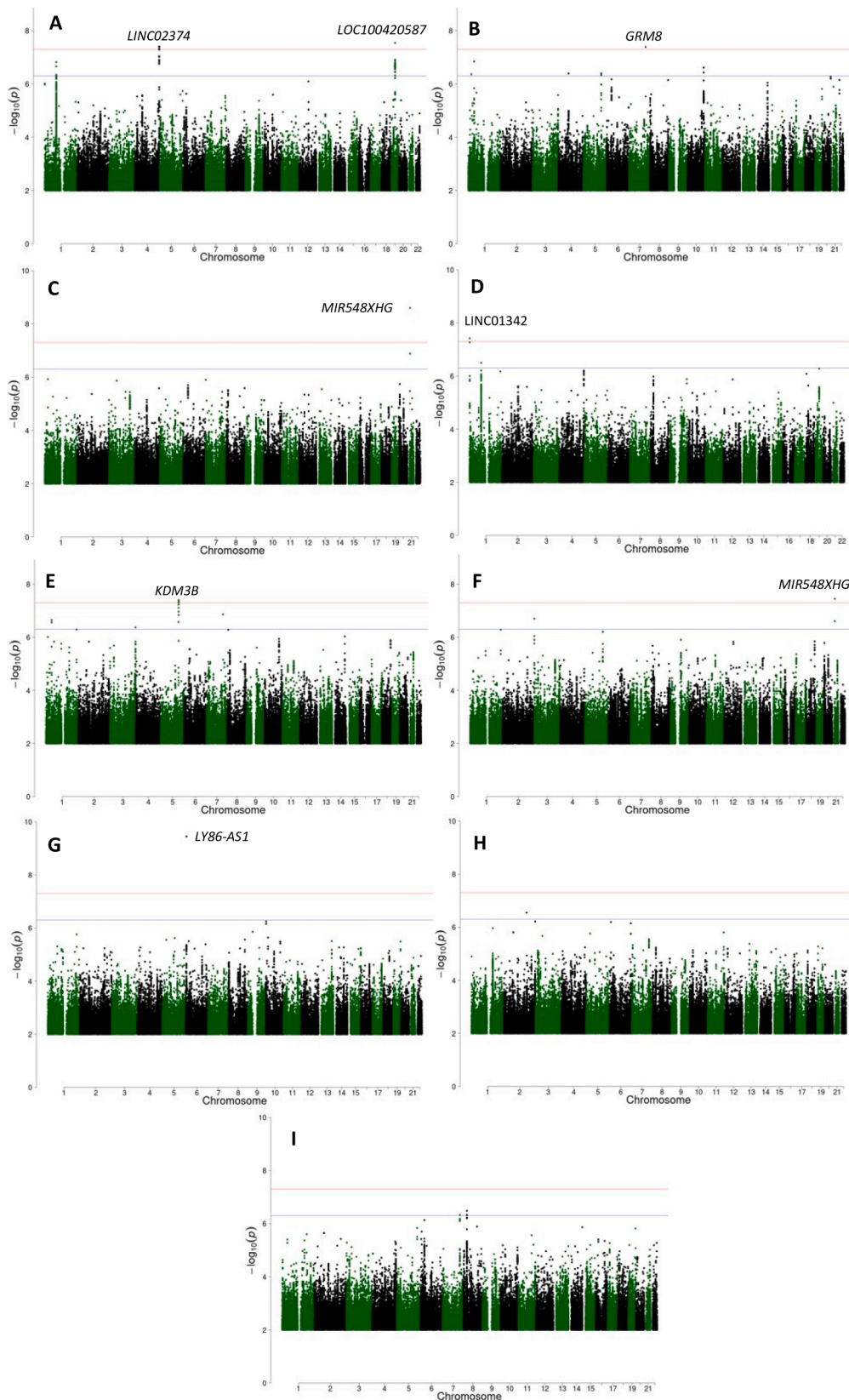
Fig. 1. Manhattan plots for the gene-air pollution interaction GWAS (A) FEV_1-PM_{10} (B) FEV_1-$PM_{2.5}$ (C) FEV_1-NO_2 (D) FVC-PM_{10} (E) FVC-$PM_{2.5}$ (F) FVC-NO_2 (G) FEV_1/FVC-PM_{10} (H) FEV_1/FVC-$PM_{2.5}$ (I) FEV_1/FVC-NO_2. The red line represents a p-value threshold of 5×10^{-8}. The blue line represents a p-value threshold of 5×10^{-7}. Each genome-wide signal is annotated by nearest gene. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

One signal was identified for air pollutant NO_2 for both lung function traits FEV_1 and FVC (in 21q21.1 [near *MIR548XHG*]). Of the seven identified SNPs, three were common (MAF > 5%) two were low frequency, (1% < MAF < 5%) and two were rare (MAF < 1%). Conditional analysis did not identify any additional signals in each region.

To aid with the interpretation of statistically significant interaction effects, we have presented the association between air pollution and lung function stratified by genotype group (number of copies of coded

Table 1

Seven identified genome-wide gene-air pollution interaction signals Note: CAF = Coded Allele Frequency, INFO = Imputation quality, AP = Air Pollutant, LF = Lung Function, BP = Base Position. A negative BETA (interaction effect) suggests a more deleterious effect on lung function per unit increase of air pollutant as the coded allele increases. A positive BETA (interaction effect) suggests a more protective effect. Interaction effect is per 10 μg/m³ increase in air pollutant NO_2 and per 5 μg/m³ increase for air pollution variables PM_{10} and $PM_{2.5}$ as the coded allele increases. Lung function effects are the product of the BETA value and the observed standard deviation of the lung function trait within the analysed sample.

LF trait	AP	SNP	CHR	BP	Coded allele	Non-Coded allele	CAF	INFO	BETA	SE	BETA effect in units of LF (ml for FEV_1, FVC or percentage points for FEV_1/FVC)	P	Locus
FVC	PM_{10}	rs74048016	1	1,068,280	C	G	0.98	0.97	−0.139	0.025	−133.1	3.83×10^{-8}	C1orf159 (dist = 16811), LINC01342 (dist = 4117)
FEV_1	PM_{10}	rs28666788	4	188,078,645	G	A	0.096	0.99	−0.065	0.012	−49.2	3.92×10^{-8}	FAT1 (dist = 433635), LINC02374 (dist = 44495)
FVC	$PM_{2.5}$	rs192415220	5	137,726,002	C	T	0.006	0.94	−0.479	0.087	−459.3	3.96×10^{-8}	KDM3B
FEV_1/FVC	PM_{10}	rs137914543	6	6,414,006	G	GTCTC	0.06	0.83	−0.101	0.016	−0.65	3.55×10^{-10}	LY86-AS1
FEV_1	$PM_{2.5}$	rs138235384	7	125,969,169	C	T	0.994	0.90	−0.466	0.085	−352.1	4.10×10^{-8}	LOC101928283 (dist = 949794), GRM8 (dist = 109483)
FEV_1	PM_{10}	rs762101031	19	29,112,275	CAAT	C	0.96	0.79	−0.120	0.022	−90.5	2.87×10^{-8}	LOC100420587
FEV_1	NO_2	rs2825255	21	20,362,376	T	C	0.83	1.00	−0.027	0.005	−20.6	2.53×10^{-9}	MIR548XHG (dist = 230246), LINC01683 (dist = 903217)
FVC									−0.025	0.005	−24.2	3.52×10^{-8}	

allele) for each of the seven genome-wide interaction signals (Fig. 2) and interaction plots of predicted lung function against air pollution for each genotype group (**Supplementary Fig. 5**). In some instances, statistically significant association between lung function and air pollutant is observed in all genotype groups. For others, the association only reaches statistical significance for certain genotype groups.

Signals were deemed suggestively statistically significant using the same signal selection procedure with a threshold of $P < 5 \times 10^{-7}$ (**Supplementary Table 4 and Supplementary Fig. 6**). Region plots after conditional analysis suggested only one signal per 2 Mb region. Ten suggestive signals were identified that were independent of the seven genome-wide significant signals, all were either intergenic or mapped to the intronic region of the mapped gene. Eight were represented by common SNPs, and two by low frequency SNPs.

3.2. Credible sets and causal genes

To identify the gene (or genes) via which the SNPs (for genome-wide and suggestively significant signals) might be exerting their effects on lung function, we used a Bayesian method to refine our signal to define the 95% credible set of causal SNPs (assuming the causal SNP was included in the analysis, **Supplementary Table 5**). We then investigated whether credible set and sentinel SNPs were associated with changes in gene expression in GTEx, Blood eQTL and Open Target Genetics databases (**Supplementary Table 6**). Genome-wide significant signal, rs74048016, whose C allele had a larger deleterious effect on lung function as the measurement of PM_{10} increased, was associated with decreased expression of *HES4* and increased expression of *C1orf159* and *RP11-465B22.3* in blood. However, there was no statistical support (using COLOC (Giambartolomei et al., 2014) and the eQTLgen database (Võsa et al., 2021) that the interaction signal and gene expression association signals originated from the same SNPs. Credible set SNPs for suggestive signals rs769937512, rs111552599, rs139556451, rs200460259 and rs10082259 were associated in various tissues for genes *AL445991.1, FRAS1, PNMA2/DPYSL, MUC4/MUC20* and *UROD* respectively (**Supplementary Table 6**). These signals did not colocalise, suggesting again that the observed gene expression and interaction signals were not driven by the same SNPs. There was no association with protein expression and no overlap with regions that had strong evidence for interaction with gene promoters.

3.3. Association with other phenotypes

The sentinel SNPs for the 17 genome-wide and suggestively significant signals were queried in PhenoScanner and Open Targets Genetics resources (**Supplementary Table 7**), to explore their association with related phenotypes e.g. asthma that might support a causal interpretation. Five signals were found to be associated with at least one trait at $P < 1 \times 10^{-3}$, three genome-wide signals (rs28666788, rs192415220 and rs138235384) and two suggestive signals (rs10082259/rs6661026 and rs769937512), but none of the associations had reached genome-wide significance ($P < 5 \times 10^{-8}$). For the genome-wide signals rs28666788, rs192415220 and rs138235384 the strongest associations (at $P < 5 \times 10^{-6}$) were with alcohol consumption, self-reported cervical polyps and sexual dysfunction respectively.

3.4. Tissue-specificity of interaction signals

When looking for evidence that the interaction signals were over-represented in tissue-specific functionally active regions of the genome (DNase I hypersensitive sites (DHS) indicative of open chromatin) using GARFIELD or responsible for chromatin effects using DeepSEA, only SNPs showing $SNP-NO_2$ interaction effects on lung function phenotype FVC were enriched in various tissues including fetal lung, using a threshold of $P < 5 \times 10^{-5}$ to select contributing SNPs
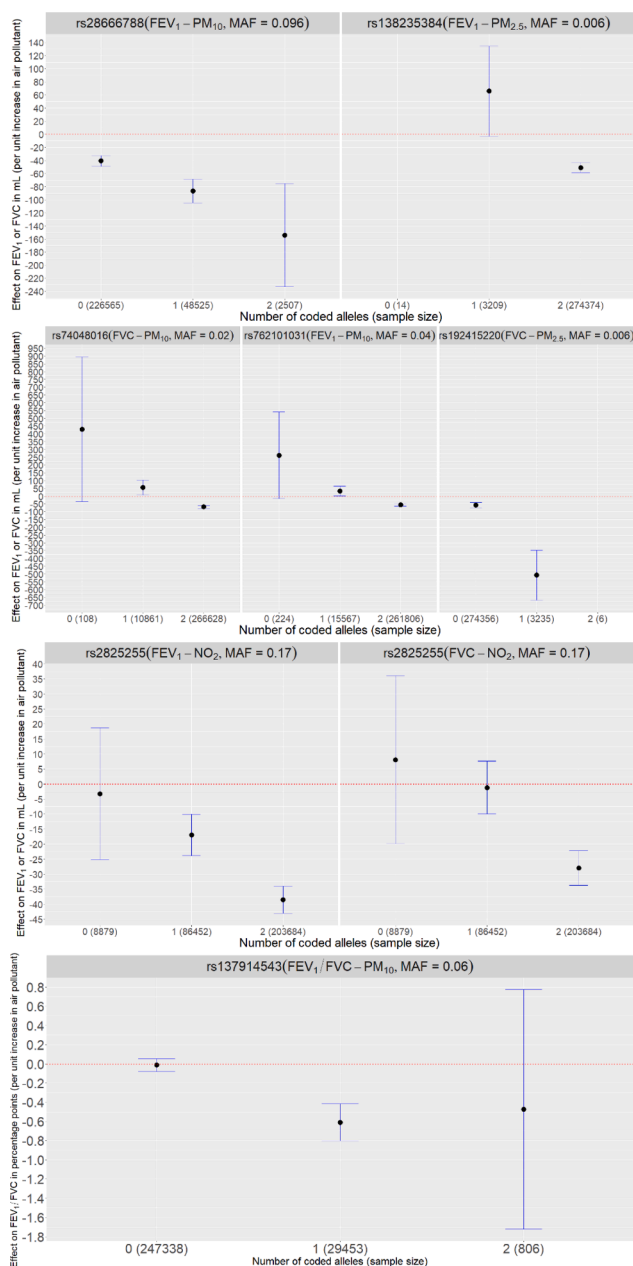
Fig. 2. Association between lung function trait and air pollutant (effect size and confidence intervals) for the seven genome-wide signals. Note: For SNPs rs138235384 and rs192415220 the effect size for 0 copies and 2 copies of the effect allele respectively are not presented due to the low minor allele frequency and small sample size. Effect sizes will not be exactly consistent with Table 1 due to rounding error when converting from dosage to direct genotypes. Units are per 5 μg/m^3 for PM$_{10}$ / PM$_{2.5}$ and per 10 μg/m^3 for NO$_2$.

(**Supplementary Fig. 7 and Supplementary Table 8**).

3.5. Sensitivity analyses

3.5.1. Effects of Socio-Economic status

When adjusting for socio-economic status variables educational status and income status, sample sizes were reduced to 259,130 and 240,202 for the NO$_2$ and PM$_{10}$/PM$_{2.5}$ analyses respectively. Effect sizes were largely consistent with the primary analysis with minimal reductions in effect size for rs74048016 and rs192415220 (**Supplementary Table 9 and Supplementary Fig. 8**), suggesting that the interactions identified were not due to confounding by SES factors.

Interaction effects were generally larger in magnitude (but not significantly due to overlapping confidence intervals) for those in the lower educational group (**Supplementary Fig. 9**). When stratifying by income group (**Supplementary Fig. 10**), overlapping confidence intervals again suggested no significant effect of income status on air pollution and lung function association across genotype groups. A slight inverse correlation between magnitude of interaction effect and income group was observed for rs2825255 for both lung function traits (higher income group, smaller interaction effect magnitude) with a positive correlation observed for rs762101031 (higher income group, larger interaction effect magnitude).

3.5.2. Exposure misspecification

There was very little effect on effect estimates and statistical significance for identified genome-wide interaction signals when modelling a non-linear effect of air pollution on lung function (**Supplementary Table 10**).

3.6. Lung function associated signals

To determine whether any signals previously shown to be associated with lung function produced an interaction effect with air pollution variables, we performed a look up of the 304 variants (279 lung function signals from Shrine et al. (Shrine et al., 2019) and 24 COPD signals from Sakornsakolpat et al. (Sakornsakolpat et al., 2019)) in our genome-wide analysis. Of the 304 signals, one signal, rs10841302, near *AEBP2*, for which the G allele is associated with lower values of FEV$_1$/FVC, met a Bonferroni threshold of P < 1.6 × 10^{-4} for an interaction with PM$_{2.5}$ for FEV$_1$/FVC (interaction β: -0.0569; 95% CI: − 0.0826, -0.0312; interaction P = 9.65x10^{-6}) (**Supplementary Table 11**), suggesting a larger deleterious effect of PM$_{2.5}$ on FEV$_1$/FVC as copies of the G allele increased (Fig. 3). This is equivalent to an FEV$_1$/FVC effect of −0.363 percentage points (CI: − 0.529, -0.200) per 5μ g/m^3 increase in PM$_{2.5}$. The interaction can also be interpreted by air pollution and lung function association stratified by genotype group. For genotype groups CC, CG and GG for SNP rs10841302, a 5μ g/m^3 increase in PM$_{2.5}$ resulted in a reduction of FEV$_1$/FVC by 0.16 (95% CI: 0.13–0.19; P = 1.21 × 10^{-22}), 0.17 (95% CI: 0.14–0.200; P = 5.52 × 10^{-39}) and 0.28 (95% CI: 0.24–0.32; P = 2.43 × 10^{-43}) standard deviations. This equates to direct FEV$_1$/FVC effects of 1.024 (95% CI: 0.83–1.22.), 1.08 (95% CI: 0.90–1.28.) and 1.79 (95% CI: 1.54–2.04) percentage points respectively per 5μ g/m^3 of PM$_{2.5}$.

We tested the interaction between a weighted GRS for lung function (based on the effect sizes of 279 lung function signals reported in Shrine et al. (Shrine et al., 2019) and each air pollution measure on FEV$_1$, FVC and FEV$_1$/FVC (**Supplementary Table 12**). None of the interaction effects were statistically significant (all P > 0.05).

3.7. Antioxidant genes and their interaction with air pollution

We performed a look up of the 13 variants corresponding to seven commonly evaluated antioxidant genes and/or those analysed in previous studies of antioxidant gene-air pollution interaction analyses, as reviewed by Fuertes et al. (Fuertes et al., 2020) (**Supplementary Table 13**). None of the SNPs reached the Bonferroni significant adjusted threshold used to determine statistical significance (P < 3.85 × 10^{-3}). One SNP, rs1001179 in *CAT* approached this threshold (P = 0.009) for an interaction with NO$_2$ for FEV$_1$/FVC.

4. Discussion

We carried out the largest genome-wide gene-air pollution interaction study of lung function and identified seven genome-wide statistically significant signals, as well as identifying a small interaction with air pollution for one previously identified lung function signal.
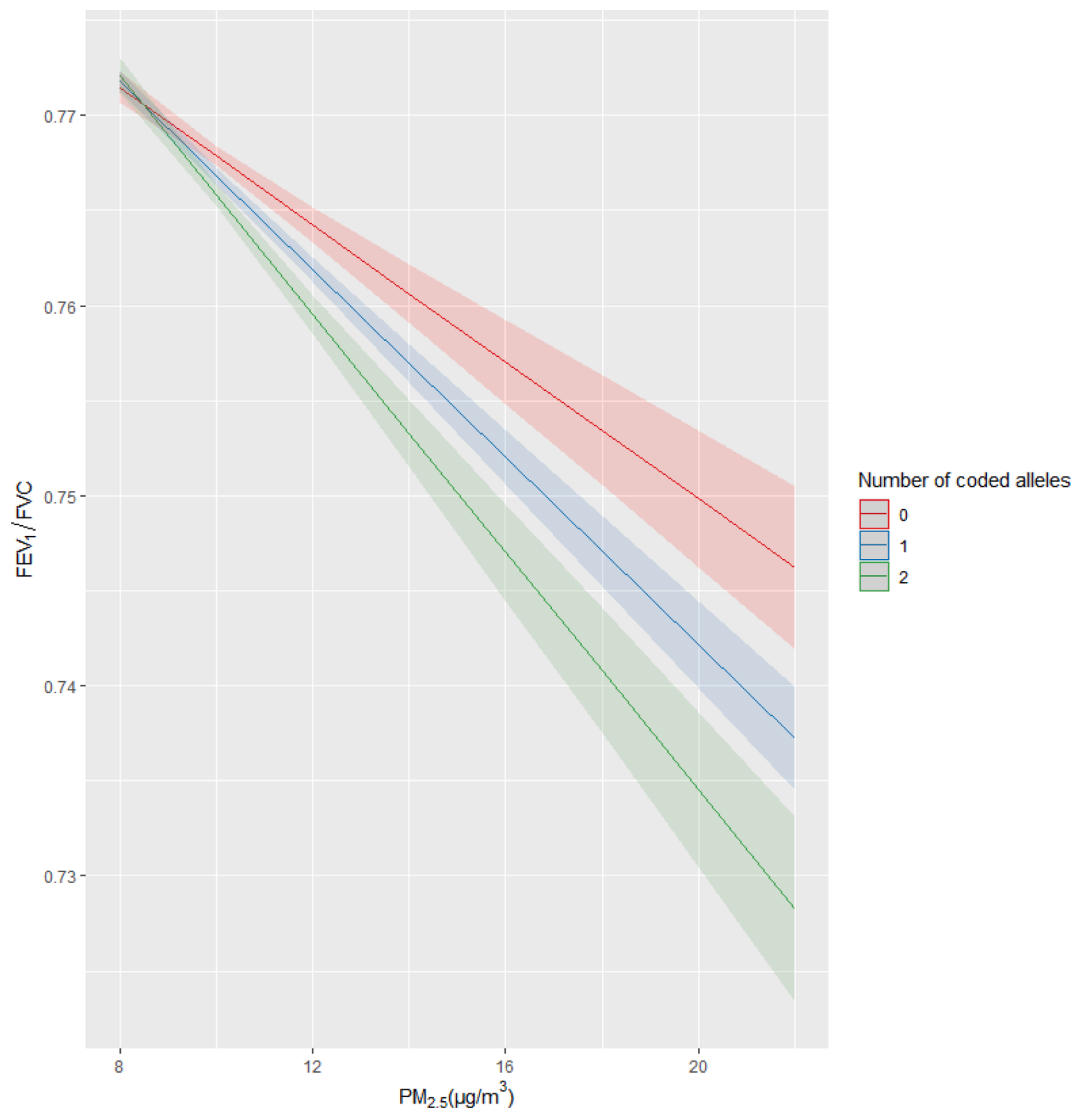
Fig. 3. Interaction plot of FEV_1/FVC predicted values against $PM_{2.5}$ values across genotype groups (with coded allele G) for the previously identified lung function signal rs10841302.

Independent replication is required to confirm these results. There were no interactions detected between air pollution and a weighted genetic risk score for lung function (using previously identified lung function signals), nor with seven commonly evaluated antioxidant genes. Further, we did not see convincing evidence of effect modification by social class.

For the signals identified, ascribing the biological mechanisms proves a challenge and further biological studies of gene function for those implicated are needed. For genome-wide SNP rs74048016, as the number of copies of the coded allele increases the effect of air pollutant PM_{10} on FVC becomes more negative, suggesting that those with two copies of the effect allele are at increased susceptibility of air pollution effects. The coded allele is associated with decreased expression of *HES4* and increased expression of *C1orf159* in blood in Open Targets Genetics. The signals for genome-wide association and gene expression signals did not colocalise (there was insufficient evidence of a shared causal variant between the two analyses) in this genomic region (using data from eQTLgen). Expression of *HES4 (*hes family bHLH transcription factor 4) has been implicated in poor outcomes for patients with Triple Negative Breast Cancer (TNBC) (Stoeck et al., 2014) and both *HES4* and *C1orf159* (chromosome 1 open reading frame 159*)* have been implicated via functional annotation (nearest gene) of other genome-wide significant

loci for several traits and diseases, including peak expiratory flow (PEF) (Ghoussaini et al., 2021; Neale Lab, 2021). There is also evidence of colocalisation between gene expression and genome-wide analyses for these genes in certain tissues for height phenotypes (standing and sitting) (Ghoussaini et al., 2021; Neale Lab, 2021).

We identified a further ten signals (independent of the primary genome-wide signals) at suggestive statistical significance, which would be important to take forward in future replication analyses. Genes implicated include *PNMA2*, *DPYSL2* and *BNIP3L*, all via functional annotation of other genome-wide significant loci for height, and additionally for educational attainment phenotypes (Kichaev et al., 2019; Lee et al., 2018). There was however no attenuation of suggestive signal rs139556451 (which implicated the aforementioned genes in our analysis) when re-analysing with adjustment for education and income status (in the subset for which this data was available). *BNIP3L* expression has also been linked with lung cancer (Sun, et al., 2004). Additionally, gene *FRAS1* identified by eQTL associations for SNPs in the rs111552599 suggestive signal credible set has been implicated by other genome-wide signals for lung function, specifically for trait FEV_1/FVC (Kichaev et al., 2019; Shrine et al., 2019) and mutations in *FRAS1* have been observed amongst individuals with Fraser syndrome, which can cause airway abnormalities (Pitera et al., 2008; van Haelst et al., 2007).

MUC4 (identified by credible set eQTL associations for rs200460259), which encodes airway mucins (Copin et al., 2000) is associated with severity of lung disease in cystic fibrosis (through functional annotation of another genome-wide signal) (Corvol et al., 2015) and risk of lung cancer (association with variants in the gene) (Zhang et al., 2013). We were however unable to determine whether the association signal for the genes described here were driven by the same causal variant as the interaction signal.

We identified an interaction effect between SNP rs10841302 (a previously identified lung function signal associated with FEV_1/FVC) and $PM_{2.5}$ for lung function trait FEV_1/FVC. Previous work has shown that the rs10841302 G-allele is associated with a deleterious effect on FEV_1/FVC. We found that this deleterious effect increased in magnitude as the exposure to $PM_{2.5}$ increased. A causative gene for the association between rs10841302 and lung function has not been determined. The SNP is near *AEBP2* (AE Binding Protein 2), a transcriptional repressor with a possible contribution to histone methylation and the G allele is associated with increased expression of both *RP11-405A12.2* (in pancreas and subcutaneous adipose tissues) and *RP11-664H17.1* (in pancreas and tibial nerve tissues) in GTEx (GTEx Consortium, 2013). There was no evidence of an interaction between air pollution measures and a combined effect from all previously identified lung function signals represented by a genetic risk score.

A particular strength of this study is the discovery sample size available for the interaction analysis, despite resulting in relatively few findings. This is likely indicative of the fact that several environmental and other exposures are at interplay across an individual's life course, and these are not addressed in our analysis. Moreover, interactions are challenging to identify due to the requirement of much larger sample sizes than GWAS efforts exploring the marginal effects of genetic variants (Thomas, 2010). This strength is however unfortunately a contributor to its biggest limitation, which is identifying suitable independent datasets of sufficient sample size with lung function data in European ancestry populations to replicate discovery interaction signals. We calculated that sample sizes to replicate three of our novel genome-wide interaction signals when considering the reported interaction effect, main genetic effect and air pollution variable effect (chosen from each MAF frequency group of common, low frequency and rare), signals rs28666788 (MAF = 10%), rs74048016 (MAF = 2%) and rs192415220 (MAF = 0.6%) would be ∼ 72 k, ∼ 71 k and ∼ 66 k respectively to detect the effect at 80% power. However, these sample sizes are indeed sensitive to any observed error in interaction effect estimates, such that when using lower and upper confidence interval effect estimates, sample sizes required could range from ∼ 35 k to ∼ 194 k.

The discovery of gene-air pollution interactions which affect lung function susceptibility is limited, likely due to the aforementioned difficulty in identifying suitable sample sizes to provide adequate power for replication studies, which is a limitation of our present analysis. Previous genome-wide interaction studies are either attributed to related phenotypes, such as asthma (Gref et al., 2017) or have focussed on candidate genes, such as those with a role in oxidative stress, where conclusions drawn are often inconsistent with respect to direction of effect or presence of interaction (Minelli et al., 2011; Romieu et al., 2010). Previous studies of interactions between genes and smoking behaviour, the largest risk factor for poor lung function and COPD, have also been largely unsuccessful in identifying interaction signals. This has been of interest as not all smokers develop restrictive lung problems. Candidate gene-smoking interactions have been identified, however utilising small sample sizes with absence of replication (Sadeghnejad et al., 2007; Hunninghake et al., 2009; He et al., 2004) and none of the previously identified lung function signals produced an interaction with smoking behaviour (Shrine et al., 2019). Genome-wide interaction analysis efforts have also been considered for lung function (Hancock et al., 2012) however with little success, and although a recent study of gene-smoking interaction effects for COPD found a genome-wide

significant interaction at *15q25.1* (Kim et al., 2020), this is likely driven by the strong association between this locus and smoking behaviour (Thorgeirsson et al., 2010; Liu et al., 2010; The Tobacco and Genetics Consortium, 2010). There has however been some evidence of interaction between smoking behaviour and genetic risk scores, when combining the effects of SNPs associated with lung function (Aschard et al., 2017; Shrine et al., 2019). To the best of our knowledge, no genome-wide significant smoking interaction signals for lung function have been identified, highlighting the impact of identifying novel genome-wide gene-air pollution interaction signals.

Should the interaction effects be replicated in future analyses, the magnitude of effects observed here suggest potential for clinically relevant impacts on those with certain genotypes. Results (Table 1, Fig. 2) are expressed per 5 μg/m³ for air pollutants PM_{10} and $PM_{2.5}$ and per 10 μg/m³ for NO_2. For context, average annual concentrations of PM_{10} in 2018 were 14.7 μg/m³ in 2018 at urban background air quality monitoring sites (likely to represent where most of the UK population live) (GOV.UK, 2021). Corresponding concentrations for $PM_{2.5}$ and NO_2 was 10.0 μg/m³ and 20.1 μg/m³ respectively. Taking genome-wide signal rs28666788 as an example, (with coded allele G frequency of 0.096), effects on FEV_1 per 5 μg/m³ increase in PM_{10} were statistically significant for all genotype groups. For those with zero, one and two copies of the effect allele, lung function effects of approximately -40 ml, -87.5 ml and -150 ml were observed per 5 μg/m³ PM_{10} respectively (Fig. 2). Therefore, when subjected to the average concentrations of 14.7 μg/m³ of PM_{10}, this equates to respective reductions of approximately 118 ml, 260 ml and 440 ml. Average declines in FEV_1 per year could be up to 46 ml for individuals aged 30 onwards (Quanjer et al., 2012), so these effects are approximately equivalent to nine years of normal loss of lung function for those with two copies of the coded allele (4 and 7 more than those with one and zero copies respectively). For other SNPs, such as rs2825255, with coded allele (T) frequency of 0.83, association between lung function and air pollutant is observed for certain genotype groups. Using the average NO_2 measure, those with one and two copies of the effect allele could be subject to reductions in FEV_1 of approximately 35 ml and 75 ml (approximately equivalent to 0.75 and 1.5 years of normal lung function decline respectively), as opposed to those with zero copies, where there was no observed statistically significant effect of air pollutant on FEV_1 (confidence interval overlaps 0).

There were approximately 40,000 individuals with clean lung function data with missing data for education and income status. We expect that those with higher SES and higher income are more likely to have complete data thus the data is not missing at random. We did not carry out imputation as it is difficult to know which might introduce more bias, imputation or exclusion and thus carried out a complete-case analysis. Further studies are required in this respect. Previous studies have reported modification of air pollution effects on lung function when considering SES (Doiron et al., 2017; Wheeler & Ben-Shlomo, 2005; Forastiere et al., 2007; Doiron et al., 2019) possibly due to differences in housing conditions, indoor air quality, nutrition and occupation (Forastiere et al., 2007). Adjusting for SES and presenting interaction effects across educational and income groups did not produce a notable modification of interaction effects in our analyses, suggesting that observed differences in the effect of air pollution across genotype groups are not mediated or confounded by socio-economic status.

There are other limitations with this study. We only had air pollution data at baseline with some limitations in the availability and did not have follow-up data. An analysis of a German cohort of 601 elderly women (mainly non-smokers) with three follow-ups from 1985 to 2013 suggested that changes in air pollution over time was associated with improvements in lung function, modified by genetic factors (Hüls et al., 2019). In addition, there are limitations with the ESCAPE models (Eeftens et al., 2012; Beelen et al., 2013). Exposure estimates are based on place of residence so will not capture variability in exposure related to

work and leisure activities outside the home, which may have led to exposure misclassification bias making it harder to detect effects. Furthermore, it must be noted that our analysis includes imputed genetic dosages alongside directly genotyped data and we only considered an additive genetic model for our analysis. Previous studies for certain antioxidant gene SNPs such as rs1695 in *GSTP1* have also considered the suitability of alternative genetic models (Wang, et al., 2019; Song et al., 2016).

In conclusion, we have identified genetic variants whose effect on lung function is dependent on air pollution exposure levels. This could help identify high-risk genetic subgroups whose lung function could be more susceptible to the effects of outdoor air pollution. While this is the largest study of this type to date, we highlight the need for replication in independent datasets with recorded lung function, for which availability is currently limited. We hope that future replication and further biological studies of gene function will help to establish the genes and biological pathways involved.

CRediT authorship contribution statement

Carl A. Melbourne: Conceptualization, Methodology, Software, Validation, Investigation, Formal analysis, Data curation, Writing – original draft, Writing – review & editing, Visualization. **A. Mesut Erzurumluoglu:** Conceptualization, Methodology, Software, Writing – review & editing. **Nick Shrine:** Software, Resources, Writing – review & editing. **Jing Chen:** Software, Resources, Writing – review & editing. **Martin D. Tobin:** Conceptualization, Resources, Supervision, Funding acquisition, Writing – review & editing, Project administration. **Anna L. Hansell:** Conceptualization, Investigation, Resources, Supervision, Project administration, Funding acquisition, Writing – review & editing. **Louise V. Wain:** Conceptualization, Project administration, Supervision, Funding acquisition, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envint.2021.107041.

References

Agustí, A., Hogg, J.C., 2019. Update on the Pathogenesis of Chronic Obstructive Pulmonary Disease. New Engl. J. Med. 381 (13), 1248–1256.

Aschard, H., Tobin, M.D., Hancock, D.B., Skurnik, D., Sood, A., James, A., Vernon Smith, A., Manichaikul, A.W., Campbell, A., Prins, B.P., Hayward, C., Loth, D.W., Porteous, D.J., Strachan, D.P., Zeggini, E., O'Connor, G.T., Brusselle, G.G., Boezen, H.M., Schulz, H., Deary, I.J., Hall, I.P., Rudan, I., Kaprio, J., Wilson, J.F., Wilk, J.B., Huffman, J.E., Hua Zhao, J., de Jong, K., Lyytikäinen, L., Wain, L.V., Jarvelin, M., Kähönen, M., Fornage, M., Polasek, O., Cassano, P.A., Barr, R.G., Rawal, R., Harris, S.E., Gharib, S.A., Enroth, S., Heckbert, S.R., Lehtimäki, T., Gyllensten, U., Jackson, V.E., Gudnason, V., Tang, W., Dupuis, J., Soler Artigas, M., Joshi, A.D., London, S.J., Kraft, P., 2017. Evidence for large-scale gene-by-smoking interaction effects on pulmonary function. Int. J. Epidemiol. 46, 894–904.

Beelen, R., Hoek, G., Vienneau, D., Eeftens, M., Dimakopoulou, K., Pedeli, X., Tsai, M.-Y., Künzli, N., Schikowski, T., Marcon, A., Eriksen, K.T., Raaschou-Nielsen, O., Stephanou, E., Patelarou, E., Lanki, T., Yli-Tuomi, T., Declercq, C., Falq, G., Stempfelet, M., Birk, M., Cyrys, J., von Klot, S., Nádor, G., Varró, M.J., Dèdelè, A., Grażulevičienė, R., Mölter, A., Lindley, S., Madsen, C., Cesaroni, G., Ranzi, A., Badaloni, C., Hoffmann, B., Nonnemacher, M., Krämer, U., Kuhlbusch, T., Cirach, M., de Nazelle, A., Nieuwenhuijsen, M., Bellander, T., Korek, M., Olsson, D., Strömgren, M., Dons, E., Jerrett, M., Fischer, P., Wang, M., Brunekreef, B., de Hoogh, K., 2013. Development of NO2 and NOx land use regression models for estimating air pollution exposure in 36 study areas in Europe – The ESCAPE project. Atmos. Environ. 72, 10–23.

Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., Cortes, A., Welsh, S., Young, A., Effingham, M., McVean, G., Leslie, S., Allen, N., Donnelly, P., Marchini, J., 2018. The UK Biobank resource with deep phenotyping and genomic data. Nature 562 (7726), 203–209.

Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., Lee, J.J., 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience 4, 7–8 eCollection 2015.

Copin, M.-C., Devisme, L., Buisine, M.-P., Marquette, C.H., Wurtz, A., Aubert, J.-P., Gosselin, B., Porchet, N., 2000. From normal respiratory mucosa to epidermoid carcinoma: expression of human mucin genes. Int. J. Cancer 86 (2), 162–168.

Corvol, H., Blackman, S.M., Boëlle, P., Gallins, P.J., Pace, R.G., Stonebraker, J.R., Accurso, F.J., Clement, A., Collaco, J.M., Dang, H., Dang, A.T., Franca, A., Gong, J., Guillot, L., Keenan, K., Li, W., Lin, F., Patrone, M.V., Raraigh, K.S., Sun, L., Zhou, Y., O'Neal, W.K., Sontag, M.K., Levy, H., Durie, P.R., Rommens, J.M., Drumm, M.L., Wright, F.A., Strug, L.J., Cutting, G.R., Knowles, M.R., 2015. Genome-wide association meta-analysis identifies five modifier loci of lung disease severity in cystic fibrosis. Nat. Commun. 6, 8382.

Doiron, D., de Hoogh, K., Probst-Hensch, N., Fortier, I., Cai, Y., De Matteis, S., Hansell, A. L., 2019. Air pollution, lung function and COPD: results from the population-based UK Biobank study. Eur. Respir J. 54 (1), 1802140. https://doi.org/10.1183/13993003.02140-2018. Print 2019 Jul.

Doiron, D., de Hoogh, K., Probst-Hensch, N., Mbatchou, S., Eeftens, M., Cai, Y., Schindler, C., Fortier, I., Hodgson, S., Gaye, A., Stolk, R., Hansell, A., 2017. Residential Air Pollution and Associations with Wheeze and Shortness of Breath in Adults: A Combined Analysis of Cross-Sectional Data from Two Large European Cohorts. Environ. Health Perspect. 125 (9), 097025. https://doi.org/10.1289/EHP1353.

Eeftens, M., Beelen, R., de Hoogh, K., Bellander, T., Cesaroni, G., Cirach, M., Declercq, C., Dèdelè, A., Dons, E., de Nazelle, A., Dimakopoulou, K., Eriksen, K., Falq, G., Fischer, P., Galassi, C., Grażulevičienė, R., Heinrich, J., Hoffmann, B., Jerrett, M., Keidel, D., Korek, M., Lanki, T., Lindley, S., Madsen, C., Mölter, A., Nádor, G., Nieuwenhuijsen, M., Nonnemacher, M., Pedeli, X., Raaschou-Nielsen, O., Patelarou, E., Quass, U., Ranzi, A., Schindler, C., Stempfelet, M., Stephanou, E., Sugiri, D., Tsai, M.-Y., Yli-Tuomi, T., Varró, M.J., Vienneau, D., Klot, S.V., Wolf, K., Brunekreef, B., Hoek, G., 2012. Development of Land Use Regression models for PM (2.5), PM(2.5) absorbance, PM(10) and PM(coarse) in 20 European study areas; results of the ESCAPE project. Environ. Sci. Technol. 46 (20), 11195–11205.

Forastiere, F., Stafoggia, M., Tasco, C., Picciotto, S., Agabiti, N., Cesaroni, G., Perucci, C. A., 2007. Socioeconomic status, particulate air pollution, and daily mortality: differential exposure or differential susceptibility. Am. J. Ind. Med. 50 (3), 208–216.

Fuertes, E., van der Plaat, D.A., Minelli, C., 2020. Antioxidant genes and susceptibility to air pollution for respiratory and cardiovascular health. Free Radical Biol. Med. 151, 88–98.

Gbd, 2016. Causes of Death Collaborators, 2017. Global, regional, and national age-sex specific mortality for 264 causes of death, 1980–2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet (London, England). 390, 1151–1210.

Ghoussaini, M., Mountjoy, E., Carmona, M., Peat, G., Schmidt, E.M., Hercules, A., Fumis, L., Miranda, A., Carvalho-Silva, D., Buniello, A., Burdett, T., Hayhurst, J., Baker, J., Ferrer, J., Gonzalez-Uriarte, A., Jupp, S., Karim, M.A., Koscielny, G., Machlitt-Northen, S., Malangone, C., Pendlington, Z.M., Roncaglia, P., Suveges, D., Wright, D., Vrousgou, O., Papa, E., Parkinson, H., MacArthur, J.A.L., Todd, J.A., Barrett, J.C., Schwartzentruber, J., Hulcoop, D.G., Ochoa, D., McDonagh, E.M., Dunham, I., 2021. Open Targets Genetics: systematic identification of trait-associated genes using large-scale genetics and functional genomics. Nucleic Acids Res. 49, D1311–D1320.

Giambartolomei, C., Vukcevic, D., Schadt, E.E., Franke, L., Hingorani, A.D., Wallace, C., Plagnol, V., 2014. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genetics 10.

GOV.UK, Air quality statistics. [online]. Available at: https://webarchive.nationalarchives.gov.uk/20200303040317/https://www.gov.uk/government/statistics/air-quality-statistics [accessed 22/02/21].

Gref, A., Merid, S.K., Gruzieva, O., Ballereau, S., Becker, A., Bellander, T., Bergström, A., Bossé, Y., Bottai, M., Chan-Yeung, M., Fuertes, E., Ierodiakonou, D., Jiang, R., Joly, S., Jones, M., Kobor, M.S., Korek, M., Kozyrskyj, A.L., Kumar, A., Lemonnier, N., MacIntyre, E., Ménard, C., Nickle, D., Obeidat, Ma'en, Pellet, J., Standl, M., Sääf, A., Söderhäll, C., Tiesler, C.M.T., van den Berge, M., Vonk, J.M., Vora, H., Xu, C.-J., Antó, J.M., Auffray, C., Brauer, M., Bousquet, J., Brunekreef, B., Gauderman, W.J., Heinrich, J., Kere, J., Koppelman, G.H., Postma, D., Carlsten, C., Pershagen, G., Melén, E., 2017. Genome-Wide Interaction Analysis of Air Pollution Exposure and Childhood Asthma with Functional Follow-up. Am. J. Respir. Crit. Care Med. 195 (10), 1373–1383.

GTEx Consortium, 2013. The Genotype-Tissue Expression (GTEx) project. Nat. Genet. 45, 580–585.

Gulliver, J., Hoogh, K.d., 2015. Environmental exposure assessment: modelling air pollution concentrations. [e-book]. Oxford University Press. Available from : https://oxfordmedicine.com/view/10.1093/med/9780199661756.001.0001/med-9780199661756-chapter-135 [cited Nov 23, 2021].

Hancock, D.B., Soler Artigas, M., Gharib, S.A., Henry, A., Manichaikul, A., Ramasamy, A., Loth, D.W., Imboden, M., Koch, B., McArdle, W.L., Smith, A.V., Smolonska, J., Sood, A., Tang, W., Wilk, J.B., Zhai, G., Zhao, J.H., Aschard, H., Burkart, K.M., Curjuric, I., Eijgelsheim, M., Elliott, P., Gu, X., Harris, T.B., Janson, C., Homuth, G., Hysi, P.G., Liu, J.Z., Loehr, L.R., Lohman, K., Loos, R.J.F., Manning, A.K., Marciante, K.D., Obeidat, M., Postma, D.S., Aldrich, M.C., Brusselle, G.G., Chen, T., Eiriksdottir, G., Franceschini, N., Heinrich, J., Rotter, J.I., Wijmenga, C., Williams, O.D., Bentley, A.R., Hofman, A., Laurie, C.C., Lumley, T., Morrison, A.C., Joubert, B.R., Rivadeneira, F., Couper, D.J., Kritchevsky, S.B., Liu, Y., Wjst, M., Wain, L.V., Vonk, J.M., Uitterlinden, A.G., Rochat, T., Rich, S.S., Psaty, B.M., O'Connor, G.T., North, K.E., Mirel, D.B., Meibohm, B., Launer, L.J., Khaw, K., Hartikainen, A., Hammond, C.J., Gläser, S., Marchini, J., Kraft, P., Wareham, N.J., Völzke, H., Stricker, B.H.C., Spector, T.D., Probst-Hensch, N.M., Jarvis, D., Jarvelin, M., Heckbert, S.R., Gudnason, V., Boezen, H.M., Barr, R.G., Cassano, P.A., Strachan, D.P., Fornage, M., Hall, I.P., Dupuis, J., Tobin, M.D., London, S.J., 2012. Genome-wide joint meta-analysis of SNP and SNP-by-smoking interaction identifies novel loci for pulmonary function. PLoS Genetics 8.

He, J.-Q., Connett, J.E., Anthonisen, N.R., Paré, P.D., Sandford, A.J., 2004. Glutathione S-transferase variants and their interaction with smoking on lung function. Am. J. Respir. Crit. Care Med. 170 (4), 388–394.

Huang, J., Howie, B., McCarthy, S., Memari, Y., Walter, K., Min, J.L., Danecek, P., Malerba, G., Trabetti, E., Zheng, H.-F., Gambaro, G., Richards, J.B., Durbin, R., Timpson, N.J., Marchini, J., Soranzo, N., 2015. Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel. Nat. Commun. 6 (1) https://doi.org/10.1038/ncomms9111.

Hüls, A., Sugiri, D., Abramson, M.J., Hoffmann, B., Schwender, H., Ickstadt, K., Krämer, U., Schikowski, T., 2019. Benefits of improved air quality on ageing lungs: impacts of genetics and obesity. Eur. Respir. J. 53.

Hunninghake, G.M., Cho, M.H., Tesfaigzi, Y., Soto-Quiros, M.E., Avila, L., Lasky-Su, J., Stidley, C., Melén, E., Söderhäll, C., Hallberg, J., Kull, I., Kere, J., Svartengren, M., Pershagen, G., Wickman, M., Lange, C., Demeo, D.L., Hersh, C.P., Klanderman, B.J., Raby, B.A., Sparrow, D., Shapiro, S.D., Silverman, E.K., Litonjua, A.A., Weiss, S.T., Celedón, J.C., 2009. MMP12, lung function, and COPD in high-risk populations. New Eng. J. Med. 361 (27), 2599–2608.

Iotchkova, V., Ritchie, G.R.S., Geihs, M., Morganella, S., Min, J.L., Walter, K., Timpson, N.J., Dunham, I., Birney, E., Soranzo, N., 2019. GARFIELD classifies disease-relevant genomic features through integration of functional annotations with association signals. Nat. Genet. 51 (2), 343–353.

Kichaev, G., Bhatia, G., Loh, P.-R., Gazal, S., Burch, K., Freund, M.K., Schoech, A., Pasaniuc, B., Price, A.L., 2019. Leveraging Polygenic Functional Enrichment to Improve GWAS Power. Am. J. Hum. Genet. 104 (1), 65–75.

Kim, W., Prokopenko, D., Sakornsakolpat, P., Hobbs, B.D., Lutz, S.M., Hokanson, J.E., Wain, L.V., Melbourne, C.A., Shrine, N., Tobin, M.D., Silverman, E.K., Cho, M.H., Beaty, T.H., 2020. Genome-wide Gene-by-smoking Interaction Study of Chronic Obstructive Pulmonary Disease. Am. J. Epidemiol.

Lee, J.J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., Nguyen-Viet, T.A., Bowers, P., Sidorenko, J., Karlsson Linnér, R., Fontana, M.A., Kundu, T., Lee, C., Li, H., Li, R., Royer, R., Timshel, P.N., Walters, R.K., Willoughby, E.A., Yengo, L., Alver, M., Bao, Y., Clark, D.W., Day, F.R., Furlotte, N.A., Joshi, P.K., Kemper, K.E., Kleinman, A., Langenberg, C., Mägi, R., Trampush, J.W., Verma, S.S., Wu, Y., Lam, M., Zhao, J.H., Zheng, Z., Boardman, J.D., Campbell, H., Freese, J., Harris, K.M., Hayward, C., Herd, P., Kumari, M., Lencz, T., Luan, J., Malhotra, A.K., Metspalu, A., Milani, L., Ong, K.K., Perry, J.R.B., Porteous, D.J., Ritchie, M.D., Smart, M.C., Smith, B.H., Tung, J.Y., Wareham, N.J., Wilson, J.F., Beauchamp, J.P., Conley, D.C., Esko, T., Lehrer, S.F., Magnusson, P.K.E., Oskarsson, S., Pers, T.H., Robinson, M.R., Thom, K., Watson, C., Chabris, C.F., Meyer, M.N., Laibson, D.I., Yang, J., Johannesson, M., Koellinger, P.D., Turley, P., Visscher, P.M., Benjamin, D. J., Cesarini, D., 2018. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nat. Genet. 50 (8), 1112–1121.

Liu, J.Z., Tozzi, F., Waterworth, D.M., Pillai, S.G., Muglia, P., Middleton, L., Berrettini, W., Knouff, C.W., Yuan, X., Waeber, G., Vollenweider, P., Preisig, M., Wareham, N.J., Zhao, J.H., Loos, R.J.F., Barroso, I., Khaw, K.-T., Grundy, S., Barter, P., Mahley, R., Kesaniemi, A., McPherson, R., Vincent, J.B., Strauss, J., Kennedy, J.L., Farmer, A., McGuffin, P., Day, R., Matthews, K., Bakke, P., Gulsvik, A., Lucae, S., Ising, M., Brueckl, T., Horstmann, S., Wichmann, H.-E., Rawal, R., Dahmen, N., Lamina, C., Polasek, O., Zgaga, L., Huffman, J., Campbell, S.,

Kooner, J., Chambers, J.C., Burnett, M.S., Devaney, J.M., Pichard, A.D., Kent, K.M., Satler, L., Lindsay, J.M., Waksman, R., Epstein, S., Wilson, J.F., Wild, S.H., Campbell, H., Vitart, V., Reilly, M.P., Li, M., Qu, L., Wilensky, R., Matthai, W., Hakonarson, H.H., Rader, D.J., Franke, A., Wittig, M., Schäfer, A., Uda, M., Terracciano, A., Xiao, X., Busonero, F., Scheet, P., Schlessinger, D., Clair, D.S., Rujescu, D., Abecasis, G.R., Grabe, H.J., Teumer, A., Völzke, H., Petersmann, A., John, U., Rudan, I., Hayward, C., Wright, A.F., Kolcic, I., Wright, B.J., Thompson, J. R., Balmforth, A.J., Hall, A.S., Samani, N.J., Anderson, C.A., Ahmad, T., Mathew, C. G., Parkes, M., Satsangi, J., Caulfield, M., Munroe, P.B., Farrall, M., Dominiczak, A., Worthington, J., Thomson, W., Eyre, S., Barton, A., Mooser, V., Francks, C., Marchini, J., 2010. Meta-analysis and imputation refines the association of 15q25 with smoking quantity. Nat. Genet. 42 (5), 436–440.

Manichaikul, A., Mychaleckyj, J.C., Rich, S.S., Daly, K., Sale, M., Chen, W.M., 2010. Robust relationship inference in genome-wide association studies. Bioinformatics (Oxford, England). 26, 2867–2873.

McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A.R., Teumer, A., Kang, H. M., Fuchsberger, C., Danecek, P., Sharp, K., Luo, Y., Sidore, C., Kwong, A., Timpson, N., Koskinen, S., Vrieze, S., Scott, L.J., Zhang, H., Mahajan, A., Veldink, J., Peters, U., Pato, C., van Duijn, C.M., Gillies, C.E., Gandin, I., Mezzavilla, M., Gilly, A., Cocca, M., Traglia, M., Angius, A., Barrett, J.C., Boomsma, D., Branham, K., Breen, G., Brummett, C.M., Busonero, F., Campbell, H., Chan, A., Chen, S., Chew, E., Collins, F.S., Corbin, L.J., Smith, G.D., Dedoussis, G., Dorr, M., Farmaki, A.E., Ferrucci, L., Forer, L., Fraser, R.M., Gabriel, S., Levy, S., Groop, L., Harrison, T., Hattersley, A., Holmen, O.L., Hveem, K., Kretzler, M., Lee, J.C., McGue, M., Meitinger, T., Melzer, D., Min, J.L., Mohlke, K.L., Vincent, J.B., Nauck, M., Nickerson, D., Palotie, A., Pato, M., Pirastu, N., McInnis, M., Richards, J.B., Sala, C., Salomaa, V., Schlessinger, D., Schoenherr, S., Slagboom, P.E., Small, K., Spector, T., Stambolian, D., Tuke, M., Tuomilehto, J., Van den Berg, L.H., Van Rheenen, W., Volker, U., Wijmenga, C., Toniolo, D., Zeggini, E., Gasparini, P., Sampson, M.G., Wilson, J.F., Frayling, T., de Bakker, P.I., Swertz, M.A., McCarroll, S., Kooperberg, C., Dekker, A., Altshuler, D., Willer, C., Iacono, W., Ripatti, S., Soranzo, N., Walter, K., Swaroop, A., Cucca, F., Anderson, C.A., Myers, R.M., Boehnke, M., McCarthy, M.I., Durbin, R., 2016. A reference panel of 64,976 haplotypes for genotype imputation. Nature Genet. 48, 1279–1283.

Minelli, C., Wei, I., Sagoo, G., Jarvis, D., Shaheen, S., Burney, P., 2011. Interactive effects of antioxidant genes and air pollution on respiratory function and airway disease: a HuGE review. Am. J. Epidemiol. 173, 603–620.

Neale Lab, UK Biobank. [online]. Available at: http://www.nealelab.is/uk-biobank [accessed Feb 17, 2021].

Pitera, J.E., Scambler, P.J., Woolf, A.S., 2008. Fras1, a basement membrane-associated protein mutated in Fraser syndrome, mediates both the initiation of the mammalian kidney and the integrity of renal glomeruli. Hum. Mol. Genet. 17, 3953–3964.

Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S., Gliedt, T.P., Boehnke, M., Abecasis, G.R., Willer, C.J., 2010. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics (Oxford, England). 26, 2336–2337.

Quanjer, P.H., Stanojevic, S., Cole, T.J., Baur, X., Hall, G.L., Culver, B.H., Enright, P.L., Hankinson, J.L., Ip, M.S.M., Zheng, J., Stocks, J., 2012. Multi-ethnic reference values for spirometry for the 3–95-yr age range: the global lung function 2012 equations. Eur. Respir. J. 40, 1324–1343.

Rabe, K.F., Watz, H., 2017. Chronic obstructive pulmonary disease. Lancet (London, England). 389, 1931–1940.

Romieu, I., Moreno-Macias, H., London, S.J., 2010. Gene by environment interaction and ambient air pollution. Proc. Am. Thoracic Soc. 7, 116–122.

Sadeghnejad, A., Meyers, D.A., Bottai, M., Sterling, D.A., Bleecker, E.R., Ohar, J.A., 2007. IL13 promoter polymorphism 1112C/T modulates the adverse effect of tobacco smoking on lung function. Am. J. Respir. Crit. Care Med. 176, 748–752.

Sakornsakolpat, P., Prokopenko, D., Lamontagne, M., Reeve, N.F., Guyatt, A.L., Jackson, V.E., Shrine, N., Qiao, D., Bartz, T.M., Kim, D.K., Lee, M.K., Latourelle, J.C., Li, X., Morrow, J.D., Obeidat, M., Wyss, A.B., Bakke, P., Barr, R.G., Beaty, T.H., Belinsky, S. A., Brusselle, G.G., Crapo, J.D., de Jong, K., DeMeo, D.L., Fingerlin, T.E., Gharib, S. A., Gulsvik, A., Hall, I.P., Hokanson, J.E., Kim, W.J., Lomas, D.A., London, S.J., Meyers, D.A., O'Connor, G.T., Rennard, S.I., Schwartz, D.A., Sliwinski, P., Sparrow, D., Strachan, D.P., Tal-Singer, R., Tesfaigzi, Y., Vestbo, J., Vonk, J.M., Yim, J.J., Zhou, X., Bossé, Y., Manichaikul, A., Lahousse, L., Silverman, E.K., Boezen, H.M., Wain, L.V., Tobin, M.D., Hobbs, B.D., Cho, M.H., SpiroMeta Consortium, International COPD Genetics Consortium, 2019. Genetic landscape of chronic obstructive pulmonary disease identifies heterogeneous cell-type and phenotype associations. Nat. Genet. 51, 494-505.

Shrine, N., Guyatt, A.L., Erzurumluoglu, A.M., Jackson, V.E., Hobbs, B.D., Melbourne, C. A., Batini, C., Fawcett, K.A., Song, K., Sakornsakolpat, P., Li, X., Boxall, R., Reeve, N. F., Obeidat, M., Zhao, J.H., Wielscher, M., Weiss, S., Kentistou, K.A., Cook, J.P., Sun, B.B., Zhou, J., Hui, J., Karrasch, S., Imboden, M., Harris, S.E., Marten, J., Enroth, S., Kerr, S.M., Surakka, I., Vitart, V., Lehtimäki, T., Allen, R.J., Bakke, P.S., Beaty, T.H., Bleecker, E.R., Bossé, Y., Brandsma, C.A., Chen, Z., Crapo, J.D., Danesh, J., DeMeo, D.L., Dudbridge, F., Ewert, R., Gieger, C., Gulsvik, A., Hansell, A.L., Hao, K., Hoffman, J.D., Hokanson, J.E., Homuth, G., Joshi, P.K., Joubert, P., Langenberg, C., Li, X., Li, L., Lin, K., Lind, L., Locantore, N., Luan, J., Mahajan, A., Maranville, J.C., Murray, A., Nickle, D.C., Packer, R., Parker, M.M., Paynton, M.L., Porteous, D.J., Prokopenko, D., Qiao, D., Rawal, R., Runz, H., Sayers, I., Sin, D.D., Smith, B.H., Soler Artigas, M., Sparrow, D., Tal-Singer, R., Timmers, P. R. H. J., Van den Berge, M., Whittaker, J.C., Woodruff, P.G., Yerges-Armstrong, L.M., Troyanskaya, O.G., Raitakari, O.T., Kähönen, M., Polašek, O., Gyllensten, U., Rudan, I., Deary, I.J., Probst-Hensch, N.M., Schulz, H., James, A.L., Wilson, J.F., Stubbe, B., Zeggini, E., Jarvelin, M.R., Wareham, N., Silverman, E.K., Hayward, C., Morris, A.P., Butterworth, A.S., Scott, R.A., Walters, R.G., Meyers, D.A., Cho, M.H., Strachan, D.P.,

Hall, I.P., Tobin, M.D., Wain, L.V., Understanding Society Scientific Group, 2019. New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries. Nat. Genet. 51, 481-493.

Song, Z., Shao, C., Feng, C., Lu, Y., Gao, Y., Dong, C., 2016. Association of glutathione S-transferase T1, M1, and P1 polymorphisms in the breast cancer risk: a meta-analysis. Ther. Clin. Risk Manag. 12, 763–769.

Staley, J.R., Blackshaw, J., Kamat, M.A., Ellis, S., Surendran, P., Sun, B.B., Paul, D.S., Freitag, D., Burgess, S., Danesh, J., Young, R., Butterworth, A.S., 2016. PhenoScanner: a database of human genotype-phenotype associations. Bioinformatics (Oxford, England). 32, 3207–3209.

Stoeck, A., Lejnine, S., Truong, A., Pan, L., Wang, H., Zang, C., Yuan, J., Ware, C., MacLean, J., Garrett-Engele, P.W., Kluk, M., Laskey, J., Haines, B.B., Moskaluk, C., Zawel, L., Fawell, S., Gilliland, G., Zhang, T., Kremer, B., Knoechel, B., Bernstein, B. E., Pear, W.S., Liu, X.S., Aster, J.C., Sathyanarayanan, S., 2014. Discovery of biomarkers predictive of GSI response in triple negative breast cancer and adenoid cystic carcinoma. Cancer Discovery 4, 1154–1167.

Sun, J., He, X., Yu, Y., Chen, Z., 2004. [Expression and structure of BNIP3L in lung cancer]. Ai Zheng = Aizheng = Chinese Journal of. Cancer 23, 8–14.

Sun, R., Carroll, R.J., Christiani, D.C., Lin, X., 2018. Testing for gene-environment interaction under exposure misspecification. Biometrics 74, 653–662.

Tchetgen Tchetgen, E.J., Kraft, P., 2011. On the robustness of tests of genetic associations incorporating gene-environment interaction when the environmental exposure is misspecified. Epidemiology (Cambridge, Mass.) 22, 257–261.

The Tobacco and Genetics Consortium, 2010. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. Nat. Genet. 42, 441–447.

Thomas, D., 2010. Gene–environment-wide association studies: emerging approaches. Nat. Rev. Genet. 11, 259–272.

Thorgeirsson, T.E., Gudbjartsson, D.F., Surakka, I., Vink, J.M., Amin, N., Geller, F., Sulem, P., Rafnar, T., Esko, T., Walter, S., Gieger, C., Rawal, R., Mangino, M., Prokopenko, I., Mägi, R., Keskitalo, K., Gudjonsdottir, I.H., Gretarsdottir, S., Stefansson, H., Thompson, J.R., Aulchenko, Y.S., Nelis, M., Aben, K.K., den Heijer, M., Dirksen, A., Ashraf, H., Soranzo, N., Valdes, A.M., Steves, C., Uitterlinden, A.G., Hofman, A., Tönjes, A., Kovacs, P., Hottenga, J.J., Willemsen, G., Vogelzangs, N., Döring, A., Dahmen, N., Nitz, B., Pergadia, M.L., Saez, B., De Diego, V., Lezcano, V., Garcia-Prats, M.D., Ripatti, S., Perola, M., Kettunen, J., Hartikainen, A., Pouta, A., Laitinen, J., Isohanni, M., Huei-Yi, I., Allen, M., Krestyaninova, M., Hall, A.S., Jones, G.T., van Rij, A.M., Mueller, T., Dieplinger, B., Haltmayer, M., Jonsson, S., Matthiasson, S.E., Oskarsson, H., Tyrfingsson, T., Kiemeney, L.A., Mayordomo, J.I., Lindholt, J.S., Pedersen, J.H., Franklin, W.A., Wolf, H., Montgomery, G.W., Heath, A.C., Martin, N.G., Madden, P.A.F., Giegling, I., Rujescu, D., Järvelin, M., Salomaa, V., Stumvoll, M., Spector, T.D., Wichmann, H., Metspalu, A., Samani, N.J., Penninx, B.W., Oostra, B.A., Boomsma, D.I., Tiemeier, H., van Duijn, C.M., Kaprio, J., Gulcher, J.R., McCarthy, M.I., Peltonen, L., Thorsteinsdottir, U., Stefansson, K., 2010. Sequence variants at CHRNB3-CHRNA6 and CYP2A6 affect smoking behavior. Nature Genet. 42, 448–453.

van Haelst, M.M., Scambler, P.J., Hennekam, R.C.M., 2007. Fraser syndrome: a clinical study of 59 cases and evaluation of diagnostic criteria. Am. J. Med. Genet.. Part A 143A, 3194–3203.

Võsa, U., Claringbould, A., Westra, H., Bonder, M.J., Deelen, P., Zeng, B., Kirsten, H., Saha, A., Kreuzhuber, R., Yazar, S., Brugge, H., Oelen, R., de Vries, D.H., van der Wijst, Monique G. P., Kasela, S., Pervjakova, N., Alves, I., Favé, M., Agbessi, M., Christiansen, M.W., Jansen, R., Seppälä, I., Tong, L., Teumer, A., Schramm, K., Hemani, G., Verlouw, J., Yaghootkar, H., Sönmez Flitman, R., Brown, A., Kukushkina, V., Kalnapenkis, A., Rüeger, S., Porcu, E., Kronberg, J., Kettunen, J., Lee, B., Zhang, F., Qi, T., Hernandez, J.A., Arindrarto, W., Beutner, F., Dmitrieva, J., Elansary, M., Fairfax, B.P., Georges, M., Heijmans, B.T., Hewitt, A.W., Kähönen, M., Kim, Y., Knight, J.C., Kovacs, P., Krohn, K., Li, S., Loeffler, M., Marigorta, U.M., Mei, H., Momozawa, Y., Müller-Nurasyid, M., Nauck, M., Nivard, M.G., Penninx, Brenda W. J. H., Pritchard, J.K., Raitakari, O.T., Rotzschke, O., Slagboom, E.P., Stehouwer, C.D.A., Stumvoll, M., Sullivan, P., 't Hoen, Peter A. C., Thiery, J., Tönjes, A., van Dongen, J., van Iterson, M., Veldink, J.H., Völker, U., Warmerdam, R., Wijmenga, C., Swertz, M., Andiappan, A., Montgomery, G.W., Ripatti, S., Perola, M., Kutalik, Z., Dermitzakis, E., Bergmann, S., Frayling, T., van Meurs, J., Prokisch, H., Ahsan, H., Pierce, B.L., Lehtimäki, T., Boomsma, D.I., Psaty, B.M., Gharib, S.A., Awadalla, P., Milani, L., Ouwehand, W.H., Downes, K., Stegle, O., Battle, A., Visscher, P.M., Yang, J., Scholz, M., Powell, J., Gibson, G., Esko, T., Franke, L., 2021. Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. Nat. Gen. 53, 1300-1310.

Wakefield, J., 2007. A Bayesian measure of the probability of false discovery in genetic epidemiology studies. Am. J. Hum. Genet. 81, 208–227.

Wang, K., Li, M., Hakonarson, H., 2010. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 38, e164.

Wang, S., Zhang, J., Jun, F., Bai, Z., 2019. Glutathione S-transferase pi 1 variant and squamous cell carcinoma susceptibility: a meta-analysis of 52 case-control studies. BMC Med. Genet. 20, 22.

Westra, H., Peters, M.J., Esko, T., Yaghootkar, H., Schurmann, C., Kettunen, J., Christiansen, M.W., Fairfax, B.P., Schramm, K., Powell, J.E., Zhernakova, A., Zhernakova, D.V., Veldink, J.H., Van den Berg, Leonard H., Karjalainen, J., Withoff, S., Uitterlinden, A.G., Hofman, A., Rivadeneira, F., Hoen, Peter A. C. 't, Reinmaa, E., Fischer, K., Nelis, M., Milani, L., Melzer, D., Ferrucci, L., Singleton, A.B., Hernandez, D.G., Nalls, M.A., Homuth, G., Nauck, M., Radke, D., Völker, U., Perola, M., Salomaa, V., Brody, J., Suchy-Dicey, A., Gharib, S.A., Enquobahrie, D.A., Lumley, T., Montgomery, G.W., Makino, S., Prokisch, H., Herder, C., Roden, M., Grallert, H., Meitinger, T., Strauch, K., Li, Y., Jansen, R.C., Visscher, P.M., Knight, J.C., Psaty, B. M., Ripatti, S., Teumer, A., Frayling, T.M., Metspalu, A., van Meurs, Joyce B. J., Franke, L., 2013. Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat. Genet. 45, 1238-1243.

Wheeler, B.W., Ben-Shlomo, Y., 2005. Environmental equity, air quality, socioeconomic status, and respiratory health: a linkage analysis of routine data from the Health Survey for England. J. Epidemiol. Community Health 59, 948–954.

Zhang, Z., Wang, J., He, J., Zheng, Z., Zeng, X., Zhang, C., Ye, J., Zhang, Y., Zhong, N., Lu, W., 2013. Genetic variants in MUC4 gene are associated with lung cancer risk in a Chinese population. PLoS ONE 8, e77723.

Zhou, J., Troyanskaya, O.G., 2015. Predicting effects of noncoding variants with deep learning-based sequence model. Nat. Methods 12, 931–934.