



Pathway analysis of a genome-wide gene by air pollution interaction study in asthmatic children

Despo Ierodiakonou^{1,2} · Brent A. Coull³ · Antonella Zanobetti⁴ · Dirkje S. Postma^{2,5} · H. Marike Boezen^{1,2} · Judith M. Vonk^{1,2} · Edward F. McKone⁶ · Jonathan S. Schildcrout⁷ · Gerard H. Koppelman^{2,8} · Damien C. Croteau-Chonka⁹ · Thomas Lumley¹⁰ · Petros Koutrakis⁴ · Joel Schwartz⁴ · Diane R. Gold^{4,9} · Scott T. Weiss⁹

Received: 9 June 2018 / Revised: 23 November 2018 / Accepted: 8 March 2019
© Springer Nature America, Inc. 2019

Abstract

Objectives We aimed to investigate the role of genetics in the respiratory response of asthmatic children to air pollution, with a genome-wide level analysis of gene by nitrogen dioxide (NO₂) and carbon monoxide (CO) interaction on lung function and to identify biological pathways involved.

Methods We used a two-step method for fast linear mixed model computations for genome-wide association studies, exploring whether variants modify the longitudinal relationship between 4-month average pollution and post-bronchodilator FEV₁ in 522 Caucasian and 88 African-American asthmatic children. Top hits were confirmed with classic linear mixed-effect models. We used the improved gene set enrichment analysis for GWAS (*i-GSEA4GWAS*) to identify plausible pathways.

Results Two SNPs near the *EPHA3* (rs13090972 and rs958144) and one in *TXNDC8* (rs7041938) showed significant interactions with NO₂ in Caucasians but we did not replicate this locus in African-Americans. SNP–CO interactions did not reach genome-wide significance. The *i-GSEA4GWAS* showed a pathway linked to the HO-1/CO system to be associated with CO-related FEV₁ changes. For NO₂-related FEV₁ responses, we identified pathways involved in cellular adhesion, oxidative stress, inflammation, and metabolic responses.

Conclusion The host lung function response to long-term exposure to pollution is linked to genes involved in cellular adhesion, oxidative stress, inflammatory, and metabolic pathways.

Keywords Air pollution · Asthma · Genome-wide · Gene–environment interaction · Lung function · Pathways.

Introduction

Epidemiological studies have demonstrated a strong association between exposure to ambient air pollution and adverse effects on childhood respiratory health [1–3], with asthmatic children being more susceptible to the negative effects of air pollution [4–6]. Lower lung function levels in asthmatic and non-asthmatic children have been associated with short-term exposure to air pollution [3, 7, 8], but the

long-term effects of pollution on lung function are less well studied in asthmatic children [9–12].

Known biological mechanisms by which air pollution can impair health include autonomic dysfunction, oxidative stress, and systemic inflammatory responses [13–16]. Respiratory response to air pollution varies between individuals suggesting that genetic susceptibility likely plays a role [17]. Recent genome-wide interaction analyses of chronic air pollution exposure indicated that gene–environment interactions are important for asthma development [18] and for lung function decline in non-asthmatic adults [19].

In asthma, also genes play a role in determining the susceptibility to the harmful effects of air pollution [20] but the underlying biological mechanisms of air pollution-mediated health effects are not fully understood, warranting further examination of the genes and pathways that might be involved.

Supplementary information The online version of this article (<https://doi.org/10.1038/s41370-019-0136-3>) contains supplementary material, which is available to authorized users.

✉ Despo Ierodiakonou
desierod@gmail.com

Extended author information available on the last page of the article

We previously investigated the longitudinal relationship between the 4-month average exposure to air pollution and post-bronchodilator (BD) forced expiratory volume in 1 second (FEV₁) and showed that among the measured air pollutants, long-term exposures to carbon monoxide (CO) and nitrogen dioxide (NO₂) are associated with reduced levels of FEV₁ in children with asthma [21]. In the current study, we use a hypothesis-free, genome-wide analysis to investigate whether genetic variants modify the long-term effects of CO and NO₂ on lung function in children with asthma, and with a pathway analysis we explore further plausible underlying biological pathways of CO and NO₂-mediated effects on lung function in asthmatic children.

Materials and methods

The Childhood Asthma Management Program (CAMP; ClinicalTrials.gov Identifier: NCT00000575) study design and methods have been described elsewhere [22]. Additional details on all methods used in the present report are provided in an online data supplement. In summary, children enrolled in CAMP were 5–12 years of age and had airway hyper-responsiveness to methacholine at study entry. In total, 1041 children entered the randomization (RZ) phase and 311, 312, and 418 children received budesonide, nedocromil, and placebo, respectively. All subjects were treated and followed for 4 years with visits at 2 and 4 months after RZ and at 4-month intervals thereafter. Each parent or guardian signed a consent form and participants of 7 years of age and older signed an assent form approved by each clinical center's institutional review board.

Spirometry before and after the administration of two puffs of albuterol (bronchodilator) was conducted at RZ and at follow-up visits ($n = 13$) according to the American Thoracic Society Standards [23]. Twenty-four hours average concentrations of CO and NO₂ were estimated for each metropolitan area using data from the United States Environmental Protection Agency's Atmospheric Integrated Research Monitoring Network. The ZIP or postal code centroid coordinates were used to link participants to daily concentrations from the nearest monitor within 50 km that did not have missing data on that day (December 1993 through June 1999). Averaging the daily pollution concentrations for the 4-month intervals between the clinic visits for lung function measurement created the moving averages.

Genome-wide single-nucleotide polymorphisms (SNP) genotyping for CAMP subjects (their families and iControlDB controls) was performed on Illumina's Human-Hap550 Genotyping BeadChip (Illumina, Inc., San Diego, CA).

Statistical analysis

Genome-wide interaction study

In a genome-wide interaction analysis, the computational effort needed to evaluate the effects of hundreds of thousands SNPs on the longitudinally measured trait is prohibitively large with a classic linear mixed model (LMM) approach. We followed the Sikorska et al. conditional two-step approach for fast LMM computations for genome-wide association studies (GWAS) [24], a method to explore whether the longitudinal relationship between 4-month averaged pollution (CO and NO₂) and post-BD FEV₁ % predicted is modified by SNPs in the human genome. The practical application of this approach is to be used as a surrogate of classic LMM, hence we performed the genome-wide scan for hundreds of thousands SNPs in a fast manner.

In summary, in the first step we fitted a LMM with subject-specific (random) intercept and slope for pollution exposure with all SNP terms omitted (main effect and interaction with pollutant) from the model. LMM tests were performed in the R programming language (version 3.5.0 (2018-04-23)), and code availability can be requested by the corresponding author.

At the second step, simple linear regression tests of SNPs (genome-wide) with the individual's FEV₁ response to CO and NO₂ (provided as subject-specific random slopes of pollution by LMM in step 1), respectively, were performed in PLINK [25], using an additive allelic model. SNPs included in the genome-wide analysis had a minor allele frequency >5% ($n = 474,792$).

For estimating the exact effect size of the interactions and confirm statistical significance, top signals for SNP–pollution interaction as given by two-step approach (P -value < 10^{-5}) were tested with the classic LMM including terms of pollution, SNP and SNP–pollution interaction, (e.g., Bonferroni corrected minimally significant P -value being $0.05/474,792 = 1.05\text{E-}07$). Non-Hispanic white (Caucasian) CAMP subjects ($n = 522$) were used as the primary study population and African-American CAMP subjects ($n = 88$) served as the replication study population.

Pathway-level analysis for the genome-wide SNP by pollutant interaction analysis

To analyze pathway-level SNP–pollutant interactions, we used the improved gene set enrichment analysis for GWAS (*i*-GSEA4GWAS; <http://gsea4gwas.psych.ac.cn/inputPage.jsp>) [26], GSEA evaluates whether the distribution of genes sharing a biochemical or cellular function is different from the distribution of a ranked genome-wide gene list

[26, 27]. Details on the *i-GSEA4GWAS* method are given in the Supplementary material.

Input data to perform the pathway-level analysis of the SNP–pollution interaction analysis were *P*-values of the two-step genome-wide SNP–pollution interaction analysis in Caucasian CAMP subjects. We changed the default settings and selected specific parameters for the gene set enrichment analysis; to avoid overrepresentation of SNPs in more than one gene we restricted mapping SNPs to ± 20 kb around a gene. We selected additional filtering for gene set size, set to at least five genes, so any narrow functional categories would not be missed. Next, the default canonical pathway method of gene sets was used for further analysis. These canonical pathways were extracted and curated from Molecular Signatures Database from a variety of online resources (MSigDB v2.5; <http://www.broadinstitute.org/gsea/msigdb/>). The genome-wide *P*-values were transformed to $-\log$ (*P*-values), represented genes were mapped based on SNPs *P*-values, and the enrichment score was calculated.

Significant genes in a pathway are defined as the genes mapped with at least one of the top 5% *P*-values of all SNPs ($0.05 \times 474,792 = 23,740$ SNPs). Each significant gene was represented by the SNP in that gene with the lowest genome-wide SNP–pollutant interaction *P*-value (top SNP per significant gene). We selected the top SNPs of all given pathways and with classic LMM, we estimated the interaction effect size for the gene sets most significant SNP–pollutant interactions in Caucasians.

Results

All subjects in CAMP considered in this analysis were randomized and followed up during the trial period. A total of 1003 of the 1041 randomized children (96.3%) had pollution data available of which 610 were studied in the genetic analysis. At study entry, the mean (SD) age was 9 (2.1) and geometric mean (min–max) PC_{20} 1.1 (0.02–2.5) mg/ml. Table 1 shows the main characteristics of the participants. In all, 82.5% of the children attended all visits during the 4-year trial (median number of completed visits = 14 (range: 1–14)). Each participant had a median of 10 (range: 1–10) post-BD lung function measurements. Repeated FEV_1 measurements increase the power of our statistical analysis to detect significant differences between means (8200 and 8600 observations for NO_2 and CO analysis, respectively). Tables S1 and S2 summarize the 4-month moving averages pollutant concentrations during December 1993–June 1999, with number of observations, percentiles, and interquartile range (IQR). CO and NO_2 were weakly correlated (spearman $\rho = 0.30$).

Table 1 Population characteristics

<i>N</i> = 1003	
City, <i>n</i> (%)	
Albuquerque	121 (12.1)
Baltimore	126 (12.6)
Boston	123 (12.3)
Denver	141 (14.1)
San Diego	122 (12.2)
Seattle	136 (13.6)
Saint Louis	133 (13.3)
Toronto	101 (10.1)
Sex, <i>n</i> (%)	
Males	602 (60)
Females	401 (40)
Treatment group; <i>n</i> (%)	
Placebo	407 (40.6)
Budesonide	298 (29.7)
Nedocromil	298 (29.7)
Ethnicity, <i>n</i> (%)	
Caucasians	677 (67.5)
African-Americans	137 (13.7)
Hispanics	97 (9.7)
Other	92 (9.2)
Annual income $\geq 30K$ USD, <i>n</i> (%)	
Yes	728 (76)
No	235 (24)
In utero smoking exposure, <i>n</i> (%)	
Yes	114 (14)
No	854 (86)
Pre-bronchodilator lung function at randomization, mean (SD)	
FEV_1 %predicted	93.8 (14.3)
FVC %predicted	104.0 (13.1)
$FEV_1/FVC\%$	79.7 (8.3)
Post-bronchodilator lung function at randomization, mean (SD)	
FEV_1 %predicted	103.0 (12.8)
FVC %predicted	106.5 (12.8)
$FEV_1/FVC\%$	85.5 (6.5)

FEV₁ forced expiratory volume in 1 s, *FVC*: forced vital capacity, *SD* standard deviation, $\geq 30K$ USD equal or more than 30,000 United State Dollars

Two-step genome-wide SNP by pollutant interaction analysis

Figure 1 presents an overview of our study design and results of the genome wide interaction study (GWIS). After minor allele frequency (MAF) pruning, 474,792 SNPs were included in the primary analysis, and the smallest *P*-values for SNP– NO_2 and SNP–CO interactions with the two-step approach were $1.37E-06$ and $2.04E-06$, respectively, showing only suggestive evidence for genome-wide

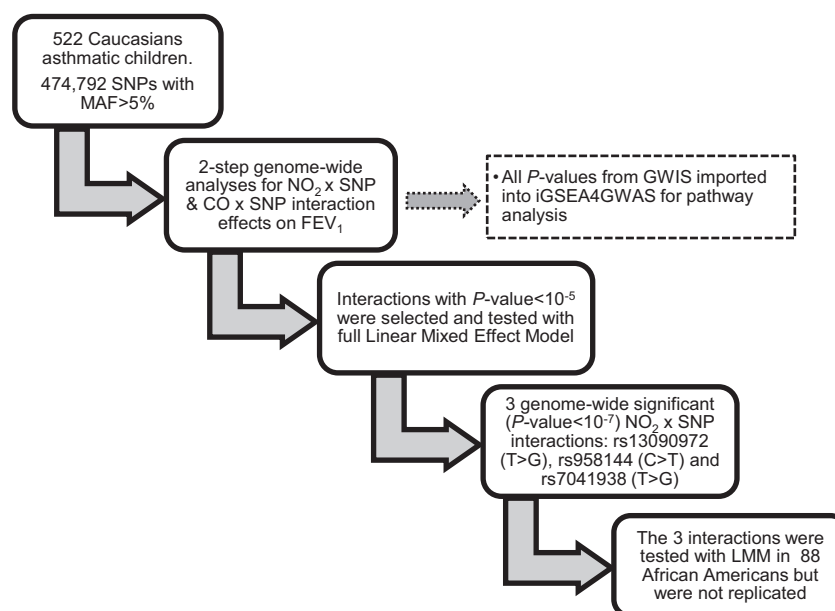


Fig. 1 The flowchart with the analytic steps and summary of results of the genome-wide gene by pollutant(s) interaction study. Top hit SNPs ($P < 10^{-5}$) interacting with pollutants in Caucasians were selected and with LMM we assessed the interaction effect size and P -values. Genome-wide significant interactions ($P < 10^{-7}$) were tested for

replication in African-Americans. CO carbon monoxide, NO₂ nitrogen dioxide, LMM linear mixed model, SNP single-nucleotide polymorphism, MAF minor allele frequency, GWAS genome-wide interaction study, iGSEA4GWAS improved gene set enrichment analyses for GWAS, FEV₁ forces expiratory volume in 1 s

SNP–pollutant interactions (Table 2 and Supplementary Tables S3A and S4). The quantile–quantile (QQ) plots of the two-step GWIS are presented in Supplementary Figs. S1 and S2, showing that the distribution of association P -values was similar to that expected for a null distribution, and that no P -values met the conventional genome-wide statistically significant levels (e.g., Bonferroni corrected minimally significant P -value being $0.05/474,792 = 1.05\text{E}-07$; see Supplementary Figs. S1 and S2).

Confirmation by classic LMM testing

We selected the six top (P -value $< 10^{-5}$) SNP–NO₂ interactions given by the two-step approach and with the classic LMM model we assessed the effect size of these interactions and compared their P -value as given by the two approaches. In Caucasians, change in post-BD FEV₁ % predicted per IQR increase in NO₂ level ranged from -1.3 to 1.1 for the six SNP–NO₂ interactions. With the classic LMM models, the P -values decreased for five out of six SNP–NO₂ interactions with values ranging from $1.3\text{E}-08$ to $8.5\text{E}-06$ (Supplementary Table S3A). Three SNP–NO₂ interactions reached genome-wide significance with the classic LMM: rs13090972 (80 kb 5' of EPHA3) and rs958144 (162 kb 5' of EPHA3) near EPHA3 (linkage disequilibrium (LD) between two SNPs $r^2 = 0.55$) and rs7041938 in TXNDC8—the latter in high LD ($r^2 = 0.8$) with rs12684188 in SVEP1 (Table 2). Similarly, in African-

Americans the P -values of associations were lower with LMM, but none reached genome-wide statistical significance (all P -values > 0.05 ; see table S3B). Supplementary Table S4 shows that the seven top signals (P -value $< 10^{-5}$) SNP–CO interactions as given by the two-step approach did not reach genome-wide statistical significance with LMM. The change in post-BD FEV₁ % predicted per IQR increase in CO level ranged from -0.98 to 0.83 and P -values range from $9.69\text{E}-07$ to $1.26\text{E}-05$.

Pathway-level analysis for the two-step genome-wide SNP by pollutant interaction analysis on FEV₁ %predicted

For the *i-GSEA4GWAS* in Caucasian CAMP subjects, $-\log(P$ -values) of 474,792 gene variants were imported and 265,485 variants were mapped on genes ± 20 kb (total number of genes: 16,854). We identified one pathway interacting with CO (P -value = 0.001) and 23 pathways interacting with NO₂ (P -values: 0.0001–0.01). Table S5 presents the *i-GSEA4GWAS* suggested pathways for the two pollutants. Details for each individual pathway (SNPs, mapped genes, gene sets, FDR, P -value, description) of NO₂ and CO-mediated effects can be found http://gsea4gwas.psych.ac.cn/getResult.do?result=13F3A972887892430E6A5C369D76FEAD_1372284527739 and http://gsea4gwas.psych.ac.cn/getResult.do?result=13F3A972887892430E6A5C369D76FEAD_1372283303807,

Table 2 Top genome-wide gene by nitrogen dioxide interaction loci and suggested functions

Top genome-wide interaction locus	SNP	Two-step approach <i>P</i> -values	Classic LMM <i>P</i> -value	LMM change per IQR for SNPs of <i>EPHA3</i>	Function(s) related to genes ^a	Identified pathways ^b linked to those gene functions
Near <i>EPHA3</i> ; chr3:88994672 and 89076896	rs13090972 (T>G) rs958144 (C>T)	1.37E-06 4.81E-06	1.33E-08* 2.94E-08*	-1.33 -1.28	Receptor tyrosine kinase of Eph family; cell adhesion; immune surveillance; tissue remodeling	Cell adhesion molecules; calcium regulation; glycosphingolipids metabolism, glycosaminoglycan (chondroitin) biosynthesis
<i>TXNDC8</i> ; chr9:113091523	rs7041938 (T>G)	7.35E-06	1.04E-07*	1.14	Thioredoxin reductase family; cell redox homeostasis	HSP27, iNOS, IL10, heme biosynthesis-heme oxygenase-1/CO, calcium regulation
In LD=0.8 with <i>SEVP1</i> ; chr9:113133588	rs12684188 (C>T)	>1E-05	3.88E-07	1.20	Cell adhesion; immune surveillance	Cell adhesion molecules, glycosphingolipids metabolism, calcium regulation

SNP single-nucleotide polymorphism, LMM linear mixed model, IQR interquartile range of 4-month average nitrogen dioxide concentration (4 parts per billion).

*Genome-wide significance ($P < 1.05E-07$)

^aBased on coding protein's function(s). Details for each gene are given in the text

^bBased on pathway analysis

respectively. All the pathways we present in our findings had false discovery rate (FDR) < 0.25. In summary, the *i-GSEA4GWAS* showed a pathway (PAC1R; receptor of pituitary adenylate cyclase-activating polypeptide (PACAP)) to be associated with CO-related FEV₁ changes. For NO₂-related FEV₁ responses, we identified several pathways involved in inflammation, oxidative stress, the HO-1/CO system, calcium homeostasis, cellular adhesion and metabolic responses.

Within each gene set/pathway there were significant genes (genes mapped with at least one of the top 5% of all SNPs-pollutant interactions in the two-step genome-wide analysis). Each significant gene is represented by the SNP in that gene with the lowest genome-wide *P*-value of SNP by pollutant interaction (the top SNP per significant gene). Effect sizes of interaction of those SNPs with pollutants as given by LMM are shown in the Supplementary material (see Supplementary Tables S6 and S7).

Discussion

Most gene-air pollution studies have focused on a few candidate genetic variations and investigated short-term exposures to pollution [17]. Although these small hypothesis-driven studies can contribute to our understanding of specific gene-pollution effects, they often fail to uncover novel disease-causing mechanisms and in some cases have not been replicated by subsequent studies [28, 29]. To the best of our knowledge, this is the first longitudinal GWIS on lung function response to ambient air pollution in asthmatic children. We used the two-step approach as a screening tool to identify genes that may interact with air pollution while gaining computational time, and we confirmed the top hits of the two-step approach with the classic LMM; we used the genome-wide output for a *iGSEA4GWAS*. Below we discuss the putative genes involved in air pollution effects on lung function in childhood asthma and the identified pathways.

At SNP level, two loci, the *EPHA3* (receptor tyrosine kinase of Eph family; location 3p11.2) and *TXNDC8* (thioredoxin domain containing 8 (spermatozoa) or Spermatoocyte/Spermatid-Specific Thioredoxin-3; location 9q31.3) genes showed genome-wide statistical evidence for interaction with NO₂ (with the classic LMM). The best-documented function of the Eph-receptor/ephrin-A signaling is the regulation of cell adhesion and migration processes critical for a wide variety of normal and pathological processes, including tissue remodeling and immune surveillance [30, 31]. Recent findings suggest that Eph signaling is involved in pathological conditions such as lung cancer, yet its role in asthma is unknown [32, 33]. The fact that receptor tyrosine kinase pathways contribute to aspects

of airway inflammation, airway hyper-responsiveness and remodeling of asthma [34], suggests that we may have identified a novel receptor tyrosine kinases (EPHA3) important for the pathogenesis of asthma in response to NO₂ in Caucasian children.

The second top signal locus, *TXNDC8*, belongs to the thioredoxin reductase enzymes, a well-characterized sub-family of selenoproteins that perform an essential redox role in immune cells [35]. Recent studies indicated that thioredoxin system may contribute to the pathogenesis of COPD, asthma and lung injury and suggest that this pathway may be used in future therapeutic applications [36]. The genome-wide top hit SNP (rs7041938) in *TXNDC8* found to modify the NO₂ effects on FEV₁ in Caucasian subjects is in high linkage disequilibrium ($r^2 > 0.8$) with rs12684188 in *SVEPI* (sushi, von Willebrand factor type A, EGF and pentraxin domain containing 1; location 9q32). In a recent GWAS, a locus containing the *SVEPI* gene showed signals of association with FEV₁ decline in non-asthmatic adults [37]. In our asthmatic children, the interaction *P*-value of the *SVEPI* variant did not reach significance. *SVEPI* codes for a protein called polydom, which is recognized as a cell adhesion molecule with a biological role in cellular adhesion and/or in the immune system [38, 39]; but its role in asthma has not been investigated so far. We were unable to replicate these loci in African-Americans and it would be important to replicate our finding in other populations in the future.

The pathway analysis helps to clarify biological plausible connections for our GWIS hits with one another. Some of the pathways identified from our *iGSEA4GWAS* analysis have been previously found to play a role in asthma and be related to cellular adhesion and immune response, as do so our GWIS top loci. The first genome-wide gene by air pollution interaction study on asthma development identified genes involved in glycosphingolipids biosynthesis, G-protein coupled receptor signaling and adhesion [18]. Similarly, our *iGSEA4GWAS* identified sphingolipid (glycosphingolipid metabolism pathway), G-coupled receptor (gs-pathway, agpcr-pathway, plce-pathway), and epithelial adhesion (HSA04514 cell adhesion molecules pathway) pathways in lung function response to NO₂, pointing to the same direction. Sphingolipids and altered sphingolipid metabolism have emerged as potential key contributors to the pathogenesis of asthma [40]. Orosomucoid-like 3 gene (*ORMDL3*) and the asthma susceptibility locus 17q21 have been strongly and reproducibly linked to childhood asthma [41].

The role of airway epithelial barrier function (HSA04514 cell adhesion molecules pathway) in the susceptibility to develop allergic asthma has been extensively studied and polymorphisms in adhesion molecules genes have been associated with asthma and asthma severity [42–44]. It is

plausible that exposure to NO₂ induces oxidative stress with cellular barrier damage and inflammatory responses. In the Supplementary material, we describe in more detail how pathways involved in inflammation and oxidative stress (NOS1, HSP27, IL10, heme biosynthase) may be linked to NO₂ exposure and how they are inter-related.

Metabolic pathways (feeder of glycolysis and obesity pathways) are activated to compensate the cellular demands to stress and the HO-1/CO system may protect against oxidative stress and inflammation. In line with our findings, a GWIS study of non-asthmatic adults, identified a mechanistic link between adiponectin (a metabolic biomarker with modulating action on inflammatory processes systemically and locally in the lung) and cadherin 13 as a biologically plausible pathway for modifying the air pollution exposure effect on lung decline [19].

Oxidative stress has been associated with calcium influx regulation, two responses observed in our pathway analysis as well [45, 46]. Interestingly, a proteomic-based study has shown that allergen-induced early asthma response in rats is associated with glycolysis, calcium binding, and mitochondrial activity [47], supporting our identified underlying molecular mechanisms for response to environmental toxicants in asthma.

The *iGSEA4GWAS* of CO interactions suggested the neuropeptide pituitary adenylate cyclase-activating peptide receptor (PAC1R) pathway in CO-related response. The ligand of PAC1R (PACAP) can induce bronchodilation and endogenous regulation of airway tone by means of a CO-dependent mechanism with local HO-1/CO release in the airway smooth muscle, and it also has pro-inflammatory functions that require calcium regulation [48–50]. Furthermore, PACAP, acting through type 1 PACAP receptor, exerts a potent protective effect against oxidative stress-induced apoptosis [51].

Our childhood asthma study had the advantage of having a long follow-up period with high attendance of the subjects and repeated lung function measurements, air pollution levels during that period and genomic data. The two-step approach used for longitudinal data [24] provided shorter processing time and we confirmed its accuracy, i.e., at a second stage the genome-wide top signals found by the two-step approach were confirmed by LMM testing.

We could not find a second study of asthmatic children with similar design, repeated lung function measurements, population characteristics, genome-wide genotyping, and air pollution data. Although population stratification is less likely to bias estimates of gene–environment interaction effects [52], we used as our primary study only Caucasian CAMP subjects and found no evidence of stratification in our Q/Q plots. For replication studies, definition and measurement of the exposure and/or outcome is critical to the success of gene–environment investigations, therefore we

decided to use the second largest ethnic subgroup of the CAMP as our replication population (although of relative small size), to ensure that the genotyping, outcome, and exposure were measured reliably and consistently. This reduced power and potential different LD patterns in the replication population represent limitations of this study.

After testing for pollution effect modification at the SNP level, we performed the pathway approach to assess the overall evidence of interaction of pollution with a group of functionally related genes, thus incorporating prior biological knowledge. Our pathway-level analysis of SNP–pollution interactions identified biological plausible mechanisms for pollution-mediated asthma progression in children that are generally consistent with the SNP-level analysis.

Our findings highlight the promise of pursuing genome-wide gene–environment interaction studies in smaller populations with high-quality longitudinal exposure information by showing that they can identify biologically relevant effects of these exposures. We conclude that genetic susceptibility to traffic-related air pollutants such as CO and NO₂ are linked to oxidative stress and inflammation pathways, while metabolic pathways, calcium homeostasis and the HO-1/CO pathway may play a cytoprotective role against oxidative stress and inflammation. Our findings may represent the first step for functional research and pharmacological developments for protection against the detrimental effects of air pollution on asthma severity and progression.

Acknowledgements We would like to thank Steve Melly for his contribution on the air pollution database preparation and our colleagues Paul V. Williams, Teal S. Hallstrand, and Anne N. Fuhlbrigge for the Childhood Management Asthma (CAMP) Program Group. We dedicate this manuscript to the memory of our friend and colleague Dr. Gail G. Shapiro who passed away unexpectedly during this study. Dr. Shapiro dedicated her life to understanding the causes of childhood asthma and determining the best treatments for asthma. She is deeply missed by her colleagues, patients, and the asthma community. A special thank you to all participants of the CAMP study and their families.

Author contributions Each author participated sufficiently in the current work. All authors were involved in the conception, hypotheses delineation, and design of the present article. DI wrote the article and all authors had a substantial involvement in its revision prior to submission. Management of the data and the analysis was performed by DI in consultation with BAC, AZ, DRG, and STW. DRG, PK, and JS provided comprehensive input on air pollution exposure assessment and modeling. STW, DSP, JV, HMB, and GHK supported the genome-wide and pathway analyses. DC C-C provided input on bioinformatic tools. EFMck, JSS, TL, and STW represent the CAMP research group who designed, conducted, and completed the study.

Funding The Childhood Asthma Management Program trial and CAMP Continuation Study were supported by contracts NO1-HR-16044, 16045, 16046, 16047, 16048, 16049, 16050, 16051, and 16052 with the National Heart, Lung, and Blood Institute and General Clinical Research Center grants M01RR00051, M01RR0099718-24, M01RR02719-14,

and RR00036 from the National Center for Research Resources. The CAMP Continuation Study/Phases 2 and 3 were supported by grants U01HL075232, U01HL075407, U01HL075408, U01HL075409, U01HL075415, U01HL075416, U01HL075417, U01HL075419 and U01HL075420 from the National Heart, Lung, and Blood Institute. The National Jewish Health site was also supported in part by Colorado CTSA grant UL1RR025780 from NCRR/NIH and UL1TR000154. In addition, all work on data collected from the CAMP Genetic Ancillary Study was conducted at the Channing Laboratory of the Brigham and Women's Hospital under appropriate CAMP policies and human subject's protections. The CAMP Genetics Ancillary Study is supported by U01 HL075419, U01 HL65899, P01 HL083069, R01 HL086601, and RC2 HL101543 from the National Heart, Lung and Blood Institute, National Institutes of Health. This study was also funded by: the National Institutes of Health (NHLBI P01 HL083069, U01 HL075419, U01 HL65899, R01 HL086601; NIEHS P01 ES09825, R21 ES020194, P30 ES000002); the U.S. Environmental Protection Agency (RD 83241601, RD 83479801), and the International Initiative for Environment and Public Health Cyprus Program of HSPH. The contents of this publication are solely the responsibility of the grantee and do not necessarily represent the official views of the US EPA. Further, US EPA does not endorse the purchase of any commercial products or services mentioned in the publication.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

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

References

1. Searing DA, Rabinovitch N. Environmental pollution and lung effects in children. *Curr Opin Pediatr*. 2011;23:314–8.
2. Schwela D. Air pollution and health in urban areas. *Rev Environ Health*. 2000;15:13–42.
3. World Health Organization. Effects of air pollution on children's health and development. A review of the evidence. Bonn. http://www.euro.who.int/__data/assets/pdf_file/0010/74728/E86575.pdf. 2005. Date last accessed: April 2013.
4. Schildcrout JS, Sheppard L, Lumley T, Slaughter JC, Koenig JQ, Shapiro GG. Ambient air pollution and asthma exacerbations in children: an eight-city analysis. *Am J Epidemiol*. 2006;164:505–17.
5. Sunyer J, Spix C, Querel P, Ponce-de-Leon A, Ponka A, Barumandzadeh T, et al. Urban air pollution and emergency admissions for asthma in four European cities: the APHEA project. *Thorax*. 1997;52:760–5.
6. Romieu I, Meneses F, Ruiz S, Sienra JJ, Huerta J, White MC, et al. Effects of air pollution on the respiratory health of asthmatic children living in Mexico city. *Am J Respir Crit Care Med*. 1996;154:300–7.
7. HEI panel on the Health Effects of Traffic-Related Air Pollution. Traffic-related air pollution: a critical review of the literature on emissions, exposure, and health effects. Boston, MA: 2010. Report No.: HEI Special Report 17.
8. Li S, Williams G, Jalaludin B, Baker P. Panel studies of air pollution on children's lung function and respiratory symptoms: a literature review. *J Asthma*. 2012;49:895–910.
9. Gao Y, Chan EY, Li LP, He QQ, Wong TW. Chronic effects of ambient air pollution on lung function among Chinese children. *Arch Dis Child*. 2013;98:128–35.

10. Rosenlund M, Forastiere F, Porta D, De Sario M, Badaloni C, Perucci CA. Traffic-related air pollution in relation to respiratory symptoms, allergic sensitisation and lung function in school-children. *Thorax*. 2009;64:573–80.
11. Gauderman WJ, Vora H, McConnell R, Berhane K, Gilliland F, Thomas D, et al. Effect of exposure to traffic on lung development from 10 to 18 years of age: a cohort study. *Lancet*. 2007;369:571–7.
12. Schwartz J. Lung function and chronic exposure to air pollution: a cross-sectional analysis of NHANES II. *Environ Res*. 1989;50:309–21.
13. Kelly FJ. Oxidative stress: its role in air pollution and adverse health effects. *Occup Environ Med*. 2003;60:612–6.
14. Patel MM, Chillrud SN, Deepti KC, Ross JM, Kinney PL. Traffic-related air pollutants and exhaled markers of airway inflammation and oxidative stress in new york city adolescents. *Environ Res*. 2013;121:71–8.
15. Emmerechts J, Hoylaerts MF. The effect of air pollution on haemostasis. *Hamostaseologie*. 2012;32:5–13.
16. Auerbach A, Hernandez ML. The effect of environmental oxidative stress on airway inflammation. *Curr Opin Allergy Clin Immunol*. 2012;12:133–9.
17. Yang IA, Fong KM, Zimmerman PV, Holgate ST, Holloway JW. Genetic susceptibility to the respiratory effects of air pollution. *Postgrad Med J*. 2009;85:428–36.
18. Gref A, Kebede Merid S, Gruziova O, Ballereau S, Becker A, Bellander T, et al. Genome-wide interaction analysis of air pollution exposure and childhood asthma with functional follow-up. *Am J Respir Crit Care Med*. 2016;195:1373–83.
19. Imboden M, Kumar A, Curjuric I, Adam M, Thun GA, Haun M, et al. Modification of the association between PM10 and lung function decline by cadherin 13 polymorphisms in the SAPALDIA cohort: a genome-wide interaction analysis. *Environ Health Perspect*. 2015;123:72–9.
20. Moreno-Macias H, Dockery DW, Schwartz J, Gold DR, Laird NM, Sienra-Monge JJ, et al. Ozone exposure, vitamin C intake, and genetic susceptibility of asthmatic children in Mexico city: a cohort study. *Respir Res*. 2013;14:14.
21. Ierodiakonou D, Zanobetti A, Coull BA, Melly S, Postma DS, Boezen HM, et al. Ambient air pollution, lung function, and airway responsiveness in asthmatic children. *J Allergy Clin Immunol*. 2016;137:390–9.
22. The Childhood Asthma Management Program Research Group. The childhood asthma management program (CAMP): design, rationale, and methods. childhood asthma management program research group. *Control Clin Trials*. 1999;20:91–120.
23. American Thoracic Society. Standardization of spirometry, 1994 update. *Am J Respir Crit Care Med*. 1995;152:1107–36.
24. Sikorska K, Rivadeneira F, Groenen PJ, Hofman A, Uitterlinden AG, Eilers PH, et al. Fast linear mixed model computations for genome-wide association studies with longitudinal data. *Stat Med*. 2013;32:165–80.
25. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81:559–75.
26. Zhang K, Cui S, Chang S, Zhang L, Wang J. i-GSEA4GWAS: a web server for identification of pathways/gene sets associated with traits by applying an improved gene set enrichment analysis to genome-wide association study. *Nucleic Acids Res*. 2010;38:W90–5. (Web Server issue).
27. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA*. 2005;102:15545–50.
28. Romieu I, Moreno-Macias H, London SJ. Gene by environment interaction and ambient air pollution. *Proc Am Thorac Soc*. 2010;7:116–22.
29. Bowatte G, Lodge CJ, Perret JL, Matheson MC, Dharmage SC. Interactions of GST polymorphisms in air pollution exposure and respiratory diseases and allergies. *Curr Allergy Asthma Rep*. 2016;16:85.
30. Miao H, Wang B. EphA receptor signaling—complexity and emerging themes. *Semin Cell Dev Biol*. 2012;23:16–25.
31. Arvanitis D, Davy A. Eph/ephrin signaling: networks. *Genes Dev*. 2008;22:416–29.
32. Pasquale EB. Eph-ephrin bidirectional signaling in physiology and disease. *Cell*. 2008;133:38–52.
33. Zhuang G, Song W, Amato K, Hwang Y, Lee K, Boothby M, et al. Effects of cancer-associated EPHA3 mutations on lung cancer. *J Natl Cancer Inst*. 2012;104:1182–97.
34. Guntur VP, Reinero CR. The potential use of tyrosine kinase inhibitors in severe asthma. *Curr Opin Allergy Clin Immunol*. 2012;12:68–75.
35. Huang Z, Rose AH, Hoffmann PR. The role of selenium in inflammation and immunity: from molecular mechanisms to therapeutic opportunities. *Antioxid Redox Signal*. 2012;16:705–43.
36. Xu J, Li T, Wu H, Xu T. Role of thioredoxin in lung disease. *Pulm Pharmacol Ther*. 2012;25:154–62.
37. Imboden M, Bouzigon E, Curjuric I, Ramasamy A, Kumar A, Hancock DB, et al. Genome-wide association study of lung function decline in adults with and without asthma. *J Allergy Clin Immunol*. 2012;129:1218–28.
38. Schwanzer-Pfeiffer D, Rossmanith E, Schildberger A, Falkenhagen D. Characterization of SVEP1, KIAA, and SRPX2 in an in vitro cell culture model of endotoxemia. *Cell Immunol*. 2010;263:65–70.
39. Gilges D, Vinit MA, Callebaut I, Coulombel L, Cacheux V, Romeo PH, et al. Polydom: a secreted protein with pentraxin, complement control protein, epidermal growth factor and von willebrand factor A domains. *Biochem J*. 2000;352:49–59. Pt 1.
40. Ono JG, Worgall TS, Worgall S. Airway reactivity and sphingolipids-implications for childhood asthma. *Mol Cell Pediatr*. 2015;2:13.
41. Zhao CN, Fan Y, Huang JJ, Zhang HX, Gao T, Wang C, et al. The association of GSDMB and ORMDL3 gene polymorphisms with asthma: a meta-analysis. *Allergy Asthma Immunol Res*. 2015;7:175–85.
42. Heijink IH, Nawijn MC, Hackett TL. Airway epithelial barrier function regulates the pathogenesis of allergic asthma. *Clin Exp Allergy*. 2014;44:620–30.
43. Faura Tellez G, Willemse BW, Brouwer U, Nijboer-Brinksma S, Vandepoele K, Noordhoek JA, et al. Protocadherin-1 localization and cell-adhesion function in airway epithelial cells in asthma. *PLoS ONE*. 2016;11:e0163967.
44. Ierodiakonou D, Postma DS, Koppelman GH, Boezen HM, Gerlitsen J, Ten Hacken N, et al. E-cadherin gene polymorphisms in asthma patients using inhaled corticosteroids. *Eur Respir J*. 2011;38:1044–52.
45. Jiang LH, Yang W, Zou J, Beech DJ. TRPM2 channel properties, functions and therapeutic potentials. *Expert Opin Ther Targets*. 2010;14:973–88.
46. Xu R, Li Q, Zhou XD, Perelman JM, Kolosov VP. Oxidative stress mediates the disruption of airway epithelial tight junctions through a TRPM2-PLCgamma1-PKCalpha signaling pathway. *Int J Mol Sci*. 2013;14:9475–86.
47. Xu YD, Cui JM, Wang Y, Yin LM, Gao CK, Liu YY, et al. The early asthmatic response is associated with glycolysis, calcium binding and mitochondria activity as revealed by proteomic analysis in rats. *Respir Res*. 2010;11:9921–11–107.

48. Kinhult J, Andersson JA, Uddman R, Stjarne P, Cardell LO. Pituitary adenylate cyclase-activating peptide 38 a potent endogenously produced dilator of human airways. *Eur Respir J*. 2000;15:243–7.
49. Kinhult J, Uddman R, Cardell LO. The induction of carbon monoxide-mediated airway relaxation by PACAP 38 in isolated guinea pig airways. *Lung*. 2001;179:1–8.
50. Linden A, Cardell LO, Yoshihara S, Nadel JA. Bronchodilation by pituitary adenylate cyclase-activating peptide and related peptides. *Eur Respir J*. 1999;14:443–51.
51. Douiri S, Bahdoudi S, Hamdi Y, Cubi R, Basille M, Fournier A, et al. Involvement of endogenous antioxidant systems in the protective activity of pituitary adenylate cyclase-activating polypeptide against hydrogen peroxide-induced oxidative damages in cultured rat astrocytes. *J Neurochem*. 2016;137: 913–30.
52. Thomas D. Gene–environment-wide association studies: emerging approaches. *Nat Rev Genet*. 2010;11:259–72.

Affiliations

Despo Ierodiakonou^{1,2} · Brent A. Coull³ · Antonella Zanobetti⁴ · Dirkje S. Postma^{2,5} · H. Marika Boezen^{1,2} · Judith M. Vonk^{1,2} · Edward F. McKone⁶ · Jonathan S. Schildcrout⁷ · Gerard H. Koppelman^{2,8} · Damien C. Croteau-Chonka⁹ · Thomas Lumley¹⁰ · Petros Koutrakis⁴ · Joel Schwartz⁴ · Diane R. Gold^{4,9} · Scott T. Weiss⁹

¹ Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

² Groningen Research Institute for Asthma and COPD, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

³ Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, United States

⁴ Environmental Epidemiology and Risk Program, Harvard T.H. Chan School of Public Health, Boston, MA, United States

⁵ Department of Pulmonology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

⁶ Department of Respiratory Medicine, St. Vincent University

Hospital, Dublin, Ireland

⁷ Department of Environmental and Occupational Health Sciences, School of Public Health, University of Washington, Seattle, WA, United States

⁸ Department of Pediatric Pulmonology and Pediatric Allergy-Beatrix Children Hospital, University of Groningen, University Medical Center, Groningen, The Netherlands

⁹ Channing Division of Network Medicine, Brigham and Women's Hospital, Department of Medicine, Harvard Medical School, Boston, MA, United States

¹⁰ Department of Biostatistics, University of Auckland, Auckland, New Zealand

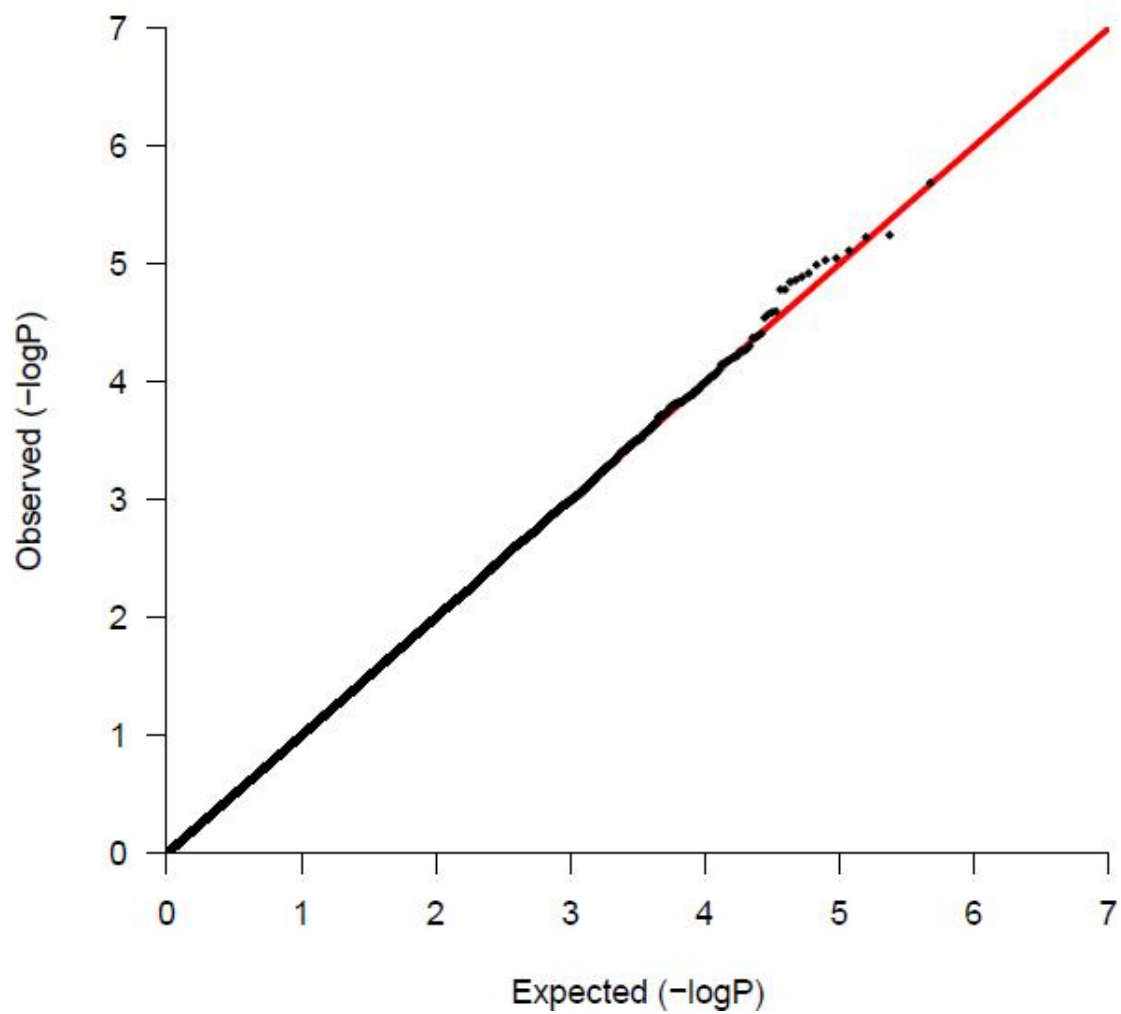


Figure S1: The quantile-quantile (QQ) plot of the genome-wide SNP by carbon monoxide interaction study on post-BD FEV₁ %predicted in Caucasian CAMP subjects

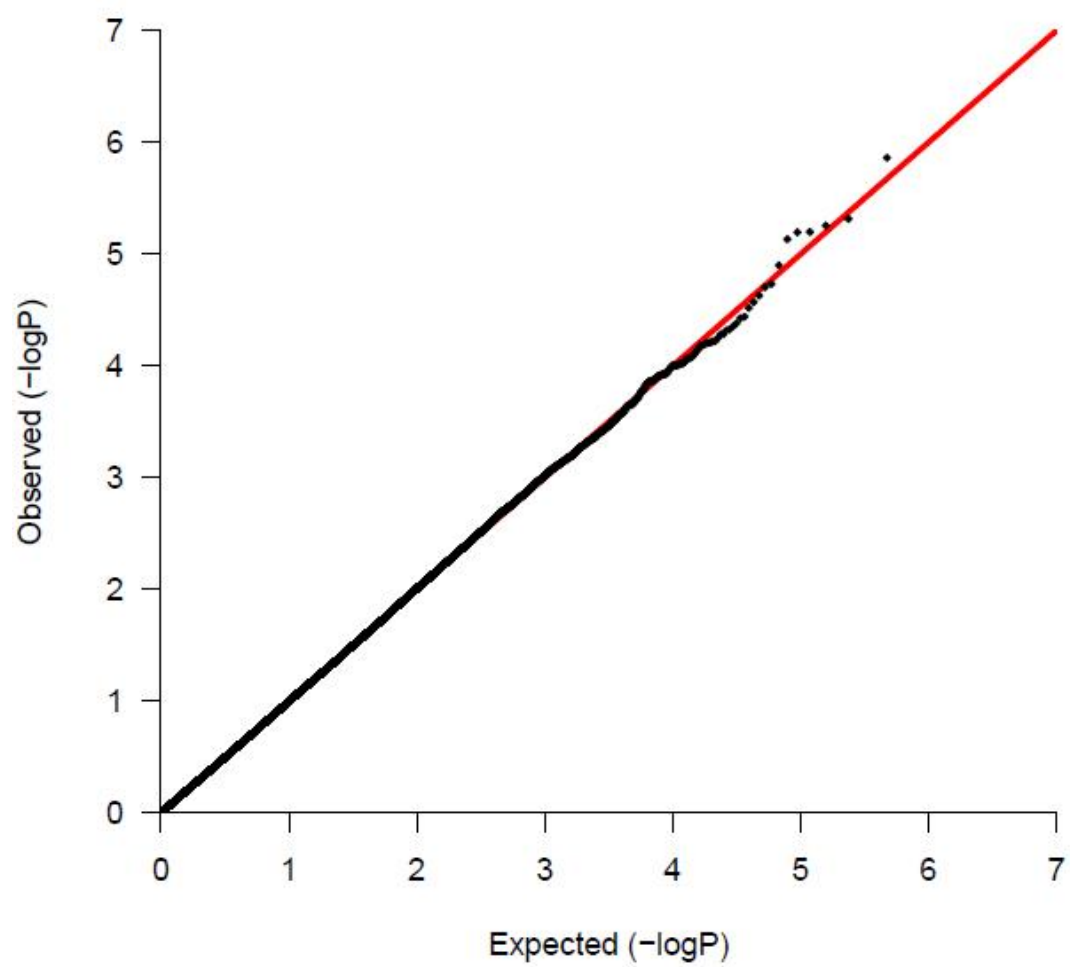


Figure S2: The quantile-quantile (QQ) plot of the genome-wide SNP by nitrogen dioxide interaction study on post-BD FEV₁ %predicted in Caucasian CAMP subjects

Online Supplement – materials-methods & discussion

Title: Pathway analysis of a genome-wide gene by air pollution interaction study in asthma

Methods

The Childhood Asthma Management Program (CAMP): is a multi-center, double-blinded clinical trial involving eight clinical centers in North America (Albuquerque, New Mexico; Baltimore, Maryland; Boston, Massachusetts; Denver, Colorado; San Diego, California; Seattle, Washington; St. Louis, Missouri; and Toronto-Ontario, Canada). Its main goal was to evaluate the long-term effectiveness and safety of daily-inhaled anti-inflammatory medication in children diagnosed with asthma. Details of the CAMP study design and methods have been described elsewhere¹. In summary, children enrolled in CAMP were 5–12 years of age at the initial interview, and completed a 5-16 week pre-randomization screening period (1993-1995). All participants had mild-to-moderate asthma with increased airway responsiveness to methacholine (a provocation concentration causing a 20% reduction in FEV₁ [PC₂₀] ≤ 12.5 mg/ml) at study entry. 1,041 children entered the randomization phase and 311, 312, 418 children received budesonide, nedocromil, and placebo, respectively. All subjects were treated and followed for four years with follow-up visits at CAMP centers at two and four months after randomization and at four-month intervals thereafter. Each parent or guardian signed a consent form and participants of 7 years of age and older signed an assent form approved by each clinical center's institutional review board.

Outcomes Measures

Spirometry, before and after the administration of two puffs of albuterol, was conducted at randomization (RZ) and at follow up visits (n=13) according to the American Thoracic Society Standards². The focus of the trial on long-term effects of the treatments on lung function was reflected in the choice of post-bronchodilator FEV₁ percentage of predicted (post-BD FEV₁ % predicted) as the primary outcome measure³. Use of post-BD FEV₁ rather than the pre-BD measure minimized fluctuations caused by diurnal variation and day-to-day variability of airway tone.

Air Pollution Exposure Assessment

Data on concentrations of 2 gaseous pollutants (carbon monoxide (CO) and nitrogen dioxide (NO₂)) were obtained from the Aerometric Information Retrieval System (www.epa.gov/air/data/index.html) for US cities and from the Ontario Ministry of the Environment (www.ene.gov.on.ca/environment; Air Quality Assessment and Reporting Unit) for Toronto. For each metropolitan area, multiple monitors were used to measure ambient pollution concentrations. Monitor-specific concentrations were 24-hour averages for carbon monoxide and nitrogen dioxide. In most cases, for US participants we were able to link ZIP codes of their address at study entry to US Census 2000 ZIP Code Tabulation Areas to identify the latitudes and longitudes of the centroids of those areas. In a few cases ZIP Code centroid coordinates were obtained from www.zip-codes.com. Coordinates for Canadian postal codes were obtained from www.batchgeo.com. The ZIP or postal code centroid coordinates were used to link participants to daily concentrations from the nearest monitor within 50 km that did not have missing data on that day (December 1993 through June 1999). If none of the

monitors within 50 km radius was operating the air pollution concentration of that day was set as missing.

Exclusion criteria

Participants were excluded if we were unable to assign their ZIP or postal codes to latitudes and longitudes or if we verified that the centroid of the ZIP or postal code in which they lived was greater than 50 km from the nearest pollutant monitor. For the Canadian participants, a few monitors northwest of Toronto became operational after 2000 and subjects living in postal codes near those monitors were excluded. In total 38 subjects of the 1041 trial participants (3.7%) were excluded.

Genotyping and Quality Control

Genome-wide single nucleotide polymorphisms (SNP) genotyping for CAMP subjects (their families and iControlDB controls) was performed on Illumina's HumanHap550 Genotyping BeadChip (Illumina, Inc., San Diego, CA). CAMP samples and markers passed stringent quality-control standards; 6,257 markers were removed as a result of low clustering scores. Markers whose flanking sequences did not map to a unique position on the HG17 reference genome sequence were removed (n = 1,329). Further quality control was performed with PLINK (version 1.07; 2009-09-10). The average completion rate for each marker was more than 99%. Monomorphic markers (n = 3,790) and those with five or more Mendel errors (n = 2,445) were removed. The average genotyping completion rate for each subject was 99.75%.

Statistical Analysis

Linear Mixed Model used in the present study

We examined the relation between the level of post-BD FEV₁ %predicted and ZIP-code level 4-month moving average concentrations of carbon monoxide and nitrogen dioxide. Averaging the daily pollution concentrations for the 4-month intervals between the clinic visits for lung function measurement created the moving averages. For any given averaging period, we required that 75% of the daily pollution data were available.

We fitted a LMM, with random intercepts and random slope for pollutant for each subject, to estimate the associations between FEV₁ and ambient air pollutant concentrations. $FEV_1 = b_0 + b_1 \text{ between-subject pollution} + b_2 \text{ within-subject pollution} + b_3 \text{ covariates} + e \mid \text{subject-specific random slope and intercept for pollution.}$

The number of days from randomization was used as the time trend in the model. To estimate associations across all cities, we constructed a model including city as a covariate. Potential for confounding by a number of factors was considered carefully, basing choice of covariates on prior CAMP experience⁴: we controlled for gender, ethnicity, *in utero* smoking exposure, annual family income (dichotomized at \$15K per annum), treatment group and treatment by time interaction. Confounding due to seasonal factors is a primary consideration in air pollution epidemiology, and we adjusted for “season” by using sine and cosine functions of time⁵ and their interactions with city. In addition, we decomposed daily ambient pollution concentrations into between- and within-subject exposures. Specifically, subject *i*’s exposure to pollutant x_1 on day *t* was decomposed as follows: The between-subject

component, is subject i 's average exposure during the time in which he or she was observed, and the within-subject component, is the deviation from the average on day t . Within-subject effects are likely to be far less confounded by season or by covariates that vary primarily across individual (e.g. in utero smoking exposure, etc.) than between-subject effects⁶, and they are more comparable to the parameters estimated in a traditional panel study, where all subjects are observed over a single time period. For further discussion regarding the decomposition of time-varying exposures into between- and within-subject components, see the papers by Sheppard⁶ and Neuhaus and Kalbfleisch⁷.

We used SAS[®] software (version 9.3; SAS Institute Inc. 2011, Cary, NC USA) and IBM SPSS statistics (version 22; Armonk, NY USA: IBM Corp 2013) to manage all data. Statistical analysis was performed in IBM SPSS and the R programming language (version 3.2.2 (2015-08-14)).

Pathway-level analysis for the genome-wide SNP by pollutant interaction analysis

Gene set enrichment analysis (GEAS) evaluates whether the distribution of genes sharing a biochemical or cellular function is different from the distribution of a ranked genome-wide gene list^{8, 9}. Input data to perform the pathway-level analysis of the SNP-pollution interaction analysis were P -values of the two-step genome-wide SNP-pollution interaction analysis in Caucasian CAMP subjects. With the *i*-GSEA4GWAS method the maximum $-\log(P\text{-value})$ of all SNPs mapped to (and around) a gene is used to represent the gene and the enrichment score was calculated. To avoid overrepresentation of SNPs in more than one gene we restricted mapping SNPs to +/-

20kb around a gene. Next, the canonical pathway method of gene-sets was used for further analysis. These canonical pathways were extracted and curated from Molecular Signatures Database (MSigDB) v2.5 (<http://www.broadinstitute.org/gsea/msigdb/>). MSigDB included gene-sets denoting canonical pathways integrated from a variety of online resources. We set our analysis to pathways with 5–200 genes. All genome-wide represented genes are ranked by decreasing values and their distribution is compared to the distribution of genes sharing a biochemical or cellular function. Kolmogorov-Smirnov like statistics gives to each gene set an enrichment score that reflects the trend that genes of a gene set tend to be located at the top of the entire genome-wide gene list. Then, *i-GSEA4GWAS* performs SNP-label permutations to assess the significance of the pathway-based enrichment score and to give the false discovery rate (FDR) for multiple testing corrections. *i-GSEA4GWAS* focuses on the pathways or gene-sets with the highest proportions of significant genes instead of relying solely on the total significance generated from either a few or many significant genes, thus improving the sensitivity to identify pathways/gene sets that represent the combined effects of all possibly modest SNPs/genes⁸.

Significant genes in a pathway are defined as the genes mapped with at least one of the top 5% *P*-values of all SNPs ($0.05 \times 474,792 = 23,740$ SNPs). Each significant gene was represented by the SNP in that gene with the lowest genome-wide SNP-pollutant interaction *P*-value (top SNP per significant gene). A significant gene - through its related top SNP- can be involved in more than one pathway. We selected the top SNPs of all given pathways and with classic LMM we estimated the interaction effect size for the gene-sets most significant SNP-pollutant interactions in Caucasians.

Discussion

Pathway analysis of genome-wide gene by pollution interaction study on FEV₁ in Caucasian CAMP subjects

The lung is rich in nitric oxide synthases (iNOS), and NO is normally present in the exhaled air (FeNO). Nitrogen dioxide is a pollutant tightly linked with increased exhaled nitric oxide in children with asthma and is considered as a pro-inflammatory mediator associated with airway inflammation and reduced lung function^{10, 10, 11}. NO₂ has long been found to injure the airway epithelial lining, leading to decrements in barrier and active ion transport properties¹²⁻¹⁴, and findings indicate a role for the iNOS/NO pathway in nitrogen dioxide-mediated cytotoxicity characterized by interruption of tight junctions/adhesion molecules^{13, 15, 16} and increased cytokine and ICAM-1 production with influx of inflammatory cells^{15, 17} and late phase apoptosis¹⁶. Oxidative stress, through production of reactive oxygen species, triggers disturbance in intracellular calcium homeostasis, which has been identified as an important factor in the tight junction defects and the pathogenesis of inflammatory diseases¹⁸⁻²⁰. Interleukin-10 is an anti-inflammatory cytokine induced by NO₂ exposure¹⁷ found to alleviate the mucosal injury²¹ probably via p38/MAPK activation of the heme oxygenase-1 (HO-1) pathway^{16, 22, 23}. HO-1 catalyzes the oxidation of heme to carbon monoxide (CO) and biliverdin, which play a concerted action in cytoprotection against oxidative stress and in the modulation of cell proliferation and differentiation dependent on the modulation of the p38/MAPK-signaling pathway²⁴. Ryter, Alam and Choi extensively review the HO-1/CO (heme oxygenase/carbon monoxide) system and reported that among other, glucose and lipids metabolism, agents promoting ROS generation, Angiotensin II, hypoxia, hyperthermia (heat shock proteins) and a number

170 of cytokines including IL10 induce HO-1/CO production for protection in oxidative
171 tissue injury such as lung injury models^{24, 25}. The HO-1 pathway has also been found
172 to interact with the iNOS pathway in the control of cellular homeostasis²⁶.

173

References

1. The Childhood Asthma Management Program Research Group. The childhood asthma management program (CAMP): Design, rationale, and methods. childhood asthma management program research group. Control Clin Trials. 1999 Feb;20(1):91-120.
2. American Thoracic Society. Standardization of spirometry, 1994 update. Am J Respir Crit Care Med. 1995 Sep;152(3):1107-36.
3. Strunk RC, Childhood Asthma Management Program Research Group. Childhood asthma management program: Lessons learned. J Allergy Clin Immunol. 2007 Jan;119(1):36-42.
4. The Childhood Asthma Management Program Research Group. Long-term effects of budesonide or nedocromil in children with asthma. the childhood asthma management program research group. N Engl J Med. 2000 Oct 12;343(15):1054-63.
5. Stolwijk AM, Straatman H, Zielhuis GA. Studying seasonality by using sine and cosine functions in regression analysis. J Epidemiol Community Health. 1999 Apr;53(4):235-8.
6. Sheppard L. Insights on bias and information in group-level studies. Biostatistics. 2003 Apr;4(2):265-78.
7. Neuhaus JM, Kalbfleisch JD. Between- and within-cluster covariate effects in the analysis of clustered data. Biometrics. 1998 Jun;54(2):638-45.

- 194 8. Zhang K, Cui S, Chang S, Zhang L, Wang J. i-GSEA4GWAS: A web server for
 195 identification of pathways/gene sets associated with traits by applying an improved
 196 gene set enrichment analysis to genome-wide association study. *Nucleic Acids Res.*
 197 2010 Jul;38(Web Server issue):W90-5.

- 198 9. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, *et*
 199 *al.* Gene set enrichment analysis: A knowledge-based approach for interpreting
 200 genome-wide expression profiles. *Proc Natl Acad Sci U S A.* 2005 Oct
 201 25;102(43):15545-50.

- 202 10. Holguin F, Flores S, Ross Z, Cortez M, Molina M, Molina L, *et al.* Traffic-related
 203 exposures, airway function, inflammation, and respiratory symptoms in children. *Am*
 204 *J Respir Crit Care Med.* 2007 Dec 15;176(12):1236-42.

- 205 11. Flamant-Hulin M, Caillaud D, Sacco P, Penard-Morand C, Annesi-Maesano I. Air
 206 pollution and increased levels of fractional exhaled nitric oxide in children with no
 207 history of airway damage. *J Toxicol Environ Health A.* 2010;73(4):272-83.

- 208 12. Persinger RL, Poynter ME, Ckless K, Janssen-Heininger YM. Molecular
 209 mechanisms of nitrogen dioxide induced epithelial injury in the lung. *Mol Cell*
 210 *Biochem.* 2002 May-Jun;234-235(1-2):71-80.

- 211 13. Robison TW, Kim KJ. Dual effect of nitrogen dioxide on barrier properties of
 212 guinea pig tracheobronchial epithelial monolayers cultured in an air interface. *J*
 213 *Toxicol Environ Health.* 1995 Jan;44(1):57-71.

- 214 14. Gordon RE, Solano D, Kleinerman J. Tight junction alterations of respiratory
215 epithelium following long-term NO₂ exposure and recovery. *Exp Lung Res.*
216 1986;11(3):179-93.
- 217 15. Ayyagari VN, Januszkiewicz A, Nath J. Pro-inflammatory responses of human
218 bronchial epithelial cells to acute nitrogen dioxide exposure. *Toxicology.* 2004 Apr
219 15;197(2):149-64.
- 220 16. Ayyagari VN, Januszkiewicz A, Nath J. Effects of nitrogen dioxide on the
221 expression of intercellular adhesion molecule-1, neutrophil adhesion, and
222 cytotoxicity: Studies in human bronchial epithelial cells. *Inhal Toxicol.* 2007
223 Feb;19(2):181-94.
- 224 17. Pathmanathan S, Krishna MT, Blomberg A, Helleday R, Kelly FJ, Sandstrom T,
225 *et al.* Repeated daily exposure to 2 ppm nitrogen dioxide upregulates the expression
226 of IL-5, IL-10, IL-13, and ICAM-1 in the bronchial epithelium of healthy human
227 airways. *Occup Environ Med.* 2003 Nov;60(11):892-6.
- 228 18. Xu R, Li Q, Zhou XD, Perelman JM, Kolosov VP. Oxidative stress mediates the
229 disruption of airway epithelial tight junctions through a TRPM2-PLCgamma1-
230 PKCalpha signaling pathway. *Int J Mol Sci.* 2013 Apr 29;14(5):9475-86.
- 231 19. Simet SM, Wyatt TA, DeVasure J, Yanov D, Allen-Gipson D, Sisson JH. Alcohol
232 increases the permeability of airway epithelial tight junctions in beas-2B and NHBE
233 cells. *Alcohol Clin Exp Res.* 2012 Mar;36(3):432-42.
- 234 20. Jiang LH, Yang W, Zou J, Beech DJ. TRPM2 channel properties, functions and
235 therapeutic potentials. *Expert Opin Ther Targets.* 2010 Sep;14(9):973-88.

- 236 21. Emami CN, Chokshi N, Wang J, Hunter C, Guner Y, Goth K, *et al.* Role of
237 interleukin-10 in the pathogenesis of necrotizing enterocolitis. *Am J Surg.* 2012
238 Apr;203(4):428-35.
- 239 22. Drechsler Y, Dolganiuc A, Norkina O, Romics L, Li W, Kodys K, *et al.* Heme
240 oxygenase-1 mediates the anti-inflammatory effects of acute alcohol on IL-10
241 induction involving p38 MAPK activation in monocytes. *J Immunol.* 2006 Aug
242 15;177(4):2592-600.
- 243 23. Lee TS, Chau LY. Heme oxygenase-1 mediates the anti-inflammatory effect of
244 interleukin-10 in mice. *Nat Med.* 2002 Mar;8(3):240-6.
- 245 24. Ryter SW, Alam J, Choi AM. Heme oxygenase-1/carbon monoxide: From basic
246 science to therapeutic applications. *Physiol Rev.* 2006 Apr;86(2):583-650.
- 247 25. Ryter SW, Otterbein LE, Morse D, Choi AM. Heme oxygenase/carbon monoxide
248 signaling pathways: Regulation and functional significance. *Mol Cell Biochem.* 2002
249 May-Jun;234-235(1-2):249-63.
- 250 26. Foresti R, Motterlini R. The heme oxygenase pathway and its interaction with
251 nitric oxide in the control of cellular homeostasis. *Free Radic Res.* 1999
252 Dec;31(6):459-75.
- 253

1 ***Table S1. Distribution of the 4-month average pollution***
 2 ***concentrations (December 1993 through June 1999).***

		Carbon Monoxide (ppm x10)	Nitrogen Dioxide (ppb)
Number of observations	Valid	12289	12217
	Missing	1233 (10%)	1825 (15.9%)
Percentiles	10	0	13
	25	6	18
	50	8	21
	75	11	26
	90	14	30
IQR	25-75 percentile	5	8

3

4 IQR: interquartile range, ppb: part per billion; ppm: parts per million

5

6 ***Table S2. Interquartile ranges (25th to 75th percentile) of pollution***
7 ***concentrations (December 1993 through June 1999).***

IQR 4-month average	Overall	Within-subject
Nitrogen Dioxide (ppb)	8	4
Carbon Monoxide (10 x ppm)	5	3

8

9 IQR: interquartile range, ppb: part per billion; ppm: parts per million

10

- 11 ***Table S3: Top signals of genome-wide SNP by nitrogen dioxide***
- 12 ***interactions and post-BD FEV₁ %predicted in A) Caucasian and B)***
- 13 ***African American CAMP subjects***

A) Caucasian CAMP subjects			
Nitrogen dioxide by SNP_minor allele	2-Step P values	LMM P value	LMM Change per IQR^s
rs13090972_T	1.37E-06	1.33E-08*	-1.33
rs958144_T	4.81E-06	2.94E-08*	-1.28
rs13157529_T	5.54E-06	8.46E-06	-1.08
rs4571247_G	6.31E-06	1.08E-07	-1.25
rs11656155_T	6.35E-06	1.66E-07	0.91
rs7041938_G	7.35E-06	1.04E-07*	1.14
rs1875137_G	1.85E-05	NT	NT
rs4704701_G	4.43E-05	NT	NT
rs4378142_A	2.35E-05	NT	NT
rs12684188_T	3.01E-05	NT	NT
rs1205047_A	8.08E-05	NT	NT
rs11159689_T	2.67E-05	NT	NT
rs6733550_G	4.14E-05	NT	NT
rs10782467_C	3.73E-05	NT	NT
rs1510646_T	6.57E-05	NT	NT
rs2621949_A	3.61E-05	NT	NT
rs711173_C	1.97E-05	NT	NT
rs4875101_A	7.17E-05	NT	NT

Pathways in FEV₁ response to air pollution...

rs12566794_G	5.98E-05	NT	NT
rs4896716_T	5.62E-05	NT	NT
rs8039508_T	8.58E-05	NT	NT
rs3769767_A	5.19E-05	NT	NT
rs17780310_T	9.73E-05	NT	NT
rs1844723_G	4.71E-05	NT	NT
rs2142301_T	6.64E-05	NT	NT
rs2490591_G	9.10E-05	NT	NT
rs7356801_G	9.90E-05	NT	NT
rs7642606_T	4.81E-05	NT	NT
rs939590_A	8.45E-05	NT	NT
rs10984545_A	5.17E-05	NT	NT
rs12486049_G	6.45E-05	NT	NT
rs12119540_C	6.17E-05	NT	NT
rs10512256_G	9.91E-05	NT	NT
rs1059003_G	9.91E-05	NT	NT
rs10984516_T	9.91E-05	NT	NT
rs4742708_A	9.91E-05	NT	NT
rs10058249_A	1.26E-05	NT	NT
rs3793490_G	9.44E-05	NT	NT
rs9323351_G	6.01E-05	NT	NT
rs2157401_C	9.53E-05	NT	NT
rs4672884_A	6.20E-05	NT	NT
rs897877_A	7.85E-05	NT	NT
rs6009565_A	9.42E-05	NT	NT

rs831208_T	7.34E-05	NT	NT
rs9528794_C	6.24E-05	NT	NT
rs12943384_G	8.62E-05	NT	NT
rs9988260_T	8.38E-05	NT	NT
rs10485399_T	9.69E-05	NT	NT

14

B) African American CAMP subjects			
Nitrogen dioxide by SNP_minor allele	2-Step P values	LMM P value	LMM Change per IQR
rs13090972_T	0.5822	0.5687	0.39
rs958144_T	0.9247	0.8681	0.2
rs13157529_T	0.4205	0.303	0.64
rs4571247_G	0.9596	0.7932	0.31
rs11656155_T	0.6218	0.3954	0.47
rs7041938_G	0.2615	0.1627	1.2

15

16 NT: not tested, SNP: single nucleotide polymorphisms, LMM: linear mixed model:

17 $FEV_1 = b_0 + b_1 \text{ between-subject pollution} + b_2 \text{ within-subject pollution} + b_3$

18 $SNP(\text{additive}) + b_4 \text{ SNP} \times \text{pollutant} + b_5 \text{ covariates} + e \mid \text{subject-specific (random)}$

19 intercept

20 IQR: interquartile range of 4-month average nitrogen dioxide concentration (0.004

21 parts per million)

22 *genome-wide significant (Bonferroni corrected minimally significant *P*-value being

23 $0.05/474,792 = 1.05E-07$)

24 \$direction of association same as in the 2-step approach

25 **Table S4: Top signals of genome-wide SNP by carbon monoxide**
 26 **interactions and post-BD FEV₁ %predicted in Caucasian CAMP**
 27 **subjects**

Caucasian CAMP subjects			
Carbon monoxide by SNP_minor allele	2-Step P value	LMM P value	LMM Change per IQR ^{\$}
rs883398_A	2.04E-06	2.11E-06	-0.98
rs4915866_G	5.63E-06	5.16E-06	-0.94
rs2403547_A	5.87E-06	1.17E-05	0.79
rs1986254_G	7.65E-06	1.26E-05	0.83
rs1861415_A	8.88E-06	4.41E-06	0.82
rs6569651_C	9.23E-06	9.69E-07	0.88
rs6545694_G	1.01E-05	4.17E-06	0.83
rs5742743_A	1.19E-05	NT	NT
rs13121814_G	1.28E-05	NT	NT
rs1507559_T	1.37E-05	NT	NT
rs1770011_G	1.42E-05	NT	NT
rs12130070_A	1.64E-05	NT	NT
rs2878079_G	1.64E-05	NT	NT
rs9317963_A	2.51E-05	NT	NT
rs10204566_G	2.56E-05	NT	NT
rs2237113_A	2.66E-05	NT	NT
rs6679072_G	2.84E-05	NT	NT

Pathways in FEV₁ response to air pollution...

rs2518023_T	3.86E-05	NT	NT
rs9388766_T	4.07E-05	NT	NT
rs799974_C	4.18E-05	NT	NT
rs2720625_A	4.21E-05	NT	NT
rs7519097_A	4.92E-05	NT	NT
rs1476162_G	5.24E-05	NT	NT
rs6899976_G	5.41E-05	NT	NT
rs576874_C	5.47E-05	NT	NT
rs6446330_G	5.50E-05	NT	NT
rs4388690_C	5.95E-05	NT	NT
rs3848726_T	6.09E-05	NT	NT
rs11904039_G	6.17E-05	NT	NT
rs7600221_A	6.29E-05	NT	NT
rs2039056_T	6.49E-05	NT	NT
rs10241629_G	6.59E-05	NT	NT
rs11078734_T	6.69E-05	NT	NT
rs1605461_G	6.97E-05	NT	NT
rs1470148_A	6.98E-05	NT	NT
rs9811139_C	7.17E-05	NT	NT
rs10491310_C	7.83E-05	NT	NT
rs6759519_A	8.06E-05	NT	NT
rs11621122_T	8.50E-05	NT	NT
rs11126981_A	8.74E-05	NT	NT
rs2194654_A	8.85E-05	NT	NT
rs6660832_A	8.93E-05	NT	NT

Pathways in FEV₁ response to air pollution...

rs13413446_C	9.02E-05	NT	NT
rs9388769_G	9.41E-05	NT	NT
rs10174379_T	9.71E-05	NT	NT
rs2041813_A	9.78E-05	NT	NT
rs4677169_A	9.96E-05	NT	NT

28

29 NT: not tested, SNP: single nucleotide polymorphisms, LMM: linear mixed model:

30 $FEV_1 = b_0 + b_1 \text{ between-subject pollution} + b_2 \text{ within-subject pollution} + b_3 \text{ SNP}$

31 (additive) + $b_4 \text{ SNP} \times \text{pollutant} + b_5 \text{ covariates} + e \mid \text{subject-specific (random)}$

32 intercept

33 IQR: interquartile range of 4-month average carbon monoxide concentration (0.3 parts

34 per million)

35 *genome-wide significant (Bonferroni corrected minimally significant *P*-value being

36 $0.05/474,792 = 1.05E-07$)

37 ^{\$}direction of association same as in the 2-step approach

38

39

40

41

42

43

44

45

46

47

49 **Table S5: Improved-Gene Set Enrichment Analysis for GWAS: suggested pathways for the gene by pollution**
 50 **interactions on FEV₁ (Caucasian CAMP subjects)**

CARBON MONOXIDE ¹				
	Gene Set Name	Gene Set Description	Gene Set P Value	N*
1	ST_PAC1_RECEPTOR	The signaling peptide PACAP binds to its receptor, PAC1R, which activates adenylyl cyclase and phospholipase C.	0.001	4/6/8
NITROGEN DIOXIDE ²				
	Gene Set Name	Gene Set Description	Gene Set P Value	N*
1	NOS1	Glutamate stimulates NMDA-	0.001	12/20/22

Pathways in FEV₁ response to air pollution...

		mediates calcium influx, which promotes nitric oxide synthesis from arginine by neuronal nitric oxide synthase, activating guanylate cyclase.		
2	HSP27	Hsp27 oligomers have molecular chaperone activity and protect heat-stressed cells against apoptosis.	0.002	6/14/16
3	HSA04614_RENIN_ANGIOTENSIN_SYSTEM	Genes involved in renin-angiotensin system	0.001	8/17/17
4	HSA04514_CELL_ADHESION_MOLECULES	Genes involved in cell adhesion molecules (CAMs)	0.002	58/124/134
5	GS	Activated G-protein coupled receptors stimulate cAMP	< 0.001	4/6/6

Pathways in FEV₁ response to air pollution...

		production and thus activate protein kinase A, involved in a number of signal transduction pathways.		
6	RAB	Rab family GTPases regulate vesicle transport, endocytosis and exocytosis, and vesicle docking via interactions with the rabphilins.	0.005	4/9/12
7	FEEDER	Sugars such as mannose, galactose, and fructose are enzymatically converted to glucose via feeder pathways that lead to glycolysis.	0.003	4/8/9
8	AGPCR	G-protein coupled receptors (GPCRs) transduce extracellular signals across the plasma membrane;	0.001	7/10/13

Pathways in FEV₁ response to air pollution...

		attenuation occurs by signal molecule degradation or receptor-mediated endocytosis.		
9	HSA05214_GLIOMA	Genes involved in glioma	0.002	26/59/64
10	HSA05218_MELANOMA	Genes involved in melanoma	0.001	29/66/71
11	HEME_BIOSYNTHESIS	NA	0.004	4/8/9
12	RARRXR	RXR and RAR suppress transcription in the absence of ligand and, on binding trans- or 9-cis-retinoic acid, are ubiquitinated to allow transcription to proceed.	0.006	6/14/15

Pathways in FEV₁ response to air pollution...

13	IL10	The cytokine IL-10 inhibits the inflammatory response by macrophages via activation of <i>heme oxygenase 1</i> .	0.01	6/13/13
14	CALCIUM_REGULATION_IN _CARDIAC_CELLS	NA	0.003	60/125/149
15	HSA00532_CHONDROITIN _SULFATE_BIOSYNTHESIS	Genes involved in chondroitin sulfate biosynthesis	0.005	9/15/22
16	PLCE	Gs-coupled receptors activate adenylyl cyclase, which activates Epac1, leading to the stimulation of PLC and subsequent DAG and IP3 production.	0.005	6/10/12
17	GLYCOSPHINGOLIPID_METABOLISM	NA	0.004	9/24/24

18	VOBESITY	The adipose tissue of obese individuals overexpresses a key glucocorticoid-metabolizing enzyme, activating inactive circulating corticosteroids and inducing insulin resistance.	0.005	4/7/8
19	CREB	CREB is a transcription factor that binds to cAMP-responsive elements (CREs) to activate transcription in response to extracellular signaling.	0.004	13/14/27
20	DNA_REPLICATION_REACTOME	NA	0.009	12/39/49
21	MPR	Progesterone binding to its	0.005	10/21/23

Pathways in FEV₁ response to air pollution...

		intracellular receptor activates the MAPK pathway and induces oocyte maturation; binding to membrane receptor inhibits adenylyl cyclase.		
22	TRANSLATION_FACTORS	NA	0.003	10/31/53
23	TEL	Telomerase is a ribonucleotide protein that adds telomeric repeats to the 3' ends of chromosomes.	0.013	8/15/18

51

52 ¹⁻²Details on the genes and mapped top SNPs involved in each pathway can be found in links provided by the *i*-GSEA4GWAS server under the
53 label “view details”:

54 ¹Carbon Monoxide-related gene set/pathway:

55 http://gsea4gwas.psych.ac.cn/getResult.do?result=13F3A972887892430E6A5C369D76FEAD_1372283303807

Pathways in FEV₁ response to air pollution...

56 ²Nitrogen Dioxide-related gene set/pathway:

57 http://gsea4gwas.psych.ac.cn/getResult.do?result=13F3A972887892430E6A5C369D76FEAD_1372284527739

58 N*: The number of Significant genes/Selected genes/All genes

59 Significant genes: genes mapped with at least one of the top 5% of all SNPs; Selected genes: genes included in the *i*-GSEA analysis; All genes:

60 all genes of the gene set.

61 NA: not applicable (by the *i*-GSEA4GWAS server)

62 GWAS: genome wide association study; FEV₁: forced expiratory volume in 1 second

63

64

Table S6: Pathway analysis of top SNPs with effect modification of the long-term effects of nitrogen dioxide on post-BD FEV₁ %predicted in Caucasian subjects

Top SNP_minor allele	Significant Gene	Chr	Change per IQR	P-value of SNP by nitrogen dioxide interactions
rs1521481_C	IGF1R	15	0.78	2.32E-06
rs3755377_G	TGFA	2	-0.74	3.61E-06
rs12119540_C	CHRM3	1	0.73	1.42E-05
rs7304453_T	ITPR2	12	0.71	1.64E-05
rs13219957_A	UST	6	0.90	2.02E-05
rs13077495_C	PPARG	3	-1.27	3.62E-05
rs10151613_C	NRXN3	14	0.67	4.86E-05
rs1800896_C	IL10	1	0.66	5.29E-05
rs11820322_T	NRXN2	11	1.77	1.13E-04
rs8_T	CDK6	7	-0.73	1.16E-04
rs717099_A	PRKAR2B	7	-0.60	1.18E-04
rs4899406_T	RGS6	2	-1.04	1.59E-04
rs3087474_T	ORC1L	1	-1.12	1.61E-04
rs17409602_C	EIF2AK1	7	0.67	1.67E-04
rs4952447_C	SLC8A1	10	0.57	1.70E-04
rs1732664_T	SYT1	12	0.62	2.15E-04

Pathways in FEV₁ response to air pollution...

rs1200138_C	ATP1B1	1	0.64	2.19E-04
rs6904263_T	GMNN	6	-0.74	2.21E-04
rs12144820_A	ATP1A4	1	-0.57	2.45E-04
rs7178_G	CCND1	11	-1.01	2.46E-04
rs1868289_G	GRIN2A	16	0.63	2.86E-04
rs1025298_C	ADCY2	5	-0.61	2.90E-04
rs1642763_A	ATP1B2	17	0.66	3.10E-04
rs658132_G	NLGN1	3	0.66	3.97E-04
rs4899329_C	SLC8A3	14	-0.57	4.41E-04
rs13164785_G	VCAN	5	0.67	4.41E-04
rs1981429_G	SDC4	20	-0.58	4.54E-04
rs199933_A	CACNA1E	1	-0.65	4.57E-04
rs8102695_A	CACNA1A	19	-0.68	4.88E-04
rs1805352_C	RXRA	9	0.59	5.07E-04
rs3774463_T	CACNA1D	3	1.12	5.16E-04
rs478839_G	PTEN	10	-0.58	5.30E-04
rs6956460_G	CNTNAP2	7	1.05	5.43E-04
rs2581_T	HLA-DOA	6	0.54	5.43E-04
rs2025818_T	PRKCQ	10	-0.54	5.79E-04
rs9300729_C	FGF14	13	-0.82	5.81E-04
rs1131857_G	CPOX	3	0.76	6.42E-04
rs4678009_T	ADCY5	3	0.60	7.00E-04
rs12695902_G	AGTR1	3	0.88	7.15E-04
rs3830180_G	MPZL1	1	0.59	7.91E-04

Pathways in FEV₁ response to air pollution...

rs181997_A	HLA-DMB	6	-0.57	9.48E-04
rs13229270_A	RPA3	7	-0.53	1.01E-03
rs8036393_T	CHSY1	15	-0.51	1.03E-03
rs2850971_T	PPP3CA	4	-0.58	1.18E-03
rs165979_G	PDGFRB	5	-0.52	1.20E-03
rs2849380_T	BCL2	18	-0.63	1.22E-03
rs3790112_T	GNAO1	16	0.55	1.30E-03
rs17676564_A	NCOA2	8	-1.01	1.31E-03
rs4254322_G	PLCG2	16	0.59	1.40E-03
rs629079_C	NEGR1	1	0.57	1.46E-03
rs4724420_T	ADCY1	7	-0.52	1.47E-03
rs2236561_G	GNG2	14	0.52	1.49E-03
rs797999_T	CD2	1	-0.59	1.51E-03
rs12026958_C	RGS7	1	-0.84	1.61E-03
rs11198973_G	RGS10	10	0.52	1.63E-03
rs1485237_C	MME	3	-0.52	1.64E-03
rs9842138_G	FGF12	3	0.68	1.68E-03
rs2443502_G	PPP3CC	8	-0.48	1.71E-03
rs10259426_C	GNGT1	7	-0.65	1.72E-03
rs17041884_T	ENPEP	4	-0.57	1.80E-03
rs10515521_A	NR3C1	5	0.65	1.83E-03
rs1489024_G	PVRL3	3	-0.51	1.98E-03
rs4783446_G	EEF2K	16	-0.99	2.06E-03
rs2835370_C	CLDN14	21	0.84	2.10E-03

Pathways in FEV₁ response to air pollution...

rs733299_T	CHST14	15	-0.52	2.15E-03
rs2281390_T	HLA-DPB1	6	-0.59	2.17E-03
rs3767248_G	EIF4G3	1	0.48	2.21E-03
rs454006_C	PRKCG	19	0.60	2.25E-03
rs13422838_C	ITGAV	2	-0.83	2.27E-03
rs12086956_G	RYR2	1	-1.55	2.30E-03
rs12423712_A	NCOR2	12	0.95	2.52E-03
rs3806206_T	CAP1	1	0.74	2.77E-03
rs12589478_A	SPTLC2	14	-0.53	2.84E-03
rs11258248_A	MCM10	10	-0.90	2.84E-03
rs7306482_A	FGF6	12	-0.48	2.91E-03
rs173681_T	CAMK4	5	-0.59	2.94E-03
rs16916456_T	UGCG	9	0.92	2.97E-03
rs2575710_C	SDC2	8	0.46	3.03E-03
rs2292719_A	GTF2E1	3	-0.53	3.13E-03
rs4405588_C	RPA1	17	0.46	3.22E-03
rs2531992_A	ADCY9	16	0.58	3.30E-03
rs6743452_C	NRXN1	2	-0.46	3.32E-03
rs10066754_T	PPAP2A	5	-0.77	3.32E-03
rs3897583_C	NCAM1	11	0.48	3.32E-03
rs7623208_T	CPA3	3	-0.82	3.39E-03
rs6797758_G	ATP2B2	3	-0.59	3.41E-03
rs7255066_C	PVR	19	0.52	3.55E-03
rs12626592_A	CLDN8	21	-0.98	3.64E-03

Pathways in FEV₁ response to air pollution...

rs744143_A	E2F3	6	0.72	3.68E-03
rs3783834_A	RPS6KA5	14	-0.49	3.69E-03
rs17172446_A	EGFR	7	0.56	3.69E-03
rs3785905_A	PRKAR1A	17	-0.73	3.70E-03
rs1819741_C	RGS2	1	-0.56	3.86E-03
rs10423648_G	CD22	19	0.90	3.95E-03
rs11122574_T	AGT	1	-1.05	3.98E-03
rs10042299_G	ANKHD1	5	-0.44	4.01E-03
rs12471357_T	PRKCE	2	0.45	4.12E-03
rs11216689_A	IL10RA	11	0.98	4.14E-03
rs558190_T	DSE	6	0.52	4.14E-03
rs2104977_A	TEP1	14	0.60	4.17E-03
rs3766074_A	NME7	1	0.70	4.19E-03
rs4670185_G	EIF2AK2	2	0.88	4.41E-03
rs7623610_A	MITF	3	-0.44	4.51E-03
rs4339559_T	HGF	7	-0.68	4.61E-03
rs2859228_G	FASLG	1	-0.47	4.67E-03
rs252649_C	NLN	5	-0.91	4.79E-03
rs260849_G	PDGFD	11	-0.56	4.83E-03
rs944260_G	CDH4	20	-0.52	4.83E-03
rs11666282_T	MAG	19	0.51	4.93E-03
rs2049050_T	RAB3A	19	-0.48	4.97E-03
rs730005_C	LCT	2	-0.57	5.07E-03
rs16909222_G	GRIN2B	12	-0.79	5.13E-03

Pathways in FEV₁ response to air pollution...

rs7515374_C	MAPKAPK 2	1	0.46	5.21E-03
rs17407577_C	FGF2	4	-1.04	5.34E-03
rs869789_A	FXYD2	11	-0.64	5.43E-03
rs2075067_A	CAMK2B	7	-0.74	5.45E-03
rs237484_A	KCNB1	20	0.44	5.50E-03
rs16923198_A	PDCD1LG 2	9	-0.56	5.73E-03
rs4665716_G	NCOA1	2	0.87	5.85E-03
rs8068961_C	XYLT2	17	-0.46	5.88E-03
rs11927393_G	EIF2B5	3	-0.77	5.97E-03
rs982804_C	FGF5	4	-0.45	6.06E-03
rs2517912_T	HLA-F	6	0.45	6.06E-03
rs9904113_A	ATP2A3	17	0.46	6.25E-03
rs352165_A	ALAS1	3	0.45	6.34E-03
rs6891913_T	FGF1	5	-0.47	6.36E-03
rs10487841_C	NRCAM	7	0.92	6.53E-03
rs16875961_G	ARSB	5	-0.66	6.57E-03
rs2876981_C	MAPK1	22	0.43	6.67E-03
rs6425689_C	SDC3	1	-0.47	6.76E-03
rs6885505_G	CAMK2A	5	0.46	6.90E-03
rs3773729_C	PRKCD	3	0.43	6.98E-03
rs6709175_C	CD8B	2	0.52	7.01E-03
rs2393374_T	RFC5	12	-0.46	7.22E-03

Pathways in FEV₁ response to air pollution...

rs7886134_C	NLGN3	X	-0.35	7.52E-03
rs1033908_C	PRKCH	14	0.53	7.66E-03
rs7298903_C	NOS1	12	0.61	7.84E-03
rs6804441_G	CD80	3	0.54	7.95E-03
rs1886700_T	CDH3	16	0.59	8.12E-03
rs893952_T	B3GAT1	11	0.45	8.27E-03
rs3760055_A	GLG1	16	-0.48	8.34E-03
rs9526475_T	RB1	13	-0.47	8.35E-03
rs4072067_G	CALML3	10	-0.54	8.42E-03
rs4316697_A	FGF7	15	0.44	8.74E-03
rs2007044_G	CACNA1C	12	-0.44	8.75E-03
rs1009897_A	CD34	1	0.43	8.86E-03
rs6092704_C	GNAS	20	0.70	9.07E-03
rs999297_A	CD6	11	-0.47	9.13E-03
rs10025791_C	CAMK2D	4	0.42	9.71E-03
rs10032688_A	PDGFRA	4	0.60	9.77E-03
rs1956918_A	CMA1	14	0.41	9.81E-03
rs10413199_A	SHC2	19	0.55	9.86E-03
rs11636875_G	SHC4	15	-0.42	1.01E-02
rs17208112_A	CD226	18	-0.45	1.02E-02
rs733018_G	ITPR1	3	-0.47	1.04E-02
rs651084_G	CLDN17	21	0.50	1.04E-02
rs10796316_A	ITGA8	10	2.63	1.04E-02
rs10799902_A	RGS5	1	0.42	1.04E-02

Pathways in FEV₁ response to air pollution...

rs583720_A	PRKACB	1	0.51	1.05E-02
rs6982126_T	TERF1	8	0.44	1.05E-02
rs2475193_A	ITGB1	10	0.52	1.08E-02
rs1784304_A	HMBS	11	-0.49	1.09E-02
rs2352865_C	GJA5	1	0.50	1.10E-02
rs1619379_T	HLA-G	6	-0.43	1.10E-02
rs10491087_T	PIK3R5	17	-0.72	1.11E-02
rs12125947_C	CDC7	1	-0.41	1.12E-02
rs11247963_A	RPS6KA1	1	-0.50	1.13E-02
rs1367728_A	HLA-DPA1	6	-0.54	1.14E-02
rs527834_A	CHRM5	15	0.80	1.17E-02
rs2289059_G	RAB27A	15	0.84	1.20E-02
rs367689_A	ANXA6	5	-0.74	1.20E-02
rs739999_G	RGS11	16	-0.70	1.21E-02
rs12477594_A	ITGA6	2	1.18	1.29E-02
rs10820938_A	SPTLC1	9	0.74	1.33E-02
rs263238_A	ADCY8	8	-0.53	1.34E-02
rs4698804_C	EGF	4	-2.14	1.34E-02
rs7066678_A	RAB9A	X	-0.34	1.41E-02
rs8091758_T	PTPRM	18	-1.03	1.42E-02
rs919699_A	APAF1	12	-0.63	1.43E-02
rs12982646_A	MADCAM 1	19	-0.41	1.45E-02
rs932757_A	RFC3	13	-0.37	1.47E-02

Pathways in FEV₁ response to air pollution...

rs2271026_C	ERCC3	2	-0.89	1.48E-02
rs899480_G	NEO1	15	-0.81	1.48E-02
rs477549_C	PYGM	11	-0.61	1.48E-02
rs3767498_A	CACNA1S	1	-0.41	1.49E-02
rs11861415_G	XYLT1	16	-0.41	1.50E-02
rs7111082_A	KCNJ5	11	0.39	1.52E-02
rs578130_A	ARRB1	11	0.41	1.52E-02
rs715849_A	PVRL1	11	-0.41	1.52E-02
rs3002208_G	PABPC3	13	0.48	1.53E-02
rs1122356_G	CDH2	18	-0.43	1.53E-02
rs2922666_T	POLD2	7	-0.51	1.58E-02
rs13191198_G	PKIB	6	0.38	1.62E-02
rs2524005_A	HLA-A	6	-0.50	1.69E-02
rs4855074_T	GNB4	3	-0.52	1.75E-02
rs10496444_G	IL1A	2	0.40	1.75E-02
rs2993577_G	CLDN10	13	1.79	1.84E-02
rs7167430_C	RYR3	15	0.54	1.85E-02
rs742403_G	PDGFB	22	0.37	1.86E-02
rs9958503_A	FECH	18	-0.64	1.87E-02
rs10823343_G	HK1	10	-0.41	1.89E-02
rs745966_C	ATP1B3	3	-0.61	1.92E-02
rs206187_A	ITGB8	7	-0.60	1.94E-02
rs1956908_A	CTSG	14	0.38	1.97E-02
rs3170368_G	EEF2	19	0.55	2.01E-02

Pathways in FEV₁ response to air pollution...

rs1693892_G	CALM2	2	0.39	2.01E-02
rs715643_T	B3GALT6	1	-0.71	2.14E-02
rs2665742_G	CHST11	12	0.36	2.16E-02
rs6429206_T	GNG4	1	-0.37	2.24E-02
rs17126224_G	ASAH1	8	-0.80	2.29E-02
rs7207345_C	PRKCA	17	-0.39	2.29E-02
rs785520_C	PIK3R3	1	-0.35	2.50E-02
rs197770_G	ITGA9	3	-0.56	2.59E-02
rs3730064_A	ADCY6	12	-0.62	2.60E-02
rs3746821_T	CD40	20	-0.57	2.67E-02
rs1014982_T	DIAPH2	X	0.28	2.76E-02
rs1572057_T	PPAP2B	1	0.47	2.86E-02
rs4925479_C	ACTA1	1	0.39	2.89E-02
rs2652423_T	KCNJ3	2	-0.39	2.93E-02
rs7277902_T	IL10RB	21	0.54	2.94E-02
rs1398259_C	CADM1	11	-0.35	2.96E-02
rs7550839_T	MPZ	1	-0.46	2.96E-02
rs10489612_A	PGM1	1	-0.54	3.03E-02
rs263955_T	EIF2B3	1	-0.35	3.05E-02
rs9616802_T	ARSA	22	-0.34	3.20E-02
rs900781_T	FGF17	8	0.56	3.26E-02
rs7255050_A	GRIN2D	19	-0.33	3.31E-02
rs2307373_T	ORC3L	6	-0.34	3.34E-02
rs7776857_G	IL6	7	0.34	3.38E-02

Pathways in FEV₁ response to air pollution...

rs2273814_G	PRKD1	14	-0.34	3.51E-02
rs6790988_A	CLDN11	3	-0.35	3.66E-02
rs16833157_A	STAT1	2	0.81	3.87E-02
rs2826877_T	NCAM2	21	-0.32	4.05E-02
rs5030390_A	ICAM1	19	-0.56	4.07E-02
rs7401414_A	EIF5	14	0.47	4.09E-02
rs6949887_A	ORC5L	7	-0.72	4.64E-02
rs11179589_C	CNTN1	12	-0.55	4.69E-02
rs270_A	LPL	8	0.41	4.95E-02
rs2494738_A	AKT1	14	0.59	5.34E-02
rs12820100_C	FGF23	12	-0.32	5.43E-02
rs13323993_A	RAP2B	3	0.33	5.56E-02
rs1443237_A	PDGFC	4	-0.35	5.68E-02
rs679305_T	FGF13	X	-0.56	6.20E-02
rs1449264_C	ITGA4	2	0.30	6.66E-02

\$ Linear mixed model: $FEV_1 = b_0 + b_1 \text{ between-subject pollution} + b_2 \text{ within-subject pollution}$

+ $b_3 \text{ SNP(additive)} + b_4 \text{ SNP} \times \text{pollutant} + b_5 \text{ covariates} + e \mid \text{subject-specific (random)}$

intercept

IQR: interquartile range of 4-month average nitrogen dioxide concentration = 0.004 parts per

million; Chr: chromosome

Table S7: Pathway analysis top SNPs and effect modification of the long-term effects of carbon monoxide on post-BD FEV₁ %predicted in Caucasian CAMP subjects

Top SNPs _minor allele	Significant Gene	Chr	Change per IQR ^{\$}	P-value ^{\$}
rs11130199_C	DAG1	3	-0.55	0.002
rs17515291_G	ASAH1	8	-0.64	0.001
rs3768405_A	ITPKB	1	0.66	0.001
rs7035689_A	GNAQ	9	0.45	0.022

Caucasian Childhood Asthma Management Program subjects (n=522); ^{\$} Linear mixed model:

FEV₁=bo + b1 between-subject pollution + b2 within-subject pollution + b3 SNP (additive) + b4 SNP x pollutant + b5 covariates + e | subject-specific (random) intercept

IQR: interquartile range of 4-month average carbon monoxide concentration = 0.3 parts per million