



ARTICLE

Understanding the role of the chromosome 15q25.1 in COPD through epigenetics and transcriptomics

Ivana Nedeljkovic¹ · Elena Carnero-Montoro^{1,2} · Lies Lahousse^{1,3,4} · Diana A. van der Plaats^{5,6} · Kim de Jong^{5,6} · Judith M. Vonk^{5,6} · Cleo C. van Diemen^{5,7} · Alen Faiz⁶ · Maarten van den Berge⁸ · Ma'een Obeidat⁹ · Yohan Bossé¹⁰ · David C. Nickle¹¹ · BIOS Consortium · Andre G. Uitterlinden^{1,12} · Joyce J. B. van Meurs¹² · Bruno C. H. Stricker¹ · Guy G. Brusselle^{1,4,13} · Dirkje S. Postma^{6,8} · H. Marika Boezen^{5,6} · Cornelia M. van Duijn¹ · Najaf Amin¹

Received: 14 March 2017 / Revised: 6 November 2017 / Accepted: 19 December 2017 / Published online: 8 February 2018
© European Society of Human Genetics 2018

Abstract

Chronic obstructive pulmonary disease (COPD) is a major health burden in adults and cigarette smoking is considered the most important environmental risk factor of COPD. Chromosome 15q25.1 locus is associated with both COPD and smoking. Our study aims at understanding the mechanism underlying the association of chromosome 15q25.1 with COPD through epigenetic and transcriptional variation in a population-based setting. To assess if COPD-associated variants in 15q25.1 are methylation quantitative trait loci, epigenome-wide association analysis of four genetic variants, previously associated with COPD ($P < 5 \times 10^{-8}$) in the 15q25.1 locus (rs12914385:C>T-*CHRNA3*, rs8034191:T>C-*HYKK*, rs13180:C>T-*IREB2* and rs8042238:C>T-*IREB2*), was performed in the Rotterdam study ($n = 1489$). All four variants were significantly associated ($P < 1.4 \times 10^{-6}$) with blood DNA methylation of *IREB2*, *CHRNA3* and *PSMA4*, of which two, including *IREB2* and *PSMA4*, were also differentially methylated in COPD cases and controls ($P < 0.04$). Further additive and multiplicative effects of smoking were evaluated and no significant effect was observed. To evaluate if these four genetic variants are expression quantitative trait loci, transcriptome-wide association analysis was performed in 1087 lung samples. All four variants were also significantly associated with differential expression of the *IREB2* 3'UTR in lung tissues ($P < 5.4 \times 10^{-95}$). We conclude that regulatory mechanisms affecting the expression of *IREB2* gene, such as DNA methylation, may explain the association between genetic variants in chromosome 15q25.1 and COPD, largely independent of smoking.

Electronic supplementary material The online version of this article (<https://doi.org/10.1038/s41431-017-0089-8>) contains supplementary material, which is available to authorized users.

✉ Najaf Amin
n.amin@erasmusmc.nl

¹ Department of Epidemiology, Erasmus Medical Centre, Rotterdam, The Netherlands

² Pfizer University of Granada, GENYO Centre for Genomics and Oncological Research, Andalusian Region Government, Granada, Spain

³ Department of Respiratory Medicine, Ghent University Hospital, Ghent, Belgium

⁴ Department of Bioanalysis Pharmaceutical Care Unit, Ghent University Hospital, Ghent, Belgium

⁵ Department of Epidemiology University of Groningen, University Medical Centre Groningen, Groningen, The Netherlands

⁶ Groningen Research Institute for Asthma and COPD (GRIAC), University of Groningen, Groningen, The Netherlands

⁷ Department of Genetics, University of Groningen, University Medical Centre Groningen, Groningen, The Netherlands

⁸ Department of Pulmonology University of Groningen, University Medical Centre Groningen, Groningen, The Netherlands

⁹ Centre for Heart Lung Innovation, University of British Columbia, Vancouver, BC, Canada

¹⁰ Department of Molecular Medicine, Laval University, Institut universitaire de cardiologie et de pneumologie de Québec, Québec, QC, Canada

¹¹ Genetics and Pharmacogenomics (GpGx), Merck Research Laboratories, Seattle, WA, USA

¹² Department of Internal Medicine, Erasmus Medical Centre, Rotterdam, The Netherlands

¹³ Department of Respiratory Medicine, Erasmus Medical Centre, Rotterdam, The Netherlands

Introduction

Chronic obstructive pulmonary disease (COPD) affects over 300 million people and is the third leading cause of death worldwide, which makes it a major public health burden [1]. COPD is characterised by airflow limitation and chronic, inflammatory response of the airways to cigarette smoke, occupational exposures, air pollution etc. [2]. Systemic inflammation and complications, together with comorbid conditions, add to its complexity [3].

COPD is determined by both genetic and environmental factors. Genetic factors explain 20–40% of the variance in the disease [4], while the most important environmental risk factor is smoking. Smokers have 10-fold increased incidence of COPD compared to never-smokers [5]. Genome-wide association studies (GWASs) revealed genetic variants associated with COPD and lung function [6, 7]. From a genetic–epidemiological perspective, the chromosome 15q25.1 locus is of interest, harbouring three nicotinic receptors (*CHRNA3*, *CHRNA5* and *CHRNA4*) and other genes that could have potential impact on COPD, including *IREB2*, *PSMA4* and *HYKK* [8–11]. The 15q25.1 region has also been associated with smoking [12, 13] and lung cancer [14] in large GWASs. Because smoking is a risk factor for both COPD and lung cancer, the association of this locus with COPD and lung cancer might be mediated through smoking [15], which is in line with the only longitudinal study investigating this hypothesis thus far [16].

The mechanism through which these single-nucleotide polymorphisms (SNPs) and smoking are involved in COPD and related outcomes remains obscure. SNPs in the 15q25.1 locus are shown to be *cis*- expression quantitative trait loci (*cis*-eQTLs) in blood [17], brain [18], sputum [19] and lungs [20, 21]. This raises the question of whether the variants are involved in COPD through the regulatory mechanisms. DNA methylation is a heritable, dynamic, epigenetic mark that plays a critical role in the regulation of gene expression [22]. Despite having a strong genetic component, DNA methylation is known to respond to changes in environmental factors [23], and its role in mediating genetic risk effect and the interaction with environmental exposure has been widely proposed [24].

Recent studies have shown the association between genome-wide patterns of DNA methylation variation with smoking behaviour [25, 26] and COPD [22, 27]. Though differential DNA methylation sites (CpG) in *CHRNA3* (15q25.1) were associated with COPD status and lung cancer [28], the role of DNA methylation as a mechanism through which the genetic variants may be involved in COPD and related outcomes remains unexplored.

In this study we selected all SNPs in the 15q25.1 region, associated with COPD in GWAS: rs12914385:C>T (*CHRNA3*), rs8034191:T>C (*HYKK*, *CHRNA3*, *CHRNA5*),

rs13180:C>T (*IREB2*) and rs8042238:C>T (*IREB2*) [8–11]. The rs12914385:C>T is a top hit in the largest GWAS of COPD (odds ratio (OR) = 1.39, $P = 2.7 \times 10^{-16}$) [8]. SNPs rs12914385:C>T and rs8034191:T>C are in moderate linkage disequilibrium (LD; $r^2 = 0.723$), while rs13180:C>T and rs8042238:C>T are in perfect LD ($r^2 = 0.997$). Rs8034191:T>C is estimated to explain 12.2% of COPD risk in the general population and 14.3% in current smokers [11]. Rs13180:C>T is associated with COPD, independently of smoking and of rs8034191:T>C [10, 15, 21]. Evidence suggests that the association of COPD with rs8034191:T>C may be mediated by smoking, while the association with rs13180:C>T is independent of smoking [21, 29]. We examined whether SNPs in the 15q25.1 locus are also associated with differential DNA methylation in the population-based Rotterdam study (RS). Further, we tested if methylation patterns associated with these variants are associated with COPD and FEV₁/FVC (the ratio of the forced expiratory volume in first second (FEV₁) over forced vital capacity of the lungs (FVC)). We also tested whether the variants are associated with a differential expression in non-tumour lung tissue from the Lung eQTL study (LES).

Methods

Study population

Participants of the discovery and replication cohorts were part of the RS, a prospective, population-based study, designed to investigate the occurrence and determinants of diseases in the elderly, as described elsewhere [30]. The discovery cohort of our epigenetic analysis is a random sample of 723 participants from RS with complete phenotype, genome-wide genotype and methylation data available. An independent sample of 766 participants from RS were included as a replication cohort. RS is part of the Biobanking and Biomolecular Resources Research Infrastructure for The Netherlands (BBMRI-NL), BIOS (Biobank-based Integrative Omics Studies) project [31]. The epigenome-wide association study (EWAS) data of RS were made publicly available as a Rainbow Project (RP3; BIOS) of the BBMRI-NL (data access link: http://wiki.bbmri.nl/wiki/BIOS_bios). Results of this study are available through dbGaP (accession number phs000930, <https://www.ncbi.nlm.nih.gov/gap>).

Detailed information on spirometry measures, COPD diagnosis, COPD SNP selection, genotyping and DNA methylation assessment in RS and RNA array in LES is provided in the Supplementary information. LES included patients from three participating sites: University of Groningen (GRN), Laval University (Laval) and University of British Columbia (UBC).

Statistical analyses

First, we tested the association of the four selected SNPs with COPD and FEV₁/FVC in our discovery and replication cohorts using logistic and linear regression models, respectively, adjusted for age and sex in model 1 and additionally adjusting for current smoking and pack-years in model 2. Results from the two cohorts were then meta-analysed using fixed effects models with “rmeta” package in R. Further, in the RS discovery cohort ($n = 723$), we performed four EWASs to assess the relationship between dosages of each SNP, as independent variable, and epigenome-wide DNA methylation in blood as dependent variable. We applied linear regression methods using two models. One adjusted for age, sex, technical covariates to correct for batch effects (array number and position on array), and white blood cell types to correct for the cellular heterogeneity of blood (number of lymphocytes, monocytes and granulocytes) (model 1). The other was adjusted additionally for current smoking and pack-years, the number of cigarette packs smoked in 1 year (model 2). The false discovery rate (FDR) of <0.05 was used to declare epigenome-wide significance. Significant sites were tested in the replication cohort ($n = 766$) using the same models. Since the 15q25.1 region is also associated with smoking behaviour, significant CpG sites were also tested in a third model including “SNP \times current smoking” and “SNP \times pack-years” interaction terms to assess possible genetic–environment interaction between the tested SNPs and smoking per cohort. Per cohort results were meta-

analysed using fixed effects models with “rmeta” package in R. Associations of the identified CpG sites with COPD and FEV₁/FVC were further performed using logistic and linear regression, respectively, adjusted for age, sex, technical covariates and white blood cell counts in both the discovery and replication cohorts, and meta-analysed as mentioned above.

Finally, we assessed whether the identified SNPs were acting as eQTLs in lung tissue. We performed genome-wide eQTL analysis in 1087 samples from GRN, UBC and Laval, compared to Nguyen et al. [20] who used Laval ($N = 420$) as the discovery and UBC and GRN samples for replication. First, cohort-specific (GRN, Laval and UBC) principal components (PCs) explaining at least 1% were calculated based on residuals from linear regression models on genome-wide 2-log transformed gene expression levels (of each probe separately) adjusted for COPD status, age, sex and smoking status. Second, in each cohort separately, linear regression analysis was used to test for association between the SNPs and genome-wide 2-log transformed gene expression levels. SNPs were tested in an additive genetic model and the models were adjusted for disease status (COPD, alpha-1 antitrypsin deficiency, idiopathic pulmonary fibrosis, pulmonary hypertension, cystic fibrosis and other disease), age, sex, smoking status and the cohort-specific number of PCs (14 PCs for GRN and Laval, and 16 for UBC). Finally, SNP effect estimates of the three cohorts were meta-analysed using fixed effects models with effect estimates weighted by the reciprocal of the estimated variance. We used FDR < 0.05 to correct for multiple testing.

Table 1 Characteristics of the discovery and replication cohorts and per COPD status

	Discovery cohort			Replication cohort			P-value
	COPD	Controls	All	COPD	Controls	All	
N (% of all)	114 (15.8)	541 (74.8)	723	93 (12.1)	591 (77.2)	766	0.054*
Age (years)	61.9 \pm 8.6	59.3 \pm 7.9	59.9 \pm 8.2	68.2 \pm 5.7	67.6 \pm 5.9	67.7 \pm 5.9	$< 5 \times 10^{-6}$
Gender (% males)	59.6	43.1	45.8	58.1	42.1	42.3	0.176
FEV ₁ /FVC	0.63 \pm 0.07	0.78 \pm 0.04	0.75 \pm 0.08	0.63 \pm 0.07	0.78 \pm 0.05	0.76 \pm 0.08	0.08
Smoking status							$< 5 \times 10^{-6}$
Current smokers (%)	37.7	19.8	23.2	21.5	8.8	10.4	
Ex-smokers (%)	48.2	44.2	44.4	54.8	55.2	55.7	
Never smokers (%)	14	36	32.4	23.7	36	33.8	
Pack-years ^a	34.3 \pm 26.9	19.9 \pm 19.6	23.2 \pm 22.0	33.7 \pm 18.7	19.6 \pm 20.1	21.9 \pm 20.6	0.261

Data presented as percent or mean \pm SD; for COPD status, which was not available for all participants, the valid percentage is denoted in brackets (percent of all); in the discovery cohort 68 patients and in the replication cohort 82 patients were excluded from the association analyses with COPD due to possible asthma

All all participants included in EWAS, COPD chronic obstructive pulmonary disease cases, P-value P-value of the difference of discovery and replication cohorts

*P-value of the difference of COPD status in discovery and replication cohorts

^aPack-years calculated in current and ex-smokers only

Results

The discovery set comprised 723 participants of RS, with genotype and DNA methylation data from whole blood, including 114 COPD cases and 541 controls (68 excluded due to possible asthma). The replication set comprised 766 independent participants of Rotterdam study, with genotype and DNA methylation data, including 93 COPD cases and 591 controls (82 excluded due to possible asthma). The characteristics of the discovery and replication cohorts are shown in Table 1. COPD cases were more often male and

smokers and had smoked on average more pack-years compared to controls.

Three of the four selected SNPs (rs12914385:C>T, rs13180:C>T and rs8042238:C>T) on chromosome 15q25.1 were nominally associated with COPD in RS ($n = 1339$), while only rs12914385:C>T was nominally associated with COPD in a considerably smaller dataset from LES ($n = 512$) (Table S1). None of the SNPs were associated with FEV₁/FVC (Table S1). To determine whether the four SNPs (rs12914385:C>T, rs8034191:T>C, rs13180:C>T and rs8042238:C>T) in chromosome

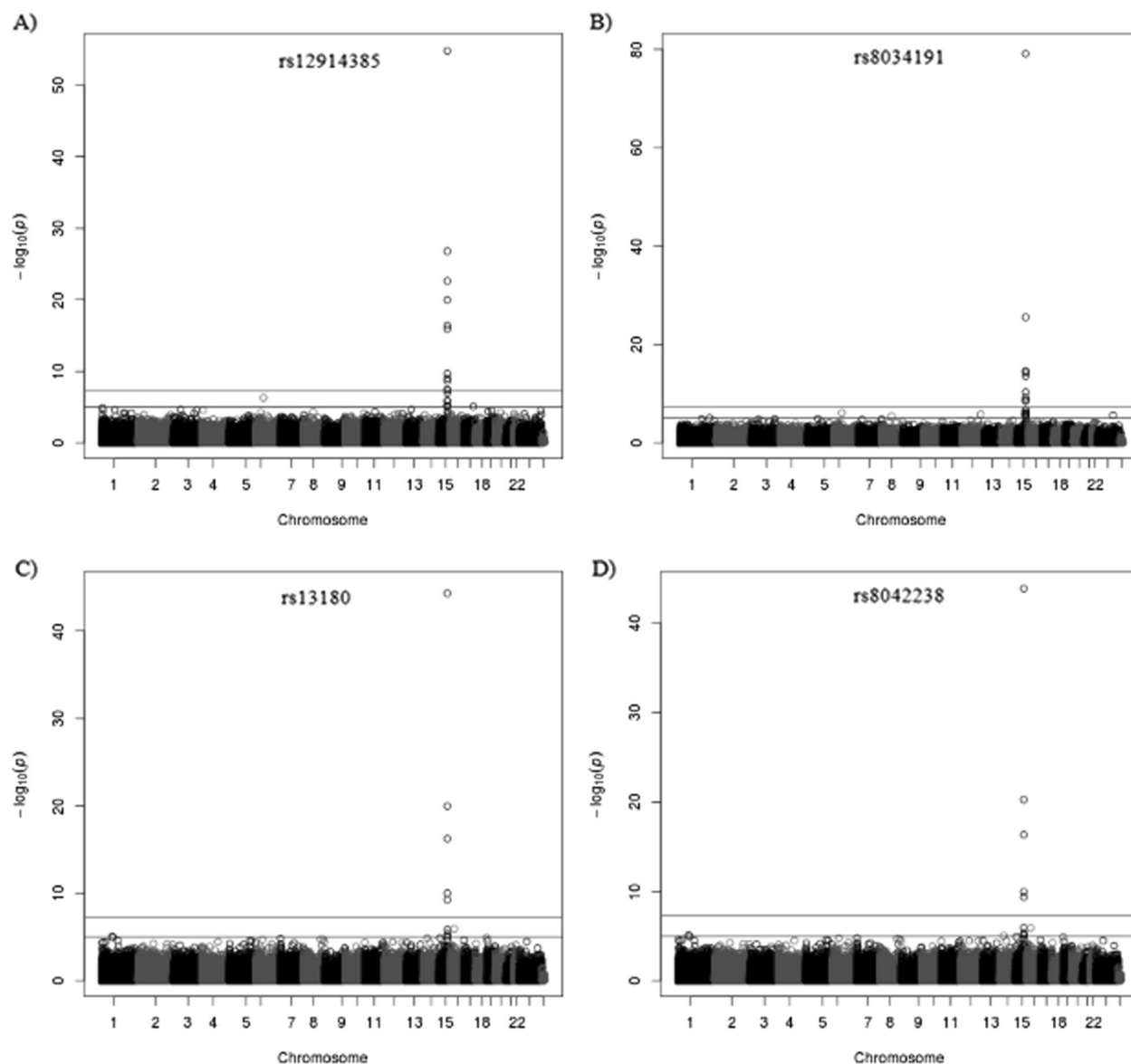


Fig. 1 Plotted results of the discovery EWAS of the four SNPs at chromosome 15. In circles are represented all CpGs throughout the genome. The x-axis shows all chromosomes. The y-axis shows negative logarithm of the P -value of the associations of **a** rs12914385:C>T,

b rs8034191:T>C, **c** rs13180:C>T and **d** rs8042238:C>T with epigenome-wide DNA methylation levels. Upper line represents the significance threshold (FDR < 0.05). $N = 723$

15q25.1 are methylation QTLs (meQTLs), we performed EWAS in RS for each SNP. Significant associations (FDR < 0.05) were detected at 14 unique CpG sites (Fig. 1), 12 sites in *cis* (within the window of 400 kb) (Fig. 2), and two in different chromosomes 6 and 12 (Table 2, model 1). Of these 14 CpG sites, 10 were significantly replicated in the independent sample from RS at a significance level of $P < 0.0019$, corresponding to the Bonferroni correction for number of tests performed in the replication sample ($n = 26$) (Table 2, model 1). All four SNPs were significantly associated with differential methylation at three CpG sites (cg18825076, cg04882995 and cg04140906) in *IREB2*, *CHRNA3* and *PSMA4*, respectively (Table 2, Fig. 2). Addition of smoking as a confounder did not change the results (Table 2, model 2), suggesting that the effects of the SNPs on DNA methylation are independent of smoking. However, significant genetic–environment interaction between pack-years and rs12914385:C>T (*CHRNA3*) was observed for DNA methylation levels at two CpG sites including the top hit, cg18825076 (*IREB2*; $P_{\text{interaction}} = 5.0 \times 10^{-4}$) and cg00540400 (*ADAMTS7-MORF4L1*; $P_{\text{interaction}} = 1.7 \times 10^{-3}$) (Table S2). The association between rs12914385:C>T (*CHRNA3*) and cg18825076 and cg00540400 remained significant, albeit slightly decreased (Table S2). However, the direction of the effect of both

significant interactions was opposite in the discovery and replication cohorts (Table S3), suggesting this is likely a false positive finding. Interestingly these two CpG sites, i.e., cg18825076 and cg00540400, and a third one, cg04140906 in *PSMA4*, were nominally associated ($P < 0.04$) with COPD (Table 3, Table S4, model 1). When correcting for smoking (Table 3, Table S4, model 2) the association between cg18825076 and COPD disappeared ($P = 0.16$), while the association of the other two CpG sites (cg00540400 and cg04140906) became stronger ($P < 0.03$). None of the CpGs were associated with FEV₁/FVC (Table S5).

Results involving differential DNA methylation at *IREB2* are of special interest, as we have detected association with COPD risk allele (rs12914385:C>T) and with COPD status. These relationships are illustrated in Fig. 3. We found a higher frequency of T allele among the COPD cases, compared to the controls in our dataset (Fig. 3a). Further, the T allele was associated with lower DNA methylation of the cg18825076 (Fig. 3b). Lastly, COPD cases showed lower DNA methylation at cg18825076 compared to controls (Fig. 3c).

Finally, in the eQTL analysis in 1087 lung tissue samples from LES, significant association (FDR < 0.05) was observed at 15 expression probe sets both in *cis* and *trans*

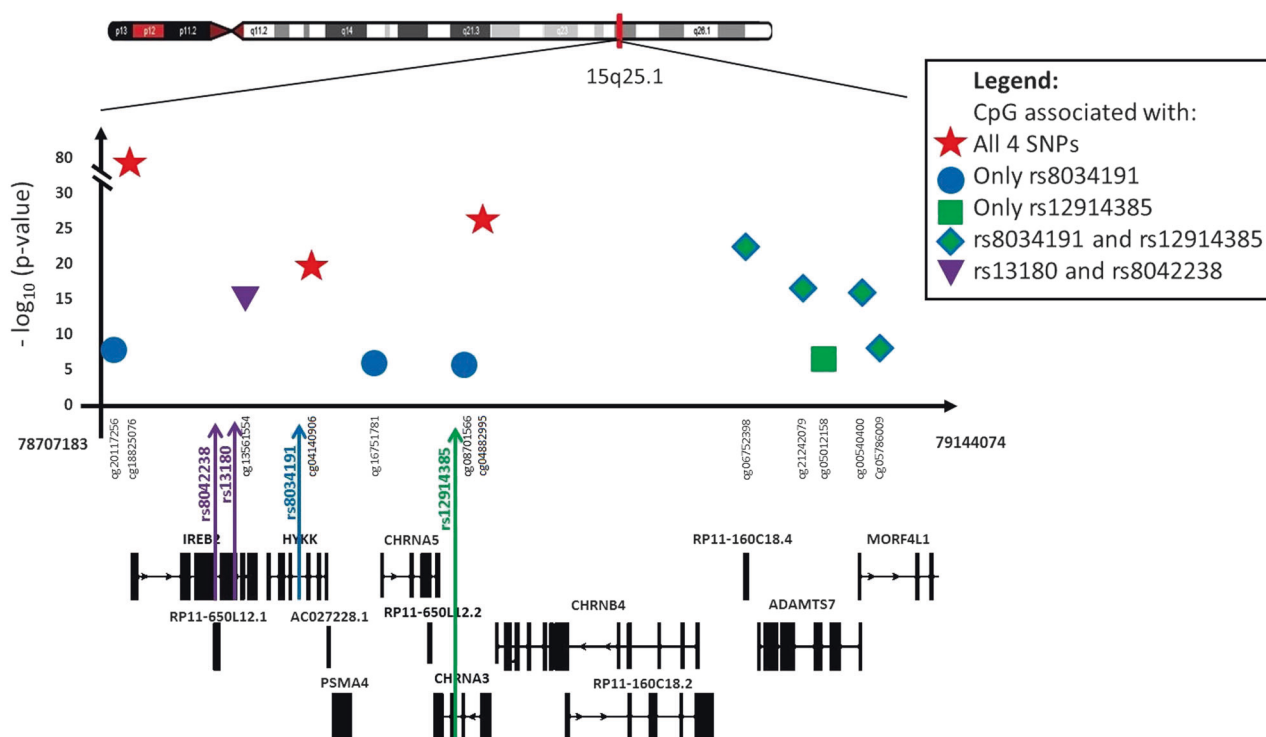


Fig. 2 Regional P -value plot of chromosome 15q25.1 with significant SNP-CpG associations. Every point in the plot represents one SNP-CpG association ($n = 12$). The x -axis shows all genes in the region as well as the position of SNPs of interest. The y -axis shows negative

logarithm of the P -values of the associations of CpGs with the corresponding SNP (distinguished by different colours). In the case a CpG associates with more than one SNP, the smaller P -value is taken into account

Table 2 Association of SNPs with epigenome-wide DNA methylation in two models in discovery and replication cohorts

SNP	CpG	Annotation of the CpG (hg19)	Discovery (<i>N</i> = 723)						Replication (<i>N</i> = 766)					
			Model 1			Model 2			Model 1			Model 2		
			β	SE	<i>P</i>	β	SE	<i>P</i>	β	SE	<i>P</i>	β	SE	<i>P</i>
rs12914385 <i>CHRNA3</i> intronic variant chr15:g.78606381C>T	cg18825076 ^a	chr15:g.78729989, <i>IREB2</i>	-0.0225	0.0013	1.54×10^{-55}	-0.0225	0.0013	1.19×10^{-55}	-0.0309	0.0016	1.39×10^{-68}	-0.0307	0.0016	2.05×10^{-67}
cg04882995 ^a chr15:g.78912658, <i>CHRNA3</i>	cg04882995 ^a	chr15:g.78912658, <i>CHRNA3</i>	0.0124	0.0011	1.59×10^{-27}	0.0123	0.0011	6.34×10^{-27}	0.0148	0.0014	4.61×10^{-25}	0.0148	0.0014	7.46×10^{-25}
	cg06752398	chr15:g.79053858, <i>ADAMTS7</i>	0.0165	0.0016	2.34×10^{-23}	0.0164	0.0016	6.62×10^{-23}	NA	NA	NA	NA	NA	NA
	cg21242079	chr15:g.79101063, <i>ADAMTS7</i>	0.0195	0.0022	3.90×10^{-17}	0.0194	0.0022	4.64×10^{-17}	0.0224	0.0034	8.36×10^{-11}	0.022	0.0034	1.95×10^{-10}
	cg00540400	chr15:g.79124168, <i>ADAMTS7-MORF4L1</i>	-0.0235	0.0028	1.19×10^{-16}	-0.0239	0.0027	3.07×10^{-17}	-0.0334	0.0039	1.51×10^{-16}	-0.0334	0.004	2.13×10^{-16}
	cg04140906 ^a	chr15:g.78833505, <i>PSMA4</i>	-0.0061	0.0011	3.25×10^{-8}	-0.0062	0.0011	2.08×10^{-8}	-0.0125	0.0021	3.49×10^{-9}	-0.0125	0.0021	4.35×10^{-9}
	cg05786009	chr15:g.79152474, <i>ADAMTS7-MORF4L1</i>	0.0037	0.0007	3.96×10^{-8}	0.0037	0.0007	5.15×10^{-8}	0.0048	0.0011	3.28×10^{-5}	0.0048	0.0011	3.16×10^{-5}
	cg05012158	chr15:g.79051863, <i>ADAMTS7</i>	-0.0058	0.0011	1.04×10^{-7}	-0.0058	0.0011	1.37×10^{-7}	NA	NA	NA	NA	NA	NA
	cg21436520	chr6:g.53413189, <i>GCLC-LOC101929136</i>	0.006	0.0012	4.74×10^{-7}	0.0061	0.0012	2.75×10^{-7}	-0.0007	0.0021	0.75139	-0.0004	0.0021	0.830983
	cg18825076 ^a	chr15:g.78729989, <i>IREB2</i>	-0.0277	0.0013	7.41×10^{-80}	-0.0279	0.0013	3.11×10^{-80}	-0.0353	0.0015	2.39×10^{-87}	-0.0352	0.0015	4.51×10^{-86}
	cg04882995 ^a	chr15:g.78912658, <i>CHRNA3</i>	0.0129	0.0012	2.72×10^{-26}	0.0128	0.0012	1.24×10^{-25}	0.016	0.0014	1.96×10^{-27}	0.0161	0.0014	1.89×10^{-27}
rs8034191 <i>HYKK</i> intronic variant chr15:g.78513681T>C	cg06752398	chr15:g.79053858, <i>ADAMTS7</i>	0.014	0.0017	5.08×10^{-15}	0.0139	0.0018	1.32×10^{-14}	NA	NA	NA	NA	NA	NA
	cg21242079	chr15:g.79101063, <i>ADAMTS7</i>	0.0189	0.0024	2.49×10^{-14}	0.0188	0.0024	2.92×10^{-14}	0.022	0.0035	8.71×10^{-10}	0.0217	0.0036	1.85×10^{-9}
	cg20117256	chr15:g.78726576, 4 kb 5' <i>IREB2</i>	0.024	0.0039	1.98×10^{-9}	0.0235	0.004	4.31×10^{-9}	0.0268	0.0034	1.44×10^{-14}	0.0266	0.0034	3.47×10^{-14}
	cg00540400	chr15:g.79124168, <i>ADAMTS7-MORF4L1</i>	-0.0181	0.003	3.02×10^{-9}	-0.0185	0.003	1.29×10^{-9}	-0.0304	0.0041	6.34×10^{-13}	-0.0304	0.0042	9.67×10^{-13}
	cg16751781	chr15:g.78858589, <i>CHRNA5</i>	-0.0037	0.0007	1.89×10^{-7}	-0.0036	0.0007	2.40×10^{-7}	-0.0055	0.0016	0.000467	-0.0057	0.0016	0.000312
	cg21436520	chr6:g.53413189, <i>GCLC-LOC101929136</i>	0.0063	0.0013	7.75×10^{-7}	0.0065	0.0013	4.12×10^{-7}	-0.0002	0.0021	0.909182	-0.0001	0.0021	0.970673
	cg08701566	chr15:g.78911099, <i>CHRNA3</i>	-0.0037	0.0007	8.90×10^{-7}	-0.0038	0.0007	5.35×10^{-7}	-0.0077	0.0013	3.56×10^{-9}	-0.0078	0.0013	2.55×10^{-9}
	cg05786009	chr15:g.79152474, <i>ADAMTS7-MORF4L1</i>	0.0035	0.0007	1.09×10^{-6}	0.0035	0.0007	1.44×10^{-6}	0.0045	0.0012	0.000151	0.0045	0.0012	0.000167
	cg04140906 ^a	chr15:g.78833505, <i>PSMA4</i>	-0.0057	0.0012	1.37×10^{-6}	-0.0059	0.0012	8.65×10^{-7}	-0.011	0.0022	5.75×10^{-7}	-0.0111	0.0022	5.67×10^{-7}
	cg23556238	chr12:g.125298876, <i>SCARBI</i>	-0.0071	0.0015	1.53×10^{-6}	-0.007	0.0015	2.14×10^{-6}	-0.0034	0.0012	0.005374	-0.0037	0.0012	0.002745
rs13180 <i>IREB2</i> synonymous variant chr15:g.78497146C>T	cg18825076 ^a	chr15:g.78729989, <i>IREB2</i>	-0.0201	0.0013	4.84×10^{-45}	-0.0201	0.0013	1.32×10^{-44}	-0.0257	0.0017	8.23×10^{-47}	-0.0256	0.0016	1.2×10^{-46}
rs8042238 <i>IREB2</i>	cg04140906 ^a	chr15:g.78833505, <i>PSMA4</i>	-0.01	0.001	1.02×10^{-20}	-0.0101	0.001	4.65×10^{-21}	-0.0126	0.002	1.23×10^{-9}	-0.0125	0.002	1.28×10^{-9}
	cg13561554	chr15:g.78795944, <i>IREB2-HYKK</i>	-0.0245	0.0028	5.59×10^{-17}	-0.0244	0.0029	8.73×10^{-17}	-0.0207	0.0021	1.26×10^{-21}	-0.0206	0.0021	2×10^{-21}
	cg04882995 ^a	chr15:g.78912658, <i>CHRNA3</i>	0.0072	0.0011	5.44×10^{-10}	0.0071	0.0011	9.84×10^{-10}	0.0095	0.0014	2.79×10^{-11}	0.0095	0.0014	3.16×10^{-11}
	cg18825076 ^a	chr15:g.78729989, <i>IREB2</i>	-0.0201	0.0013	1.35×10^{-44}	-0.02	0.0013	3.64×10^{-44}	-0.0258	0.0016	1.47×10^{-47}	-0.0256	0.0016	2.12×10^{-47}

Table 2 (continued)

SNP	CpG	Annotation of the CpG (hg19)	Discovery (N = 723)						Replication (N = 766)					
			Model 1			Model 2			Model 1			Model 2		
			β	SE	P	β	SE	P	β	SE	P	β	SE	P
intronic variant chr15:g.78481929C>T	cg04140906 ^a	chr15:g.78833505, <i>PSMA4</i>	-0.01	0.001	5.47×10^{-21}	-0.0102	0.001	2.43×10^{-21}	-0.0129	0.002	4.09×10^{-10}	-0.0129	0.002	4×10^{-10}
	cg13561554	chr15:g.78795944, <i>IREB2-HYKK</i>	-0.0246	0.0028	4.64×10^{-17}	-0.0245	0.0029	7.24×10^{-17}	-0.0203	0.0021	4.77×10^{-21}	-0.0202	0.0021	8.11×10^{-21}
	cg04882995 ^a	chr15:g.78912658, <i>CHRNA3</i>	0.0072	0.0011	4.37×10^{-10}	0.0071	0.0011	8.02×10^{-10}	0.0095	0.0014	2.56×10^{-11}	0.0095	0.0014	2.73×10^{-11}

In model 1, coefficients are corrected for sex, age, technical covariates and different white blood cellular proportions; in model 2, coefficients are additionally adjusted for current smoking and pack-years smoked; significant results are shown in bold

β Regression coefficient estimates from linear regression model regressing DNA methylation levels on indicated SNPs, N number of participants in the analysis, P P-value of the significance

^aMethylation sites significantly associated with all four SNPs

(Table 4). All four SNPs were significantly associated with probe set 100154936_TGI_at (3' untranslated region (UTR) of the *IREB2* gene, $P < 3.2 \times 10^{-98}$) (Fig. 4). The T allele of rs12914385:C>T was associated with higher expression of this probe. *Trans*-eQTL effects were observed in the chromosomes 2, 3, 4, 13 and 14 (Table 4). The rs8034191:T>C was associated with gene expression of *FAM13A* intron in chromosome 4 (100158626_TGI_at), a very well-known gene involved in COPD [10].

Discussion

In the current study we show that four COPD-associated SNPs (rs12914385:C>T, *CHRNA3*; rs8034191:T>C, *HYKK*; rs13180:C>T and rs8042238:C>T, *IREB2*) in the 15q25.1 locus are also blood meQTLs that regulate mostly nearby DNA methylation levels. All variants are associated with differential DNA methylation of *IREB2*, *CHRNA3* and *PSMA4* genes, independently of smoking. We further show that DNA methylation at two sites in genes *IREB2* and *PSMA4*, together with the site between *ADAMTS7* and *MORF4L1*, are associated with COPD. Finally, we show that all four SNPs are also lung *cis*- and *trans*-eQTLs, affecting the expression of several genes including *IREB2*, *PSMA4*, *CHRNA3*, *CHRNA5*, *HYKK*, *FAM13A*, *KLC1* and *TRIM13*. Our results demonstrate that COPD-SNPs shape the epigenetic regulatory landscape in the 15q25.1 locus in blood and lung tissues, and suggest that the genetic risk of these SNPs on COPD might be mediated and/or modified by DNA methylation levels in this region. Overall, our findings put forward the role of DNA methylation in COPD as an important mechanism in the complex regulation of the 15q25.1 locus.

Rs12914385:C>T and rs8034191:T>C are in moderate LD (Table S6), and hence as expected they yielded multiple overlapping results. The rs13180:C>T and rs8042238:C>T are in perfect LD (Table S6) and thus showed almost the same effects. We show that rs12914385:C>T, rs8034191:T>C and rs13180:C>T are associated with COPD in RS, but only rs12914385:C>T in LES, which is expected given the lower sample size. While rs12914385:C>T was not vastly studied, rs8034191:T>C is a well-studied SNP with regards to COPD and lung function in different populations [32–35]. It has been shown that rs8034191:T>C is associated with an increased risk for COPD independent of rs13180:C>T and pack-years of smoking [33]. On the contrary, the same study shows that the association of rs13180:C>T with severe COPD is possibly driven by moderate LD with rs8034191:T>C ($r^2 = 0.21$). However, a study in a Chinese population shows association of both SNPs with lung function but not with COPD, and the association of rs8034191:T>C with pack-years in COPD

Table 3 Association of COPD with DNA methylation at 10 replicated sites from the meQTL analysis; meta-analysed results ($N = 1339$)

Trait	CpG	Position (hg19)	Gene	Model 1			Model 2		
				β	SE	P	β	SE	P
COPD	cg18825076	78729989	<i>IREB2</i>	−5.165	2.514	0.04	−3.65	2.58	0.158
	cg04882995	78912658	<i>CHRNA3</i>	1.6	3.098	0.606	−0.57	3.225	0.86
	cg21242079	79101063	<i>ADAMTS7</i>	2.321	1.354	0.086	2.454	1.389	0.077
	cg00540400	79124168	<i>ADAMTS7-MORF4L1</i>	2.577	1.27	0.042	3.228	1.313	0.014
	cg04140906	78833505	<i>PSMA4</i>	−5.674	2.714	0.037	−6	2.772	0.03
	cg05786009	79152474	<i>ADAMTS7-MORF4L1</i>	2.043	4.307	0.635	1.723	4.469	0.7
	cg20117256	78726576	<i>IREB2</i>	−1.227	1.188	0.302	−1.54	1.23	0.212
	cg16751781	78858589	<i>CHRNA5</i>	0.891	3.57	0.803	2.092	3.707	0.573
	cg08701566	78911099	<i>CHRNA3</i>	6.785	3.971	0.087	6.333	4.136	0.126
	cg13561554	78795944	<i>IREB2-HYKK</i>	−1.185	1.521	0.436	−0.62	1.559	0.689

Model 1 adjusted for age, sex, technical covariates and different white blood cellular proportions; model 2 additionally adjusted for current smoking and pack-years; nominally significant results are shown in bold

β coefficient estimates from the logistic regression models, P P -value of the significance

cases [35]. A study including 3424 COPD cases and 1872 controls showed that the association of rs8034191:T>C with COPD is 30% mediated by pack-years, and this mediation increases to 42% when adjusted for rs13180:C>T [15]. In our study we show that the rs8034191:T>C yielded more epigenome-wide significant results and that some overlapped with rs13180:C>T (Fig. 2).

The most interesting finding of this study is that all four SNPs influence the three CpG sites in the *IREB2*, *CHRNA3* and *PSMA4* genes. Furthermore, the same site in *PSMA4* is also associated with COPD, independent of smoking. Site in *IREB2*, our top hit, is also associated with COPD, but this association drops after correcting for smoking. This suggests that the four genetic variants may influence COPD susceptibility through changes in DNA methylation of *IREB2*, *PSMA4* and *CHRNA3*.

We focused on chromosome 15q25.1 region hits from COPD GWASs. However, this region has also been reported in the association with smoking by several large smoking genetics consortia [12, 13]. They showed that locus in 15q25.1, represented by rs16969968:G>A, and other SNPs are mostly associated with smoking quantity. Saccone et al. [36] showed no significant association between rs16969968:G>A and COPD in smokers, adjusted for cigarettes per day. This SNP was in high LD with our SNPs, rs12914385:C>T ($r^2 = 0.84$) and rs8034191:T>C ($r^2 = 0.93$), but in this study we show that the proposed pathway, in which SNPs act as meQTLs, is mainly independent of smoking. Additionally, none of our four SNPs were found to be significant meQTLs in the brain in a study of nicotine dependence [37]. Despite the well-described role

of smoking behaviour in shaping the methylome at multiple tissues [25, 26], our SNP-DNA methylation associations could not be explained by exposure to smoking, making a mediating effect of smoking improbable. In line with our observations, the studies of the role of smoking on DNA methylation have failed to show an effect in this region [25, 38–41]. In a recent large smoking EWAS meta-analysis [41], none of our 10 replicated sites were associated with smoking, strengthening our claims. Yet, the failure to detect an association in these analyses does not necessarily represent absence of a true effect, it may rather reflect the lack of the statistical power to detect a true interaction. Further larger studies are needed to elucidate this question.

In blood, rs12914385:C>T, rs13180:C>T and rs8042238:C>T have been previously associated with differential expression of *IREB2* and *PSMA4*, while rs8034191:T>C was associated with the expression of *IREB2*, *PSMA4* and *CHRNA5* [17]. However, in the present study we examined the gene expression in the lung tissue and found that all four SNPs were also lung eQTLs for multiple genes in *cis*, including *IREB2*, *PSMA4*, *CHRNA3*, *CHRNA5*, and *HYKK*, as well as with other genes in *trans* involved in COPD pathogenesis, such as *FAM13A* (chromosome 4) [10]. Compared to Nguyen et al. [20], who only used a part of this dataset as discovery, we showed that in addition to *CHRNA3*, *CHRNA5* and *PSMA4*, the four SNPs were also associated with differential expression of *IREB2* and *HYKK*. Again, all SNPs were associated with differential expression of *IREB2* gene, suggesting that possibly the genetic variants are involved in the pathogenesis of the

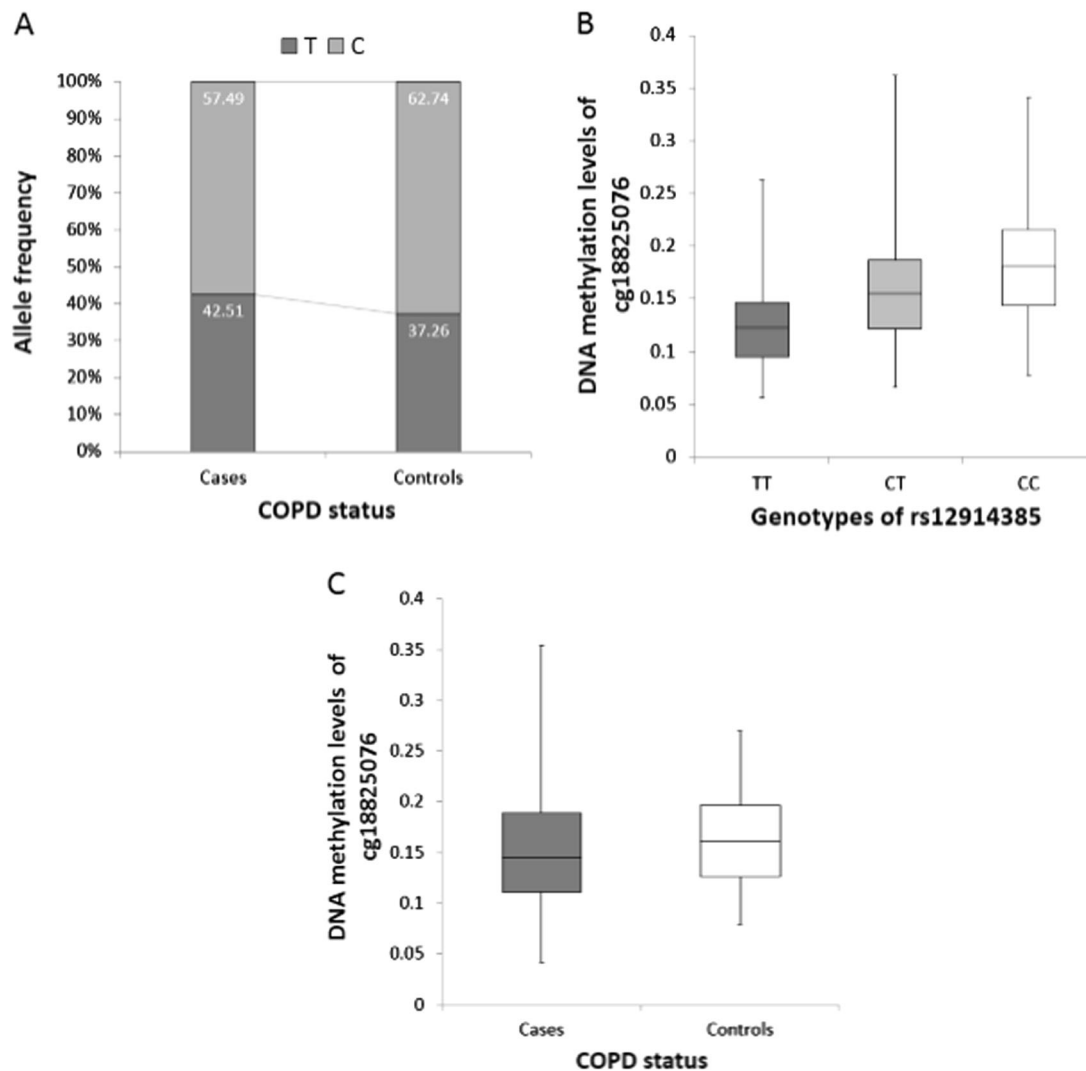


Fig. 3 Interplay between genetics, DNA methylation and COPD status at *IREB2* gene. **a** The frequency of T and C alleles of rs12914385:C>T (y-axis) among COPD cases and controls (x-axis) ($N = 1339$, $P = 0.043$). Homozygotes were counted as carrying two and heterozygotes as carrying one copy of a given allele. **b** Differences in cg18825076

DNA methylation levels (y-axis) between rs12914385:C>T genotypes (x-axis) ($N = 1489$, $P = 1.05 \times 10^{-125}$). **c** Differences in cg18825076 DNA methylation levels (y-axis) in COPD cases and controls (x-axis) ($N = 1339$, $P = 0.04$)

disease through differential DNA methylation and regulation of expression.

Based on our results in relation to *IREB2*, we propose a disease model in which the COPD-risk allele (rs12914385: C>T, *CHRNA3*) exerts its risk by lowering the DNA methylation level at *IREB2* gene and subsequently increasing its expression in COPD patients. The lower level of *IREB2* DNA methylation in blood from COPD cases and the positive effect of the risk allele on gene expression in lungs support this scenario. As we do not have the methylation and expression data in the same tissue, we were not able to validate this hypothesis directly, but future integrative studies in lung tissue should elucidate this further. *IREB2* gene is coding the RNA-binding protein that binds

to iron-responsive elements and can regulate the expression of transferrin receptor and ferritin by changing its own protein expression and thereby regulate iron metabolism, important in pathogenesis of lung diseases [42]. It is shown that *IREB2* gene interacts with *MYC* and *MAX* genes involved in the regulation of the gene transcription through epigenetic changes [43].

The strength of this study is that our findings are based on a large and unique sample of patients who are in-depth characterised genomically and epigenetically and our findings on the role of the genetic variants in blood corroborate with the changes in the transcriptome in lung. However, there are some limitations to our study. The first and main limitation is the use of blood in the DNA methylation

Table 4 Association of COPD-SNPs with transcriptome-wide gene expression in lungs ($N = 1087$)

SNP	Tested allele	Probe	Chr	Position (hg19)	Annotation	β	SE	P
rs12914385 CHRNA3 intronic variant chr15:g.78606381C>T	T	100154936_TGI_at ^a	15	78792408-78792890	3'UTR <i>IREB2</i> (AF147302)	0.23002	0.0091	2.13×10^{-142}
		100148404_TGI_at	15	78832747-78841563	<i>PSMA4</i> (NM_002789)	-0.06463	0.0058	1.62×10^{-28}
		100156434_TGI_at	15	78857862-78887611	<i>CHRNA5</i> (NM_000745)	0.13973	0.0139	1.00×10^{-23}
		100313804_TGI_at	15	78828532-78828986	<i>mRNA intron HYKK</i> (AA532820)	-0.09604	0.0152	2.38×10^{-10}
		100132425_TGI_at	15	78799906-78829715	<i>HYKK</i> (NM_001083612)	0.082	0.0147	2.56×10^{-8}
		100123649_TGI_at	15	78887647-78913637	<i>CHRNA3</i> (NM_000743)	0.11452	0.0255	7.25×10^{-6}
rs8034191 HYKK intronic variant chr15:g.78513681T>C	C	100154936_TGI_at ^a	15	78792408-78792890	3'UTR <i>IREB2</i> (AF147302)	-0.2436	0.0084	3.27×10^{-184}
		100148404_TGI_at	15	78832747-78841563	<i>PSMA4</i> (NM_002789)	0.0719	0.0056	3.17×10^{-37}
		100156434_TGI_at	15	78857862-78887611	<i>CHRNA5</i> (NM_000745)	-0.1594	0.0136	9.11×10^{-32}
		100313804_TGI_at	15	78828532-78828986	<i>mRNA intron HYKK</i> (AA532820)	0.1108	0.0149	9.15×10^{-14}
		100132425_TGI_at	15	78799906-78829715	<i>HYKK</i> (NM_001083612)	-0.1009	0.0145	2.90×10^{-12}
		100146326_TGI_at	3	31745705-31763038	<i>OSBPL10</i> (BX103488)	0.0491	0.011	7.57×10^{-6}
		100312267_TGI_at	15	78887350-78889074	<i>CHRNA3</i> (AJ584709)	-0.0901	0.0203	9.42×10^{-6}
		100123649_TGI_at	15	78887647-78913637	<i>CHRNA3</i> (NM_000743)	-0.108	0.0253	2.02×10^{-5}
		100158626_TGI_at	4	89658713-89659241	<i>mRNA intron FAM13A</i> (AI922966)	-0.0632	0.015	2.37×10^{-5}
		100143446_TGI_at	15	78730518-78793798	<i>IREB2</i> (NM_004136)	-0.0259	0.0062	2.76×10^{-5}
rs13180 IREB2 synonymous variant chr15:g.78497146C>T	T	100154936_TGI_at ^a	15	78792408-78792890	3'UTR <i>IREB2</i> (AF147302)	-0.2046	0.0097	3.17×10^{-98}
		100129239_TGI_at	15	78780991-78786607	<i>IREB2</i> (AW118658)	0.059	0.0114	2.46×10^{-7}
		100143446_TGI_at	15	78730518-78793798	<i>IREB2</i> (NM_004136)	-0.0309	0.0064	1.41×10^{-6}
		100161812_TGI_at	2	242943360-242948160	(AK126530)	-0.0381	0.0086	8.78×10^{-6}
		100140544_TGI_at	14	104145786-104168220	<i>KLC1</i> (CO247613)	0.0422	0.0095	9.38×10^{-6}
		100144578_TGI_at	2	37869087-37871279	UTR <i>CDC42EP3</i> (AK055915)	-0.0428	0.0099	1.44×10^{-5}
		100141776_TGI_at	13	50571143-50592603	<i>TRIM13</i> (NM_052811)	0.0412	0.0098	2.83×10^{-5}
rs8042238 IREB2 intronic variant chr15:g.78481929C>T	T	100154936_TGI_at ^a	15	78792408-78792890	3'UTR <i>IREB2</i> (AF147302)	-0.2049	0.0097	1.10×10^{-98}
		100129239_TGI_at	15	78780991-78786607	<i>IREB2</i> (AW118658)	0.0587	0.0114	2.84×10^{-7}

Table 4 (continued)

SNP	Tested allele	Probe	Chr	Position (hg19)	Annotation	β	SE	<i>P</i>
		100143446_TGI_at	15	78730518-78793798	<i>IREB2</i> (NM_004136)	-0.0312	0.0064	1.09×10^{-6}
		100161812_TGI_at	2	242943360-242948160	(AK126530)	-0.0389	0.0086	5.73×10^{-6}
		100140544_TGI_at	14	104145786-104168220	<i>KLC1</i> (CO247613)	0.0423	0.0095	8.79×10^{-6}
		100144578_TGI_at	2	37869087-37871279	UTR <i>CDC42EP3</i> (AK055915)	-0.0429	0.0099	1.36×10^{-5}
		100141776_TGI_at	13	50571143-50592603	<i>TRIM13</i> (NM_052811)	0.0415	0.0098	2.47×10^{-5}

Significant results are shown in bold

 β Regression coefficient estimates from linear regression models regressing gene expression on indicated SNPs (see model in the Methods section), *P* *P*-value of the significance^aProbe set associated with all four SNPs

analysis as a proxy for clinically and biologically relevant changes that develop in the lungs. In the absence of lung tissue DNA methylation measurements, blood is the most reasonable surrogate for examination of methylation changes related to COPD and smoking. This is because the disease, apart from affecting lung tissue, also induces systemic changes and has been associated with elevation in markers of systemic inflammation [44]. Furthermore, studies comparing the DNA methylation patterns in multiple tissues confirm that there is a great overlap in patterns, encouraging us to believe that blood is a good surrogate to study differences that occur in lungs [45, 46]. The second limitation is the use of COPD definition based on the pre-bronchodilator spirometry. This measure demonstrates the variability of the smooth muscle contraction, while by using post-bronchodilator spirometry we can observe the irreversibility of the airflow limitation, the main characteristic of COPD [47]. However, in the attempt to minimise potential misclassification, we have identified and excluded all possible asthmatic patients. In addition, some epidemiological studies show that both pre- and post-bronchodilator spirometry predicted mortality related to COPD with a similar degree of accuracy [48]. The third limitation to this study is the use of the whole lung tissue for gene expression analysis, which is very heterogeneous, instead of identifying the source cells for our eQTL signals. Finally, we formulated hypothesis based on results obtained from a cross-sectional study, in which inferences on directionality of effects are complicated. Future results based on longitudinal studies will help to support the role of DNA methylation and gene expression as regulatory mediators for COPD genetic risk. In addition, further replication of our results in other ethnicities and more diverse studies may corroborate our results. Since genetic variants identified through GWASs usually map to non-coding intergenic and intronic regulatory regions, the functional role is often unclear [49, 50]. They are more likely to modulate gene expression through regulatory mechanisms and epigenetic modifications (e.g., DNA methylation), as we show in this study. However, we do not show significant association of all SNPs with the disease, but this is expected since our sample size is considerably smaller than that used in the original GWAS. Our study did not aim to prioritise between the four variants in terms of relevance to COPD, but to investigate if the variants in 15q25.1 region, associated to COPD, are involved in regulatory mechanisms. The changes in DNA methylation levels that we observed in blood are small, most likely because we assess differential methylation in a variety of cell types. More substantial changes may be found in lung tissue in future integrative studies.

In summary, we found evidence suggesting that genetic variations underlying *IREB2*, *HYKK* and *CHRNA3* act as *cis*

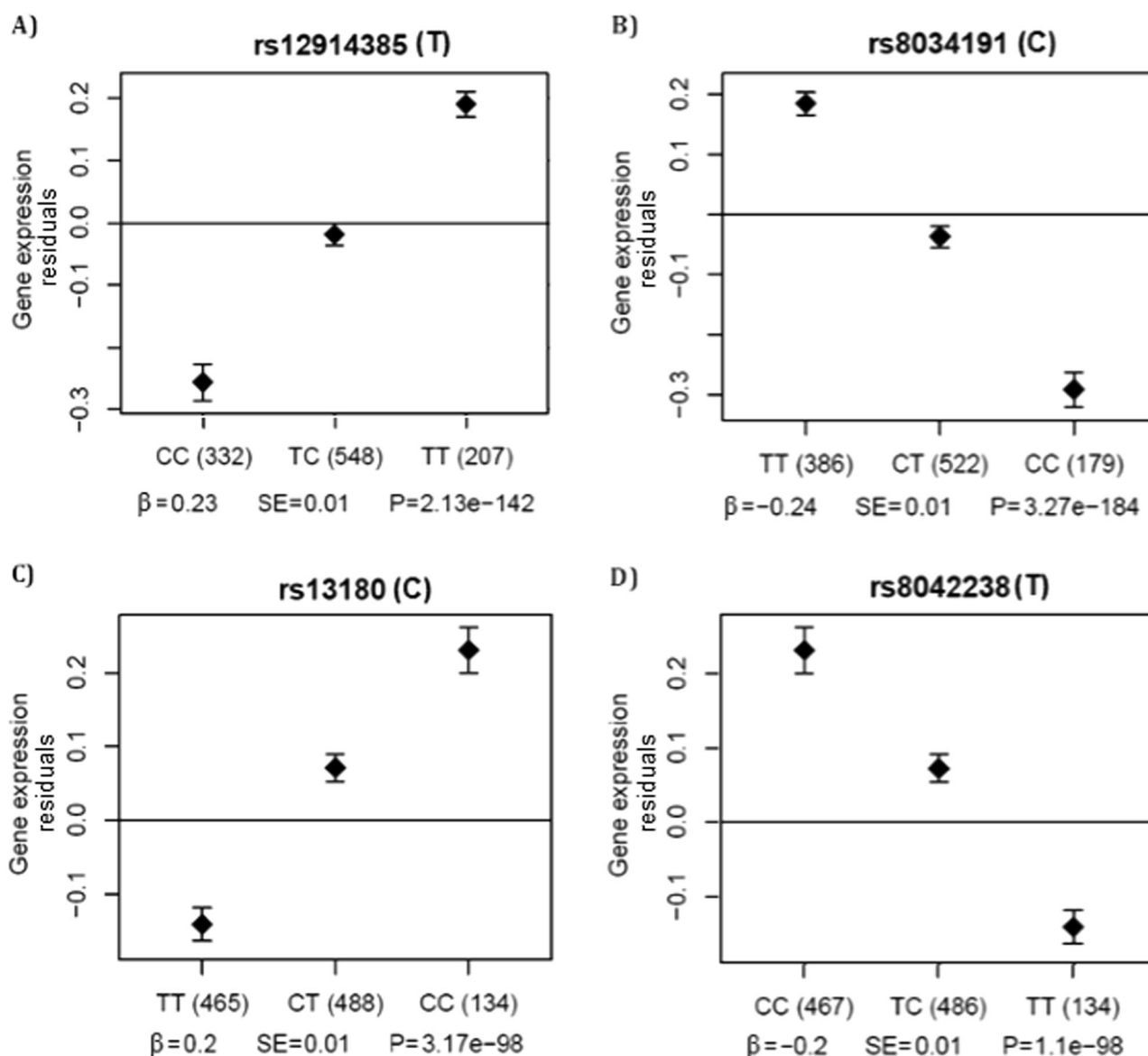


Fig. 4 3'UTR *IREB2* expression plot in regards to the genotypes of four SNPs. Plots of genotype-specific mean residuals (with 95% confidence intervals) from the linear regression models on gene expression levels adjusted for disease status, age, sex, smoking status and the cohort-specific number of PCs. The association of residuals of the expression levels (probe set 100154936_TGI_at, 3'UTR *IREB2*) on

the y-axis with genotypes of **a** rs12914385:C>T, **b** rs8034191:T>C, **c** rs13180:C>T and **d** rs8042238:C>T on the x-axis, risk/tested allele is given in brackets; β is regression coefficient estimate, P is P -value of the significance, TT/TC/CC are different genotypes of the given SNP and number of carriers is given in brackets

meQTLs and eQTLs. They all affect the DNA methylation and expression of *IREB2*, which also contributes to the risk of having COPD. We did not find evidence that smoking mediates these relationships, although this should be corroborated in larger sample sizes. This finding is compatible with the hypothesis that the genetic variants are involved in the pathogenesis of COPD through differential methylation and regulation of expression. Future integrative studies quantifying both DNA methylation and gene expression in lung tissue, as well as functional studies, are needed to confirm suggested hypothesis.

Declarations

Ethics approval and consent to participate

The Rotterdam Study has been approved by the Medical Ethics Committee of the Erasmus MC and by the Ministry of Health, Welfare and Sport of the Netherlands, implementing the Population Studies Act: Rotterdam Study. All participants provided written informed consent to participate in the study and to obtain information from their treating physicians. Patients from LES provided written

informed consent and the study was approved by the ethics committees of the Institut universitaire de cardiologie et de pneumologie de Québec and the UBC-Providence Health Care Research Institute Ethics Board for Laval and UBC, respectively. The study protocol was consistent with the Research Code of the University Medical Centre Groningen and Dutch national ethical and professional guidelines.

Acknowledgements The generation and management of the Illumina 450K methylation array data (EWAS data) and the Illumina exome chip v1.0 array data for the Rotterdam Study were executed by the Human Genotyping Facility of the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, Rotterdam, The Netherlands. The Exome chip array data set was funded by the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, from the Netherlands Genomics Initiative (NGI)/Netherlands Organisation for Scientific Research (NWO)-sponsored Netherlands Consortium for Healthy Aging (NCHA; project no. 050-060-810); the Netherlands Organization for Scientific Research (NWO; project number 184021007) and by the Rainbow Project (RP10; Netherlands Exome Chip Project) of the Biobanking and Biomolecular Research Infrastructure Netherlands (BBMRI-NL; www.bbMRI.nl). We thank Ms. Mila Jhamai, Ms. Sarah Higgins and Mr. Marijn Verkerk for their help in creating the exome chip database, and Carolina Medina-Gomez, Lennard Karsten and Linda Broer for QC and variant calling. We thank Michael Verbiest, Mila Jhamai, Sarah Higgins, Marijn Verkerk and Lisette Stolk for their help in creating the methylation database. The authors are grateful to the study participants, the staff from the Rotterdam Study and the participating general practitioners and pharmacists. The authors would like to thank the staff at the Respiratory Health Network Tissue Bank of the FRQS for their valuable assistance with the Lung eQTL dataset at Laval University. We would like to acknowledge members of the BIOS consortium (<https://www.bbMRI.nl/?p=259>).

Funding This study is sponsored by grant number 4.1.13.007 of Lung Foundation (Longfonds), the Netherlands. DAvdP, KdJ and NA are supported by grant number 4.1.13.007 of Lung Foundation. IN is supported by the ERAWEB scholarship. The Rotterdam Study is funded by Erasmus Medical Centre and Erasmus University, Rotterdam, Netherlands Organization for the Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam. The EWAS data was funded by the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, and by the Netherlands Organization for Scientific Research (NWO; project number 184021007). The Lung eQTL study at Laval University was supported by the Chaire de pneumologie de la Fondation JD Bégin de l'Université Laval, the Fondation de l'Institut universitaire de cardiologie et de pneumologie de Québec, the Respiratory Health Network of the FRQS, the Canadian Institutes of Health Research (MOP – 123369), and the Cancer Research Society and Read for the Cure. YB holds a Canada Research Chair in Genomics of Heart and Lung Diseases. The sponsors of this study played no role in the design of the study; in data collection, analysis, and interpretation; or in the writing and submission of the manuscript.

Author contributions IN and EC-M participated in the study design, analysis and interpretation of the data and drafting of the manuscript, tables and figures. HMB, DSP and CMvD obtained funding and supervised research. DAvdP, KdJ, JMV, CCvD, HMB, DSP, CMvD and NA determined the conception and study design, participated in the analyses and interpretation of data. LL performed statistical

analysis of Rotterdam Study data and contributed to the interpretation and writing of the manuscript. AGU, JBJvM, BHCS and GGB participated in the COPD data collection within the Rotterdam Study and contributed to the interpretation of the results. AF, MvdB, MO, YB and DCN participated in setting up the lung eQTL study and analyses. All authors contributed to the writing of the manuscript and approved the final version of the manuscript.

Compliance with ethical standards

Conflict of interest LL reports personal fees from Boehringer Ingelheim GmbH, non-financial support from Novartis, grants from AstraZeneca, grants and non-financial support from European Respiratory Society, grants and non-financial support from Belgian Respiratory Society, outside the submitted work. MvdB reports research grants paid to The University from Chiesi, GlaxoSmithKline (GSK), Astra Zeneca and TEVA, outside the submitted work. DCN is employed by Merck & Co. The University of Groningen has received money for Professor D. Postma regarding a grant for research from Astra Zeneca, Chiesi, Genentec, GSK and Roche. Fees for consultancies were given to the University of Groningen by Astra Zeneca, Boehringer Ingelheim, Chiesi, GSK, Takeda and TEVA. The other authors declare that they have no conflict of interest.

References

- Lozano R, Bin Abdulhak A, Birbeck G, Blyth F, Bolliger I, Boufous S, Bucello C, Burch M, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*. 2012;380:2095.
- de Jong K, Vonk JM, Timens W, et al. Genome-wide interaction study of gene-by-occupational exposure and effects on FEV1 levels. *J Allergy Clin Immunol*. 2015;136:1664–72. e1614
- Vestbo J, Hurd SS, Agustí AG, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Am J Respir Crit Care Med*. 2013;187:347–65.
- Zhou JJ, Cho MH, Castaldi PJ, Hersh CP, Silverman EK, Laird NM. Heritability of chronic obstructive pulmonary disease and related phenotypes in smokers. *Am J Respir Crit Care Med*. 2013;188:941–7.
- Gibson GJ, Loddenkemper R, Sibille Y, Lundback B, editors. The European lung white book: respiratory health and disease in Europe. European Respiratory Society; 2013 Sep 1. Sheffield, S10 2PX, UK.
- Soler Artigas M, Loth DW, Wain LV, et al. Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. *Nat Genet*. 2011;43:1082–90.
- Wilk JB, Shrine NR, Loehr LR, et al. Genome-wide association studies identify CHRNA5/3 and HTR4 in the development of airflow obstruction. *Am J Respir Crit Care Med*. 2012;186:622–32.
- Cho MH, McDonald M-LN, Zhou X, et al. Risk loci for chronic obstructive pulmonary disease: a genome-wide association study and meta-analysis. *Lancet Respir Med*. 2014;2:214–25.
- Lee JH, Cho MH, Hersh CP, et al. Genetic susceptibility for chronic bronchitis in chronic obstructive pulmonary disease. *Respir Res*. 2014;15:1.
- Cho MH, Boutaoui N, Klanderman BJ, et al. Variants in FAM13A are associated with chronic obstructive pulmonary disease. *Nat Genet*. 2010;42:200–2.

11. Pillai SG, Ge D, Zhu G, et al. A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. *PLoS Genet.* 2009;5:e1000421.
12. Tobacco and Genetics Consortium. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat Genet.* 2010;42:441–7.
13. Liu JZ, Tozzi F, Waterworth DM, et al. Meta-analysis and imputation refines the association of 15q25 with smoking quantity. *Nat Genet.* 2010;42:436–40.
14. Hung RJ, McKay JD, Gaborieau V, et al. A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. *Nature.* 2008;452:633–7.
15. Siedlinski M, Tingley D, Lipman PJ, et al. Dissecting direct and indirect genetic effects on chronic obstructive pulmonary disease (COPD) susceptibility. *Hum Genet.* 2013;132:431–41.
16. Budulac SE, Vonk JM, Postma DS, Siedlinski M, Timens W, Boezen MH. Nicotinic acetylcholine receptor variants are related to smoking habits, but not directly to COPD. *PLoS ONE.* 2012;7:e33386.
17. Westra H-J, Peters MJ, Esko T, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet.* 2013;45:1238–43.
18. Wang JC, Cruchaga C, Saccone NL, et al. Risk for nicotine dependence and lung cancer is conferred by mRNA expression levels and amino acid change in CHRNA5. *Hum Mol Genet.* 2009;18:3125–35.
19. Qiu W, Cho MH, Riley JH, et al. Genetics of sputum gene expression in chronic obstructive pulmonary disease. *PLoS ONE.* 2011;6:e24395.
20. Nguyen JD, Lamontagne M, Couture C, et al. Susceptibility loci for lung cancer are associated with mRNA levels of nearby genes in the lung. *Carcinogenesis.* 2014;35:2653–9.
21. DeMeo DL, Mariani T, Bhattacharya S, et al. Integration of genomic and genetic approaches implicates IREB2 as a COPD susceptibility gene. *Am J Human Genet.* 2009;85:493–502.
22. Vucic EA, Chari R, Thu KL, et al. DNA methylation is globally disrupted and associated with expression changes in chronic obstructive pulmonary disease small airways. *Am J Respir Cell Mol Biol.* 2014;50:912–22.
23. Yet I, Tsai PC, Castillo-Fernandez JE, Camero-Montoro E, Bell JT. Genetic and environmental impacts on DNA methylation levels in twins. *Epigenomics.* 2016;8:105–17.
24. Ladd-Acosta C, Fallin MD. The role of epigenetics in genetic and environmental epidemiology. *Epigenomics.* 2016;8:271–83.
25. Zeilinger S, Kuhnel B, Klopp N, et al. Tobacco smoking leads to extensive genome-wide changes in DNA methylation. *PLoS ONE.* 2013;8:e63812.
26. Hillemacher T, Frieling H, Moskau S, et al. Global DNA methylation is influenced by smoking behaviour. *Eur Neuropsychopharmacol.* 2008;18:295–8.
27. Qiu W, Baccarelli A, Carey VJ, et al. Variable DNA methylation is associated with chronic obstructive pulmonary disease and lung function. *Am J Respir Crit Care Med.* 2012;185:373–81.
28. Paliwal A, Vaissière T, Krais A, et al. Aberrant DNA methylation links cancer susceptibility locus 15q25.1 to apoptotic regulation and lung cancer. *Cancer Res.* 2010;70:2779–88.
29. Bossé Y. Updates on the COPD gene list. *Int J Chron Obstruct Pulmon Dis.* 2012;7:607–31.
30. Ligthart S, Steenaard RV, Peters MJ, et al. Tobacco smoking is associated with DNA methylation of diabetes susceptibility genes. *Diabetologia.* 2016;59:998–1006.
31. Netherlands BaBriot: BIOS 2016, Vol 2016.
32. Cui K, Ge X, Ma H. Four SNPs in the CHRNA3/5 alpha-neuronal nicotinic acetylcholine receptor subunit locus are associated with COPD risk based on meta-analyses. *PLoS ONE.* 2014;9:e102324.
33. Hardin M, Zielinski J, Wan ES, et al. CHRNA3/5, IREB2, and ADCY2 are associated with severe chronic obstructive pulmonary disease in Poland. *Am J Respir Cell Mol Biol.* 2012;47:203–8.
34. Kim WJ, Wood AM, Barker AF, et al. Association of IREB2 and CHRNA3 polymorphisms with airflow obstruction in severe alpha-1 antitrypsin deficiency. *Respir Res.* 2012;13:16.
35. Zhou H, Yang J, Li D, et al. Association of IREB2 and CHRNA3/5 polymorphisms with COPD and COPD-related phenotypes in a Chinese Han population. *J Hum Genet.* 2012;57:738–46.
36. Saccone NL, Culverhouse RC, Schwantes-An TH, et al. Multiple independent loci at chromosome 15q25.1 affect smoking quantity: a meta-analysis and comparison with lung cancer and COPD. *PLoS Genet.* 2010;6:e1001053.
37. Hancock DB, Wang J-C, Gaddis NC, et al. A multi-ancestry study identifies novel genetic associations with CHRNA5 methylation in human brain and risk of nicotine dependence. *Human Mol Genet.* 2015;24:5940–54.
38. Wan ES, Qiu W, Carey VJ, et al. Smoking-associated site-specific differential methylation in buccal mucosa in the COPD Gene Study. *Am J Respir Cell Mol Biol.* 2015;53:246–54.
39. Tspirouni LG, Yang T-P, Bell J, et al. Cigarette smoking reduces DNA methylation levels at multiple genomic loci but the effect is partially reversible upon cessation. *Epigenetics.* 2014;9:1382–96.
40. Shenker NS, Ueland PM, Polidoro S, et al. DNA methylation as a long-term biomarker of exposure to tobacco smoke. *Epidemiology.* 2013;24:712–6.
41. Joehanes R, Just AC, Marioni RE, et al. Epigenetic signatures of cigarette smoking. *Circ Cardiovasc Genet.* 2016;9:436–47.
42. Cheng Z, Dai LL, Song YN, et al. Regulatory effect of iron regulatory protein-2 on iron metabolism in lung cancer. *Genet Mol Res.* 2014;13:5514–22.
43. Ning B, Liu G, Liu Y, et al. 5-aza-2'-deoxycytidine activates iron uptake and heme biosynthesis by increasing c-Myc nuclear localization and binding to the E-boxes of transferrin receptor 1 (TfR1) and ferrochelatase (Fech) genes. *J Biol Chem.* 2011;286:37196–206.
44. Agustí A. Systemic effects of chronic obstructive pulmonary disease: what we know and what we don't know (but should). *Proc Am Thorac Soc.* 2007;4:522–5.
45. Lokk K, Modhukur V, Rajashekar B, et al. DNA methylome profiling of human tissues identifies global and tissue-specific methylation patterns. *Genome Biol.* 2014;15:r54.
46. Yang X, Shao X, Gao L, Zhang S. Systematic DNA methylation analysis of multiple cell lines reveals common and specific patterns within and across tissues of origin. *Hum Mol Genet.* 2015;24:4374–84.
47. Sterk PJ. Let's not forget: the GOLD criteria for COPD are based on post-bronchodilator FEV1. *Eur Respir J.* 2004;23:497–8.
48. Mannino DM, Diaz-Guzman E, Buist S. Pre- and post-bronchodilator lung function as predictors of mortality in the Lung Health Study. *Respir Res.* 2011;12:1.
49. Hindorf LA, Sethupathy P, Junkins HA, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci USA.* 2009;106:9362–7.
50. Maurano MT, Humbert R, Rynes E, et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science.* 2012;337:1190–5.