December 11, 2024

EPIGENOME-WIDE ASSOCIATION STUDIES OF COPD AND LUNG FUNCTION: A SYSTEMATIC REVIEW

Sandra Casas-Recasens[1,2]*, Raisa Cassim[3]*, Núria Mendoza[1,2], Alvar Agusti[1,2,4,5], Caroline Lodge[3], Shuai Li[6,7,8,9], Dinh Bui[3], David Martino[10,11], Shyamali C Dharmage[3‡], Rosa Faner[1,2,5.12‡]

* Shared first and ‡ shared last authors.

1. Fundació Clinic Recerca Biomedica-Institut d'Investigacions Biomediques August Pi i Sunyer (FCRB-IDIBAPS), Barcelona, Spain.
2. Centro de Investigación Biomédica en Red de Enfermedades Respiratorias (CIBERES), Spain.
3. Allergy and Lung Health Unit, School of Population and Global Health, The University of Melbourne, Melbourne, Australia.
4. Respiratory Institute, Hospital Clinic, Barcelona, Spain
5. Catedra Salud Respiratoria, University of Barcelona, Barcelona, Spain.
6. Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Australia
7. Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK
8. Precision Medicine, School of Clinical Sciences at Monash Health, Monash University, Clayton, Victoria, Australia
9. Murdoch Children's Research Institute, Royal Children's Hospital, Parkville, Victoria, Australia

10. Walyun Respiratory Research Centre, Telethon Kids Institute, University of Western Australia, Perth, Australia.
11. Centre for Food and Allergy Research, Murdoch Children's Research Institute, University of Melbourne.
12. University of Barcelona, Biomedicine Department. Barcelona, Spain.

Corresponding author:
Rosa Faner, PhD. University of Barcelona, IDIBAPS-CELLEX, CIBERES. c/Casanova 143, P2A, Barcelona 08036, Spain. Tel.: +34 93 227 1715; Fax: +34 93 227 1716; e-mail: rfaner@ub.edu

Funding: This work was supported in part by Instituto de Salud Carlos III (PI17/00379, PI18/00018, PI21/00735), European Research Council (ERC) under the Horizon Europe research and innovation programme (Grant Agreement No. 101044387) and SEPAR grants. Shuai Li is supported by SL, a National Health and Medical Research Council Emerging Leadership Fellow (GNT2017373). Rosa Faner is a Serra Hunter Fellow, Sandra Casas is recipient of a P-FIS (FI18/00106) grant. SD is supported by the Australian National Health Medical Research Council.

Main text: 4482 words**; References** 45**; Tables 1; Figures** 4
Online supplement: 24 Tables; 1 Figure
Key words: methylation, omics, COPD, emphysema, chronic bronchitis
Author contribution: Study conception and design, SCD, DM, RF, AA, EHW, CL, SL, DB, RC, SCR. Data acquisition: SCR, RC, NM. Data analysis: SCD, DM, SCR, RF, RC, EHW, CL. Manuscript draft preparation: All. Manuscript revision: All.

ABSTRACT

Background: Chronic Obstructive Pulmonary Disease (COPD) results from gene-environment interactions over the lifetime. These interactions are captured by epigenetic changes, such as DNA methylation. This systematic review synthesizes evidence from epigenome-wide association studies (EWAS) related to COPD and lung function.

Methods: A systematic literature search on PubMed, Embase and CINAHL database, identified 1947 articles that investigated epigenetic changes associated with COPD and lung function; 17 of them met our eligibility criteria from which data was manually extracted. Differentially methylated positions (DMPs) and/or annotated genes, were considered replicated if identified by ≥2 studies with a $p<1 \times 10\text{-}4$.

Results: Ten studies profiled DNA methylation changes in blood and 7 in respiratory samples, including surgically resected lung tissue (n=3), small airways epithelial brushings (n=2), bronchoalveolar lavage (n=1) and sputum (n=1). Main results showed: *(1)* high variability in study design, covariates considered and effect sizes, which prevented a formal meta-analysis; *(2)* in blood samples, 51 DMPs were replicated in relation to lung function and 12 related to COPD ; *(3)* in respiratory samples, 42 DMPs were replicated in relation to COPD but none in relation lung function; and, *(4)* in COPD vs. control studies, 123 genes (2.6% of total) were shared between ≥1 blood and ≥1 respiratory samples and associated with chronic inflammation, ion transport and coagulation.

Conclusions: There is high heterogeneity across published COPD/lung function EWAS studies. A few genes (n=123; 2.6%) were replicated in blood and respiratory samples, suggesting that blood can recapitulate some changes in respiratory tissues. These findings have implications for future research.

Trial registration: Open Science Frameworks (OSF)

Word count 254 (max=250)

INTRODUCTION

Chronic Obstructive Pulmonary Disease (COPD) is a major public health problem. COPD is considered to be the end result of dynamic and cumulative gene-environment interactions across the lifespan of the individual (1-5) . Epigenetic changes, such as DNA methylation, capture the resulting pressure dynamics of the exposome in a given individual (2). Over the past few years, several EWAS studies have assessed the methylome in relation to COPD and lung function. These studies provided initial evidence that molecular signatures may represent targetable pathways in COPD (6-8) and identified DNA methylation sites associated with early-life environmental factors and adult lung function changes (9). In 2017, Machin *et al.* published a systematic review of studies investigating *blood* DNA methylation in the general population in relation to lung function and COPD (10). Yet, whether epigenetic changes seen in blood correlated with those occurring in lung tissue was not explored and is currently unknown (11, 12). Further, the systematic review by Machin *et al.* could not include studies using the new 850k Illumina array platform because it was not available at the time. This 850k Illumina array platform doubles the number of screened CpGs and includes 350,000 additional CpGs located in enhancer regions identified by the FANTOM5 project across tissue types (10). A robust synthesis of the evidence from current EWAS studies can inform potential mechanistic studies aiming to identify plausible biological pathways and also novel drug targets.

Accordingly, this review sought to: *(1)* systematically review currently available EWAS studies (some of them using this new 850k Illumina array platform) determined in blood and/or in respiratory tissue samples (surgically resected, airway epithelial brushings, bronchoalveolar lavage fluid or sputum) in relation to COPD and lung function; *(2)* investigate the reproducibility of findings in and between these different tissue samples, considering both differentially methylated positions (DMPs, i.e. CpGs) and/or annotated genes, and (3) explore potential mechanistic pathways behind these associations.

METHODS

Full methods are provided in the online supplement. This systematic review was prospectively registered in Open Science Frameworks (OSF) and its methodology and reporting adheres to the PRISMA (13) guidelines for systematic reviews. A search was developed to identify publications that have investigated the relationship of DNA methylation with COPD and/or lung function using EWAS.

Search strategy

Two reviewers (SC and NM) independently searched PubMed, Embase and CINAHL online databases for pertinent original articles using the keywords and MeSH terms listed in Table S1, bound together with the Boolean operators "OR" and "AND". The final search was conducted on August 11th, 2023, when PubMed also included searches in preprint databases as BioRxiv. A flow diagram of the selection procedure is included in Figure 1 and the reasons of the final exclusion of articles is in Table S2. The overall quality and risk of bias was assessed using the Newcastle-Ottawa Scale (NOS) (Table S3).

Data extraction

The following data were extracted from each article: first author, publishing date, size and age of samples, clinical characteristics, number of differentially methylated sites detected, statistical threshold used and methylation platform (Table 1 and Table S4). Specific details on DMPs used for the reproducibility analysis are described in the online supplement and Table S5. For comparison across studies a p value cutoff p-value<1 x 10-4, was used, because it approximately corresponds to the significant threshold of the Illumina 27k platform (FDR<0.05).

RESULTS

Literature search and study selection

Our search strategy identified 2657 potential articles, 1947 after duplicated articles were excluded (Figure 1). After screening titles and abstracts, 35 articles underwent full text review,

from which 17 studies analyzing more than 30,000 individuals were included in this systematic review. Table 1 provides the list of these 17 included studies, with their participant profiles, year of publication, tissue studied, outcomes considered and main results. Tables S4-5 present detailed information on each study.

Tissue studied and study design

No study investigated epigenetic changes in blood and respiratory samples obtained simultaneously in the same individual.

Blood

As shown in Table 1, ten studies profiled whole blood DNA; of these: *(1)* two compared COPD patients with controls cross-sectionally (14, 15); *(2)* three investigated the association of DNA methylation with FEV_1 and/or FEV_1/FVC (16-18). One of them was cross-sectional(18), while two considered temporal changes (16, 17); and *(3)* five studies contrasted COPD vs. controls and investigated the association with FEV_1 and/or FEV_1/FVC cross-sectionally (19-23).

Respiratory tissues samples

Three different studies (24-26) compared lung tissue biopsies in COPD vs. controls (Table 1), and one of them (25) also evaluated the association with the severity of the disease as determined by FEV_1 % predicted. Two COPD vs. control studies profiled small airway epithelia (27, 28) with the first one also reporting the association with FEV_1 % predicted. Finally, we included one COPD vs controls study using sputum (29) and one using bronchoalveolar lavage (30).

Study participant characteristics

As detailed in Table 1, 15 studies compared the methylation of patients with COPD to that of controls. Yet the definition of COPD varied across studies: *(1)* six used pre-bronchodilator FEV_1/FVC<0.7 (14, 16, 18, 22, 23, 28); *(2)* five used post-bronchodilator FEV_1/FVC<0.7 (15, 19, 21, 25, 30); and *(3)* six did not specify whether pre- or post-bronchodilator values were used (17, 20, 24, 26, 27, 29).

The severity of airflow limitation in COPD patients included in these studies also varied, and in some cases specific FEV_1 information from this group of patients was not available. COPD was: *(1)* mild to moderate (FEV_1 66.1 ± 8.6 % ref.) in five studies (21, 25, 27, 28, 30); *(2)* severe-very severe (FEV_1 28.7 ± 7.2 % predicted) in three studies,(15, 24, 26); *(3)* FEV_1 was only presented in liters (not as % predicted) in five studies (16-18, 22, 23); *(4)* one study only specified that seven from 10 COPD patients were GOLD2 (moderate) and three were GOLD3 (severe) (29); *(5)* three studies presented the mean FEV_1 % predicted for all individuals (cases and controls) (14, 19, 20).

COPD cases and controls had similar ages in six studies (21, 23-25, 27, 30), but age profiles were not reported in five studies (14, 16-18, 20, 22), while four studies (15, 19, 21, 28, 29) included younger controls, and in one study the COPD patients were younger (26). *A*ll studies included males and females except Carmona *et al*. who included males only (17). Only four studies included non-Caucasian populations (15, 21, 22, 24).

The studies varied in inclusion of participants according to smoking status: *(1)* seven included never- and ever-smokers (16, 17, 19, 21-23, 28); *(2)* three studies restricted to ever-smokers only (15, 20, 30); *(3)* two studies included former smokers only (24, 27); *(4)* two studies investigated never and current smokers (14, 26); *(5)* one study comprised never-smokers only (18); and *(6)* two studies were conducted in never and former smokers (25, 29). Eleven studies adjusted their results by smoking history: *(1)* four for smoking status and pack years (19, 21-23); *(2)* three for pack years only (14, 15, 24) ; *(3)* one for smoking status only (28); *(4)* one for pack years and years from quitting (27), *(5)* one for pack years and smoking intensity (30); and *(6)* two studies presented the results before and after adjusting for smoking (by status and cumulative exposure) (17, 20) (Table 1).

DNA methylation assessment platform

All studies performed EWAS using microarrays (14-22, 24-30) (Table 1). Three Illumina BeadChip platforms were used: *(1)* three studies used the 'Infinium HumanMethylation27 BeadChip array' (15, 20, 27); *(2)* seven studies used the 'Infinium HumanMethylation450 BeadChip array' (14, 16-

18, 21, 24, 26, 29); *(3)* five used the 'Infinium HumanMethylationEPIC BeadChip array' (19, 23, 25, 28, 30), and (4) 2 used both 'Infinium HumanMethylation450 BeadChip array' and 'Infinium HumanMethylationEPIC BeadChip array' (16, 22).

Cell populations adjustment

Cell populations adjustment is an important source of variability on DNA methylation analyses. In blood, cell-type deconvolution has been widely studied since 2012 (31), but in tissues a DNA methylation atlas enabling the decomposition of different human tissue cell-types has just been recently introduced (32). Deconvoluted cell-type composition was included as a covariate in the analysis of 8 blood studies (14-18, 21-23). In four of them (15, 17, 21, 22) they used cell type deconvolution using Houseman et al. methodology (31), in Bermingham et al (19), and Imboden et al. (16) studies the 'estimateCellCounts' function in minfi which implements Jaffe and Irizarry's(33) modified version of Houseman's algorithm(31) was used, Domingo-relloso et al. (23)used the package FlowSorted.Blood. EPIC based in samples assayed by Brock Christensen and colleagues (34) and Vries et al. (14, 18) used measured cell counts. Qiu et al. (20) is the only blood study that was not adjusted for cell composition. For the studies on respiratory tissues, only the Ström et al. study (30) performed on bronchoalveolar lavage samples was adjusted for measured cell type count, and the Groth et al study (29) on Sputum estimated cell type-specific molecular profiles with a regression by quadratic programming based on sputum differential cell counts.

Risk of bias assessment

For the selection of controls and cases 94% of the studies used an adequate definition for COPD, but 65% presented a potential for selection biases or the representativeness was not stated. Among the COPD-control studies, only 53% of the studies used community controls, while the rest used hospital controls or the selection was not clear. 65% used controls with no history of disease. Furthermore, confounding factors were considered in 88% of studies.

Bias due to baseline confounding or lack of investigation for effect-modification was a major issue as the adjustment of smoking varied across studies. Individual study bias scores are provided in Table S3.

DNA methylation findings

Due to the heterogeneity of study designs and the lack of estimates for some of the associations reported, a formal meta-analysis of EWAS results was not possible. So, to compare their findings, we classified the 17 studies analyzed here by the type of biological sample used (blood or respiratory tissues samples) and by their design (i.e. COPD vs control studies, cross-sectional or longitudinal association with lung function (FEV_1 and/or FEV_1/FVC). For analysis we only used differentially methylated CpG positions (DMPs) found in the discovery cohorts, filtered by p<1 x 10-4 threshold. We did it this way because: *(1)* not all studies performed validation, *(2)* some used as validation published discovery cohorts which are also included in this analysis (Table 1), and *(3)* significance thresholds used in each study were different. The protocol used to select CpGs from each study to be analyzed in this systematic review is described in the online supplement (Table S4). Also, in Table S6 we show the number of significant DMPs and genes, in each study, for each outcome, using different p-value thresholds. Blood studies present a lower number of DMPs and less variability depending on the threshold since these DMPs present high significant p-values. Instead, Respiratory tissue studies present a higher number of significant DMPs but only some of them with high significant p-values.

COPD v control studies of blood DNA methylation

Table S7 list the DMPs and genes described in the discovery cohort of each study. Briefly:

(1) Vries *et al.*(14) in never or current smokers, using a genome-wide p value cut-off (p<1.19 x 10-7) did not identify DMPs associated with COPD but reported that the highest associations were those with cg14972228 (annotated to SIPAL3, at p=5.66×10−6) in never smokers, and with cg08282037 (annotated to EPS8L1, at p=1.45×10−5) in current smokers; *(2)* Bermingham *et al*. (19), who excluded GOLD1 COPD patients, reported 28 DMPs (p < 3.6E−08) from 26 annotated genes significantly associated with FEV_1, FVC, FEV_1/FVC and/or COPD after adjusting for confounders (Table 1, S3); *(3)* Qiu *et al*. (20) not adjusting for confounders identified 3,565 DMPs at FDR < 0.05 in COPD, of them 349 overlapped also for association with FEV_1 and FEV_1/FVC; *(4)* Busch *et al*. (15) found 5 DMPs at FDR< 0.05 and 7 additional DMPs at FDR<0.1 associated with

252 COPD among smoker African-Americans from the PA-SCOPE cohort; *(5)* Lee *et al*. (21) reported
253 no significant differentially methylated probes associated with COPD at p< 1.2 x 10-7; *(6)* in the
254 meta-analysis performed by Lee *et al.*(22) 323 DMPs were identified at pl<0.05 overlapping for
255 COPD and FEV_1 or FEV_1/FVC or FVC; and finally, *(7)* Domingo-Relloso et al. (23) using an elastic
256 net model found 1118 DMPs associated with presence of airflow limitation, 36 of them
257 overlapping with level of with FEV_1 association and 143 DMPs with the FEV_1/FVC association.
258
259 Table S8 lists the 12 DMPs (and 14 genes) with a p value < 1 x 10-4 described in ≥2 blood studies.
260 Of these, 11 DMPs were shared by two studies (Qiu (20) and Busch (15)), both profiled patients
261 with reduced FEV_1 and used the same Illumina platform (27k). Figure 2A and Table S9, show the
262 ontologies associated with these 14 genes, as erythrocyte differentiation and inorganic anion
263 transmembrane transport.
264
265 *Cross- sectional association between blood DNA methylation and lung function*
266 Tables S10 and S11 include all DMPs and genes described in these studies. Briefly: *(1)* Vries *et al.*
267 (18) did not find any DMP associated with FEV_1/FVC in never-smokers at p <1.19 x 10-7; *(2)*
268 Bermingham *et al*. (19) reported 17 DMPs associated with FEV_1, 3 with FEV_1/FVC, and 7 with FVC
269 at p < 3.6E−08, after adjusting for confounders; *(3)* Qiu *et al*. (20) described 4,798 DMPs
270 associated with FEV_1/FVC, and 4,899 associated with FEV_1 at FDR < 0.05, without adjustment for
271 confounders; *(4)* Lee *et al*. (21) found one DMP (cg03559389 (DIP2C)) and nine suggestive
272 associations (p< 1 x 10-5) with FEV_1/FVC, but none for FEV_1 or FVC at p < 1.2 x 10-7; *(5)* Imboden
273 *et al*. (16) identified 111 DMPs associated with either FEV_1, FVC or FEV_1/FVC at p<5 x 10-7; *(6)*
274 Carmona *et al.* (17) found 2 DMPs, cg05575921 (annotated to AHRR) and cg06126421 (annotated
275 to IER3), associated with FEV_1 but these results varied when they adjusted for smoking; *(7)* Lee
276 *et al.* (22) reported 1267 DMPs associated to any of the pulmonary function parameters (FEV_1,
277 FEV_1/FVC, FVC) at FDR< 0.025, and specifically, 1118 DMPs associated with FEV_1 % predicted and
278 finally, *(8)* Domingo-Relloso et al. (23), using an elastic net model, found 838 DMPs associated
279 with FEV_1, 762 DMPs with FVC and 545 DMPs with FEV_1/FVC. In summary, Table S12 shows that
280 51 DMPs and 69 genes were associated with FEV_1 with a p < 1 x 10-4 in ≥2 studies. Figure 2B

shows the associated ontologies to these genes, with the predominant ones being the regulation of the CD4 and T-helper cell activation, immune response, and cytokine production. Table S13 show the full list of these ontologies.

Finally, Table S14 shows that 11 DMPs (and 12 genes) were associated with FEV_1/FVC with a p<1 x 10-4 in ≥2 studies. Comparing Tables S8, S12 and S14, we found that in blood, four genes (F2RL3, AHRR, GFI1, GPR15) were associated with both FEV_1 % predicted and FEV_1/FVC, and one (AHRR) with COPD as well as these spirometric parameters.

Longitudinal studies of Blood DNA methylation and lung function changes over time

Two studies investigated potential associations between changes of lung function over time and DNA methylation: *(1)* Imboden *et al.* (16) identified six DMPs (5 genes) associated with changes in FEV_1/FVC, 20 DMPs (13 genes) with changes in FEV_1 and 4 DMPs (4 genes) with changes in FVC ($p < 5 \times 10^{-7}$); and, *(2)* Carmona *et al.* (17) found 12 DMPs (9 genes) associated FEV_1 or FVC changes (FDR<0.01). The RUNX3 gene was replicated in both studies, but with changes in FEV_1 in the former and changes in FVC in the latter (Table S15).

COPD v control studies of DNA methylation in respiratory tissues

Four COPD vs control studies profiled lung tissue, one sputum, two small airway epithelia (SAE) and one bronchoalveolar lavage (BAL) (Table 1). Table S16 lists the DMPs and genes identified in these four studies:

Lung tissue: *(1)* Morrow *et al*. (24) identified 2,456 DMPs associated with severe to very severe COPD at FDR< 0.05, and in 535 of them (21.7%) with >5% fold difference in methylation between COPD and controls >5%; *(2)* Casas-Recasens *et al*.(25) found 139 DMPs at $p < 1 \times 10^{-8}$ (and 2,007 at FDR< 0.05) associated with COPD. In this study, associations with the severity of airflow limitation were also investigated reporting 36 DMPs at $p < 1 \times 10^{-8}$ in patients with mild to moderate COPD and 99 DMPs at $p < 10^{-8}$ in those with severe COPD; *(3)* Sundar *et al*. (26)

identified 280 DMPs associated with COPD (vs. never-smokers) at raw p< 0.001, and 1,961 DMPs at raw p<0.01.

On the other hand, in sputum, Groth *et al.* (29) found 7,300 DMPs associated with COPD using a Benjamini Hochberg (BH) adjusted p < 0.05 and delta beta ≥ 0.1. In small airway epithelia, Vucic *et al*. (27) and Hernandez *et al*. (28) detected 1,120 genes (1,260 DMPs) and 2,744 DMPs associated to COPD, respectively, both with FDR p <0.05. Finally in BAL, Ström *et al*. (30) found 1,155 DMPs associated with COPD (at a p < 6.74 x 10-8). Table S17 lists the 42 DMPs (and 662 genes) with a p value < 1 x 10-4 described in ≥2 respiratory tissue studies. Figure 2C shows the significant ontology association to those genes: axonogenesis, regulation of cell morphogenesis, regulation of cell adhesion, and of ATPs. Table S18 shows the full list of these ontologies.

Figure 3 compares the results of COPD-control lung studies stratified by tissue type, filtered by a p < 1 x 10-4 and shows that: *(1)* 11 genes (0.21%) were associated with COPD in all lung tissue types; and *(2)* from the 1,410 genes found to be associated with COPD in lung biopsies, 257 genes (18%) were also found associated with COPD in airway epithelia, 129 genes (9.1%) in sputum and 103 genes (7.3%) in BAL.

Cross-sectional associations of DNA methylation with airflow limitation severity in respiratory samples

The association of *DNA methylation* with the severity of airflow limitation (FEV_1 % predicted) was reported in lung tissue of one COPD vs control study (25) (Table S19A), identifying two DMPs (cg01333650 and cg05796331) at p < 1 x 10-8, and in one of small airway epithelia (27), identifying 62 DMPs and genes (Table S19B) at BH-adjusted p < 0.05 with a delta beta of 0.1. No DMPs or genes were found common between both studies once filtered by a p < 1 x 10-4 significance threshold.

Shared DMPs and genes between blood and lung studies

Panel A of Figure 4 compares the results at p < 1 x 10-4 of COPD-control studies both in blood and respiratory tissues, showing that 123 genes (2.6% of total) were commonly shared between at least one blood and one respiratory tissue study (Table S20). Figure 4A shows the ontologies associated to these genes: chronic inflammatory responses, neutrophil activation, apoptosis, among other biological processes (enrichment score is provided in Table S21). Next we explored the evidence of lung and blood expression of these genes using the GTEx platform (35), and found for most (n=98) similar reads per kilobase of transcript per million reads (RPKM) in both tissues (Figure S1) (35). We also explored the transcriptomic data provided by some of the studies and we identified 33 of the 123 genes with reported expression (Table S23).

Finally, Figure 4 Panel B shows that 26 genes (1.7% of total genes reported) were associated with the severity of the airflow limitation both in blood and lung tissue (Table S22), and they were associated to ontologies related to cell cycle and chromosome organization (Table S24).

Relationship between risk of bias and review findings

Looking at CpGs and genes found associated with the different outcomes in the different tissues we saw that: (1) studies done in blood presenting a replication of CpGs and genes associated with FEV1, FEV1/FVC or COPD (15-20, 22, 23), were mainly the ones with a higher score (score >4), except Qiu et al. study (score = 3) (20), on the risk of bias assessment; (2) on the other hand, all studies done in respiratory tissues presented some replication of CpGs and genes associated with COPD, but all of them (24-30) presented medium (score = 4) or low (score<3) scores on the risk of bias assessment.

DISCUSSION

The main results of this systematic review are that: *(1)* there is substantial heterogeneity in study population, design, and analytical criteria across the 17 EWAS studies included, which prevented formal meta-analysis; *(2)* differences in the study populations and in smoking exposure may have confounded the associations of identified DMPs and/or annotated genes with COPD and lung function; *(3) COPD* DMPs replicated in blood are involved in to inorganic ion transport, while those of respiratory samples are involved in axonogenesis and cell morphogenesis; *(4)* COPD

genes shared between ≥1 blood and respiratory tissue studies were associated to chronic inflammation, ion transport and coagulation.

Blood DNA methylation

COPD

Eleven out of twelve COPD associated blood DMPs (92%) found to be replicated, were present in the studies by Qiu *et al.* and Busch *et al.* (15, 20). These two studies profiled a similar group of severe COPD patients recruited from specialized respiratory clinics and used the same methylation platform, likely explaining the very high reproducibility of their findings. Additionally, Bermingham *et al*. (19) and Lee *et al*. (22) reported that a DMP annotated to aryl hydrocarbon receptor repressor gene (AHRR), a well-known smoking related gene, was associated with COPD. Of note, both studies adjusted for smoking status and pack/years, supporting an additional role of AHRR in the pathogenesis of COPD.

Lung function

Fifty-one DMPs (and 69 genes) were reproducibly associated with FEV_1 in ≥2 blood studies, most of them (n=45) described previously (16, 19, 22). Five of them (cg05575921, cg03636183, cg21566642, cg01940273 and cg19859270) were also associated with FEV_1/FVC, two of which (cg05575921 (AHRR) and cg03636183 (F2RL3, Factor II receptor-like 3 gene)) are known smoking associated DMPs (26, 36-39).

Only two studies (16, 17) analyzed lung function changes over time and identified that methylation of the RUNX3 gene was associated with temporal changes in either FEV_1 (16) or FVC (17). RUNX3 is a member of the Runt family related, among other things, to immune regulation, lung development and pulmonary angiogenesis (40). An important conclusion is that studies investigating methylation changes and lung function trajectories over the lifetime are clearly needed (41).

DNA methylation in respiratory samples

Studies profiling respiratory samples identified 662 replicated genes associated with COPD (Table S17). However, only 11 genes (of the total 4,511; 0.2%) were shared across the different types of respiratory samples investigated (lung parenchyma, small airway epithelia, bronchoalveolar lavage fluid, and sputum) (Figure 2). Interestingly, the insulin-like growth factor gene (IGF1R) was one of them. IGF1R is related to regulation of cell growth, adhesion, migration, metabolism and senescence (42), and plays an important role in lung development (43). Epithelial mesenchymal transition (EMT) has been highly implicated in COPD airway pathogenesis (44), which would fit with this. Further, and again in line with this concept, the highest methylation reproducibility across respiratory tissues was found between lung biopsies and small airway epithelia (18%, Figure 2), likely due to the presence of epithelial cells in both.

Shared CpGs and genes in blood and respiratory samples

We identified 123 genes associated with COPD in ≥1 blood *and* ≥1 respiratory sample study (Table S20). Interestingly, these genes were enriched in 20 gene ontologies related to chronic inflammatory responses, coagulation and cardiac muscle cell membrane repolarization. All these pathways have been previously associated with the pathobiology of COPD. On the other hand, we found 26 genes shared between blood and lung in relation to lung function (FEV_1) (Table S22). These genes were significantly enriched by ontologies related to chromosome organization but, also some of them have been associated with smoking, COPD, epithelial injury (NOTCH4) (45), emphysema severity (SERPINF1) (46), chronic inflammation (PIK3R5, CPNE5) and lung development (AMPD3) (47). Collectively, these findings indicate that blood can be used to identify biomarkers of some biological processes occurring in the lung itself.

Strengths and potential limitations

To our knowledge, this is the first systematic review of studies examining associations of DNA methylation profiles and including: *(1)* use of cutting-edge EWAS platforms in different tissue samples, with COPD, lung function and lung function changes over time; and *(2)* a critical appraisal enabling a systematic evaluation of the quality of evidence, which shows that blood studies present a lower risk of bias (higher assessment scores) compared to studies done on

respiratory tissues, and that studies with the lowest risk of bias present a better replication of the results. However, we acknowledge several potential limitations. First, the lack of consistency in the reporting practices, adjustment of confounders and arrays employed in the different studies identified, preclude us from performing a formal meta-analysis of results. We would strongly advocate for some consensus for future epigenetic studies which can eventually facilitate a meta-analysis of available data.

Reporting practices to ensure comparability of future studies

We highlight below a few key aspects, to enable the comparison of results across studies. Additionally, we recommend following the TRIPOD statement for prediction modelling. To ensure comparability and potential for meta-analysis, studies should provide: *(1)* A clear outcome definition, including all relevant scales, thresholding and cut-point details. *(2)* Participant selection criteria and FAIR metadata description, and a data access statement. *(3)* Use a prediction/training and a test/validation population. Avoid systematic differences between these populations (i.e. age or sex, genetic ancestry). *(4)* Report the quality control steps performed, the number of CpG finally analyzed, and the package/software and version used. *(5)* Describe the rationale for model selection, ensuring the specific model prerequisites. *(6)* Describe the covariates included in the model, and the cell decomposition method used. In general adjustment by age, pack/years, smoking status, BMI and cell counts should be explored. It would be desirable to provide the adjusted and unadjusted results. *(7)* For each analysis, provide the CpG, the effect size, the raw p value and the adjusted p value.

Conclusions

This systematic review illustrates the significant lack of harmonization and associated difficulties in comparing the information currently available in the field of epigenetics in relation to COPD and lung function. Yet, the identification of some shared differentially methylated genes between blood-respiratory samples, in the immune response ontologies, suggest that blood could be used to identify circulating biomarkers associated with biologic processes in the lungs. Even with the limitations highlighted, some plausible pathogenic mechanisms do emerge from the analysis.

Additionally, this systematic review identifies methylation changes and lung function changes over time as a clear and critical research gap. Overall, this systematic review provides some clear directions for future research in this field.

ACKNOWLEDGMENTS

Authors thank all funding bodies for their support.

REFERENCES

1. Agusti A, Faner R. COPD beyond smoking: new paradigm, novel opportunities. Lancet Respir Med. 2018;6(5):324-6.
2. Agusti A, Melen E, DeMeo DL, Breyer-Kohansal R, Faner R. Pathogenesis of chronic obstructive pulmonary disease: understanding the contributions of gene-environment interactions across the lifespan. Lancet Respir Med. 2022;10(5):512-24.
3. Faner R, Cruz T, Agusti A. Immune response in chronic obstructive pulmonary disease. Expert Rev Clin Immunol. 2013;9(9):821-33.
4. Vogelmeier CF, Criner GJ, Martinez FJ, Anzueto A, Barnes PJ, Bourbeau J, et al. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Lung Disease 2017 Report. GOLD Executive Summary. Am J Respir Crit Care Med. 2017;195(5):557-82.
5. Lange P, Celli B, Agusti A, Boje Jensen G, Divo M, Faner R, et al. Lung-Function Trajectories Leading to Chronic Obstructive Pulmonary Disease. N Engl J Med. 2015;373(2):111-22.
6. Schamberger AC, Mise N, Meiners S, Eickelberg O. Epigenetic mechanisms in COPD: Implications for pathogenesis and drug discovery. Expert Opinion on Drug Discovery. 2014;9(6):609-28.
7. Wu DD, Song J, Bartel S, Krauss-Etschmann S, Rots MG, Hylkema MN. The potential for targeted rewriting of epigenetic marks in COPD as a new therapeutic approach. Pharmacology and Therapeutics. 2018;182(August 2017):1-14.
8. Zhang L, Valizadeh H, Alipourfard I, Bidares R, Aghebati-Maleki L, Ahmadi M. Epigenetic Modifications and Therapy in Chronic Obstructive Pulmonary Disease (COPD): An Update Review. COPD: Journal of Chronic Obstructive Pulmonary Disease. 2020;17(3):333-42.
9. Li S, Ye Z, Mather KA, Nguyen TL, Dite GS, Armstrong NJ, et al. Early life affects late-life health through determining DNA methylation across the lifespan: A twin study. EBioMedicine. 2022;77:103927.
10. Machin M, Amaral AF, Wielscher M, Rezwan FI, Imboden M, Jarvelin MR, et al. Systematic review of lung function and COPD with peripheral blood DNA methylation in population based studies. BMC Pulm Med. 2017;17(1):54.
11. Benincasa G, DeMeo DL, Glass K, Silverman EK, Napoli C. Epigenetics and pulmonary diseases in the horizon of precision medicine: a review. Eur Respir J. 2021;57(6).
12. Comer BS, Ba M, Singer CA, Gerthoffer WT. Epigenetic targets for novel therapies of lung diseases. Pharmacology and Therapeutics. 2015;147:91-110.
13. https://prisma-statement.org// [Available from: https://prisma-statement.org//.
14. De Vries M, Van Der Plaat DA, Vonk JM, Boezen M. No association between DNA methylation and COPD in never and current smokers Chronic obstructive pulmonary disease. BMJ Open Resp Res. 2018;5:282-.
15. Busch R, Qiu W, Lasky-Su J, Morrow J, Criner G, DeMeo D. Differential DNA methylation marks and gene comethylation of COPD in African-Americans with COPD exacerbations. Respir Res. 2016;17(1):143.
16. Imboden M, Wielscher M, Rezwan FI, Amaral AFS, Schaffner E, Jeong A, et al. Epigenome-wide association study of lung function level and its change. The European respiratory journal. 2019;54(1):35-.
17. Carmona JJ, Barfield RT, Panni T, Nwanaji-Enwerem JC, Just AC, Hutchinson JN, et al. Metastable DNA methylation sites associated with longitudinal lung function decline and aging in humans: an epigenome-wide study in the NAS and KORA cohorts. Epigenetics. 2018;13(10-11):1039-55.
18. de Vries M, Nedeljkovic I, van der Plaat DA, Zhernakova A, Lahousse L, Brusselle GG, et al. DNA methylation is associated with lung function in never smokers. Respir Res. 2019;20(1):268.

19. Bermingham ML, Walker RM, Marioni RE, Morris SW, Rawlik K, Zeng Y, et al. Identification of novel differentially methylated sites with potential as clinical predictors of impaired respiratory function and COPD. EBioMedicine. 2019;43:576-86.
20. Qiu W, Baccarelli A, Carey VJ, Boutaoui N, Bacherman H, Klanderman B, et al. Variable DNA methylation is associated with chronic obstructive pulmonary disease and lung function. Am J Respir Crit Care Med. 2012;185(4):373-81.
21. Lee MK, Hong Y, Kim SY, Kim WJ, London SJ. Epigenome-wide association study of chronic obstructive pulmonary disease and lung function in Koreans. Epigenomics. 2017;9(7):971-84.
22. Lee M, Huan T, McCartney DL, Chittoor G, de Vries M, Lahousse L, et al. Pulmonary Function and Blood DNA Methylation: A Multi-Ancestry Epigenome-Wide Association Meta-Analysis. Am J Respir Crit Care Med. 2022.
23. Domingo-Relloso A, Riffo-Campos AL, Powers M, Tellez-Plaza M, Haack K, Brown RH, et al. An epigenome-wide study of DNA methylation profiles and lung function among American Indians in the Strong Heart Study. Clin Epigenetics. 2022;14(1):75.
24. Morrow JD, Cho MH, Hersh CP, Pinto-Plata V, Celli B, Marchetti N, et al. DNA methylation profiling in human lung tissue identifies genes associated with COPD. Epigenetics. 2016:1-10.
25. Casas-Recasens S, Noell G, Mendoza N, Lopez-Giraldo A, Garcia T, Guirao A, et al. Lung DNA Methylation in Chronic Obstructive Pulmonary Disease: Relationship with Smoking Status and Airflow Limitation Severity. Am J Respir Crit Care Med. 2021;203(1):129-34.
26. Sundar IK, Yin Q, Baier BS, Yan L, Mazur W, Li D, et al. DNA methylation profiling in peripheral lung tissues of smokers and patients with COPD. Clin Epigenetics. 2017;9:38.
27. Vucic EA, Chari R, Thu KL, Wilson IM, Cotton AM, Kennett JY, 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(5):912-22.
28. Hernández Cordero AI, Yang CX, Yang J, Horvath S, Shaipanich T, MacIsaac J, et al. Airway Aging and Methylation Disruptions in HIV-associated Chronic Obstructive Pulmonary Disease. Am J Respir Crit Care Med. 2022.
29. Groth EE, Weber M, Bahmer T, Pedersen F, Kirsten A, Bornigen D, et al. Exploration of the sputum methylome and omics deconvolution by quadratic programming in molecular profiling of asthma and COPD: the road to sputum omics 2.0. Respir Res. 2020;21(1):274.
30. Eriksson Ström J, Kebede Merid S, Pourazar J, Blomberg A, Lindberg A, Ringh MV, et al. COPD is Associated with Epigenome-wide Differential Methylation in BAL Lung Cells. Am J Respir Cell Mol Biol. 2022.
31. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics. 2012;13:86.
32. Zhu T, Liu J, Beck S, Pan S, Capper D, Lechner M, et al. A pan-tissue DNA methylation atlas enables in silico decomposition of human tissue methylomes at cell-type resolution. Nat Methods. 2022;19(3):296-306.
33. Jaffe AE, Irizarry RA. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. Genome Biol. 2014;15(2):R31.
34. Salas LA, Koestler DC, Butler RA, Hansen HM, Wiencke JK, Kelsey KT, et al. An optimized library for reference-based deconvolution of whole-blood biospecimens assayed using the Illumina HumanMethylationEPIC BeadArray. Genome Biol. 2018;19(1):64.
35. Consortium G. The Genotype-Tissue Expression (GTEx) project. Nat Genet. 2013;45(6):580-5.
36. Obeidat M, Dvorkin-Gheva A, Li X, Bosse Y, Brandsma CA, Nickle DC, et al. The Overlap of Lung Tissue Transcriptome of Smoke Exposed Mice with Human Smoking and COPD. Sci Rep. 2018;8(1):11881.

552 37. Grieshober L, Graw S, Barnett MJ, Thornquist MD, Goodman GE, Chen C, et al. AHRR
553 methylation in heavy smokers: associations with smoking, lung cancer risk, and lung cancer mortality.
554 BMC Cancer. 2020;20(1):905.
555 38. Tantoh DM, Wu MC, Chuang CC, Chen PH, Tyan YS, Nfor ON, et al. AHRR cg05575921
556 methylation in relation to smoking and PM2.5 exposure among Taiwanese men and women. Clin
557 Epigenetics. 2020;12(1):117.
558 39. Breitling LP, Yang R, Korn B, Burwinkel B, Brenner H. Tobacco-smoking-related differential DNA
559 methylation: 27K discovery and replication. Am J Hum Genet. 2011;88(4):450-7.
560 40. Lee SH, Bae SC, Kim KW, Lee YM. RUNX3 inhibits hypoxia-inducible factor-1alpha protein
561 stability by interacting with prolyl hydroxylases in gastric cancer cells. Oncogene. 2014;33(11):1458-67.
562 41. Agustí A, Melén E, DeMeo DL, Breyer-Kohansal R, Faner R. Pathogenesis of chronic obstructive
563 pulmonary disease: understanding the contributions of gene-environment interactions across the
564 lifespan. Lancet Respir Med. 2022;10(5):512-24.
565 42. López IP, Piñeiro-Hermida S, Pais RS, Torrens R, Hoeflich A, Pichel JG. Involvement of Igf1r in
566 Bronchiolar Epithelial Regeneration: Role during Repair Kinetics after Selective Club Cell Ablation. PLoS
567 One. 2016;11(11):e0166388.
568 43. He H, Snowball J, Sun F, Na CL, Whitsett JA. IGF1R controls mechanosignaling in myofibroblasts
569 required for pulmonary alveologenesis. JCI Insight. 2021;6(6).
570 44. Walters EH, Shukla SD, Mahmood MQ, Ward C. Fully integrating pathophysiological insights in
571 COPD: an updated working disease model to broaden therapeutic vision. Eur Respir Rev. 2021;30(160).
572 45. Tilley AE, Harvey BG, Heguy A, Hackett NR, Wang R, O'Connor TP, et al. Down-regulation of the
573 notch pathway in human airway epithelium in association with smoking and chronic obstructive
574 pulmonary disease. Am J Respir Crit Care Med. 2009;179(6):457-66.
575 46. Kawanami Y, Morimoto Y, Kim H, Nakamura T, Machida K, Kido T, et al. Calcitonin gene-related
576 peptide stimulates proliferation of alveolar epithelial cells. Respir Res. 2009;10:8.
577 47. Mukherjee N, Arathimos R, Chen S, Kheirkhah Rahimabad P, Han L, Zhang H, et al. DNA
578 methylation at birth is associated with lung function development till age 26 years. European
579 Respiratory Journal. 2020:2003505-.

580

581

582 ***TABLE 1.*** Description of the articles included in this systematic review: tissue, platform, design, significant p value threshold, smoking

583 status, population, outcomes and summary of results.

584

Author (year), PMID	Tissue	Methylation [illegible]	Discovery P-value [illegible]	Discovery Cohort	Validation Cohort	All individuals	Controls	COPD		CpGs descrived at pval<10-4						Genes descrived at pval<10-4					
						N, mean (SD) age, sex, ethnicity	N, mean (SD) age, sex, ethnicity	N, mean (SD) age, sex, ethnicity	Mean (SD) FEV_1% ref. or FEV_1(l)	Case-control	FEV1	FEV1/FVC	LF change (FEV1)	LF change (FEV1/FVC)	LF change (FVC)	Case-control	FEV1	FEV1/FVC	LF change (FEV1)	LF change (FEV1/FVC)	LF change (FVC)
de Vries (2018), 30018765	Blood	Illumina 450K	p<1.19×10−7	Lifelines		n= 903 never-smokers, 46(18-80) years, 43.7% ♀	n= 587	n= 316	92.58% (14.3)	2						2					
						n= 658 current-smokers, 46(22-79) years, 43% ♀	n= 379	n= 279	85.82% (14.0)												
Busch (2016), [illegible]	Blood	Illumina 27K	FDR<0.05 and FDR<0.1	PA-SCOPE, IGCN for [illegible]	IGCN	n= 362	n= 269, 48.9(6.4) years 27.5%♀; African-Americans	n= 93, 61.4(7.7) years 46.2% ♀; African-Americans	36.8% (15.4)	12						12					
Imboden (2019), 31073081	Blood	Illumina 450k and EPIC	p<5×10−7	Aging Lungs in European [illegible]	KORA, LifeLines, NSPHS, [illegible]	n= 962 (SAPALDIA), 58.8(11.3) years, 53.5% ♀	n= 678	n= 284	3.0L (0.8)*		62	29	20	6	4		41	17	13	5	4
						n= 470 (ECRHS), 54.5(6.8) years, 56% ♀	n= 394	n= 76	3.0L (0.8)*												

						n= 611 (NFBC1966), 46.3(0.4) years, 55.3% ♀	n= 545	n= 66	3.5L (0.7)*												
Carmona	Blood	Illumina 450K	Holm level	NAS	KORA	n= 633, 72 years, 0% ♀	n= 410, 47 Asthma	n= 176	2.54L (0.58)*		6		4		8		2		2		7
de Vries (2019), 31791327	Blood	Illumina 450K	p<1.26 × 10^− 7 = FDR<0.05; p<0.0001	LL COPD&C, LLDEEP, RS-III-1, RS-BIOS		n= 903 (LL COPD&C), 46(18-80) years, 43.7% ♀	n= 587	n= 316	3.5(0.9)L*			30						25			
						n= 166 (LLDEEP), 42(20-78) years, 57.2% ♀	n= 151	n= 15	3.6(0.9)L*												
						n= 150 (RS-III-1), 63(53-93) years, 50.7% ♀	n= 137	n= 13	3.2(0.8)L*												
						n= 206 (RS-BIOS), 68(52-79) years, 61.2% ♀	n= 187	n= 19	2.7(0.7)L*												
Bermingham (2019),	Blood	Illumina EPIC	p<3.6 × 10−8	Generation Scotland:	LBC	n= 3193	n= 2919, 46.98(13.37) years 60.2% ♀;	n= 274, 53.96(13.59) years 66.1% ♀	66.96% (11.44)†	5	24	3				4	22	3			
Qiu (2011),	Blood	Illumina 27K	FDR<0.05	IGCN	EOCOPD	n= 945, 57.3(8.1) years, 45.6% ♀	n= 325	n= 620	64.2% (32.7)*	347	349	349				319	321	321			

Lee (2017), 28621160	Blood	Illumina 450K	p < 1.2 × 10-7 = FDR<0.05 for	Korean cohort		n= 100	n= 40, 71.8(6.9) years 40%♀	n= 60, 73.5(5.8) years 30%♀. Asian	75% (18)	0	4	10				0	4	10			
Lee. (2022), 35536696	Blood	Illumina 450k and EPIC	FDR<0.025	ALHS, ARIC, CHS, FHS, GS, LifeLines, LBC, MESA, RS, and TwinsUK, within the		n= 2268 (ALHS), 63 years; 49%; European	n= 1332	n= 378	2.549L*	13	1118	117				13	930	97			
						n= 787 (ARIC), 59 years; 61%♀ ; European	n= 561	n= 90	2.723L*												
						n= 2261 (ARIC), 56 years; 63%♀; African	n= 1716	n= 156	2.427L*												
						n= 218 (CHS) , 76 years; 63%♀ ; European	n= 100	n= 43	2.004L*												
						n= 181 (CHS), 73 years; 65%♀ ; African	n= 95	n= 39	1.743L*												
						n= 3205(FHS), 59 years; 54%♀ ; European	n= 2227	n= 275	2.882L*												

						n= 1700(GS), 55 years; 60%♀; European	n= 1270	n= 147	2.874L*															
						n= 2954 (GS), 56 years; 54%♀; European	n= 2193	n= 264	2.883L*															
						n= 435 (LBC1921), 79 years; 60%♀; European	n= 256	n= 57	1.874L*															
						n= 905 (LBC1936), 70 years; 50%♀; European	n= 609	n= 91	2.36L*															
						n= 1155 (Lifelines), 51 years; 44%♀; European	n= 622	n= 138	3,292*															
						n= 246 (MESA), 70 years; 47%♀; European	n= 144	n= 33	2.511L*															
						n= 107 (MESA), 69 years; 58%♀; African	n= 72	n= 15	2.046L*															
						n= 193 (MESA), 67 years; 53%♀; Hispanic	n= 135	n= 22	2.338L*															

						n= 716 (RS), 68 years; 56%♀; European	n= 564	n= 54	2.657L*												
						n= 172 (TwinsUK), 59 years; 100%♀; European	n= 150	n= 7	2.488L*												
Domingo-Relloso (2022), 35681244	Blood	Illumina EPIC	Elastic net analysis	SHS		n= 1677), 55 years; 61%♀; American Indians	n= 1096, 53 years; 62%♀; American Indians	n= 352, 59 years; 51%♀; American Indians	2.1L	57	84	81				36	59	52			
								n= 229 (restriction), 56 years; 70%♀; American Indians	1.9L												
Morrow (2016),27564456	Lung	Illumina 450K	FDR<0.05 and bdiff>5%			n= 160	n= 46, 65.3(8.9) years 63%♀; African-American 13% and white 84.8%	n= 114, 63.4(6.7) years 52.6%♀; African-American 16.7% and white 80.7%	26.3% (9.1)	219						150					
Casas-Recasens [illegible]	Lung	Illumina EPIC	P value<10-8 or FDR 0.05			n= 150	n= 26, 66.3(9.55) years 57.7%♀;	1n= 24, 65.8(7.58) years 16.1%♀	56.1% (23.8)	2007	372					1148	284				

Sundar (2017), 28416970	Lung	Illumina 450K	pval<0.001, and pval<0.01			n= 24	8 never-smokers 62.62(1.87) years 25%♀ 8 current-smokers 61(2.33) years 12.5%♀;	n= 8, 56.1(2.57) years 37.5%♀.	23% (4.41)	289						169					
Groth (2020),	Sputum	Illumina 450K	B-H P value <0.05	ALLIANCE(asyhma and		n= 17	n= 10‡, 44(20) years 33.3%♀;	n= 10, 68(10) years 10%♀.	7 GOLD 2 and 3 GOLD 3	118 9						709					
Vucic (2013),	Bronchial small	Illumina 27K	B-H P value			n= 38	n= 23, 64(4.8) years 34.7%♀;	n= 15, 65(5.76) years 33.3%♀	58% (15.6)	936	0					851	0				
Hernández	Bronchial	Illumina EPIC	FDR<0.05 and			n= 42	n= 20, 64(56-68) years 60%♀;	n= 22, 69(64-73) years 45%♀	73% (59.88-78.53)	206 4						315					
Eriksson	Bronchoalve	Illumina EPIC	P < 6.74E-8	OLIN		n= 33	n= 15, 65(63-72.5) years 22.2%♀;	n= 18, 63(62-72) years 53.3%♀	68.3 (44.1-72.3)	115 5						174					

*mean FEV1 is from all individuals, not only COPD

† GOLD1 patientes were excluded

‡ 3 control samples lost

585

587

FIGURE LEGENDS

Figure 1: PRISMA 2009, flow-diagram illustrating the study selection process; i.e. number of articles that were retrieved in each step, and the number of excluded articles.

Figure 2: Panel A: REVIGO diagram with a visual summary of the enriched ontologies from the genes found in ≥2 COPD association blood studies. Panel B: REVIGO diagram with a visual summary of the enriched ontologies from the genes found ≥2 FEV1 association blood studies. Panel C: REVIGO diagram with a visual summary of the enriched ontologies from the the ≥2 COPD association studies in respiratory tissues.

Figure 3: Venn diagram showing the genes differentially methylated in the COPD analysis of airways by profiled tissue (lung, sputum, bronchoalveolar lavage, small airway epithelia). In each tissue type, all the identified genes have been pooled and used for the comparison. The % of the common genes are presented and the names of the 11 genes found common in all respiratory tissue studies are listed.

Figure 4: Panel A: Venn diagram showing the genes differentially methylated in the COPD analysis comparing blood versus airways (lung, sputum, bronchoalveolar lavage, small airway epithelia). REVIGO diagram with a visual summary of the ontologies found enriched in the common genes.

Panel B: Venn diagram showing the genes differentially methylated in the FEV1 analysis comparing blood versus lung tissue. REVIGO diagram with a visual summary of the ontologies found enriched in the common genes.

PRISMA 2009 flow diagram

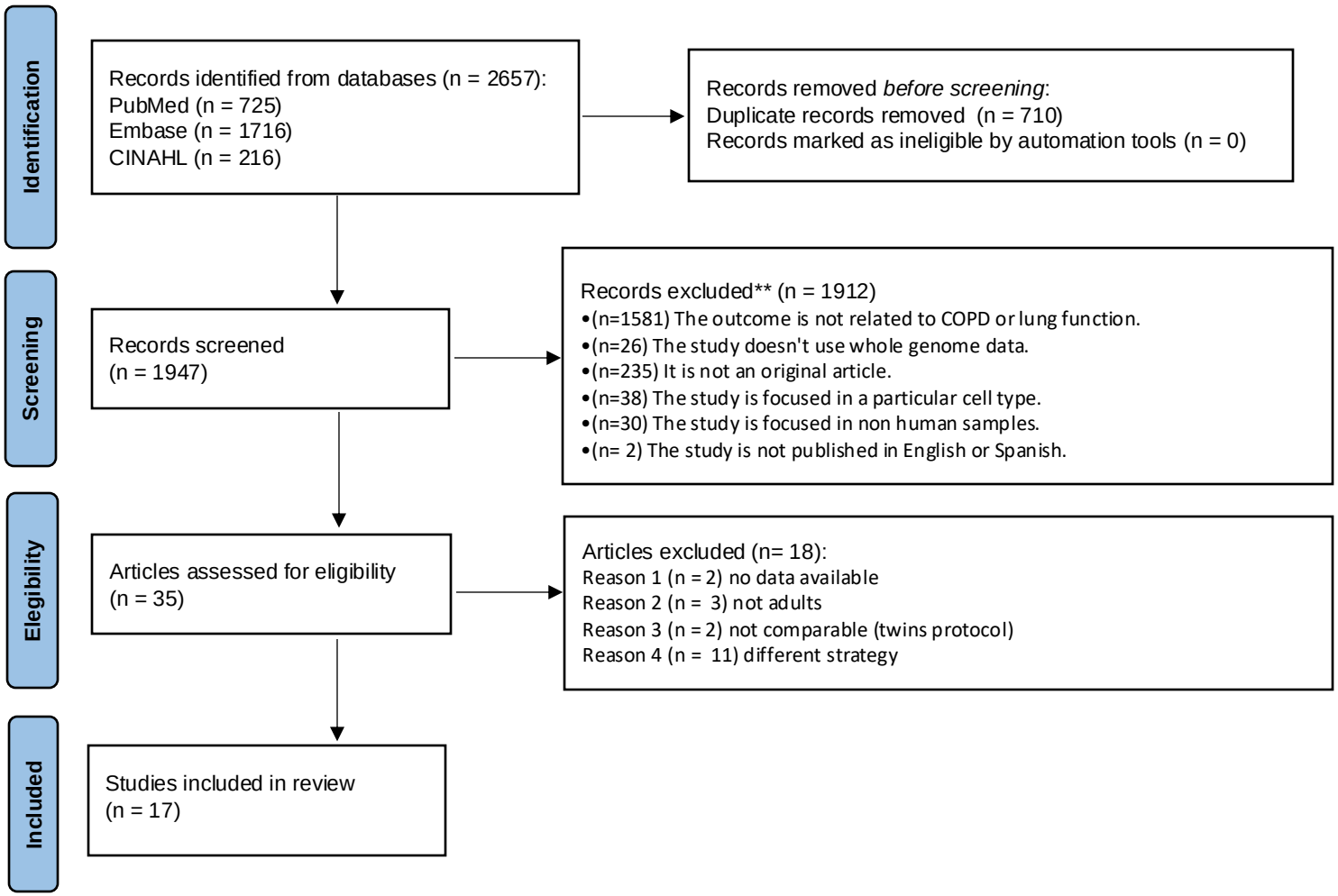
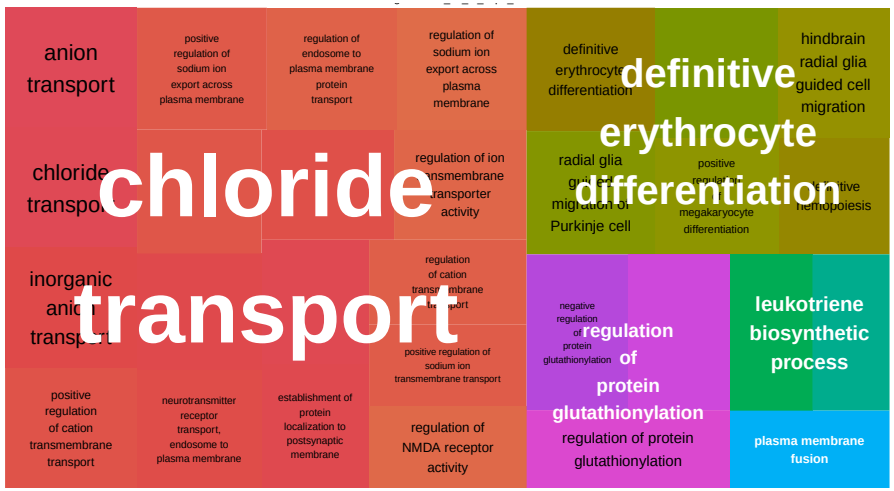
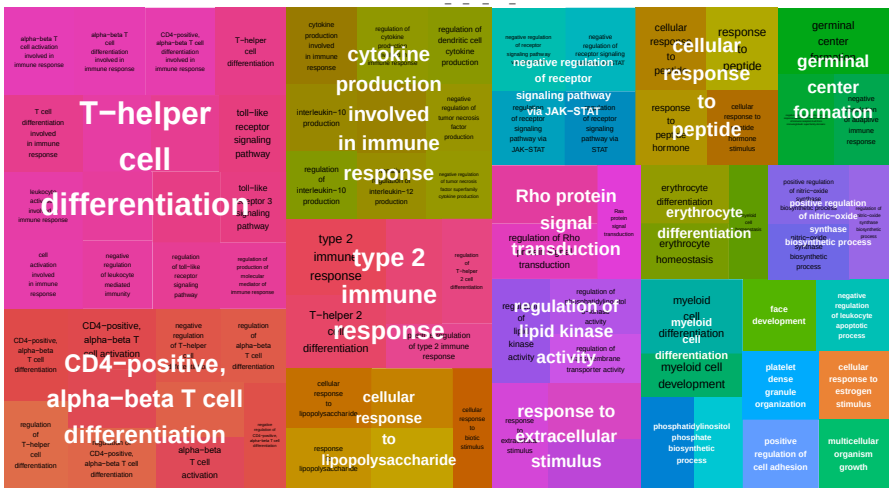


Figure 1

A) Genes common in ≥2 BLOOD **COPD** association studies



B) Genes common in ≥2 BLOOD **FEV1** association studies



C) Genes common in ≥2 Respiratory tissues **COPD** association studies

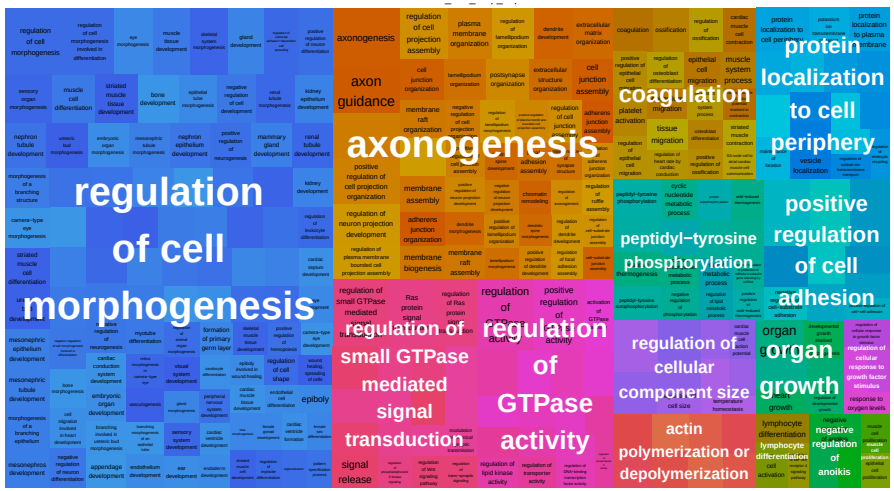


Figure 2

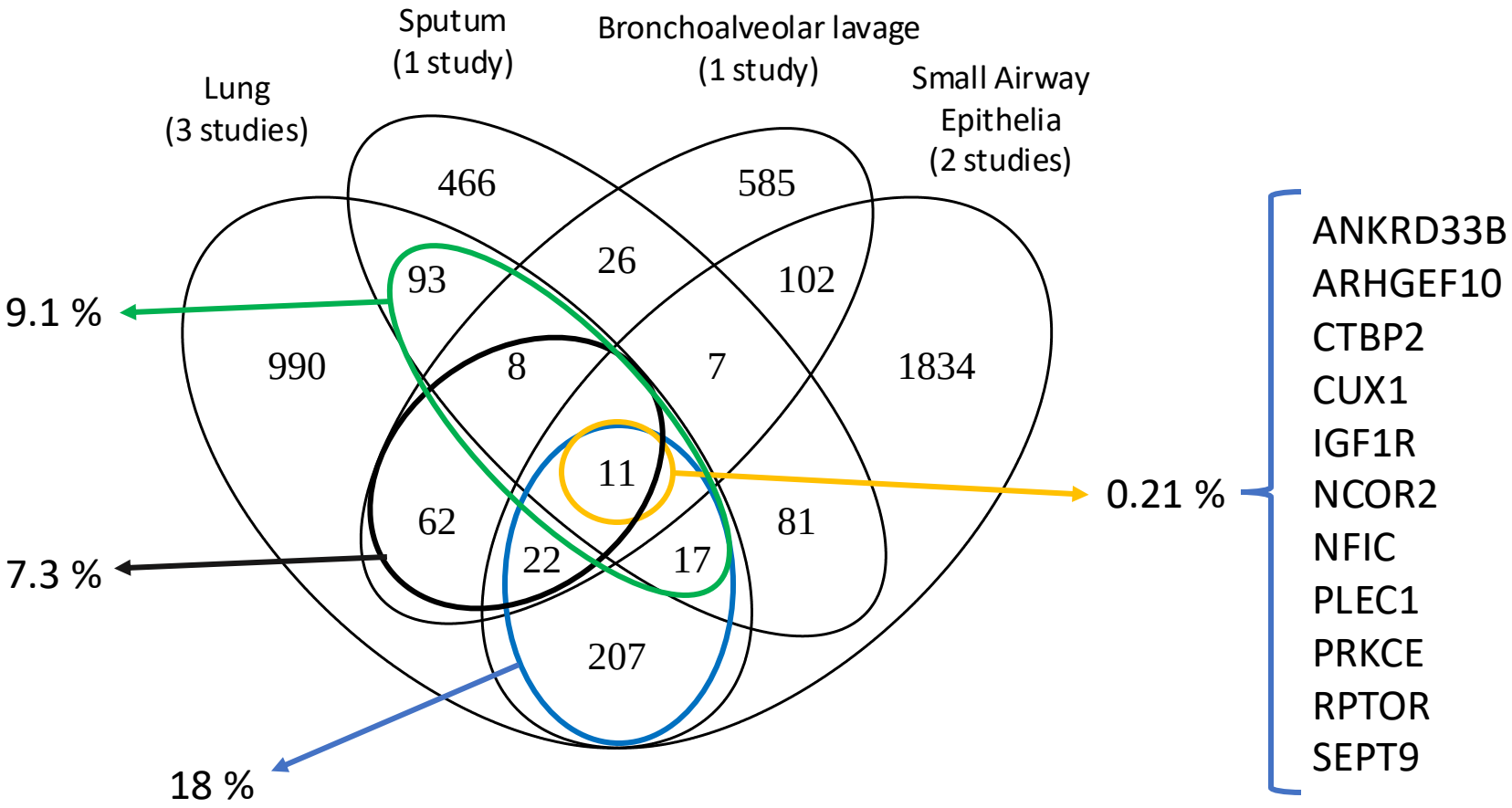

Lung
(3 studies)
Sputum
(1 study)
Bronchoalveolar lavage
(1 study)
Small Airway
Epithelia
(2 studies)
466
585
93
26
102
9.1 %
990
8
7
1834
11
0.21 %
62
81
7.3 %
22
17
207
18 %
ANKRD33B
ARHGEF10
CTBP2
CUX1
IGF1R
NCOR2
NFIC
PLEC1
PRKCE
RPTOR
SEPT9

Figure 3

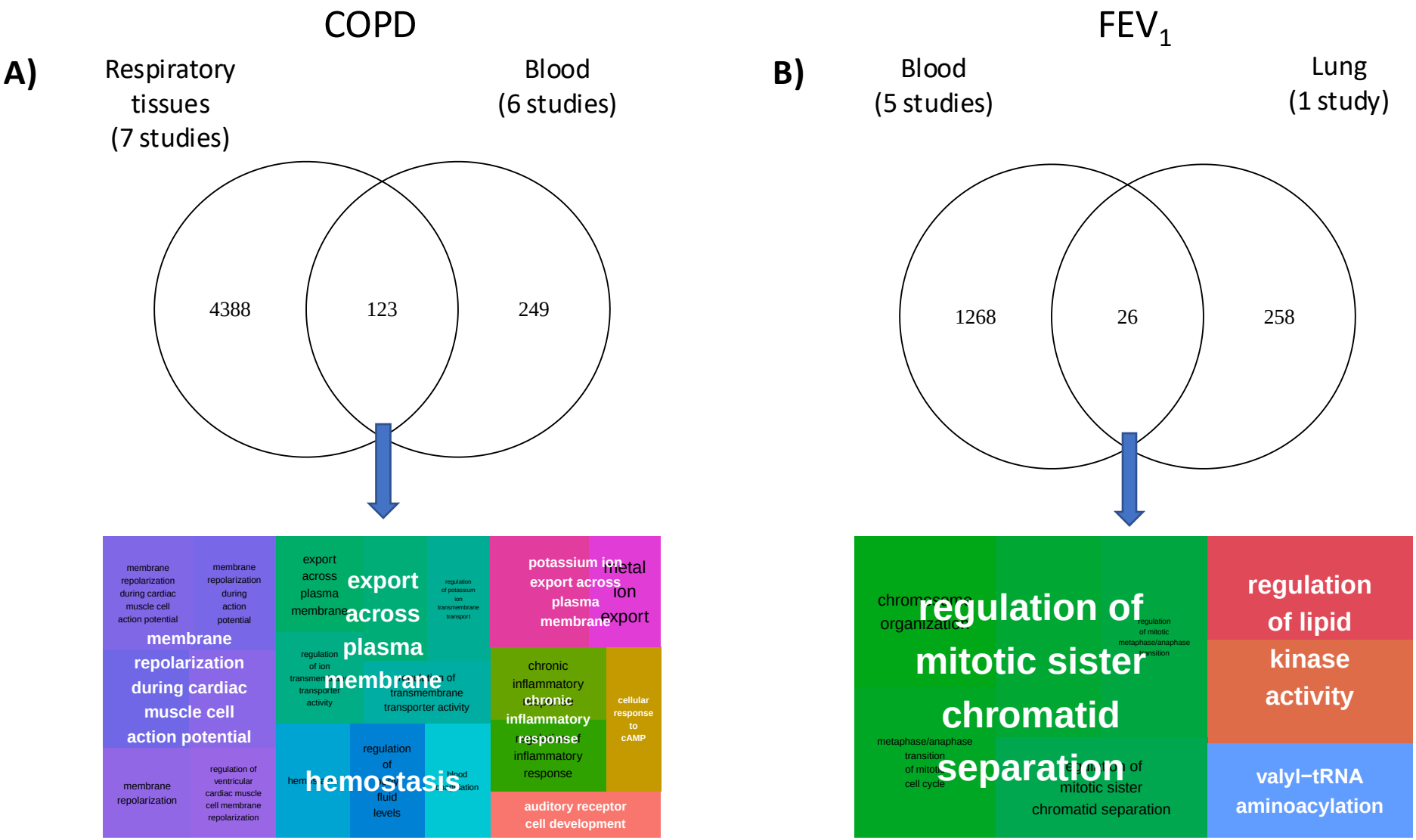

COPD
A)
Respiratory tissues (7 studies)
Blood (6 studies)
4388
123
249
membrane repolarization during cardiac muscle cell action potential
membrane repolarization during action potential
membrane repolarization
membrane repolarization during cardiac muscle cell action potential
export across plasma membrane
export across plasma membrane
regulation of potassium ion transmembrane transport
potassium ion export across plasma membrane
metal ion export
regulation of ion transmembrane transporter activity
regulation of transmembrane transporter activity
chronic inflammatory response
chronic inflammatory response
cellular response to cAMP
regulation of inflammatory response
membrane repolarization
regulation of ventricular cardiac muscle cell membrane repolarization
hemostasis
regulation of body fluid levels
blood coagulation
auditory receptor cell development
FEV1
B)
Blood (5 studies)
Lung (1 study)
1268
26
258
chromosome organization
regulation of mitotic metaphase/anaphase transition
regulation of mitotic sister chromatid separation
metaphase/anaphase transition of mitotic cell cycle
regulation of mitotic sister chromatid separation
regulation of lipid kinase activity
valyl−tRNA aminoacylation

Figure 4

ONLINE SUPPLEMENT

EPIGENOME-WIDE ASSOCIATION STUDIES OF COPD AND LUNG FUNCTION: A SYSTEMATIC REVIEW

Sandra Casas-Recasens[1,2]*, Raisa Cassim[3]*, Núria Mendoza[1,2], Alvar Agusti[1,2,4,5], Caroline Lodge[3], Shuai Li[6,7,8,9], Dinh Bui[3], David Martino[10,,11], Shyamali C Dharmage[3‡], Rosa Faner[1,2,5.12‡]

SUPPLEMENTARY METHODS

Eligibility Criteria

For inclusion in this systematic review, publications must have investigated the relationship of DNA methylation with COPD and/or, lung function using EWAS. The outcomes of interest were: *(1)* presence of COPD, as defined by either GOLD (FEV1/FVC<0.7) or by FEV_1/FVC lower than the lower limit of normal (LLN) of the corresponding reference values; and *(2)* lung function values (FEV1 % predicted, FVC % predicted or FEV1/FVC (%)) or their decline over time.

Selection of included articles

Titles and abstracts of articles were screened by two reviewers and only articles assessing the association of DNA methylation with outcomes of interest were reviewed. The reference list of all studies reviewed was screened to identify further potentially relevant studies. Articles without an abstract in English were excluded. Papers with no statistical measure of the relationship between methylation and outcomes were excluded, as were studies that did not

measure DNA methylation genome wide. Further, reviews, comments, case-series and conference abstracts were all excluded too.

Critical appraisal

We performed a critical appraisal of all the studies using a modified version of The Newcastle-Ottawa Scale (NOS) (1). This enabled an unbiased systematic evaluation of the quality of evidence from each study (Table S2).

Data extraction

Identification data (author, date, title and PubMed ID), study design details (study setting, study and control groups, population demographics), methods used (tissue type, method of measurement, statistical analysis, correction methods), results (β coefficients, P values) were extracted into a standardized template (Table S3). To compare studies, we: *(1)* extracted the lists of DMPs and p values (raw p values whenever is possible (Tables S3 and S4) (with the less restrictive adjustment when adjustments were provided) from the discovery cohort of each study; *(2)* considered only the DMPs with a raw $p<10^{-4}$ (where available); *(3)* re-annotated all the DMPs/CpG from all studies to genes using the Illumina database (EPIC Genome build 37 (HG19)); and *(4)* identified the DMPs or genes reported in ≥ 2 studies with the same outcome and/or tissue.

Enrichment analysis

To identify over-represented pathways in the genes annotated to DMPs of interest, we used hypergeometric tests in the R Bioconductor package GeneAnswers (2) or clusterProfiler (3). Significant enrichment was defined as FDR< 0.05, with five or more genes associated to the

term. To evaluate the term overlap and obtain a visual summary we used R Bioconductor package rrvgo (4).

GTEx expression data

To check the expression of the genes annotated to DMPs of interest, we used The Genotype-Tissue Expression (GTEx) which was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. Figures were obtained from: the GTEx Portal on 08/14/23 (https://www.gtexportal.org/home/multiGeneQueryPage).

SUPPLEMENTARY TABLES

Table S1: Pubmed Keyword search for the systematic review.

Table S2: Comments, and scores for the Critical appraisal.

Table S3: Additional details of the selected studies.

Table S4: Methodology of DMP extraction for the analisis in the systematic review.

Table S5: A) DMPs in blood COPD-control studies. B) Genes in blood COPD-control studies.

Table S6: DMPs and genes differentially methylated ≥ 2 different COPD-control studies in blood.

Table S7: A) DMPs in blood FEV1 studies. B) Genes in blood FEV1 studies.

Table S8: A) DMPs in blood FEV1/FVC studies. B) Genes in blood FEV1/FVC studies.

Table S9: DMPs and genes differentially methylated ≥ 2 different FEV1 association studies in blood.

Table S10: DMPs and genes differentially methylated ≥ 2 different FEV1/FVC association studies in blood.

Table S11: A) DMPs in blood change in lung function studies. B) Genes in blood change in lung function studies.

Table S12: A) DMPs in respiratory tissues COPD-control studies. B) Genes in respiratory tissues COPD-control studies.

Table S13: A) DMPs at p <10E-4 in ≥2 respiratory tissue COPD-control studies. B) Genes at p <10E-4 in ≥2 respiratory tissue COPD-control studies.

Table S14: A) DMPs at p <10E-4 in respiratory tissue FEV1 studies. B) Genes at p <10E-4 in respiratory tissue FEV1 studies.

Table S15: Genes associated with COPD commonly shared between at least one blood and one respiratory tissue study.

Table S16: Biological ontologies enriched on the 108 genes found common in the COPD-control analysis between blood and respiratory tissues studies.

Table S17: Genes associated with FEV1 commonly shared between at least one blood and one lung tissue study.

SUPPLEMENTARY FIGURES

Figure S1: Heatmap showing the list of genes from the 108 genes found common in the COPD-control analysis between blood and respiratory tissues studies, that present evidences of lung and blood expression on GTEx platform (33). Colours indicate the amount reads per kilobase of transcript per million reads (TPM).

REFERENCES

1. *Wells G, Shea, B*, *O'Connell D*, *Peterson J*, *Welch V*, *Losos M*, *Tugwell P*. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses [Available from: https://www.ohri.ca/programs/clinical_epidemiology/oxford.asp.
2. Feng G, Du P, Krett NL, Tessel M, Rosen S, Kibbe WA, et al. A collection of bioconductor methods to visualize gene-list annotations. BMC Res Notes. 2010;3:10.
3. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012;16(5):284-7.
4. Sayols S. rrvgo: a Bioconductor package to reduce and visualize Gene Ontology terms. 2020.

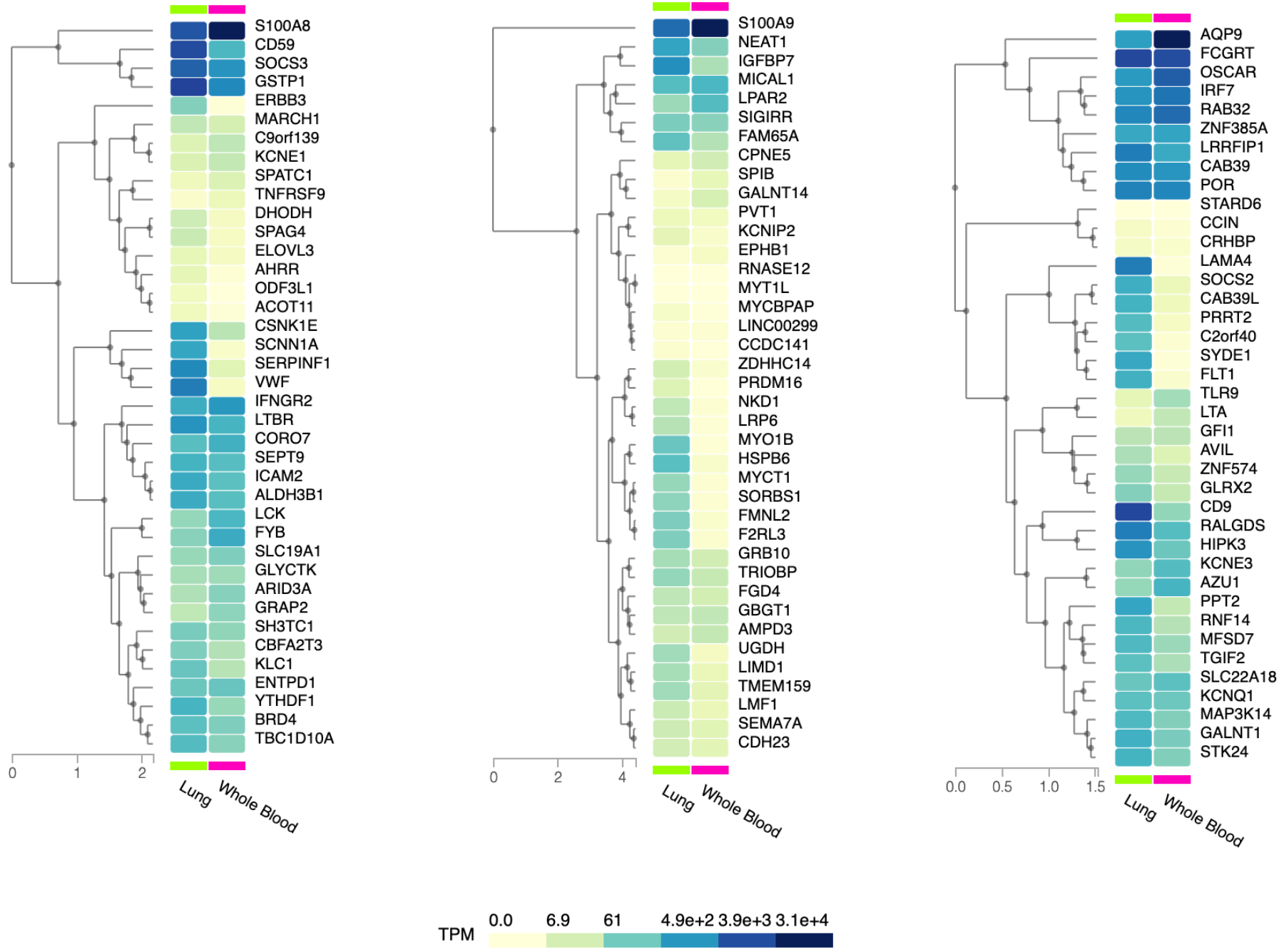

S100A8
CD59
SOCS3
GSTP1
ERBB3
MARCH1
C9orf139
KCNE1
SPATC1
TNFRSF9
DHODH
SPAG4
ELOVL3
AHRR
ODF3L1
ACOT11
CSNK1E
SCNN1A
SERPINF1
VWF
IFNGR2
LTBR
CORO7
SEPT9
ICAM2
ALDH3B1
LCK
FYB
SLC19A1
GLYCTK
ARID3A
GRAP2
SH3TC1
CBFA2T3
KLC1
ENTPD1
YTHDF1
BRD4
TBC1D10A
0
1
2
Lung
Whole Blood
S100A9
NEAT1
IGFBP7
MICAL1
LPAR2
SIGIRR
FAM65A
CPNE5
SPIB
GALNT14
PVT1
KCNIP2
EPHB1
RNASE12
MYT1L
MYCBPAP
LINC00299
CCDC141
ZDHHC14
PRDM16
NKD1
LRP6
MYO1B
HSPB6
MYCT1
SORBS1
FMNL2
F2RL3
GRB10
TRIOBP
FGD4
GBGT1
AMPD3
UGDH
LIMD1
TMEM159
LMF1
SEMA7A
CDH23
0
2
4
Lung
Whole Blood
AQP9
FCGRT
OSCAR
IRF7
RAB32
ZNF385A
LRRFIP1
CAB39
POR
STARD6
CCIN
CRHBP
LAMA4
SOCS2
CAB39L
PRRT2
C2orf40
SYDE1
FLT1
TLR9
LTA
GFI1
AVIL
ZNF574
GLRX2
CD9
RALGDS
HIPK3
KCNE3
AZU1
PPT2
RNF14
MFSD7
TGIF2
SLC22A18
KCNQ1
MAP3K14
GALNT1
STK24
0.0
0.5
1.0
1.5
Lung
Whole Blood
TPM
0.0
6.9
61
4.9e+2
3.9e+3
3.1e+4

Figure S1