



# The DNA repair transcriptome in severe COPD

Maor Sauler <sup>1</sup>, Maxime Lamontagne<sup>2</sup>, Eric Finnemore<sup>1</sup>, Jose D. Herazo-Maya<sup>1</sup>, John Tedrow<sup>3</sup>, Xuchen Zhang<sup>4</sup>, Julia E. Morneau<sup>1</sup>, Frank Sciarba<sup>3</sup>, Wim Timens<sup>5</sup>, Peter D. Paré<sup>6</sup>, Patty J. Lee<sup>1</sup>, Naftali Kaminski<sup>1</sup>, Yohan Bossé <sup>2,7</sup> and Jose L. Gomez <sup>1</sup>

**Affiliations:** <sup>1</sup>Dept of Medicine, Yale School of Medicine, New Haven, CT, USA. <sup>2</sup>Centre de Recherche Institut Universitaire de Cardiologie et de Pneumologie de Québec, Laval University, Quebec, QC, Canada. <sup>3</sup>Dept of Medicine, University of Pittsburgh, Pittsburgh, PA, USA. <sup>4</sup>Dept of Pathology, Yale School of Medicine, New Haven, CT, USA. <sup>5</sup>Dept of Pathology and Medical Biology, University Medical Center Groningen, GRIAC Research Institute, University of Groningen, Groningen, The Netherlands. <sup>6</sup>The University of British Columbia Centre for Heart Lung Innovation, St. Paul's Hospital, Vancouver, BC, Canada. <sup>7</sup>Dept of Molecular Medicine, Laval University, Quebec, QC, Canada.

**Correspondence:** Maor Sauler, The Anlyan Center, 300 Cedar Street, Ste South 455C, New Haven, CT 06519, USA. E-mail: maor.sauler@yale.edu



@ERSpublications

**Severe COPD is associated with reduced transcription of genes involved in the nucleotide excision repair pathway** <http://ow.ly/TNoa3019j2y>

**Cite this article as:** Sauler M, Lamontagne M, Finnemore E, *et al.* The DNA repair transcriptome in severe COPD. *Eur Respir J* 2018; 52: 1701994 [<https://doi.org/10.1183/13993003.01994-2017>].

**ABSTRACT** Inadequate DNA repair is implicated in the pathogenesis of chronic obstructive pulmonary disease (COPD). However, the mechanisms that underlie inadequate DNA repair in COPD are poorly understood. We applied an integrative genomic approach to identify DNA repair genes and pathways associated with COPD severity.

We measured the transcriptomic changes of 419 genes involved in DNA repair and DNA damage tolerance that occur with severe COPD in three independent cohorts (n=1129). Differentially expressed genes were confirmed with RNA sequencing and used for patient clustering. Clinical and genome-wide transcriptomic differences were assessed following cluster identification. We complemented this analysis by performing gene set enrichment analysis, Z-score and weighted gene correlation network analysis to identify transcriptomic patterns of DNA repair pathways associated with clinical measurements of COPD severity.

We found 15 genes involved in DNA repair and DNA damage tolerance to be differentially expressed in severe COPD. K-means clustering of COPD cases based on this 15-gene signature identified three patient clusters with significant differences in clinical characteristics and global transcriptomic profiles. Increasing COPD severity was associated with downregulation of the nucleotide excision repair pathway.

Systematic analysis of the lung tissue transcriptome of individuals with severe COPD identified DNA repair responses associated with disease severity that may underlie COPD pathogenesis.

---

This article has supplementary material available from [erj.ersjournals.com](http://erj.ersjournals.com)

Received: Sept 29 2017 | Accepted after revision: July 25 2018

Copyright ©ERS 2018

## Introduction

Chronic obstructive pulmonary disease (COPD) is a leading cause of global mortality [1]. Chronic exposure to cigarette smoke (CS) is a leading modifiable risk factor for COPD, but COPD is a complex and heterogeneous disease, and the clinical and pathologic consequences of chronic CS exposure vary amongst smokers. The factors that underlie COPD heterogeneity are not well understood, but may include the cellular responses to DNA damage [2–6]. CS is a well-characterised genotoxin, and CS-mediated DNA damage contributes to COPD pathogenesis [7, 8]. Lung cells and peripheral blood cells from COPD patients demonstrate increased global and telomeric DNA damage, and cellular responses to CS-mediated DNA damage can result in pathogenic events involved in disease progression, including apoptosis, cellular senescence, inflammation and mutagenesis [9–11].

DNA damage is sensed and repaired by a diverse, integrated network of cellular signalling pathways collectively known as the DNA damage response, involving multiple DNA repair and DNA damage tolerance pathways [12–14]. Direct repair (DR) reverses covalently modified nucleotides *via* a single enzymatic reaction, base excision repair (BER) repairs incorrect or damaged bases, mismatch repair (MMR) repairs aberrant nucleotide insertions or deletions, and nucleotide excision repair (NER) repairs “bulky” lesions *via* the excision and repair of multi-base oligonucleotides that are sensed by either stalled RNA polymerase or helix distortions. Double-stranded DNA breaks are potent inducers of cellular dysfunction and are repaired *via* homologous recombination (HR) or non-homologous end-joining (NHEJ). HR requires template sister chromatids and predominately occurs during replication, whereas NHEJ occurs throughout the cell cycle but is more error prone. The Fanconi anaemia (FA) pathway integrates multiple DNA repair pathways to repair interstrand crosslinks, while certain enzymes are needed to repair or elongate shortened and/or damaged telomeres. Translesion synthesis (TLS) refers to the use of specialised polymerases that allow for DNA replication past DNA lesions. In addition, many enzymes are involved in the remodelling of chromatin in response to DNA damage. Collectively, these pathways constitute mechanisms *via* which eukaryotic cells repair or tolerate DNA damage.

Inadequate DNA repair has been observed in the context of COPD. CS inhibits DNA repair *in vitro* and cells acquired from individuals with COPD demonstrate a lower capacity for DNA repair [15, 16]. Several polymorphisms in DNA repair genes have been associated with COPD susceptibility, and decreased expression of specific DNA repair genes has been demonstrated in the lungs of subjects with COPD [2, 16, 17]. However, a systematic characterisation of DNA repair mechanisms in COPD is lacking. We hypothesised that severe COPD is associated with an impaired response to DNA damage. To evaluate this hypothesis, we analysed the expression of genes involved in DNA repair and DNA damage tolerance in lung tissue from patients with COPD to identify differentially expressed genes (DEGs) and pathways associated with severe COPD.

## Methods

We analysed microarray mRNA expression data from lung tissue samples from three independent patient cohorts: Lung Genomics Research Consortium (LGRC), Ohio State University (OSU), and Lung Expression Quantitative Trait Loci Consortium (Lung eQTL). Basic summary data are provided in table 1. Normalised gene expression values were adjusted for age, smoking status (current, former, never) and sex in the LGRC and Lung eQTL study, but not in the OSU study owing to sample size [18]. Details describing tissue procurement, cohort characteristics, gene expression normalisation and adherence to institutional review board guidelines have been previously described, and further details are provided in the supplementary methods [19–22]. An outline of the study design is shown in figure 1. We identified 419 genes constituting 10 pathways involved in DNA repair and DNA damage tolerance (DDRT) (supplementary table E1) [23, 24]. Using data from the LGRC, OSU and Lung eQTL studies, we compared the expression of these genes in patients with severe COPD (Global Initiative for Chronic Obstructive Lung Disease (GOLD) IV) *versus* non-severe disease (GOLD I, II) and severe COPD (GOLD IV) *versus* control (GOLD 0) using significance analysis of microarrays [25, 26]. DDRT genes were included for further analysis if they were differentially expressed in all three cohorts and shared the same direction of effect (false discovery rate (FDR) <0.1) (supplementary table E2). DDRT genes were validated based on RNA sequencing (RNAseq) of lung tissue from a subset of 57 LGRC patient samples. Complete details for this cohort have been previously described (supplementary table E3) [27]. We clustered all LGRC patients with COPD (GOLD I–IV) based on the 15 DDRT consensus genes using K-means. Following cluster identification, we identified clinical characteristics associated with each cluster, and we performed genome-wide transcriptomic analysis to identify specific pathways associated with each cluster. Gene set enrichment analysis (GSEA) [28], Z-score [29] and weighted gene correlation network analysis (WGCNA) [30, 31] were applied to genome-wide transcriptomic data from the LGRC cohort to identify transcriptional changes of known DDRT pathways that correlated with disease severity. For detailed methods, please refer to the supplementary methods.

TABLE 1 Demographic characteristics of study patients

	LGRC				OSU			Lung eQTL		
	GOLD 0	GOLD I, II	GOLD III	GOLD IV	GOLD 0	GOLD I, II	GOLD IV	GOLD 0	GOLD I, II	GOLD IV
<b>Subjects n</b>	93	97	27	45	9	13	10	389	389	57
<b>Male sex</b>	41 (44.1)	61 (62.9)	16 (59.3)	18 (40.0)	4 (44.4)	6 (46.2)	6 (60.0)	189 (48.3)	245 (63.0)	16 (28.1)
<b>Age years</b>	63.7±11.7	69.1±8.1	65.1±8.7	57.7±8.5	63.2±11.4	69.2±7.3	50.9±5.6	60.9±10.6	65.4±9.4	54.0±5.5
<b>Ever-smoker</b>	52 (61.9) [9]	93 (95.9)	25 (93.8)	44 (97.8)	9 (100)	13 (100)	10 (100)	330 (84.8)	377 (96.9)	55 (94.7)
<b>Pack-years</b>	22±33 [9]	57±40	49±43	49±23	26±17	40±23	56±36	40±23 [37]	50±28 [23]	33±16 [2]
<b>Coexisting malignancy</b>	83 (89.2)	81 (83.5)	22 (81.3)	3 (6.7)	0 (0)	0 (0)	0 (0)	364 (93.6)	379 (97.4)	3 (5.3)
<b>Percentage emphysema</b>	0.3±0.7	6.3±8.2	22.8±16.0	37.6±13.4						
<b>D<sub>Lco</sub> % pred</b>	84.9±16.5 [11]	67.1±20.2 [5]	42.6±13.0 [2]	31.8±10.0 [6]						
<b>FEV<sub>1</sub> % pred</b>	95.2±12.3	65.1±13.3	34.5±5.7	20.2±3.9						
<b>SGRQ score</b>	12.5±16.1 [11]	22.4±18.1 [16]	40.2±19.8	61.3±12.5						
<b>BODE index</b>	0.8±1.2	1.5±1.4	5.2±1.8	6.8±1.3						
<b>6MWD m</b>	437±118 [34]	403±35 [16]	325±97 [5]	277±88 [2]						
<b>SF-12 score</b>	48±11 [11]	45±11 [1]	37±10	29±7						

Data are presented as n (%) or mean±SD, unless otherwise stated. [x] represents missing samples. LGRC: Lung Genomics Research Consortium; OSU: Ohio State University; eQTL: expression quantitative trait loci; GOLD: Global Initiative for Chronic Obstructive Lung Disease; D<sub>Lco</sub>: diffusing capacity of the lungs for carbon monoxide; FEV<sub>1</sub>: forced expiratory volume in 1 s; SGRQ: St George's Respiratory Questionnaire; BODE: body mass index, airflow obstruction, dyspnoea and exercise capacity; 6MWD: 6-min walk distance; SF-12: 12-item Short-Form Health Survey.

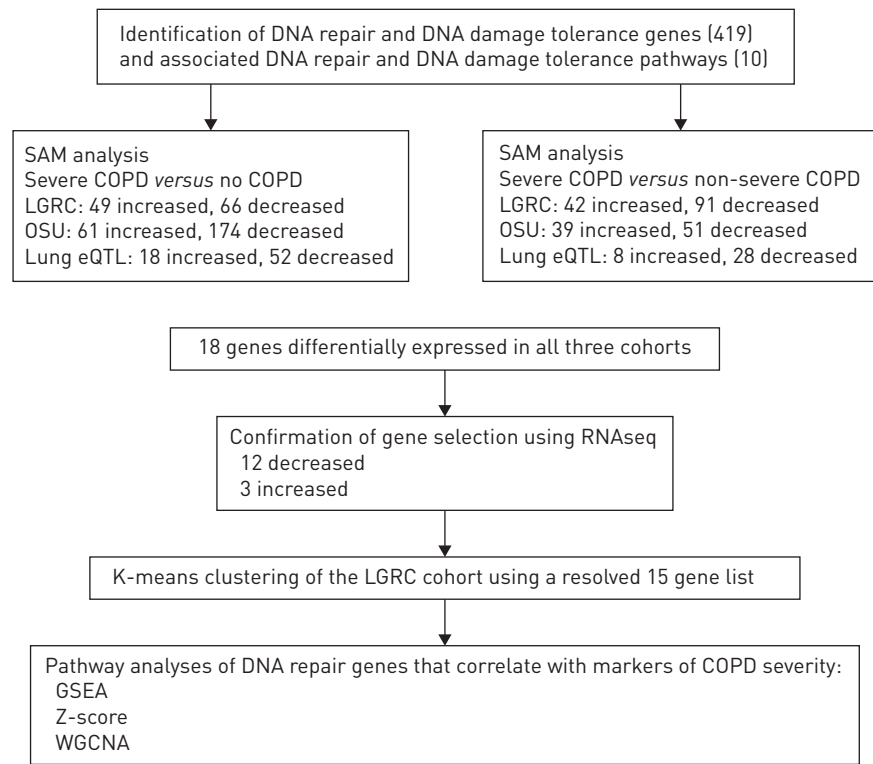


FIGURE 1 Study workflow. COPD: chronic obstructive pulmonary disease; SAM: significance analysis of microarrays; LGRC: Lung Genomics Research Consortium; OSU: Ohio State University; Lung eQTL: Lung expression quantitative trait loci; GSEA: gene set enrichment analysis; WGCNA: weighted gene correlation network analysis.

## Results

### *A DNA repair signature of 15 genes is associated with severe COPD*

We analysed 419 DDRT genes in three cohorts: LGRC, OSU and Lung eQTL. These cohorts were chosen to overcome the potential confounding effect of coexisting malignancy that might occur if we studied the LGRC cohort alone. We chose one COPD cohort with a high prevalence of coexisting malignancy (Lung eQTL) and one COPD cohort without coexisting malignancy (OSU). GOLD III patients were not included in the OSU study, and therefore were not included in these analyses. We identified 18 differentially expressed DDRT genes present in the comparisons between severe COPD, non-severe COPD and controls in the three cohorts (supplementary table E2). A second filtering step was implemented to test these 18 DDRT genes on a subset of patients in the LGRC cohort using RNAseq, a non-array method, to confirm gene expression changes in the lungs of patients with COPD. Of the 18 identified genes in the array-based cohorts, 15 DDRT genes were confirmed in the RNAseq subgroup (figure 2).

### *Identification of three COPD clusters using the 15-DDRT gene signature*

To characterise distinct DNA repair patient clusters in COPD, we performed K-means clustering using the 15-DDRT signature in the LGRC cohort. We identified three distinct clusters of COPD using this approach (figure 2), and compared clinical differences amongst clusters. Clinical measurements of disease included the percentage of emphysema present based on high resolution computed tomography, forced expiratory volume in 1 s (FEV<sub>1</sub>) % predicted, diffusing capacity for carbon monoxide (DLCO) % predicted, 6-min walk distance (6MWD), St George's Respiratory Questionnaire (SGRQ), BODE index (body mass index, airflow obstruction, dyspnoea and exercise capacity), and the 12-item Short-Form Health Survey. Patients in Cluster 1 (n=65) had milder disease than patients in Cluster 2 and Cluster 3, characterised by less emphysema, less impairment in DLCO and increased FEV<sub>1</sub> (figure 3). Similarly, compared to patients in Cluster 2 and Cluster 3, patients in Cluster 1 had better functional status and higher quality of life as measured by 6MWD, BODE index and SGRQ scores. There were no statistically significant differences in the clinical characteristics of patients in severe-disease Cluster 2 and severe-disease Cluster 3. There were no differences in sex, pack-years, race or average age amongst the three clusters (supplementary table E4). There were no differences in the rates of coexisting malignancy between Cluster 2 and Cluster 3, but there were increased rates of coexisting malignancy in Cluster 1. These data suggest that clustering of COPD cases based on a DNA repair gene signature identifies three clusters, with Cluster 2 and Cluster 3 characterised by increased disease severity.

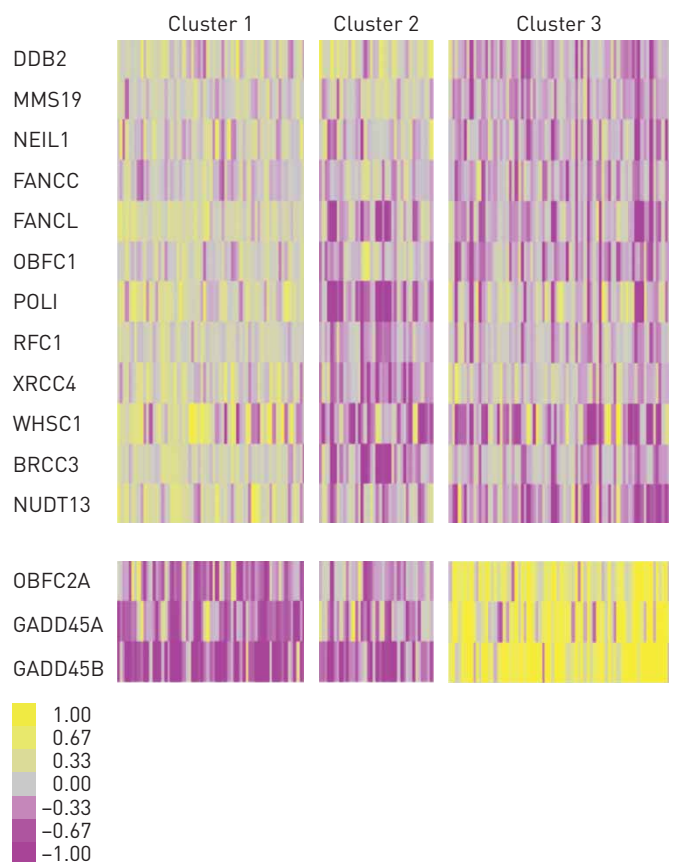


FIGURE 2 K-means clustering of patients based on 15 gene consensus signature. Yellow denotes an increase over the sample mean, and purple denotes a decrease over sample mean.

### Global gene expression profiling of DNA repair clusters in COPD

To characterise the global gene expression patterns of these three clusters, we compared the global transcriptomic profiles of patients in the three clusters with control samples in the LGRC cohort. Cluster 1 had 361 DEGs, Cluster 2 had 3109 DEGs and Cluster 3 had 2219 DEGs. A total of 73 DEGs were dysregulated in all three clusters, and 22% of these common DEGs ( $n=16$ ) had changes in the same direction in all three clusters. To identify non-DNA repair pathways associated with these clusters, pathway enrichment analyses were performed (supplementary table E5). The top enriched pathways for both Cluster 1 and Cluster 3 were related to cytokine signalling. In Cluster 1, several interleukin pathways (IL-1, IL-3, IL-5, IL-6, IL-17 and IL-18) were amongst the top 10 enriched pathways. Similarly, in Cluster 3, interleukin pathways (IL-3, IL-5, IL-10 and IL-17) were amongst the top 10 enriched pathways. The most enriched pathway in both Cluster 1 and Cluster 3 was IL-5 but with the opposite direction of effect: Cluster 1 showed downregulation and Cluster 3 showed upregulation of genes in the IL-5 pathway. In contrast to Cluster 1 and Cluster 3, Cluster 2 was characterised by upregulation of several pathways involved in cell adhesion and cytoskeletal remodelling, including transforming growth factor- $\beta$  (TGF- $\beta$ ) and WNT pathways. The most significant DEGs amongst the top 50 signalling pathways for Cluster 2 and Cluster 3 are shown in figure 4. These pathway enrichment data suggest that Cluster 2 is associated with increased expression of genes involved in tissue remodelling, and Cluster 3 is associated with increased expression of genes involved in inflammation.

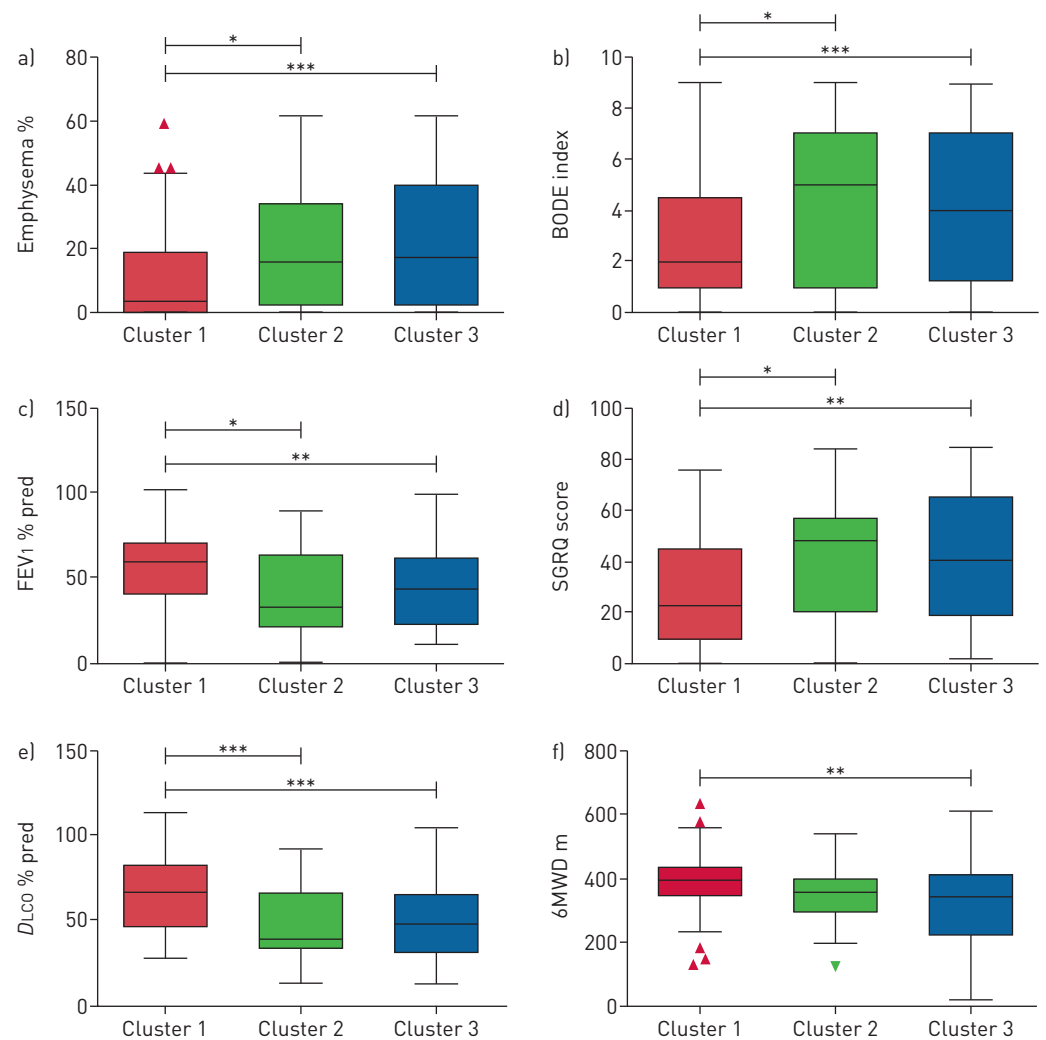


FIGURE 3 Clinical characteristics of chronic obstructive pulmonary disease by cluster. a) Box and whiskers of percentage emphysema by cluster. b) Box and whiskers of BODE index (body mass index, airflow obstruction, dyspnoea and exercise capacity) by cluster. c) Box and whiskers of forced expiratory volume in 1 s (FEV1) % predicted by cluster. d) Box and whiskers of St George's Respiratory Questionnaire (SGRQ) score by cluster. e) Box and whiskers of diffusing capacity of the lung for carbon monoxide (DLco) % predicted by cluster. f) Box and whiskers of 6-min walk distance (6MWD) by cluster. \*:  $p<0.05$ ; \*\*:  $p<0.005$ ; \*\*\*:  $p<0.0005$ .

To confirm the location of selected DNA repair proteins, we performed immunohistochemistry on lung tissue samples for Endonuclease 8-like 1 (NEIL1), X-ray repair cross-complementing protein 4 (XRCC4) and DNA damage-binding protein 2 (DDB2). Nuclear staining was identified in epithelial cells, endothelial cells and macrophages. Bronchiolar epithelial cells demonstrated the most prominent staining intensity for all three proteins (figure 5). There was marked heterogeneity in staining intensity between samples for all three proteins. We did not identify a clear difference amongst clusters when evaluating XRCC4; however, we did identify decreased epithelial staining for DDB2 and NEIL1 in samples from patients in severe-disease Cluster 3 when compared to samples from patients in mild Cluster 1. Notably, there was more DDB2 staining in samples from patients with a history of smoking than in those of never smokers. These data suggest that transcriptional changes identified in whole lung tissue samples are also associated with cellular protein level differences in patients with severe COPD.

**DDRT pathways in patients with COPD**

While our initial analysis identified individual genes associated with severe COPD, we sought to determine if DDRT pathways were differentially expressed in patients with severe COPD using three different approaches. First, we performed a genome-wide analysis to identify genes that correlated with clinical measurements of COPD severity, and then used GSEA to identify the DDRT pathways that were significantly enriched amongst the most correlated genes. We found that TLS, NER and FA pathways were inversely correlated (*i.e.* protective) with multiple measurements of COPD severity (figure 6a–d and supplementary table E6A). Second, a Z-score analysis was performed using the transcriptomic profiles of

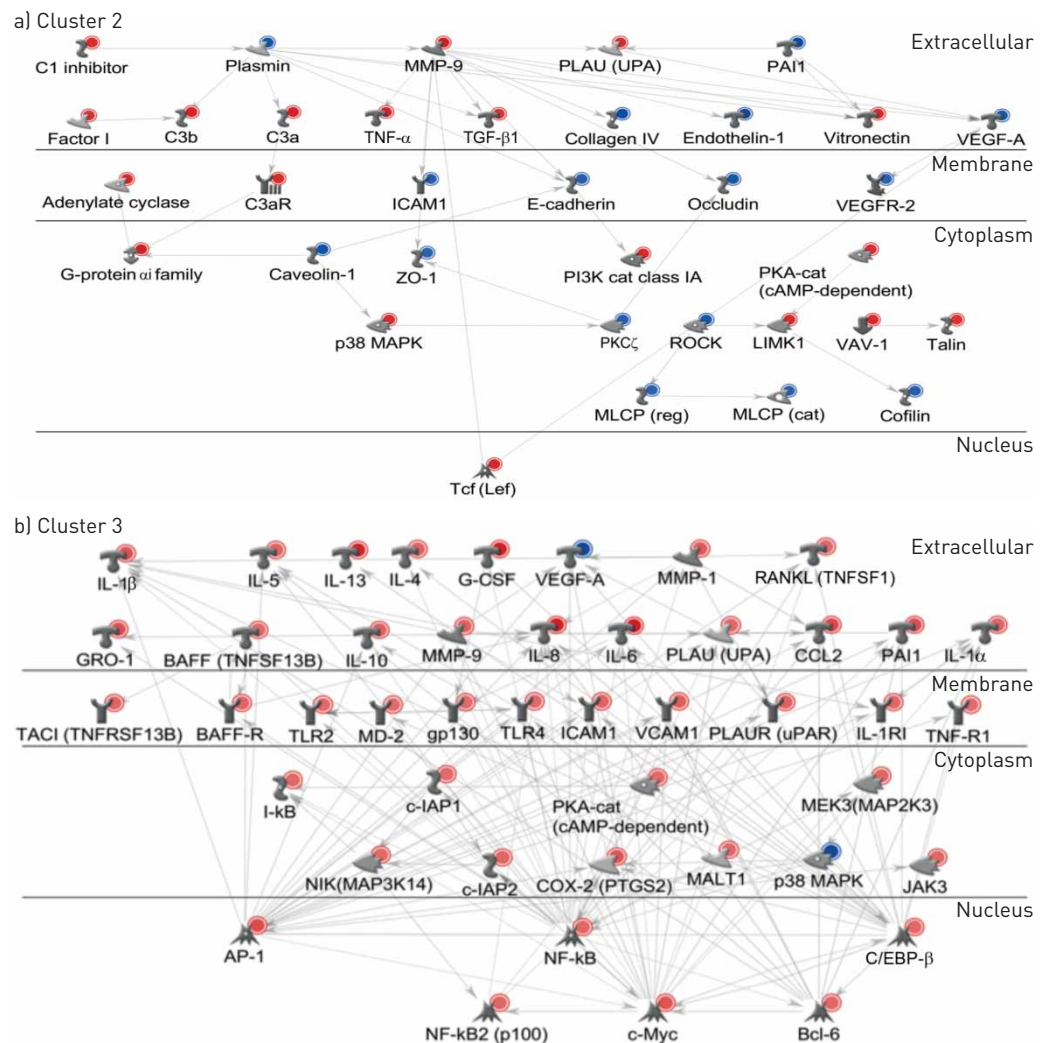


FIGURE 4 Overrepresented genes in the top 50 enriched pathways. a) Comparison between Cluster 2 and controls. b) Comparison between Cluster 3 and controls. Red denotes upregulated genes. Blue denotes downregulated genes. Lines represent curated associations between genes.

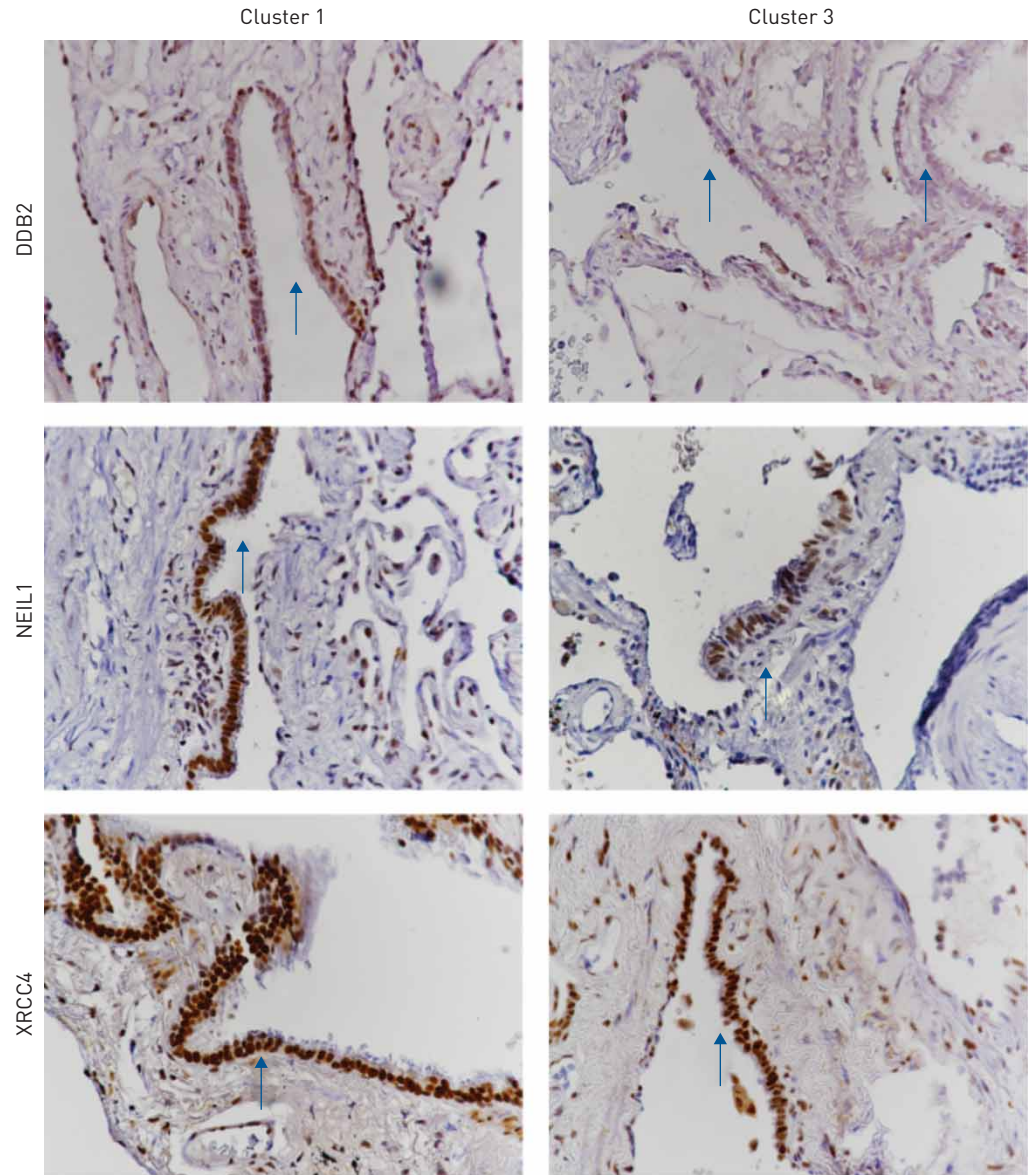


FIGURE 5 Immunohistochemistry for DDB2, NEIL1 and XRCC4. Immunohistochemistry demonstrating localisation and staining intensity for DDB2, NEIL1 and XRCC4 (identified by brown chromogen) performed on lung tissue samples from patients in Cluster 1 and Cluster 3. Nuclear staining appeared particularly localised to bronchiole epithelial cells (arrows), although other cells also demonstrated nuclear staining. Images acquired using a 40× objective lens.

lung tissue from COPD patients. We generated DDRT pathway coefficients (Z-scores) for each individual with COPD, and correlated these coefficients with clinical characteristics of disease. The NER, TLS, FA, MMR and HR pathways were inversely correlated with multiple measures of COPD severity (figure 6e-h and supplementary table E6B). In both methods, the DR pathway was the only one that showed a positive correlation with clinical measurements of COPD severity. However, this pathway was the smallest (n=8 genes), making it more susceptible to the influence of the weights used to generate the coefficient. Finally, we performed WGCNA using whole transcriptome data to determine if DDRT pathways were co-expressed and correlated with indices of disease severity. We identified 40 modules of co-expressed gene, and multiple modules correlated with disease severity (figure 7). To ensure that the makeup of our DNA repair pathway gene lists was not biasing our results, we used Metacore to identify gene set enrichment across the full complement of cellular pathways (supplementary table E7). The module with the strongest negative correlation with measurements of disease severity, Yellow, was also most enriched for the NER-BER pathway. The Yellow module correlated with the percentage of emphysema

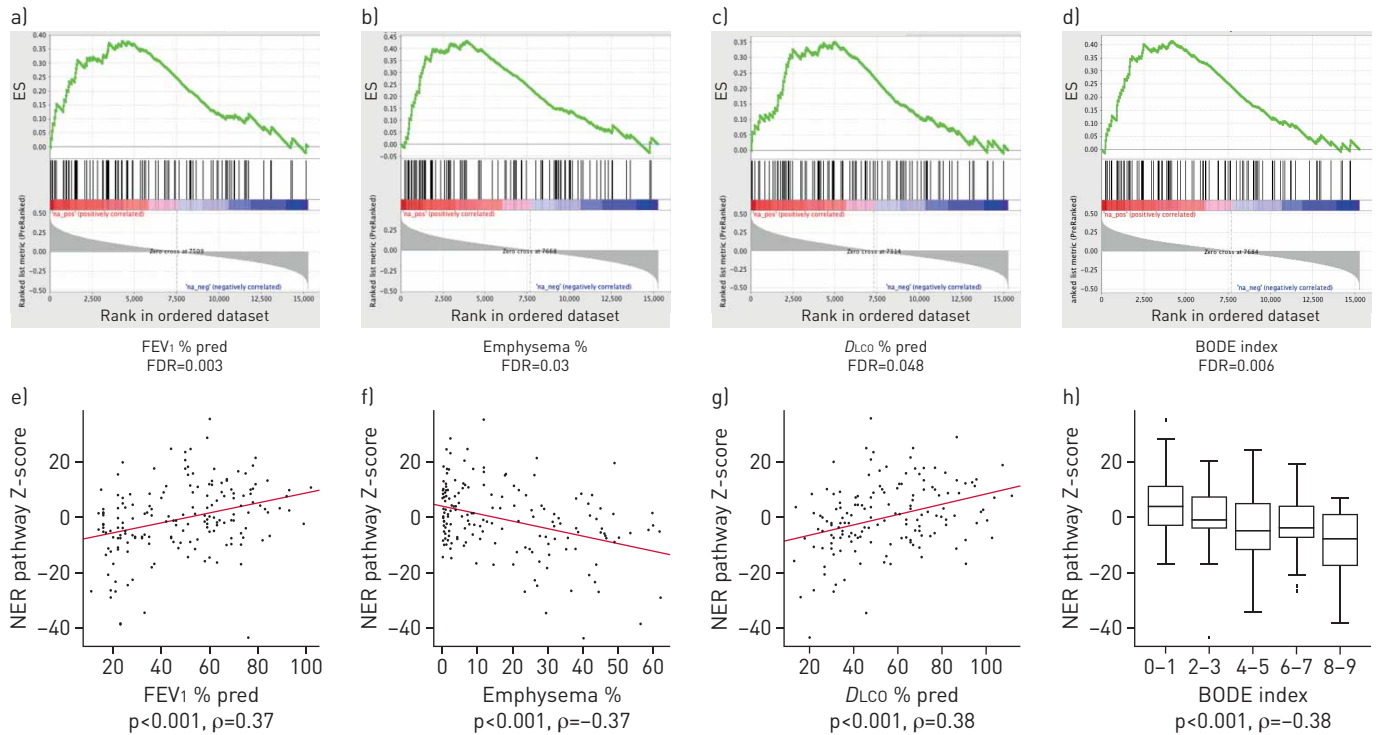


FIGURE 6 The nucleotide excision repair (NER) pathway is downregulated in severe chronic obstructive pulmonary disease. a–d) Enrichment plots from gene set enrichment analysis. The enrichment plots contain profiles of the running enrichment scores (ES) and the barcode plot indicates the position of the genes in each gene set; red represents Spearman correlations with more severe disease, blue represents Spearman correlations with less severe disease. False discovery rates (FDRs) for NER gene set enrichment are reported for a) forced expiratory volume in 1 s (FEV1), b) percentage emphysema, c) diffusing capacity of the lung for carbon monoxide ( $DL_{CO}$ ) % pred and d) BODE index (body mass index, airflow obstruction, dyspnoea and exercise capacity). e–h) NER pathway Z-score coefficients for each patient plotted against e) FEV1 % pred, f) percentage emphysema, g)  $DL_{CO}$  % pred and h) BODE index.

(correlation =  $-0.4$ ,  $p=1 \times 10^{-7}$ ), BODE index (correlation =  $-0.4$ ,  $p=4 \times 10^{-8}$ ), FEV1 % pred (correlation =  $0.34$ ,  $p=5 \times 10^{-6}$ ), SGRQ (correlation =  $-0.43$ ,  $p=4 \times 10^{-9}$ ),  $DL_{CO}$  % pred (correlation =  $0.38$ ,  $p=3 \times 10^{-7}$ ) and 6MWD (correlation =  $0.37$ ,  $p=8 \times 10^{-7}$ ). There were multiple canonical NER genes within the yellow module that demonstrated both high module membership and gene significance for clinical indices of COPD severity, including Xeroderma pigmentosum group a-complementing protein (XPA) and Excision repair cross-complementation group 5 (ERCC5) (supplementary figure E8). The combination of these three approaches demonstrated that downregulation of the NER pathway was associated with COPD severity.

## Discussion

In this study, we identified 15 DEGs that were common to three independent cohorts in the largest assessment of DDRT genes to date. Transcriptional changes of these 15 genes were heterogeneous amongst COPD patients. However, subsequent clustering of patients based on these 15 genes identified three clusters with different clinical characteristics and gene expression profiles correlating with important mechanisms of disease pathogenesis, suggesting a potential relationship between DNA repair, inflammation and tissue remodelling. Our data also suggest that multiple DDRT pathways are downregulated in patients with COPD, with the strongest evidence being demonstrated for the NER pathway. Taken together, these data support the hypothesis that diminished DNA repair underlies the complex and heterogeneous manifestations of COPD.

Severe COPD was associated with upregulation of three of the 15 DDRT genes, *GADD45A*, *GADD45B* and *OBFC2A*. These three genes are relevant to COPD pathogenesis because they are implicated in cell cycle arrest, apoptosis and cellular senescence. We also found an association between severe COPD and downregulation of 12 DDRT genes. Amongst these 12 DDRT genes were two FA genes (*FANCC* and *FANCL*), two NER genes (*DDB2* and *MMS19*) and three genes involved in HR and NHEJ pathways (*WHSC1*, *BRCC3* and *XRCC4*). Additionally, we identified *OBFC1*, which is implicated in the maintenance of telomere length; *POLI*, which has an exonuclease function; and *NEIL1*, a canonical BER gene which is also implicated in NER. Previous studies of *NEIL1* and *POLI* have shown that genotoxic stress increases the expression of these genes; however, our data show decreased expression of these genes in severe COPD



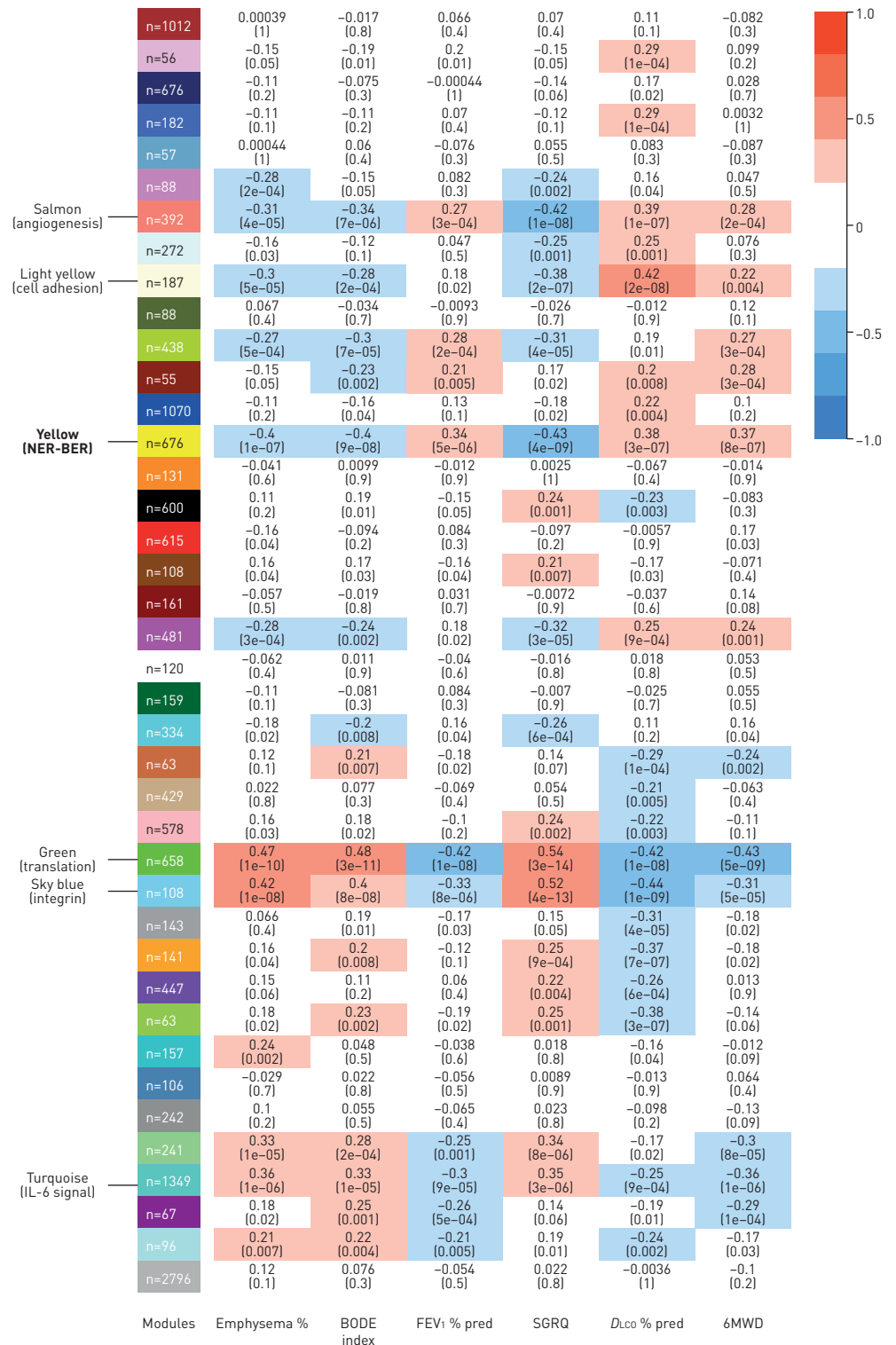


FIGURE 7 Weighted gene co-expression network analysis (WGCNA). WGCNA identified 40 gene modules as demonstrated in this WGCNA heatmap. n represents the number of genes within each module. Positive correlations are red, and negative correlations are blue. The green, sky blue, and turquoise were the three modules most positively correlated with indices of decreased disease severity, and the yellow, salmon, and light yellow were the three modules most negatively correlated with indices of increased disease severity. The most highly enriched process is indicated for each of the top six modules, including the nucleotide excision repair (NER)-base excision repair (BER) process in the yellow module. IL-6: interleukin 6; BODE: body mass index, airflow obstruction, dyspnoea and exercise capacity; FEV1: forced expiratory volume in 1 s; SGRQ: St George's Respiratory Questionnaire; DLco: diffusing capacity of the lung for carbon monoxide; 6MWD: 6-min walk distance.

[32]. There are many potential reasons for such differences pertinent to the pathogenesis of COPD, including histone modifications, dysregulation of homeostatic signalling due to oxidative stress, and interference with gene transcription by DNA lesions. For example, *NEIL1* has been frequently found to be hypermethylated in head and neck cancer [33]. Therefore, our data support the hypothesis that a maladaptive response to genotoxic stress contributes to disease progression in COPD.

The 15-DDRT gene signature identified three patient clusters of COPD differentiated by disease severity and distinct non-DNA repair pathway expression profiles. Cluster 1 was characterised by mild clinical disease, whereas Clusters 2 and 3 had severe disease. Cluster 2 showed enrichment for pathways associated with cytoskeletal remodelling, including TGF- $\beta$  and WNT signalling. Cluster 3 showed enrichment for NF- $\kappa$ B, IL-5 and IL-17 pathways. Excess inflammation and aberrant remodelling are well-described mechanisms of COPD pathogenesis. The relationship between DNA damage and chronic inflammation is well described, because defective DNA repair contributes to autoimmunity, chronic inflammation and tissue remodelling [34, 35]. Based on these findings we suggest that future studies of COPD pathogenesis consider the DNA damage response in conjunction with assessments of these inflammatory and tissue remodelling pathways.

We applied multiple methods to characterise transcriptional changes in DDRT pathways, and identified transcriptional changes in the NER pathway as most consistently associated with increased disease severity across multiple clinical features and all analytical methods. Interestingly, certain genes that appeared differentially expressed between Cluster 2 and Cluster 3, including *NEIL1*, *DDB2* and *MMS19*, are implicated in NER. Furthermore, immunohistochemistry confirmed that both *DDB2* and *NEIL1* were decreased in severe-disease Cluster 3. This is significant for COPD pathogenesis because the NER pathway is primarily responsible for detecting and removing bulky DNA adducts caused by CS, and it is therefore critical for protecting against tobacco-induced carcinogenesis. Previous studies have demonstrated impaired NER capacity with CS, and diminished NER capacity has been implicated as a risk for lung cancer [36, 37]. Inadequate NER may also lead to excess DNA damage and subsequent susceptibility to cell death, tissue destruction and/or inflammation, and emphysema [38]. It is likely that other DDRT pathways are dysregulated in COPD given our findings. Importantly, almost all observed associations between COPD severity and DDRT pathways suggest that downregulation of genes involved in DNA repair and DNA damage tolerance occurs in severe COPD.

There are certain limitations to our study. The influence of coexisting malignancy on the transcriptomic profile of DDRT genes in patients with COPD is an important confounding variable. While tissue samples were taken from non-malignant tissue, changes have been identified in “normal” lung tissue from patients with COPD and coexisting malignancies [39, 40]. This is a challenging problem because lung tissue is not commonly obtained from patients with normal lung function, unless there is a suspicion of cancer. To account for the potential influence of a “field of cancerisation”, we included the OSU cohort that excluded patients with a coexisting malignancy to generate our consensus DNA repair signature. Other potential limitations are that many DDRT genes are not primarily regulated at the transcriptional level, and that we profiled whole lung tissue and therefore differential gene expression may be due to differences in tissue composition of various cell types. To address this concern, we performed immunohistochemistry for various DNA repair proteins and identified decreased *DDB2* and *NEIL1* in severe-disease Cluster 3 lung tissue samples. We did not see profound differences amongst clusters when analysing *XRCC4*; however, there was significant heterogeneity amongst samples and our study was likely underpowered to detect a difference. Future studies will require analyses with additional molecular readouts including protein concentration, modifiers (*i.e.* phosphorylation, ubiquitination), cell type and nuclear co-localisation.

We used a multistep, complementary analytical approach to study DDRT genes and pathways and their association with disease severity in three independent cohorts. At the individual gene level, we found that a 15-DDRT gene signature enabled the identification of three disease clusters characterised by clinical differences in severity and distinct non-DNA repair gene pathways associated with increased inflammation and tissue remodelling. We also identified a consistent downregulation of the NER pathway in severe COPD. These findings suggest that transcriptional changes in DDRT genes contribute to disease heterogeneity and may underlie distinct pathogenic responses in COPD.

**Acknowledgements:** The authors would like to thank the staff at the Respiratory Health Network Tissue Bank of the Fonds de Recherche Québec - Santé for their valuable assistance with the lung eQTL dataset at Laval University.

**Conflict of interest:** W. Timens reports unrestricted institutional grants from Merck, during the conduct of the study; fees paid to institution for consultancy from Pfizer, fees paid to the institution for lecturing from GSK, Chiesi, Lilly Oncology and Boehringer Ingelheim, fees paid to the institution for consultancy and lecturing, and travel costs, from Roche Diagnostics/Ventana, grants from Dutch Asthma Fund, fees for travel paid to the institution from Biotest, and fees paid to the institution for consultancy and lecturing from Merck Sharp Dohme, AstraZeneca and Novartis, outside

the submitted work. N. Kaminski reports grants and personal fees for consultancy from Biogen Idec, personal fees for consultancy from Boehringer Ingelheim, Third Rock and MMI, non-financial support from Actelion and Miragen, personal fees for advisory board work from Pliant, unpaid consultancy work for Samumed, and personal fees from Numedii, outside the submitted work; in addition, N. Kaminski has a patent New Therapies in Pulmonary Fibrosis licensed, and a patent Peripheral Blood Gene Expression issued, and is a Member of the Scientific Advisory Committee, the Research Advisory Forum and the Board of the Pulmonary Fibrosis Foundation, and also serves as Deputy Editor of *Thorax*. None of the above relate to COPD.

Support statement: The authors would like to acknowledge grants to M. Sauler (K08HL135402-01; FAMRI YCSA 142017) and J.L. Gomez (K01HL125474-03; FAMRI YCSA 113393). M. Lamontagne was the recipient of a doctoral studentship from the Fonds de Recherche Québec - Santé (FRQS). Y. Bossé holds a Canada Research Chair in Genomics of Heart and Lung Diseases. Funding information for this article has been deposited with the Crossref Funder Registry.

## References

- 1 GBD 2015 Chronic Respiratory Disease Collaborators. Global, regional, and national deaths, prevalence, disability-adjusted life years, and years lived with disability for chronic obstructive pulmonary disease and asthma, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet Respir Med* 2017; 5: 691–706.
- 2 Caramori G, Adcock IM, Casolari P, *et al*. Unbalanced oxidant-induced DNA damage and repair in COPD: a link towards lung cancer. *Thorax* 2011; 66: 521–527.
- 3 Anderson GP, Bozinovski S. Acquired somatic mutations in the molecular pathogenesis of COPD. *Trends Pharmacol Sci* 2003; 24: 71–76.
- 4 Pastukh VM, Zhang L, Ruchko MV, *et al*. Oxidative DNA damage in lung tissue from patients with COPD is clustered in functionally significant sequences. *Int J Chron Obstruct Pulmon Dis* 2011; 6: 209–217.
- 5 Aoshiba K, Zhou F, Tsuji T, *et al*. DNA damage as a molecular link in the pathogenesis of COPD in smokers. *Eur Respir J* 2012; 39: 1368–1376.
- 6 Aguilera-Aguirre L, Hosoki K, Bacsı A, *et al*. Whole transcriptome analysis reveals a role for OGG1-initiated DNA repair signaling in airway remodeling. *Free Radic Biol Med* 2015; 89: 20–33.
- 7 Liu X, Conner H, Kobayashi T, *et al*. Cigarette smoke extract induces DNA damage but not apoptosis in human bronchial epithelial cells. *Am J Respir Cell Mol Biol* 2005; 33: 121–129.
- 8 Nakayama T, Kaneko M, Kodama M, *et al*. Cigarette smoke induces DNA single-strand breaks in human cells. *Nature* 1985; 314: 462–464.
- 9 Maluf SW, Mergener M, Dalcanele L, *et al*. DNA damage in peripheral blood of patients with chronic obstructive pulmonary disease (COPD). *Mutat Res* 2007; 626: 180–184.
- 10 Savale L, Chaouat A, Bastuji-Garin S, *et al*. Shortened telomeres in circulating leukocytes of patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2009; 179: 566–571.
- 11 Rodier F, Coppe JP, Patil CK, *et al*. Persistent DNA damage signalling triggers senescence-associated inflammatory cytokine secretion. *Nat Cell Biol* 2009; 11: 973–979.
- 12 Friedberg EC. DNA damage and repair. *Nature* 2003; 421: 436–440.
- 13 Jackson SP, Bartek J. The DNA-damage response in human biology and disease. *Nature* 2009; 461: 1071–1078.
- 14 Ciccia A, Elledge SJ. The DNA damage response: making it safe to play with knives. *Mol Cell* 2010; 40: 179–204.
- 15 Hang B. Formation and repair of tobacco carcinogen-derived bulky DNA adducts. *J Nucleic Acids* 2010; 2010: 709521.
- 16 da Silva AL, da Rosa HT, Karnopp TE, *et al*. Evaluation of DNA damage in COPD patients and its correlation with polymorphisms in repair genes. *BMC Med Genet* 2013; 14: 93.
- 17 Bosse Y. Updates on the COPD gene list. *Int J Chron Obstruct Pulmon Dis* 2012; 7: 607–631.
- 18 Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 2007; 8: 118–127.
- 19 Ezzie ME, Crawford M, Cho JH, *et al*. Gene expression networks in COPD: microRNA and mRNA regulation. *Thorax* 2012; 67: 122–131.
- 20 Bauer Y, Tedrow J, de Bernard S, *et al*. A novel genomic signature with translational significance for human idiopathic pulmonary fibrosis. *Am J Respir Cell Mol Biol* 2015; 52: 217–231.
- 21 Hao K, Bosse Y, Nickle DC, *et al*. Lung eQTLs to help reveal the molecular underpinnings of asthma. *PLoS Genet* 2012; 8: e1003029.
- 22 Yang IV, Pedersen BS, Rabinovich E, *et al*. Relationship of DNA methylation and gene expression in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2014; 190: 1263–1272.
- 23 Kassambara A, Gourzones-Dmitriev C, Sahota S, *et al*. A DNA repair pathway score predicts survival in human multiple myeloma: the potential for therapeutic strategy. *Oncotarget* 2014; 5: 2487–2498.
- 24 Mjelle R, Hegre SA, Aas PA, *et al*. Cell cycle regulation of human DNA repair and chromatin remodeling genes. *DNA Repair (Amst)* 2015; 30: 53–67.
- 25 Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci USA* 2001; 98: 5116–5121.
- 26 Reiner A, Yekutieli D, Benjamini Y. Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics* 2003; 19: 368–375.
- 27 Kusko RL, Brothers IJ, Tedrow J, *et al*. Integrated genomics reveals convergent transcriptomic networks underlying COPD and IPF. *Am J Respir Crit Care Med* 2016; 194: 948–960.
- 28 Subramanian A, Tamayo P, Mootha VK, *et al*. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 2005; 102: 15545–15550.
- 29 Cheadle C, Vawter MP, Freed WJ, *et al*. Analysis of microarray data using Z score transformation. *J Mol Diagn* 2003; 5: 73–81.
- 30 Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008; 9: 559.

- 31 Modena BD, Bleecker ER, Busse WW, *et al.* Gene expression correlated with severe asthma characteristics reveals heterogeneous mechanisms of severe disease. *Am J Respir Crit Care Med* 2017; 195: 1449–1463.
- 32 Christmann M, Kaina B. Transcriptional regulation of human DNA repair genes following genotoxic stress: trigger mechanisms, inducible responses and genotoxic adaptation. *Nucleic Acids Res* 2013; 41: 8403–8420.
- 33 Chaisaingmongkol J, Popanda O, Warta R, *et al.* Epigenetic screen of human DNA repair genes identifies aberrant promoter methylation of NEIL1 in head and neck squamous cell carcinoma. *Oncogene* 2012; 31: 5108–5116.
- 34 Kidane D, Chae WJ, Czocho J, *et al.* Interplay between DNA repair and inflammation, and the link to cancer. *Crit Rev Biochem Mol Biol* 2014; 49: 116–139.
- 35 Alder JK, Barkauskas CE, Limjunyawong N, *et al.* Telomere dysfunction causes alveolar stem cell failure. *Proc Natl Acad Sci USA* 2015; 112: 5099–5104.
- 36 Planchard D, Domont J, Taranchon E, *et al.* The NER proteins are differentially expressed in ever smokers and in never smokers with lung adenocarcinoma. *Ann Oncol* 2009; 20: 1257–1263.
- 37 Li W, Hu J, Adebali O, *et al.* Human genome-wide repair map of DNA damage caused by the cigarette smoke carcinogen benzo[a]pyrene. *Proc Natl Acad Sci USA* 2017; 114: 6752–6757.
- 38 Sears CR, Zhou H, Justice MJ, *et al.* Xeroderma pigmentosum group C deficiency alters cigarette smoke DNA damage cell fate and accelerates emphysema development. *Am J Respir Cell Mol Biol* 2018; 58: 402–411.
- 39 Spira A, Beane JE, Shah V, *et al.* Airway epithelial gene expression in the diagnostic evaluation of smokers with suspect lung cancer. *Nat Med* 2007; 13: 361–366.
- 40 Korde A, Jin L, Zhang JG, *et al.* Lung endothelial microRNA-1 regulates tumor growth and angiogenesis. *Am J Respir Crit Care Med* 2017; 196: 1443–1455.