ERJ Express. Published on September 6, 2018 as doi: 10.1183/13993003.01994-2017

EUROPEAN RESPIRATORY *journal*

FLAGSHIP SCIENTIFIC JOURNAL OF ERS

Early View

Original article

The DNA Repair Transcriptome in Severe COPD

Maor Sauler, Maxime Lamontagne, Eric Finnemore, Jose D. Herazo-Maya, John Tedrow, Xuchen Zhang, Julia E. Morneau, Frank Sciurba, Wim Timens, Peter D Paré, Patty J. Lee, Naftali Kaminski, Yohan Bossé, Jose L. Gomez

Please cite this article as: Sauler M, Lamontagne M, Finnemore E, *et al*. The DNA Repair Transcriptome in Severe COPD. *Eur Respir J* 2018; in press (https://doi.org/10.1183/13993003.01994-2017).

The DNA Repair Transcriptome in Severe COPD

Authors:
Maor Sauler[1], Maxime Lamontagne[2], Eric Finnemore[1], Jose D. Herazo-Maya[1], John Tedrow[3], Xuchen Zhang[4], Julia E. Morneau[1], Frank Sciurba[3], Wim Timens[5], Peter D Paré[6], Patty J. Lee[1], Naftali Kaminski[1], Yohan Bossé[2,7], Jose L. Gomez[1]

[1]Department of Medicine, Yale School of Medicine, New Haven, CT,
[2]Centre de recherche Institut universitaire de cardiologie et de pneumologie de Québec, Laval University, Quebec, Canada
[3]Department of Medicine, University of Pittsburgh, Pittsburgh, PA,
[4]Department of Pathology, Yale School of Medicine, New Haven, CT,
[5]University of Groningen, University Medical Center Groningen, Department of Pathology and Medical Biology, GRIAC research institute, University of Groningen, Groningen, The Netherlands,
[6]The University of British Columbia Centre for Heart Lung Innovation, St. Paul's Hospital, Vancouver, British Columbia, Canada
[7]Department of Molecular Medicine, Laval University, Quebec, Canada

Corresponding author
Maor Sauler
The Anlyan Center
300 Cedar Street, Ste South 455C
New Haven, CT 06519
maor.sauler@yale.edu

Take Home message: Severe COPD is associated with reduced transcription of genes involved in the nucleotide excision repair pathway.

Twitter: DNA repair genes are downregulated in severe COPD, particularly the NER pathway, and these changes may underlie COPD heterogeneity

ABSTRACT

BACKGROUND: Inadequate DNA repair is implicated in the pathogenesis of COPD. However, the mechanisms that underlie inadequate DNA repair in COPD are poorly understood.

OBJECTIVES: We applied an integrative genomic approach to identify DNA repair genes and pathways associated with COPD severity.

METHODS: 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=1,129). 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 GSEA, Z-score, and WGCNA methods to identify transcriptomic patterns of DNA repair pathways associated with clinical measurements of COPD severity.

RESULTS: Fifteen genes involved in DNA repair and DNA damage tolerance were 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.

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

INTRODUCTION

Chronic Obstructive Pulmonary Disease (COPD) is currently the third 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 characterized 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 signaling 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, while NHEJ occurs throughout the cell cycle, but is more error prone. The Fanconi anemia pathway (FA) integrates multiple DNA repair pathways to repair interstrand crosslinks, while certain enzymes are needed to repair or elongate shortened and/or damaged telomeres (TR). Translesion synthesis (TLS) refers to the use of specialized polymerases that allow for DNA replication past DNA lesions. In addition, many enzymes are involved in the remodeling of chromatin (CR) 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 have been demonstrated in the lungs of subjects with COPD.[2, 16, 17] However, a systematic characterization of DNA repair mechanisms in COPD is lacking. We hypothesized that severe COPD is associated with an impaired response to DNA damage. To evaluate this hypothesis, we analyzed the expression of genes involved in DNA repair and DNA damage tolerance in lung tissue from patients with COPD to identify differentially expressed genes and pathways associated with severe COPD.

METHODS

We analyzed microarray mRNA expression data performed on lung tissue samples from three independent patient cohorts: Lung Genomics Research Consortium (LGRC), Ohio State University (OSU), and Lung expression quantitative trait loci (eQTL): Basic summary data is provided in **Table 1.** Normalized gene expression values were adjusted for age, smoking status (current, former, never), and gender in the LGRC and Lung eQTL study, but not in the OSU study due to sample size.[18] Complete details describing tissue procurement, cohort characteristics, gene expression normalization, and adherence to institutional review board guidelines have been previously described, and further details are provided in the **Supplemental 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) **(Supplemental Table E1).**[23, 24] In the LGRC, OSU, and Lung eQTL study, we compared the expression of these genes in patients

with severe COPD (GOLD IV) vs nonsevere disease (GOLD I,II) and severe COPD (GOLD IV) vs. control (GOLD 0) using Significance Analysis of Microarrays (SAM).[25, 26] DNA repair genes were included for further analysis if they were differentially expressed in all three cohorts and shared the same direction of effect (FDR < 0.1) **(Supplemental 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 **(Supplemental 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 see **Supplemental Methods**.

	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**
n	93	97	27	45	9	13	10	389	389	57
Sex (male)	41 (44.1)	61 (62.9)	16 (59.3)	18 (40.0)	4 (44.4)	6 (46.2)	6 (60.0)	188 (48.3)	245 (63.0)	16 (28.1)
Age	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
Ever-smoker	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)
Pack-years	22±33 [9]	57±40	49±43	49±23	26±17	40±23	56±36	40±23 [37]	50±28 [23]	33±16 [2]
Coexisting Malignancy	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)
Percent emphysema	0.3±0.7	6.3±8.2	22.8±16.0	37.6±13.4						
DLCO % predicted	84.9±16.5 [11]	67.1±20.2 [5]	42.6±13.0 [2]	31.8±10.0 [6]						
FEV1 % predicted	95.2±12.3	65.1±13.3	34.5±5.7	20.2±3.9						
SGRQ score	12.5±16.1 [11]	22.4±18.1 [16]	40.2±19.8	61.3±12.5						
BODE index	0.8±1.2	1.5±1.4	5.2±1.8	6.8±1.3						
6MWD (meters)	437±118 [34]	403±35 [16]	325±97 [5]	277±88 [2]						
SF12 score	48±11 [11]	45±11 [1]	37±10	29±7						

Table 1: Demographic characteristics of study patients. Data are expressed as n (%) or mean ± standard deviation unless otherwise stated [] represents missing samples.
Definition of abbreviations: FEV_1 – Forced expiratory volume in one second. DLCO - diffusing capacity of the lungs for carbon monoxide (DLCO). SGRQ - St. George's Respiratory Questionnaire, BODE - body mass index, airflow obstruction, dyspnea, and exercise capacity, 6-minute walk distance (6MWD), and SF-12 - Short Form Healthy Survey-12.

RESULTS

A DNA Repair Signature of 15 Genes is Associated with Severe COPD

We analyzed 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 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, nonsevere COPD and controls in the three cohorts **(Supplemental 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 characterize 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: percent emphysema based on high resolution computed tomography (HRCT), forced expiratory volume in one second (FEV_1) percent predicted, diffusing capacity for carbon monoxide (DLCO) percent predicted, 6-minute walk testing (6MWD), St. George's Respiratory Questionnaire (SGRQ), body mass index, airflow obstruction, dyspnea, and exercise capacity index (BODE), and the Short Form Healthy Survey-12 (SF-12). Compared to patients in Cluster 2 and Cluster 3, patients in Cluster 1 (n=65) had milder disease, characterized by less emphysema, less impairment in DLCO, and increased FEV_1 **(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 Cluster 2 and severe Cluster 3. There were no differences in gender, pack-years, race, or average age amongst all three clusters (**Supplemental 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 characterized by increased disease severity.

Global Gene Expression Profiling of DNA Repair Clusters in COPD

To characterize 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 differentially expressed genes (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 **(Supplemental Table E5).** The top enriched pathways for both Cluster 1 and Cluster 3 were related to cytokine signaling. In Cluster 1, several interleukin pathways (IL-1, IL-3, IL-5, IL-6, IL-17 and IL-18) were among the top 10 enriched pathways. Similarly, Cluster 3 also showed enrichment of interleukin pathways (IL-3, IL5, IL-10 and IL-17), among the top 10 enriched pathways. The most enriched pathway in both Cluster 1 and Cluster 3 was IL-5, however with the opposite direction of effect; Cluster 1 showed downregulation of genes in the IL-5 pathway and Cluster 3 showed upregulation of genes in the IL-5 pathway. In contrast to Cluster 1 and Cluster 3, Cluster 2 was characterized by upregulation of several pathways involved in cell adhesion and cytoskeletal remodeling, including TGF-β and WNT pathways. The most significant DEGs

amongst the top 50 signaling pathways, for Cluster 2 and Cluster 3 are shown in **(Figure 4)**. This pathway enrichment data suggests Cluster 2 is associated with increased expression of genes involved in tissue remodeling, and Cluster 3 is associated with increased expression of genes involved with inflammation.

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. In particular, 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 the severe Cluster 3 patients when compared to samples from the mild Cluster 1 patients. Notably, DDB2 staining was increased in patients with a history of smoking, compared to never smokers. This data suggests 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 **(Supplemental Table E6A and Figures 6A-D)**. Second, a Z-score analysis was performed using the transcriptomic profiles of 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 measurements of COPD severity. **(Supplemental Table E6B and Figures 6E-6H)**. In both methods, the DR pathway was the only one that showed a positive correlation with clinical measurements of COPD severity, however, this was the smallest pathway (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. 40 modules of co-expressed gene were identified, 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 **(Supplemental 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 percent emphysema (correlation=-0.4,p=1e-07), BODE index (correlation=-0.4,p=4e-08), FEV_1 percent predicted (correlation=0.34,p=5e-06), SGRQ (correlation=-0.43,p=4e-09), DLCO percent predicted (correlation=0.38,p=3e-07), and 6MWD (correlation=0.37,p=8e-07) Within the yellow module were multiple canonical NER genes that demonstrated both high module membership and gene significance for clinical indices of COPD severity, including Xeroderma Pigmentosum Group A-Complementing Protein (XPA) and ERCC5 (Excision Repair Cross-Complementation Group 5) **(Supplemental Figure E8)**. The combination of these three approaches demonstrated that down-regulation of the NER pathway was associated with COPD severity.

DISCUSSION:

In this study, we identified 15 differentially expressed genes 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 remodeling. Our data also suggests that multiple DDRT pathways are downregulated in patients with COPD, with the strongest evidence being demonstrated for the NER pathway. Taken together, this data supports the hypothesis that diminished DNA repair may underlie 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 as 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 shows decreased expression of these genes in severe COPD [32] There are many potential reasons for such differences pertinent to the pathogenesis of COPD including histone modifications, dysregulation of homeostatic signaling 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 may contribute 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 characterized by mild clinical disease. The severe Cluster 2 showed enrichment for pathways associated with cytoskeletal remodeling, including TGF-β and WNT signaling, while the severe Cluster 3 showed enrichment for NF-κB, IL-5, and IL-17 pathways. Excess inflammation and aberrant remodeling are well described mechanisms of COPD pathogenesis. The relationship between DNA damage and chronic inflammation is well described, as defective DNA repair contributes to autoimmunity and chronic inflammation, while WNT and TGF-β are well described regulators of the DNA damage response. [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 remodeling pathways.

We applied multiple methods to characterize 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 appear differentially expressed between Cluster 2 and Cluster 3, including NEIL1, DDB2, and MMS19 are implicated in NER. Furthermore, both DDB2 and NEIL1 were confirmed to be decreased by immunohistochemistry in the severe Cluster 3. This is significant for COPD pathogenesis as the NER pathway is primarily responsible for detecting and removing bulky DNA-adducts caused by CS, and therefore critical for protecting against tobacco-induced carcinogenesis. Previous studies have demonstrated an impairment of NER capacity

with CS, and diminished NER capacity has been implicated as a risk 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, and importantly, almost all observed associations between COPD severity and DDRT pathways suggested that down regulation of genes involved in DNA repair and DNA damage tolerance occur in severe COPD.

There are certain limitations to our study. The influence of co-existing 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 as lung tissue is not commonly obtained from patients with normal lung function, unless there is a concern for cancer. To account for the potential influence of a "field of cancerization", 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 by various cell types. To address this concern, we performed immunohistochemistry of various DNA repair proteins and identified decreased DDB2 and NEIL1 in severe Cluster 3 lung tissue samples. We did not see profound differences amongst clusters when analyzing 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, etc.), cell type, and nuclear colocalizaiton.

We used a multi-step, 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 characterized by clinical differences of severity and distinct non-DNA repair gene pathways associated with increased inflammation and tissue remodeling. 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.

Acknowledgments

The authors would like to acknowledge M.S. (K08HL135402-01), M.S. FAMRI (YCSA 142017), J.G. FAMRI (YCSA 113393), and J.G. (K01HL125474-03). The authors would like to thank the staff at the Respiratory Health Network Tissue Bank of the FRQS for their valuable assistance with the lung eQTL dataset at Laval University. M.L. was the recipient of a doctoral studentship from the Fonds de recherche Québec - Santé (FRQS). Y.B. holds a Canada Research Chair in Genomics of Heart and Lung Diseases.

Supplemental Methods

I. Lung Genomics Research Consortium (LGRC): Gene expression profiles were obtained using the Agilent-014850 Whole Human Genome Microarray 4x44K G4112F-Probe number version (Agilent, Agilent Technologies, Santa Clara, CA, USA). Normalized gene expression values were adjusted for age, smoking status (current, former, never), and gender. Individuals with known interstitial lung disease and alpha-1 antitrypsin deficiency were excluded. The Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo/) accession number for this study is GSE47460.[20, 22]

II. Ohio State University (OSU): We retrieved normalized mRNA expression microarray data from the GEO GSE38974 dataset.[19] Briefly, gene expression profiles were generated with the Agilent-014850 Whole Human Genome Microarray 4x44K G4112F-Feature number version array (Agilent, Agilent Technologies, Santa Clara, CA, USA). Gene expression was not adjusted for covariates due to the small sample size.

III. Lung expression quantitative trait loci (eQTL): Lung tissue samples were collected at three sites: Laval University (Quebec, Canada), University of British-Columbia (Vancouver, Canada) and Groningen University (Groningen, The Netherlands). Gene expression profiles were obtained using a custom Affymetrix microarray (GPL10379) (Affymetrix, Santa Clara, CA, USA).[21] Normalized gene expression data from these three sites were combined using the ComBat adjustment methods and were used for analyses.[18] RMA expression values were adjusted for age, smoking status (current, former, never), and gender, in the same manner as in the LGRC cohort. Individuals with known interstitial lung disease and alpha-1 antitrypsin deficiency were excluded. The GEO accession numbers for this study is GSE23546.

DNA Repair Pathways: We identified 419 DNA damage repair and tolerance genes (DDRT), in the online databases REPAIRTOIRE (www.repairtoire.com), GO Pathways (www.gopathways.com), DNA repair and chromatin remodeling genes (www.dnarepairgenes.com), and the Wood laboratory website (https://sciencepark.mdanderson.org/labs/wood/DNA_Repair_Genes.html). Entrez IDs for all 419 DNA repair genes were mapped to the three-gene expression platform used and assigned to one of 10 categories of DDRT pathways: DR, BER, MMR, NER, HR, NHEJ, TLS, FA, CR, and TR. [24, 25] **(Supplemental Table E1)**.

Identification of a consensus DNA repair gene list: In the OSU, Lung eQTL, and LGRC cohorts, patients with severe COPD (GOLD IV) were compared with patients with nonsevere disease (GOLD I,II) and control (GOLD 0). All genes were ranked based on Significant Analysis of Microarray (SAM) score (d). SAM analysis was performed using BRB Array Tools v 4.1.[25, 26] DNA repair genes were included for further analysis if they were differentially expressed in all three cohorts and shared the same direction of effect (FDR < 0.1).

DNA repair gene validation with RNAseq: We chose to validate the consensus genes using on a subset of lung tissue samples from the LGRC cohort that underwent gene expression profiling by RNA sequencing. (**Supplemental Table E3)**. Complete details have been previously described.[30] Briefly lung tissue samples were sequenced on the Illumina GAIIx. Samples were aligned with TopHat to hg19. Gene expression was quantified using Cufflinks and $\log_2$ transformed FPKM gene expression values were used for analysis. Genes were considered valid

if they were differentially expressed between severe COPD (GOLD IV) and control (GOLD 0) or severe COPD (GOLD IV) and nonsevere disease (GOLD I,II). The RNAseq data is available for download (https://www.lung-genomics.org/research/).

K-means of LGRC samples by the 15-DDRT gene list: Cluster 3.0 software was used for K-means clustering of patients with COPD from the LGRC cohort (GOLD I-IV) based on the 15 consensus genes (bonsai.hgc.jp/~mdehoon/software/cluster/software.htm). To justify the number of clusters: five models were evaluated using different numbers of clusters between 1 and 5, and the best number of clusters was determined by their ability to capture patients with discrete subgroups of DDRT genes. These clusters were used to evaluate the clinical and genome-wide expression differences between DNA repair expression clusters.

Pathway Enrichment Analysis for DNA repair clusters: Genome-wide mRNA expression differences were evaluated in the three DNA repair clusters of patients identified in the LGRC cohort. Pairwise comparisons using the unpaired t-test were performed between individuals in each cluster and a control cluster from the LGRC (GOLD 0) using Genespring version 12.6 (Agilent Technologies, Santa Clara, CA, USA). Transcripts with ≥ 1.2-fold change between conditions were selected for pathway enrichment analyses with MetaCore version 6.23 build 67496 (Thomson Reuters, New York, NY, USA). Pathways with a FDR < 0.05 were considered significant.

Immunohistochemistry: Deidentified formalin-fixed paraffin-embedded tissues from a subset of LGRC patient samples used for microarray expression profiling. For IHC, sections were incubated with rabbit IgG directed against Endonuclease 8-like 1 (NEIL1) (HPA054084**,** Sigma-Aldrich, St. Louis, MO, USA) X-ray repair cross-complementing protein 4 (XRCC4), (ab97351, Abcam, Cambridge, United Kingdom), and DNA damage-binding protein 2 (DDB2) (HPA058406, Sigma-Aldrich, St. Louis, MO). Expose Rabbit specific HR/DAB detection IHC kit was used to detect the primary antibody per protocol (Abcam, Cambridge, United Kingdom). Sections were counterstained with hematoxylin. Images were photographed with a Nikon DS-Ri2 microscope, using a 40x objective. Blinded comparison studies of at least 5 immunohistochemistry samples from each cluster and controls were used to assess for differences in tissue staining.

Gene Set Enrichment Analysis (GSEA): Gene set enrichment of the 10 DDRT pathway gene sets were performed using GSEA v3.0, using 1000 permutations (http://www.broad.mit.edu/gsea). [28] Gene ranking was based on Spearman correlations with clinical measurements of COPD severity amongst patients with COPD (GOLD I-IV), in the LGRC cohort. Clinical measurements of disease included: percent emphysema based on high resolution computed tomography (HRCT), forced expiratory volume in one second (FEV_1) percent predicted, diffusing capacity for carbon monoxide (DLCO) percent predicted, 6-minute walk distance (6MWD), St. George's Respiratory Questionnaire (SGRQ), body mass index, airflow obstruction, dyspnea, and exercise capacity index (BODE), and the Short Form Healthy Survey-12 (SF-12).

DNA Repair Pathway Expression Coefficients (Z-Score): Amongst patients with COPD (GOLD I-IV) in the LGRC cohort, we correlated the expression of genes within a given DDRT pathway with clinical features of COPD. Z-scores were generated for each of the 419 DNA repair genes across all COPD samples. An average Z-score value for the DDRT genes within each of the 10 pathways were used to generate a unique coefficient for all patients.[29] Spearman correlation analyses between the pathway coefficients and clinical features of disease were performed.

Weighted Gene Co-expression Network analysis (WGCNA): WGCNA version 1.42 was used to identify gene co-expression networks. [30] Using the whole transcriptome microarray data from LGRC patients, we identified genes with expression profiles that correlated across sample, and grouped those genes into gene modules. Every module is represented by an eigengene, and each module's eigengene was correlated with clinical traits. For each gene in a module, module membership values were generated, representing the similarity between an individual's gene expression and the module's eigengene. Metacore Process networks were used for module enrichment analyses. Process networks with a FDR <0.05 were considered significant.

Statistical Analysis: Basic summary measures were calculated: medians, means, and standard errors for continuous variables and counts and percentages for categorical variables as appropriate. Parametric data were compared with a students' t-test, nonparametric data were compared by Mann-Whitney, and categorical data were compared with a χ^2 statistic. D'agostino and Pearson test was used to determine if data was normally distributed. Unless otherwise mentioned, two-sided *p* values less than 0.05 were considered significant. Graphs and basic statistical comparisons were performed with GraphPad (GraphPad Software, La Jolla, CA, USA).

References

1. Collaborators GBDCRD. 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(9): 691-706.
2. Caramori G, Adcock IM, Casolari P, Ito K, Jazrawi E, Tsaprouni L, Villetti G, Civelli M, Carnini C, Chung KF, Barnes PJ, Papi A. Unbalanced oxidant-induced DNA damage and repair in COPD: a link towards lung cancer. *Thorax* 2011: 66(6): 521-527.
3. Anderson GP, Bozinovski S. Acquired somatic mutations in the molecular pathogenesis of COPD. *Trends Pharmacol Sci* 2003: 24(2): 71-76.
4. Pastukh VM, Zhang L, Ruchko MV, Gorodnya O, Bardwell GC, Tuder RM, Gillespie MN. 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, Nagai A. DNA damage as a molecular link in the pathogenesis of COPD in smokers. *Eur Respir J* 2012: 39(6): 1368-1376.
6. Aguilera-Aguirre L, Hosoki K, Bacsi A, Radak Z, Sur S, Hegde ML, Tian B, Saavedra-Molina A, Brasier AR, Ba X, Boldogh I. 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, Kim H, Wen F, Abe S, Fang Q, Wang X, Hashimoto M, Bitterman P, Rennard SI. Cigarette smoke extract induces DNA damage but not apoptosis in human bronchial epithelial cells. *Am J Respir Cell Mol Biol* 2005: 33(2): 121-129.
8. Nakayama T, Kaneko M, Kodama M, Nagata C. Cigarette smoke induces DNA single-strand breaks in human cells. *Nature* 1985: 314(6010): 462-464.
9. Maluf SW, Mergener M, Dalcanale L, Costa CC, Pollo T, Kayser M, da Silva LB, Pra D, Teixeira PJ. DNA damage in peripheral blood of patients with chronic obstructive pulmonary disease (COPD). *Mutat Res* 2007: 626(1-2): 180-184.
10. Savale L, Chaouat A, Bastuji-Garin S, Marcos E, Boyer L, Maitre B, Sarni M, Housset B, Weitzenblum E, Matrat M, Le Corvoisier P, Rideau D, Boczkowski J, Dubois-Rande JL, Chouaid C, Adnot S. Shortened telomeres in circulating leukocytes of patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2009: 179(7): 566-571.
11. Rodier F, Coppe JP, Patil CK, Hoeijmakers WA, Munoz DP, Raza SR, Freund A, Campeau E, Davalos AR, Campisi J. Persistent DNA damage signalling triggers senescence-associated inflammatory cytokine secretion. *Nat Cell Biol* 2009: 11(8): 973-979.
12. Friedberg EC. DNA damage and repair. *Nature* 2003: 421(6921): 436-440.
13. Jackson SP, Bartek J. The DNA-damage response in human biology and disease. *Nature* 2009: 461(7267): 1071-1078.

14. Ciccia A, Elledge SJ. The DNA damage response: making it safe to play with knives. *Mol Cell* 2010: 40(2): 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, Charlier CF, Ellwanger JH, Moura DJ, Possuelo LG, Valim AR, Guecheva TN, Henriques JA. 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(1): 118-127.
19. Ezzie ME, Crawford M, Cho JH, Orellana R, Zhang S, Gelinas R, Batte K, Yu L, Nuovo G, Galas D, Diaz P, Wang K, Nana-Sinkam SP. Gene expression networks in COPD: microRNA and mRNA regulation. *Thorax* 2012: 67(2): 122-131.
20. Bauer Y, Tedrow J, de Bernard S, Birker-Robaczewska M, Gibson KF, Guardela BJ, Hess P, Klenk A, Lindell KO, Poirey S, Renault B, Rey M, Weber E, Nayler O, Kaminski N. A novel genomic signature with translational significance for human idiopathic pulmonary fibrosis. *Am J Respir Cell Mol Biol* 2015: 52(2): 217-231.
21. Hao K, Bosse Y, Nickle DC, Pare PD, Postma DS, Laviolette M, Sandford A, Hackett TL, Daley D, Hogg JC, Elliott WM, Couture C, Lamontagne M, Brandsma CA, van den Berge M, Koppelman G, Reicin AS, Nicholson DW, Malkov V, Derry JM, Suver C, Tsou JA, Kulkarni A, Zhang C, Vessey R, Opiteck GJ, Curtis SP, Timens W, Sin DD. Lung eQTLs to help reveal the molecular underpinnings of asthma. *PLoS Genet* 2012: 8(11): e1003029.
22. Yang IV, Pedersen BS, Rabinovich E, Hennessy CE, Davidson EJ, Murphy E, Guardela BJ, Tedrow JR, Zhang Y, Singh MK, Correll M, Schwarz MI, Geraci M, Sciurba FC, Quackenbush J, Spira A, Kaminski N, Schwartz DA. Relationship of DNA methylation and gene expression in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2014: 190(11): 1263-1272.
23. Kassambara A, Gourzones-Dmitriev C, Sahota S, Reme T, Moreaux J, Goldschmidt H, Constantinou A, Pasero P, Hose D, Klein B. A DNA repair pathway score predicts survival in human multiple myeloma: the potential for therapeutic strategy. *Oncotarget* 2014: 5(9): 2487-2498.
24. Mjelle R, Hegre SA, Aas PA, Slupphaug G, Drablos F, Saetrom P, Krokan HE. 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 U S A* 2001: 98(9): 5116-5121.
26. Reiner A, Yekutieli D, Benjamini Y. Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics* 2003: 19(3): 368-375.
27. Kusko RL, Brothers Ii JF, Tedrow J, Pandit K, Huleihel L, Perdomo C, Liu G, Juan-Guardela B, Kass D, Zhang S, Lenburg M, Martinez F, Quackenbush J, Sciurba F, Limper A, Geraci M, Yang I, Schwartz DA, Beane J, Spira A, Kaminski N. Integrated Genomics Reveals Convergent Transcriptomic Networks Underlying COPD and IPF. *Am J Respir Crit Care Med* 2016.
28. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-

based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005: 102(43): 15545-15550.
29. Cheadle C, Vawter MP, Freed WJ, Becker KG. Analysis of microarray data using Z score transformation. *J Mol Diagn* 2003: 5(2): 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, Erzurum SC, Gaston BM, Jarjour NN, Meyers DA, Milosevic J, Tedrow JR, Wu W, Kaminski N, Wenzel SE. Gene Expression Correlated with Severe Asthma Characteristics Reveals Heterogeneous Mechanisms of Severe Disease. *Am J Respir Crit Care Med* 2017: 195(11): 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(18): 8403-8420.
33. Chaisaingmongkol J, Popanda O, Warta R, Dyckhoff G, Herpel E, Geiselhart L, Claus R, Lasitschka F, Campos B, Oakes CC, Bermejo JL, Herold-Mende C, Plass C, Schmezer P. Epigenetic screen of human DNA repair genes identifies aberrant promoter methylation of NEIL1 in head and neck squamous cell carcinoma. *Oncogene* 2012: 31(49): 5108-5116.
34. Kidane D, Chae WJ, Czochor J, Eckert KA, Glazer PM, Bothwell AL, Sweasy JB. Interplay between DNA repair and inflammation, and the link to cancer. *Crit Rev Biochem Mol Biol* 2014: 49(2): 116-139.
35. Alder JK, Barkauskas CE, Limjunyawong N, Stanley SE, Kembou F, Tuder RM, Hogan BL, Mitzner W, Armanios M. Telomere dysfunction causes alveolar stem cell failure. *Proc Natl Acad Sci U S A* 2015: 112(16): 5099-5104.
36. Planchard D, Domont J, Taranchon E, Monnet I, Tredaniel J, Caliandro R, Validire P, Besse B, Soria JC, Fouret P. The NER proteins are differentially expressed in ever smokers and in never smokers with lung adenocarcinoma. *Ann Oncol* 2009: 20(7): 1257-1263.
37. Li W, Hu J, Adebali O, Adar S, Yang Y, Chiou YY, Sancar A. Human genome-wide repair map of DNA damage caused by the cigarette smoke carcinogen benzo[a]pyrene. *Proc Natl Acad Sci U S A* 2017: 114(26): 6752-6757.
38. Sears CR, Zhou H, Justice MJ, Fisher AJ, Saliba J, Lamb I, Wicker J, Schweitzer KS, Petrache I. Xeroderma Pigmentosum Group C Deficiency Alters Cigarette Smoke DNA Damage Cell Fate and Accelerates Emphysema Development. *Am J Respir Cell Mol Biol* 2018: 58(3): 402-411.
39. Spira A, Beane JE, Shah V, Steiling K, Liu G, Schembri F, Gilman S, Dumas YM, Calner P, Sebastiani P, Sridhar S, Beamis J, Lamb C, Anderson T, Gerry N, Keane J, Lenburg ME, Brody JS. Airway epithelial gene expression in the diagnostic evaluation of smokers with suspect lung cancer. *Nat Med* 2007: 13(3): 361-366.
40. Korde A, Jin L, Zhang JG, Ramaswamy A, Hu B, Kolahian S, Juan Guardela B, Herazo-Maya J, Siegfried JM, Stabile L, Pisani MA, Herbst RS, Kaminski N, Elias JA, Puchalski JT, Takyar SS. Lung Endothelial MicroRNA-1 Regulates Tumor Growth and Angiogenesis. *Am J Respir Crit Care Med* 2017.

	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
n	93	97	27	45	9	13	10	389	389	57
Sex (male)	41 (44.1)	61 (62.9)	16 (59.3)	18 (40.0)	4 (44.4)	6 (46.2)	6 (60.0)	188 (48.3)	245 (63.0)	16 (28.1)
Age	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
Ever-smoker	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)
Pack-years Mean (SD)	22±33 [9]	57±40	49±43	49±23	26±17	40±23	56±36	40±23 [37]	50±28 [23]	33±16 [2]
Coexisting Malignancy	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)
Percent emphysema	0.3±0.7	6.3±8.2	22.8±16.0	37.6±13.4						
DLCO % predicted	84.9±16.5 [11]	67.1±20.2 [5]	42.6±13.0 [2]	31.8±10.0 [6]						
FEV1	95.2±12.3	65.1±13.3	34.5±5.7	20.2±3.9						
SGRQ score	12.5±16.1 [11]	22.4±18.1 [16]	40.2±19.8	61.3±12.5						
BODE index	0.8±1.2	1.5±1.4	5.2±1.8	6.8±1.3						
6-minute walk (meters)	437±118 [34]	403±35 [16]	325±97 [5]	277±88 [2]						
SF12 score	48±11 [11]	45±11 [1]	37±10	29±7						

Figure 1

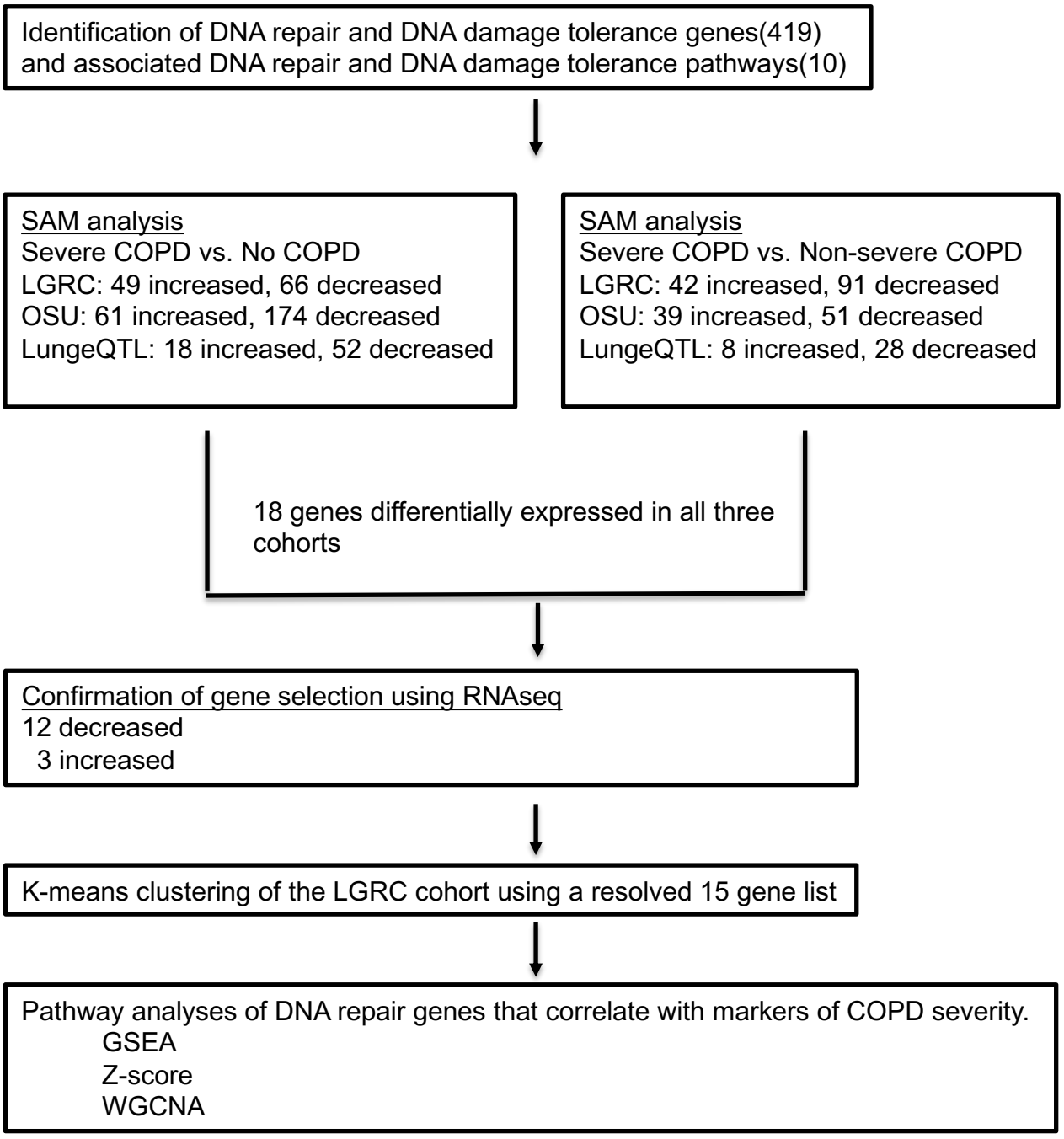


Figure 1: Study Workflow.

Figure 2

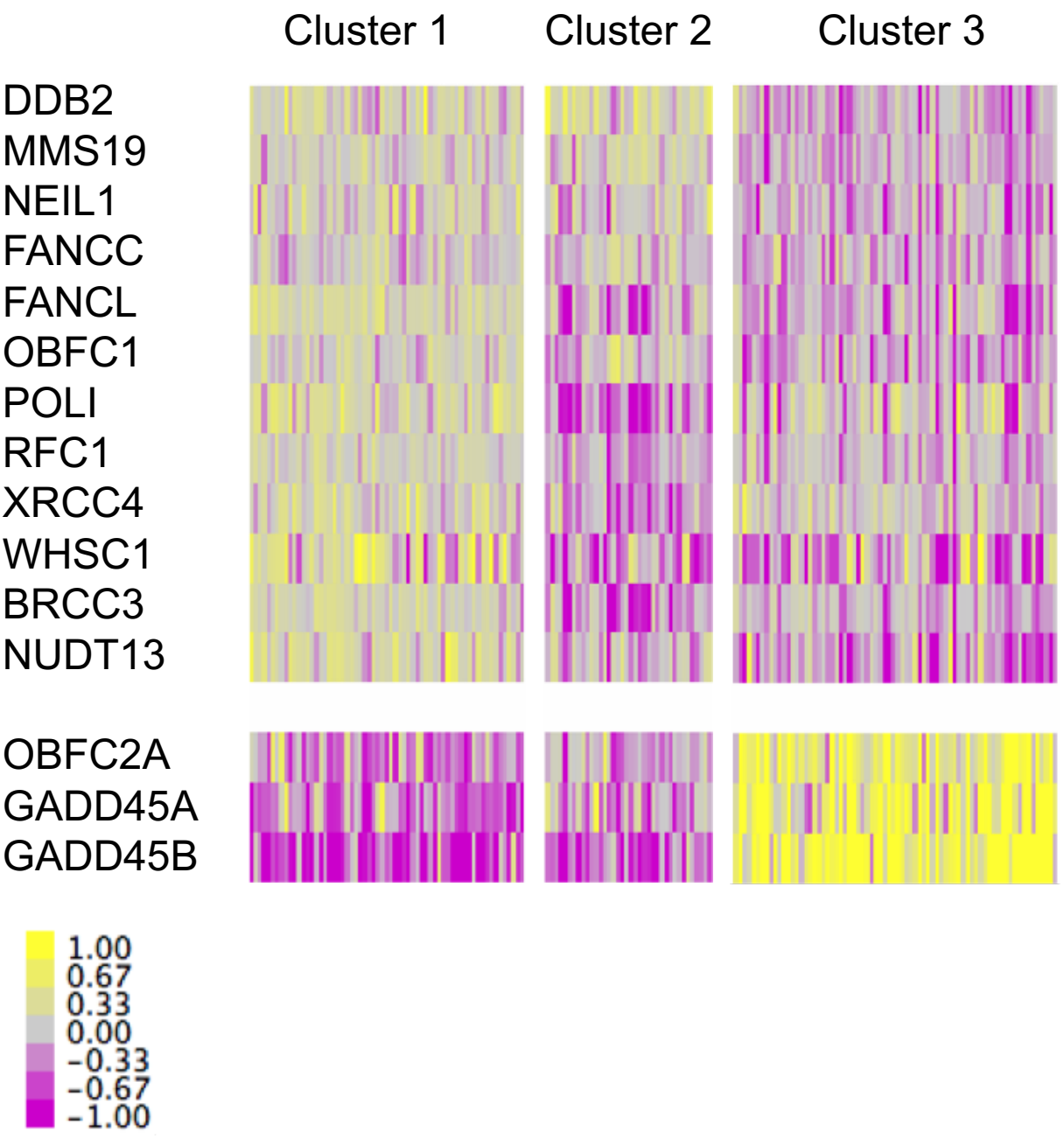


Figure 2: K-means clustering of patients based on 15 gene consensus signature. Yellow denotes an increase over the sample mean, and purple denotes a decrease over sample mean.

Figure 3

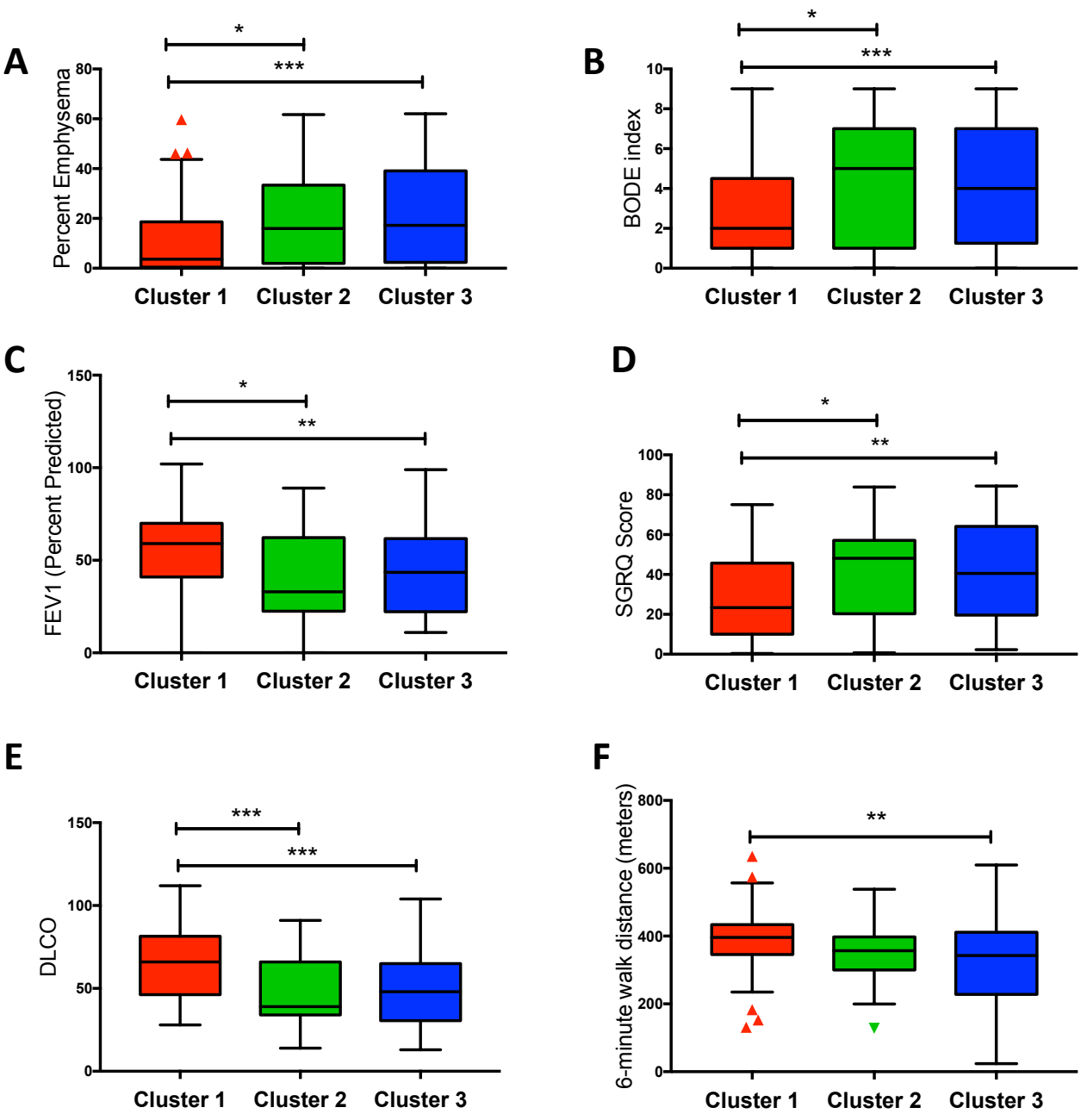
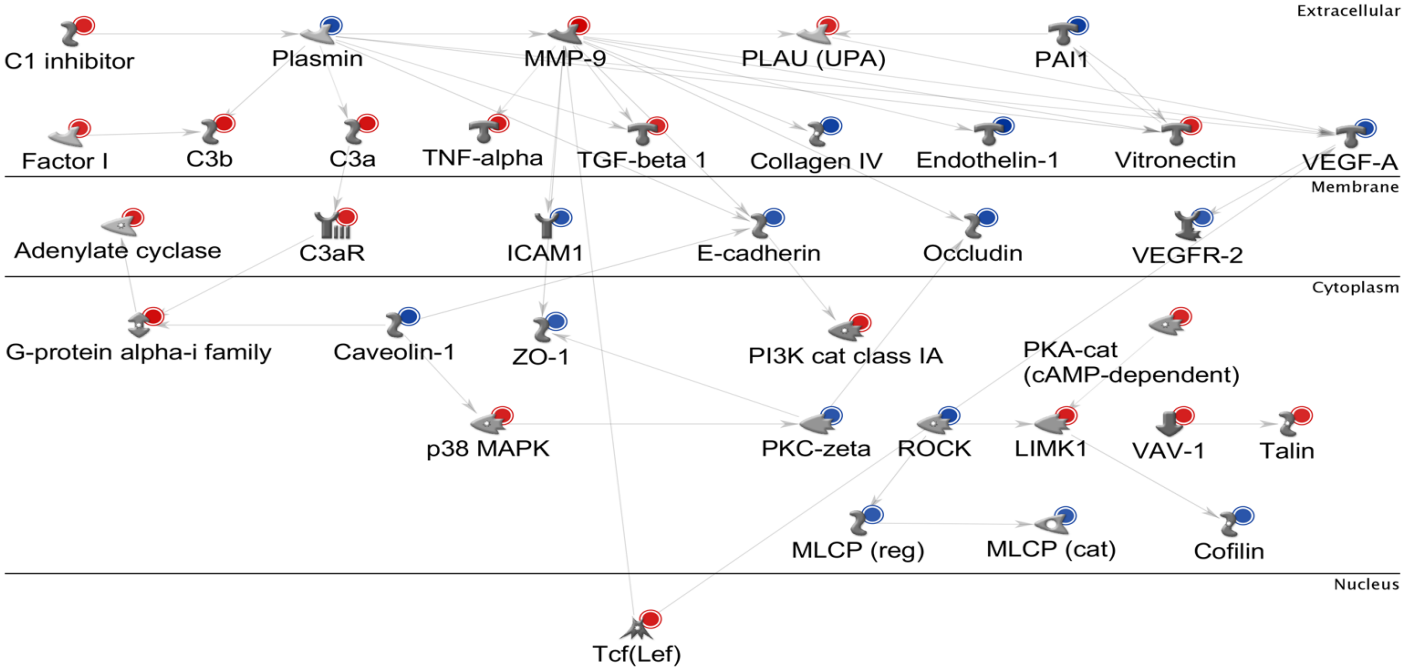


Figure 3: Clinical characteristics of COPD by cluster. A) Box and whiskers of percent emphysema by cluster. **B)** Box and whiskers of BODE index by cluster. **C)** Box and whiskers of FEV_1 percent predicted by cluster. **D)** Box and whiskers of SGRQ by cluster. **E)** Box and whiskers of DLCO percent predicted by cluster. **F)** Box and whiskers of 6-minute walk distance (meters) by cluster. (*** $p<0.0005$, ** $p<0.005$, * $p<0.05$)

Figure 4

A Cluster 2



B Cluster 3

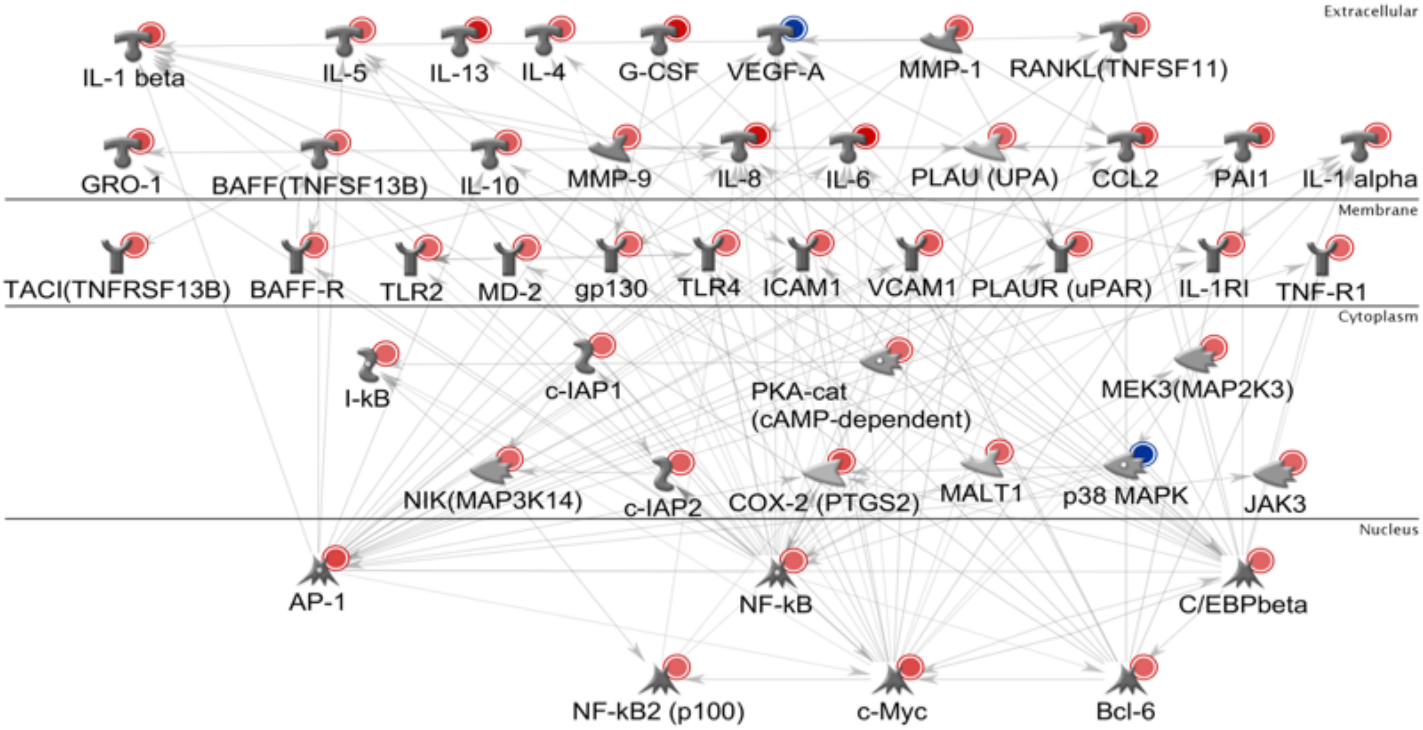


Figure 4: Overrepresented genes in the Top 50 enriched pathways. A) Comparison between Cluster 2 and controls. B) Comparison between Cluster 3 and controls. Red denotes upregulated genes. Blue denotes downregulated genes. Lines represent curated associations between genes.

Figure 5

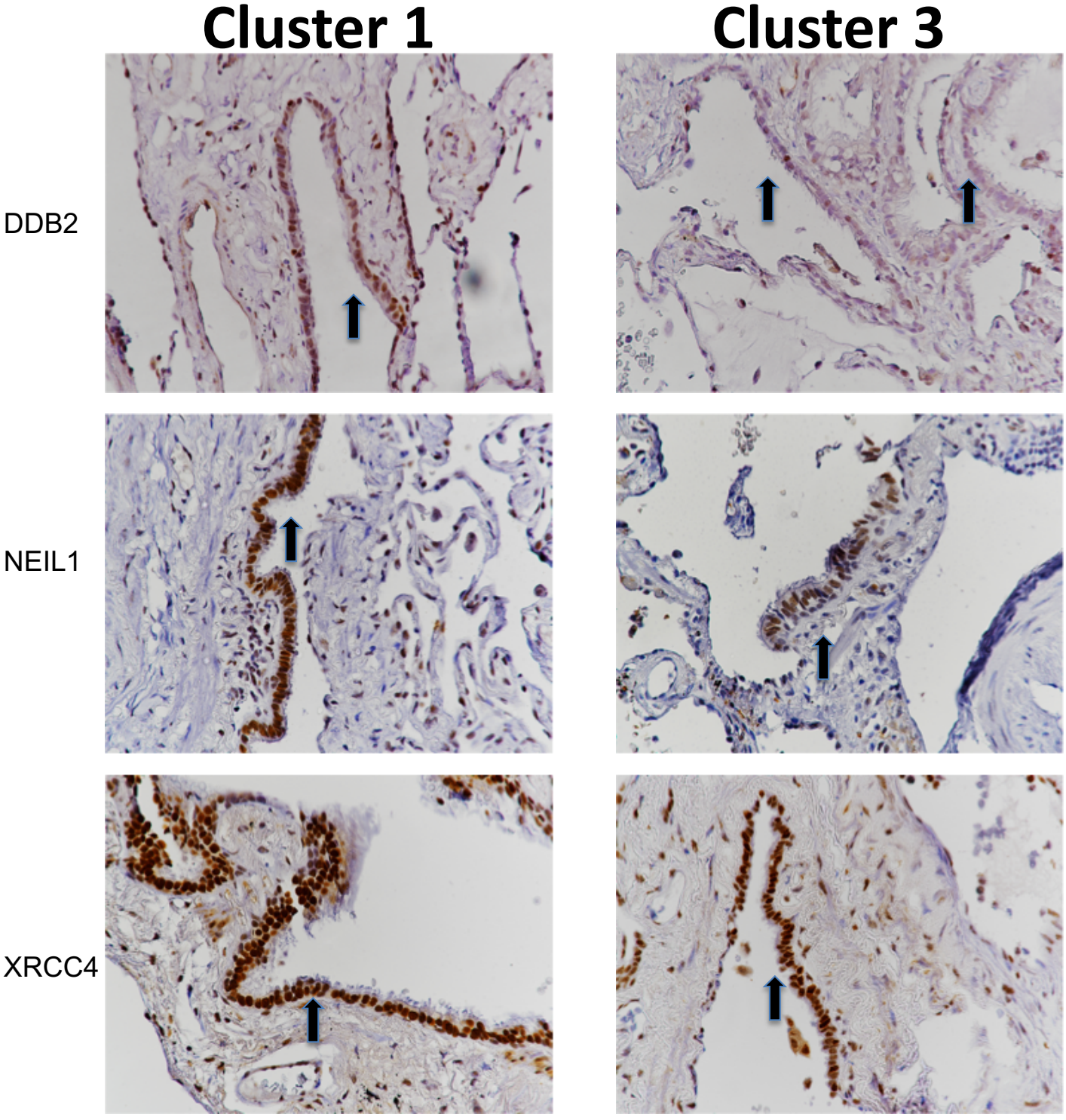


Figure 5: Immunohistochemistry for DDB2, NEIL1, and XRCC4. Immunohistochemistry demonstrating localization and staining intensity for DDB2, NEIL1, and XRCC4 (identified by brown chromogen) performed on lung tissue samples from patients in Cluster 1 and Cluster 3. Nuclear staining appeared particularly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

Figure 6

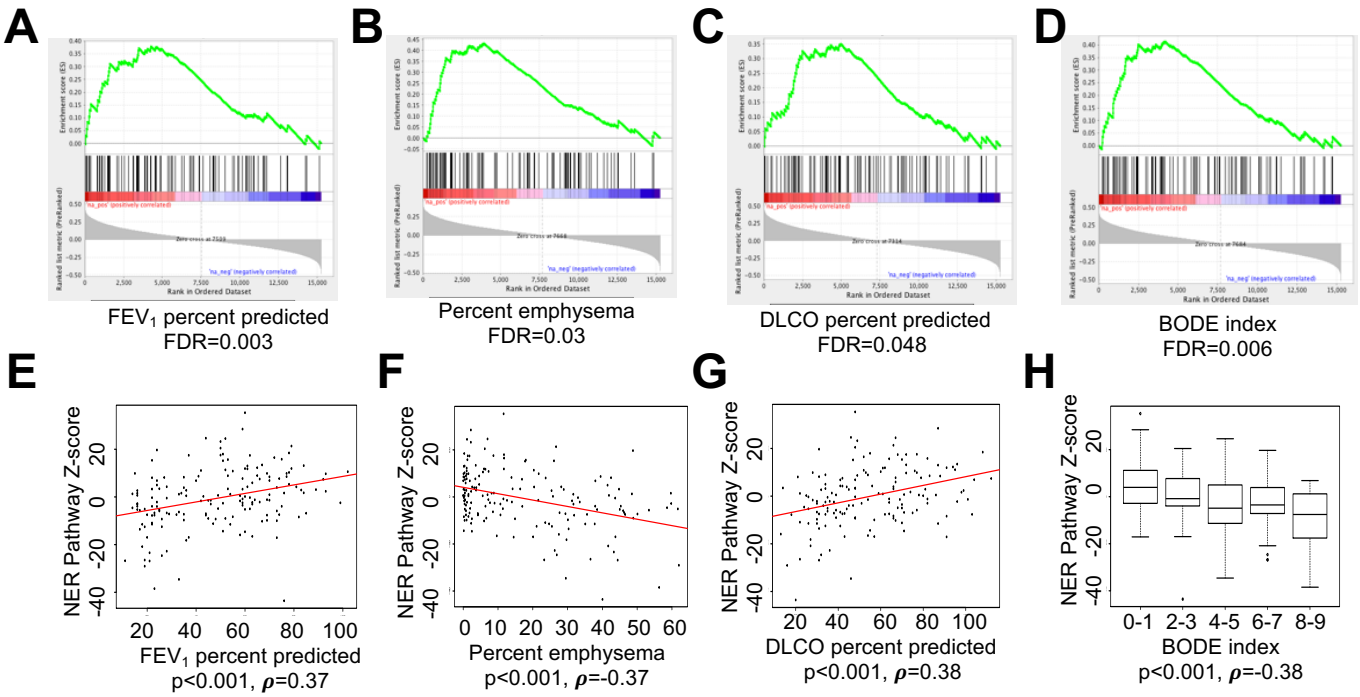


Figure 6: The Nucleotide Excision Repair (NER) pathway is downregulated in severe COPD. **A-D)** Enrichment plots from gene set enrichment analysis (GSEA).The enrichment plots contain profiles of the running enrichment scores (ES) and the barcode plot indicates the position of the genes in each gene set; red represents Spearman correlations with more severe disease, blue represents Spearman correlations with less severe disease. FDRs for NER gene set enrichment are reported for **A)** FEV_1 **B)** percent emphysema, **C)** DLCO percent predicted, and **D)** BODE Index. **E-H)** NER pathway Z-score coefficients for each patient plotted against **E)** FEV_1 percent predicted **F)** percent emphysema, **G)** DLCO percent predicted, and **H)** BODE Index.

Figure 7

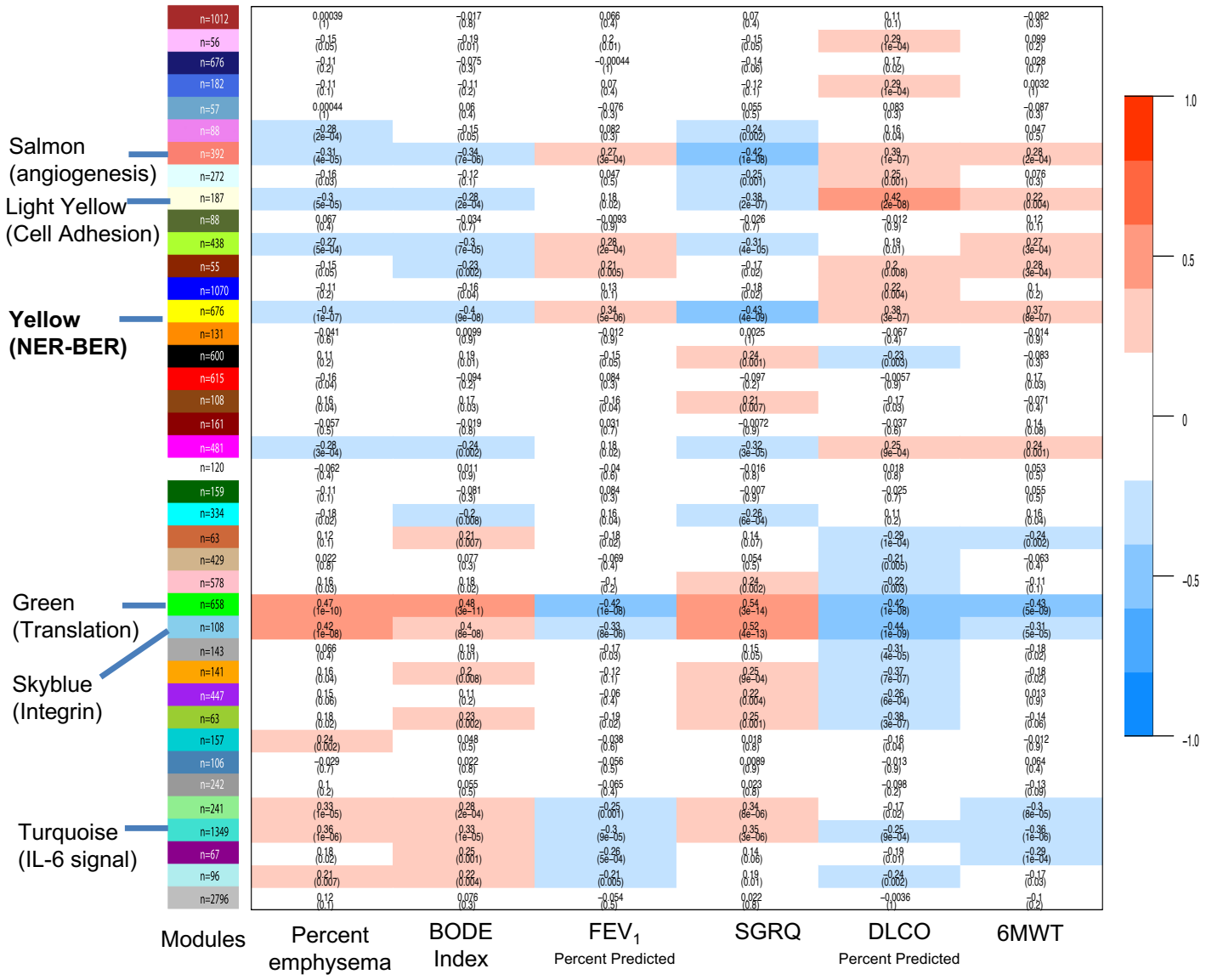


Figure 7: Weighted gene co-expression network analysis (WGCA). WGCNA identified 40 gene modules as demonstrated in this WGCNA heatmap. n represents the number of genes within each module. Positive correlations are red, and negative correlations are blue. The Green, Skyblue, and Turquoise were the three modules most positively correlated with indices of decreased disease severity, and the Yellow, Salmon, and Lightyellow 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 6 modules, including the NER-BER process in the Yellow module.

Table E1

Fanconi Anemia	
Entrez ID	Symbol
6233	RPS27A
7311	UBA52
378708	CENPS
84126	ATRIP
84464	SLX4
199990	FAAP20
1325	CORT
64858	DCLRE1B
64789	EXO5
353497	POLN
2067	ERCC1
91442	FAAP24
2175	FANCA
2187	FANCB
2176	FANCC
2177	FANCD2
2178	FANCE
2188	FANCF
2189	FANCG
55215	FANCI
55120	FANCL
57697	FANCM
27033	ZBTB32
146956	EME1
22909	FAN1
80198	MUS81
64710	NUCKS1
10635	RAD51AP1
5892	RAD51D
165918	RNF168
9025	RNF8
6117	RPA1
6118	RPA2
6119	RPA3
5888	RAD51
9937	DCLRE1A
201254	CENPX
7314	UBB
7316	UBC
29089	UBE2T
7398	USP1
7507	XPA
2072	ERCC4
7517	XRCC3
80233	FAAP100
548593	SLX1A
675	BRCA2
83990	FANCJ
79728	PALB2
5889	RAD51C
672	BRCA1
5810	RAD1

Translesion Synthesis	
Entrez ID	Symbol
6233	RPS27A
7311	UBA52
199990	FAAP20
201973	PRIMPOL
5424	POLD1
5425	POLD2
10714	POLD3
57804	POLD4
5427	POLE2
5429	POLH
11201	POLI
51426	POLK
353497	POLN
5980	REV3L
51514	DTL
7706	TRIM25
9636	ISG15
10459	MAD2L2
55666	NPLOC4
5111	PCNA
25898	RCHY1
51455	REV1
5982	RFC2
5983	RFC3
5984	RFC4
5985	RFC5
5981	RFC1
6117	RPA1
6118	RPA2
6119	RPA3
7353	UFD1L
9100	USP10
124739	USP43
7415	VCP
22890	ZBTB1
9768	KIAA0101
7335	UBE2V1
7336	UBE2V2
27434	POLM
10721	POLQ
55339	WDR33
389898	UBE2NL

Homologous Recombination	
Entrez ID	Symbol
10213	PSMD14
472	ATM
641	BLM
84464	SLX4
672	BRCA1
675	BRCA2
119392	SRR1
126074	SWSAP1
55317	AP5S1
253714	MMS22L
375757	SWI5
8318	CDC45
8317	CDC7
135458	HUS1B
11144	DMC1
3980	LIG3
353497	POLN
10721	POLQ
51659	GINS2
7979	SHFM1
642636	RAD21L1
7334	UBE2N
55218	EXD2
2187	FANCB
84893	FBXO18
2237	FEN1
147872	CCDC155
348654	GEN1
548593	SLX1A
113510	HELQ
3364	HUS1
3014	H2AFX
54617	INO80
23514	SPIDR
9907	AP5Z1
84515	MCM8
254394	MCM9
286053	NSMCE2
4361	MRE11
10933	MORF4L1
4683	NBN
10111	RAD50
22909	FAN1
80198	MUS81
64859	OBCF2A
4796	TONSL
64710	NUCKS1
197370	NSMCE1
220064	ORAOV1
79728	PALB2
142	PARP1
10635	RAD51AP1
6421	SFPQ
5890	RAD51B
5889	RAD51C
5892	RAD51D
25788	RAD54B
8438	RAD54L
5932	RBBP8
5965	RECQL
51444	RNF138
6117	RPA1
6118	RPA2
6119	RPA3
5888	RAD51
5893	RAD52
9400	RECQL5
84296	GINS4
23137	SMC5
79677	SMC6
57599	WDR48
2072	ERCC4
7515	XRCC1
7516	XRCC2
7517	XRCC3
7528	YY1
23503	ZFYVE26
125150	ZSWIM7
146956	EME1
197342	EME2
2139	EYA2
5810	RAD1
5884	RAD17
23049	SMG1
63967	CLSPN
1017	CDK2
79035	NABP2
5536	PPP5C
8243	SMC1A
9126	SMC3
7156	TOP3A
151987	PPP4R2

Mismatch Repair	
Entrez ID	Symbol
3978	LIG1
5424	POLD1
5425	POLD2
10714	POLD3
57804	POLD4
2067	ERCC1
9156	EXO1
29072	SETD2
4292	MLH1
27030	MLH3
4436	MSH2
4437	MSH3
4438	MSH4
4439	MSH5
2956	MSH6
5395	PMS2
5378	PMS1
4595	MUTYH
5111	PCNA
5379	PMS2P1
5380	PMS2P2
5383	PMS2P5
10535	RNASEH2A
6117	RPA1
6118	RPA2
6119	RPA3
6996	TDG
11277	TREX1
25	ABL1
7161	TP73

Base Excision Repair	
Entrez ID	Symbol
328	APEX1
27301	APEX2
3978	LIG1
3980	LIG3
5423	POLB
5424	POLD1
5426	POLE
27343	POLL
10721	POLQ
1763	DNA2
2074	ERCC6
267004	PGBD3
2237	FEN1
3146	HMGB1
8091	HMGA2
3159	HMGA1
10075	HUWE1
8930	MBD4
4350	MPG
4595	MUTYH
79661	NEIL1
252969	NEIL2
55247	NEIL3
4913	NTHL1
4968	OGG1
142	PARP1
10038	PARP2
5111	PCNA
5428	POLG
55170	PRMT6
6117	RPA1
6118	RPA2
6119	RPA3
9401	RECQL4
8034	SLC25A16
23583	SMUG1
51548	SIRT6
6996	TDG
7374	UNG
10309	CCNO
55031	USP47
7486	WRN
7515	XRCC1
200558	APLF
4521	NUDT1
143	PARP4
25961	NUDT13

Non-Homologous End Joining	
Entrez ID	Symbol
10213	PSMD14
472	ATM
64421	DCLRE1C
79184	BRCC3
9577	BRE
580	BARD1
672	BRCA1
78996	C7orf49
286257	C9orf142
84142	FAM175A
79840	NHEJ1
64858	DCLRE1B
3978	LIG1
3980	LIG3
3981	LIG4
5422	POLA1
27343	POLL
27434	POLM
10721	POLQ
5591	PRKDC
7334	UBE2N
22992	KDM2A
8924	HERC2
8290	HIST3H3
29086	BABAM1
3014	H2AFX
8359	HIST1H4A
8366	HIST1H4B
8364	HIST1H4C
8360	HIST1H4D
8367	HIST1H4E
8361	HIST1H4F
8365	HIST1H4H
8294	HIST1H4I
8368	HIST1H4L
2547	XRCC6
7520	XRCC5
4292	MLH1
22976	PAXIP1
7336	UBE2V2
286053	NSMCE2
4361	MRE11
4683	NBN
10111	RAD50
9656	MDC1
27339	PRPF19
10524	KAT5
51588	PIAS4
5720	UIMC1
55183	RIF1
165918	RNF168
9025	RNF8
6419	SETMAR
23137	SMC5
9937	DCLRE1A
7341	SUMO1
7405	UVRAG
7468	WHSC1
7518	XRCC4
91419	ATP23
7158	TP53BP1
5965	RECQL
2072	ERCC4
7486	WRN
2139	EYA2
10432	RBM14
28990	ASTE1
4841	NONO
9232	PTTG1
5536	PPP5C
9126	SMC3

Chromatin Remodeling	
Entrez ID	Symbol
60	ACTB
86	ACTL6A
51412	ACTL6B
57180	ACTR3B
79913	ACTR5
64431	ACTR6
8289	ARID1A
196528	ARID2
29028	ATAD2
546	ATRX
22893	BAHD1
11176	BAZ2A
23476	BRD4
2186	BPTF
25855	BRMS1
10036	CHAF1A
8208	CHAF1B
1105	CHD1
9557	CHD1L
1106	CHD2
1107	CHD3
1108	CHD4
26038	CHD5
84181	CHD6
55636	CHD7
57680	CHD8
80205	CHD9
1487	CTBP1
10664	CTCF
140690	CTCFL
30827	CXXC1
7913	DEK
55929	DMAP1
8193	DPF1
5977	DPF2
8110	DPF3
57634	EP400
2074	ERCC6
1647	GADD45A
4616	GADD45B
10912	GADD45G
3054	HCFC1
3070	HELLS
6596	HLTF
10362	HMG20B
54556	ING3
54617	INO80
83444	INO80B
125476	INO80C
54891	INO80D
283899	INO80E
10524	KAT5
222229	LRWD1
114785	MBD6
9112	MTA1
57504	MTA3
10514	MYBBP1A
4641	MYO1C
4678	NASP
8467	SMARCA5
56916	SMARCAD1
50485	SMARCAL1
6599	SMARCC1
6601	SMARCC2
6602	SMARCD1
6603	SMARCD2
6604	SMARCD3
6605	SMARCE1
10847	SRCAP
29844	TFPT
7153	TOP2A
7155	TOP2B
7161	TP73
6944	VPS72
7528	YY1
11244	ZHX1
10467	ZNHIT1
84083	ZRANB3
8607	RUVBL1
10856	RUVBL2
26122	EPC2
23054	NCOA6
29128	UHRF1
25961	NUDT13
6594	SMARCA1

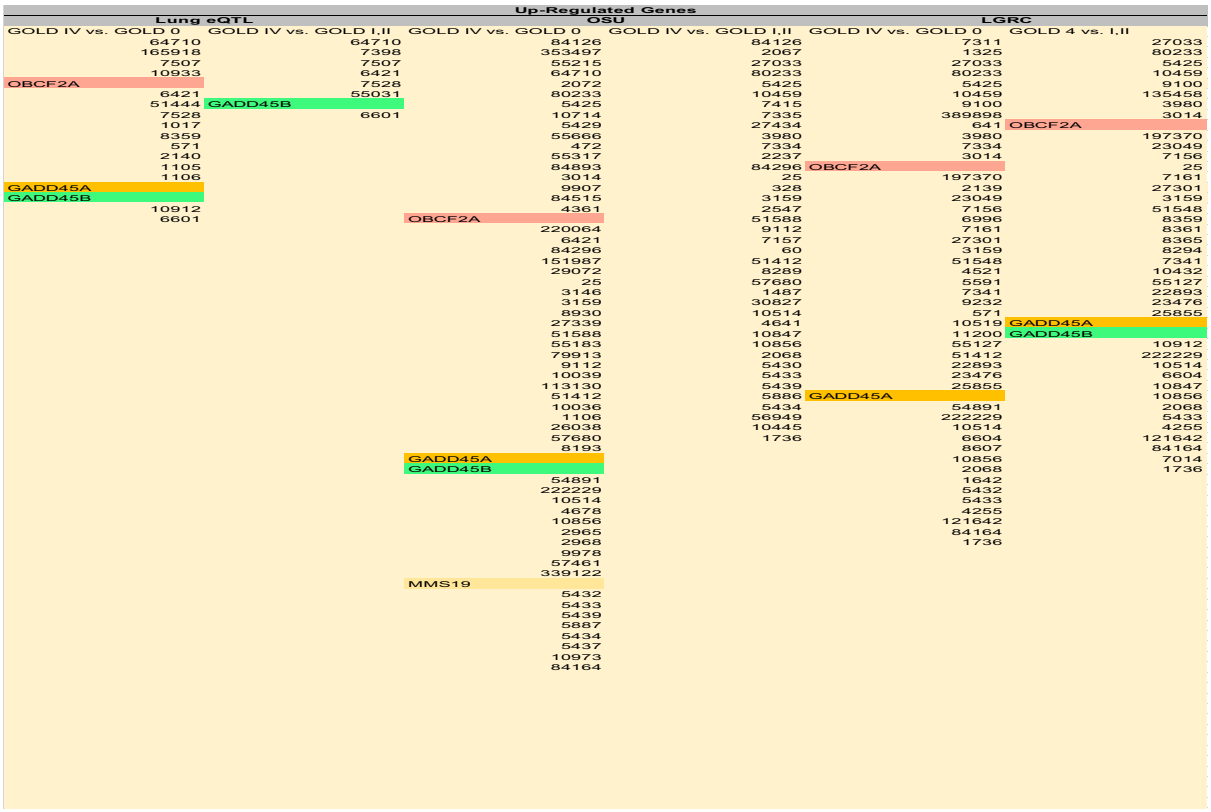
Nucleotide Excision Repair	
Entrez ID	Symbol
6233	RPS27A
7311	UBA52
9716	AQR
54778	RNF111
54841	BIVM
84464	SLX4
675	BRCA2
902	CCNH
1022	CDK7
4331	MNAT1
91603	ZNF830
9557	CHD1L
2873	GPS1
1069	CETN2
135458	HUS1B
2068	ERCC2
2071	ERCC3
2965	GTF2H1
2967	GTF2H3
2968	GTF2H4
404672	GTF2H5
9978	RBX1
8451	CUL4A
8450	CUL4B
1642	DDB1
10450	PPIE
64858	DCLRE1B
1643	DDB2
3978	LIG1
3980	LIG3
3981	LIG4
5423	POLB
5424	POLD1
5425	POLD2
10714	POLD3
57804	POLD4
5426	POLE
51426	POLK
27343	POLL
7329	UBE2I
7334	UBE2N
2067	ERCC1
2074	ERCC6
267004	PGBD3
1161	ERCC8
2176	FANCC
3150	HMGN1
10463	SLC30A9
3364	HUS1
57461	ISY1
339122	RAB43
57654	UVSSA
4287	ATXN3
64210	MMS19
7336	UBE2V2
22909	FAN1
79661	NEIL1
252969	NEIL2
55247	NEIL3
27339	PRPF19
4913	NTHL1
4968	OGG1
220064	ORAOV1
142	PARP1
5111	PCNA
5430	POLR2A
5431	POLR2B
5432	POLR2C
5433	POLR2D
5436	POLR2G
5438	POLR2I
5439	POLR2J
5886	RAD23A
5887	RAD23B
5892	RAD51D
5932	RBBP8
5982	RFC2
5983	RFC3
5984	RFC4
5985	RFC5
5981	RFC1
5434	POLR2E
5435	POLR2F
5437	POLR2H
5440	POLR2K
5441	POLR2L
6117	RPA1
6118	RPA2
6119	RPA3
29935	RPA4
9937	DCLRE1A
7341	SUMO1
6613	SUMO2
6612	SUMO3
23411	SIRT1
6917	TCEA1
7314	UBB
7316	UBC
85015	USP45
7874	USP7
56949	XAB2
7507	XPA
7508	XPC
2072	ERCC4
2073	ERCC5
7515	XRCC1
2033	EP300
7157	TP53
5037	HCNP
93973	ACTR8
8607	RUVBL1
25842	ASF1A
10445	MCRS1
4798	NFRKB

Direct Repair	
Entrez ID	Symbol
4255	MGMT
121642	ALKBH2
221120	ALKBH3
54784	ALKBH5
10973	ASCC3
51008	ASCC1
79068	FTO
84164	ASCC2

Telomere Repair	
Entrez ID	Symbol
80119	PIF1
23049	SMG1
23293	SMG6
79991	OBFC1
54386	TERF2IP
5976	UPF1
7013	TERF1
7014	TERF2
5884	RAD17
1017	CDK2
10728	PTGES3
1736	DKC1
5558	PRIM2
5557	PRIM1
23649	POLA2
80351	TNKS2

Table E1 List of DNA repair pathways genes. *Definition of abbreviations*: BER - Base excision repair, CR - Chromatin remodeling, DR - Direct repair, FA - Fanconi anemia, HR - Homologous recombination, MMR - Mismatch repair, NER - Nucleotide excision repair, NHEJ - Non-homologous end joining, TR - telomere repair, TLS - Translesion synthesis pathway.

Table E2 Upregulated Genes in Severe COPD



Up-Regulated Genes						
Lung eQTL		OSU		LGRC		
GOLD IV vs. GOLD 0	GOLD IV vs. GOLD I,II	GOLD IV vs. GOLD 0	GOLD IV vs. GOLD I,II	GOLD IV vs. GOLD 0	GOLD 4 vs. I,II	
64710	64710	84126	84126	7311	27033	
165918	7398	353497	2067	1325	80233	
7507	7507	55215	27033	27033	5425	
10933	6421	64710	80233	80233	10459	
OBCF2A	7528	2072	5425	5425	9100	
6421	55031	80233	10459	10459	135458	
51444	GADD45B	5425	7415	9100	3980	
7528	6601	10714	7335	389898	3014	
1017		5429	27434	641	OBCF2A	
8359		55666	3980	3980	197370	
571		472	7334	7334	23049	
2140		55317	2237	3014	7156	
1105		84893	84296	OBCF2A	25	
1106		3014	25	197370	7161	
GADD45A		9907	328	2139	27301	
GADD45B		84515	3159	23049	3159	
10912		4361	2547	7156	51548	
6601		OBCF2A	51588	6996	8359	
		220064	9112	7161	8361	
		6421	7157	27301	8365	
		84296	60	3159	8294	
		151987	51412	51548	7341	
		29072	8289	4521	10432	
		25	57680	5591	55127	
		3146	1487	7341	22893	
		3159	30827	9232	23476	
		8930	10514	571	25855	
		27339	4641	10519	GADD45A	
		51588	10847	11200	GADD45B	
		55183	10856	55127	10912	
		79913	2068	51412	222229	
		9112	5430	22893	10514	
		10039	5433	23476	6604	
		113130	5439	25855	10847	
		51412	5886	GADD45A	10856	
		10036	5434	54891	2068	
		1106	56949	222229	5433	
		26038	10445	10514	4255	
		57680	1736	6604	121642	
		8193		8607	84164	
		GADD45A		10856	7014	
		GADD45B		2068	1736	
		54891		1642		
		222229		5432		
		10514		5433		
		4678		4255		
		10856		121642		
		2965		84164		
		2968		1736		
		9978				
		57461				
		339122				
		MMS19				
		5432				
		5433				
		5439				
		5887				
		5434				
		5437				
		10973				
		84164				

Table E2 Differentially expressed genes. Entrez IDs for differentially expressed genes in GOLD IV vs. GOLD 0 and GOLD IV vs. GOLD I, II in the Lung eQTL, LGRC, and OSU cohorts. For the 15 genes that were identified as differentially expressed across three cohorts, Entrez IDs have been replaced with gene symbols and color coded for easy identification.

Table E2 (cont) Downregulated Genes in Severe COPD

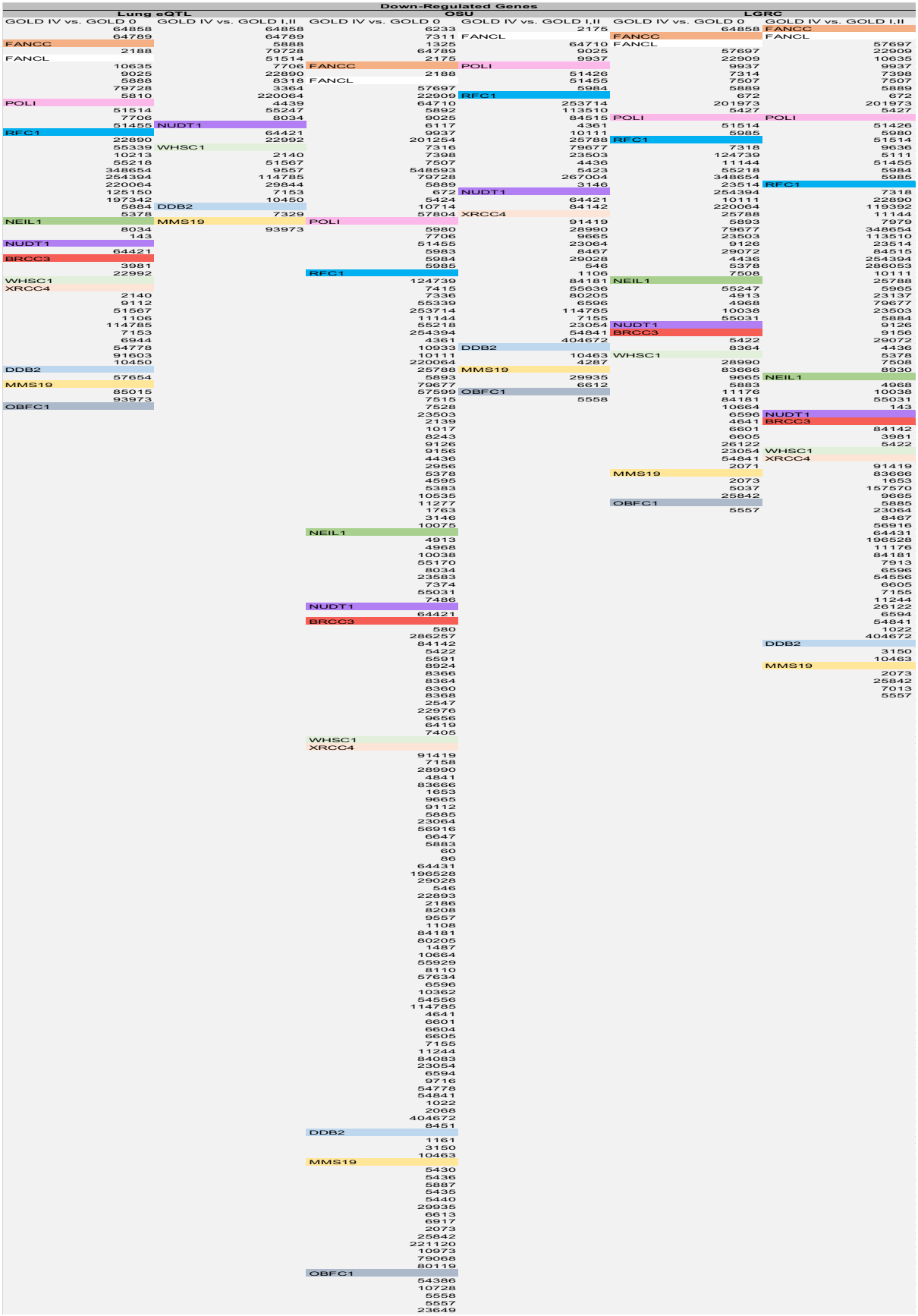


Table E3

	GOLD 0	GOLD I, II	GOLD IV
n	20	14	23
Age (years)	62.7±9.8	70.0±10.3	54.2±6.9
Male Gender	11 (55)	10 (71.4)	16 (69.2)
Pack-years	30±23 [4]	50±23	54±7
Smoking Status (n)	[2]	[2]	
Current	1	2	0
Former	15	10	22
Never	2	0	1
Coexisting Malignancy	19 (95)	13 (93)	0 (0)

Table E3 Demographic characteristics of lung tissue samples used for RNAseq. Data are expressed as n (%) or mean ± standard deviation unless otherwise stated [] represents missing samples.

Table E4

	Cluster 1	Cluster 2	Cluster 3	p-value
n	65	32	72	
Age (years)	66.2 ±10.1	65.2± 8.3	64.9±9.8	0.72
Male Gender	38 (58.4)	18 (56.3)	39 (54.2)	0.88
Pack-years	60±35	50±31	54±38	0.28
Smoking Status (n)				0.98
Current	5	2	4	
Former	58	29	65	
Never	2	1	3	
Coexisting Malignancy	48 (73.8)	14 (43.8)	35 (48.6)	0.003

Table E4 Demographic characteristics of patient clusters. Data are expressed as n (%) or mean ± standard deviation unless otherwise stated.

Table E5

	PATHWAYS	FDR
Cluster 3 vs. Control	Immune response_IL-5 signaling via JAK/STAT	2.58416E-14
	Development_PEDF signaling	1.74796E-08
	Renal tubulointerstitial injury in Lupus Nephritis	1.74796E-08
	Immune response_IL-17 signaling pathways	1.74796E-08
	Immune response_IL-3 signaling via JAK/STAT, p38, JNK and NF-kB	1.74796E-08
	Signal transduction_NF-kB activation pathways	1.74796E-08
	Immune response_IL-10 signaling pathway	2.43031E-08
	Immune response_HSP60 and HSP70/ TLR signaling pathway	4.70008E-08
	Th17 cells in CF	2.98976E-07
	Apoptosis and survival_Anti-apoptotic TNFs/NF-kB/Bcl-2 pathway	8.5043E-07
	Neurogenesis_NGF/ TrkA MAPK-mediated signaling	9.74958E-07
	Cell adhesion_ECM remodeling	1.04107E-06
	Immune response_TLR2 and TLR4 signaling pathways	4.58091E-06
	Development_Regulation of epithelial-to-mesenchymal transition (EMT)	4.91049E-06
	Myeloid-derived suppressor cells and M2 macrophages in cancer	4.91049E-06
	Immune response_HMGB1/RAGE signaling pathway	4.91049E-06
	T follicular helper cell dysfunction in SLE	6.89479E-06
	Apoptosis and survival_APRIL and BAFF signaling	7.88042E-06
	Transcription_Role of VDR in regulation of genes involved in osteoporosis	8.35924E-06
	Rheumatoid arthritis (general schema)	8.35924E-06
	PDE4 regulation of cyto/chemokine expression in inflammatory skin diseases	8.35924E-06
	Immune response_Oncostatin M signaling via JAK-Stat	1.19474E-05
	Role of B cells in SLE	1.60612E-05
	Immune response_Oncostatin M signaling via MAPK	1.93912E-05
	Immune response_IL-6 signaling pathway via JAK/STAT	2.52669E-05
	Glomerular injury in Lupus Nephritis	2.86506E-05
	Development_ERBB-family signaling	3.64516E-05
	Immune response_OX40L/ OX40 signaling pathway	4.92197E-05
	Immune response_B cell antigen receptor (BCR) pathway	5.87773E-05
	Immune response_Role of HMGB1 in dendritic cell maturation and migration	7.06489E-05
	Dysregulation of germinal center response in SLE	7.61744E-05
	Apoptosis and survival_Lymphotoxin-beta receptor signaling	8.41162E-05
	Expression targets of Tissue factor signaling in cancer	8.41162E-05
	Stimulation of TGF-beta signaling in lung cancer	9.00069E-05
	Immune response_IL-18 signaling	9.01493E-05
	Signal transduction_PTMs in BAFF-induced non-canonical NF-kB signaling	9.40188E-05
	Th17 cells in CF (mouse model)	0.000102011
	PDE4 regulation of cyto/chemokine expression in arthritis	0.000102011
	Immune response_Bacterial infections in normal airways	0.000102011
	Role and regulation of Prostaglandin E2 in gastric cancer	0.000102011
	Immune response_MIF-mediated glucocorticoid regulation	0.000111028
	Immune response_TNF-R2 signaling pathways	0.000159031
	Immune response_PGE2 signaling in immune response	0.000159031
	Colorectal cancer (general schema)	0.000250931
	Cell adhesion_Chemokines and adhesion	0.000267804
	Immune response_IL-6-induced acute-phase response in hepatocytes	0.000300058
	Immune response_ICOS signaling pathway in T-helper cell	0.000301549
	PGE2 pathways in cancer	0.000402922
	Apoptosis and survival_Anti-apoptotic TNFs/NF-kB/IAP pathway	0.00050595
	Immune response_Role of PKR in stress-induced antiviral cell response	0.000614013

Table E5 cont.

	PATHWAYS	FDR
Cluster 2 vs. Control	Cytoskeleton remodeling_TGF, WNT and cytoskeletal remodeling	1.85677E-06
	Cytoskeleton remodeling_Cytoskeleton remodeling	2.62659E-05
	Cell adhesion_Integrin inside-out signaling in T cells	6.659E-05
	Immune response_CCL2 signaling	6.659E-05
	Cell adhesion_Histamine H1 receptor signaling in the interruption of cell barrier integrity	6.91542E-05
	Immune response_Classical complement pathway	0.000141904
	Immune response_Alternative complement pathway	0.000141904
	Cell adhesion_Plasmin signaling	0.000206069
	FGF signaling in pancreatic cancer	0.000303336
	Development_MAG-dependent inhibition of neurite outgrowth	0.000360669
	Immune response_C3a signaling	0.000466874
	Cell adhesion_Chemokines and adhesion	0.000466874
	Renal tubulointerstitial injury in Lupus Nephritis	0.000476109
	Signal transduction_IP3 signaling	0.000476109
	Immune response _CCR3 signaling in eosinophils	0.000476109
	Colorectal cancer (general schema)	0.000549706
	Immune response_Lectin induced complement pathway	0.000549706
	Alternative complement cascade disruption in age-related macular degeneration	0.000747639
	Development_Regulation of cytoskeleton proteins in oligodendrocyte differentiation and myelination	0.000968822
	Cytoskeleton remodeling_Regulation of actin cytoskeleton by Rho GTPases	0.001015361
	Cell adhesion_ECM remodeling	0.001015361
	Dysregulation of germinal center response in SLE	0.001085358
	Muscle contraction_GPCRs in the regulation of smooth muscle tone	0.001085358
	Complement pathway disruption in thrombotic microangiopathy	0.001605382
	Cytoskeleton remodeling_Substance P mediated membrane blebbing	0.001637494
	Airway smooth muscle contraction in asthma	0.001652474
	Impaired inhibitory action of lipoxins on neutrophil migration in CF	0.001652474
	B cell signaling in hematological malignancies	0.001652474
	Immune response_T cell subsets: secreted signals	0.001726088
	Development_Regulation of epithelial-to-mesenchymal transition (EMT)	0.002397985
	Cell adhesion_Endothelial cell contacts by junctional mechanisms	0.002397985
	Apoptosis and survival_NGF/ TrkA PI3K-mediated signaling	0.002637393
	Chemotaxis_Inhibitory action of lipoxins on IL-8- and Leukotriene B4-induced neutrophil migration	0.002637393
	Muscle contraction_Regulation of eNOS activity in endothelial cells	0.002642032
	Immune response_B cell antigen receptor (BCR) pathway	0.002642032
	Cell adhesion_Tight junctions	0.002649148
	G-protein signaling_H-RAS regulation pathway	0.002649148
	Muscle contraction_Relaxin signaling pathway	0.002649148
	Neurophysiological process_ACM regulation of nerve impulse	0.002649148
	Cell cycle_Role of Nek in cell cycle regulation	0.00271413
	Cell adhesion_Gap junctions	0.007040667
	Resolution of inflammation in healing myocardial infarction	0.007040667
	Development_Regulation of lung epithelial progenitor cell differentiation	0.007040667
	Oxidative stress_Activation of NADPH oxidase	0.007231906
	ENaC regulation in normal and CF airways	0.007231906
	Glomerular injury in Lupus Nephritis	0.007746357
	Cytoskeleton remodeling_Reverse signaling by Ephrin-B	0.008199773
	Chemotaxis_CCR1 signaling	0.008199773
	Stimulation of TGF-beta signaling in lung cancer	0.008199773
	Development_Role of G-CSF in hematopoietic stem cell mobilization	0.008331283

Table E5 cont.

	PATHWAYS	FDR
Cluster 1 vs. Control	Immune response_IL-5 signaling via JAK/STAT	2.46432E-08
	Neurogenesis_NGF/ TrkA MAPK-mediated signaling	1.80016E-05
	Immune response_IL-6 signaling pathway via JAK/STAT	2.26837E-05
	Immune response_IL-3 signaling via JAK/STAT, p38, JNK and NF-kB	2.26837E-05
	Immune response_IL-18 signaling	3.16719E-05
	Development_Transcription regulation of granulocyte development	3.93554E-05
	Immune response_IL-1 signaling pathway	8.65083E-05
	Reproduction_Gonadotropin-releasing hormone (GnRH) signaling	9.73401E-05
	Immune response_IL-17 signaling pathways	0.000206079
	K-RAS signaling in pancreatic cancer	0.000230174
	Immune response_CD40 signaling	0.000311612
	Immune response_TLR2 and TLR4 signaling pathways	0.000427075
	Development_GM-CSF signaling	0.000427075
	Immune response_HSP60 and HSP70/ TLR signaling pathway	0.00066774
	Immune response_Oncostatin M signaling via MAPK	0.000681542
	Immune response_MIF-mediated glucocorticoid regulation	0.000733887
	Immune response_TLR5, TLR7, TLR8 and TLR9 signaling pathways	0.000777029
	Development_ERBB-family signaling	0.000777029
	Immune response_IL-33 signaling pathway	0.000793299
	Immune response_IL-10 signaling pathway	0.00117069
	Immune response_Substance P-stimulated expression of proinflammatory cytokines via MAPKs	0.00118276
	Renal tubulointerstitial injury in Lupus Nephritis	0.001449424
	Glomerular injury in Lupus Nephritis	0.001933506
	Stimulation of TGF-beta signaling in lung cancer	0.001954887
	Development_PEDF signaling	0.00211156
	Development_EGFR signaling pathway	0.002168513
	Immune response_CCL2 signaling	0.003388931
	Immune response_Platelet activating factor/ PTAFR pathway signaling	0.003496915
	PGE2 pathways in cancer	0.003496915
	Immune response_IL-9 signaling pathway	0.003617684
	Development_YAP/TAZ-mediated co-regulation of transcription	0.003617684
	Transcription_Role of AP-1 in regulation of cellular metabolism	0.00454154
	IGF family signaling in colorectal cancer	0.004981025
	Immune response_Oncostatin M signaling via JAK-Stat	0.005028124
	Myeloid-derived suppressor cells and M2 macrophages in cancer	0.006685753
	Immune response_IL-27 signaling pathway	0.006698825
	T follicular helper cell dysfunction in SLE	0.006698825
	Muscle contraction_Regulation of eNOS activity in endothelial cells	0.006698825
	Development_Leptin signaling via JAK/STAT and MAPK cascades	0.007129372
	Transcription_Transcription regulation of aminoacid metabolism	0.007129372
	Development_Cross-talk between VEGF and Angiopoietin 1 signaling pathways	0.008122061
	Immune response_MIF-induced cell adhesion, migration and angiogenesis	0.008557024
	Development_TGF-beta-dependent induction of EMT via MAPK	0.009237917
	Immune response_Histamine H1 receptor signaling in immune response	0.009953739
	Immune response_CD137 signaling in immune cell	0.011248497
	Development_TGF-beta receptor signaling	0.011248497
	Immune response_C5a signaling	0.011248497
	Development_G-CSF-induced myeloid differentiation	0.011964976
	Immune response_IL-5 signaling via PI3K, MAPK and NF-kB	0.011964976
	Signal transduction_Additional pathways of NF-kB activation (in the cytoplasm)	0.012653182

Table E5 Global gene expression pathways. Analysis of global gene expression pathways associated with Clusters 1, 2, and 3 compared to control patients. The top 50 enriched pathways are listed for each comparison.

Table E6

(A)

	BER		CR		DR		FA		HR		MMR		NER		NHEJ		TR		TLS	
	FDR	NES	FDR	NES	FDR	NES	FDR	NES	FDR	NES	FDR	NES	FDR	NES	FDR	NES	FDR	NES	FDR	NES
FEV1 Percent Predicted	0.14	1.28	0.31	1.09	0.18	-1.2	0.076	1.51	0.13	1.37	0.15	1.26	0.003	1.95	0.12	1.33	0.14	1.33	0.005	1.84
DLCO Percent Predicted	0.57	0.25	0.15	1.31	0.02	-1.61	0.042	1.62	0.18	1.26	0.2	1.22	0.048	1.57	0.55	0.23	0.15	-1.4	0.02	1.73
BODE	0.28	-1.11	0.23	-1.17	0.11	1.27	0.08	-1.43	0.06	-1.5	0.08	-1.4	0.006	-1.86	0.08	-1.38	0.045	-1.56	0.011	-1.88
SF-12 Score	0.37	1.13	0.64	0.65	0.38	-1.04	0.021	1.66	0.1	1.41	0.23	1.25	0.021	1.66	0.48	1.01	0.35	1.11	0.0045	1.92
SGRQ Score	0.67	-0.88	0.21	-1.23	0.22	1.16	0.058	-1.61	0.23	-1.24	0.5	-1	0.058	-1.61	0.39	-1.08	0.15	-1.34	0.092	-1.64
6MWD (meters)	0.16	1.26	0.25	1.13	0.17	-1.21	0.08	1.58	0.04	1.36	0.25	1.15	0.06	1.56	0.14	1.33	0.14	1.29	0.03	1.78
Percent Emphysema	0.13	-1.33	0.38	-1.04	0.04	1.44	0.017	-1.77	0.12	-1.34	0.36	-1.08	0.03	-1.67	0.13	-1.36	0.12	-1.39	0.019	-1.85

Correlation with Features of Decreased COPD severity (FDR)
<0.01
≥0.01 &<0.05
≥0.05 &<0.1
Not Significant
≥0.05 &<0.1
≥0.01 &<0.05
Correlation with Features of Increased COPD severity (FDR)

(B)

	BER		CR		DR		FA		HR		MMR		NER		NHEJ		TR		TLS	
	p-value	ρ	p-value	ρ	p-value	ρ	p-value	ρ	p-value	ρ	p-value	ρ	p-value	ρ	p-value	ρ	p-value	ρ	p-value	ρ
FEV1 Percent Predicted	0.402	0.14	0.888	0.02	0.151	-0.13	0.020	0.20	0.056	0.17	0.044	0.18	<0.001	0.37	0.015	0.21	0.578	0.06	<0.001	0.33
DLCO Percent Predicted	0.523	0.08	0.543	0.12	0.017	-0.23	0.020	0.22	0.079	0.15	0.036	0.20	<0.001	0.38	0.135	0.13	0.204	0.13	<0.001	0.42
BODE	0.484	-0.11	0.888	-0.04	0.152	0.12	0.020	-0.20	0.052	-0.21	0.036	-0.20	<0.001	-0.38	0.015	-0.23	0.117	-0.16	<0.001	-0.37
SF-12 Score	0.580	0.06	0.888	-0.04	0.221	-0.10	0.013	0.23	0.052	0.18	0.053	0.17	<0.001	0.32	0.104	0.13	0.094	0.18	<0.001	0.32
SGRQ Score	0.826	0.02	0.888	0.01	0.006	0.26	0.020	-0.19	0.056	-0.16	0.122	-0.12	<0.001	-0.32	0.046	-0.17	0.094	-0.19	<0.001	-0.31
6MWD(meters)	0.523	0.08	0.888	-0.02	0.120	-0.16	0.020	0.20	0.056	0.17	0.111	0.14	0.002	0.25	0.015	0.22	0.117	0.16	0.001	0.28
Percent Emphysema	0.343	-0.16	0.543	-0.09	0.017	0.21	0.002	-0.28	0.052	-0.18	0.036	-0.20	<0.001	-0.37	0.015	-0.21	0.312	-0.10	<0.001	-0.37

Correlation with Features of Decreased COPD severity (ρ)
≥0.30
≥0.20 & <0.30
<0.20
Not Significant
<0.20
≥0.20 & <0.30
Correlation with Features of Increased COPD severity (ρ)

Table E6 Enrichment analysis. DNA repair pathways associated with markers of COPD severity based on **A)** GSEA (Normalized Enrichment Score (NES) and FDR) and **B)** Z-score (Correlation Coefficient (***ρ***) and p-value). *Definition of abbreviations*: FEV_1 – Forced expiratory volume in one second, DLCO - diffusing capacity of the lungs for carbon monoxide, SGRQ - St. George's Respiratory Questionnaire, BODE - body mass index, airflow obstruction, dyspnea, and exercise capacity , and SF-12 - Short Form Healthy Survey-12, BER - Base excision repair, CR - Chromatin remodeling, DR - Direct repair, FA - Fanconi anemia, HR - Homologous recombination, MMR - Mismatch repair, NER - Nucleotide excision repair, NHEJ - Non-homologous end joining, TR - telomere repair, TLS - Translesion synthesis pathway. 6-minute walk distance – (6MWD)

Table E7

Yellow	**FDR**
DNA damage_BER-NER repair	7.42E-03
Transcription_Chromatin modification	1.64E-02
DNA damage_DBS repair	1.88E-02
Salmon	
Development_Regulation of angiogenesis	1.68E-03
Transcription_Chromatin modification	8.39E-02
Light Yellow	
Cell adhesion_Amyloid proteins	5.31E-04
Cell adhesion_Cadherins	9.48E-03
Development_Neurogenesis_Synaptogenesis	3.72E-02
Cell adhesion_Cell junctions	3.87E-02
Signal transduction_WNT signaling	
Green	
Translation_Translation initiation	1.92E-02
Sky Blue	
Cell adhesion_Integrin-mediated cell-matrix adhesion	5.79E-02
Turqoise	
Inflammation_IL-6 signaling	8.32E-05
Cytoskeleton_Intermediate filaments	1.73E-04
Inflammation_Amphoterin signaling	2.35E-04
Apoptosis_Apoptotic nucleus	2.35E-04
Inflammation_Protein C signaling	4.86E-04
Cell cycle_G1-S Interleukin regulation	6.16E-04
Inflammation_Histamine signaling	6.16E-04
Development_Ossification and bone remodeling	6.16E-04
Apoptosis_Death Domain receptors & caspases in apoptosis	6.68E-04
Immune response_TCR signaling	1.13E-03
Immune response_Th17-derived cytokines	1.93E-03
Cell adhesion_Platelet-endothelium-leucocyte interactions	2.37E-03
Inflammation_IgE signaling	2.37E-03
Inflammation_Neutrophil activation	2.37E-03
Inflammation_MIF signaling	2.87E-03
Inflammation_IL-2 signaling	2.87E-03
Inflammation_IL-10 anti-inflammatory response	2.87E-03
Cardiac development_Role of NADPH oxidase and ROS	3.43E-03
Inflammation_TREM1 signaling	3.79E-03
Inflammation_Inflammasome	4.10E-03
Immune response_Innate immune response to RNA viral infection	4.93E-03
Reproduction_Feeding and Neurohormone signaling	5.13E-03
Inflammation_Innate inflammatory response	5.14E-03
Cell cycle_G1-S Growth factor regulation	7.22E-03
Immune response_BCR pathway	8.43E-03
Chemotaxis	9.71E-03
Apoptosis_Apoptotic mitochondria	1.46E-02
Apoptosis_Anti-Apoptosis mediated by external signals via PI3K/AKT	1.95E-02
Cell cycle_G1-S	2.63E-02
Inflammation_NK cell cytotoxicity	2.73E-02
Inflammation_IL-4 signaling	3.22E-02
Inflammation_IL-12,15,18 signaling	3.66E-02

Table E7 Pathway enrichment for weighted gene co-expression network modules (WGCNA). The enriched processes associated with the WGCNA modules most closely associated with clinical indices of COPD severity (FDR <0.05).

Table E8

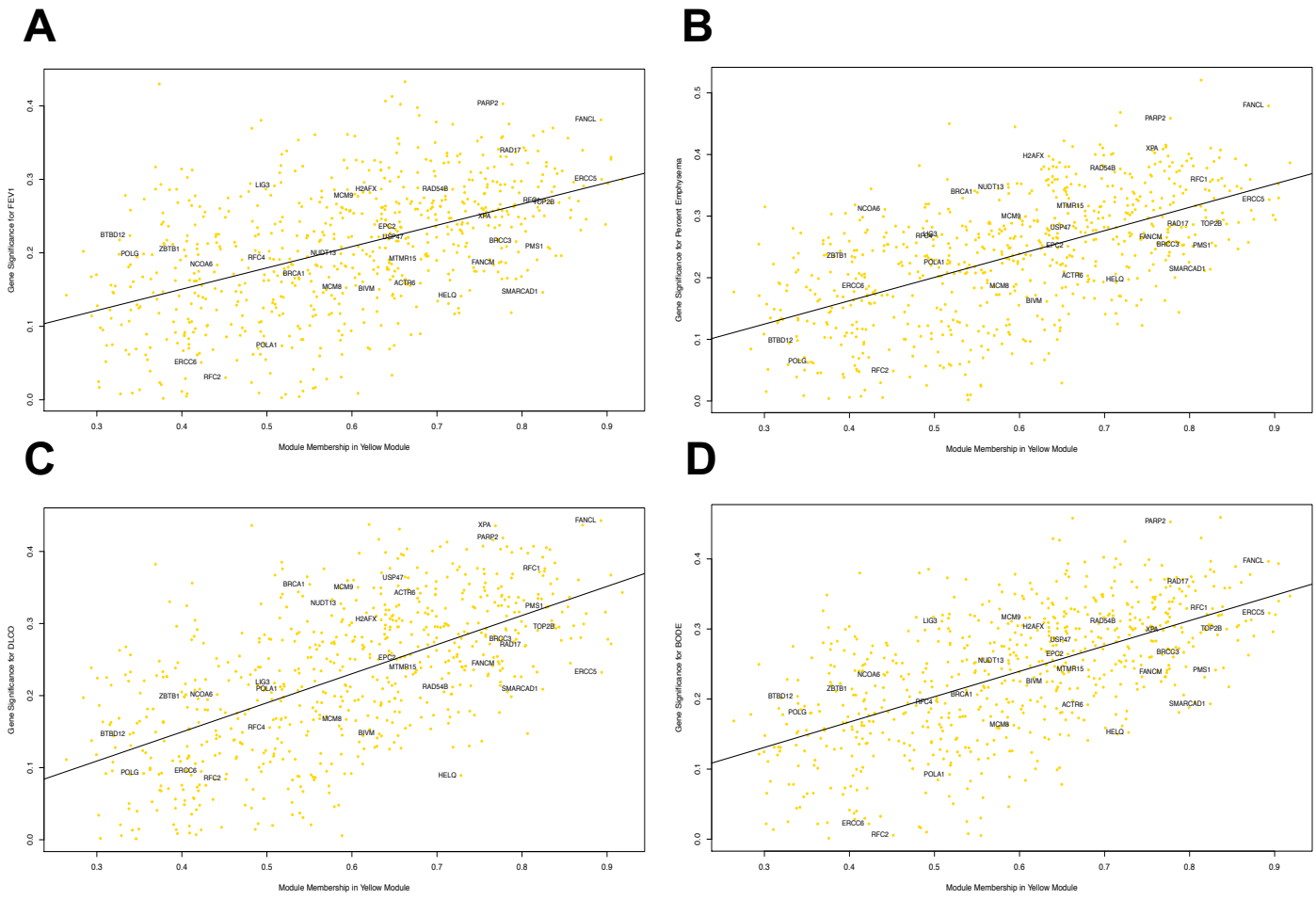


Figure E8 Gene significance (GS) versus module membership (MM) for the Yellow module. Weighted gene co-expression network analysis (WGCNA) derived MM were plotted against GS for the following traits: **A)** FEV_1 percent predicted **B)** percent emphysema **C)** DLCO percent predicted, and **D)** BODE index. All DDRT genes **(Supplemental Table 1)** within the module are identified.