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.