

RESEARCH ARTICLE

Increased Transcript Complexity in Genes Associated with Chronic Obstructive Pulmonary Disease

Lela Lackey, Evonne McArthur, Alain Laederach*

Department of Biology, University of North Carolina, Chapel Hill, North Carolina, United States of America

* alain@unc.edu



Abstract

Genome-wide association studies aim to correlate genotype with phenotype. Many common diseases including Type II diabetes, Alzheimer's, Parkinson's and Chronic Obstructive Pulmonary Disease (COPD) are complex genetic traits with hundreds of different loci that are associated with varied disease risk. Identifying common features in the genes associated with each disease remains a challenge. Furthermore, the role of post-transcriptional regulation, and in particular alternative splicing, is still poorly understood in most multigenic diseases. We therefore compiled comprehensive lists of genes associated with Type II diabetes, Alzheimer's, Parkinson's and COPD in an attempt to identify common features of their corresponding mRNA transcripts within each gene set. The *SERPINA1* gene is a well-recognized genetic risk factor of COPD and it produces 11 transcript variants, which is exceptional for a human gene. This led us to hypothesize that other genes associated with COPD, and complex disorders in general, are highly transcriptionally diverse. We found that COPD-associated genes have a statistically significant enrichment in transcript complexity stemming from a disproportionately high level of alternative splicing, however, Type II Diabetes, Alzheimer's and Parkinson's disease genes were not significantly enriched. We also identified a subset of transcriptionally complex COPD-associated genes (~40%) that are differentially expressed between mild, moderate and severe COPD. Although the genes associated with other lung diseases are not extensively documented, we found preliminary data that idiopathic pulmonary disease genes, but not cystic fibrosis modulators, are also more transcriptionally complex. Interestingly, complex COPD transcripts are more often the product of alternative acceptor site usage. To verify the biological importance of these alternative transcripts, we used RNA-sequencing analyses to determine that COPD-associated genes are frequently expressed in lung and liver tissues and are regulated in a tissue-specific manner. Additionally, many complex COPD-associated genes are spliced differently between COPD and non-COPD patients. Our analysis therefore suggests that post-transcriptional regulation, particularly alternative splicing, is an important feature specific to COPD disease etiology that warrants further investigation.

OPEN ACCESS

Citation: Lackey L, McArthur E, Laederach A (2015) Increased Transcript Complexity in Genes Associated with Chronic Obstructive Pulmonary Disease. PLoS ONE 10(10): e0140885. doi:10.1371/journal.pone.0140885

Editor: Oliver Eickelberg, Helmholtz Zentrum München, GERMANY

Received: June 5, 2015

Accepted: September 30, 2015

Published: October 19, 2015

Copyright: © 2015 Lackey et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Disease-associated gene data are available from the National Human Genome Research Institute catalog of published GWAS (<https://www.genome.gov/26525384>). UniProt-GOA reference human genes from the GO Consortium's Reference Genome initiative (<http://www.ebi.ac.uk/GOA/RGI>). Gene position and annotation data are available from the UCSC Genome browser (hg19) with RefSeq transcript IDs (<https://genome.ucsc.edu/cgi-bin/hgTables>). Microarray data analyzing gene expression changes in COPD patients are available from the Gene Expression Omnibus (GEO) (gds4265). Alternative

Splicing profiles developed with the Illumina BodyMap RNA-seq are available (<http://ccb.jhu.edu/software/ASprofile/data/>). RNA sequencing data from COPD patients and subjects with normal spirometry are available on GEO (GSE57148).

Funding: National Institutes of Health/National Heart Lung and Blood Institute: RO1 HL111527-01. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Introduction

Disease predisposition is likely driven by disturbances at multiple levels of gene regulation. Genetic variation affecting RNA in both the coding and non-coding portions of the genome has the potential to regulate transcript localization, degradation, splicing and translation, which contribute to disease [1–4]. A role for post-transcriptional regulatory processes in multi-genic diseases is supported by the disproportionate number of single nucleotide polymorphisms (SNPs) mapping to non-coding regions of genes that are associated with disease [5]. However, the importance of post-transcriptional regulatory processes in multigenic disease etiologies has yet to be fully investigated. Therefore, we are interested in determining whether transcript complexity at gene loci could reveal underlying molecular mechanisms of specific multi-gene diseases.

The basic premise for this study stems from a simple observation in the *SERPINA1* gene. This gene produces α -1-anti-trypsin, a protein that regulates the proteolytic activity of elastase, and broadly affects the inflammatory response in the lung [6, 7]. Individuals with severe α -1-anti-trypsin deficiency may develop COPD even if they are not smokers [8]. However, genetic disruptions to *SERPINA1* account for less than 5% of COPD cases [9]. Thus, α -1-anti-trypsin deficiency is predictive of COPD in only a small subset of individuals even among smokers. One particularly striking feature of *SERPINA1* is its high transcriptional complexity—i.e. the gene yields eleven alternative transcripts (Fig 1A). Several different types of alternative splicing mechanisms generate this complexity of transcripts, including exon skipping and alternative 5' and 3' splice site usage. The gene is unusual in terms of its alternative splicing—in comparison, 95.5% of loci within a reference set of human genes annotated by RefSeq have less than five transcripts. Thus, *SERPINA1* is in the top 1% in terms of number of transcripts. In addition, *SERPINA1* harbors non-coding SNPs in its 5' untranslated region (UTR) that are associated with COPD, suggesting post-transcriptional processes likely play a role in its regulation [10].

COPD is a lung disease defined as progressive airflow limitation [11]. It is the fourth leading cause of death in the world, and is a major burden for 64 million people globally [12]. Although smoking is a strong environmental factor for the development of COPD, whether or not an individual will develop the disease if he or she smokes is highly dependent on family history [9, 13, 14]. We investigate here whether the high transcriptional complexity of *SERPINA1* (Fig 1A) is a common feature of COPD associated genes and if high transcriptional complexity is a broader characteristic of other complex multigenic diseases.

Results

Identification of disease-associated gene loci

Since the cause of COPD can only be attributed to α -1-anti-trypsin deficiency in ~5% of patients, but many other genomic loci are associated with COPD development and progression, we manually curated a list of COPD associated genes to investigate their transcriptional complexity. First, we combined genes implicated in COPD from recent COPD literature reviews (Fig 1B, left) [14–17]. The subsequent list was merged with search results from the National Human Genome Research Institute (NHGRI) Genome Wide Association Study (GWAS) catalog to yield a total set of 206 putative COPD-associated gene loci (Fig 1B, right, and S1 Table) [18, 19]. Literature reviews contributed the majority of the gene list for COPD, as these represent the best form of manual curation for further analysis. We did not include the PubMed text-miner, Glad4U, and the SNP database, SNP4Disease in our gene list generation, although we did use these resources in a confirmatory manner (S1 Fig, S2 Table) [20, 21].

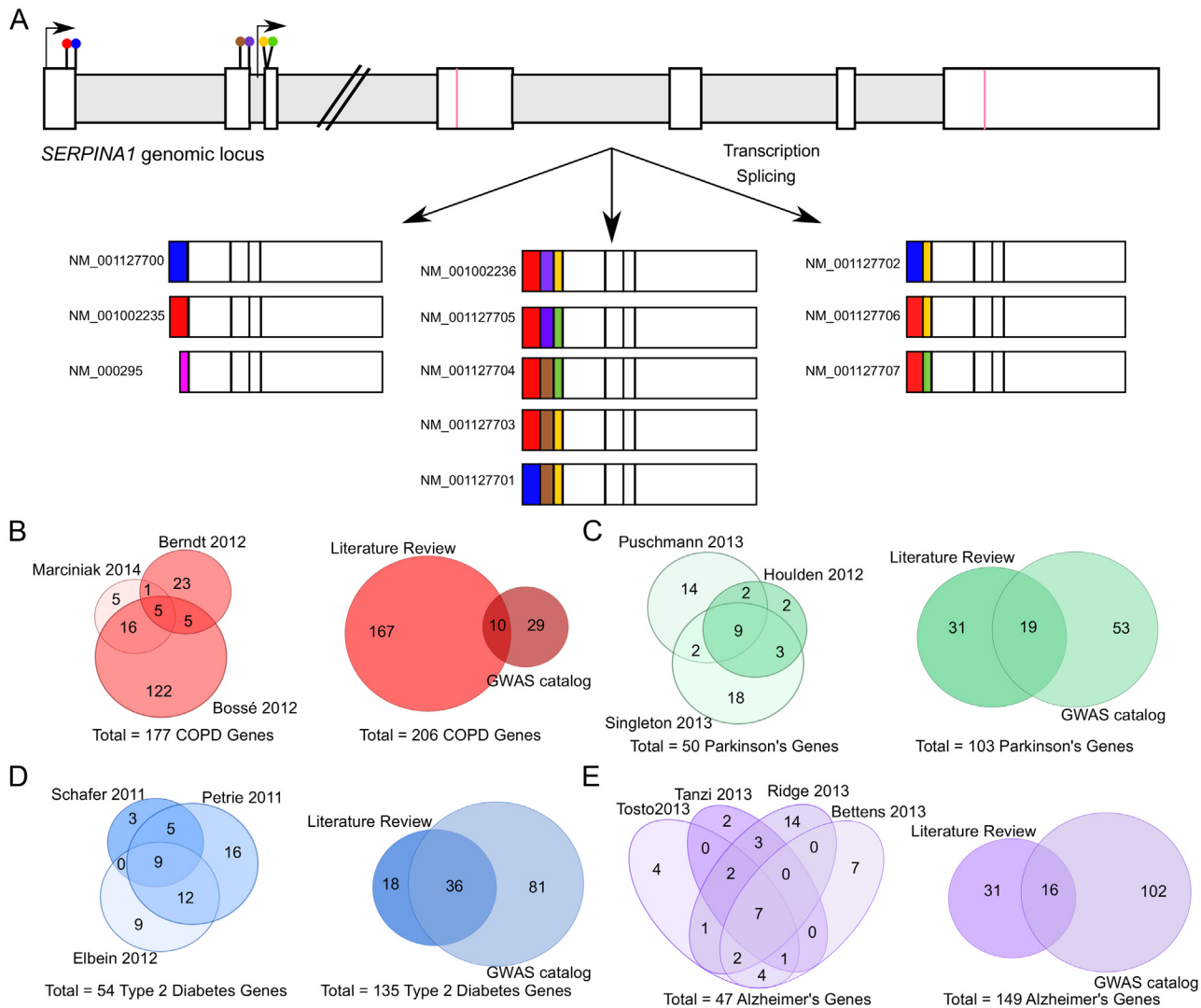


Fig 1. Combining sources to identify genes associated with COPD. (A) The COPD-associated gene, *SERPINA1*, is alternatively spliced to produce 11 different transcripts. Two transcription start sites are indicated by arrows; splice sites are shown as colored lollipops. Exons are indicated by white bars and introns by horizontal gray bars. The coding sequence start and stop codons are indicated with pink lines. 11 transcripts are depicted and colored by splice site selection. The 11 transcripts make *SERPINA1* a particularly complex gene in terms of alternative splicing, 99.5% of human genes have fewer transcripts. (B) COPD-associated genes were identified by merging disease-associated genes from different literature reviews (left) and combining them with genes from the NHGRI GWAS catalog (right). Other comparative disease lists were compiled in the same manner, including (C) Parkinson's disease-associated genes, (D) Type 2 diabetes-associated genes and (E) Alzheimer's disease-associated genes. The number of genes from each source is indicated in the Venn diagrams.

doi:10.1371/journal.pone.0140885.g001

To assess transcript complexity in other complex diseases, we curated similar gene lists for Parkinson's disease (PRK), Type 2 Diabetes (T2D) and Alzheimer's disease (ALZ). These three diseases have different etiologies, but are also complex chronic diseases that include significant genetic components. We followed the same protocol to create lists of genes associated with PRK, T2D and ALZ from recent literature reviews and the NHGRI GWAS resource. We identified 103 PRK gene loci (Fig 1C, S3 Table) [18, 22–25] and 135 T2D gene loci (Fig 1D, S4 Table) [18, 26–29] as well as 149 ALZ gene loci (Fig 1E, S5 Table) [18, 30–35]. Interestingly, although all reviews were within similar time frames, the genes reported were not identical. While the reviews and the GWAS identified genes have more genes in common than would be expected

randomly ([S6](#), [S7](#) and [S8](#) Tables), they also each contribute unique genes for analysis. Our strategy of curating gene lists therefore guarantees broad pools of COPD, T2D, ALZ and PRK-associated genes to investigate alternative transcript usage.

COPD-associated gene loci have a high level of transcriptional complexity

While a majority of gene loci (55.7%) have only one annotated RNA transcript (based on RefSeq annotation), others produce upwards of twenty or thirty different transcripts through alternative splicing and transcription start site usage. These transcripts include non-coding RNAs that regulate their own and other loci and alternatively spliced mRNAs that produce various protein isoforms [36]. We were particularly interested in these high transcript complexity loci as we hypothesize they are indicative of post-transcriptional regulatory mechanisms analogous to SERPINA1 ([Fig 1A](#)). In order to analyze the transcript complexity of the COPD list and our other gene lists, we corrected for the length bias inherent in GWAS studies as well as the number of genes in each list ([S2 Fig](#)). To do so, we generated control gene sets with similar length distributions for each of the disease lists by randomly selecting length-matched genes from a set of annotated human genes [37]. Then, we tested whether these loci-length controlled measurements of transcript complexity would reveal a difference in COPD-associated gene lists versus randomized control lists.

When we performed transcript complexity analysis on our COPD gene list, we discovered that these loci produce significantly more (2.65 transcripts per gene, $p = 0.001$) than the expected number of transcripts from each locus ([Fig 2A](#)). This is unique to COPD-associated genes, as no significant enrichment is observed in comparable PRK, T2D, and ALZ gene lists, although these genes do tend to have higher than average transcript complexity than randomized control lists. To better characterize the bias in COPD-associated loci we categorized the loci by how many transcripts they produced. We found that COPD-associated genes have significantly fewer loci with only one transcript and significantly more loci with greater than 5 transcripts ([Fig 2B](#), $p = 0.005$ and 0.001). The transcript complexity observed in SERPINA1 ([Fig 1A](#)) appears to be a feature of many COPD-associated genes, suggesting that alternative splicing may be an important component of COPD predisposition.

Since our list of COPD-associated genes includes those genes with varying degrees of experimental support, we decided to use patient data to solidify our confidence in our gene list and re-test our hypothesis. As part of the Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoints (ECLIPSE) Study sputum samples from ex-smokers with varying degrees of COPD (level 2, 3 or 4) were collected to compare patient gene expression via microarray with additional independent sputum samples used for PCR confirmation (gds4265) [38]. We identified 85 genes from our original list that are differentially expressed between mild, moderate and severe COPD patients ([Fig 3A](#), [Methods](#)). We confirmed that these differentially expressed COPD genes were more transcriptionally complex than control gene sets ($p = 0.001$) ([Fig 3B](#)). Examples of transcriptionally complex genes, their differential expression values and their connection to COPD are listed in [Table 1](#).

We removed genes that had weaker experimental support to check a subset of 151 COPD-associated genes. These strongly associated COPD genes were significantly more transcriptionally complex than control gene lists ($p = 0.012$) ([S3 Fig](#)). Finally, we determined how much of the transcriptional complexity within each list was due to a few genes by systematically removing the most complex genes from the full COPD list and the top three normalized control lists. We found that the original COPD list remained significantly more enriched for transcriptionally complex genes even with the top three complex genes removed ($p = 0.002$ to $p = 0.017$) ([S3](#)

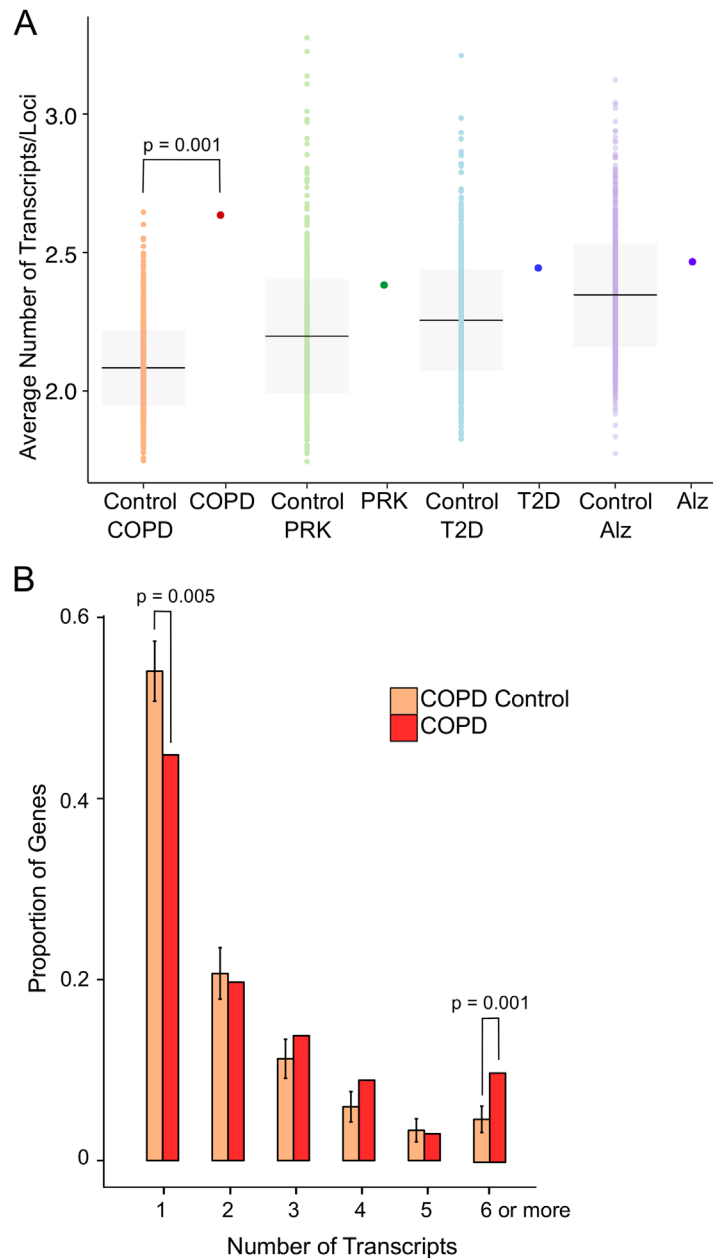


Fig 2. COPD-associated genes are transcriptionally complex. (A) We calculated the average number of transcripts per loci for length-normalized control loci and each disease-associated gene list. The control sets used as comparisons are labeled. The mean and standard deviation of the control lists are shown (black bar, grey box). Only the COPD-associated gene list is significantly different from the normalized control ($p = 0.001$). (B) We analyzed the number of gene loci that produce 1, 2, 3, 4, 5, and 6 or more transcripts and plotted these gene loci by their proportion in each list. The COPD-associated gene set has significantly fewer loci with 1 transcript and significantly more loci with more than 5 transcripts ($p = 0.005$ and $p = 0.001$).

doi:10.1371/journal.pone.0140885.g002

(Fig). This was not true of the control lists—with the top three complex genes removed none of these lists remained significantly transcriptionally complex ($p = 0.231, 0.781$ and 0.093) (S3 Fig). These experiments support our hypothesis that COPD-associated genes are transcriptionally complex.

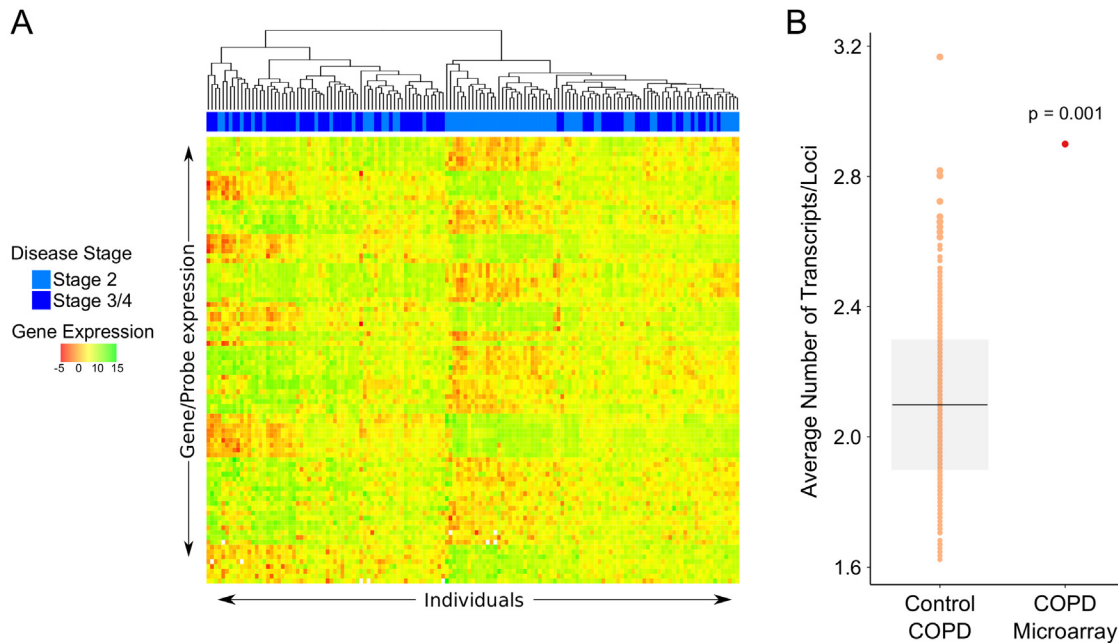


Fig 3. The transcript complexity enrichment of COPD-associated genes is robust. (A) Microarray data from the sputum of ex-smokers diagnosed with COPD stage 2, 3, or 4 identified 85 COPD-associated genes with significant differential expression (gds4265). These 85 genes are shown on the y-axis and individuals are shown on the x-axis with their COPD status marked by disease stage along the top. (B) We calculated the average number of transcripts per loci for these 85 differentially expressed COPD-associated genes along with length-normalized control gene sets. The mean and standard deviation of the control lists are shown (black bar, grey box). These expressed genes are significantly more transcriptionally complex than controls ($p = 0.001$).

doi:10.1371/journal.pone.0140885.g003

Table 1. Transcriptionally complex loci associated with COPD from genes expressed in stage 2 versus stages 3 and 4 COPD patients [38].

HGNC Name	Full Name	RefSeq Isoforms	Stage 2 vs 3 differential expression	Stage2 vs 4 differential expression	adj.P. Val	COPD References	Gene Function
CSF3	Colony-stimulating factor 3	5	1.16	1.02	2.62E-03	[39]	[40], [41], [42]
FAM13A	Family with sequence similarity 13, member A	5	0.35	-0.31	3.70E-03	[43], [44], [45], [46], [47], [48], [49]	[50], [51]
PDE4D	3',5'-cyclic-AMP phosphodiesterase activity and degrades cAMP	9	0.48	0.56	1.00E-04	[52], [53], [54], [55]	[56], [57]
AGER	Advanced glycosylation end product-specific receptor	10	0.49	-0.16	3.80E-02	[58], [59], [60]	[61], [62]
SERPINA1	Serpin peptidase inhibitor, clade A [alpha-1 antiproteinase, antitrypsin], member 1	11	-0.06	-0.15	1.25E-02	[55], [63], [64]	[65], [66]
TP53	Tumor protein p53	15	-0.73	-1.04	3.83E-07	[67], [68], [69], [70]	[71], [72], [73]
FUT8*	Fucosyltransferase 8 [alpha [1,6] fucosyltransferase]	5	-0.41	-0.51	6.52E-03	[74]	[75], [76]
NR3C1*	Nuclear receptor subfamily 3, group C, member 1 [glucocorticoid receptor]	15	0.51	0.27	2.00E-03		[77], [78], [79], [80], [81]
VEGFA*	Vascular endothelial growth factor A	19	0.34	0.42	6.49E-05	[82]	[83]

* Genes not included in the strongly supported COPD gene list

doi:10.1371/journal.pone.0140885.t001

Transcriptional complexity in cystic fibrosis modulatory genes and idiopathic pulmonary fibrosis-associated genes

We chose T2D, ALZ and PRK to compare with COPD transcript complexity since we were able to curate similar sized gene lists for these diseases. Nonetheless, these diseases have very different pathologically affected tissues, so we additionally analyzed gene lists associated with the lung diseases idiopathic pulmonary fibrosis (IPF) and cystic fibrosis (CF). COPD is characterized by progressive airflow limitation while IPF is a progressive fibrosis of the lung that restricts normal breathing [84]. In comparison, cystic fibrosis is caused by mutations in CFTR, resulting in abnormal mucus and impaired lung function. We found 45 genes associated with IPF [18, 85–87] (S9 Table) and 80 genes that modulate CF [18, 88–90] Both IPF associated genes and CF modulators separately share ~31% different COPD-associated genes (S9 and S10 Tables). IPF-associated genes may be more transcriptionally complex since 20% have 6 or more transcripts, while only 8.25% of CF genes have 6 or more transcripts (S8 Fig). Likewise, the average number of transcripts per gene is 2.65 for COPD-associated genes, 2.04 for CF-associated modulators, and 2.75 for IPF-associated genes. Due to the small number of genes currently associated with IPF and CF we cannot statistically establish the role of transcriptional complexity in these diseases. However, the high levels of transcriptional complexity in COPD and IPF suggest that post-transcriptional processes in the lung should be further investigated as potential mediators of disease.

COPD-associated transcript complexity is not generated from non-canonical splicing, exon size or transcript size

The high transcript complexity observed in COPD-associated genes suggests that alternative splicing is an important aspect of the post-transcriptional regulatory program of COPD-associated genes. Therefore, we further scrutinized our COPD gene list to identify transcript features that would lead to alternative splicing. First, we analyzed the percentage of strict canonical splice sites within the donor and acceptor regions of control gene sets. A difference in the percentage of canonical splice sites within COPD-associated genes would suggest that the observed transcript complexity is generated through non-canonical splicing. However, we discovered that COPD-associated gene splice sites are not significantly different from control splice sites (S4 Fig). We also found that there are no significant differences between COPD-associated genes and the length of the mature mRNAs or the number of large exons within the genes (S5 and S6 Figs). Thus, the high levels of alternative splicing in these genes must arise from other, unidentified, characteristics.

COPD-associated gene loci have unusual GC content and splicing patterns

Next, we explored the GC content around the first exon and splice sites. Within a gene, the region around the transcription start site (TSS) generally has high GC content, as do the splice junctions surrounding the first intron [27, 91, 92]. We determined the GC content within the sixty nucleotides surrounding the TSS, the first 5' donor splice site (Donor), the first 3' acceptor splice site (Accept) and the remaining splice acceptor and donor populations (DonorMid and AcceptMid) (Fig 4A). While COPD-associated genes are not significantly different than control genes at each individual boundary, they trend toward less abrupt differences between splice regions. To quantify this we analyzed the change in GC content surrounding the first intron (the difference between the first 5' donor splice site and the first 3' acceptor splice site). In COPD-associated loci this range narrows to 9.52%, and is significantly smaller than control

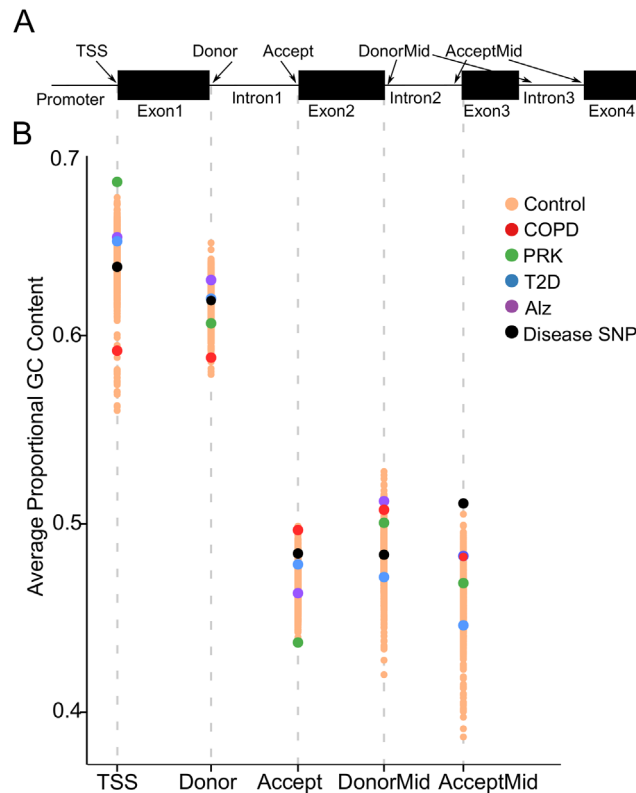


Fig 4. COPD splice junctions have lower GC content than expected. (A) Diagram of a representative gene with the regions of interest labeled, including the transcription start site (TSS), the first donor splice site (Donor), the first acceptor splice site (Accept) and the subsequent internal donor and acceptor splice sites (DonorMid and AcceptMid). (B) We calculated the average GC content for the 60 nucleotides surrounding each region in COPD-associated genes as well as genes associated with other diseases (PRK, T2D and ALZ). As a comparison, we also analyzed the GC content of genes with disease-associated SNPs within the TSS or splice junctions (black dots).

doi:10.1371/journal.pone.0140885.g004

samples (17.1%, $p = 0.004$). Unlike COPD gene loci, the other gene sets we analyzed have a broader range of GC content (Fig 4B, Table 2). In addition, known SNPs associated with disease that fall within splice junctions had a normal Donor-Acceptor breadth (15.2%) (Table 2). As differences in GC distribution are linked to gene evolutionary age and tissue-specific expression, we propose that this characteristic sets apart a group of complex genes with important roles in COPD etiology [93].

Table 2. In COPD the difference between the average GC content of the transcription start site and the first acceptor splice site is significantly smaller than expected ($p = 0.004$).

List	Average Difference	Standard Deviation
Control	17.1	1.76
COPD	9.52	-
Parkinson's Disease	24.4	-
Type 2 Diabetes	17.1	-
Alzheimer's Disease	18.8	-
Disease associated SNPs	15.2	-

doi:10.1371/journal.pone.0140885.t002

Our observations raise the question as to whether COPD-associated genes produce transcripts with atypical splicing patterns. Alternative splicing in higher eukaryotes is characterized primarily by exon skipping, but also includes alternative acceptor and donor splice sites and a variety of other complex events incorporating these and other types of splicing [94]. The *SERPINA1* gene is characterized by a very large number of potential splice sites (Fig 5A). For example, *SERPINA1* transcripts contain exon skipping events (Fig 5B) as well as alternative donor (Fig 5C) and alternative acceptor sites (Fig 5D). We analyzed control sets of genes to determine what a typical splicing pattern is like for gene loci. On average, we ascertained that alternatively spliced genes normally have around 41% skipped exons, 11% alternative donor sites, 15% alternative acceptor sites and 34% more complex events (Fig 5E). In comparison, alternative transcripts produced from the COPD-associated genes have fewer skipped exons (28%) and more alternative acceptor sites (22%), such as the event depicted in *SERPINA1* transcript (Fig 5D).

COPD-associated splicing isoforms are expressed in relevant tissues in a regulated fashion

Although many genes have documented alternative transcripts, we do not currently have a complete picture of when or where these isoforms are expressed. We have extensively demonstrated that COPD-associated genes are documented to have an unusually high number of transcripts per loci, but we wanted to know whether these transcripts were used in the lung or liver tissue (where *SERPINA1* is expressed and secreted) [65]. We analyzed the alternative splicing profiles generated from the Illumina BodyMap dataset. Looking at the highest expressed splice variant from each loci we saw that each was generally dominant in one or two tissues, rather than broadly expressed across the whole human body (Fig 6, S1 Table: **BodyMap Tissues for a list of genes and tissue distribution**). In addition, expression of the highest expressed variant was frequently found in lung tissue (14.5%), as expected for genes with a role in COPD. This enrichment of COPD-associated highly expressed variants is significant ($p < 0.001$) as the average enrichment of highly expressed variants in lung tissue from control gene lists is 4.82% (S9 Fig). The next enriched tissues in COPD associated-gene transcripts were white blood cells, testes and liver tissue (9.9%, 9.0% and 8.7% respectively). The tissue where the most prevalent transcript isoform is expressed is listed for each gene in S1 Table. Our results support the relevance of our COPD list to the disease.

In addition to examining general characteristics of COPD-associated gene expression, we wanted to analyze several genes in depth to inspect all of their splicing variants. To do so, we chose to scrutinize our archetypical COPD gene—*SERPINA1*. The highest expression of *SERPINA1* occurs in the liver, and liver expression is mainly confined to two splice variations. However, in other tissues these splice variants are not used and alternative splicing occurs (Fig 7A). We then examined *advanced glycosylation end product-specific receptor* (*AGER*). We chose this gene because of the evidence underlying its association with COPD, as well as its plethora of splice variants. The protein product of *AGER* is a multiligand cell surface receptor with links to many diseases involved with inflammation, including COPD [96–99] (Table 1). In addition, *AGER* is well documented to produce many alternative splice variants [100]. Similar to *SERPINA1*, we saw that *AGER* splice variants are highly regulated between tissues (Fig 7B). Several splice variations are present in COPD relevant tissues, like the lung, white blood cells and liver.

Furthermore, we decided to determine if alternatively spliced transcripts are differentially expressed in COPD patient lung tissue. Using RNA-seq data from the lung tissue of 26 COPD subjects and 26 controls with normal spirometry, a subset of data generated by Kim et al [102], we analyzed differential expression of highly complex, COPD-associated genes. We tested each

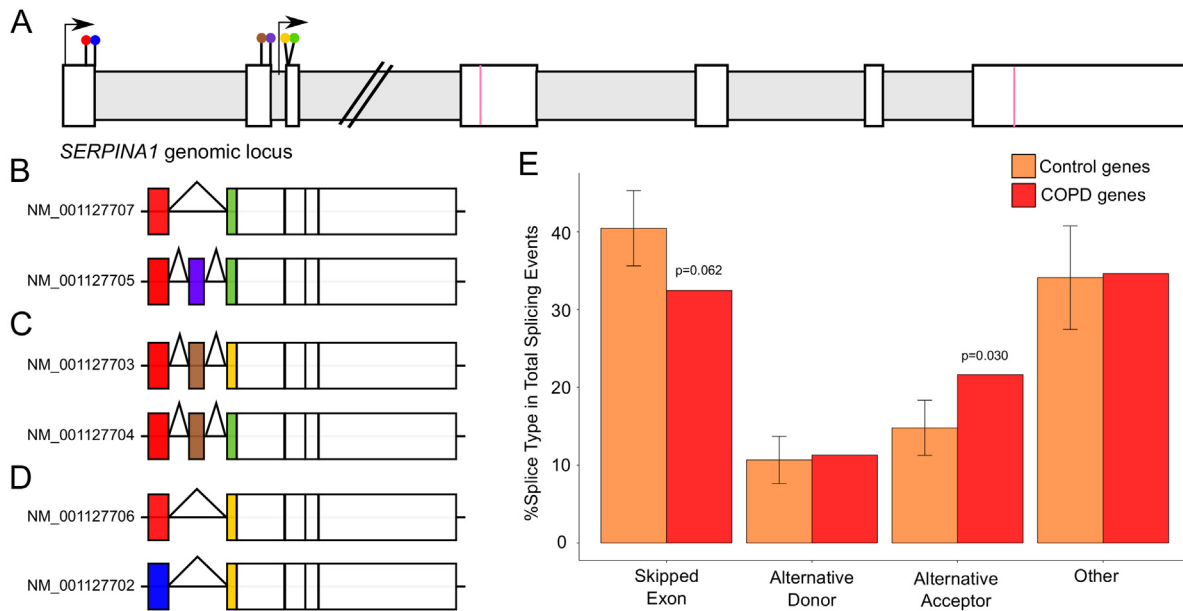


Fig 5. Skipped exons and alternative acceptor splicing events are differentially enriched in COPD genes. (A) *SERPINA1* contains a number of different splicing events including (B) skipped exons, (C) alternative donors and (D) alternative acceptors. (E) On average, COPD genes contain fewer skipped exons and significantly more alternative acceptor splice events ($p = 0.030$) per total splice events in comparison with normalized reference genes.

doi:10.1371/journal.pone.0140885.g005

gene in our COPD gene list with DEXSeq for differential exon usage (DEU), which measures the change in the relative exon usage between sample sets [101]. The transcript complexity contributing to DEU represents alternative TSS and polyadenylation sites as well as alternative splicing. For each gene, we report the total number of exons and the number of these exons that are differentially expressed in COPD and control subjects ($FDR < 0.05$) (S1 Table). Many COPD-associated genes, such as *AGER* and *SERPINA1*, exhibit significant DEU in multiple exons (Fig 7C and 7D), supporting our conclusions that transcriptional complexity is a feature of COPD etiology. Previously, Kim et al., identified specific alternative splicing patterns that were prevalent in COPD subjects compared to control subjects. Our findings are consistent with this finding, which further supports the hypothesis that transcript complexity is an important aspect of COPD [102].

Discussion

Current databases have been refining gene annotations for decades and, more recently, incorporating a flood of RNA sequencing data. One aspect arising from this work is a plethora of transcripts produced from alternative splicing. We provide evidence that COPD-associated genes are enriched for highly complex mRNAs, implying that the development of COPD might be influenced by malfunctions in alternative splicing. This transcriptional complexity is evident in two commonly used transcript reference databases, NCBI RefSeq and Ensembl (S10 Fig). We expected that other chronic diseases might also be enriched for transcriptionally complex genes. However, we did not find any significant enrichment in our disease gene sets for Type 2 Diabetes, Parkinson's or Alzheimer's disease even though they all trended higher than the average of their control sets. We did find that idiopathic pulmonary fibrosis (which shares more than 30% of the same genes as COPD) trends toward increased transcript complexity, while cystic fibrosis (which also shares more than 30% of the same genes as COPD) does not (S9 and S10 Tables). However, these data sets are small and will require further investigation.

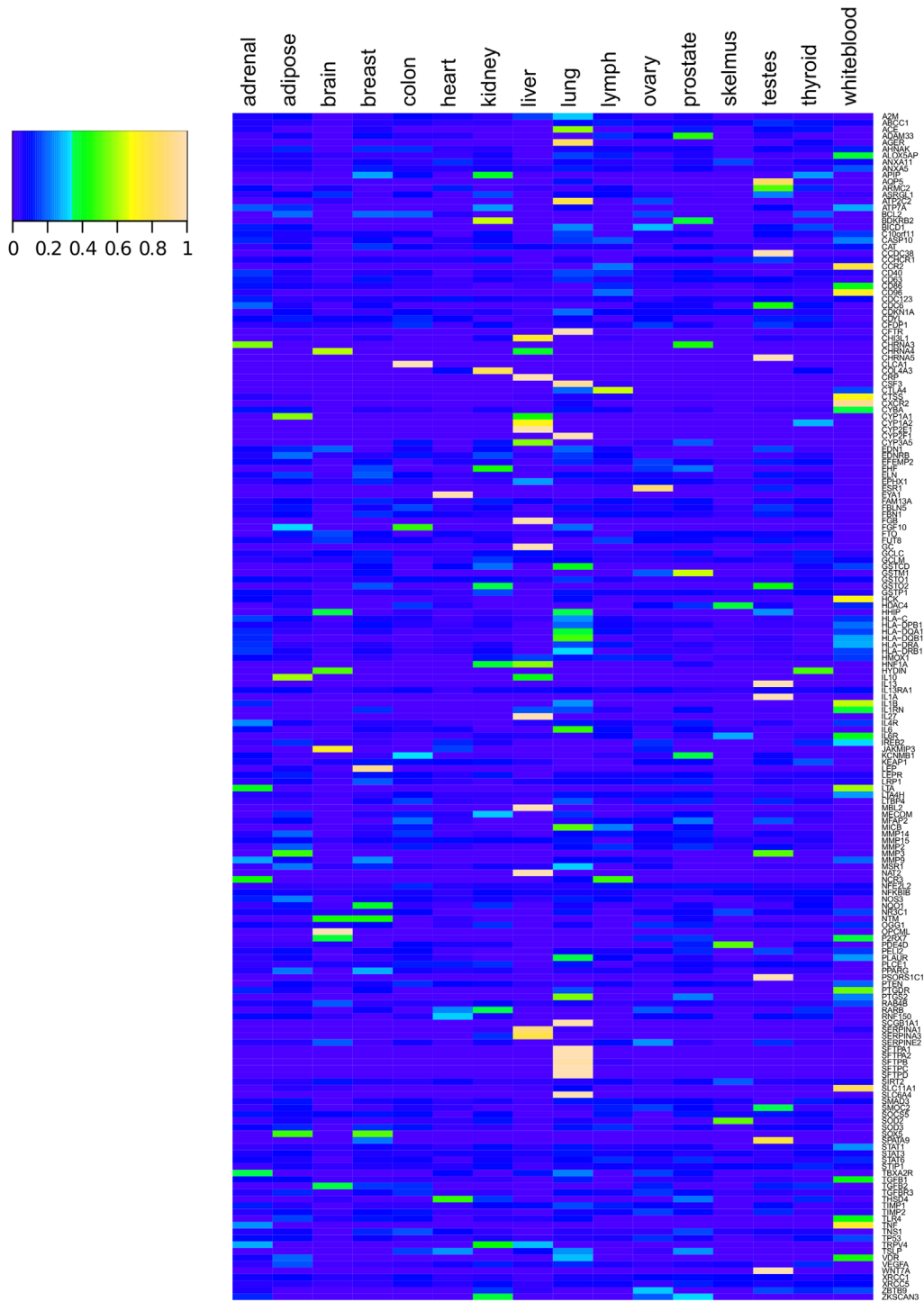


Fig 6. Highly expressed COPD-associated gene transcripts are enriched in lung tissue. The highest expressed transcript from each COPD-associated gene is shown as a percentage of expression in each of the 16 tissues from BodyMap. COPD-associated gene expression is highest in lung, white blood cells, testes and liver (14.5%, 9.9%, 9.0% and 8.7%), respectively. Other COPD-associated transcripts not in these tissues are still tissue specific and may be detrimental if expressed in the lung. Splice variants were determined as alternative splice sites through ASprofile [95].

doi:10.1371/journal.pone.0140885.g006

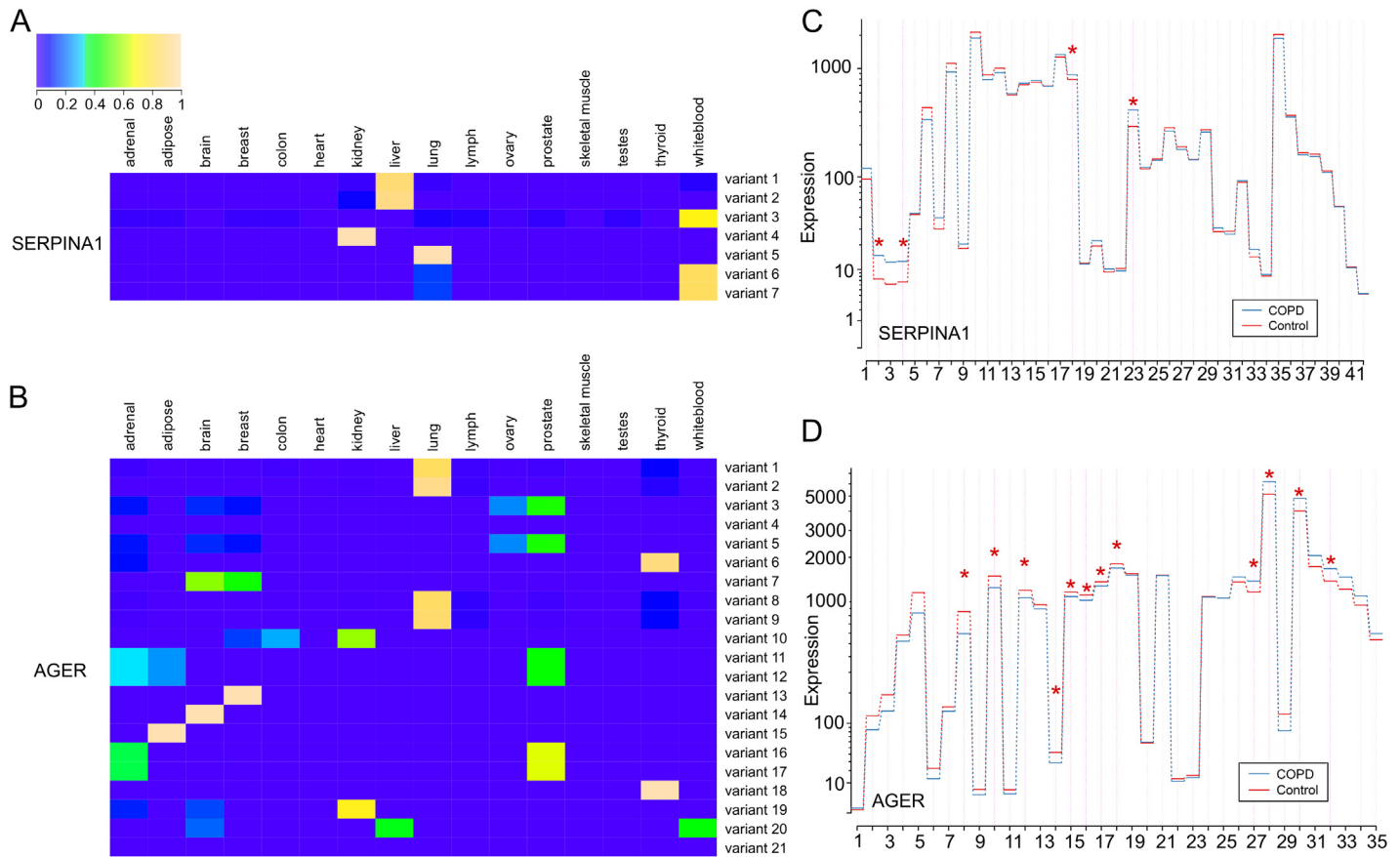


Fig 7. Usage of *SERPINA1* and *AGER* splice variants by tissue and COPD status. (A) The expression of each splice variant in *SERPINA1* was normalized across tissues. The pattern of expression shows that these splice variants are not broadly used in every tissue, but specific to the kidney, liver, lung and white blood cells. (B) Likewise, *AGER* is expressed in nearly every tissue tested, but splice variants are have specific patterns that imply regulation. Splice variants were determined as alternative splice sites through ASprofile [95]. (C) RNA-seq data from control subjects and COPD patients indicate that in *SERPINA1* four exons are significantly differentially used ($p < 0.05$). (D) In the *AGER* gene twelve exons are differentially expressed in COPD patients compared to normal controls ($p < 0.05$). Exon usage was generated with DEXSeq [101].

doi:10.1371/journal.pone.0140885.g007

Interestingly, an analysis of all genes with six or more transcripts indicates that transcriptionally complex genes are generally more enriched in disease databases, like the NHGRI catalog and SNP4Disease, than genes with less than six transcripts (S7 Fig). These databases collect information from a wide variety of diseases, including cancers and chronic diseases. The enrichment of highly transcriptionally complex genes in these disease-focused databases could reflect the susceptibility of these genes to mis-regulation that can lead to disease.

COPD-associated genes were significantly enriched with transcriptionally complex genes. COPD is a long-term degenerative disease with a strong environmental component; specifically damage caused by cigarette smoking. We do not fully understand how COPD develops, but it is clear that the immune system plays an important role. For example, T cells are found at higher numbers in the lungs of COPD patients and damaged epithelial cells secrete a variety of inflammatory signals [103]. The human adaptive immune system has been developing since the early vertebrates, but is still a relatively new and complex biological system, with significant dependence on alternative splicing for fine-tuning response [104–106]. It is possible that although transcriptional complexity allows for greater flexibility and control in complex systems, it is more likely misregulated, particularly in systems that depend heavily on alternative

splicing. However, a deeper understanding of the biological causes of COPD, as well as further research into the function and regulation of transcript variants, will be required to understand why COPD is associated with more transcriptionally complex genes.

The mechanisms that control alternative splicing are still an active area of investigation. We explored several genetic aspects of COPD-associated genes to narrow down the features that can affect a gene's ability to produce multiple isoforms from the same or similar transcripts. Unexpectedly, we found that COPD-associated genes trend toward constant GC-rich regions even though alternative splicing is generally correlated with high GC content around splice junctions. We also inspected transcript size, number of large exons and canonical/noncanonical splice sites, but did not find any differences between COPD-associated genes and control gene sets. It is possible that COPD genes may rely on enhancer elements or other regions outside of splice junction to control splicing activity. When we analyzed the types of splicing taking place within COPD-associated transcripts we discovered an increase in alternative acceptor sites. However, what drives splicing in these genes is not yet clear.

We used microarray expression data to detect broad gene expression changes during COPD progression as a way to confirm the importance of a subset of genes in COPD. However, since most microarray platforms do not detect splice isoforms, we also used the Illumina BodyMap RNA-seq alternative splicing profile data set to explore tissue specific expression of mRNAs. As expected, COPD-associated genes are highly expressed in lung tissue. More importantly, we saw tissue specific isoform expression—supporting the notion that splice variant expression is regulated and that transcripts have different biological roles. We also looked at alternative exon usage from a recent RNA-seq experiment with COPD and control patients [102]. When we looked at transcripts from COPD-associated genes in these people, we found that many of the exons were significantly differentially expressed in COPD patients in comparison to people with normal lung function.

COPD is both prevalent and incurable, but studies to identify the molecular causes of COPD lag behind other diseases like diabetes or cancer. Based on our findings that COPD-associated genes have a high level of transcript complexity, we propose that alternative splicing may be of particular interest in the study of COPD etiology.

Materials and Methods

Disease List Creation

We searched NCBI PubMed for reviews with the terms 'genetics' and each disease. We progressed from the most recent review article to the next, compiling a list of genes. After three reviews minimal unique genes were added to the gene lists for COPD, Parkinson's disease and Type 2 Diabetes. To identify Alzheimer's-associated genes we used four recent reviews, as a significant number of new disease-associated genes were added with another publication. The identified reviews spanned 2009 to 2014. Gene candidates were updated with the HUGO Gene Nomenclature Committee (HGNC) nomenclature and combined into literature review lists. In addition, we used the National Human Genome Research Institute (NHGRI) catalog of published GWAS [18]. We queried the catalog using supplied disease/trait terms, such as 'Parkinson's disease' and derivatives, such as 'Parkinson's disease (age of onset)'. We updated these genes with recent HGNC nomenclature and combined them with each literature review list.

Creation of Normalized Control Lists

To generate control gene sets normalized by length, we obtained genomic coordinates from the UCSC Genome Browser (hg19) [107, 108]. Length for each loci was determined as the maximum span encompassing all transcripts of the gene. Misannotated and pseudogene loci with

excessively large sizes were eliminated by removing loci with more than 1 million base-pairs. We then randomly selected genes from the UniProt-GOA reference human genes from the GO Consortium's Reference Genome initiative [37] so as to obtain a comparable length distribution (within 10% variance) to the disease test with the same number of genes. This operation was repeated 1000 times to establish statistics on how often the disease set was more transcriptionally complex than the random gene list. The reference gene set used is composed of over 17,000 gene loci, allowing us to establish independent gene lists for statistical evaluation. A small subset of disease genes had very small loci (e.g. miR499) and were not considered in the analysis.

Gene to Transcript Calculations

The number of transcripts from a locus was calculated by counting the number of unique UCSC Genome browser (hg19) RefSeq transcript IDs linked to the HGNC gene name [107, 108]. Alternatively, we used Ensembl transcript IDs, available from BioMart, linked to HGNC names to develop a pool of control genes. To calculate the statistical significance of the controls versus the disease-associated lists we performed boot-strapping with 1000 different randomized control lists for each disease. We used this data to calculate the number of genes producing 1, 2, 3, 4, and 5 or more transcripts.

Differential Gene Expression Analyses

Microarray data analyzing gene expression changes in COPD patients are available from the Gene Expression Omnibus (GEO) NCBI browser. We chose to analyze a large dataset (148 samples) with gene expression information collected from the sputum of ex-smokers diagnosed with COPD stage 2, 3 or 4 (gds4265) [38]. To analyze the data we used R bioconductor with Biobase, GEOquery to download and modify to microarray to useable form and empirical Bayes methods (limma) to determine which genes were differentially expressed by disease state [109–111]. We compared genes with significant expression (adjusted p-value greater than 0.05) to genes associated with COPD and tested these 85 genes for transcript complexity. The p-value adjustment performed by Singh, et al. corrected for age, gender and batch of patients [38].

GC content analyses

To obtain sequence information around splice junctions, we downloaded exonic sequences with additional 30 nucleotides upstream and downstream from the UCSC database (hg19) [107, 108]. The first 60 nucleotides of the first exon were used to calculate the percentage of G and Cs around the transcription start site (TSS). The last 60 nucleotides of the first exon were the first donor site (Donor) and the first 60 nucleotides of the second exon were the first acceptor site (Accept). All the remaining 5' junctions were combined as the DonorMid calculation. The AcceptMid value was the combination of all the 3' junctions with the exception of the first and last exons. Control calculations were based on non-normalized lists of 206 gene loci from the UniProt-GOA reference human gene set [37]. To calculate the statistical significance of the controls versus COPD we performed boot-strapping with 1000 different randomized control lists.

Splice type analyses

We employed the AStalavista program to analyze splicing patterns [112]. The UCSC Genome browser (hg19) was used to obtain gtf files for the COPD-associated genes as well as control

length-normalized genes in each of 1000 lists [107, 108]. These genes were analyzed for all splicing events with AStalavista and the number of skipped exon; alternative donor, alternative acceptor and other splice types were extracted from this information. The gene UTY was removed from the control set because of its abnormally large number of splice events. Splice type values were normalized by the total number of splicing events. We analyzed 1000 different randomized control lists to bootstrap the significant between the control and COPD genes.

BodyMap Tissue Specific Expression of mRNA Isoforms

We used the Alternative Splicing profiles developed with the Illumina BodyMap RNA-seq dataset to investigate how COPD-associated gene transcripts are expressed across tissues. The BodyMap data was prepared from 16 different human tissues, subjected to polyA purification, fragmented and random primed before 2x50 sequencing on HiSeq2000 with 1 lane for each tissue (raw data available from ArrayExpress, E-MTAB-513). The approximately 80 million reads for each tissue were aligned to the human genome, hg19, and turned into fragments with Cufflinks [95] (alternative splicing profiles available from <http://ccb.jhu.edu/software/ASprofile/data/>). ASprofile uses pairwise comparison of splicing events to define splice sites within each gene based on Ensembl annotations and sequencing data. Additional programs contained in the ASprofile suite calculate normalized expression for each tissue in a comparable fashion. We used ASprofile to map the percent expression in each tissue of the highest expressed transcript from each of the COPD-associated genes. We also investigated *SERPINA1* and *AGER* and mapped the percent expression in each tissue for all of the documented splice variants in these genes. To calculate the statistical significance of COPD lung enrichment we performed bootstrapping with 1000 different randomized control lists, selected the ASprofiles for the genes in each list, and computed the number of highest expressed variants expressed in the lung from each list. We used the same protocol to compute T2D, ALZ and PRK lung enrichment.

Differential Exon Usage in COPD Patients versus Control Subjects

We used publically available data generated as part of a study of COPD patients in comparison to subjects with normal spirometry [102]. The samples were prepared from lung tissue from 98 COPD patients and 91 controls, subjected to polyA-selected RNA extraction, fragmented, and primed with random hexamers before 2x50 sequencing on HiSeq2000 (raw data available from NCBI Gene Expression Omnibus (GEO) through accession number GSE57148). Read quality and alignment were verified with FastQC and Picard. They were aligned to the human genome, hg19, and expression was measured with Cufflinks 2.0.0. A subset of 26 control samples and 26 COPD-affected samples were randomly selected for analysis of differential exon usage (DEU) in our gene list using DEXSeq [101]. For each gene, a list of all transcripts were flattened into exon "counting bins", which are whole exons or fragments of exons that have differing boundaries between transcripts. For each exon bin, the number of reads that map to it were counted; if the read overlapped several exons, it was counted in multiple bins. The relative exon usage was calculated as a proportion of the number of transcripts from the gene that contain the exon per the number of all transcripts from the gene. Using read coverage for each bin, dispersion values were calculated to then test for significant DEU between COPD and control. To correct for the multiple hypothesis testing, the p-value was adjusted by the Benjamini-Hochberg algorithm and reported as a false discovery rate.

GLAD4U and SNP4Disease COPD List Creation

We accessed the SNP4Disease database of disease-associated SNPs (<http://snp4disease.mpi-bn.mpg.de/>) as an additional source of disease information. To develop the comprehensive

database list of genes, we queried the catalog using all supplied disease terms, such as ‘Respiratory Tract Diseases’ and ‘Virus Diseases’. For the COPD-associated genes, we searched the database for ‘Pulmonary Disease, Chronic Obstructive’. We used the gene symbol associated with each SNP. To explore text-mining results, we used Glad4U (<http://bioinfo.vanderbilt.edu/glad4u/>) [20]. We searched PubMed through this application for the term ‘Chronic Obstructive Pulmonary Disease’. We updated all gene lists with the most recent HGNC nomenclature before further analysis.

Removal of Weakly Correlated COPD Genes

We took advantage of a comprehensive literature review to sort COPD-associated genes by evidence [15]. We removed all genes with more negative than positively correlating studies and all genes with equal numbers of negatively and positively correlating studies. This resulted in a subset of 151 remaining COPD-associated genes, and eliminated genes with negative or conflicting evidence. We did not attempt to remove genes based on journal of publication or study type.

Transcript and Exon Length Comparisons

Genomic coordinates from UCSC detailing the position of all exons within a gene for control and disease gene lists were used to count how many exons larger than 200 base-pairs were in each transcript [107, 108]. We compared the average number of large exons between the COPD-associated transcripts and control transcripts using boot-strapping with 1000 control lists. In addition, we used these coordinates to analyze the total length of transcripts in control and disease lists.

Canonical vs. Non-canonical Splice Site Analysis

We downloaded exonic sequences with additional 30 nucleotides upstream and downstream from UCSC database. We selected the 60 nucleotides 5’ and 3’ of each exon and looked for the presence of a strict splice site within this region. The 5’ region of the first exon and 3’ region of the last exon were excluded. For the 5’ site we used the sequence GGTA / GGTGA. For the 3’ site we used the sequence CAGGT / TAGGT. Control calculations are based on non-normalized lists of 206 gene loci from the UniProt-GOA reference human gene set using boot-strapping with 1000 random lists.

Analysis of Disease Enrichment in Transcriptionally Diverse Genes

We separated a set of 804 genes with more than 5 transcripts from the UniProt-GOA reference human gene set. Using genes with 1 to 5 transcripts to make control sets of genes, we calculated the number of disease-associated genes within 1000 control sets and within the gene set with more than 5 transcripts to boot-strap the significance of our findings. We performed these calculations for both the NHGRI GWAS catalog of disease-associated genes and the SNP4Disease database of disease-associated genes.

Supporting Information

S1 Fig. Non-curated COPD-associated gene sources contribute unique candidates.

SNP4Disease results for genes associated COPD and other related disorders, resulting in a large pool of candidate genes that overlap with text-mining Glad4U results for COPD (left). The COPD-associated gene list used in the main body of this paper is not fully identified by either Glad4U or SNP4Disease results. Both SNP4Disease and Glad4U contribute additional

genes that are not commonly referenced in COPD literature or the NHGRI database (right). (TIFF)

S2 Fig. GWAS studies select for longer gene loci. (A) We calculated the average length for equal sets of genes from the UniProt-GOA reference human gene list (orange) and the NHGRI GWAS catalog (yellow) and plotted a histogram of their length distribution. The NHGRI GWAS catalog has much larger average gene loci lengths, therefore in all subsequent analyses we controlled for length. (B) We calculated the average length for equal sets of genes from the UniProt-GOA reference human gene list (orange) and genes from the SNP4Disease GWAS database (yellow). Lists of gene loci selected from the SNP4Disease database are larger than gene loci selected from the reference set. (C) We measured the average gene loci size of reference gene lists as well as the combined disease lists (All) and literature review (L) and GWAS (G) components for all four diseases. The mean and standard deviation of the average loci size of the control lists are shown (black bar, gray box). Compared to PRK, T2D and ALZ, COPD associated loci are the shortest and fall within the expected size of control gene sets. (TIFF)

S3 Fig. The transcript complexity enrichment of COPD genes is robust. (A) Genes with weak or mixed connection to COPD were removed from the original list and the resulting set was significantly more transcriptionally complex than length-normalized control gene lists. (B) The transcript complexity in the complete COPD list remains significant even when the top transcript producing genes are removed, unlike the top three control lists treated in the same fashion. The p-values are shown in all panels. (TIFF)

S4 Fig. COPD-associated genes transcript splice sites are composed of the expected number of canonical splice sites. COPD-associated 5' transcript splice sites contain similar percentages of canonical donor splice sites as control 5' transcript splice sites (left). The percentage of canonical COPD-associated transcript splice sites at the acceptor site also falls within the expected percentage range of canonical sites, as determined by analyzing the 3' acceptor splice sites from reference gene sets (right). (TIFF)

S5 Fig. Length-normalization selects control lists with similar length distributions as their template disease lists. (A) We measured the average gene loci size of normalized reference gene lists and the combined disease lists. The mean and standard deviation of the average loci size of the control lists are shown (black bar, grey box). (B) The average size of COPD-associated mRNAs falls within the range of the average mRNA size produced by length-normalized reference genes. Similar calculations were performed for Parkinson's Disease, Type 2 Diabetes and Alzheimer's Disease with controls normalized to each disease list. Disease-associated mRNA average size is within the expected values for all diseases tested. (TIFF)

S6 Fig. COPD-associated transcripts contain the expected number of large exons. The number of large exons (>200 bps) was calculated for each mRNA and then averaged for all mRNAs produced from each gene list. The average number of large exons per mRNA for the COPD-associated gene list falls within the range of the average number of large exons found in reference gene sets. (TIFF)

S7 Fig. Genes with more than 5 transcripts are more likely to be associated with disease. We counted the number of disease-associated genes (identified from public GWAS databases)

with 1–5 transcripts compared to those with more than 5 transcripts. Using NHGRI defined disease genes (left), we found that genes with >5 transcripts were significantly enriched for disease-associated genes ($p < 0.001$). We discovered similar effects when we compared these genes to the SNP4Disease gene list (right), with the gene list of genes with >5 transcripts containing significantly more disease-associated genes ($p < 0.001$).

(TIFF)

S8 Fig. IPF has a high number of genes with 6 or more transcripts. 20% of IPF-associated genes (red) and 10.68% of COPD-associated genes (green) have 6 or more transcripts, while only 8.25% of CF genes (orange) have 6 or more transcripts. The average percentage of total genes with each number of transcripts is shown for 1000 randomized control lists (blue) with the standard deviation shown as a grey ribbon. The low number of genes in the IPF and CF lists impede calculating statistical significance.

(TIF)

S9 Fig. Expressed COPD transcript isoforms are enriched in lung tissue. Expression of the most prevalent isoforms of COPD-associated gene transcripts occurred in lung tissue (14.5%) while the percent of isoforms expressed in lung tissue from randomized controls as well as from ALZ, PRK and T2D was much lower (4.82, 4.32, 4.25 and 5.74% respectively). This enrichment in COPD-associated isoforms is significant ($p < 0.001$).

(TIF)

S10 Fig. COPD-associated genes are significantly more transcriptionally complex than control datasets. (A) The number of transcripts per gene was calculated using Ensembl defined genes and transcripts for 1000 lists of 206 Ensembl genes as well as the COPD-associated genes. (B) The number of genes with 1,2,3,4,5 and 6 or more transcripts are displayed as an average for 1000 lists of control genes as well as the list of COPD-associated genes.

(TIF)

S1 Table. COPD-associated genes and information.

(XLS)

S2 Table. SNP4Disease and Glad4U genes associated with COPD.

(XLSX)

S3 Table. Genes associated with Parkinson's disease.

(XLSX)

S4 Table. Genes associated with Type 2 Diabetes.

(XLSX)

S5 Table. Genes associated with Alzheimer's Disease.

(XLSX)

S6 Table. Overlap between Literature Reviews in disease-associated gene loci and control sets.

(XLSX)

S7 Table. Overlap between Literature Reviews in Alzheimer's associated gene loci and control sets.

(XLSX)

S8 Table. Overlap between Literature Review and GWAS identified genes in disease and control sets.

(XLSX)

S9 Table. Genes associated with idiopathic pulmonary fibrosis.

(XLSX)

S10 Table. Cystic fibrosis modulatory genes.

(XLSX)

S11 Table. FPKM expression of exons stratified by expression in COPD and control patients.

(XLSX)

Author Contributions

Conceived and designed the experiments: LL AL. Performed the experiments: LL EM. Analyzed the data: LL EM. Wrote the paper: LL AL EM.

References

1. Baralle D, Baralle M. Splicing in action: assessing disease causing sequence changes. *Journal of medical genetics*. 2005; 42(10):737–48. doi: [10.1136/jmg.2004.029538](https://doi.org/10.1136/jmg.2004.029538) PMID: [16199547](https://pubmed.ncbi.nlm.nih.gov/16199547/); PubMed Central PMCID: PMC1735933.
2. Chatterjee S, Pal JK. Role of 5'- and 3'-untranslated regions of mRNAs in human diseases. *Biology of the cell / under the auspices of the European Cell Biology Organization*. 2009; 101(5):251–62. doi: [10.1042/BC20080104](https://doi.org/10.1042/BC20080104) PMID: [19275763](https://pubmed.ncbi.nlm.nih.gov/19275763/).
3. Lu ZX, Jiang P, Xing Y. Genetic variation of pre-mRNA alternative splicing in human populations. *Wiley interdisciplinary reviews RNA*. 2012; 3(4):581–92. doi: [10.1002/wrna.120](https://doi.org/10.1002/wrna.120) PMID: [22095823](https://pubmed.ncbi.nlm.nih.gov/22095823/); PubMed Central PMCID: PMC3339278.
4. Singh RK, Cooper TA. Pre-mRNA splicing in disease and therapeutics. *Trends in molecular medicine*. 2012; 18(8):472–82. doi: [10.1016/j.molmed.2012.06.006](https://doi.org/10.1016/j.molmed.2012.06.006) PMID: [22819011](https://pubmed.ncbi.nlm.nih.gov/22819011/); PubMed Central PMCID: PMC3411911.
5. Manolio TA. Genomewide association studies and assessment of the risk of disease. *The New England journal of medicine*. 2010; 363(2):166–76. doi: [10.1056/NEJMr0905980](https://doi.org/10.1056/NEJMr0905980) PMID: [20647212](https://pubmed.ncbi.nlm.nih.gov/20647212/).
6. Brebner JA, Stockley RA. Recent advances in alpha-1-antitrypsin deficiency-related lung disease. *Expert review of respiratory medicine*. 2013; 7(3):213–29; quiz 30. doi: [10.1586/ers.13.20](https://doi.org/10.1586/ers.13.20) PMID: [23734645](https://pubmed.ncbi.nlm.nih.gov/23734645/).
7. Gooptu B, Ekeowa UI, Lomas DA. Mechanisms of emphysema in alpha1-antitrypsin deficiency: molecular and cellular insights. *The European respiratory journal*. 2009; 34(2):475–88. doi: [10.1183/09031936.00096508](https://doi.org/10.1183/09031936.00096508) PMID: [19648523](https://pubmed.ncbi.nlm.nih.gov/19648523/).
8. Stoller JK, Lacbawan FL, Aboussouan LS. Alpha-1 Antitrypsin Deficiency. In: Pagon RA, Adam MP, Ardinger HH, Bird TD, Dolan CR, Fong CT, et al., editors. *GeneReviews(R)*. Seattle (WA)1993.
9. Decramer M, Janssens W, Miravittles M. Chronic obstructive pulmonary disease. *Lancet*. 2012; 379(9823):1341–51. doi: [10.1016/S0140-6736\(11\)60968-9](https://doi.org/10.1016/S0140-6736(11)60968-9) PMID: [22314182](https://pubmed.ncbi.nlm.nih.gov/22314182/).
10. Halvorsen M, Martin JS, Broadaway S, Laederach A. Disease-associated mutations that alter the RNA structural ensemble. *PLoS Genet*. 2010; 6(8):e1001074. Epub 2010/09/03. doi: [10.1371/journal.pgen.1001074](https://doi.org/10.1371/journal.pgen.1001074) PMID: [20808897](https://pubmed.ncbi.nlm.nih.gov/20808897/); PubMed Central PMCID: PMC2924325.
11. Silverman EK, Palmer LJ, Mosley JD, Barth M, Senter JM, Brown A, et al. Genomewide linkage analysis of quantitative spirometric phenotypes in severe early-onset chronic obstructive pulmonary disease. *Am J Hum Genet*. 2002; 70(5):1229–39. Epub 2002/03/27. doi: [10.1086/340316](https://doi.org/10.1086/340316) PMID: [11914989](https://pubmed.ncbi.nlm.nih.gov/11914989/); PubMed Central PMCID: PMC447597.
12. Mathers C, Fat DM, Boerma JT, World Health Organization. *The global burden of disease: 2004 update*. Geneva, Switzerland: World Health Organization; 2008. vii, 146 p. p.
13. Bossé Y. Genetics of chronic obstructive pulmonary disease: a succinct review, future avenues and prospective clinical applications. *Pharmacogenomics*. 2009; 10(4):655–67. doi: [10.2217/pgs.09.10](https://doi.org/10.2217/pgs.09.10) PMID: [19374520](https://pubmed.ncbi.nlm.nih.gov/19374520/).

14. Berndt A, Leme AS, Shapiro SD. Emerging genetics of COPD. *EMBO molecular medicine*. 2012; 4(11):1144–55. doi: [10.1002/emmm.201100627](https://doi.org/10.1002/emmm.201100627) PMID: [23090857](https://pubmed.ncbi.nlm.nih.gov/23090857/); PubMed Central PMCID: PMC3494872.
15. Bossé Y. Updates on the COPD gene list. *International journal of chronic obstructive pulmonary disease*. 2012; 7:607–31. doi: [10.2147/COPD.S35294](https://doi.org/10.2147/COPD.S35294) PMID: [23055711](https://pubmed.ncbi.nlm.nih.gov/23055711/); PubMed Central PMCID: PMC3459654.
16. Marciniak SJ, Lomas DA. Genetic susceptibility. *Clinics in chest medicine*. 2014; 35(1):29–38. doi: [10.1016/j.ccm.2013.10.008](https://doi.org/10.1016/j.ccm.2013.10.008) PMID: [24507835](https://pubmed.ncbi.nlm.nih.gov/24507835/).
17. Nakamura H. Genetics of COPD. *Allergology international: official journal of the Japanese Society of Allergology*. 2011; 60(3):253–8. doi: [10.2332/allergolint.11-RAI-0326](https://doi.org/10.2332/allergolint.11-RAI-0326) PMID: [21778810](https://pubmed.ncbi.nlm.nih.gov/21778810/).
18. Hindorf LA, MacArthur J, Morales J, Junkins HA, Hall PN, Klemm AK, et al. A Catalog of Published Genome-Wide Association Studies [cited 2014 March]. Available from: <http://www.genome.gov/gwastudies>.
19. Chang CQ, Yesupriya A, Rowell JL, Pimentel CB, Clyne M, Gwinn M, et al. A systematic review of cancer GWAS and candidate gene meta-analyses reveals limited overlap but similar effect sizes. *European journal of human genetics: EJHG*. 2014; 22(3):402–8. doi: [10.1038/ejhg.2013.161](https://doi.org/10.1038/ejhg.2013.161) PMID: [23881057](https://pubmed.ncbi.nlm.nih.gov/23881057/); PubMed Central PMCID: PMC3925284.
20. Jourquin J, Duncan D, Shi Z, Zhang B. GLAD4U: deriving and prioritizing gene lists from PubMed literature. *BMC genomics*. 2012; 13 Suppl 8:S20. doi: [10.1186/1471-2164-13-S8-S20](https://doi.org/10.1186/1471-2164-13-S8-S20) PMID: [23282288](https://pubmed.ncbi.nlm.nih.gov/23282288/); PubMed Central PMCID: PMC3535723.
21. Research MPIfHaL. SNP4Disease [cited 2014 March]. Available from: <http://snp4disease.mpi-bn.mpg.de/index.php>.
22. Houlden H, Singleton AB. The genetics and neuropathology of Parkinson's disease. *Acta neuropathologica*. 2012; 124(3):325–38. doi: [10.1007/s00401-012-1013-5](https://doi.org/10.1007/s00401-012-1013-5) PMID: [22806825](https://pubmed.ncbi.nlm.nih.gov/22806825/); PubMed Central PMCID: PMC3589971.
23. Lesage S, Brice A. Parkinson's disease: from monogenic forms to genetic susceptibility factors. *Human molecular genetics*. 2009; 18(R1):R48–59. doi: [10.1093/hmg/ddp012](https://doi.org/10.1093/hmg/ddp012) PMID: [19297401](https://pubmed.ncbi.nlm.nih.gov/19297401/).
24. Puschmann A. Monogenic Parkinson's disease and parkinsonism: clinical phenotypes and frequencies of known mutations. *Parkinsonism & related disorders*. 2013; 19(4):407–15. doi: [10.1016/j.parkreldis.2013.01.020](https://doi.org/10.1016/j.parkreldis.2013.01.020) PMID: [23462481](https://pubmed.ncbi.nlm.nih.gov/23462481/).
25. Singleton AB, Farrer MJ, Bonifati V. The genetics of Parkinson's disease: progress and therapeutic implications. *Movement disorders: official journal of the Movement Disorder Society*. 2013; 28(1):14–23. doi: [10.1002/mds.25249](https://doi.org/10.1002/mds.25249) PMID: [23389780](https://pubmed.ncbi.nlm.nih.gov/23389780/); PubMed Central PMCID: PMC3578399.
26. Bonnefond A, Froguel P, Vaxillaire M. The emerging genetics of type 2 diabetes. *Trends in molecular medicine*. 2010; 16(9):407–16. doi: [10.1016/j.molmed.2010.06.004](https://doi.org/10.1016/j.molmed.2010.06.004) PMID: [20728409](https://pubmed.ncbi.nlm.nih.gov/20728409/).
27. Elbein SC, Gamazon ER, Das SK, Rasouli N, Kern PA, Cox NJ. Genetic risk factors for type 2 diabetes: a trans-regulatory genetic architecture? *Am J Hum Genet*. 2012; 91(3):466–77. doi: [10.1016/j.ajhg.2012.08.002](https://doi.org/10.1016/j.ajhg.2012.08.002) PMID: [22958899](https://pubmed.ncbi.nlm.nih.gov/22958899/); PubMed Central PMCID: PMC3512001.
28. Petrie JR, Pearson ER, Sutherland C. Implications of genome wide association studies for the understanding of type 2 diabetes pathophysiology. *Biochemical pharmacology*. 2011; 81(4):471–7. doi: [10.1016/j.bcp.2010.11.010](https://doi.org/10.1016/j.bcp.2010.11.010) PMID: [21111713](https://pubmed.ncbi.nlm.nih.gov/21111713/).
29. Schäfer SA, Machicao F, Fritsche A, Häring HU, Kantartzis K. New type 2 diabetes risk genes provide new insights in insulin secretion mechanisms. *Diabetes research and clinical practice*. 2011; 93 Suppl 1:S9–24. doi: [10.1016/S0168-8227\(11\)70008-0](https://doi.org/10.1016/S0168-8227(11)70008-0) PMID: [21864758](https://pubmed.ncbi.nlm.nih.gov/21864758/).
30. Bettens K, Sleegers K, Van Broeckhoven C. Genetic insights in Alzheimer's disease. *Lancet neurology*. 2013; 12(1):92–104. doi: [10.1016/S1474-4422\(12\)70259-4](https://doi.org/10.1016/S1474-4422(12)70259-4) PMID: [23237904](https://pubmed.ncbi.nlm.nih.gov/23237904/).
31. Loy CT, Schofield PR, Turner AM, Kwok JB. Genetics of dementia. *Lancet*. 2014; 383(9919):828–40. doi: [10.1016/S0140-6736\(13\)60630-3](https://doi.org/10.1016/S0140-6736(13)60630-3) PMID: [23927914](https://pubmed.ncbi.nlm.nih.gov/23927914/).
32. Ridge PG, Ebbert MT, Kauwe JS. Genetics of Alzheimer's disease. *BioMed research international*. 2013; 2013:254954. doi: [10.1155/2013/254954](https://doi.org/10.1155/2013/254954) PMID: [23984328](https://pubmed.ncbi.nlm.nih.gov/23984328/); PubMed Central PMCID: PMC3741956.
33. Ringman JM, Coppola G. New genes and new insights from old genes: update on Alzheimer disease. *Continuum*. 2013; 19(2 Dementia):358–71. PMID: [23558482](https://pubmed.ncbi.nlm.nih.gov/23558482/); PubMed Central PMCID: PMC3915548.
34. Tanzi RE. A brief history of Alzheimer's disease gene discovery. *Journal of Alzheimer's disease: JAD*. 2013; 33 Suppl 1:S5–13. doi: [10.3233/JAD-2012-129044](https://doi.org/10.3233/JAD-2012-129044) PMID: [22986781](https://pubmed.ncbi.nlm.nih.gov/22986781/).
35. Tosto G, Reitz C. Genome-wide association studies in Alzheimer's disease: a review. *Current neurology and neuroscience reports*. 2013; 13(10):381. doi: [10.1007/s11910-013-0381-0](https://doi.org/10.1007/s11910-013-0381-0) PMID: [23954969](https://pubmed.ncbi.nlm.nih.gov/23954969/); PubMed Central PMCID: PMC3809844.

36. Gebauer F, Preiss T, Hentze MW. From cis-regulatory elements to complex RNPs and back. *Cold Spring Harbor perspectives in biology*. 2012; 4(7):a012245. doi: [10.1101/cshperspect.a012245](https://doi.org/10.1101/cshperspect.a012245) PMID: [22751153](https://pubmed.ncbi.nlm.nih.gov/22751153/); PubMed Central PMCID: PMC3385959.
37. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature genetics*. 2000; 25(1):25–9. PMID: [10802651](https://pubmed.ncbi.nlm.nih.gov/10802651/); PubMed Central PMCID: PMC3037419.
38. Singh D, Fox SM, Tal-Singer R, Plumb J, Bates S, Broad P, et al. Induced sputum genes associated with spirometric and radiological disease severity in COPD ex-smokers. *Thorax*. 2011; 66(6):489–95. doi: [10.1136/thx.2010.153767](https://doi.org/10.1136/thx.2010.153767) PMID: [21441172](https://pubmed.ncbi.nlm.nih.gov/21441172/).
39. He JQ, Shumansky K, Connett JE, Anthonisen NR, Pare PD, Sandford AJ. Association of genetic variations in the CSF2 and CSF3 genes with lung function in smoking-induced COPD. *The European respiratory journal*. 2008; 32(1):25–34. Epub 2008/03/21. doi: [10.1183/09031936.00040307](https://doi.org/10.1183/09031936.00040307) PMID: [18353856](https://pubmed.ncbi.nlm.nih.gov/18353856/).
40. Lin WR, Yen TH, Lim SN, Perng MD, Lin CY, Su MY, et al. Granulocyte colony-stimulating factor reduces fibrosis in a mouse model of chronic pancreatitis. *PLoS One*. 2014; 9(12):e116229. Epub 2015/01/01. doi: [10.1371/journal.pone.0116229](https://doi.org/10.1371/journal.pone.0116229) PMID: [25551560](https://pubmed.ncbi.nlm.nih.gov/25551560/); PubMed Central PMCID: PMC4281240.
41. Zhang F, Zhang L, Jiang HS, Chen XY, Zhang Y, Li HP, et al. Mobilization of bone marrow cells by CSF3 protects mice from bleomycin-induced lung injury. *Respiration*. 2011; 82(4):358–68. Epub 2011/07/23. doi: [10.1159/000328762](https://doi.org/10.1159/000328762) PMID: [21778693](https://pubmed.ncbi.nlm.nih.gov/21778693/).
42. Perrotti D, Cesi V, Trotta R, Guerzoni C, Santilli G, Campbell K, et al. BCR-ABL suppresses C/EBPalpha expression through inhibitory action of hnRNP E2. *Nature genetics*. 2002; 30(1):48–58. Epub 2001/12/26. doi: [10.1038/ng791](https://doi.org/10.1038/ng791) PMID: [11753385](https://pubmed.ncbi.nlm.nih.gov/11753385/).
43. Pillai SG, Kong X, Edwards LD, Cho MH, Anderson WH, Coxson HO, et al. Loci identified by genome-wide association studies influence different disease-related phenotypes in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 2010; 182(12):1498–505. Epub 2010/07/27. doi: [10.1164/rccm.201002-0151OC](https://doi.org/10.1164/rccm.201002-0151OC) PMID: [20656943](https://pubmed.ncbi.nlm.nih.gov/20656943/); PubMed Central PMCID: PMC3029936.
44. Cho MH, Boutaoui N, Klanderman BJ, Sylvia JS, Ziniti JP, Hersh CP, et al. Variants in FAM13A are associated with chronic obstructive pulmonary disease. *Nature genetics*. 2010; 42(3):200–2. Epub 2010/02/23. doi: [10.1038/ng.535](https://doi.org/10.1038/ng.535) PMID: [20173748](https://pubmed.ncbi.nlm.nih.gov/20173748/); PubMed Central PMCID: PMC2828499.
45. Xie J, Wu H, Xu Y, Wu X, Liu X, Shang J, et al. Gene susceptibility identification in a longitudinal study confirms new loci in the development of chronic obstructive pulmonary disease and influences lung function decline. *Respir Res*. 2015; 16:49. Epub 2015/05/01. doi: [10.1186/s12931-015-0209-3](https://doi.org/10.1186/s12931-015-0209-3) PMID: [25928290](https://pubmed.ncbi.nlm.nih.gov/25928290/); PubMed Central PMCID: PMC4427922.
46. Kim S, Kim H, Cho N, Lee SK, Han BG, Sull JW, et al. Identification of FAM13A gene associated with the ratio of FEV1 to FVC in Korean population by genome-wide association studies including gene-environment interactions. *J Hum Genet*. 2015; 60(3):139–45. Epub 2015/01/23. doi: [10.1038/jhg.2014.118](https://doi.org/10.1038/jhg.2014.118) PMID: [25608829](https://pubmed.ncbi.nlm.nih.gov/25608829/).
47. Chen W, Brehm JM, Manichaikul A, Cho MH, Boutaoui N, Yan Q, et al. A genome-wide association study of chronic obstructive pulmonary disease in Hispanics. *Ann Am Thorac Soc*. 2015; 12(3):340–8. Epub 2015/01/15. doi: [10.1513/AnnalsATS.201408-380OC](https://doi.org/10.1513/AnnalsATS.201408-380OC) PMID: [25584925](https://pubmed.ncbi.nlm.nih.gov/25584925/); PubMed Central PMCID: PMC4418314.
48. Castaldi PJ, Cho MH, Zhou X, Qiu W, McGeachie M, Celli B, et al. Genetic control of gene expression at novel and established chronic obstructive pulmonary disease loci. *Human molecular genetics*. 2015; 24(4):1200–10. Epub 2014/10/16. doi: [10.1093/hmg/ddu525](https://doi.org/10.1093/hmg/ddu525) PMID: [25315895](https://pubmed.ncbi.nlm.nih.gov/25315895/).
49. Lee JH, Cho MH, Hersh CP, McDonald ML, Crapo JD, Bakke PS, et al. Genetic susceptibility for chronic bronchitis in chronic obstructive pulmonary disease. *Respir Res*. 2014; 15:113. Epub 2014/09/23. doi: [10.1186/s12931-014-0113-2](https://doi.org/10.1186/s12931-014-0113-2) PMID: [25241909](https://pubmed.ncbi.nlm.nih.gov/25241909/); PubMed Central PMCID: PMC4190389.
50. Corvol H, Hodges CA, Drumm ML, Guillot L. Moving beyond genetics: is FAM13A a major biological contributor in lung physiology and chronic lung diseases? *Journal of medical genetics*. 2014; 51(10):646–9. Epub 2014/08/29. doi: [10.1136/jmedgenet-2014-102525](https://doi.org/10.1136/jmedgenet-2014-102525) PMID: [25163686](https://pubmed.ncbi.nlm.nih.gov/25163686/).
51. Jin Z, Chung JW, Mei W, Strack S, He C, Lau GW, et al. Regulation of nuclear-cytoplasmic shuttling and function of Family with sequence similarity 13, member A (Fam13a), by B56-containing PP2As and Akt. *Mol Biol Cell*. 2015; 26(6):1160–73. Epub 2015/01/23. doi: [10.1091/mbc.E14-08-1276](https://doi.org/10.1091/mbc.E14-08-1276) PMID: [25609086](https://pubmed.ncbi.nlm.nih.gov/25609086/); PubMed Central PMCID: PMC4357514.
52. Homma S, Sakamoto T, Hegab AE, Saitoh W, Nomura A, Ishii Y, et al. Association of phosphodiesterase 4D gene polymorphisms with chronic obstructive pulmonary disease: relationship to interleukin 13 gene polymorphism. *Int J Mol Med*. 2006; 18(5):933–9. Epub 2006/10/04. PMID: [17016624](https://pubmed.ncbi.nlm.nih.gov/17016624/).

53. de Jong K, Vonk JM, Timens W, Bosse Y, Sin DD, Hao K, et al. Genome-wide interaction study of gene-by-occupational exposure and effects on FEV levels. *J Allergy Clin Immunol*. 2015. Epub 2015/05/17. doi: [10.1016/j.jaci.2015.03.042](https://doi.org/10.1016/j.jaci.2015.03.042) PMID: [25979521](https://pubmed.ncbi.nlm.nih.gov/25979521/).
54. Yoon HK, Hu HJ, Rhee CK, Shin SH, Oh YM, Lee SD, et al. Polymorphisms in PDE4D are associated with a risk of COPD in non-emphysematous Koreans. *Copd*. 2014; 11(6):652–8. Epub 2014/06/14. doi: [10.3109/15412555.2014.898045](https://doi.org/10.3109/15412555.2014.898045) PMID: [24926854](https://pubmed.ncbi.nlm.nih.gov/24926854/).
55. Obeidat M, Wain LV, Shrine N, Kalsheker N, Soler Artigas M, Repapi E, et al. A comprehensive evaluation of potential lung function associated genes in the SpiroMeta general population sample. *PLoS One*. 2011; 6(5):e19382. Epub 2011/06/01. doi: [10.1371/journal.pone.0019382](https://doi.org/10.1371/journal.pone.0019382) PMID: [21625484](https://pubmed.ncbi.nlm.nih.gov/21625484/); PubMed Central PMCID: PMC3098839.
56. Himes BE, Hunninghake GM, Baurley JW, Rafaels NM, Sleiman P, Strachan DP, et al. Genome-wide association analysis identifies PDE4D as an asthma-susceptibility gene. *American journal of human genetics*. 2009; 84(5):581–93. Epub 2009/05/12. doi: [10.1016/j.ajhg.2009.04.006](https://doi.org/10.1016/j.ajhg.2009.04.006) PMID: [19426955](https://pubmed.ncbi.nlm.nih.gov/19426955/); PubMed Central PMCID: PMC2681010.
57. Konrad FM, Bury A, Schick MA, Ngamsri KC, Reutershan J. The unrecognized effects of phosphodiesterase 4 on epithelial cells in pulmonary inflammation. *PLoS One*. 2015; 10(4):e0121725. Epub 2015/04/25. doi: [10.1371/journal.pone.0121725](https://doi.org/10.1371/journal.pone.0121725) PMID: [25909327](https://pubmed.ncbi.nlm.nih.gov/25909327/); PubMed Central PMCID: PMC4409344.
58. Repapi E, Sayers I, Wain LV, Burton PR, Johnson T, Obeidat M, et al. Genome-wide association study identifies five loci associated with lung function. *Nature genetics*. 2010; 42(1):36–44. Epub 2009/12/17. doi: [10.1038/ng.501](https://doi.org/10.1038/ng.501) PMID: [20010834](https://pubmed.ncbi.nlm.nih.gov/20010834/); PubMed Central PMCID: PMC2862965.
59. Hancock DB, Eijgelsheim M, Wilk JB, Gharib SA, Loehr LR, Marcianti KD, et al. Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. *Nature genetics*. 2010; 42(1):45–52. Epub 2009/12/17. doi: [10.1038/ng.500](https://doi.org/10.1038/ng.500) PMID: [20010835](https://pubmed.ncbi.nlm.nih.gov/20010835/); PubMed Central PMCID: PMC2832852.
60. Hancock DB, Artigas MS, Gharib SA, Henry A, Manichaikul A, Ramasamy A, et al. Genome-wide joint meta-analysis of SNP and SNP-by-smoking interaction identifies novel loci for pulmonary function. *PLoS genetics*. 2012; 8(12):e1003098. Epub 2013/01/04. doi: [10.1371/journal.pgen.1003098](https://doi.org/10.1371/journal.pgen.1003098) PMID: [23284291](https://pubmed.ncbi.nlm.nih.gov/23284291/); PubMed Central PMCID: PMC3527213.
61. Sessa L, Gatti E, Zeni F, Antonelli A, Catucci A, Koch M, et al. The receptor for advanced glycation end-products (RAGE) is only present in mammals, and belongs to a family of cell adhesion molecules (CAMs). *PLoS One*. 2014; 9(1):e86903. Epub 2014/01/30. doi: [10.1371/journal.pone.0086903](https://doi.org/10.1371/journal.pone.0086903) PMID: [24475194](https://pubmed.ncbi.nlm.nih.gov/24475194/); PubMed Central PMCID: PMC3903589.
62. Soler Artigas M, Wain LV, Repapi E, Obeidat M, Sayers I, Burton PR, et al. Effect of five genetic variants associated with lung function on the risk of chronic obstructive lung disease, and their joint effects on lung function. *Am J Respir Crit Care Med*. 2011; 184(7):786–95. Epub 2011/10/04. doi: [10.1164/rccm.201102-0192OC](https://doi.org/10.1164/rccm.201102-0192OC) PMID: [21965014](https://pubmed.ncbi.nlm.nih.gov/21965014/); PubMed Central PMCID: PMC3398416.
63. Thun GA, Imboden M, Ferrarotti I, Kumar A, Obeidat M, Zorzetto M, et al. Causal and synthetic associations of variants in the SERPINA gene cluster with alpha1-antitrypsin serum levels. *PLoS genetics*. 2013; 9(8):e1003585. Epub 2013/08/31. doi: [10.1371/journal.pgen.1003585](https://doi.org/10.1371/journal.pgen.1003585) PMID: [23990791](https://pubmed.ncbi.nlm.nih.gov/23990791/); PubMed Central PMCID: PMC3749935.
64. Carrell RW, Jeppsson JO, Laurell CB, Brennan SO, Owen MC, Vaughan L, et al. Structure and variation of human alpha 1-antitrypsin. *Nature*. 1982; 298(5872):329–34. Epub 1982/07/22. PMID: [7045697](https://pubmed.ncbi.nlm.nih.gov/7045697/).
65. Matamala N, Martinez MT, Lara B, Perez L, Vazquez I, Jimenez A, et al. Alternative transcripts of the SERPINA1 gene in alpha-1 antitrypsin deficiency. *J Transl Med*. 2015; 13:211. Epub 2015/07/05. doi: [10.1186/s12967-015-0585-y](https://doi.org/10.1186/s12967-015-0585-y) PMID: [26141700](https://pubmed.ncbi.nlm.nih.gov/26141700/); PubMed Central PMCID: PMC4490674.
66. Ghose R, Chu A, Wang Y, Perlmutter DH. Mysteries of alpha1-antitrypsin deficiency: emerging therapeutic strategies for a challenging disease. *Dis Model Mech*. 2014; 7(4):411–9. Epub 2014/04/11. doi: [10.1242/dmm.014092](https://doi.org/10.1242/dmm.014092) PMID: [24719116](https://pubmed.ncbi.nlm.nih.gov/24719116/); PubMed Central PMCID: PMC3974452.
67. Arif E, Vibhuti A, Deepak D, Singh B, Siddiqui MS, Pasha MA. COX2 and p53 risk-alleles coexist in COPD. *Clin Chim Acta*. 2008; 397(1–2):48–50. Epub 2008/08/12. doi: [10.1016/j.cca.2008.07.010](https://doi.org/10.1016/j.cca.2008.07.010) PMID: [18692035](https://pubmed.ncbi.nlm.nih.gov/18692035/).
68. Dumur V, Lafitte JJ, Gervais R, Debaecker D, Kesteloot M, Lalau G, et al. Abnormal distribution of cystic fibrosis delta F508 allele in adults with chronic bronchial hypersecretion. *Lancet*. 1990; 335(8701):1340. Epub 1990/06/02. PMID: [1971393](https://pubmed.ncbi.nlm.nih.gov/1971393/).
69. Damico R, Simms T, Kim BS, Tekeste Z, Amankwan H, Damarla M, et al. p53 mediates cigarette smoke-induced apoptosis of pulmonary endothelial cells: inhibitory effects of macrophage migration inhibitor factor. *Am J Respir Cell Mol Biol*. 2011; 44(3):323–32. Epub 2010/05/08. doi: [10.1165/rcmb.2009-0379OC](https://doi.org/10.1165/rcmb.2009-0379OC) PMID: [20448056](https://pubmed.ncbi.nlm.nih.gov/20448056/); PubMed Central PMCID: PMC3095933.

70. Lee YL, Chen W, Tsai WK, Lee JC, Chiou HL, Shih CM, et al. Polymorphisms of p53 and p21 genes in chronic obstructive pulmonary disease. *J Lab Clin Med*. 2006; 147(5):228–33. Epub 2006/05/16. doi: [10.1016/j.lab.2005.12.008](https://doi.org/10.1016/j.lab.2005.12.008) PMID: [16697770](https://pubmed.ncbi.nlm.nih.gov/16697770/).
71. Hancox RJ, Poulton R, Welch D, Olova N, McLachlan CR, Greene JM, et al. Accelerated decline in lung function in cigarette smokers is associated with TP53/MDM2 polymorphisms. *Hum Genet*. 2009; 126(4):559–65. Epub 2009/06/13. doi: [10.1007/s00439-009-0704-z](https://doi.org/10.1007/s00439-009-0704-z) PMID: [19521721](https://pubmed.ncbi.nlm.nih.gov/19521721/); PubMed Central PMCID: PMC3740961.
72. D'Anna C, Cigna D, Costanzo G, Ferraro M, Siena L, Vitulo P, et al. Cigarette smoke alters cell cycle and induces inflammation in lung fibroblasts. *Life Sci*. 2015; 126:10–8. Epub 2015/02/01. doi: [10.1016/j.lfs.2015.01.017](https://doi.org/10.1016/j.lfs.2015.01.017) PMID: [25637683](https://pubmed.ncbi.nlm.nih.gov/25637683/).
73. Shetty SK, Bhandary YP, Marudamuthu AS, Abernathy D, Velusamy T, Starcher B, et al. Regulation of airway and alveolar epithelial cell apoptosis by p53-Induced plasminogen activator inhibitor-1 during cigarette smoke exposure injury. *Am J Respir Cell Mol Biol*. 2012; 47(4):474–83. Epub 2012/05/18. doi: [10.1165/rcmb.2011-0390OC](https://doi.org/10.1165/rcmb.2011-0390OC) PMID: [22592924](https://pubmed.ncbi.nlm.nih.gov/22592924/); PubMed Central PMCID: PMC3488631.
74. Yamada M, Ishii T, Ikeda S, Naka-Mieno M, Tanaka N, Arai T, et al. Association of fucosyltransferase 8 (FUT8) polymorphism Thr267Lys with pulmonary emphysema. *J Hum Genet*. 2011; 56(12):857–60. Epub 2011/10/21. doi: [10.1038/jhg.2011.118](https://doi.org/10.1038/jhg.2011.118) PMID: [22011814](https://pubmed.ncbi.nlm.nih.gov/22011814/).
75. Kamio K, Yoshida T, Gao C, Ishii T, Ota F, Motegi T, et al. alpha1,6-Fucosyltransferase (Fut8) is implicated in vulnerability to elastase-induced emphysema in mice and a possible non-invasive predictive marker for disease progression and exacerbations in chronic obstructive pulmonary disease (COPD). *Biochem Biophys Res Commun*. 2012; 424(1):112–7. Epub 2012/06/27. PMID: [22732410](https://pubmed.ncbi.nlm.nih.gov/22732410/).
76. Gao C, Maeno T, Ota F, Ueno M, Korekane H, Takamatsu S, et al. Sensitivity of heterozygous alpha1,6-fucosyltransferase knock-out mice to cigarette smoke-induced emphysema: implication of aberrant transforming growth factor-beta signaling and matrix metalloproteinase gene expression. *J Biol Chem*. 2012; 287(20):16699–708. Epub 2012/03/22. doi: [10.1074/jbc.M111.315333](https://doi.org/10.1074/jbc.M111.315333) PMID: [22433854](https://pubmed.ncbi.nlm.nih.gov/22433854/); PubMed Central PMCID: PMC3351343.
77. Newton R. Anti-inflammatory glucocorticoids: changing concepts. *Eur J Pharmacol*. 2014; 724:231–6. Epub 2013/06/12. doi: [10.1016/j.ejphar.2013.05.035](https://doi.org/10.1016/j.ejphar.2013.05.035) PMID: [23747654](https://pubmed.ncbi.nlm.nih.gov/23747654/).
78. Rider CF, King EM, Holden NS, Giembycz MA, Newton R. Inflammatory stimuli inhibit glucocorticoid-dependent transactivation in human pulmonary epithelial cells: rescue by long-acting beta2-adrenoceptor agonists. *J Pharmacol Exp Ther*. 2011; 338(3):860–9. Epub 2011/05/31. doi: [10.1124/jpet.111.181016](https://doi.org/10.1124/jpet.111.181016) PMID: [21622733](https://pubmed.ncbi.nlm.nih.gov/21622733/).
79. Lajoie M, Hsu YC, Gronostajski RM, Bailey TL. An overlapping set of genes is regulated by both NFIB and the glucocorticoid receptor during lung maturation. *BMC genomics*. 2014; 15:231. Epub 2014/03/26. doi: [10.1186/1471-2164-15-231](https://doi.org/10.1186/1471-2164-15-231) PMID: [24661679](https://pubmed.ncbi.nlm.nih.gov/24661679/); PubMed Central PMCID: PMC4023408.
80. Corvol H, Nathan N, Charlier C, Chadelat K, Le Rouzic P, Tabary O, et al. Glucocorticoid receptor gene polymorphisms associated with progression of lung disease in young patients with cystic fibrosis. *Respir Res*. 2007; 8:88. Epub 2007/12/01. doi: [10.1186/1465-9921-8-88](https://doi.org/10.1186/1465-9921-8-88) PMID: [18047640](https://pubmed.ncbi.nlm.nih.gov/18047640/); PubMed Central PMCID: PMC2217522.
81. Barnes PJ. Corticosteroid resistance in patients with asthma and chronic obstructive pulmonary disease. *J Allergy Clin Immunol*. 2013; 131(3):636–45. Epub 2013/01/31. doi: [10.1016/j.jaci.2012.12.1564](https://doi.org/10.1016/j.jaci.2012.12.1564) PMID: [23360759](https://pubmed.ncbi.nlm.nih.gov/23360759/).
82. Zanini A, Spanevello A, Baraldo S, Majori M, Della Patrona S, Gumiero F, et al. Decreased maturation of dendritic cells in the central airways of COPD patients is associated with VEGF, TGF-beta and vascularization. *Respiration*. 2014; 87(3):234–42. Epub 2014/01/18. doi: [10.1159/000356749](https://doi.org/10.1159/000356749) PMID: [24435103](https://pubmed.ncbi.nlm.nih.gov/24435103/).
83. Sharma S, Murphy AJ, Soto-Quiros ME, Avila L, Klanderman BJ, Sylvia JS, et al. Association of VEGF polymorphisms with childhood asthma, lung function and airway responsiveness. *The European respiratory journal*. 2009; 33(6):1287–94. Epub 2009/02/07. doi: [10.1183/09031936.00113008](https://doi.org/10.1183/09031936.00113008) PMID: [19196819](https://pubmed.ncbi.nlm.nih.gov/19196819/); PubMed Central PMCID: PMC3725278.
84. Murray JF, Mason RJ. Murray and Nadel's textbook of respiratory medicine. 5th ed. Philadelphia, PA: Saunders/Elsevier; 2010.
85. Kropski JA, Blackwell TS, Loyd JE. The genetic basis of idiopathic pulmonary fibrosis. *The European respiratory journal*. 2015; 45(6):1717–27. Epub 2015/04/04. doi: [10.1183/09031936.00163814](https://doi.org/10.1183/09031936.00163814) PMID: [25837031](https://pubmed.ncbi.nlm.nih.gov/25837031/).
86. Hambly N, Shimbori C, Kolb M. Molecular classification of idiopathic pulmonary fibrosis: Personalized medicine, genetics and biomarkers. *Respirology*. 2015. Epub 2015/06/26. doi: [10.1111/resp.12569](https://doi.org/10.1111/resp.12569) PMID: [26109466](https://pubmed.ncbi.nlm.nih.gov/26109466/).

87. Spagnolo P, Grunewald J, du Bois RM. Genetic determinants of pulmonary fibrosis: evolving concepts. *Lancet Respir Med*. 2014; 2(5):416–28. Epub 2014/05/13. doi: [10.1016/S2213-2600\(14\)70047-5](https://doi.org/10.1016/S2213-2600(14)70047-5) PMID: [24815806](https://pubmed.ncbi.nlm.nih.gov/24815806/).
88. Tsui LC, Dorfman R. The cystic fibrosis gene: a molecular genetic perspective. *Cold Spring Harb Perspect Med*. 2013; 3(2):a009472. Epub 2013/02/05. doi: [10.1101/cshperspect.a009472](https://doi.org/10.1101/cshperspect.a009472) PMID: [23378595](https://pubmed.ncbi.nlm.nih.gov/23378595/); PubMed Central PMCID: PMC3552342.
89. Knowles MR. Gene modifiers of lung disease. *Curr Opin Pulm Med*. 2006; 12(6):416–21. Epub 2006/10/21. PMID: [17053491](https://pubmed.ncbi.nlm.nih.gov/17053491/).
90. Cutting GR. Modifier genes in Mendelian disorders: the example of cystic fibrosis. *Ann N Y Acad Sci*. 2010; 1214:57–69. Epub 2010/12/24. doi: [10.1111/j.1749-6632.2010.05879.x](https://doi.org/10.1111/j.1749-6632.2010.05879.x) PMID: [21175684](https://pubmed.ncbi.nlm.nih.gov/21175684/); PubMed Central PMCID: PMC3040597.
91. Zhang J, Kuo CC, Chen L. GC content around splice sites affects splicing through pre-mRNA secondary structures. *BMC genomics*. 2011; 12:90. doi: [10.1186/1471-2164-12-90](https://doi.org/10.1186/1471-2164-12-90) PMID: [21281513](https://pubmed.ncbi.nlm.nih.gov/21281513/); PubMed Central PMCID: PMC3041747.
92. Shepard PJ, Hertel KJ. Conserved RNA secondary structures promote alternative splicing. *Rna*. 2008; 14(8):1463–9. doi: [10.1261/ma.1069408](https://doi.org/10.1261/ma.1069408) PMID: [18579871](https://pubmed.ncbi.nlm.nih.gov/18579871/); PubMed Central PMCID: PMC2491482.
93. Hao L, Ge X, Wan H, Hu S, Lercher MJ, Yu J, et al. Human functional genetic studies are biased against the medically most relevant primate-specific genes. *BMC evolutionary biology*. 2010; 10:316. doi: [10.1186/1471-2148-10-316](https://doi.org/10.1186/1471-2148-10-316) PMID: [20961448](https://pubmed.ncbi.nlm.nih.gov/20961448/); PubMed Central PMCID: PMC2970608.
94. Sammeth M, Foissac S, Guigo R. A general definition and nomenclature for alternative splicing events. *PLoS computational biology*. 2008; 4(8):e1000147. doi: [10.1371/journal.pcbi.1000147](https://doi.org/10.1371/journal.pcbi.1000147) PMID: [18688268](https://pubmed.ncbi.nlm.nih.gov/18688268/); PubMed Central PMCID: PMC2467475.
95. Florea L, Song L, Salzberg SL. Thousands of exon skipping events differentiate among splicing patterns in sixteen human tissues. *F1000Research*. 2013; 2:188. doi: [10.12688/f1000research.2-188.v2](https://doi.org/10.12688/f1000research.2-188.v2) PMID: [24555089](https://pubmed.ncbi.nlm.nih.gov/24555089/); PubMed Central PMCID: PMC3892928.
96. Smith DJ, Yerkovich ST, Towers MA, Carroll ML, Thomas R, Upham JW. Reduced soluble receptor for advanced glycation end-products in COPD. *The European respiratory journal*. 2011; 37(3):516–22. doi: [10.1183/09031936.00029310](https://doi.org/10.1183/09031936.00029310) PMID: [20595148](https://pubmed.ncbi.nlm.nih.gov/20595148/).
97. Miniati M, Monti S, Basta G, Cocci F, Fornai E, Bottai M. Soluble receptor for advanced glycation end products in COPD: relationship with emphysema and chronic cor pulmonale: a case-control study. *Respir Res*. 2011; 12:37. doi: [10.1186/1465-9921-12-37](https://doi.org/10.1186/1465-9921-12-37) PMID: [21450080](https://pubmed.ncbi.nlm.nih.gov/21450080/); PubMed Central PMCID: PMC3072955.
98. Guo WA, Knight PR, Raghavendran K. The receptor for advanced glycation end products and acute lung injury/acute respiratory distress syndrome. *Intensive care medicine*. 2012; 38(10):1588–98. doi: [10.1007/s00134-012-2624-y](https://doi.org/10.1007/s00134-012-2624-y) PMID: [22777515](https://pubmed.ncbi.nlm.nih.gov/22777515/).
99. Simm A, Bartling B, Silber RE. RAGE: a new pleiotropic antagonistic gene? *Ann N Y Acad Sci*. 2004; 1019:228–31. doi: [10.1196/annals.1297.038](https://doi.org/10.1196/annals.1297.038) PMID: [15247020](https://pubmed.ncbi.nlm.nih.gov/15247020/).
100. Hudson BI, Carter AM, Harja E, Kalea AZ, Arriero M, Yang H, et al. Identification, classification, and expression of RAGE gene splice variants. *FASEB journal: official publication of the Federation of American Societies for Experimental Biology*. 2008; 22(5):1572–80. doi: [10.1096/fj.07-9909com](https://doi.org/10.1096/fj.07-9909com) PMID: [18089847](https://pubmed.ncbi.nlm.nih.gov/18089847/).
101. Anders S, Reyes A, Huber W. Detecting differential usage of exons from RNA-seq data. *Genome Res*. 2012; 22(10):2008–17. doi: [10.1101/gr.133744.111](https://doi.org/10.1101/gr.133744.111) PMID: [22722343](https://pubmed.ncbi.nlm.nih.gov/22722343/); PubMed Central PMCID: PMC3460195.
102. Kim WJ, Lim JH, Lee JS, Lee SD, Kim JH, Oh YM. Comprehensive Analysis of Transcriptome Sequencing Data in the Lung Tissues of COPD Subjects. *Int J Genomics*. 2015; 2015:206937. doi: [10.1155/2015/206937](https://doi.org/10.1155/2015/206937) PMID: [25834810](https://pubmed.ncbi.nlm.nih.gov/25834810/); PubMed Central PMCID: PMC4365374.
103. Barnes PJ. Mediators of chronic obstructive pulmonary disease. *Pharmacol Rev*. 2004; 56(4):515–48. Epub 2004/12/17. doi: [10.1124/pr.56.4.2](https://doi.org/10.1124/pr.56.4.2) PMID: [15602009](https://pubmed.ncbi.nlm.nih.gov/15602009/).
104. Chen L, Bush SJ, Tovar-Corona JM, Castillo-Morales A, Urrutia AO. Correcting for differential transcript coverage reveals a strong relationship between alternative splicing and organism complexity. *Mol Biol Evol*. 2014; 31(6):1402–13. Epub 2014/04/01. doi: [10.1093/molbev/msu083](https://doi.org/10.1093/molbev/msu083) PMID: [24682283](https://pubmed.ncbi.nlm.nih.gov/24682283/); PubMed Central PMCID: PMC4032128.
105. Lynch KW. Consequences of regulated pre-mRNA splicing in the immune system. *Nat Rev Immunol*. 2004; 4(12):931–40. Epub 2004/12/02. doi: [10.1038/nri1497](https://doi.org/10.1038/nri1497) PMID: [15573128](https://pubmed.ncbi.nlm.nih.gov/15573128/).
106. Flajnik MF, Kasahara M. Origin and evolution of the adaptive immune system: genetic events and selective pressures. *Nat Rev Genet*. 2010; 11(1):47–59. Epub 2009/12/10. doi: [10.1038/nrg2703](https://doi.org/10.1038/nrg2703) PMID: [19997068](https://pubmed.ncbi.nlm.nih.gov/19997068/); PubMed Central PMCID: PMC3805090.

107. International Human Genome Sequencing C. Initial sequencing and analysis of the human genome. *Nature*. 2001; 409(6822):860–921. doi: [10.1038/35057062](https://doi.org/10.1038/35057062) PMID: [11237011](https://pubmed.ncbi.nlm.nih.gov/11237011/).
108. Karolchik D, Hinrichs AS, Furey TS, Roskin KM, Sugnet CW, Haussler D, et al. The UCSC Table Browser data retrieval tool. *Nucleic acids research*. 2004; 32(Database issue):D493–6. doi: [10.1093/nar/gkh103](https://doi.org/10.1093/nar/gkh103) PMID: [14681465](https://pubmed.ncbi.nlm.nih.gov/14681465/); PubMed Central PMCID: PMC308837.
109. McCarthy DJ, Smyth GK. Testing significance relative to a fold-change threshold is a TREAT. *Bioinformatics*. 2009; 25(6):765–71. doi: [10.1093/bioinformatics/btp053](https://doi.org/10.1093/bioinformatics/btp053) PMID: [19176553](https://pubmed.ncbi.nlm.nih.gov/19176553/); PubMed Central PMCID: PMC2654802.
110. Davis S, Meltzer PS. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics*. 2007; 23(14):1846–7. doi: [10.1093/bioinformatics/btm254](https://doi.org/10.1093/bioinformatics/btm254) PMID: [17496320](https://pubmed.ncbi.nlm.nih.gov/17496320/).
111. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome biology*. 2004; 5(10):R80. doi: [10.1186/gb-2004-5-10-r80](https://doi.org/10.1186/gb-2004-5-10-r80) PMID: [15461798](https://pubmed.ncbi.nlm.nih.gov/15461798/); PubMed Central PMCID: PMC545600.
112. Foissac S, Sammeth M. ASTALAVISTA: dynamic and flexible analysis of alternative splicing events in custom gene datasets. *Nucleic acids research*. 2007; 35(Web Server issue):W297–9. doi: [10.1093/nar/gkm311](https://doi.org/10.1093/nar/gkm311) PMID: [17485470](https://pubmed.ncbi.nlm.nih.gov/17485470/); PubMed Central PMCID: PMC1933205.