Suppressed Expression of T-Box Transcription Factors Is Involved in Senescence in Chronic Obstructive Pulmonary Disease

George K. Acquaah-Mensah[1][9]*, Deepti Malhotra[2,3][9], Madhulika Vulimiri[2,4], Jason E. McDermott[5], Shyam Biswal[2]

1 Department of Pharmaceutical Sciences, Massachusetts College of Pharmacy and Health Sciences, Worcester, Massachusetts, United States of America, **2** Department of Environmental Health Sciences, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland, United States of America, **3** Genetic Disease Research Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, United States of America, **4** University of North Carolina at Chapel Hill, Raleigh-Durham, North Carolina, United States of America, **5** Computational Biology and Bioinformatics Group Pacific Northwest National Laboratory, Richland, Washington, United States of America

Abstract

Chronic obstructive pulmonary disease (COPD) is a major global health problem. The etiology of COPD has been associated with apoptosis, oxidative stress, and inflammation. However, understanding of the molecular interactions that modulate COPD pathogenesis remains only partly resolved. We conducted an exploratory study on COPD etiology to identify the key molecular participants. We used information-theoretic algorithms including Context Likelihood of Relatedness (CLR), Algorithm for the Reconstruction of Accurate Cellular Networks (ARACNE), and Inferelator. We captured direct functional associations among genes, given a compendium of gene expression profiles of human lung epithelial cells. A set of genes differentially expressed in COPD, as reported in a previous study were superposed with the resulting transcriptional regulatory networks. After factoring in the properties of the networks, an established COPD susceptibility locus and domain-domain interactions involving protein products of genes in the generated networks, several molecular candidates were predicted to be involved in the etiology of COPD. These include COL4A3, CFLAR, GULP1, PDCD1, CASP10, PAX3, BOK, HSPD1, PITX2, and PML. Furthermore, T-box (TBX) genes and cyclin-dependent kinase inhibitor 2A (CDKN2A), which are in a direct transcriptional regulatory relationship, emerged as preeminent participants in the etiology of COPD by means of senescence. Contrary to observations in neoplasms, our study reveals that the expression of genes and proteins in the lung samples from patients with COPD indicate an increased tendency towards cellular senescence. The expression of the anti-senescence mediators TBX transcription factors, chromatin modifiers histone deacetylases, and sirtuins was suppressed; while the expression of TBX-regulated cellular senescence markers such as CDKN2A, CDKN1A, and CAV1 was elevated in the peripheral lung tissue samples from patients with COPD. The critical balance between senescence and anti-senescence factors is disrupted towards senescence in COPD lungs.

Introduction

Chronic obstructive pulmonary disease (COPD) is characterized by a progressive decline in lung function, with an irreversible airflow obstruction, caused either by chronic bronchitis, emphysema or both [1]. It is a leading cause of morbidity and mortality worldwide, and thus a global health problem [2,3]. COPD affects around 210 million people worldwide and is predicted to become the third leading cause of death worldwide by the year 2020 [4,5]. The pathogenesis of COPD involves chronic inflammatory response in airways and lung parenchyma that results in pulmonary tissue injury, repair, and abnormal remodeling processes [6–9]. The pathobiology of COPD involves persistent inflammation, oxidative, and nitrosative stress, impaired cell repair and cell death manifested as senescence and apoptosis, and destruction of extracellular matrix due to protease-antiprotease imbalance in the lung tissues [1,10].

Cigarette smoking remains the primary preventable environmental risk factor for COPD [11]. Other factors such as air pollution, respiratory infections, and aging are being recognized as critical environmental contributors to disease pathogenesis as many more people are exposed to biomass pollutants compared to

Author Summary

Chronic obstructive pulmonary disease or COPD is among the most lethal of respiratory diseases. While this disease has been well characterized, more studies are needed to learn the interaction of macromolecules involved in the progression towards illness. We explored possible interactions involved in the disease process using a compendium of gene expression data from frontline cells of the respiratory airways of the lung. The gene expression data were generated under a variety of experimental conditions. Application of computational schemes, which robustly detect enduring patterns, among sections of the genes represented across the varying experimental perturbations, revealed important regulatory relationships. When gene expression data from lungs of patients with COPD were factored into these networks of regulatory relationships, certain highly connected nodes (hubs) representing differentially expressed genes emerged. Notably included are members of the T-box (TBX) family of genes and CDKN2A, which regulate cellular aging. These findings were confirmed in studies using lung samples from COPD patients. Novel genes linked to TBX and CDKN2A include COL4A3, CFLAR, GULP1, PDCD1, CASP10, PAX3, BOK, HSPD1, PITX2, and PML, which were thus predicted to be involved in the disease process. The balance between senescence and anti-senescence factors is disrupted towards senescence in COPD lungs.

tobacco smoke [10,12]. It is unknown why only a subset of smokers as well as people exposed to similar amount of other environmental lung toxicants develops the disease. Furthermore, the disease progresses at different rates in different people exposed to similar amounts of pollutants. COPD and lung cancer are the fourth and second leading cause of deaths in the US, respectively, and it is important to understand the genes and processes that may define the bifurcations for both these debilitating diseases with high lethality [13]. A better comprehension of the host genetic susceptibility and consequent differential regulation of pathogenic processes is required to make advances in directed therapy of COPD.

Oxidative and nitrosative stress induced by cigarette smoking is thought to be responsible for corticosteroid resistance in COPD [14–18][19][20]. Oxidant–antioxidant imbalance in the lungs has been strongly implicated in COPD severity and resistance to corticosteroids [21–23]. Strong epidemiologic and genetic evidence indicates that an individual's ability to defend against cigarette smoke–induced oxidative stress through up-regulation of lung antioxidant defenses is important, presenting oxidative stress as a critical event in the pathogenesis of COPD [24].

Although, the understanding of the underlying mechanisms of COPD is constantly evolving, the absence of any novel or effective therapy aimed at this irreversible disease presents a significant challenge [8,9]. There are limited effective therapies for COPD [25,26]. Therapies such as bronchodilators provide temporary symptomatic relief, while corticosteroids are not completely effective [27] [28,29]. Interestingly, recently, it is reported that the risk of pneumonia in patients with COPD increases with corticosteroid treatments [27]. There is no current therapy to arrest the long-term decline in lung function seen in COPD. Current therapy has not significantly decreased the mortality noted in susceptible former smokers even after years of smoking cessation. Hence, the understanding of the irreversible processes important in the pathogenesis of COPD seen in smokers may

provide a means of exploiting these destructive processes and genes associated with the disease process.

To explore the molecular definition of COPD, transcriptional regulatory networks were derived from airway gene expression data. Large collections of gene expression data provide regulatory patterns that potentially bear valuable insights regarding disease mechanisms. A number of predicted molecular participants involved in COPD etiology are identified in this report, concurring with the aging hypothesis for COPD [30]. This hypothesis, based on empirical evidence as presented by Aoshiba and Nagai (2009) [30], highlights the involvement of cellular senescence in key processes that characterize COPD such as chronic inflammation, increased susceptibility to infection, emphysematous lesions, and arrested tissue repair. Notable among the identified genes and their corresponding products are members of the T-box (TBX) family transcription factors and CDKN2A, which are associated with senescence [31,32][33,34]. The findings suggest that the balance between senescence and anti-senescence factors in normal smokers is disrupted towards senescence in COPD lungs.

Results

The experimental approach is summarized in Figure 1. Using the Context Likelihood of Relatedness (CLR) algorithm [35], data from both the U133A and the U133Plus_2 Affymetrix platforms were used to generate a transcriptional regulatory network consisting of genes involved in apoptosis, response to oxidative stress, and inflammatory response. The rationale for the focus on these processes is that they have been identified as key players in molecular pathogenesis of COPD [36][37], [38]. A union of the generated network consisted of 535 genes (represented at the nodes in Figure S1A) and 1474 interactions (represented as connections or edges between the nodes in Figure S1). Further details are presented in Table S1.

For the purpose of identifying the most influential nodes within the overall network, two features were used. First, the size of each node is an indication of its connectedness within the network: the large size nodes are more connected in the network. The larger nodes include HMOX1, TGFB1, TBX3, CDKN2A, PML, NME1, NPM1, SMAD3, RELA, FOXL2, STAT1, IL1B, TP63, NOTCH2, NFX1, ELF3, HIF1A, NLRP3, NFRKB, E2F1, TIAL1, AATF, TBX5, TCF7L2, HTAPIP2, TNF, ITCH, NFAM1, and CREB1. Second, the color of each node is an indication of the alterations in gene expression in COPD. For this purpose, a study on the differential gene expression in 15 COPD cases and 18 controls was used [39]. As several genes were represented by multiple probe sets on the arrays, the probe set with the median gene expression level was used. Pink nodes represent genes whose expression levels were unchanged in the study. Olive-green nodes represent genes whose median probe set expressions are suppressed in COPD. These include SMAD3, TBX3, TBX5, AATF, TCF7L2, NFX1, SEMA6A, HIP1, TNFRSF19, TNFSF10, DNAJB6, AGER, PRDX5, RAC1, NFATC3, PAWR, MGLL, SCYE1, NAE1, GPX3, SIRT2, HDAC6, HDAC1, and CAT. White nodes represent genes whose median probe set expressions are elevated in COPD. These include SOCS3, MCL1, IL1B, IL6, IL8, IL24, IL1RN, TNFAIP6, CCL3, CCL4, PTX3, ADORA3, NFAM1, NLRP3, BCL10, PPARD, FAIM3, PAX3, MAPK1, PRKCA, CASP2, SERPINB9, BCL2L11, TRADD, CAV1, and CDKN2A.

Combining these two features facilitated the identification of the most connected nodes that were also differentially expressed in COPD. We hypothesize that these nodes represent genes that may be critical regulators in the etiology of the disease. Among these,
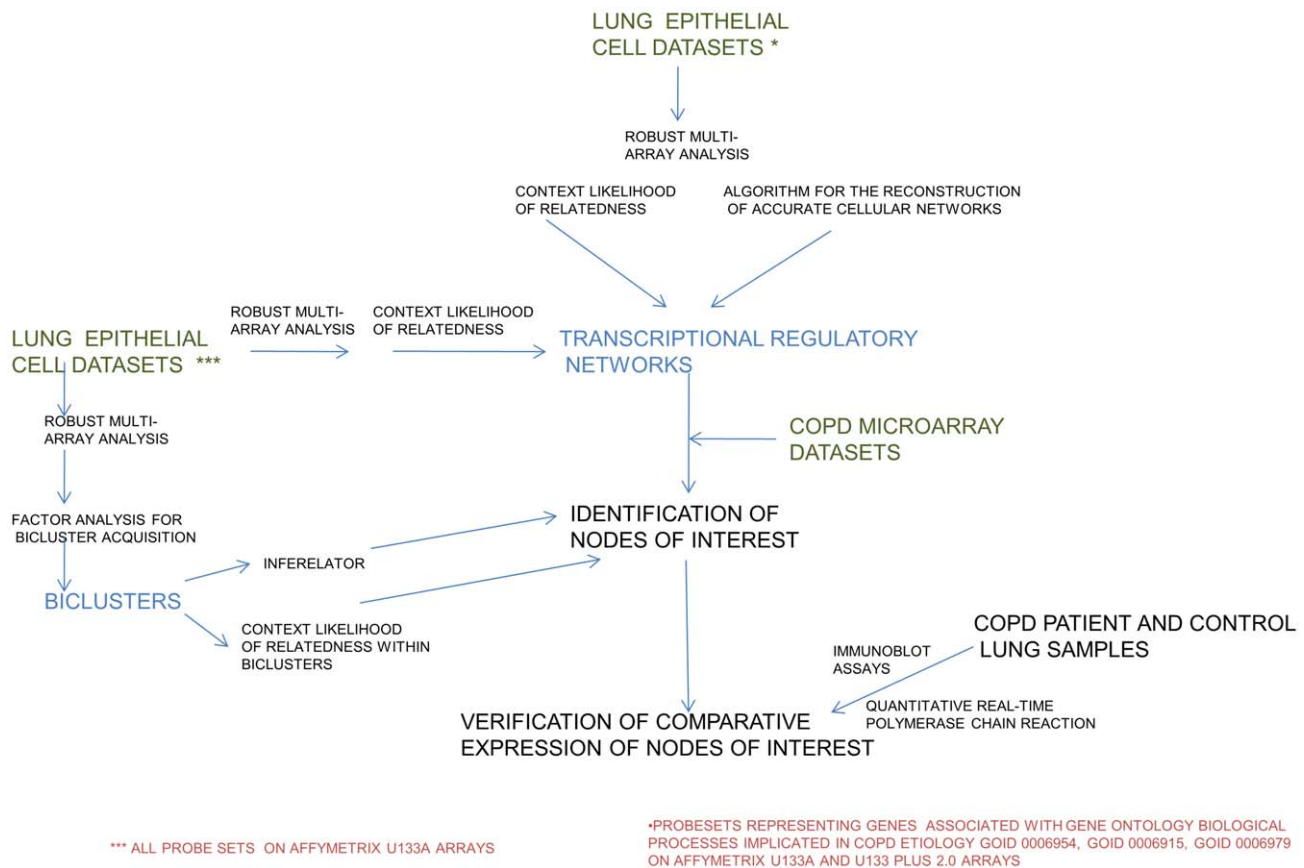
Figure 1. A view of the outline of the approaches used in this study. A combination of network inference and other algorithms applied to the datasets as described in the Materials and Methods section, led to the nodes of interest identified in the networks.
doi:10.1371/journal.pcbi.1002597.g001

the T-box (TBX) transcription factors, TBX3 and TBX5 (Figure S1B) along with the cell cycle inhibitor, CDKN2A (Figure S1B, C) were noteworthy for the extent of their connectedness. An enrichment analysis showed that genes dependent on TBX3 in the network are involved in apoptosis. However they also included genes associated with cell development, cell proliferation, and signal transduction. Of the 83 genes, 19 genes (e.g. TBX3) are involved in the cell cycle process. Similarly, of the 46 genes dependent on TBX5, 11 are involved in the regulation of the cell cycle. Quantitative real-time PCR (qRT-PCR) analysis of samples from smokers with or without COPD confirmed that relative to normal smokers without COPD, the expression of TBX3, TBX5, HDAC6, SIRT1, SIRT5, is suppressed in severe cases of COPD (Figure 2).

For confirmation purposes, a second transcriptional regulatory network was generated using the same data and an alternative algorithm, the Algorithm for the Reconstruction of Accurate Cellular Networks (ARACNE) [40]. Like CLR, ARACNE uses mutual information computed on the basis of gene expression data. ARACNE and CLR differ in their modes of discretizing and eliminating false edges. Unlike ARACNE, CLR uses B-spline functions for discretization of gene expression data [41]. Both, nonetheless, assert high connectivity for the same nodes of interest when direct as well as indirect close neighbors are considered. As summarized in Table 1, CLR was the more conservative of the two algorithms; most of the edges it predicted were predicted by ARACNE. For example, both algorithms infer the following as *direct* neighbors of TBX5 in the network generated: KNG1,

SNCA, RHOB, ALOX15B, SOCS3, RHOT1, MPO, AGTR2, BCL2, MAPK1, TBX3, SEMA6A, BBC3, PRKCA, BOK, LYST, PDCD6, PTEN, ADORA1, CD74, ACTN3, NLRP12, DNM2, PDIA2, BCL2L1, CDK5R1, TRAF7, CECR2, COL4A3, PAX3, GRM4, ACTN1, HSPD1, MAPK8, CDKN2D, TIA1, and PDCD1.

The exploratory study was then expanded to involve all probe sets available on the Affymetrix U133A platform (removing the focus away from apoptosis, oxidative stress, and inflammation genes). All 22,283 probe sets represented after robust multi-array analysis [42] on 109 human lung epithelia arrays were used. Using the CLR likelihood estimate cut-off of 2.5 (a threshold that captured known biological associations described as part of the Discussion section), a network consisting of 17,396 nodes and 127,331 edges was generated. At this cut-off, the previously established dependencies were detected along with additional ones. Using the study by Bhattacharya et al. [39], the same coloring scheme as indicated above was used for differentially expressed genes. The T-box transcription factors remained central in this enlarged network, with the TBX2 gene emerging as one of the most connected to other genes in the network (Figure S2A). The CDKN2A node remained highly connected in this network as well (Figure S2B). It is noteworthy that a large cross-section of the genes differentially expressed in COPD was found to be dependent on TBX2 and CDKN2A in this network. (Further details are presented in Tables S2 and S3 respectively.) This expanded study revealed links for additional genes such as CAV1 and certain histone deacetylases (including certain sirtuins), whose probable
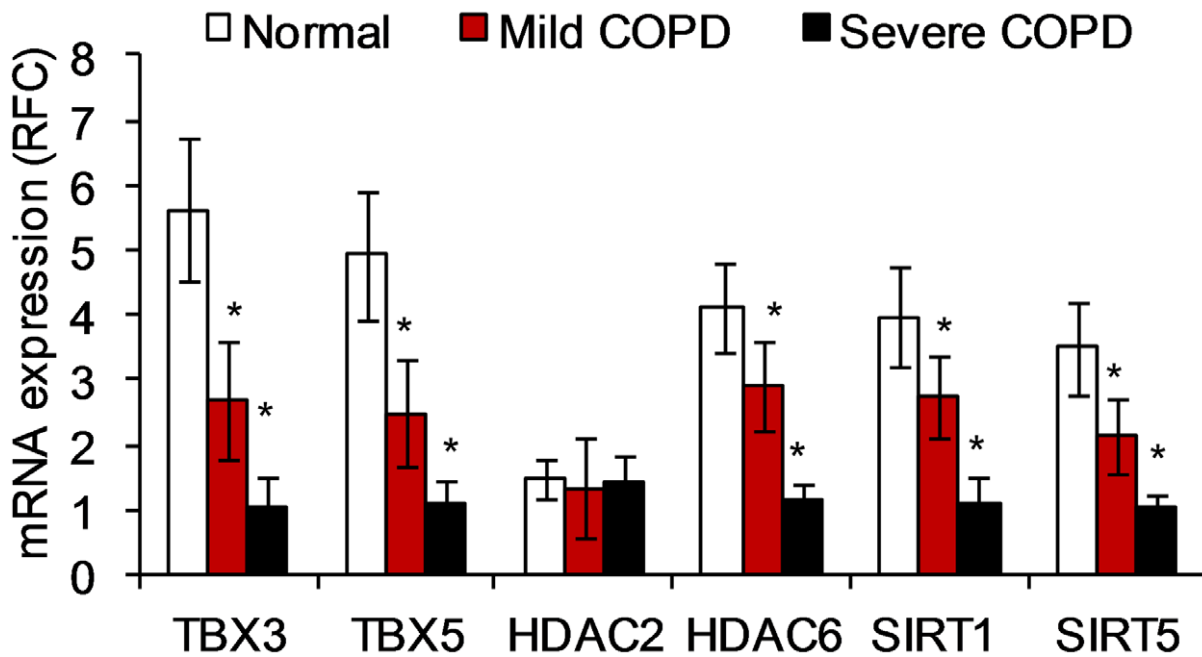
Figure 2. Quantitative PCR data indicate TBX3, TBX5, HDAC6, SIRT1, SIRT6 gene expression is suppressed in the lungs of patients with COPD, while mRNA expression of HDAC2 does not change compared to lung tissue from normal smokers. The extent of the suppression is highly significant between patients with mild and those with severe COPD. Patient diagnosis was based on the National Heart, Lung, and Blood Institute/World Health Organization Global Initiative for Chronic Obstructive Lung Disease (GOLD) [146]. Fifteen normal, nine mild COPD, and six severe COPD samples were used for this analysis. The data is represented as Mean ± S.D. The data was analyzed using student's t-test for comparing mRNA expression in the respective groups. *represents a significance of p-value<0.01.
doi:10.1371/journal.pcbi.1002597.g002

roles in COPD were indicated by altered expression in COPD lung samples (Figures 2 and 3).

For confirmatory purposes, the entire dataset of 22,283 probe sets and 109 observations (Gene Expression Omnibus datasets GDS534 and GDS999) was also subjected to a bicluster analysis. Each bicluster consists of a subset of probe sets and a collection of the observations (conditions) within which they are similar [43]. Members of the observation set are similar only when the given subset of probe sets is considered and vice versa. This unsupervised learning of subsets within the larger dataset constitutes an unbiased mechanism of identifying functional associations within the data. By helping to identify relationships among the probe sets,

the biclusters provided a means for re-examining the observations made in Figures S1 and S2 for possible corroboration. Using the Factor Analysis for Bicluster Acquisition (FABIA) algorithm [44], five, ten, then twenty biclusters were identified. Several probe sets grouped together in this way were also in transcriptional regulatory relationships per the CLR and ARACNE learning. As depicted in Table 2, CDKN2A, CDKN2B, COL4A3, COL4A3BP, CTNNA1, FOXJ2, FOXK2, FOXL1, FOXN3, HDAC6, HDAC9, IGF1, IGF2, IGF2BP3, PML, TBX1, TBX2, and TBX3 are all contained in the same bicluster (bicluster number 3) when ten biclusters were identified. Several of these (and related) genes also fall in the same bicluster when five and

Table 1. Proportion and list of direct neighbors also asserted in initial networks generated via other mutual information—based algorithm.

GENE	CLR	ARACNE	DIRECT NETWORK NEIGHBORS FOUND BY BOTH ALGORITHMS
TBX3	80/84	80/347	ACTN1 ACTN3 ACVR1B AIFM2 AVEN BAX BCL2L1 BCL2L10 BCL3 BCLAF1 BID BIRC7 BOK CASP10 CD74 CDK5R1 CDKN2D CECR2 CIDEA CLCF1 COL4A3 CRYAA DAPK2 DAPK3 DIABLO DLC1 DOCK1 F2 FOXO1 FURIN GLRX2 HIP1 IGF1R IL3 IL4I1 KIAA1967 KNG1 KRT18 LYST MAP3K10 MAPK1 MAPK8IP2 NCR1 NLRP12 NME1 NME2 NME5 NOL3 NPM1 PAX3 PCSK6 PDCD5 PDIA2 PHLDA2 PRKAA1 PRKCA PRKCZ PRLR RNF7 RTKN SCIN SEMA6A SERPINB2 SERPINB9 SFRP1 SMAD3 SOCS3 SPATA3 SST SSTR3 TBX5 TGFB1 TGFB2 TIA1 TIAL1 TLR2 TNFRSF19 TNFSF10 TP53 TRIAP1
TBX5	37/45	37/180	ACTN1 ACTN3 ADORA1 AGTR2 ALOX15B BBC3 BCL2 BCL2L1 BOK CD74 CDK5R1 CDKN2D CECR2 COL4A3 DNM2 GRM4 HSPD1 KNG1 LYST MAPK1 MAPK8 MPO NLRP12 PAX3 PDCD1 PDCD6 PDIA2 PRKCA PTEN RHOB RHOT1 SEMA6A SNCA SOCS3 TBX3 TIA1 TRAF7
CDKN2A	44/60	44/189	ADAMTSL4 ANXA1 ANXA4 API5 BBC3 BCL2 BCL2A1 BCLAF1 BTG1 CADM1 CASP8AP2 CD38 CDC2L2 CFLAR CRADD CYFIP2 DAPK2 DAXX DEDD2 EIF5A FASLG GCLC GULP1 HDAC3 HMGB1 IL1A MAPK8IP2 MCL1 NLRP1 NOTCH2 NOX5 PCSK6 PDCD1 PDIA2 PIK3R2 PML PRKCA PRKCE SPHK1 TAOK2 TGFB1 TNFRSF25 TNFRSF6B TP63

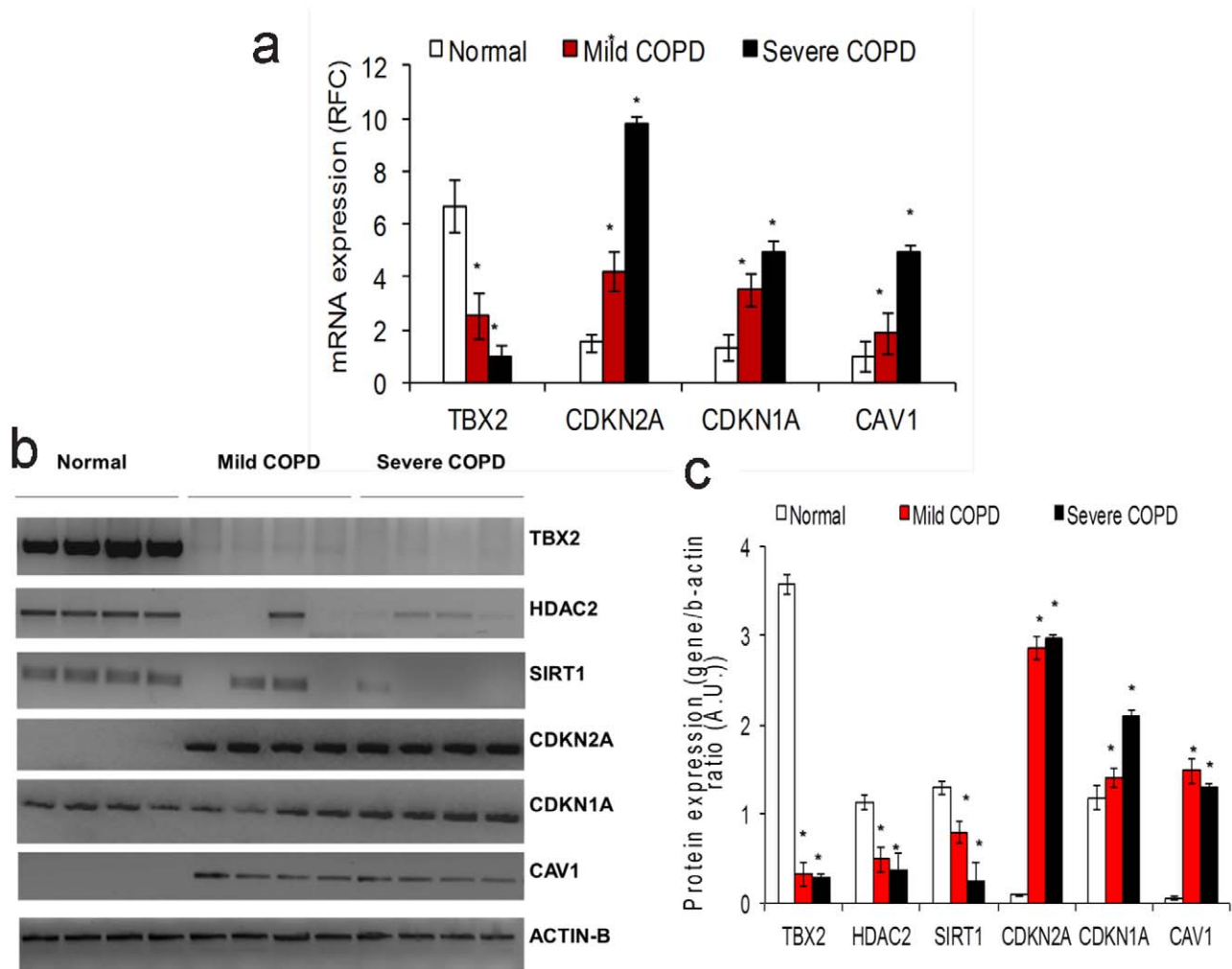doi:10.1371/journal.pcbi.1002597.t001

Figure 3. Patients with COPD have suppressed TBX2 and increased CDKN2A, CDKN1A, and caveolin-1 mRNA and protein expression. A) Quantitative PCR data indicate TBX2 gene expression is suppressed, while senescence factors, CDKN2A, CDKN1A, and caveolin-1 are induced in the lungs of patients with COPD compared to lung tissue from normal smokers. Fifteen normal, nine mild COPD, and six severe COPD samples were used for this analysis. The data is represented as Mean $\pm$ S.D. The data was analyzed using student's t-test for comparing mRNA expression in the respective groups. B) Representative Western blots showing suppressed TBX2, HDAC2, SIRT1 proteins and increased expression of CDKN2A, CDKN1A and caveolin-1 proteins in samples from patients with COPD. C) Densitometry analysis of Western blot data. Four normal, four mild COPD, and four severe COPD samples were used for this analysis. Densitometry analysis was carried out using image-J software. The data is represented as Mean $\pm$ S.D. The data was analyzed using student's t-test for comparing protein expression in the respective groups. *represents a significance of p-value<0.01.

doi:10.1371/journal.pcbi.1002597.g003

twenty biclusters are identified (Table 2). Tables S4A, S4B, and S4C contain complete listings of the membership of the various biclusters. Within-bicluster only CLR runs affirmed the regulatory relationship between TBX2 and several genes differentially expressed in COPD, as found before (Figure S3).

Furthermore, the Inferelator algorithm [45] was used to infer the transcription regulators targeting the FABIA-derived biclusters. Inferelator uses model shrinkage and standard regression to select predictive models for the expression of a gene or gene cluster, on the basis of expression levels of previously identified transcription regulators and interactions between them. Inferelator has been judged among the best performing network inference algorithms [46,47]. As shown in Table 3 (the output when twenty FABIA-generated biclusters were fed into Inferelator), TBX3 was a predicted positive regulator of bicluster 1. TBX5 was predicted to negatively regulate biclusters 2 and 7; it also negatively regulates bicluster 20 (along with HSF1). TBX2 cooperates with PML in a predicted negative regulation of bicluster 7; it also cooperates with EP300 in a predicted negative regulation of bicluster 11. Besides the TBX family members, other regulators of note identified in Table 3 include STAT1, STAT5A, STAT5B, RUNX3, SMAD3, PML, HSF1, JUN, NFKBIB, TP53, TP63, NFE2L2, PAX3, PAX7, ATF1, ATF2, CDKN1A and FOXO3. Similar outcomes were obtained when ten FABIA-generated biclusters were fed into Inferelator; four of the ten biclusters were predicted to be TBX gene product-regulated (Figure S4).

On the basis of these results, samples obtained from patients with COPD and normal subjects without COPD were examined for the relative levels of TBX2 and CDKN2A mRNA and protein expression. As shown in Figure 3, patients with COPD had elevated expression of CDKN2A and suppressed expression of TBX2 mRNA and protein. These findings are consistent with the findings from our exploratory studies (Figures S1 and S2). In addition, other senescence factors such as CDKN1A [48] and

Table 2. Examples of genes within Biclusters (of five, ten, and twenty Biclusters identified using FABIA).

Number of Biclusters Learned	Bicluster Number	Probe sets Present	Arrays Present	Some Genes Present	Description of Array Samples In Bicluster
5	Bicluster #4	924	35	CTNNA1, FOXO3, HDAC5, IGFBP2, IGFBP5, TBX2, WNT6	*Thirty four bronchoalveolar lavage samples from lung transplant recipients (7 acute rejection; 27 with no rejection). In addition, there is one array from a sample obtained from one **former smoker.
10	Bicluster #3	1749	30	CDKN2A, CDKN2B, COL4A3, COL4A3BP, CTNNA1, FOXJ2, FOXK2, FOXL1, FOXN3, HDAC6, HDAC9, IGF1, IGF2, IGF2BP3, PML, TBX1, TBX2, TBX3	*Bronchoalveolar lavage samples only. The samples were from 27 lung transplant recipients who had no rejection, and 3 who had acute rejection.
20	Bicluster #2	2096	30	CDKN2D, COL4A3, CTNNA1, CTNND2, FOXG1, FOXK2, FOXM1, FOXO3, HDAC2, IGFBP6, TBX2, WNT16	*Bronchoalveolar lavage samples only. The samples were from 27 lung transplant recipients who had no rejection, and 3 who had acute rejection.
20	Bicluster #8	1991	44	CDKN2A, CDKN2C, CIZ1, CTNNA1, FOXB1, FOXG1, FOXL2, FOXN3, HDAC5, IGF1, IGFBP2, IGFBP5, IGFBP7, IGF2BP3, PML, TBX1	*Thirty four bronchoalveolar lavage samples from lung transplant recipients (7 acute rejection; 27 with no rejection). This set also includes three arrays from samples from former smokers, five from current smokers, and two from ''never smokers''**.
20	Bicluster #13	567	56	COL4A3BP, FOXO3, IGFBP3, TBX2	*Thirty bronchoalveolar lavage samples from lung transplant recipients (3 acute rejection; 27 with no rejection). This set also includes seven arrays from samples from former smokers, eight from current smokers, and eleven from ''never smokers''**.

*Bronchoalveolar lavage samples obtained from lung transplant recipients whose biopsies had a perivascular score of between 0 and 2, and a bronchiolar score of between 0 and 1 [147].
**Lung epithelial cell transcriptome study of 34 current smokers, 18 former smokers, and 23 subjects who had never smoked [148].
doi:10.1371/journal.pcbi.1002597.t002

Table 3. Inferelator predictions of regulators of twenty Biclusters generated using FABIA.

BICLUSTER^	NUMBER OF PROBESETS	NUMBER OF OBSERVATIONS	CLUSTER PROBABLY POSITIVELY REGULATED BY***	CLUSTER PROBABLY NEGATIVELY REGULATED BY***
1	2084	32	TBX3, ELF3	ATF2_with_HSF1
2	2096	30	CEBPG, ATF1	PML, TP63, TBX5, TBX5_with_NFKBIB, PPARD_with_SMAD3
3	1560	31		CEBPG_with_TP63, NME1-NME2, TCF7L2, TIAL1
4	1569	38	NFE2L2_with_CDKN1A, STAT1	TCF7L2
5	2619	35	STAT5B_with_TP63, SMAD3	HTATIP2
6	1705	35	RUNX3	E2F1_and_TCFL2 , HTATIP2
7	1381	38	STAT1, CREBBP	E2F1, LOC652346, PML_with_TBX2, TBX5
8	1991	44	STAT5B, NFATC3	TCF7L2, TCF7L2_with_TIAL1
9	1089	33	NLRP3, SMAD3, RUNX3	
10	967	47	SIGIRR	
11	806	32	RUNX3, RUNX3_with_SMAD3	ATF2, EP300_with_TBX2, TCF7L2
12	566	34	TP63_and_STAT1	CEBPB, TCF7L2
13	567	56	FOXO3, SMAD3_with_ERC1, ELF3	STAT5A
14	493	50	ARHGDIA_with_RUNX3, RUNX3	PML, E2F1
15	754	45	ELF3	CREB1_with_TIAL1
16	356	43	PML_with_HSF1, LOC161527, HSF1, SBNO2	
17	510	46	SBNO2, JUN, RUNX3_with_CREB1, JUN_with_TP53, TP63	NFATC4_with_HTATIP2
18	380	61	ELF3_with_BCL10	ATF1
19	347	47	SMAD3	TGFB1_with_GOLGA6L4, STAT5A
20	497	50	HIF1A, STAT1, STAT5B, RELA	HSF1_with_TBX5
	^Obtained using FABIA algorithm			****Per Inferelator predictions

doi:10.1371/journal.pcbi.1002597.t003

caveolin-1 (CAV-1) [49,50] also showed enhanced expression in samples from patients with COPD (Figures 3A, B, and C). Both these genes have critical roles in senescence pathway activation. In our regulatory network, CDKN1A is connected to TBX2 via TP53 and MMP12; TBX3 via TP53 and FH; TBX5 via TP53 and SEC14L1. A number of previous reports have shown TBX2- and TBX3-mediated regulation of senescence factor, CDKN1A [51–53]. Interestingly, we found that CAV-1 is connected to TBX2 via ARHGDIA and TMED2. Both ARHGDIA and TMED2 are directly linked to the master transcriptional regulator of the anti-oxidant response, nuclear factor erythroid 2-related factor 2 (NRF2), i.e. NFE2L2, which was predicted to be a bicluster regulator along with CDKN1A (Table 3; also a regulator in one of ten FABIA-generated biclusters, Figure S4).

It is well known that COPD clusters in families [54]. Details of the genetic susceptibility loci continue to be studied. It has been determined that a polymorphism of the type IV collagen alpha3 (COL4A3) gene is associated with the risk of developing COPD [55]. Subjects with carriers of 451HH with at least one 451R allele had a higher COPD risk, which is more pronounced in younger subjects. Interestingly, in our analysis, COL4A3 expression was found to be suppressed in COPD. An examination of its expression within the networks generated indicated dependence on TBX genes (Figure 4A and Figures S1B, and S1C). Further, Figure 4A shows that the expression of COL4A3 depends on both CDKN2A and TBX2.

Taken together, these results indicate a critical balance between senescence and anti-senescence factors in normal smokers, which is disrupted towards senescence in COPD lungs. There are previous reports of decline in telomere length, which is a hallmark

of senescence in samples from patients with COPD [49,56–58]. On the other hand, increase in anti-senescence activity is reported as a hallmark of cancer [59–64] including lung cancer which, like COPD, is associated with smoking. These observations are in agreement with recent reports that highlight aging and associated senescence pathway as the key pathogenic molecular pathways involved in chronic lung diseases including lung cancer [65] and COPD [66–69].

Discussion

Using a variety of computational approaches, a number of regulatory genes important in COPD have been identified in these studies. First, using gene expression data of genes associated with three Gene Ontology biological processes implicated in COPD, transcriptional regulatory networks were learned by way of CLR and ARACNE. The most highly connected nodes of the networks which simultaneously represented genes differentially expressed in COPD were noted as important. The basic findings were also present in an expanded CLR study of all 22,283 probesets on the U133A Affymetrix platform. Differentially expressed and highly connected nodes of note included TBX2, TBX3, TBX5, and CDKN2A.

Bicluster analyses of the entire U133A platform dataset, which were unbiased in terms of a prior determination of genes of focus, found that the genes related to those noted in the CLR and ARACNE studies clustered together, often co-occurring in more than one bi-cluster (Table 2). CLR-generated regulatory networks involving only genes occurring within the same biclusters affirmed a central regulatory role for TBX2 (Figure S3). In addition,
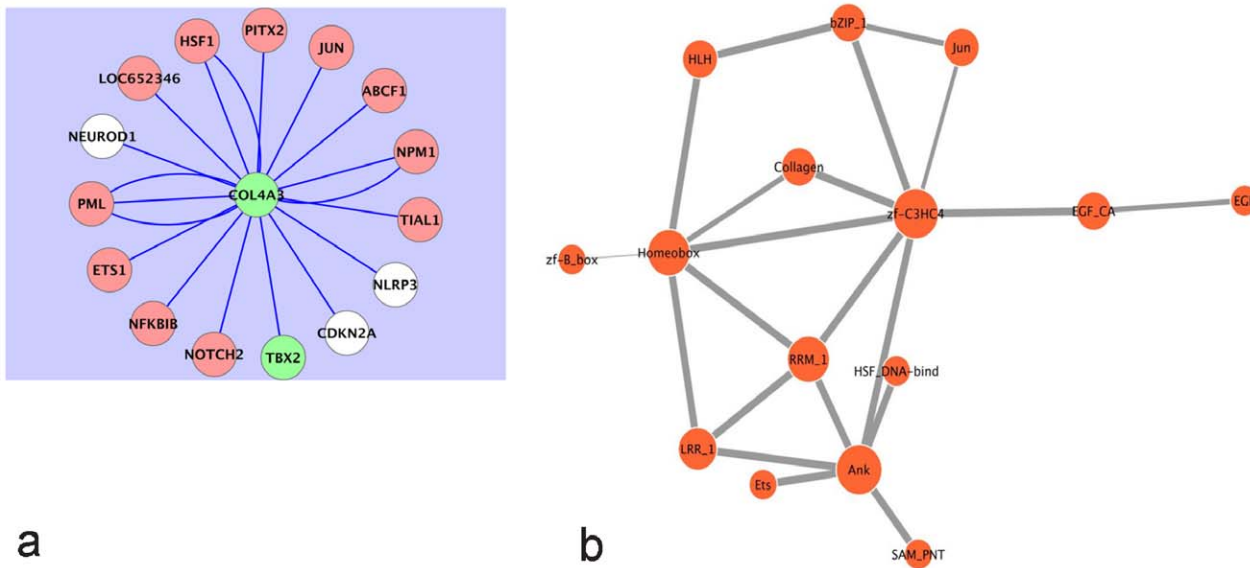

Figure 4. A) The state of the Type IV collagen alpha 3 subunit, COL4A3, depends on the states of both TBX2 and CDKN2A in human lung epithelial cells. Following Robust Multi-Array Analysis of a compendium of 109 Affymetrix arrays on the U133A platform, the Context Likelihood of Relatedness (CLR) algorithm was used to generate a transcriptional regulatory network involving all available probe sets (at a CLR likelihood estimate cut-off of 2.5). Olive-green nodes represent genes whose median probe set expressions are suppressed in COPD. White nodes represent genes whose median probe set expressions are elevated in COPD. COL4A3, whose expression is suppressed in the COPD lung, is thus statistically dependent on both TBX2 and CDKN2A. B) Evolutionarily conserved probable protein domain-domain interactions corresponding to the predictions of Table 5. The thickness of each edge is commensurate with the corresponding computed probabilities. The COL4A3 protein has the Collagen domain (Collagen in Pfam database; InterPro Database Accession IPR008160) and probably engages PML via its zf-C3HC4 domain (zf-C3HC4 in Pfam database; InterPro Database Accession IPR001841). By way of its Collagen domain, COL4A3 interacts with PITX2 via its Homeobox domain (Homeobox in Pfam database; InterPro Database Accession IPR001356). Among others, there is also a probable interaction between the zf-C3HC4 of PML and the Ankyrin repeat (Ank in Pfam database; InterPro Database Accession IPR002110) domain of CDKN2A that could impact the COPD etiology.
doi:10.1371/journal.pcbi.1002597.g004

Inferelator, which uses a very different approach from CLR and ARACNE, predicted that TBX gene products are involved in the overall regulation of between 25% and 40% of identified biclusters (Table 3, Figure S4). Thus, among a variety of computational approaches, there was a consensus regarding a regulatory role for TBX gene products in COPD.

T-box proteins are an important family of transcription factors. Over 20 genes in vertebrates have a region of homology to the DNA-binding domain of the transcription factor encoded by Brachyury, the T gene [70]. The region of homology in the gene product, called the T-box, has approximately 180 amino acid residues and is highly conserved across several species [71]. TBX transcription factors are characterized by a highly conserved DNA-binding T-box domain, which recognizes a consensus core sequence GGGTGTGA called the T-sites, and this domain is different from any other known DNA-binding motif [72]. A variety of T-box proteins serve as activators or repressors of their target genes, depending on cofactors involved [70,73] [74,75]. T-box proteins are critically important during development [76]. Importantly, with regard to COPD, T-box proteins are inhibitors of senescence [31,32]. TBX3, for instance, inhibits senescence in a process involving p53-dependent proliferation arrest [32].

CDKN2A is a mechanistic marker for cellular senescence [33,34]. The CDKN2A gene generates transcript variants that include p16INK4A and ARF. The cyclin-dependent kinase, CDK4, which is critical for cell cycle turnover, is inhibited by the p16INK4A product [77]. The other CDKN2A product, ARF, promotes the degradation of the p53 inhibitor, MDM2 and consequently promotes the accumulation and stabilization of p53 [78]. The accumulated p53 is subsequently able to activate the expression of genes involved in the arrest of the cell cycle at G1 or in apoptosis [79]. Thus, by arresting the progression of the cell cycle to promote repair of damaged DNA or by inducing apoptosis; p53 prevents accumulation of mutations that can be oncogenic.

Thus, the activities of both T-box proteins and the CDKN2A products converge on the p53 pathway. Our findings (Figure S1C) underscore the roles of T-box proteins and CDKN2A in the etiology of COPD and indicate that the expressions of these genes are linked in the human lung epithelium. Of note are the changes in expressions of TBX (-2, -3) and CDKN2A occur in opposite directions in cancer. T-box genes/proteins such as TBX2 and TBX3 are overexpressed in several neoplasms, including melanomas, breast, and pancreatic cancer [53,80,81]. On the other hand, because of its effect on MDM2, the CDKN2A product, ARF, acts as a tumor suppressor; consequently its loss is associated with neoplasms [82]. The other CDKN2A product, p16INK4A, also suppresses tumors, and is itself suppressed in neoplasms [77]. Secondly, in COPD, we show that expressions of TBX (-3, -5) are suppressed and expression of CDKN2A increases (Figures 2 and 3A and Figures S1B, and S1C). This is in distinct contrast to prior observations made in cancer. TBX2 belongs to the TBX2 sub-family of TBX transcription factors that include TBX3, TBX4, and TBX5 genes [71]. Protein expression of TBX2 also declines in the lungs of patients with COPD compared to control subjects, while the expression of CDKN2A protein is elevated in the lungs of patients with COPD compared to control subjects (Figures 3B and 3C). We also show an association between expression levels of TBX3, TBX5, and CDKN2A, an indication they are statistically associated (Figure S1C). TBX3 down-regulates the expression of the CDKN2A product, ARF [32]. Similarly, via the stress-activated p38 MAP kinase, the activated TBX2 localizes in the nucleus and represses the closely related CDKN1A (p21) promoter [51]; CDKN1A is an inhibitor of DNA repair [83,84]. In Figures

S2A and S2B, we show that both TBX2 and CDKN2A are highly connected in the human lung epithelium transcriptional regulatory network. We also show that many of the genes differentially expressed in the COPD lungs are statistically dependent on the expression of TBX2 and CDKN2A.

TBX2 and its close relative, TBX3, negatively regulate cell cycle control genes and CDKNs, specifically CDKN2A, CDKN2B, and CDKN1A [51,53] and are often upregulated in several cancers [53,80,85]. Additionally, in carcinogenesis, TBX2 destabilizes p53 by inhibition of ARF [86], while the retinoblastoma protein (Rb1) is an important modifier of TBX2 function [87]. Taken together, TBX2 and TBX3 act as suppressors of senescence factors such as CDKNs, and hence, suppress senescence especially in cancer cells; thus, TBX2 and TBX3 are thought to be critical anti-senescence factors [51,52,88,89].

As reported here, the expression of the anti-senescence T-box transcription factors are suppressed in the COPD lungs, and there is a concomitant rise in the expression of the cellular senescence markers, CDKN2A, CDKN1A, and CAV-1 (Figure 3). CDKN1A mediates cigarette smoke-induced inflammation [48]; cigarette smoke increases expression of CAV-1 which is involved in senescence and pulmonary emphysema induction [49,50]. Thus these findings support the aging hypothesis for COPD [30]. The two important relevant COPD risk factors converge. First, the incidence of COPD increases with age [90]. Secondly, cigarette smoking has been associated with the incidence of COPD. These associations suggest senescence in the lungs is probably due to the exposure to cigarette smoke. Indeed, studies in human lung epithelial cells and mice indicate that cigarette smoke induces senescence [91]. Cellular senescence is often associated with shortening of telomeres and is relevant to tissue aging [92]. One of the hallmarks of senescent cells is their tendency to generate a number of pro-inflammatory cytokines [93].

COPD Susceptibility Locus

Genome-wide studies in patients with COPD indicate that the chromosomal locus spanning 2q33.3–2q37.2 is associated with COPD [94] [95]. Among the genes linked by the CLR algorithm to TBX5 (Figure S1C), HSPD1, COL4A3, and PAX3 fall within this region, and BOK and PDCD1 lie on the edge of this locus (Table 4). HSPD1, also known as HSP60 or HSP65, is a member of the heat shock protein family. As shown in Figure S1C HSPD1, which has elevated expression in COPD is linked to the expression of TBX5, one of the central transcription factors with suppressed expression in COPD. Notably, HSF1 [96], a transcription factor responsible for the transcriptional activation of several heat shock proteins, is linked to COL4A3 (Figure 4A), which falls within a chromosomal locus of interest. Also, a polymorphism of the COL4A3 gene is associated with the risk of developing COPD [55].

COL4A3 [97] has suppressed expression in COPD (Figure S1 and Figure 4A), and is tied by the CLR algorithm to the transcription factors PML (also a tumor suppressor) and PITX2. PML and PITX2 have protein domains that likely interact with COL4A3 (Figure 4B, Table 5). By way of its Collagen domain, COL4A3 probably interacts with PITX2 via the PITX2 Homeobox domain (Homeobox in Pfam database; InterPro Database Accession IPR001356). PITX2 acts downstream of the Wnt-β-catenin pathway and responds to the activation of that pathway, by regulating the transcription of G1 cell cycle control genes such as cyclin D1 and c-Myc [98]. The dependence between PML and COL4A3 (per the CLR runs) is similarly interesting, because PML, a regulator of the cell cycle, plays a critical role in the regulation of cell proliferation, apoptosis, and senescence [99].

Table 4. Several chromosome 2 genes linked to senescence hubs in transcriptional regulatory network fall within COPD susceptibility locus, 2q33.3–2q37.2.

Hub	Direct Neighbor	Chromosomal Location
CDKN2A	ANXA4	2p13
	CFLAR	**2q33–34**
	GULP1	**2q32.3–q33**
	IL1A	2q14
	PDCD1	**2q37.3**
TBX3	**CASP10**	**2q33–34**
	PAX3	**2q35**
	RTKN	2p13.1
	TIA1	2p13
TBX5	**BOK**	**2q37.3**
	COL4A3	**2q36–37**
	HSPD1	**2q33.1**
	PAX3	**2q35**
	PDCD1	**2q37.3**
	PRKCE	2p21
	RHOB	2p24
	TIA1	2p13

Genes in bold character are in or border the region of the susceptibility locus.
doi:10.1371/journal.pcbi.1002597.t004

It induces a block of the G1 phase of the cell cycle in tumor cell lines [100] and enhances the transcriptional activity of the key tumor suppressors such as p53 and Rb [99]. The COL4A3 protein has the Collagen domain (Collagen in Pfam database; InterPro Database Accession IPR008160) and probably engages PML via its zf-C3HC4 domain (zf-C3HC4 in Pfam database; InterPro Database Accession IPR001841) (Figure 4B, Table 5). Our results also show a transcriptional regulatory relationship between PML and CDKN2A (probably by an interaction between the zf-C3HC4 of PML and the Ankyrin repeat domain of CDKN2A (Figure 4B, Table 5)). However, the significance of this observation for COPD remains unclear.

Other senescence genes. The importance of TBX2 and CDKN2A in the network has been highlighted. Other gene products associated with senescence were found via the CLR (using all available probe sets) to be statistically associated with, and thus probably dependent on, TBX2 and/or CDKN2A (Table 6). They include insulin/IGF-1 signaling genes which promote aging [101–103]; members of the forkhead family [104][105][106]; members of the WNT family and the bipartite transcription factor β-catenin/TCF [107][108][109][110][111][106]; as well as histone deacetylases (HDACs) and sirtuins (SIRTs) [112][113]. Furthermore, the association of TGBF1 with senescence-related genes in our study (Table 6) is an indication of a probable role in the etiology of COPD [114][115].

Our study furthers the paradigm on cellular senescence as its effectors and their regulation are the cross-roads of smoking-induced lung cancer or COPD [13,116,117]. Along with T-box transcription factors, other anti-aging molecules such as HDACs and SIRTs are decreased in the lungs of patients with COPD compared with smokers without COPD. This results in enhanced inflammation that furthers the progression of COPD [118,119]. The genes identified here may serve as mechanistic biomarkers for detecting impending physiological changes during the disease

Table 5. Maximum likelihood estimation indicating probabilities of protein-protein interactions based on evolutionarily conserved domain-domain interactions.

	MLE Probability	
CDKN2A	0.99	ETS1
CDKN2A	1	NFKBIB
CDKN2A	0.94	NLRP3
CDKN2A	1	NOTCH2
CDKN2A	1	PML
CDKN2A	0.99	TIAL1
COL4A3	**0.68**	**PITX2**
COL4A3	**0.9**	**PML**
ETS1	0.99	NFKBIB
ETS1	0.99	NOTCH2
HSF1	0.9	NFKBIB
HSF1	0.9	NOTCH2
JUN	0.99	NEUROD1
NEUROD1	0.99	PITX2
NLRP3	0.94	NFKBIB
NLRP3	0.94	NOTCH2
NLRP3	0.99	PITX2
NLRP3	1	TIAL1
NOTCH2	0.99	ETS1
NOTCH2	1	NFKBIB
NOTCH2	1	PML
PML	1	JUN
PML	1	NFKBIB
PML	1	NOTCH2
PML	1	PITX2
TIAL1	0.99	NFKBIB
TIAL1	0.99	NOTCH2
TIAL1	0.99	NOTCH2
TIAL1	1	PITX2
TIAL1	1	PITX2
TIAL1	1	PML

MLE = Maximum likelihood estimation.
Probable interactions involving COL4A3, which is in the susceptibility locus, are in bold character.
doi:10.1371/journal.pcbi.1002597.t005

process. Further, this understanding can help in designing directed therapies with further understanding of genomic, molecular, and physiological changes in patients with emphysema or COPD [120,121].

Materials and Methods

Ethics Statement

The study protocols were approved by the Institutional Review Board for human studies, and patients' lung function data from each of the contributing centers were obtained for this study.

Etiology of COPD and Microarray

The overall experimental approach is summarized in Figure 1. The etiology of COPD has been associated with apoptosis

Table 6. Several aging-related genes are statistically linked to the expressions of TBX2, CDKN2A, and TGFB1.

	TBX2	CDKN2A	TGFB1
Catenin-related genes	CTNNA1, CTNND1		CTNNBIP1
Forkhead transcription factor-related genes	FOXA2, FOXB1, FOXD3, FOXH1, FOX01, FOXO3		FOXA1, FOXJ1, FOXO4
Insulin Growth factor-related genes	IGF2, IGF2BP3, IGFBP7, IGFBP5	IGFALS	IGFALS, IGFBP4, IGF2PB2, IGFBP7, IGFBP5, IGF2R, IGFBP3
Interleukin-related genes	IL4, IL13, IL13RA1, IL27RA, IL1RL1, IL17RC, IL10RB, IL2RA, IL12RB1	IL1RN, IL6ST, IL1RAPL1	IL20RA, IL27RA, IL13RA1, IL2RG, IL10RA, IL4I1
Wnt-related genes	WNT3, WNT4, WNT6, WNT7B, WNT11	WNT10B	

doi:10.1371/journal.pcbi.1002597.t006

[122,123], oxidative stress [124], and inflammation [125]. The computational aspects of these studies were conducted in two phases. During the first phase, transcriptional regulatory networks of the human lung epithelium were reverse-engineered from publicly available gene expression data, using the CLR network inference algorithm [35]. Subsets of a compilation of publicly available gene expression data were generated based on genes classified under the Gene Ontology [126] term "inflammatory response" (GOID 0006954), "apoptosis" (GOID 0006915), and "response to oxidative stress" (GOID 0006979), using a program written for that purpose in lisp [127]. Thus, for each microarray platform (Affymetrix U133A and U133Plus_2), a transcription regulatory network was generated that consists of a union of statistical dependencies among the genes involved in inflammatory response, apoptosis, and response to oxidative stress. For comparison and confirmation purposes, a second transcriptional regulatory network was generated using ARACNE and the same datasets. Like CLR, ARACNE uses mutual information computed on the basis of gene expression data, as discussed below. ARACNE and CLR differ in their modes of binning and eliminating false edges. CLR was the more conservative of the two algorithms, most of the edges it asserted also having been asserted by ARACNE (Table 1).

Expanding on the findings of the first phase, the entire set of probe sets represented on 109 arrays of the U133A platform was used during the second phase in which the CLR algorithm was executed (Gene Expression Omnibus datasets GDS534 and GDS999). Subsequently to assure the reliability of the regulatory relationships just described, biclusters were identified within the dataset using the FABIA algorithm. The Inferelator algorithm was then also used to predict regulators of those biclusters.

Public Gene Expression Data

A compendium of microarray data was generated from the Gene Expression Omnibus (GEO). GEO record numbers GDS534, GDS999, GDS2604, and GDS2486 (http://www.ncbi.nlm.nih.gov/geo/) and contains the gene expression microarray experiment data from human lung epithelial cells under a variety of conditions. These arrays are based on the Affymetrix (http://www.affymetrix.com) human U133A and U133Plus_2 platforms. In all, there were 109 arrays from the human U133A and 49 arrays from the human U133Plus_2 platforms, respectively. For each platform, gene expression data files in the .CEL format were downloaded and subjected to Robust Multiarray Analysis [42], using the Bioconductor (http://www.bioconductor.org) package *affy* in R (http://cran.r-project.org/).

CLR

The CLR algorithm [35] is an improvement on the Relevance Networks algorithm [128]. Both use the concept known as mutual information to infer the state of one member of a given gene pair, given the state of the other member of the pair [129,130]. There is a need to capture biologically relevant links within the resulting reverse-engineered networks. In relevance networks, mutual information score thresholds are applied. Low thresholds tend to capture dense networks with many false positives which inevitably include misrepresentations of indirect dependencies as direct interactions. On the other hand, high thresholds result in much smaller networks albeit with fewer false positives. The CLR uses an adaptive background correction step to remove indirect influences and false correlations. It compares the mutual information value for a given pair of genes to a background distribution of mutual information scores.

Thus a likelihood estimate:

$$\sqrt{X_z{}^2 + Y_z{}^2}$$

is used (where X_z is the z-score of the mutual information between gene X and gene Y in gene X's mutual information score distribution, and Y_z is the z-score of the mutual information between gene X and gene Y in gene Y's mutual information score distribution).

Mutual information. Given two random variables, X and Y, the mutual information between them is given by:

$$M(X,Y) = H(Y) - H(Y|X) = H(X) - H(X|Y)$$

$$M(X,Y) = H(X) + H(Y) - H(X,Y)$$

(Here the Shannon entropy [130], the probability of observing a particular symbol or event, p_i,

$$H = -\, \Sigma\, p_i \log_2 p_i$$

is used as a measure of quantitative information).

In other words, the shared information between X and Y corresponds to the remaining information of one party if we remove the information of that party that is not shared with the other party. For two genes, X and Y, the mutual information is given by:

$$M(X;Y) = \sum_{i,j} P(x_i, y_j) \log \frac{p(x_i, y_j)}{p(x_i) p(y_j)}$$

x_i and y_i represent specific expression levels across a given set of measurements. The mutual information thus ranges between 0 and 1, and is a measure for dependencies in the data: negative or positive, nonlinear or linear [131,132]. The higher the mutual information score between the two genes, the greater the information inferred on the states of the first gene from the pattern of states in the second.

CLR execution. The code implementation provided by Faith et al. was used from a Linux command line [35]. During the first phase, the CLR runs were conducted on the data subsets outlined above. For each data subset, gene subsets that are also identified in the Gene Ontology as transcription factors using GeneInfoViz [133] were designated as transcription regulators as part of the execution of the algorithm. Other details of this CLR run are in Table S5.

In the second phase, all 22,283 probe sets represented after background correction and normalization on the U133A platform were used (Gene Expression Omnibus datasets GDS534 and GDS999). A likelihood estimate cut-off value of 2.5 used generated a network consisting of 17,396 nodes and 127, 331 transcription regulatory links.

ARACNE

Like CLR, ARACNE uses mutual information [40]. Unlike CLR, it uses the Data Processing Inequality (DPI) to retain only those regulatory relationships that are direct (rather than indirect) [134]. In other words, if genes g1 and g3 interact only through a third gene, g2, then DPI indicates:

$$I(g1, g3) \leq \min [I(g1, g2); I(g2, g3)]$$

Thus of the trio, the edge with the least value gets eliminated. The "DPI tolerance" used for ranking of I values, to minimize the impact of I value variance was set at 0.15 in this study. DPI tolerance values of greater than 0.2 have been determined to yield high false positive edges by the developers of ARACNE. Furthermore, the threshold p-value for establishing that the mutual information between gene pairs was significant in this study was set at 10^{-7}.

Factor Analysis for Bicluster Acquisition (FABIA)

The algorithm [44] was run using the entire dataset derived from the 109 U133A arrays. For p biclusters and additive noise, the model for the matrix X (input to biclustering method) is:

$$X = \sum_{i=1}^{p} \lambda_i z_i^T + \Upsilon,$$

where the real numbers λ, $z_{i,}$ and Υ, are the sparse prototype column vector, the sparse vector of factors (transposed as rows) with which the prototype vector is scaled for the ith bicluster, and the additive noise respectively. For each of the bicluster sets (where p is 5, 10 or 20), there were 500 iterations, and a sparseness factor of 0.1. All parameters were set at the default values.

Inferelator

The Inferelator [45] version 1.0 was used to infer a minimal set of regulators that explains the expression levels of each of the 10 and 20 biclusters identified. Potential regulators were defined (as was done for the CLR execution), and the expression data was treated as equilibrium observations (i.e. no information about temporal relationships between observations was incorporated into the inference process). The Inferelator was run with default settings.

Previous COPD Studies and Network Graphics

The human lung transcription regulatory networks generated were subsequently analyzed in the light of GEO datasets GSE1122, GSE1650, and GSE8581, representing studies on changes in gene expression between emphysema subjects and control subjects [135], patients with severe COPD and patients with symptoms ranging from mild COPD to normal [136], and patients with COPD and control patients [39]. CEL files were downloaded, and in each case, the data analyzed for differential gene expression using Partek workbench [137] after Robust Multiarray Analysis [42] processing (p-value = 0.01 and false discovery rate = 0.01) [138]. For each gene represented by differentially expressed probe sets, the median probe set value was used to represent the level of expression during visualization in Cytoscape [139]. Genes suppressed in COPD (compared to controls) were depicted using olive green-colored nodes and up-regulated genes had white-colored nodes. In our studies GSE8581 data [39] were used.

Domain Analysis

Proteins interact with each other via their component domains. An accurate prediction of domain-domain interactions would facilitate the prediction of protein-protein interactions. The Pfam database [140] contains a large collection of evolutionarily conserved protein domains and empirically determined interactions they are involved in. Based on relevant Pfam domain family data, a Maximum Likelihood Estimation method was used to infer protein-protein interactions among connected nodes of the transcriptional regulatory networks generated as described above [141][142].Although the networks were generated on the basis of gene array data, they are proxies for interactions among corresponding gene products (proteins). Indeed genes with similar expression patterns often generate proteins that interact in some fashion [143,144]. An implementation of the maximum likelihood estimation in Cytoprophet [145] was used, and probabilities of domain-domain (protein-protein) interactions were computed.

COPD Patient and Normal Lung Samples

Frozen peripheral lung tissue samples used in this study were obtained from two tissue banks: (1) the NHLBI Lung Tissue Research Consortium (University of Colorado Health Sciences Center, Denver, CO); and (2) the iCAPTURE (James Hogg iCAPTURE Centre for Cardiovascular and Pulmonary Research, St. Paul's Hospital, University of British Columbia, Vancouver, BC, Canada). We obtained data on patients' lung function from both established patient registries. Clinical information, samples size, and classification based on Global Initiative for Obstructive Lung Disease (GOLD) for Chronic Obstructive Lung Disease stages of patients with COPD and normal control subjects are summarized in Table 7. Participants in the COPD groups who smoked had similar pack-year smoking histories, where smoking for 1 pack-year refers to smoking one pack of cigarettes per day each year.

Quantitative Real-Time Polymerase Chain Reaction (qRT-PCR)

Selected genes (TBX2, TBX3, TBX5, CDKN2A, CDKN1A, HDAC2, HDAC5, SIRT1, SIRT5, and CAV1) from our analysis

Table 7. Patient characteristics.

Characteristic	Normal Samples	COPD Samples
GOLD Stage (1/2/3/4)	15	0/9/6/0
Sex (Male/Female)	15/0	15/0
Age (mean ± SD), years	69.9±14.4	68.4±15.1
Pack years smoked, (mean ± SD)	51.4±11.1	58.3±8.3
FEV$_1$ % predicted, (mean ± SD)	95.1±9.3	31.3±29.9
FVC % predicted, (mean ± SD)	88.7±8.9	60.1±18.1

FEV$_1$ = Forced expiratory volume at 1 sec; FVC = Function vital capacity; SD = Standard deviation; COPD = Chronic obstructive pulmonary disease. GOLD (Global Initiative for Chronic Obstructive Lung Disease) Stages:
1 – mild COPD: FEV$_1$≥80% predicted, FEV$_1$:FVC<70%.
2 – moderate COPD: 50%≤FEV$_1$≤80% predicted, FEV$_1$:FVC<70%.
3 – severe COPD: 30%≤FEV$_1$≤50% predicted, FEV$_1$:FVC<70%.
4 – very severe COPD: FEV1<30%predicted or FEV$_1$<50% predicted with chronic respiratory failure, FEV$_1$:FVC<70%.
Pack years: (Packs smoked per day)×(years as a smoker).
doi:10.1371/journal.pcbi.1002597.t007

were validated by qRT-PCR. Total mRNA from the peripheral lung tissues from patients with COPD and non-COPD individual's lungs were purified using the Qiagen RNeasy kit (Qiagen, Valencia, CA). qRT-PCR was then performed using inventoried Assay-on-Demand primers and probe sets from Applied Biosystems (Foster City, CA). We used the ABI 7000 Taqman system (Applied Biosystems) to perform these assays. β-actin was used as a normalization control. The analysis was run as previously described [19].

Immunoblot Assay

Immunoblots were performed using antibodies for TBX2, CDKN2A, CDKN1A, SIRT1, CAV1, HDAC2, and ACTIN-B (Santa Cruz Biotechnology, Santa Cruz, CA). ACTIN-B was used as a loading control. These immunoblots were performed using protocols as described previously [19].

Statistical Analyses for q-RT-PCR and Immunoblot Analysis

Fifteen normal, nine mild COPD, and six severe COPD samples were used for q-RTPCR analysis. Four samples per group were used for immunonoblots. All immunoblots were quantified by measuring scanned photographs in ImageJ software (NIH). All statistical analyses were done with student's t-test for comparisons of COPD groups with normal samples as control. Data in graphs were represented as mean values and error bars in the graphs represent standard deviation (SD).

Supporting Information

Figure S1 CLR-Generated transcriptional regulatory network of human lung epithelial cells. Following Robust Multi-Array Analysis of a compendium of 158 Affymetrix arrays, the Context Likelihood of Relatedness (CLR) algorithm was used to generate a transcriptional regulatory network (false discovery rate, 0.05). A) A synoptic view of the overall lung epithelial transcriptional regulatory network generated using Gene Ontology genes associated with apoptosis, response to inflammation, and response to oxidative stress. B) An up-close view of TBX3 and nodes directly connected to it in the network generated. C) An up-close view of TBX5 and nodes directly connected to it in the network generated. Larger-sized nodes represent hubs within the network,

i.e. human lung epithelium cell genes more highly connected to other genes associated with apoptosis, response to inflammation, and response to oxidative stress. Olive-green nodes represent genes whose median probe set expressions are suppressed in COPD. White nodes represent genes whose median probe set expressions are elevated in COPD.
(TIF)

Figure S2 The states of a large cross-section of human epithelial cell genes differentially expressed in COPD depend on the states of (A) TBX2 and (B) CDKN2A. Following Robust Multi-Array Analysis of a compendium of 109 Affymetrix arrays on the U133A platform, the Context Likelihood of Relatedness (CLR) algorithm was used to generate a transcriptional regulatory network involving all available probe sets (at a CLR likelihood estimate cut-off of 2.5). Olive-green nodes represent genes whose median probe set expressions are suppressed in COPD. White nodes represent genes whose median probe set expressions are elevated in COPD. TBX2 gene expression is suppressed while CDKN2A gene expression is elevated in COPD.
(TIF)

Figure S3 TBX2 is statistically associated with a large cross-section of genes differentially expressed in the COPD lung. Following Robust Multi-Array Analysis, ten biclusters were identified using FABIA from a compendium of 109 Affymetrix human lung epithelial cell microarrays. Focusing only on genes present in each cluster, the Context Likelihood of Relatedness (CLR) algorithm was used to generate Transcriptional Regulatory Networks (false discovery rate, 0.05). The ten networks are merged in this figure. Olive-green nodes represent genes whose median probe set expressions are suppressed in COPD. White nodes represent genes whose median probe set expressions are elevated in COPD. TBX2, the olive node in the center, is either directly or indirectly (by way of one or two intervening nodes) linked with a significant cross-section of genes differentially expressed in the COPD lung.
(TIF)

Figure S4 TBX gene products (captured by red arrows in figure) are predicted to be involved in the direct regulation of 40% of biclusters in the dataset. Following Robust Multi-Array Analysis, ten biclusters were identified using FABIA from a compendium of 109 Affymetrix human lung epithelial cell microarrays. Inferelator version 1.0 was used to infer a minimal set of regulators that explain the expression levels of each of 10 biclusters. TBX2 was predicted to be involved in the direct regulation of biclusters 6 and 10. TBX3 was predicted to be involved in the direct regulation of bicluster 10. TBX5 was predicted to be involved in the direct regulation of biclusters 2 and 8.
(TIF)

Table S1 Interacting nodes in CLR-generated network and their corresponding likelihood estimates. The data in this table correspond to Figure S1.
(DOC)

Table S2 Direct Connections to TBX2 in the CLR-Generated Network. The data in this table correspond to Figure S2A.
(DOC)

Table S3 Direct Connections to CDKN2A in the CLR-Generated Network. The data in this table correspond to Figure S2B.
(DOC)

Table S4 A: Membership of *Five* Biclusters Learned Using Factor Analysis for Bicluster Acquisition (FABIA). This table

identifies the genes and phenotypes that clustered together when the algorithm was applied to learn *five* biclusters from the experiments in the compendium. B: Membership of *Ten* Biclusters Learned Using Factor Analysis for Bicluster Acquisition (FABIA). This table identifies the genes and phenotypes that clustered together when the algorithm was applied to learn *ten* biclusters from the experiments in the compendium. C: Membership of *Twenty* Biclusters Learned Using Factor Analysis for Bicluster Acquisition (FABIA). This table identifies the genes and phenotypes that clustered together when the algorithm was applied to learn *twenty* biclusters from the experiments in the compendium. (DOC)

Table S5 Details of CLR Execution that Generated Results Found in Figure S1 and Table S1. (DOC)

Author Contributions

Conceived and designed the experiments: GKAM DM SB. Performed the experiments: GKAM DM MV JEM. Analyzed the data: GKAM DM JEM SB. Contributed reagents/materials/analysis tools: GKAM JEM SB. Wrote the paper: GKAM DM JEM SB.

References

1. Yoshida T, Tuder RM (2007) Pathobiology of cigarette smoke-induced chronic obstructive pulmonary disease. Physiol Rev 87: 1047–1082. 10.1152/physrev.00048.2006.
2. Rabe KF, Hurd S, Anzueto A, Barnes PJ, Buist SA, et al. (2007) Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. Am J Respir Crit Care Med 176: 532–555. 10.1164/rccm.200703-456SO.
3. Fromer L, Cooper CB (2008) A review of the GOLD guidelines for the diagnosis and treatment of patients with COPD. Int J Clin Pract 62: 1219–1236. 10.1111/j.1742-1241.2008.01807.x.
4. Mathers CD, Boerma T, Ma Fat D (2009) Global and regional causes of death. Br Med Bull 92: 7–32. 10.1093/bmb/ldp028.
5. Mathers CD, Loncar D (2006) Projections of global mortality and burden of disease from 2002 to 2030. PLoS Med 3: e442. 10.1371/journal.pmed.0030442.
6. Hogg JC, Timens W (2009) The pathology of chronic obstructive pulmonary disease. Annu Rev Pathol 4: 435–459. 10.1146/annurev.pathol.4.110807.092145.
7. Macnee W (2007) Pathogenesis of chronic obstructive pulmonary disease. Clin Chest Med 28: 479–513, v 10.1016/j.ccm.2007.06.008.
8. MacNee W, Tuder RM (2009) New paradigms in the pathogenesis of chronic obstructive pulmonary disease I. Proc Am Thorac Soc 6: 527–531. 10.1513/pats.200905-027DS.
9. Sethi S, Mallia P, Johnston SL (2009) New paradigms in the pathogenesis of chronic obstructive pulmonary disease II. Proc Am Thorac Soc 6: 532–534. 10.1513/pats.200905-025DS.
10. Salvi SS, Barnes PJ (2009) Chronic obstructive pulmonary disease in non-smokers. Lancet 374: 733–743. 10.1016/S0140-6736(09)61303-9.
11. Mannino DM, Buist AS (2007) Global burden of COPD: Risk factors, prevalence, and future trends. Lancet 370: 765–773. 10.1016/S0140-6736(07)61380-4.
12. Buist AS, McBurnie MA, Vollmer WM, Gillespie S, Burney P, et al. (2007) International variation in the prevalence of COPD (the BOLD study): A population-based prevalence study. Lancet 370: 741–750. 10.1016/S0140-6736(07)61377-4.
13. Houghton AM, Mouded M, Shapiro SD (2008) Common origins of lung cancer and COPD. Nat Med 14: 1023–1024. 10.1038/nm1008-1023.
14. Barnes PJ (2003) New concepts in chronic obstructive pulmonary disease. Annu Rev Med 54: 113–129. 10.1146/annurev.med.54.101601.152209.
15. Bowler RP, Barnes PJ, Crapo JD (2004) The role of oxidative stress in chronic obstructive pulmonary disease. COPD 1: 255–277.
16. Kostikas K, Papatheodorou G, Psathakis K, Panagou P, Loukides S (2003) Oxidative stress in expired breath condensate of patients with COPD. Chest 124: 1373–1380.
17. Malhotra D, Thimmulapapa R, Navas-Acien A, Sandford A, Elliott M, et al. (2008) Decline in NRF2-regulated antioxidants in chronic obstructive pulmonary disease lungs due to loss of its positive regulator, DJ-1. Am J Respir Crit Care Med 178: 592–604. 10.1164/rccm.200803-380OC.
18. Osoata GO, Hanazawa T, Brindicci C, Ito M, Barnes PJ, et al. (2009) Peroxynitrite elevation in exhaled breath condensate of COPD and its inhibition by fudosteine. Chest 135: 1513–1520. 10.1378/chest.08-2105.
19. Malhotra D, Thimmulapapa R, Vij N, Navas-Acien A, Sussan T, et al. (2009) Heightened endoplasmic reticulum stress in the lungs of patients with chronic obstructive pulmonary disease: The role of Nrf2-regulated proteasomal activity. Am J Respir Crit Care Med 180: 1196–1207. 10.1164/rccm.200903-0324OC.
20. Malhotra D, Thimmulapapa RK, Mercado N, Ito K, Kombairaju P, et al. (2011) Denitrosylation of HDAC2 by targeting Nrf2 restores glucocorticosteroid sensitivity in macrophages from COPD patients. J Clin Invest 121: 4289–4302. 10.1172/JCI45144; 10.1172/JCI45144.
21. Osoata GO, Yamamura S, Ito M, Vuppusetty C, Adcock IM, et al. (2009) Nitration of distinct tyrosine residues causes inactivation of histone deacetylase 2. Biochem Biophys Res Commun 384: 366–371. 10.1016/j.bbrc.2009.04.128.
22. Hansel TT, Barnes PJ (2009) New drugs for exacerbations of chronic obstructive pulmonary disease. Lancet 374: 744–755. 10.1016/S0140-6736(09)61342-8.
23. Ito K, Barnes PJ (2009) COPD as a disease of accelerated lung aging. Chest 135: 173–180. 10.1378/chest.08-1419.
24. Yang IA, Relan V, Wright CM, Davidson MR, Sriram KB, et al. (2011) Common pathogenic mechanisms and pathways in the development of COPD and lung cancer. Expert Opin Ther Targets 15: 439–456. 10.1517/14728222.2011.555400.
25. Barnes PJ (2010) Inhaled corticosteroids in COPD: A controversy. Respiration 80: 89–95. 10.1159/000315416.
26. Mroz RM, Skopinski T, Holownia A, Chyczewska E, Braszko JJ (2011) Treatment of chronic obstructive pulmonary disease–traditional bronchodilatation and targeted anti-inflammatory therapy]. Pneumonol Alergol Pol 79: 32–38.
27. Drummond MB, Dasenbrook EC, Pitz MW, Murphy DJ, Fan E (2008) Inhaled corticosteroids in patients with stable chronic obstructive pulmonary disease: A systematic review and meta-analysis. JAMA 300: 2407–2416. 10.1001/jama.2008.717.
28. Alifano M, Cuvelier A, Delage A, Roche N, Lamia B, et al. (2010) Treatment of COPD: From pharmacological to instrumental therapies. Eur Respir Rev 19: 7–23. 10.1183/09059180.00008009.
29. Antus B (2010) Role of exhaled nitric oxide in predicting steroid response in chronic obstructive pulmonary disease]. Orv Hetil 151: 2083–2088. 10.1556/OH.2010.28972.
30. Aoshiba K, Nagai A (2009) Senescence hypothesis for the pathogenetic mechanism of chronic obstructive pulmonary disease. Proc Am Thorac Soc 6: 596–601. 10.1513/pats.200904-017RM.
31. Jacobs JJ, Keblusek P, Robanus-Maandag E, Kristel P, Lingbeek M, et al. (2000) Senescence bypass screen identifies TBX2, which represses Cdkn2a (p19(ARF)) and is amplified in a subset of human breast cancers. Nat Genet 26: 291–299. 10.1038/81583.
32. Brummelkamp TR, Kortlever RM, Lingbeek M, Trettel F, MacDonald ME, et al. (2002) TBX-3, the gene mutated in ulnar-mammary syndrome, is a negative regulator of p19ARF and inhibits senescence. J Biol Chem 277: 6567–6572. 10.1074/jbc.M110492200.
33. Marcotte R, Wang E (2002) Replicative senescence revisited. J Gerontol A Biol Sci Med Sci 57: B257–69.
34. Tsuji T, Aoshiba K, Nagai A (2006) Alveolar cell senescence in patients with pulmonary emphysema. Am J Respir Crit Care Med 174: 886–893. 10.1164/rccm.200509-1374OC.
35. Faith JJ, Hayete B, Thaden JT, Mogno I, Wierzbowski J, et al. (2007) Large-scale mapping and validation of escherichia coli transcriptional regulation from a compendium of expression profiles. PLoS Biol 5: e8.
36. Brown V, Elborn JS, Bradley J, Ennis M (2009) Dysregulated apoptosis and NFkappaB expression in COPD subjects. Respir Res 10: 24. 10.1186/1465-9921-10-24.
37. Cavalcante AG, de Bruin PF (2009) The role of oxidative stress in COPD: Current concepts and perspectives. J Bras Pneumol 35: 1227–1237.
38. Beeh KM, Glaab T (2009) Is there a role for antiinflammatory treatment in COPD? COPD 6: 395–403.
39. Bhattacharya S, Srisuma S, Demeo DL, Shapiro SD, Bueno R, et al. (2009) Molecular biomarkers for quantitative and discrete COPD phenotypes. Am J Respir Cell Mol Biol 40: 359–367. 10.1165/rcmb.2008-0114OC.
40. Margolin AA, Nemenman I, Basso K, Wiggins C, Stolovitzky G, et al. (2006) ARACNE: An algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. BMC Bioinformatics 7 Suppl 1: S7. 10.1186/1471-2105-7-S1-S7.
41. Daub CO, Steuer R, Selbig J, Kloska S (2004) Estimating mutual information using B-spline functions–an improved similarity measure for analysing gene expression data. BMC Bioinformatics 5: 118. 10.1186/1471-2105-5-118.
42. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, et al. (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 4: 249–264. 10.1093/biostatistics/4.2.249.
43. Getz G, Levine E, Domany E (2000) Coupled two-way clustering analysis of gene microarray data. Proc Natl Acad Sci U S A 97: 12079–12084. 10.1073/pnas.210134797.

44. Hochreiter S, Bodenhofer U, Heusel M, Mayr A, Mitterecker A, et al. (2010) FABIA: Factor analysis for bicluster acquisition. Bioinformatics 26: 1520–1527. 10.1093/bioinformatics/btq227.

45. Bonneau R, Reiss DJ, Shannon P, Facciotti M, Hood L, et al. (2006) The inferelator: An algorithm for learning parsimonious regulatory networks from systems-biology data sets de novo. Genome Biol 7: R36. 10.1186/gb-2006-7-5-r36.

46. Greenfield A, Madar A, Ostrer H, Bonneau R (2010) DREAM4: Combining genetic and dynamic information to identify biological networks and dynamical models. PLoS One 5: e13397. 10.1371/journal.pone.0013397.

47. Madar A, Greenfield A, Vanden-Eijnden E, Bonneau R (2010) DREAM3: Network inference using dynamic context likelihood of relatedness and the inferelator. PLoS One 5: e9803. 10.1371/journal.pone.0009803.

48. Yao H, Yang SR, Edirisinghe I, Rajendrasozhan S, Caito S, et al. (2008) Disruption of p21 attenuates lung inflammation induced by cigarette smoke, LPS, and fMLP in mice. Am J Respir Cell Mol Biol 39: 7–18. 10.1165/rcmb.2007-0342OC.

49. Volonte D, Galbiati F (2009) Caveolin-1, cellular senescence and pulmonary emphysema. Aging (Albany NY) 1: 831–835.

50. Volonte D, Kahkonen B, Shapiro S, Di Y, Galbiati F (2009) Caveolin-1 expression is required for the development of pulmonary emphysema through activation of the ATM-p53-p21 pathway. J Biol Chem 284: 5462–5466. 10.1074/jbc.C800225200.

51. Abrahams A, Mowla S, Parker MI, Goding CR, Prince S (2008) UV-mediated regulation of the anti-senescence factor Tbx2. J Biol Chem 283: 2223–2230. 10.1074/jbc.M705651200.

52. Hoogaars WM, Barnett P, Rodriguez M, Clout DE, Moorman AF, et al. (2008) TBX3 and its splice variant TBX3 + exon 2a are functionally similar. Pigment Cell Melanoma Res 21: 379–387. 10.1111/j.1755-148X.2008.00461.x.

53. Vance KW, Carreira S, Brosch G, Goding CR (2005) Tbx2 is overexpressed and plays an important role in maintaining proliferation and suppression of senescence in melanomas. Cancer Res 65: 2260–2268. 10.1158/0008-5472.CAN-04-3045.

54. Kueppers F, Miller RD, Gordon H, Hepper NG, Offord K (1977) Familial prevalence of chronic obstructive pulmonary disease in a matched pair study. Am J Med 63: 336–342.

55. Kim KM, Park SH, Kim JS, Lee WK, Cha SI, et al. (2008) Polymorphisms in the type IV collagen alpha3 gene and the risk of COPD. Eur Respir J 32: 35–41. 10.1183/09031936.00076207.

56. Houben JM, Mercken EM, Ketelslegers HB, Bast A, Wouters EF, et al. (2009) Telomere shortening in chronic obstructive pulmonary disease. Respir Med 103: 230–236. 10.1016/j.rmed.2008.09.003.

57. Kong X, Cho MH, Anderson W, Coxson HO, Muller N, et al. (2011) Genome-wide association study identifies BICD1 as a susceptibility gene for emphysema. Am J Respir Crit Care Med 183: 43–49. 10.1164/rccm.201004-0541OC.

58. Morla M, Busquets X, Pons J, Sauleda J, MacNee W, et al. (2006) Telomere shortening in smokers with and without COPD. Eur Respir J 27: 525–528. 10.1183/09031936.06.00087005.

59. Capozza F, Williams TM, Schubert W, McClain S, Bouzahzah B, et al. (2003) Absence of caveolin-1 sensitizes mouse skin to carcinogen-induced epidermal hyperplasia and tumor formation. Am J Pathol 162: 2029–2039. 10.1016/S0002-9440(10)64335-0.

60. Leal JF, Fominaya J, Cascon A, Guijarro MV, Blanco-Aparicio C, et al. (2008) Cellular senescence bypass screen identifies new putative tumor suppressor genes. Oncogene 27: 1961–1970. 10.1038/sj.onc.1210846.

61. Lleonart ME, Artero-Castro A, Kondoh H (2009) Senescence induction; a possible cancer therapy. Mol Cancer 8: 3. 10.1186/1476-4598-8-3.

62. Saretzki G (2010) Cellular senescence in the development and treatment of cancer. Curr Pharm Des 16: 79–100.

63. Sarkisian CJ, Keister BA, Stairs DB, Boxer RB, Moody SE, et al. (2007) Dose-dependent oncogene-induced senescence in vivo and its evasion during mammary tumorigenesis. Nat Cell Biol 9: 493–505. 10.1038/ncb1567.

64. Williams TM, Lee H, Cheung MW, Cohen AW, Razani B, et al. (2004) Combined loss of INK4a and caveolin-1 synergistically enhances cell proliferation and oncogene-induced tumorigenesis: Role of INK4a/CAV-1 in mammary epithelial cell hyperplasia. J Biol Chem 279: 24745–24756. 10.1074/jbc.M402064200.

65. Roberson RS, Kussick SJ, Vallieres E, Chen SY, Wu DY (2005) Escape from therapy-induced accelerated cellular senescence in p53-null lung cancer cells and in human lung cancers. Cancer Res 65: 2795–2803. 10.1158/0008-5472.CAN-04-1270.

66. Babizhayev MA, Yegorov YE (2010) Smoking and health: Association between telomere length and factors impacting on human disease, quality of life and life span in a large population-based cohort under the effect of smoking duration. Fundam Clin Pharmacol 25: 425–42. 10.1111/j.1472-8206.2010.00866.x.

67. Blanchette CM, Berry SR, Lane SJ (2011) Advances in chronic obstructive pulmonary disease among older adults. Curr Opin Pulm Med 17: 84–89. 10.1097/MCP.0b013e32834316ff.

68. Gould NS, Min E, Gauthier S, Chu HW, Martin R, et al. (2010) Aging adversely affects the cigarette smoke-induced glutathione adaptive response in the lung. Am J Respir Crit Care Med 182: 1114–1122. 10.1164/rccm.201003-0442OC.

69. Lee J, Sandford A, Man P, Sin DD (2011) Is the aging process accelerated in chronic obstructive pulmonary disease? Curr Opin Pulm Med 17: 90–97.

70. Tada M, Smith JC (2001) T-targets: Clues to understanding the functions of T-box proteins. Dev Growth Differ 43: 1–11.

71. Bollag RJ, Siegfried Z, Cebra-Thomas JA, Garvey N, Davison EM, et al. (1994) An ancient family of embryonically expressed mouse genes sharing a conserved protein motif with the T locus. Nat Genet 7: 383–389. 10.1038/ng0794-383.

72. Wilson V, Conlon FL (2002) The T-box family. Genome Biol 3: REVIEWS3008.

73. Carreira S, Dexter TJ, Yavuzer U, Easty DJ, Goding CR (1998) Brachyury-related transcription factor Tbx2 and repression of the melanocyte-specific TRP-1 promoter. Mol Cell Biol 18: 5099–5108.

74. Govoni KE, Linares GR, Chen ST, Pourteymoor S, Mohan S (2009) T-box 3 negatively regulates osteoblast differentiation by inhibiting expression of osterix and runx2. J Cell Biochem 106: 482–490. 10.1002/jcb.22035.

75. Redmond KL, Crawford NT, Farmer H, D'Costa ZC, O'Brien GJ, et al. (2010) T-box 2 represses NDRG1 through an EGR1-dependent mechanism to drive the proliferation of breast cancer cells. Oncogene 29: 3252–3262. 10.1038/onc.2010.84.

76. Naiche LA, Harrelson Z, Kelly RG, Papaioannou VE (2005) T-box genes in vertebrate development. Annu Rev Genet 39: 219–239. 10.1146/annurev.genet.39.073003.105925.

77. Canepa ET, Scassa ME, Ceruti JM, Marazita MC, Carcagno AL, et al. (2007) INK4 proteins, a family of mammalian CDK inhibitors with novel biological functions. IUBMB Life 59: 419–426. 10.1080/15216540701488358.

78. Zhang Y, Xiong Y, Yarbrough WG (1998) ARF promotes MDM2 degradation and stabilizes p53: ARF-INK4a locus deletion impairs both the rb and p53 tumor suppression pathways. Cell 92: 725–734.

79. Levine AJ (1997) P53, the cellular gatekeeper for growth and division. Cell 88: 323–331.

80. Sinclair CS, Adem C, Naderi A, Soderberg CL, Johnson M, et al. (2002) TBX2 is preferentially amplified in BRCA1- and BRCA2-related breast tumors. Cancer Res 62: 3587–3591.

81. Mahlamaki EH, Barlund M, Tanner M, Gorunova L, Hoglund M, et al. (2002) Frequent amplification of 8q24, 11q, 17q, and 20q-specific genes in pancreatic cancer. Genes Chromosomes Cancer 35: 353–358. 10.1002/gcc.10122.

82. Dominguez-Brauer C, Brauer PM, Chen YJ, Pimkina J, Raychaudhuri P (2010) Tumor suppression by ARF: Gatekeeper and caretaker. Cell Cycle 9: 86–89.

83. Bendjennat M, Boulaire J, Jascur T, Brickner H, Barbier V, et al. (2003) UV irradiation triggers ubiquitin-dependent degradation of p21(WAF1) to promote DNA repair. Cell 114: 599–610.

84. Cooper MP, Balajee AS, Bohr VA (1999) The C-terminal domain of p21 inhibits nucleotide excision repair in vitro and in vivo. Mol Biol Cell 10: 2119–2129.

85. Renard CA, Labalette C, Armengol C, Cougot D, Wei Y, et al. (2007) Tbx3 is a downstream target of the Wnt/beta-catenin pathway and a critical mediator of beta-catenin survival functions in liver cancer. Cancer Res 67: 901–910. 10.1158/0008-5472.CAN-06-2344.

86. Ismail A, Bateman A (2009) Expression of TBX2 promotes anchorage-independent growth and survival in the p53-negative SW13 adrenocortical carcinoma. Cancer Lett 278: 230–240. 10.1016/j.canlet.2009.01.006.

87. Vance KW, Shaw HM, Rodriguez M, Ott S, Goding CR (2010) The retinoblastoma protein modulates Tbx2 functional specificity. Mol Biol Cell 21: 2770–2779. 10.1091/mbc.E09-12-1029.

88. Abrahams A, Parker MI, Prince S (2010) The T-box transcription factor Tbx2: Its role in development and possible implication in cancer. IUBMB Life 62: 92–102. 10.1002/iub.275.

89. Lu J, Li XP, Dong Q, Kung HF, He ML (2010) TBX2 and TBX3: The special value for anticancer drug targets. Biochim Biophys Acta 1806: 268–274. 10.1016/j.bbcan.2010.07.001.

90. Fukuchi Y, Nishimura M, Ichinose M, Adachi M, Nagai A, et al. (2004) COPD in japan: The nippon COPD epidemiology study. Respirology 9: 458–465. 10.1111/j.1440-1843.2004.00637.x.

91. Tsuji T, Aoshiba K, Nagai A (2004) Cigarette smoke induces senescence in alveolar epithelial cells. Am J Respir Cell Mol Biol 31: 643–649. 10.1165/rcmb.2003-0290OC.

92. Hornsby PJ (2002) Cellular senescence and tissue aging in vivo. J Gerontol A Biol Sci Med Sci 57: B251–6.

93. Sharma G, Hanania NA, Shim YM (2009) The aging immune system and its relationship to the development of chronic obstructive pulmonary disease. Proc Am Thorac Soc 6: 573–580. 10.1513/pats.200904-022RM.

94. Silverman EK, Palmer LJ, Mosley JD, Barth M, Senter JM, et al. (2002) Genomewide linkage analysis of quantitative spirometric phenotypes in severe early-onset chronic obstructive pulmonary disease. Am J Hum Genet 70: 1229–1239. 10.1086/340316.

95. DeMeo DL, Celedon JC, Lange C, Reilly JJ, Chapman HA, et al. (2004) Genome-wide linkage of forced mid-expiratory flow in chronic obstructive pulmonary disease. Am J Respir Crit Care Med 170: 1294–1301. 10.1164/rccm.200404-524OC.

96. Calderwood SK, Murshid A, Prince T (2009) The shock of aging: Molecular chaperones and the heat shock response in longevity and aging–a mini-review. Gerontology 55: 550–558. 10.1159/000225957.

97. Hudson BG, Kalluri R, Gunwar S, Noelken ME, Mariyama M, et al. (1993) Molecular characteristics of the goodpasture autoantigen. Kidney Int 43: 135–139.

98. Baek SH, Kioussi C, Briata P, Wang D, Nguyen HD, et al. (2003) Regulated subset of G1 growth-control genes in response to derepression by the wnt pathway. Proc Natl Acad Sci U S A 100: 3245–3250. 10.1073/pnas.0330217100.

99. Salomoni P, Pandolfi PP (2002) The role of PML in tumor suppression. Cell 108: 165–170.

100. Le XF, Vallian S, Mu ZM, Hung MC, Chang KS (1998) Recombinant PML adenovirus suppresses growth and tumorigenicity of human breast cancer cells by inducing G1 cell cycle arrest and apoptosis. Oncogene 16: 1839–1849. 10.1038/sj.onc.1201705.

101. Bluher M, Kahn BB, Kahn CR (2003) Extended longevity in mice lacking the insulin receptor in adipose tissue. Science 299: 572–574. 10.1126/science.1078223.

102. Holzenberger M, Dupont J, Ducos B, Leneuve P, Geloen A, et al. (2003) IGF-1 receptor regulates lifespan and resistance to oxidative stress in mice. Nature 421: 182–187. 10.1038/nature01298.

103. Taguchi A, Wartschow LM, White MF (2007) Brain IRS2 signaling coordinates life span and nutrient homeostasis. Science 317: 369–372. 10.1126/science.1142179.

104. Kaestner KH, Knochel W, Martinez DE (2000) Unified nomenclature for the winged helix/forkhead transcription factors. Genes Dev 14: 142–146.

105. Kops GJ, Dansen TB, Polderman PE, Saarloos I, Wirtz KW, et al. (2002) Forkhead transcription factor FOXO3a protects quiescent cells from oxidative stress. Nature 419: 316–321. 10.1038/nature01036.

106. Brunet A, Bonni A, Zigmond MJ, Lin MZ, Juo P, et al. (1999) Akt promotes cell survival by phosphorylating and inhibiting a forkhead transcription factor. Cell 96: 857–868.

107. MacDonald BT, Tamai K, He X (2009) Wnt/beta-catenin signaling: Components, mechanisms, and diseases. Dev Cell 17: 9–26. 10.1016/j.devcel.2009.06.016.

108. Behrens J, von Kries JP, Kuhl M, Bruhn L, Wedlich D, et al. (1996) Functional interaction of beta-catenin with the transcription factor LEF-1. Nature 382: 638–642. 10.1038/382638a0.

109. Rubinfeld B, Robbins P, El-Gamil M, Albert I, Porfiri E, et al. (1997) Stabilization of beta-catenin by genetic defects in melanoma cell lines. Science 275: 1790–1792.

110. Jin T, George Fantus I, Sun J (2008) Wnt and beyond wnt: Multiple mechanisms control the transcriptional property of beta-catenin. Cell Signal 20: 1697–1704. 10.1016/j.cellsig.2008.04.014.

111. Essers MA, de Vries-Smits LM, Barker N, Polderman PE, Burgering BM, et al. (2005) Functional interaction between beta-catenin and FOXO in oxidative stress signaling. Science 308: 1181–1184. 10.1126/science.1109083.

112. Harms KL, Chen X (2007) Histone deacetylase 2 modulates p53 transcriptional activities through regulation of p53-DNA binding activity. Cancer Res 67: 3145–3152. 10.1158/0008-5472.CAN-06-4397.

113. Grubisha O, Smith BC, Denu JM (2005) Small molecule regulation of Sir2 protein deacetylases. FEBS J 272: 4607–4616. 10.1111/j.1742-4658.2005.04862.x.

114. Xie S, Macedo P, Hew M, Nassenstein C, Lee KY, et al. (2009) Expression of transforming growth factor-beta (TGF-beta) in chronic idiopathic cough. Respir Res 10: 40. 10.1186/1465-9921-10-40.

115. Xu J, Gonzalez ET, Iyer SS, Mac V, Mora AL, et al. (2009) Use of senescence-accelerated mouse model in bleomycin-induced lung injury suggests that bone marrow-derived cells can alter the outcome of lung injury in aged mice. J Gerontol A Biol Sci Med Sci 64: 731–739. 10.1093/gerona/glp040.

116. Taraseviciene-Stewart L, Voelkel NF (2008) Molecular pathogenesis of emphysema. J Clin Invest 118: 394–402. 10.1172/JCI31811.

117. Tuder RM, Yun JH, Graham BB (2008) Cigarette smoke triggers code red: P21CIP1/WAF1/SDI1 switches on danger responses in the lung. Am J Respir Cell Mol Biol 39: 1–6. 10.1165/rcmb.2008-0117TR.

118. Londhe VA, Sundar IK, Lopez B, Maisonet TM, Yu Y, et al. (2011) Hyperoxia impairs alveolar formation and induces senescence through decreased histone deacetylase activity and upregulation of p21 in neonatal mouse lung. Pediatr Res 69: 371–7. 10.1203/PDR.0b013e318211c917.

119. MacNee W (2009) Accelerated lung aging: A novel pathogenic mechanism of chronic obstructive pulmonary disease (COPD). Biochem Soc Trans 37: 819–823. 10.1042/BST0370819.

120. Karrasch S, Holz O, Jorres RA (2008) Aging and induced senescence as factors in the pathogenesis of lung emphysema. Respir Med 102: 1215–1230. 10.1016/j.rmed.2008.04.013.

121. Nyunoya T, Monick MM, Klingelhutz AL, Glaser H, Cagley JR, et al. (2009) Cigarette smoke induces cellular senescence via werner's syndrome protein down-regulation. Am J Respir Crit Care Med 179: 279–287. 10.1164/rccm.200802-320OC.

122. Imai K, Mercer BA, Schulman LL, Sonett JR, D'Armiento JM (2005) Correlation of lung surface area to apoptosis and proliferation in human emphysema. Eur Respir J 25: 250–258. 10.1183/09031936.05.00023704.

123. Yokohori N, Aoshiba K, Nagai A, Respiratory Failure Research Group in Japan. (2004) Increased levels of cell death and proliferation in alveolar wall cells in patients with pulmonary emphysema. Chest 125: 626–632.

124. Oberley-Deegan RE, Regan EA, Kinnula VL, Crapo JD (2009) Extracellular superoxide dismutase and risk of COPD. COPD 6: 307–312.

125. Donaldson GC, Seemungal TA, Patel IS, Bhowmik A, Wilkinson TM, et al. (2009) Airway and systemic inflammation and decline in lung function in patients with COPD. 2005. Chest 136: e30.

126. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, et al. (2000) Gene ontology: Tool for the unification of biology. the gene ontology consortium. Nat Genet 25: 25–29.

127. Graham P (1996) ANSI common lisp, ser. Artif Intel. New Jersey, NJ: Prentice Hall.

128. Butte AJ, Kohane IS (2000) Mutual information relevance networks: Functional genomic clustering using pairwise entropy measurements. Pac Symp Biocomput : 418–429.

129. Liang S, Fuhrman S, Somogyi R (1998) Reveal, a general reverse engineering algorithm for inference of genetic network architectures. Pac Symp Biocomput 1998: 18–29.

130. Shannon CE, Weaver W (1963) The mathematical theory of communication, 111 pp. Urbana: University of Illinois Press.

131. Priness I, Maimon O, Ben-Gal I (2007) Evaluation of gene-expression clustering via mutual information distance measure. BMC Bioinformatics 8: 111. 10.1186/1471-2105-8-111.

132. Slonim N, Atwal GS, Tkacik G, Bialek W (2005) Information-based clustering. Proc Natl Acad Sci U S A 102: 18297–18302. 10.1073/pnas.0507432102.

133. Zhou M, Cui Y (2004) GeneInfoViz: Constructing and visualizing gene relation networks. In Silico Biology 4: 323–333.

134. Cover TM, Thomas JA (2006) Elements of information theory. Wiley-Interscience.

135. Golpon HA, Coldren CD, Zamora MR, Cosgrove GP, Moore MD, et al. (2004) Emphysema lung tissue gene expression profiling. Am J Respir Cell Mol Biol 31: 595–600. 10.1165/rcmb.2004-0008OC.

136. Spira A, Beane J, Pinto-Plata V, Kadar A, Liu G, et al. (2004) Gene expression profiling of human lung tissue from smokers with severe emphysema. Am J Respir Cell Mol Biol 31: 601–610. 10.1165/rcmb.2004-0273OC.

137. Downey T (2006) Analysis of a multifactor microarray study using partek genomics solution. Methods Enzymol 411: 256–270. 10.1016/S0076-6879(06)11013-7.

138. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. J R Soc Series B Stat Methodol 57: 289–300.

139. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, et al. (2003) Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Res 13: 2498–2504. 10.1101/gr.1239303.

140. Bateman A, Coin L, Durbin R, Finn RD, Hollich V, et al. (2004) The pfam protein families database. Nucleic Acids Res 32: D138–41. 10.1093/nar/gkh121.

141. Deng M, Mehta S, Sun F, Chen T (2002) Inferring domain-domain interactions from protein-protein interactions. Genome Res 12: 1540–1548. 10.1101/gr.153002.

142. Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. J R Stat Soc Series B Methodol 39: 1–38.

143. Grigoriev A (2001) A relationship between gene expression and protein interactions on the proteome scale: Analysis of the bacteriophage T7 and the yeast saccharomyces cerevisiae. Nucleic Acids Res 29: 3513–3519.

144. Ge H, Liu Z, Church GM, Vidal M (2001) Correlation between transcriptome and interactome mapping data from saccharomyces cerevisiae. Nat Genet 29: 482–486. 10.1038/ng776.

145. Morcos F, Lamanna C, Sikora M, Izaguirre J (2008) Cytoprophet: A cytoscape plug-in for protein and domain interaction networks inference. Bioinformatics 24: 2265–2266. 10.1093/bioinformatics/btn380.

146. Pauwels RA, Buist AS, Ma P, Jenkins CR, Hurd SS, et al. (2001) Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: National heart, lung, and blood institute and world health organization global initiative for chronic obstructive lung disease (GOLD): Executive summary. Respir Care 46: 798–825.

147. Gimino VJ, Lande JD, Berryman TR, King RA, Hertz MI (2003) Gene expression profiling of bronchoalveolar lavage cells in acute lung rejection. Am J Respir Crit Care Med 168: 1237–1242. 10.1164/rccm.200305-644OC.

148. Spira A, Beane J, Shah V, Liu G, Schembri F, et al. (2004) Effects of cigarette smoke on the human airway epithelial cell transcriptome. Proc Natl Acad Sci U S A 101: 10143–10148. 10.1073/pnas.0401422101.