

## ORIGINAL ARTICLE

## A large lung gene expression study identifying fibulin-5 as a novel player in tissue repair in COPD

Corry-Anke Brandsma,<sup>1,2</sup> Maarten van den Berge,<sup>2,3</sup> Dirkje S Postma,<sup>2,3</sup> Marnix R Jonker,<sup>1,2</sup> Sharon Brouwer,<sup>1,2</sup> Peter D Paré,<sup>4,5</sup> Don D Sin,<sup>4,5</sup> Yohan Bossé,<sup>6,7</sup> Michel Laviolette,<sup>6</sup> Juha Karjalainen,<sup>8</sup> Rudolf S N Fehrmann,<sup>8</sup> David C Nickle,<sup>9</sup> Ke Hao,<sup>9</sup> Anita I R Spanjer,<sup>2,10</sup> Wim Timens,<sup>1,2</sup> Lude Franke<sup>8</sup>

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/thoraxjnl-2014-205091>).

For numbered affiliations see end of article.

**Correspondence to**

Dr Corry-Anke Brandsma, Department of Pathology and Medical Biology, University Medical Center Groningen, PO Box 30.001, Groningen 9700 RB, The Netherlands; c.a.brandsma@umcg.nl

Received 6 January 2014  
Revised 15 May 2014  
Accepted 12 June 2014  
Published Online First 2 July 2014

**ABSTRACT**

**Background** Chronic obstructive pulmonary disease (COPD) is a progressive, incurable lung disease characterised by abnormal tissue repair causing emphysema and small airways fibrosis. Since current therapy cannot modify this abnormal repair, it is crucial to unravel its underlying molecular mechanisms. Unbiased analysis of genome-wide gene expression profiles in lung tissue provides a powerful tool to investigate this.

**Methods** We performed genome-wide gene expression profiling in 581 lung tissue samples from current and ex-smokers with (n=311) and without COPD (n=270). Subsequently, quantitative PCR, western blot and immunohistochemical analyses were performed to validate our main findings.

**Results** 112 genes were found to be upregulated in patients with COPD compared with controls, whereas 61 genes were downregulated. Among the most upregulated genes were fibulin-5 (*FBLN5*), elastin (*ELN*), latent transforming growth factor  $\beta$  binding protein 2 (*LTBP2*) and microfibrillar associated protein 4 (*MFAP4*), all implicated in elastogenesis. Our gene expression findings were validated at mRNA and protein level. We demonstrated higher *ELN* gene expression in COPD lung tissue and similar trends for *FBLN5* and *MFAP4*, and negative correlations with lung function. *FBLN5* protein levels were increased in COPD lung tissue and cleaved, possibly non-functional *FBLN5* protein was present. Strong coexpression of *FBLN5*, *ELN*, *LTBP2* and *MFAP4* in lung tissue and in silico analysis indicated cofunctionality of these genes. Finally, colocalisation of *FBLN5*, *MFAP4* and *LTBP2* with elastic fibres was demonstrated in lung tissue.

**Conclusions** We identified a clear gene signature for elastogenesis in COPD and propose *FBLN5* as a novel player in tissue repair in COPD.

**INTRODUCTION**

Chronic obstructive pulmonary disease (COPD) is a chronic and progressive lung disease with a large impact on patients and society and is mainly caused by cigarette smoking. There is no cure available for COPD and current drugs are mainly effective in improving symptoms and exacerbations but generally do not slow down the progression of the disease. Quitting smoking can reduce the accelerated lung function decline in some but not all

**Key messages****What is the key question?**

► Which relevant lung tissue gene expression changes can explain the mechanisms behind abnormal tissue repair in chronic obstructive pulmonary disease (COPD)?

**What is the bottom line?**

► Using genome-wide gene expression profiling in a large number of lung tissue samples from patients with COPD and non-COPD controls, we identified a clear COPD gene signature with increased expression of genes related to elastogenesis, with fibulin-5 as a main player.

**Why read on?**

► Our findings provide much needed insights in lung tissue gene expression changes in COPD and point out fibulin-5 as a potential novel player in tissue repair in COPD.

patients with COPD and it does not restore lost lung function.

Pathologically, COPD is characterised by distinct features, including emphysema with the destruction of alveolar lung tissue and fibrosis of the (small) airways with airway wall thickening.<sup>1</sup> Although emphysema and airway wall thickening occur in close proximity to each other, they seem to be driven by different processes since emphysematous destruction is characterised by insufficient extracellular matrix (ECM) production, indicating a defective repair process, whereas airway wall thickening is the result of an exaggerated repair process accompanied by a marked increase in ECM. Both processes are important contributors to loss of lung function in patients with COPD. Hogg and colleagues recently demonstrated that small airway wall thickening and loss of small airways precedes emphysematous lung tissue destruction in patients with COPD, suggesting a relation between these processes.<sup>2</sup>

The exact nature of the abnormal repair processes in COPD lungs is unknown, hence there is no effective treatment to prevent further lung tissue damage or induce repair. A better understanding of regulation of the repair processes and the underlying molecular mechanisms is thus crucial.



► <http://dx.doi.org/10.1136/thoraxjnl-2014-205887>



**To cite:** Brandsma C-A, van den Berge M, Postma DS, et al. *Thorax* 2015;**70**: 21–32.

Unbiased analysis of genome-wide gene expression in lung tissue provides a powerful tool to investigate these molecular mechanisms. However, studies using this approach have been hindered thus far by small sample sizes due to the limited availability of lung tissue samples.<sup>3–7</sup> As part of a Lung eQTL Consortium,<sup>8–10</sup> we have performed an unbiased analysis of genome-wide gene-expression profiles in lung tissue specimens derived from a large number (N=581) of well characterised current and ex-smokers with (N=311) and without (N=270) COPD.

## METHODS

Detailed methods can be found in the online supplement.

### Patients

Gene expression data were used from 311 patients with COPD and 270 non-COPD controls who were part of the Lung eQTL dataset from three academic sites.<sup>8,9</sup> Lung tissue samples were collected from patients undergoing lung surgery for various reasons, mostly therapeutic resection of lung tumours, but also lung transplantation. In case of tumour resections, macroscopically normal lung tissue was taken far distant from the tumour and histology of all samples was checked for abnormalities using standard haematoxylin and eosin staining. Lung samples were obtained in accordance with local ethical guidelines (for a detailed description, see Hao *et al*<sup>8</sup>). In the current analyses, we included current and ex-smokers >40 years with  $\geq 5$  pack-years. COPD was defined as an FEV<sub>1</sub>/FVC ratio <70%. Non-COPD control was defined as an FEV<sub>1</sub>/FVC  $\geq 70\%$  predicted. In case lung tissue samples were derived from healthy donors, no data on FEV<sub>1</sub> or FEV<sub>1</sub>/FVC ratio were available. For FEV<sub>1</sub> and FEV<sub>1</sub>/FVC, pre-bronchodilator values were used when post-bronchodilator values were not available. Subjects with other lung diseases such as asthma, cystic fibrosis or interstitial lung diseases were excluded.

### Gene expression assays

mRNA profiling was performed by Rosetta Inpharmatics Gene Expression Laboratory (Seattle, Washington, USA) using custom Affymetrix HU133 arrays (GEO platform GPL10379), consisting of 751 control and 51 627 non-control probe sets as described previously.<sup>8</sup> Gene expression normalisation was performed using Robust Multi-array Average and every probe set was normalised; that is, the mean expression was set to zero and the SD to one. We corrected for strong expression differences between the lung tissue samples due to unknown factors (eg, batch or technical effects) by correcting for the first 25 principal components, as described previously.<sup>11</sup>

We analysed the three cohorts separately, and corrected for potential confounders: age, gender, pack-years and smoking status (current vs former). We performed t tests to find differentially expressed genes between patients with COPD and controls. We then used an inverse variance meta-analysis to identify genes that were differentially expressed and behaved in the same direction across all three cohorts. We used the false discovery rate (FDR) to control for multiple testing. We report on a set of differentially expressed genes of which we accepted that 20% might be false positives. We observed that a nominal  $p < 0.001$  resulted in 252 differentially expressed genes in the real analysis, whereas we would expect 51.6 by chance, corresponding to an FDR of 0.20.

### GeneNetwork analysis

A new method was used to gain insight into the potential gene function of differentially expressed genes, that is, GeneNetwork. This method uses an independent gene expression dataset of

77 840 samples to predict the function of genes in an unbiased way, as recently employed.<sup>12</sup> We used this method to predict (currently unknown) gene functions based on known biological pathways available in the molecular signatures database MSigDB (<http://www.broadinstitute.org>) and additionally used this information for pathway enrichment analyses and cofunctionality networks. Knockout information from the Mouse Genome Informatics database was collected for genes for which human orthologues exist to predict phenotypes in mice for which currently no knockout has been described.

### Quantitative PCR validation

Quantitative PCR (qPCR) was used to confirm the expression levels of *ELN* (elastin), *FBLN5* (fibulin-5), *MFAP4* (microfibrillar associated protein) and *LTBP2* (latent transforming growth factor (TGF)- $\beta$  binding protein 2) in lung tissue samples from the Groningen cohort.

Additionally, *FBLN5* expression was analysed in primary pulmonary fibroblasts treated with or without TGF $\beta$  and cigarette smoke extract (CSE) from patients with stage IV COPD and non-COPD controls.

### Histology

Immunohistochemical staining for FBLN5, MFAP4 and LTBP2 was performed on paraffin embedded lung tissue samples from the Groningen cohort to localise protein expression in lung tissue and assess the presence of these proteins in elastic fibres.

### Western blot analysis

Western blot analyses was performed to measure total protein levels of FBLN5, ELN and MFAP4 in lung tissue and to identify the presence of the cleaved form of FBLN5.

### Statistics

Gene expression analyses were performed using the eQTLMappingPipeline as previously described.<sup>13</sup> Additional statistical analyses were performed using IBM SPSS Statistics 22 and R statistical software V.2.12.0. Differences in clinical characteristics were analysed using Mann-Whitney U (MWU) tests. For qPCR validation, MWU tests were used and increased gene expression and a one-sided p value <0.05 was considered a positive validation. Differences in *FBLN5* expression levels in COPD and control fibroblasts and differences in FBLN5 protein levels in lung tissue were assessed using MWU tests and a two-sided p value <0.05 was considered significant. The relation between FEV<sub>1</sub> and *FBLN5*, *ELN*, *LTBP2* and *MFAP4* expression and the relation between mRNA expression of these four genes in lung tissue was assessed by Spearman's correlation. A  $p < 0.05$  was considered significant.

## RESULTS

### Subject characteristics

Table 1 presents the clinical characteristics of the subjects. We included 146 subjects from the Groningen cohort, 282 from the Laval cohort and 153 from the UBC cohort. The subjects from the Groningen cohort were somewhat younger than those from Laval and UBC, had less pack-years smoked and had more severe COPD. Overall, patients with COPD and non-COPD controls did not differ in age, gender and smoking status, but patients with COPD had more pack-years smoked than the non-COPD controls. Since these phenotypic differences exist between the cohorts we performed subsequent analyses per cohort separately, followed by a meta-analysis.

**Table 1** Subject characteristics

	Groningen		Laval		UBC		All centres	
	COPD	No COPD	COPD	No COPD	COPD	No COPD	COPD	No COPD
Number	101	45	147	135	63	90	311	270
Age, years	58.0 (52.0–65.0)	60.0 (54.0–66.0)	66.0 (59.0–71.0)*	62.0 (53.5–68.0)	65.0 (58.0–71.0)*	62.0 (56.0–70.75)	64.0 (56.0–71.0)	61.5 (54.0–68.8)†
Male/female, N	56/45	20/25	83/64	76/59	43/20	49/41	182/129	145/125
Smoking								
% Current/% ex	19/81	44/56	26/74	16/84‡	51/49*	34/66	29/71	27/73
Pack-years, N	33.0 (22.1–44.5)	35.0 (21.0–41.0)	50.0 (40.0–60.0)*	38.4 (25.0–46.0)	47.5 (34.9–74.0)*	37.8 (25.0–48.9)	42.0 (30.0–58.0)	37.5 (25.0–48.0)†
GOLD stage, N								
I	0		0		1		1	
II	30		134		52		216	
III	11		12		4		27	
IV	49		1		6		56	
Not classified	11		0		0		11	
FEV <sub>1</sub> , %predicted	26.5 (18.4–57.8)	96.0 (89.2–104.8)	67.4 (59.1–73.7)*	94.3 (87.7–102.9)	67.4 (58.2–74.3)*	99.5 (89.5–108.7)	64.3 (46.6–72.4)	95.5 (88.8–106.2)†
FEV <sub>1</sub> /FVC, %	42.6 (27.6–57.4)	75.8 (72.9–80.0)	59.2 (52.8–64.7)*	75.3 (73.2–78.2)	58.1 (51.3–63.4)*	76.6 (73.3–79.5)	56.3 (45.9–63.5)	75.6 (73.2–79.1)†
TLCO, %predicted	–	–	79.0 (63.5–91.0)	92.0 (76.3–108.0)	67.8 (51.0–76.6)	85.2 (75.1–94.4)	73.0 (59.7–87.8)	89.0 (76.0–101.7)†

Data are presented as medians with IQRs unless stated otherwise. Mann–Whitney U test was used to test differences between the groups.

\*p<0.05 for difference with patients with COPD from Groningen.

†p<0.05 for difference between COPD (n=311) and non-COPD controls (n=270).

‡p<0.05 for difference with non-COPD controls from Groningen.

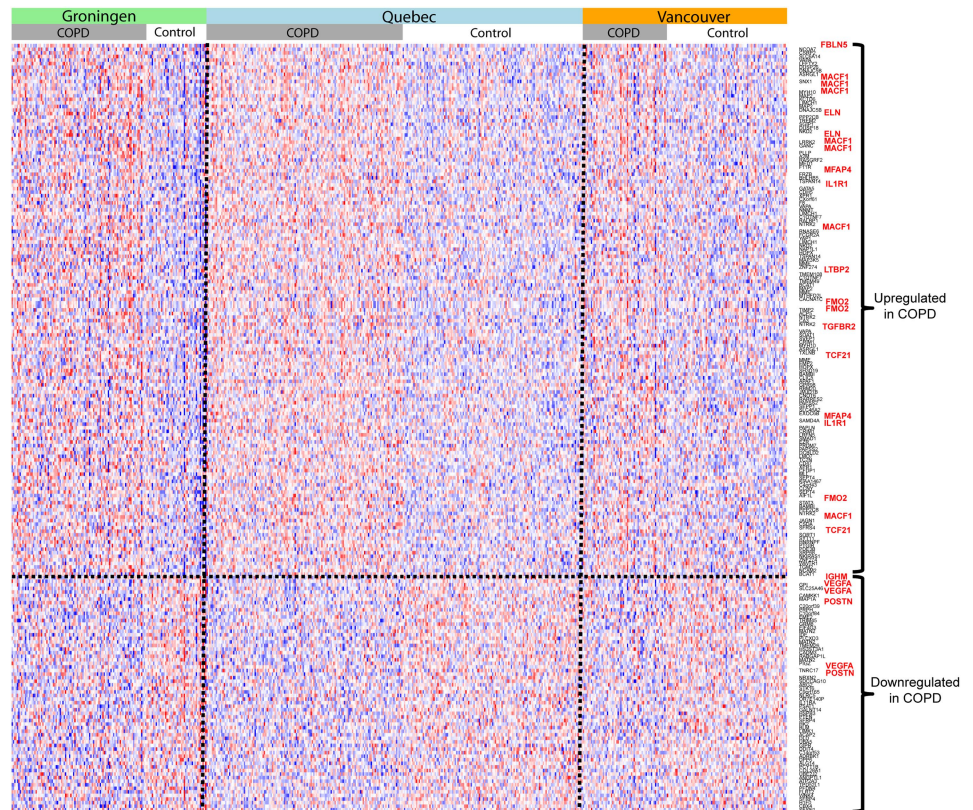
COPD, chronic obstructive pulmonary disease; GOLD, Global Initiative on Obstructive Lung Disease; TLCO, transfer factor of the lung for carbon monoxide.

### Differentially expressed genes in COPD

The expression of 171 probe sets corresponding to 112 known genes was upregulated in patients with COPD versus non-COPD controls, and the expression of 81 probe sets corresponding to 61 known genes was downregulated at a p value <0.001, corresponding to an FDR of 20%. A heatmap of these findings is shown in

figure 1. The complete list of all probe sets and statistics is shown in online supplementary table S1. Among the most upregulated genes were *FBLN5*, microtubule-actin cross linking factor 1 (*MACF1*) and *ELN*. Among the most downregulated genes were immunoglobulin heavy chain constant region  $\mu$  (*IGHM*), vascular endothelial growth factor  $\alpha$  (*VEGFA*) and periostin (*POSTN*).

**Figure 1** Heatmap of differentially expressed genes in chronic obstructive pulmonary disease (COPD). The heatmap shows differential gene expression (p<0.001) between COPD and non-COPD control lung tissue for the three patients cohorts from Groningen, Quebec and Vancouver. Each probe set is shown separately and for several genes we found more than one significant probe set. The coloured, grey and white bars above the heatmap indicate the different groups of lung tissue samples. Genes with a higher expression in COPD lung tissue are shown in the upper part of the heatmap, genes with a lower expression in the lower part. Blue represents lower and red higher relative gene expression. The genes most relevant for this manuscript are highlighted in red on the left side of the heatmap.





**Table 2** Pathway enrichment based on gene function prediction of genes upregulated in COPD

Biological processes		Mouse genome informatics	
Term	p Value	Term	p Value
Response to oxygen levels	8.60E-13	<b>Abnormal pulmonary alveolus morphology</b>	5.78E-12
Regulation of cell migration	3.72E-12	<b>Abnormal lung vasculature morphology</b>	2.03E-11
Regulation of locomotion	4.39E-12	Weight loss	4.73E-11
Regulation of anatomical structure morphogenesis	6.08E-12	<b>Thick pulmonary interalveolar septum</b>	7.60E-11
Regulation of cell motility	1.58E-11	Pericardial effusion	8.16E-11
Regulation of cellular component movement	1.66E-11	Abnormal vascular wound healing	1.05E-10
Regulation of endocytosis	2.12E-11	<b>Abnormal lung development</b>	1.14E-10
Regulation of growth	3.53E-11	Abnormal myocardium layer morphology	1.88E-10
<b>Respiratory tube development</b>	4.13E-11	Lethargy	2.68E-10
<b>Regulation of cell adhesion</b>	5.54E-11	Abnormal vascular smooth muscle physiology	3.69E-10
Circulatory system process	6.92E-11	Atelectasis	1.26E-09
Vasculature development	1.00E-10	Abnormal cardiovascular system physiology	1.61E-09
<b>Lung development</b>	1.08E-10	Abnormal heart ventricle morphology	1.80E-09
Blood vessel morphogenesis	1.47E-10	Abnormal angiogenesis	3.11E-09
Response to hypoxia	1.78E-10	Abnormal fear/anxiety-related behaviour	3.59E-09
Blood circulation	1.89E-10	Abnormal response/metabolism to endogenous compounds	3.76E-09
Regulation of system process	2.37E-10	Increased glomerular capsule space	9.22E-09
Blood vessel development	2.72E-10	Absent heartbeat	1.12E-08
Response to mechanical stimulus	3.27E-10	Decreased response of heart to induced stress	1.29E-08
Regulation of vesicle-mediated transport	3.70E-10	Abnormal production of surfactant	1.31E-08
Angiogenesis	5.56E-10	Dilated heart right ventricle	1.50E-08
Response to hyperoxia	6.36E-10	Hydronephrosis	1.64E-08
Regulation of angiogenesis	9.52E-10	Abnormal cardiac muscle contractility	1.81E-08
Vascular process in circulatory system	9.53E-10	Abnormal cell migration	2.00E-08
<b>Fibroblast migration</b>	1.10E-09	Abnormal macrophage morphology	2.00E-08
Negative regulation of cell migration	1.29E-09	Decreased cardiac muscle contractility	2.18E-08
Positive regulation of angiogenesis	1.69E-09	Abnormal stria vascularis morphology	2.41E-08
Regulation of developmental growth	1.89E-09	Ostium primum atrial septal defect	2.87E-08
Tube development	1.94E-09	Abnormal kidney morphology	3.55E-08
Negative regulation of cell motility	2.22E-09	Decreased heart rate	3.90E-08
<b>Respiratory system development</b>	2.41E-09	Enlarged heart atrium	4.03E-08
Negative regulation of cellular component movement	2.87E-09	Abnormal vascular smooth muscle morphology	4.58E-08
Cell-cell adhesion	2.88E-09	Decreased skin tensile strength	4.69E-08
Negative regulation of epithelial cell proliferation	3.25E-09	Abnormal vitelline vascular remodelling	5.68E-08
Regulation of chemotaxis	3.60E-09	Abnormal eating/drinking behaviour	6.47E-08
Developmental growth	5.69E-09	Abnormal myocardial fibre physiology	6.72E-08
Regulation of blood vessel size	6.31E-09	Abnormal heart development	6.82E-08
Negative regulation of locomotion	6.80E-09	Dilated heart left ventricle	7.58E-08
Actin filament-based process	7.89E-09	Abnormal CNS glial cell morphology	9.51E-08
Positive regulation of endocytosis	8.18E-09	Increased cranium width	1.14E-07
Glomerulus development	8.30E-09	Increased response of heart to induced stress	1.24E-07
Regulation of tube size	8.31E-09	Increased circulating creatinine level	1.31E-07
Regulation of behaviour	1.03E-08	Increased systemic arterial blood pressure	1.37E-07
Positive regulation of cell adhesion	1.27E-08	<b>Abnormal lung morphology</b>	1.57E-07
Regulation of endothelial cell migration	1.43E-08	Abnormal heart shape	1.64E-07
Response to steroid hormone stimulus	1.79E-08	Cyanosis	2.00E-07
Regulation of muscle system process	1.91E-08	Increased angiogenesis	2.74E-07
Positive regulation of cell migration	2.35E-08	Haemorrhage	2.81E-07
Striated muscle tissue development	2.59E-08	Decreased brain weight	2.92E-07
Actin cytoskeleton organisation	2.60E-08	Cardiac hypertrophy	3.10E-07
		Hypoactivity	3.60E-07
		Abnormal osteoclast morphology	3.69E-07
		Abnormal epididymal fat pad morphology	3.80E-07
		Increased type II pneumocyte number	4.25E-07
		Abnormal heart valve morphology	4.39E-07
		Pulmonary hypoplasia	4.85E-07
		Edema	5.19E-07
		Altered response to myocardial infarction	5.24E-07

Continued

Table 2 Continued

Biological processes		Mouse genome informatics	
Term	p Value	Term	p Value
		Stria vascularis degeneration	5.57E-07
		Decreased bone volume	5.59E-07
		Disorganised myocardium	6.00E-07
		Abnormal visceral yolk sac morphology	6.16E-07
		Abnormal ureter morphology	7.39E-07
		Increased urine protein level	7.77E-07
		Enlarged heart	8.47E-07
		Decreased angiogenesis	8.67E-07
		Decreased survivor rate	9.18E-07
		Partial prenatal lethality	9.89E-07
		Abnormal renal glomerulus morphology	1.06E-06
		Decreased brainstem auditory evoked potential	1.06E-06
		Abnormal astrocyte morphology	1.09E-06
		Complete lethality throughout fetal growth and development	1.28E-06
		Thin myocardium	1.39E-06
		Failure of initiation of embryo turning	1.40E-06
		Increased insulin sensitivity	1.55E-06
		<b>Abnormal type II pneumocyte morphology</b>	1.76E-06
		Congestive heart failure	1.81E-06
		Expanded mesangial matrix	1.89E-06
		<b>Emphysema</b>	2.01E-06

Biological pathways and mouse knockout phenotypes that are enriched among the genes upregulated in COPD lung tissue at  $p < 0.001$  are shown. Pathways and knockout phenotypes that are relevant for COPD pathogenesis are shown in bold. CNS, central nervous system; COPD, chronic obstructive pulmonary disease.

From the upregulated genes, *FBLN5* is most interesting in the context of COPD pathogenesis. This gene has been shown to be essential for elastic fibre formation, it is abundantly expressed during lung development,<sup>14</sup> and *FBLN5* knockout mice were shown to develop severe emphysema.<sup>15 16</sup> Moreover, three additional genes implicated in elastogenesis were found to be among the most upregulated genes in COPD, that is, *ELN*, *LTBP2* and *MFAP4*.<sup>17 18</sup>

Other upregulated genes considered to be of potential interest for COPD pathogenesis are TGF $\beta$  receptor-2 (*TGFBR2*),  $\alpha$ -2-macroglobulin (*A2M*), interleukin 1 receptor type 1 (*IL1R1*), transcription factor 21 (*TCF21*) and metalloproteinase inhibitor 2 (*TIMP2*).

Given the potential novel roles for *FBLN5*, *LTBP2* and *MFAP4* in elastogenesis in COPD, we decided to further focus on these genes in this article.

### Pathways and phenotypes related to (abnormal) lung development and emphysema are enriched among genes upregulated in COPD

We subsequently assessed the enrichment of biological processes and predicted phenotypes among the 112 genes upregulated in patients with COPD at  $p < 0.001$ . While such an analysis is typically conducted using gene set enrichment analysis (GSEA), we performed a similar analysis but instead of using only established gene functions in GSEA, we used the predicted gene functions that were based on the independent gene expression dataset of 77 840 samples.

Enrichment was found for biological processes involved in lung development, cell adhesion and fibroblast migration, and several phenotypes related to abnormal lung development and lung morphology and emphysema (table 2).

### Overlapping gene function and cofunctionality for elastogenesis genes

To gain more insight in the function of the four genes related to elastogenesis, we used our GeneNetwork method to predict potential functions of genes (see online supplementary methods). We found a striking overlap in predicted gene function for the four elastogenesis genes, most being predicted to be involved in ECM organisation, collagen fibril organisation, cell-matrix adhesion, and the TGF $\beta$  receptor signalling pathway (table 3A). Interestingly, we also found that all four genes related to elastogenesis were predicted to give an emphysematous phenotype in knockout mice (table 3B). Other predicted phenotypes in knockout mice were overexpanded pulmonary alveoli, impaired lung alveolar development and dilated airways.

When assessing which genes share many predicted functions, we found a very consistent gene network for the genes upregulated in COPD at  $p < 0.001$ . The four genes related to elastogenesis (*FBLN5*, *ELN*, *MFAP4* and *LTBP2*) clearly cluster together and are surrounded by other genes relevant for COPD pathogenesis (figure 2). The gene network for the downregulated genes was less consistent (see online supplemental figure S1).

### Validation of gene expression findings using qPCR

qPCR for *FBLN5*, *ELN*, *MFAP4* and *LTBP2* was performed on lung tissue samples from 45 patients with COPD and 42 non-COPD controls from the Groningen cohort. Subject characteristics are given in online supplementary table S2.

We found significantly higher *ELN* expression in COPD compared with control lung tissue and similar trends for *FBLN5* ( $p = 0.08$ ) and *MFAP4* ( $p = 0.08$ ) (figure 3A). There was no difference in *LTBP2* expression. In addition, significant negative correlations were found between FEV<sub>1</sub> and *FBLN5* (figure 3B), and FEV<sub>1</sub> and *ELN*, and FEV<sub>1</sub> and *MFAP4* (see online supplemental

Table 3A Gene function prediction in biological pathways

<i>ELN</i>		<i>FBLN5</i>		<i>MFAP4</i>		<i>LTBP2</i>	
Biological pathways	p Value	Biological pathways	p Value	Biological pathways	p Value	Biological pathways	p Value
<b>Collagen fibril organisation</b>	8.46E-25	<b>Extracellular structure organisation</b>	4.66E-12	<b>Extracellular structure organisation</b>	1.43E-21	Regulation of leukocyte chemotaxis	8.61E-22
<b>Extracellular structure organisation</b>	2.35E-18	<b>Extracellular matrix organisation</b>	4.66E-12	<b>Extracellular matrix organisation</b>	1.43E-21	Cell-substrate adhesion	2.18E-21
<b>Extracellular matrix organisation</b>	2.35E-18	Regulation of systemic arterial blood pressure by hormone	5.22E-11	Negative regulation of androgen receptor signalling pathway	4.33E-13	Cell-matrix adhesion	1.20E-17
Blood vessel remodelling	6.07E-17	Response to amino acid stimulus	2.38E-10	<b>TGFβ receptor signalling pathway</b>	1.20E-11	<b>TGFβ receptor signalling pathway</b>	1.28E-17
<b>TGFβ receptor signalling pathway</b>	1.60E-15	Regulation of systemic arterial blood pressure by renin angiotensin	8.97E-10	Establishment of planar polarity of embryonic epithelium	3.45E-11	Positive regulation of leukocyte chemotaxis	4.68E-17
Transmembrane receptor protein serine/threonine kinase signalling pathway	9.40E-14	Chemokine-mediated signalling pathway	1.02E-08	Protein trimerisation	4.22E-11	Positive regulation of chemotaxis	1.18E-16
Striated muscle cell proliferation	3.68E-13	Regulation of blood pressure	1.76E-08	Response to prostaglandin E stimulus	7.79E-11	Regulation of protein processing	1.29E-16
Bone morphogenesis	9.59E-13	Response to amine stimulus	1.81E-08	Lung alveolus development	1.60E-10	<b>Extracellular structure organisation</b>	2.17E-16
Bone development	1.06E-12	Regulation of cell migration	2.66E-08	Transmembrane receptor protein serine/threonine kinase signalling pathway	1.81E-10	<b>Extracellular matrix organisation</b>	2.17E-16
Cardiac muscle cell proliferation	2.04E-12	Renal system process	4.33E-08	<b>Collagen fibril organisation</b>	2.45E-10	Regulation of leukocyte migration	2.80E-16
Cardiac muscle tissue growth	4.78E-12	Response to acid	5.40E-08	Respiratory tube development	3.61E-10	Negative regulation of TGFβ receptor signalling pathway	1.29E-15
Cardiac muscle cell apoptosis	4.88E-12	Regulation of blood volume by renin angiotensin	1.36E-07	Regulation of establishment of planar polarity	3.74E-10	Mononuclear cell migration	4.75E-15
Artery morphogenesis	1.73E-11	Regulation of cell motility	1.57E-07	Wnt receptor signalling pathway. Planar cell polarity pathway	3.74E-10	Negative regulation of transmembrane receptor protein serine/threonine kinase signalling pathway	5.67E-15
Negative regulation of transmembrane receptor protein serine/threonine kinase signalling pathway	2.58E-11	Positive regulation of angiogenesis	2.13E-07	Negative regulation of epithelial cell proliferation	5.48E-10	Vasculature development	1.65E-14
Artery development	5.77E-11	<b>Collagen fibril organisation</b>	3.97E-07	Cellular response to amino acid stimulus	5.54E-10	Angiogenesis	1.68E-14
Negative regulation of epithelial to mesenchymal transition	6.24E-11	Regulation of locomotion	4.42E-07	Lung development	7.23E-10	Positive regulation of behaviour	3.02E-14
Muscle cell apoptosis	1.02E-10	Regulation of cellular component movement	4.75E-07	Positive regulation of smoothed signalling pathway	1.06E-09	Regulation of cell migration	4.55E-14
Cartilage development involved in endochondral bone morphogenesis	1.38E-10	Negative regulation of transmembrane receptor protein serine/threonine kinase signalling pathway	5.76E-07	Establishment of planar polarity involved in neural tube closure	2.35E-09	Positive regulation of cell migration	5.35E-14
Cell-substrate adhesion	1.46E-10	Cellular response to organic nitrogen	5.87E-07	Cellular response to amine stimulus	3.16E-09	Blood vessel morphogenesis	5.62E-14
Negative regulation of cell morphogenesis involved in differentiation	2.02E-10	Response to organic nitrogen	6.55E-07	Positive regulation of filopodium assembly	3.29E-09	Blood vessel development	6.98E-14

Overlap in gene function in biological pathways is indicated in bold.

*ELN*, elastin; *FBLN5*, fibulin-5; *LTBP2*, latent transforming growth factor β binding protein 2; *MFAP4*, microfibrillar associated protein 4; MGI, mouse genome informatics; TGFβ, transforming growth factor β.

figure S2A). Moreover, strong coexpression was observed between these four elastogenesis genes (see online supplemental figure S2B), which shows the robustness of our findings and validates the similarity in the predicted function for these four genes.

### FBLN5 is expressed in primary pulmonary fibroblasts and strongly induced by TGFβ

As pulmonary fibroblasts are the main source of ECM proteins in the lung, and thus important for elastic fibre formation, we

investigated *FBLN5* gene expression levels in primary pulmonary fibroblasts derived from patients with COPD and non-COPD controls from the Groningen cohort. To this end, fibroblasts were treated with or without TGFβ and CSE (figure 4). Under basal conditions there was no significant difference in *FBLN5* expression between COPD and non-COPD control fibroblasts, although COPD fibroblasts tended to have higher expression. However, there was a clear increase in *FBLN5* expression after TGFβ treatment in pulmonary

**Table 3B** Gene function prediction in mouse knockout phenotypes

<i>ELN</i>		<i>FBLN5</i>		<i>MFAP4</i>		<i>LTBP2</i>	
MGI	p Value	MGI	p Value	MGI	p Value	MGI	p Value
Abnormal tendon morphology	5.65E-18	Abnormal aorta elastic tissue morphology	4.09E-19	<b>Overexpanded pulmonary alveoli</b>	8.32E-38	Osteoarthritis	1.01E-29
Centrally nucleated skeletal muscle fibres	2.99E-15	Abnormal aorta wall morphology	2.73E-18	Abnormal aorta elastic tissue morphology	1.70E-36	<b>Emphysema</b>	1.43E-24
Abnormal cutaneous collagen fibril morphology	3.94E-15	Abnormal cutaneous collagen fibril morphology	5.07E-17	Abnormal aorta wall morphology	4.75E-31	Abnormal aorta elastic tissue morphology	1.43E-19
Myopathy	1.22E-13	Abnormal tendon morphology	7.93E-16	Loose skin	2.52E-29	<b>Abnormal pulmonary elastic fibre morphology</b>	2.10E-18
Abnormal tracheal cartilage morphology	1.98E-13	Loose skin	1.52E-12	<b>Emphysema</b>	2.18E-27	Abnormal tendon morphology	4.66E-16
<b>Abnormal lung vasculature morphology</b>	1.20E-12	Abnormal pulmonary artery morphology	2.81E-10	Corneal thinning	8.77E-23	<b>Dilated respiratory conducting tubes</b>	2.27E-14
Loose skin	1.23E-11	<b>Emphysema</b>	3.21E-10	Abnormal tendon morphology	9.54E-23	Abnormal aorta wall morphology	6.85E-14
Osteoarthritis	4.73E-11	Dystrophic muscle	1.00E-08	Abnormal cutaneous collagen fibril morphology	1.31E-22	Abnormal cutaneous collagen fibril morphology	1.10E-13
<b>Emphysema</b>	8.53E-11	<b>Overexpanded pulmonary alveoli</b>	4.22E-08	Abnormal blood vessel physiology	3.84E-16	<b>Overexpanded pulmonary alveoli</b>	2.20E-13
<b>Dilated respiratory conducting tubes</b>	1.39E-10	Increased systemic arterial systolic blood pressure	7.93E-08	<b>Dilated respiratory conducting tubes</b>	1.70E-15	Abnormal tracheal cartilage morphology	3.99E-12
Dystrophic muscle	4.73E-10	Dilated heart atrium	1.91E-07	<b>Abnormal lung vasculature morphology</b>	3.13E-15	Abnormal lung morphology	2.58E-11
<b>Overexpanded pulmonary alveoli</b>	1.14E-09	Abnormal vascular smooth muscle physiology	3.56E-07	<b>Impaired lung alveolus development</b>	1.83E-14	Impaired wound healing	5.90E-11
Increased variability of skeletal muscle fibre size	1.29E-09	Decreased skin tensile strength	4.88E-07	Abnormal kidney medulla morphology	2.20E-14	Loose skin	2.01E-10
Muscle degeneration	1.34E-09	Abnormal artery morphology	5.82E-07	<b>Abnormal pulmonary elastic fibre morphology</b>	5.33E-14	Brain inflammation	2.20E-10
Abnormal vertebral body morphology	1.34E-09	<b>Dilated respiratory conducting tubes</b>	1.07E-06	Increased systemic arterial systolic blood pressure	9.43E-14	Abnormal joint morphology	4.30E-10
Small thoracic cavity	1.48E-09	Skeletal muscle fibre necrosis	1.74E-06	Increased type II pneumocyte number	1.03E-12	Decreased susceptibility to injury	6.72E-10
Kidney cortex atrophy	4.83E-09	Partial perinatal lethality	2.99E-06	Small lung	4.12E-12	Abnormal pulmonary alveolus morphology	1.50E-09
Pulmonary hypoplasia	6.17E-09	<b>Abnormal lung vasculature morphology</b>	3.12E-06	Impaired branching involved in ureteric bud morphogenesis	8.23E-12	Abnormal involution of the mammary gland	2.80E-09
Abnormal aorta wall morphology	9.20E-09	Polyuria	4.06E-06	Abnormal lung sacculle morphology	1.71E-11	Abnormal thoracic cage shape	3.56E-09
Respiratory distress	9.23E-09	Renal tubule atrophy	6.16E-06	Abnormal artery morphology	2.40E-11	Abnormal heart ventricle morphology	4.21E-09

Overlap in mouse knockout phenotypes is indicated in bold.

*ELN*, elastin; *FBLN5*, fibulin-5; *LTBP2*, latent transforming growth factor  $\beta$  binding protein 2; *MFAP4*, microfibrillar associated protein 4; MGI, mouse genome informatics; TGF $\beta$ , transforming growth factor  $\beta$ .

fibroblasts, indicating a repair response. CSE exposure did not affect *FBLN5* expression in pulmonary fibroblasts.

### **FBLN5, MFAP4 and LTBP2 are localised in elastic fibres in the lung**

Immunohistochemistry was performed on serial cut lung tissue sections to assess colocalisation of *FBLN5*, *MFAP4* and *LTBP2* with elastic fibres in the lung. Figure 5 shows clear staining of *FBLN5* in alveolar and vessel walls, and surrounding the airways, which was similar in localisation to elastin. *LTBP2* and *MFAP4* staining was present in vessel walls and the matrix surrounding airways and vessels. *LTBP2* and *MFAP4* were also present in areas containing elastic fibres.

### **Increased FBLN5 protein levels in COPD lung tissue**

*FBLN5*, *ELN* and *MFAP4* protein levels were analysed in lung tissue samples from 24 patients with COPD and 19 non-COPD

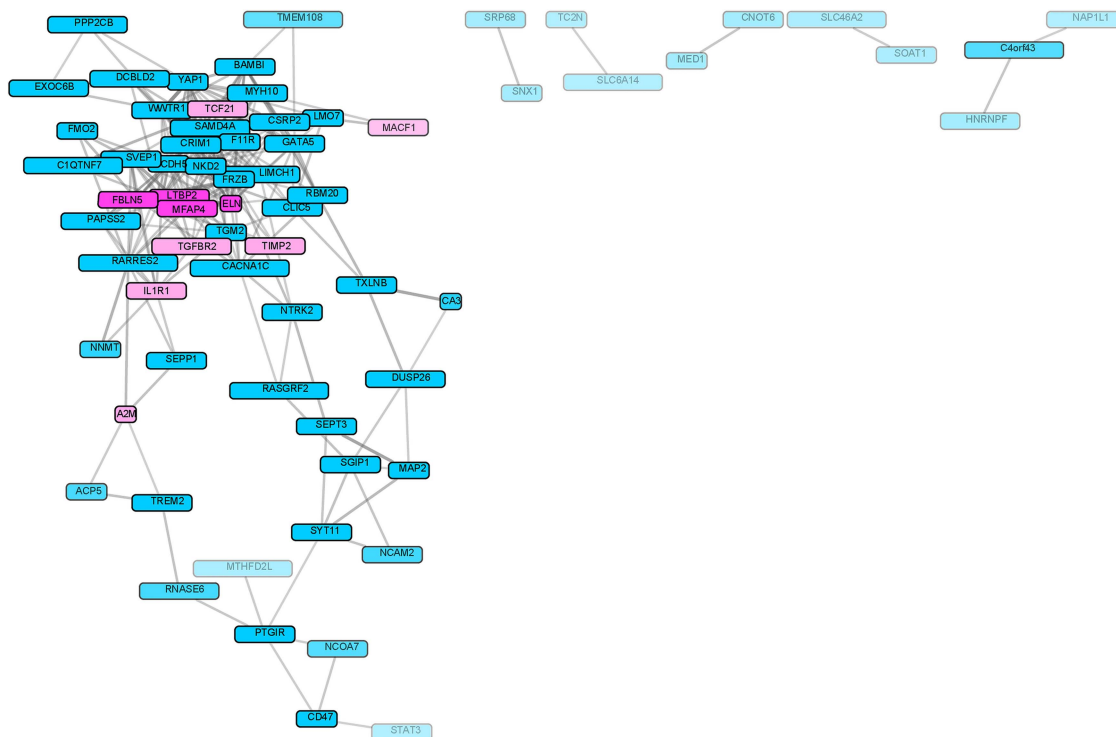
controls from the Groningen cohort (see online supplementary table S3) by western blot.

Total levels of *FBLN5* protein were significantly higher in COPD than control lung tissue (figure 6A), an effect most clearly seen when comparing control ex-smokers and COPD stage IV ex-smokers. We also demonstrated the presence of cleaved *FBLN5* protein in lung tissue and found higher levels in COPD compared with control lung tissue (figure 6B).

There were no significant differences in total protein levels of *ELN* and *MFAP4* between COPD and control lung tissue (see online supplemental figure S3), yet *ELN* protein levels were lower in current than ex-smoking controls.

### **DISCUSSION**

Using genome-wide gene expression analyses on a large dataset of lung tissue specimens from patients with COPD and non-COPD controls, we identified a clear gene signature for



**Figure 2** Cofunctionality network of genes upregulated in patients with chronic obstructive pulmonary disease (COPD). The cofunctionality network shows the clustering of all genes that are upregulated in COPD lung tissue at  $p < 0.001$ . The clustering is based on the overlap in gene function as predicted by our GeneNetwork; that is, close clustering means high overlap in predicted gene function. The four elastogenesis genes are highlighted in dark pink and cluster together in the centre of the network. Several other genes relevant for COPD pathogenesis are highlighted in light pink.

elastogenesis in COPD, with increased expression of *ELN*, *FBLN5*, *MFAP4* and *LTBP2*. Among the upregulated genes in COPD, we demonstrated a clear enrichment of genes that are known to induce an emphysematous phenotype if knocked out in mice, including *FBLN5*. Additionally, we validated part of our gene expression findings using qPCR and showed a negative correlation between *FBLN5*, *ELN* and *MFAP4* expression and lung function. We showed strong correlations between the gene expression of *ELN*, *FBLN5*, *MFAP4* and *LTBP2* and we observed colocalisation of *FBLN5*, *MFAP4* and *LTBP2* protein with elastic fibres in lung tissue. Thus the presence of these elastogenesis genes and proteins is strongly related in lung tissue, in line with the overlap in gene function that we predicted using our independent gene expression dataset. Finally, we demonstrated higher protein levels of total *FBLN5* and cleaved *FBLN5* protein in lung tissue from patients with COPD than non-COPD controls.

The gene that was most significantly upregulated in COPD, and most interesting with respect to COPD pathogenesis, was *FBLN5*. *FBLN5* is essential for elastic fibre assembly and abundantly expressed during lung and vascular development.<sup>14</sup> *FBLN5* knockout mice survive to adulthood but develop pronounced elastinopathy with human aging phenotypes like loose skin, vascular abnormalities, severe emphysema and genital prolapse.<sup>15 16</sup> This phenotype in mice is very similar to the human connective tissue disorder cutis laxa, which can be caused by mutations in *FBLN5* and is often associated with severe emphysema.<sup>19</sup> Based on these findings we consider *FBLN5* as a potential novel player in tissue repair in COPD and decided to further focus on this gene. Next to *FBLN5* and *ELN* itself, two other genes related to elastogenesis, that is, *MFAP4* and *LTBP2*, were also upregulated in COPD. *MFAP4* colocalises with elastin

in the lung<sup>18</sup> and *LTBP2* has been shown to interact with *FBLN5* and promote elastic fibre assembly.<sup>17</sup>

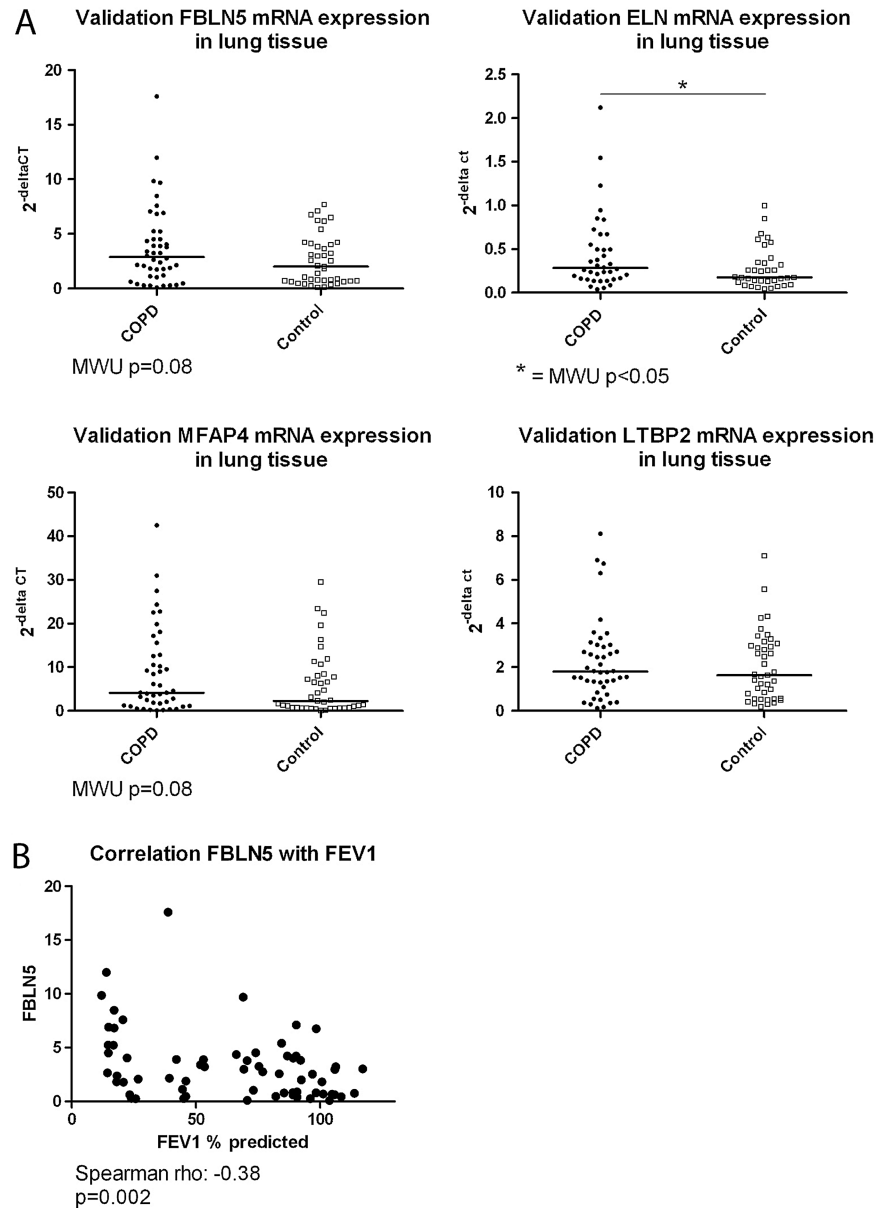
Interestingly, *FBLN5*, *LTBP2* and *MFAP4* were present in elastic fibres in all lung compartments, hence they could potentially be involved in vascular and matrix related changes associated with emphysema and COPD, which may be relevant for further studies.

Our findings are intriguing yet paradoxical as elastic fibres are destroyed with emphysema, resulting in severely impaired lung elasticity in COPD. We propose that upregulation of these elastogenesis genes in COPD is the result of an attempt to repair the damaged lung, an attempt that is not effective. This hypothesis is supported by the observation that *FBLN5* gene expression increases after elastase-induced emphysema in mice,<sup>20</sup> and is also in line with previous studies demonstrating increased elastin gene expression in severe COPD.<sup>21</sup> Furthermore, our results and those of Kuang *et al*,<sup>22</sup> showing increased *FBLN5* gene expression after TGF $\beta$  treatment in lung fibroblasts, support a role for *FBLN5* in repair responses.

A possible explanation for the fact that this repair response in COPD is not effective might be found in defects in protein translation. Two studies have now demonstrated that the *FBLN5* protein can be cleaved by serine proteases<sup>23 24</sup> and increased levels of this cleaved form of *FBLN5* protein were demonstrated in the skin of aged mice.<sup>23</sup> This is intriguing as on the one hand COPD has been regarded as an aging lung disease<sup>25 26</sup> and on the other hand increased serine protease activity is present in COPD lungs, which is thought to play an important role in emphysematous lung tissue destruction.<sup>27 28</sup> Moreover, the cleaved form of *FBLN5* does not promote elastic fibre assembly in vitro, and thus seems to be non-functional.<sup>23</sup> To investigate the role of cleaved *FBLN5*, we measured its protein expression



**Figure 3** Quantitative PCR validation. (A) Fibulin-5 (FBLN5), elastin (ELN), microfibrillar associated protein 4 (MFAP4) and latent transforming growth factor  $\beta$  binding protein 2 (LTBP2) mRNA expression relative to housekeeping gene expression in lung tissue comparing patients with chronic obstructive pulmonary disease (COPD) (closed symbols) and non-COPD controls (open symbols). The results of the Mann–Whitney U (MWU) analyses are depicted in the figure. (B) The negative correlation between the expression of *FBLN5* in lung tissue and FEV<sub>1</sub> % predicted is shown. The result of the Spearman correlation is depicted below the figure.



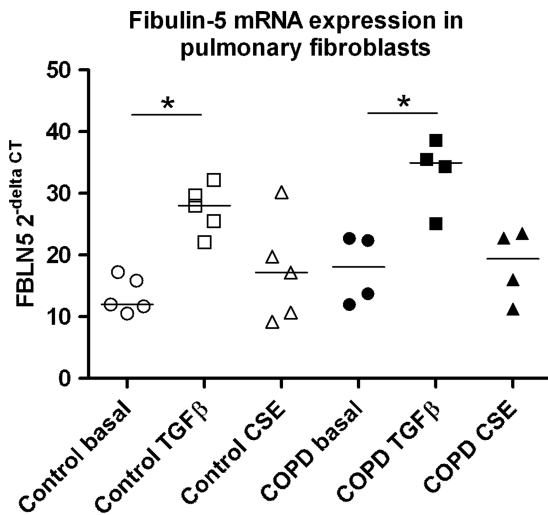
in lung tissue and found higher levels in patients with COPD than in non-COPD controls.

In addition to the four genes related to elastogenesis, we found several other genes to be upregulated in COPD lung tissue that might be relevant for COPD pathogenesis. *MACF1* is involved in the actin cytoskeleton organisation and was demonstrated to be a positive regulator of WNT signalling,<sup>29</sup> an important lung developmental pathway.<sup>30</sup> *TGFBR2* is one of the receptors for TGF $\beta$ , which is involved in many cellular processes including lung development and wound repair and often implicated in COPD.<sup>31</sup> *TCF21* is involved in epithelial–mesenchymal interactions in kidney and lung morphogenesis and is important in epithelial differentiation and branching morphogenesis.<sup>32</sup>  $\alpha$ 2-Macroglobulin (*A2M*) is a protease inhibitor and a genetic defect in this gene has been associated with COPD.<sup>33</sup>

Among the downregulated genes in COPD lung tissue, *VEGFA* and *POSTN* were most relevant for COPD pathogenesis. VEGF is an important stimulator of angiogenesis and decreased VEGF levels have been associated with emphysema in humans

and in mice.<sup>34</sup> *POSTN* is a matricellular protein that has been associated with T helper 2 related subepithelial fibrosis in asthma<sup>35</sup> and progression of fibrosis in patients with idiopathic pulmonary fibrosis.<sup>36</sup>

A strength of our study was the very stringent way of analysing gene expression data. To identify the differentially expressed genes we corrected for age, gender, smoking status and pack-years and studied each of the three cohorts separately and subsequently conducted a meta-analysis. However, we also corrected the expression data for the 25 most dominant expression effects (identified by principal component analysis). One reason to do this is that some of these components might capture unknown, confounding factors that otherwise could result in false positive associations. While this approach invariably will result in false negative findings, the small set of genes that we report on here showed a highly consistent expression difference between cases and controls that was present in each of the three cohorts, showing the robustness of our findings. Opposed to our stringent approach we decided to use a lenient FDR cut-off of 20% to assess the differentially expressed genes. This resulted in 252



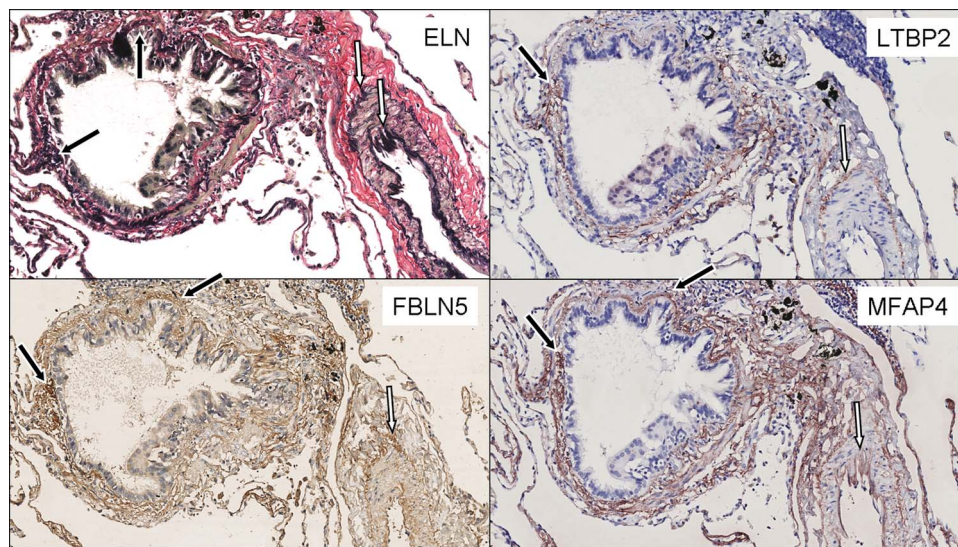
**Figure 4** Fibulin-5 (FBLN5) expression in pulmonary fibroblasts. *FBLN5* gene expression in pulmonary fibroblasts from control subjects (open symbols, n=5) and patients with chronic obstructive pulmonary disease (COPD) (closed symbols, n=4) is shown during basal conditions (circles) and after transforming growth factor  $\beta$  (TGF $\beta$ ) (squares) and cigarette smoke extract (CSE) (triangles) treatment. Differences between the groups were tested by Mann–Whitney U tests. \* $p < 0.05$ .

differentially expressed genes as opposed to 51.6 that are expected by chance. The FDR was estimated very conservatively, assuming that the 52 738 tests (ie, each probe present on the expression platform) were entirely uncorrelated. In reality, this

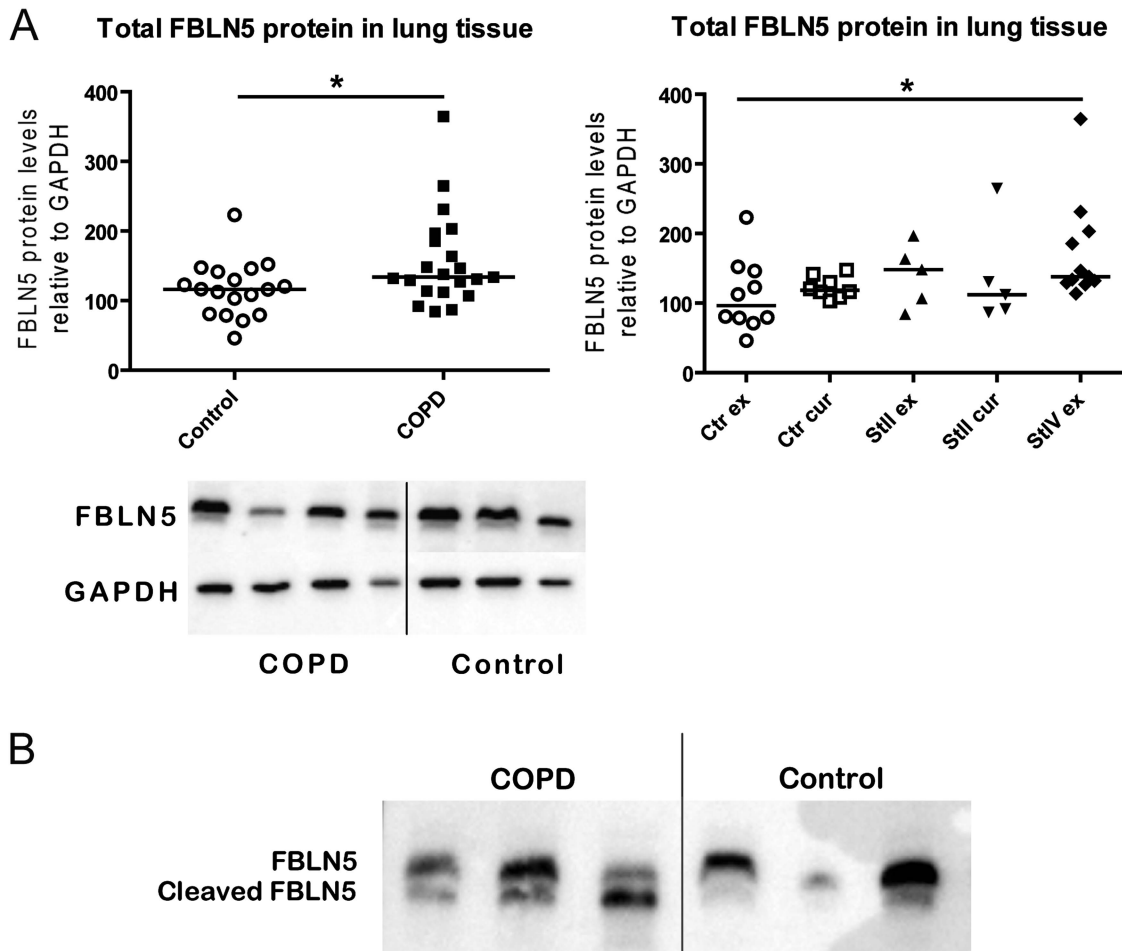
is not the case, since many probes (especially those mapping within the same gene) typically show strong coexpression. As such we feel confident that the actual FDR of these 252 differentially expressed probes is lower than 20%. We choose this FDR as a hypothesis generating cut-off, and this might give some false positive results. Therefore we performed several follow-up experiments and validated our main findings on mRNA and protein level, demonstrated colocalisation with elastic fibres in lung tissue and showed cofunctionality between the upregulated genes.

We used a novel method that can accurately predict gene function based on an independent gene expression dataset of 77 840 samples. A clear overlap in predicted gene function was found for the genes related to elastogenesis and for the other upregulated genes which are relevant for COPD pathogenesis. Visualisation in a cofunctionality network showed that these genes share many functions. Moreover, among the genes upregulated in COPD, enrichment was found for genes that are predicted to give an emphysematous phenotype in knockout mice, confirming that our results are indeed specific for COPD.

In conclusion, we identified a clear elastogenesis gene signature in COPD, with increased gene expression of *ELN*, *FBLN5*, *MFAP4* and *LTBP2* in lung tissue. We provide evidence for *FBLN5* as a novel player in abnormal tissue repair in COPD. We suggest that increased *FBLN5* mRNA expression in COPD lungs is due to a response to tissue damage and repair, and that cleavage of *FBLN5* protein is increased in COPD lungs due to high serine protease activity. Cleaved *FBLN5* is non-functional and does not contribute to elastic fibre assembly and tissue repair in COPD. Functional studies are currently ongoing to unravel the exact role of *FBLN5* in tissue repair in COPD.



**Figure 5** Elastin (ELN), fibulin-5 (FBLN5), latent transforming growth factor  $\beta$  binding protein 2 (LTBP2) and microfibrillar associated protein 4 (MFAP4) staining in lung tissue. Presence of elastic fibres (black) is demonstrated with a Verhoeff van Gieson stain in the upper left panel. LTBP2 and MFAP4 staining (red) is shown in the right upper and lower panel, respectively. FBLN5 staining (brown) is shown in the lower left panel. Examples of colocalisation with elastic fibres are indicated with black arrows in the matrix around the airways and white arrows in the vessel walls.



**Figure 6** Total and cleaved fibulin-5 (FBLN5) protein levels in lung tissue. (A) FBLN5 protein levels relative to GAPDH expression in lung tissue comparing patients with chronic obstructive pulmonary disease (COPD) and non-COPD controls. The left graph shows the comparison between all patients with COPD (closed symbols) and controls (open symbols). The right graph shows the subgroups based on smoking status and COPD stages. A representative western blot of FBLN5 and GAPDH is shown below the graphs. Differences between the groups were tested by Mann–Whitney U tests. \* $p < 0.05$ . (B) Western blot staining showing total FBLN5 and cleaved FBLN5 protein levels in lung tissue of three patients with COPD and three non-COPD controls. GAPDH, glyceraldehyde 3-phosphate dehydrogenase.

#### Author affiliations

- <sup>1</sup>Department of Pathology and Medical Biology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
- <sup>2</sup>University of Groningen, University Medical Center Groningen, Groningen Research Institute for Asthma and COPD, Groningen, The Netherlands
- <sup>3</sup>Department of Pulmonary Diseases, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
- <sup>4</sup>The University of British Columbia, Center for Heart Lung Innovation, St Paul's Hospital, Vancouver, Canada
- <sup>5</sup>Respiratory Division, University of British Columbia, Vancouver, Canada
- <sup>6</sup>Institut universitaire de cardiologie et de pneumologie de Québec, Québec, Canada
- <sup>7</sup>Department of Molecular Medicine, Laval University, Québec, Canada
- <sup>8</sup>Department of Genetics, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
- <sup>9</sup>Merck Research Laboratories, Boston, Massachusetts, USA
- <sup>10</sup>Department of Molecular Pharmacology, University of Groningen, Groningen, The Netherlands

**Contributions** C-AB was involved in lung tissue collection and compiling the lung tissue database, contributed to the study concept and design, interpreted the data, supervised the basic laboratory research and drafted the manuscript. MvdB was involved in compiling the lung tissue database and gathering of clinical data, contributed to the study concept and design, was involved in data analyses and data interpretation, and writing of the manuscript. DSP, WT, PDP, DDS, YB, ML contributed to the study concept and design, were involved in lung tissue collection and compiling the lung tissue database, including clinical information, and the writing and final editing of the manuscript. In addition, DSP and WT contributed to the data interpretation. DCN and KH were involved in the study concept and design,

edited the final manuscript and were responsible for the gene expression profiling. MRJ, AIRS and SB conducted the basic laboratory research and contributed to the writing and final editing of the manuscript. JK and RSNF developed the GeneNetwork under supervision of LF. LF had full access to all the data in the study and takes responsibility for the integrity of the data and accuracy of the data analyses. In addition, he contributed to the writing and final editing of the manuscript. C-AB and MvdB are co-first authors; WT and LF are co-last authors.

**Funding** This study was funded by the Dutch Lung Foundation (3.2.11.024) and Merck Research Laboratories.

**Competing interests** None.

**Ethics approval** Ethics committees of the Institut universitaire de cardiologie et de pneumologie de Québec (Quebec cohort) and the UBC-Providence Health Care Research Institute Ethics Board (Vancouver cohort).

**Provenance and peer review** Not commissioned; externally peer reviewed.

#### REFERENCES

- 1 Hogg JC, Timens W. The pathology of chronic obstructive pulmonary disease. *Annu Rev Pathol* 2009;4:435–59.
- 2 McDonough JE, Yuan R, Suzuki M, et al. Small-airway obstruction and emphysema in chronic obstructive pulmonary disease. *N Engl J Med* 2011;365:1567–75.
- 3 Spira A, Beane J, Pinto-Plata V, et al. Gene expression profiling of human lung tissue from smokers with severe emphysema. *Am J Respir Cell Mol Biol* 2004;31:601–10.

- 4 Ning W, Li CJ, Kaminski N, *et al.* Comprehensive gene expression profiles reveal pathways related to the pathogenesis of chronic obstructive pulmonary disease. *Proc Natl Acad Sci U S A* 2004;101:14895–900.
- 5 Golpon HA, Coldren CD, Zamora MR, *et al.* Emphysema lung tissue gene expression profiling. *Am J Respir Cell Mol Biol* 2004;31:595–600.
- 6 Wang JM, Stepanians S, Boie Y, *et al.* Gene expression profiling in patients with chronic obstructive pulmonary disease and lung cancer. *Am J Respir Crit Care Med* 2008;177:402–11.
- 7 Bhattacharya S, Srisuma S, Demeo DL, *et al.* Molecular biomarkers for quantitative and discrete COPD phenotypes. *Am J Respir Cell Mol Biol* 2009;40:359–67.
- 8 Hao K, Bosse Y, Nickle DC, *et al.* Lung eQTLs to help reveal the molecular underpinnings of Asthma. *PLoS Genet* 2012;8:e1003029.
- 9 Bosse Y, Sin D, Laviolette M, *et al.* Hypothesis-driven research on genomic data derived from a large scale lung eQTL mapping study. *Webmed Cent Lund* 2010;1:WMC00724.
- 10 Bosse Y, Postma DS, Sin DD, *et al.* Molecular signature of smoking in human lung tissues. *Cancer Res* 2012;72:3753–63.
- 11 Fehrmann RS, Jansen RC, Veldink JH, *et al.* Trans-eQTLs reveal that independent genetic variants associated with a complex phenotype converge on intermediate genes, with a major role for the HLA. *PLoS Genet* 2011;7:e1002197.
- 12 Cvejic A, Haer-Wigman L, Stephens JC, *et al.* SMIM1 underlies the Vel blood group and influences red blood cell traits. *Nat Genet* 2013;45:542–5.
- 13 Westra HJ, Peters MJ, Esko T, *et al.* Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet* 2013;45:1238–43.
- 14 Yanagisawa H, Schluterman MK, Brekken RA. Fibulin-5, an integrin-binding matricellular protein: its function in development and disease. *J Cell Commun Signal* 2009;3:337–47.
- 15 Yanagisawa H, Davis EC, Starcher BC, *et al.* Fibulin-5 is an elastin-binding protein essential for elastic fibre development in vivo. *Nature* 2002;415:168–71.
- 16 Nakamura T, Lozano PR, Ikeda Y, *et al.* Fibulin-5/DANCE is essential for elastogenesis in vivo. *Nature* 2002;415:171–5.
- 17 Hirai M, Horiguchi M, Ohbayashi T, *et al.* Latent TGF-beta-binding protein 2 binds to DANCE/fibulin-5 and regulates elastic fiber assembly. *EMBO J* 2007;26:3283–95.
- 18 Schlosser A, Thomsen T, Shipley JM, *et al.* Microfibril-associated protein 4 binds to surfactant protein A (SP-A) and colocalizes with SP-A in the extracellular matrix of the lung. *Scand J Immunol* 2006;64:104–16.
- 19 Callewaert B, Su CT, Van Damme T, *et al.* Comprehensive clinical and molecular analysis of 12 families with type 1 recessive cutis laxa. *Hum Mutat* 2013;34:111–21.
- 20 Kuang PP, Goldstein RH, Liu Y, *et al.* Coordinate expression of fibulin-5/DANCE and elastin during lung injury repair. *Am J Physiol Lung Cell Mol Physiol* 2003;285:L1147–52.
- 21 Deslee G, Woods JC, Moore CM, *et al.* Elastin expression in very severe human COPD. *Eur Respir J* 2009;34:324–31.
- 22 Kuang PP, Joyce-Brady M, Zhang XH, *et al.* Fibulin-5 gene expression in human lung fibroblasts is regulated by TGF-beta and phosphatidylinositol 3-kinase activity. *Am J Physiol Cell Physiol* 2006;291:C1412–21.
- 23 Hirai M, Ohbayashi T, Horiguchi M, *et al.* Fibulin-5/DANCE has an elastogenic organizer activity that is abrogated by proteolytic cleavage in vivo. *J Cell Biol* 2007;176:1061–71.
- 24 Budatha M, Silva S, Montoya TI, *et al.* Dysregulation of protease and protease inhibitors in a mouse model of human pelvic organ prolapse. *PLoS ONE* 2013;8:e56376.
- 25 Ito K, Barnes PJ. COPD as a disease of accelerated lung aging. *Chest* 2009;135:173–80.
- 26 Lee J, Sandford A, Man P, *et al.* Is the aging process accelerated in chronic obstructive pulmonary disease? *Curr Opin Pulm Med* 2011;17:90–7.
- 27 Hiemstra PS, van Wetering S, Stolk J. Neutrophil serine proteinases and defensins in chronic obstructive pulmonary disease: effects on pulmonary epithelium. *Eur Respir J* 1998;12:1200–8.
- 28 Barnes PJ, Shapiro SD, Pauwels RA. Chronic obstructive pulmonary disease: molecular and cellular mechanisms. *Eur Respir J* 2003;22:672–88.
- 29 Chen HJ, Lin CM, Lin CS, *et al.* The role of microtubule actin cross-linking factor 1 (MACF1) in the Wnt signaling pathway. *Genes Dev* 2006;20:1933–45.
- 30 Pongracz JE, Stockley RA. Wnt signalling in lung development and diseases. *Respir Res* 2006;7:15.
- 31 Konigshoff M, Kneidinger N, Eickelberg O. TGF-beta signaling in COPD: deciphering genetic and cellular susceptibilities for future therapeutic regimen. *Swiss Med Wkly* 2009;139:554–63.
- 32 Quaggin SE, Schwartz L, Cui S, *et al.* The basic-helix-loop-helix protein pod1 is critically important for kidney and lung organogenesis. *Development* 1999;126:5771–83.
- 33 Poller W, Barth J, Voss B. Detection of an alteration of the alpha 2-macroglobulin gene in a patient with chronic lung disease and serum alpha 2-macroglobulin deficiency. *Hum Genet* 1989;83:93–6.
- 34 Voelkel NF, Vandivier RW, Tudor RM. Vascular endothelial growth factor in the lung. *Am J Physiol Lung Cell Mol Physiol* 2006;290:L209–21.
- 35 Takayama G, Arima K, Kanaji T, *et al.* Periostin: a novel component of subepithelial fibrosis of bronchial asthma downstream of IL-4 and IL-13 signals. *J Allergy Clin Immunol* 2006;118:98–104.
- 36 Naik PK, Bozyk PD, Bentley JK, *et al.* Periostin promotes fibrosis and predicts progression in patients with idiopathic pulmonary fibrosis. *Am J Physiol Lung Cell Mol Physiol* 2012;303:L1046–56.



## Online supplement

### A large lung gene expression study identifying fibulin-5 as a novel player in tissue repair in COPD

C.A. Brandsma<sup>1,3\*</sup>, M. van den Berge<sup>2,3\*</sup>, D.S. Postma<sup>2,3</sup>, M.R. Jonker<sup>1,3</sup>, S. Brouwer<sup>1,3</sup>, P.D. Paré<sup>4,5</sup>, D.D. Sin<sup>4,5</sup>, Y. Bossé<sup>6,7</sup>, M. Laviolette<sup>6</sup>, J. Karjalainen<sup>8</sup>, R.S.N. Fehrmann<sup>8</sup>, D.C. Nickle<sup>9</sup>, K. Hao<sup>9</sup>, A.I.R. Spanjer<sup>3,10</sup>, W. Timens<sup>1,3#</sup>, L. Franke<sup>8#</sup>

\* Co-first author, # Co-last author

## Methods

### *Lung tissue samples*

Non-tumor lung tissues were collected from patients who underwent lung resection surgery at three participating sites: University of Groningen (Groningen, The Netherlands), Laval University (Quebec City, Canada), and University of British Columbia (Vancouver, Canada). At Groningen, lung specimens were provided by the local tissue bank of the Department of Pathology and the study protocol was consistent with the Research Code of the University Medical Center Groningen and Dutch national ethical and professional guidelines (“Code of conduct; Dutch federation of biomedical scientific societies”; <http://www.federa.org>). At Laval, lung specimens were collected from patients undergoing lung cancer surgery and stored at the “Institut universitaire de cardiologie et de pneumologie de Québec” (IUCPQ) site of the Respiratory Health Network Tissue Bank of the “Fonds de recherche du Québec – Santé” ([www.tissuebank.ca](http://www.tissuebank.ca)). Written informed consent was obtained from all subjects and the study was approved by the IUCPQ ethics committee. At Vancouver, the lung specimens were provided by the James Hogg Research Center Biobank at St Paul's Hospital and subjects provided written informed consent. The study was approved by the ethics committees at the UBC-Providence Health Care Research Institute Ethics Board.

### ***Immunohistochemical staining for FBLN5, MFAP4, and LTBP2***

FBLN5 staining in lung tissue was demonstrated using a mouse monoclonal antibody from Novus Biologicals (Littleton, CO, USA) in a 1:600 dilution using Tris/EDTA pretreatment. Peroxidase-labeled rabbit-anti-mouse and goat-anti-rabbit antibodies were used as secondary and tertiary steps respectively. Staining was visualized using 3,3'-diaminobenzidine (DAB). MFAP4 and LTBP2 staining in lung tissue was demonstrated using polyclonal antibodies from Proteintech (Chicago, IL, USA) and Atlas Antibodies (Stockholm, Sweden) respectively, using 1:200 (MFAP4) and 1:300 (LTBP2) dilutions and citrate (MFAP4) and Tris/HCL (LTBP2) pretreatment. Peroxidase-labeled goat-anti-rabbit was used as secondary antibody. Staining was visualized using NovaRED (Vector Labs, Burlingame, CA, USA).

### ***GeneNetwork analysis***

A new method was used to gain insight into the potential gene function of differentially expressed genes, i.e. GeneNetwork (Fehrmann *et al*, manuscript in preparation). This method uses an independent gene expression dataset of 77,840 samples to predict the function of genes in an unbiased way, and has recently been employed in a study identifying a new blood-group gene (S1) and a GWAS study on educational attainment (S2). Conceptually, prediction of gene function is conducted using a guilt-by-association approach in gene expression data; e.g. if there are 100 genes that are known to be involved in apoptosis, identification of a gene that is strongly co-expressed with these 100 genes suggests that this gene is likely to be involved in apoptosis as well. As such, we used this method to predict (currently unknown) gene functions based on known biological pathways available in the MSigDB and additionally used this information for pathway enrichment analyses and co-functionality networks. Furthermore, knock-out information from the Mouse Genome Informatics (MGI) database was collected for genes for which human orthologues exist and we used this information to predict phenotypes in mice for which currently no knock-out has been described.

In order to build our GeneNetwork we conducted a principal component analysis on an unprecedented scale. We collected gene expression data for three different species (homo sapiens, mus musculus and rattus norvegicus) from the Gene Expression Omnibus. We confined analyses to four different Affymetrix expression platforms (Affymetrix Human Genome U133A Array, Affymetrix Human Genome U133 Plus 2.0 Array, Affymetrix Mouse Genome 430 2.0 Array and Affymetrix Rat Genome 230 2.0 Array). For each of these platforms we downloaded the raw CEL files (20,108, 43,278, 18,639 and 6,124 arrays, respectively), and used RMA for normalization. We could run RMA on all samples at once for the 20,108 Human Genome U133A Array, 18,639 Mouse Genome 430 2.0 Array and 6,123 Rat Genome 230 2.0 Array. For the 43,278 Human Genome U133 Plus 2.0 Array samples we ran RMA in eight batches due to its size, by randomly assigning the samples to one of these batches. We subsequently conducted quality control (QC) on the data. We first removed duplicate samples, and subsequently conducted a principal component analysis (PCA) on the sample correlation matrix. The first principal component ( $PC_{qc}$ ) on such a matrix describes nearly always a constant pattern (dominating the data) which explains around 80-90% of the total variance. This pattern can be regarded as probe-specific variance, independent of the biological sample hybridized to the array. The correlation of each individual microarray with this  $PC_{qc}$  can be used to detect outliers, as arrays of lesser quality will have a lower correlation with the  $PC_{qc}$ . We removed samples that had a correlation  $R < 0.75$ . After QC in total 77,840 different samples remained for downstream analysis (54,736 human samples, 17,081 mouse samples, 6,023 rat samples). Although this QCed dataset can be well used for the aforementioned guilt-by-association co-expression analysis, we reasoned that the presence of profound effects on many genes will make it difficult to identify the more subtle relationships that exist between genes. Therefore we conducted a PCA on the probe correlation matrix, resulting in the identification of in total 2,206 robustly estimated principal components (377 for Human Genome U133A, 777 for Human Genome U133 Plus 2.0, 677 for Mouse Genome 430 2.0 and 375 for Rat Genome 230 2.0) by requiring a Cronbach's alpha  $> 0.70$  for each individual principal component.

Jointly these components explain between 79% and 90% of the variance in the data per Affymetrix expression platform, and many of these are well conserved across the three species.

Subsequent Gene Set Enrichment Analysis (GSEA) revealed that each of these 2,206 components are significantly enriched (False discovery rate < 0.05) for at least one GO term, KEGG, BioCarta or Reactome pathway, indicating that these components are describing biologically relevant but often diverse phenomena. While per species the very first components describe profound effects on expression (i.e. many enriched pathways and GO terms), the other components are potentially equally biologically relevant, as each of the components describe certain biological phenomena. We therefore used the individual components and integrated the different platforms and species by collapsing the probe identifiers to human Ensembl genes and used orthology information from Ensembl for the mouse and rat platform, resulting in a harmonized matrix of 19,997 unique Ensembl genes x 2,206 principal components.

We subsequently predicted the most likely Gene Ontology (GO) biological process using the following strategy: We first ascertained each individual GO term and assessed per PC whether the genes that were explicitly annotated with this GO term showed a significant difference from the genes that were not annotated with this GO term using a T-Test. We converted the resulting P-Value into an 'enrichment' Z-Score (to ensure normality). Per gene we correlated the 2,206 PC eigenvector coefficients of the gene with each GO term by taking the 2,206 'enrichment' Z-Scores as the expression profile for that GO term. A significant positive correlation means the gene has an expression profile that is comparable to the GO term. We have visualized this method at [www.genenetwork.nl/genenetwork](http://www.genenetwork.nl/genenetwork), click on "Method").

The same procedure can be applied to mouse knock-out phenotypes as well. To do so, we collected knock-out information from the Mouse Genome Informatics (MGI) for which human orthologues exist. Using the above procedure we are thus able to predict phenotypes for genes for which no mouse knock-out exists yet.



### **qPCR validation**

Quantitative PCR analysis was used to confirm the expression levels of *ELN* (elastin), *FBLN5* (fibulin-5), *MFAP4* (microfibrillar associated protein), and *LTBP2* (latent TGF $\beta$ -binding protein 2) in lung tissue samples from the Groningen cohort.

Additionally, *FBLN5* expression was analyzed in primary pulmonary fibroblasts that were treated with or without TGF $\beta$  from 4 stage IV COPD patients and 5 non-COPD controls from the Groningen cohort.

RNA was isolated using the miRNeasy mini kit (Qiagen) and RNA quantity and quality were determined using the Nanodrop D1000 (Nanodrop Technologies, Wilmington, USA) and visual inspection on agarose gel. qPCR was performed on a LightCycler<sup>®</sup> 480 (Roche Applied Science, Mannheim, Germany) according to the manufacturers' instructions using Taqman gene expression assays from Applied Biosystems (Foster City, CA, USA). Data were analyzed by the 2-delta-Ct method and *RPS9* (40S ribosomal protein S9) and *POLR2A* (Polymerase (RNA) II (DNA Directed) Polypeptide A) were used for normalization in fibroblasts and lung tissue respectively.

### **Western Blot analysis**

Western Blot analyses was performed to measure total protein levels of FBLN5, ELN and MFAP4 in lung tissue and to identify the presence of the cleaved form of FBLN5.

7.5  $\mu$ g of protein was separated using 10% polyacrylamide gels and blotted on a nitrocellulose membrane using standard methods. Nonspecific binding was blocked using 5% skim milk and membranes were incubated overnight with a mouse monoclonal antibody against FBLN5 (Novus Biologicals), a rabbit polyclonal antibody against MFAP4 (Proteintech) and a rabbit polyclonal antibody against ELN (Cedarlane, Burlington, Canada). The staining was visualized using appropriate peroxidase-labeled secondary antibodies followed by Supersignal chemiluminescent substrate (Pierce, Etten-Leur, Netherlands). Blots were scanned using the Gel Doc XR system from Bio-Rad Lifescience (Veenendaal, Netherlands) and the bands of interest were analyzed using Image Lab software 4.0.1 from

Bio-Rad. Subsequently the blots were stained for GAPDH (mouse monoclonal, Santa Cruz Biotechnology, Dallas, Texas, USA) or  $\beta$ -Actin (rabbit polyclonal, Abcam, Cambridge, UK) for protein loading control. FBLN5, ELN and MFAP4 values were expressed as percentages of GAPDH (FBLN5) or  $\beta$ -Actin (ELN and MFAP4).

***Isolation, culture and TGF $\beta$  and CSE treatment of primary pulmonary fibroblasts***

Primary lung fibroblasts were cultured from peripheral lung tissue of 4 patients with very severe stage IV COPD undergoing lung transplant surgery and 5 non-COPD controls with normal lung function undergoing tumor resection surgery. Primary fibroblasts were isolated and characterized using our explant technique as described previously (S3-S4). Fibroblasts were cultured in complete culture medium (Ham's F12 supplemented with 10% (v/v) FBS, 2 mM L-glutamine, 50 U/ml penicillin, 50  $\mu$ g/ml streptomycin and 1.5  $\mu$ g/ml amphotericin B (all from Lonza, Breda, the Netherlands)). Experiments were performed at passage 5. After reaching confluence, fibroblasts were cultured for 24 hours in culture medium with 0.5% (v/v) FBS before the stimulations started. Subsequently, cells were stimulated for 4 hours with 2 ng/ml TGF- $\beta$ 1 or 5% CSE in culture medium with 0.5% (v/v) FBS. Afterwards cells were lysed using the lysis buffer of the miRNeasy Micro Kit (Qiagen, Venlo, the Netherlands).

## Results

**Table S1; Complete list of all probe sets including statistics.**

See separate Excel file

**Table S2; Subject characteristics of the Groningen cohort used for qPCR validation**

**Subject characteristics of lung tissue samples used for validation qRT-PCR**

	<b>Number</b>	<b>Age</b>	<b>FEV1%pred.</b>	<b>FEV1/FVC</b>	<b>Packyears</b>
<b>Non-COPD control subjects</b>	42	61 (12.8)	96.0 (15.6)	75.7 (7.2)	35.0 (22.5)
<b>COPD patients</b>	45	58 (17.0)	39.1 (45.0)	46.2 (30.0)	37.5 (20.4)

*Median (interquartile range) is depicted*

**Table S3; Subject characteristics of Groningen cohort that were used for Western Blot analysis**

**Subject characteristics of lung tissue samples used for validation Western Blot**

	<b>Number</b>	<b>Age</b>	<b>FEV1%pred.</b>	<b>FEV1/FVC</b>	<b>Packyears</b>
<b>Non-COPD control ex-smokers</b>	10	64.5 (14.8)	97.6 (17.0)	76.4 (5.3)	37.5 (24.8)
<b>Non-COPD control current smokers</b>	9	62 (14.0)	92.4 (13.1)	73.9 (9.6)	36 (13.5)
<b>COPD stage II ex-smokers</b>	6	70 (13.3)	71.8 (3.9)	51.6 (15.7)	21 (11.0)
<b>COPD stage II current smokers</b>	5	56 (6.0)	76 (2.8)	58.3 (1.6)	35 (13.0)
<b>COPD stage IV ex-smokers</b>	13	58 (3.0)	20.4 (5.5)	24 (7.8)	37 (10.0)

*Median (interquartile range) is depicted*

**Supplemental Figure 1; Co-functionality network of genes downregulated in COPD patients**

The co-functionality network shows the clustering of all genes that are downregulated in COPD lung tissue at  $p < 0.001$ . The clustering is based on the overlap in gene function as predicted by our GeneNetwork; i.e. close clustering means high overlap in predicted gene function.

**Supplemental Figure 2; Correlation between expression of elastogenesis genes and FEV1 in lung tissue**

A) The negative correlation between the expression of ELN and MFAP4 in lung tissue and FEV1% predicted is shown. The result of the Spearman correlation is depicted below the figures.

B) The positive correlations between the expression of FBLN5, ELN, MFAP4 and LTBP2 in lung tissue measured by qRT-PCR is shown for the different gene combinations. The results of the Spearman correlations are depicted below the figures.

**Supplemental Figure 3; Total protein levels of ELN and MFAP4 in lung tissue**

ELN and MFAP4 protein levels relative to  $\beta$ -actin expression in lung tissue comparing COPD patients and non-COPD controls. The left graph shows the comparison between all COPD patients (closed symbols) and controls (open symbols). The right graph shows the subgroups based on smoking status and COPD stages. Differences between the groups were tested by Mann Whitney U tests. \* =  $p < 0.05$ .



## **Supplemental references**

(S1) Cvejic A, Haer-Wigman L, Stephens JC, et al. SMIM1 underlies the Vel blood group and influences red blood cell traits. *Nat Genet* 2013.

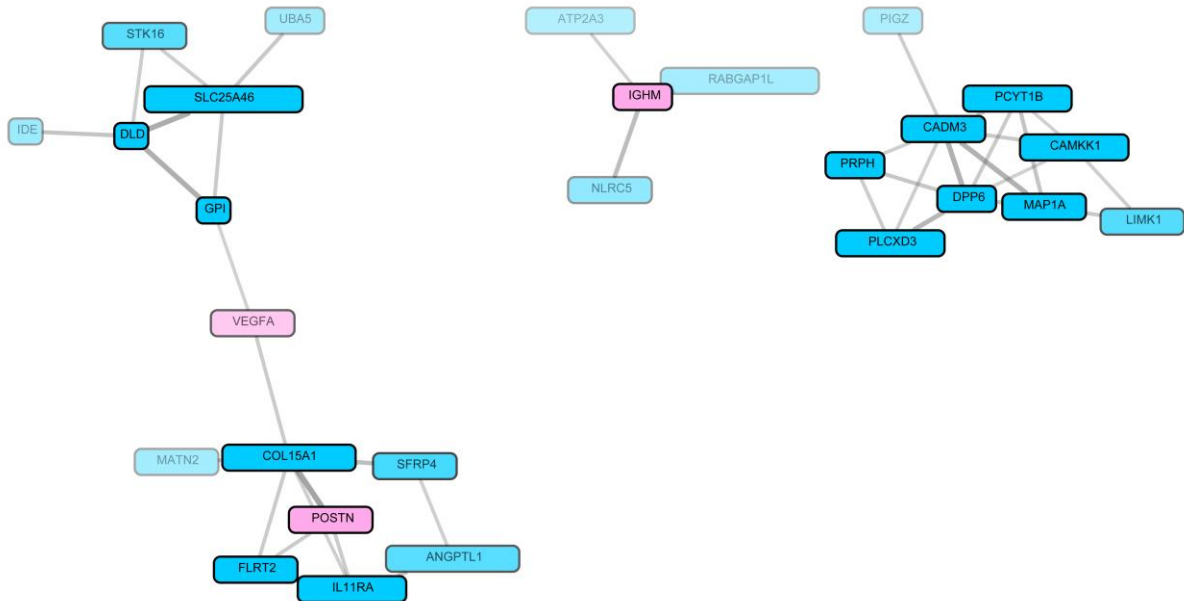
(S2) Rietveld CA, Medland SE, Derringer J, et al. GWAS of 126,559 Individuals Identifies Genetic Variants Associated with Educational Attainment. *Science* 2013.

(S3) Noordhoek J, Postma DS, Chong LL, et al. Different modulation of decorin production by lung fibroblasts from patients with mild and severe emphysema. *COPD: Journal of Chronic Obstructive Pulmonary Disease* 2005;2:17-25.

(S4) Zandvoort A, Postma DS, Jonker MR, et al. Smad gene expression in pulmonary fibroblasts: indications for defective ECM repair in COPD. *Respir Res* 2008;9:83

### Supplemental Figure 1; Co-functionality network of genes downregulated in COPD patients

The co-functionality network shows the clustering of all genes that are downregulated in COPD lung tissue at  $p < 0.001$ . The clustering is based on the overlap in gene function as predicted by our GeneNetwork; i.e. close clustering means high overlap in predicted gene function.

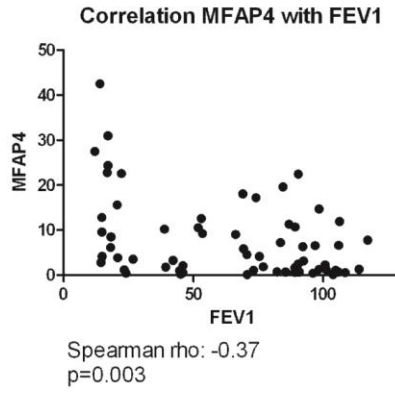
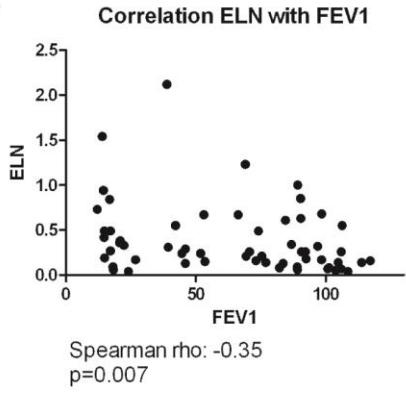


### Supplemental Figure 2; Correlation between expression of elastogenesis genes and FEV1 in lung tissue

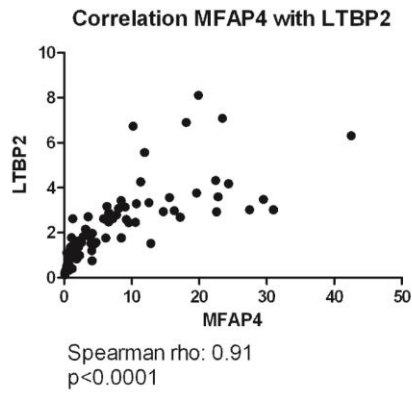
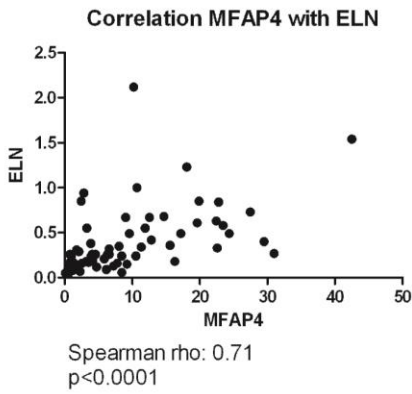
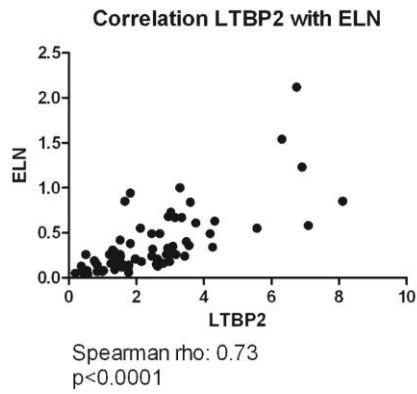
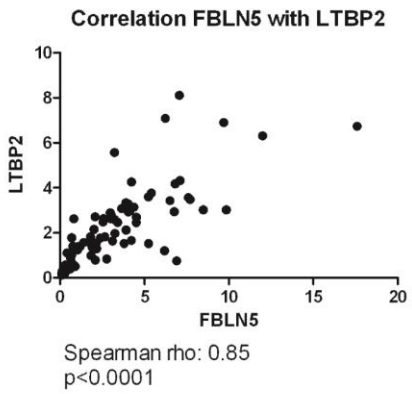
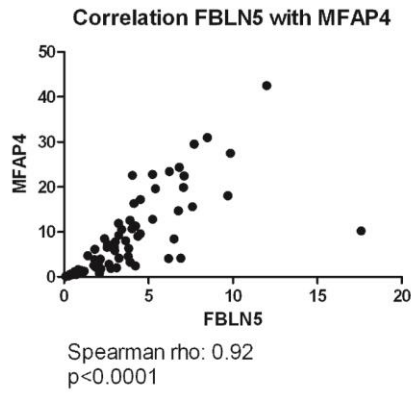
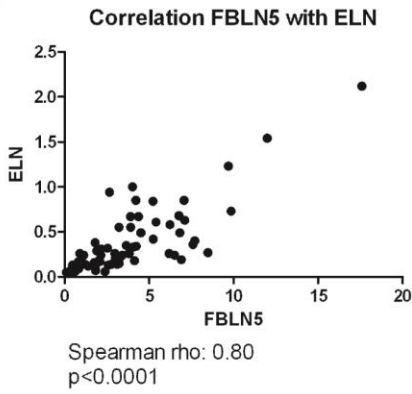
A) The negative correlation between the expression of ELN and MFAP4 in lung tissue and FEV1% predicted is shown. The result of the Spearman correlation is depicted below the figures.

B) The positive correlations between the expression of FBLN5, ELN, MFAP4 and LTBP2 in lung tissue measured by qRT-PCR is shown for the different gene combinations. The results of the Spearman correlations are depicted below the figures.

A



B



### Supplemental Figure 3; Total protein levels of ELN and MFAP4 in lung tissue

ELN and MFAP4 protein levels relative to  $\beta$ -actin expression in lung tissue comparing COPD patients and non-COPD controls. The left graph shows the comparison between all COPD patients (closed symbols) and controls (open symbols). The right graph shows the subgroups based on smoking status and COPD stages. Differences between the groups were tested by Mann Whitney U tests. \* =  $p < 0.05$ .

