Online supplement

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

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

* 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). 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**, 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 TGFβ-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 TGFβ 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® 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 µg 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 β-Actin (rabbit polyclonal, Abcam, Cambridge, UK) for protein loading control. FBLN5, ELN and MFAP4 values were expressed as percentages of GAPDH (FBLN5) or β-Actin (ELN and MFAP4).

Isolation, culture and TGFβ 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 µg/ml streptomycin and 1.5 µg/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-β1 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

	Number	Age	FEV1%pred.	FEV1/FVC	Packyears
Non-COPD control subjects	42	61 (12.8)	96.0 (15.6)	75.7 (7.2)	35.0 (22.5)
COPD patients	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

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