ONLINE SUPPLEMENT

microRNA-mRNA regulatory networks underlying chronic mucus hypersecretion in COPD

Hataitip Tasena[1,2], Alen Faiz[1,2,3], Wim Timens[1,2], Jacobien Noordhoek[1,2,3], Machteld N Hylkema[1,2], Reinoud Gosens[2,4], Pieter S Hiemstra[5], Avrum Spira[6], Dirkje S Postma[2,3], Gaik W Tew[7], Michele A Grimbaldeston[8], Maarten van den Berge[2,3], [*]Irene H Heijink[1,2,3], [*]Corry-Anke Brandsma[1,2]

([*]both authors contributed equally)

[1]University of Groningen, University Medical Centre Groningen, Department of Pathology and Medical Biology, Groningen, the Netherlands.

[2]University of Groningen, University Medical Centre Groningen, Groningen Research Institute for Asthma and COPD, Groningen, the Netherlands.

[3]University of Groningen, University Medical Centre Groningen, Department of Pulmonary Diseases, Groningen, the Netherlands.

[4]University of Groningen, Department of Molecular Pharmacology, Groningen, the Netherlands.

[5]Leiden University Medical Centre, Department of Pulmonology, Leiden, The Netherlands.

[6]Boston University Medical Centre, Department of Medicine, Division of Computational Biomedicine, Boston, Massachusetts, USA.

[7]Research & Early Development, Genentech Inc, South San Francisco, California, USA.

[8]OMNI-Biomarker Development, Genentech Inc, South San Francisco, California, USA.

Definitions of chronic mucus hypersecretion (CMH)

CMH was defined based on patient's responses to clinical questionnaires including two questions: A) "How often did you cough up sputum during the last three months?" and B) "How often did you cough up sputum during the last week?" The latter was derived from the clinical COPD questionnaire (CCQ) [1]. Question A had four optional answers: i) never, ii) sometimes, iii) almost daily, and iv) only during a cold. Patients who responded i) or iv) were classified as group 1: no CMH. Patients who responded ii) were classified as group 2: mild CMH. Patients who responded iii) were classified as group 3: moderate/severe CMH. Question B had seven optional answers: i) never, ii) sometimes, iii) once in a while, iv) often, v) most of the times, vi) regularly, and vii) always. Patients who responded i) were classified as group 1: no CMH. Patients who responded ii) and iii) were classified as group 2: mild CMH. Patients who responded iv) to vii) were classified as group 3: moderate/severe CMH.

MicroRNA (miRNA) and messenger RNA (mRNA) profiling

Both total mRNA and miRNA was extracted from bronchial biopsies at baseline (n=63) as part of the GLUCOLD study [2, 3]. mRNA profiling was performed at Boston University Microarray Resource Facility as described in GeneChip® Whole Transcript (WT) Sense Target Labeling Assay Manual (Affymetrix, Santa Clara, CA, current version available at www.affymetrix.com) as previously described [4]. Microarray hybridization was performed using Affymetrix Human Gene_ST v1.0 Arrays. Isolation of small fractions of RNA for miRNA profiling was done using the miRNeasy mini kit (Qiagen) and RNeasy MinElute Cleanup Kit (Qiagen) according to the manufacturer's protocols. MiRNA profiling was performed at Boston University Microarray Resource Facility using FlashTag™ Biotin HSR Labeling Kit (Affymetrix, Santa Clara, CA, current version available at www.affymetrix.com). miRNAs and mRNAs that were expressed in less than 50% of the

patients were excluded. The filtering of miRNAs was done using the standard Affymetrix Expression Console software, and of mRNAs was done using the Robust Multichip Average algorithm and the Entrez Gene Chip Definition File (CDF) v11.0.1[5]. As a result, the expression profiles used for all analyses included 230 miRNAs and 19,793 mRNAs. Both microarray datasets can be accessed at http://www.ncbi.nlm.nih.gov/geo/ (series accession number GSE76774 for miRNA; GSE36221 for mRNA).

To validate the findings from the microarrays, we used the mRNA profiles from the RNA-sequencing (RNA-seq) conducted on libraries belonging to biopsies obtained from 76 GLUCOLD patients. Total mRNA was extracted as described previously [2, 3]. The libraries were prepared using Ribo-Zero Gold kit and sequenced using 50bp single end read sequencing. The "FastQC" program version 0.11.5 was utilized to perform quality control checks on the raw sequence data from the RNA-Seq (http://www.bioinformatics.babraham.ac.uk). The sequences were trimmed using the java program "trimmomatic 0.33" [6] and the RNA-Seq mapping was conducted using the "Spliced Transcripts Alignment to a Reference (STAR)" program version 2.5.3a, which is an RNA-Seq aligner program [7]. We checked the quality of the libraries by calculating the raw count percentages of the forward reads. Libraries with a percentage lower than 90% were excluded from further analysis. Principal component analysis was also performed (using R) on the libraries in order to detect extreme outlier. After quality check, we decided to keep all samples for further analyses. This dataset can be made available for future potential collaboration upon request. The R package "edgeR_3.16.5" was utilized to perform differential expression analysis on the samples taken from CMH patients at baseline (n=38). We compared mRNA profiles of the patients with moderate/severe CMH (n=22 for definition A and n=21 for definition B) with the patients with no CMH (n=5 for both definitions). These analyses were corrected for age, gender and smoking status.

Identification of CMH-associated miRNAs and mRNAs

A linear regression model was used to identify miRNAs and mRNAs that were differentially expressed in patients with mild CMH or with moderate/severe CMH compared to those without CMH. The model was corrected for age, gender, smoking history and RNA integrity number scores (RIN) as described below, where Me_i represents the $\log_2$ miRNA/gene expression value for a miRNA/gene in a sample from patient i, ε_i represents the error that is assumed to be normally distributed.

$$Me_i = \beta_0 + \beta_1 X_{RIN-i} + \beta_2 X_{Smoking\ history-i} + \beta_3 X_{Age-i} + \beta_4 X_{Gender-i} + \beta_5 X_{CMH\ status-i} + \varepsilon_i$$

MiRNA-mRNA co-expression network analysis

Pearson correlation analysis was used to determine the correlation between the expression of each CMH-associated miRNA and all mRNAs in the same patient based on False Discovery Rate (FDR)<0.25. The lists of miRNA's predicted target genes based on TargetScan (v6.2), miRTarBase (v4.5) and miRDB (v5.0) were combined. Those predicted targets that were negatively correlated with the candidate miRNAs were used for creating miRNA-mRNA co-expression networks on Cytoscape (v3.4.0).

Gene Set Enrichment Analysis (GSEA)

GSEA (v2.2.2) was used for the following purposes: 1) to identify enriched pathways (Kegg, v5.2; Biocarta, v5.2), biological processes (Gene Ontology, v5.2), and molecular functions (Gene Ontology, v5.2) in which CMH-associated mRNAs may be involved, 2) to determine enrichment of a previously published set of MUC5AC-associated genes [8] among mRNAs

higher expressed with CMH, 3) to determine enrichment of CMH-associated mRNAs and MUC5AC-associated genes among miRNA-correlated mRNAs, and 4) to determine enrichment of CMH-associated mRNAs identified in the microarray dataset among the CMH-associated mRNAs identified in the RNA-seq dataset. Genes were ranked according to the strength of their t-statistic reflecting their association with CMH or their correlation with the miRNA of interest. Enrichment p-values were calculated after 1000 permutations were performed. Significant enrichment was determined by an FDR<0.01 for the first purpose and by $p<0.05$ for the second, third and fourth purposes.

STRING interaction network analysis

Of all 539 CMH-associated mRNAs, expression of 538 was significantly correlated with one or more of the 10 CMH-associated miRNAs (table S6). Interactions among these miRNAs' potential target genes were predicted using the STRING database (v10.0) and the interaction networks were created on Cytoscape (v3.4.0) with *stringApp* (v1.0.5) [9]. Sources of the interactions include text-mining, experiments, databases, co-expression, neighborhood, gene fusion and co-occurrence. Minimum required interaction score is 0.700.

Cell culture and RNA isolation

To compare the expression of candidate miRNAs in primary airway epithelial cells (PAECs) from COPD and non-COPD subjects, PAECs were obtained from 6 GOLD stage IV COPD (table S1) explanted lungs and 6 non-COPD donor lungs as described before [10]. PAECs were used in passage 3, grown to confluence in a transwell system, air-exposed for 14 days in BEGM/DMEM supplemented with 15 ng/ml retinoic acid and hormonally and growth factor-deprived overnight as described previously [10]. To compare the expression of candidate miRNAs in primary airway fibroblasts (PAFs), a separate experiment was performed. PAFs

were isolated from GOLD stage IV COPD patients (table S2) with (n=8) and without CMH (n=8) as described before [11] and used in passage 5/6 and cultured in Ham's F12 medium supplemented with 10% (v/v) fetal bovine serum (FBS) until confluence. The cells were then serum deprived for 24 hours. The study protocol followed national ethical and professional guidelines ('Code of conduct; Dutch federation of biomedical scientific societies'; http://www.federa.org) for all lung tissues and explant cell culture studies in Groningen.

Reverse transcription-quantitative PCR (RT-qPCR)

Total RNA was isolated using Tri Reagent® according to the manufacturer's protocol. For the miRNAs of interest RNA was converted to cDNA using the TaqMan microRNA reverse transcription kit (Life Technologies, Bleiswijk, Netherlands) and reverse transcription primers (Life Technologies) for let-7a-5p (assay id: 000377), miR-31-5p (002279), miR-708-5p (002341), miR-134-5p (000459), miR-146a-5p (000468), miR-193a-5p (002281), miR-500a-3p (001046) and miR-1207-5p (241060_mat). As let-7d-5p and let-7f-5p were in the same cluster as let-7a-5p and these three miRNAs shared very similar results in the previous analyses, we only investigated let-7a-5p expression. qPCR was performed using the matched PCR primers (Life Technologies) and Eurogentec qPCR MasterMix Plus (05-QP2X-03; Eurogentech, Maastricht, Netherlands) on the LightCycler 480 II (Roche, Almere, Netherlands). Expression of the miRNAs of interest was normalized to the expression of the small nuclear RNA, RNU48 (001006). To determine MUC5AC expression, RNA was converted to cDNA using iScript™ cDNA Synthesis Kit (BioRad). qPCR was then performed using MUC5AC Taqman® assay (Hs01365616_m1) and the Taqman® Master Mix according to the manufacturer's guidelines (Applied Biosystems, Foster City, CA). *MUC5AC* expression was normalized to the expression of the reference genes: *B2M* (Hs99999907_m1) and *PPIA* (Hs99999904_m1).

Table S1 Characteristics of patients from which primary airway epithelial cells (PAECs) were obtained

	COPD IV patients[#] (n=6)
Gender (male), n(%)	3(50.0)
Ex-smokers, n(%)	6(100.0)
Pack-year[*]	41.7 (23.6-50.3)
Age, year[*]	55 (51-59)
FEV_1, %predicted[*]	18.8 (17.3-19.8)
FEV_1/FVC[*]	0.24 (0.22-0.25)

[*]Data are presented as median (interquartile range); FEV_1 is forced expiratory volume in 1 second; FVC is forced vital capacity; [#]two patients were CMH-positive.

Table S2 Characteristics of patients from which primary airway fibroblast (PAFs) were obtained

	COPD patients	
	no CMH (n=8)	CMH (n=8)
Gender (male), n(%)	2(25.0)	4(50.0)
Pack-year[*]	38.0 (30.0-57.0)	30.0 (28.8-36.3)
Age, year[*]	58 (57-61)	58 (50-59)
FEV_1, %predicted[*]	20.7 (17.3-22.6)	15.2 (14.2-18.5)
FEV_1/FVC[*]	0.26 (0.24-0.31)	0.27 (0.24-0.32)

[*]Data are presented as median (interquartile range); FEV_1 is forced expiratory volume in 1 second; FVC is forced vital capacity.

Table S3 microRNAs differentially expressed with CMH and their statistics

	Severe CMH vs No CMH (definition A)				Severe CMH vs No CMH (definition B)		
miRNA	**t-value**	**FC**	**p-value**	**FDR**	**t-value**	**FC**	**p-value**
let-7d-5p	2.928	1.263	4.927E-03	0.149	2.453	1.171	1.730E-02
miR-30c-5p	2.922	1.411	5.001E-03	0.149	3.365	1.423	1.400E-03
miR-708-5p	2.794	2.061	7.116E-03	0.149	2.930	2.060	4.900E-03
miR-31-5p	2.528	1.684	1.431E-02	0.182	2.566	1.613	1.300E-02
miR-27-3p	2.433	1.452	1.821E-02	0.182	3.420	1.730	1.200E-03
miR-30b-5p	2.339	1.475	2.291E-02	0.182	3.060	1.475	3.400E-03
let-7f-5p	2.227	1.700	2.996E-02	0.230	2.766	1.789	7.700E-03
let-7a-5p	2.133	1.261	3.731E-02	0.240	2.400	1.223	1.980E-02
miR-19a-3p	2.088	1.638	4.139E-02	0.244	2.463	1.529	1.690E-02
miR-134-5p	-3.507	-2.152	9.002E-04	0.149	-2.433	-1.615	1.820E-02
miR-193a-5p	-2.774	-1.785	7.519E-03	0.149	-2.545	-1.653	1.370E-02
miR-320c	-2.681	-1.355	9.638E-03	0.171	-2.874	-1.316	5.700E-03
miR-339-3p	-2.448	-1.342	1.754E-02	0.182	-2.138	-1.269	3.690E-02
miR-320a	-2.419	-1.334	1.885E-02	0.182	-2.863	-1.320	5.900E-03
miR-320b	-2.384	-1.309	2.054E-02	0.182	-2.873	-1.299	5.700E-03
miR-1207-5p	-2.371	-1.547	2.120E-02	0.182	-3.727	-1.694	5.000E-04
miR-22-3p	-2.183	-1.209	3.326E-02	0.239	-2.250	-1.230	2.840E-02
miR-500a-3p	-2.151	-1.336	3.577E-02	0.240	-2.166	-1.239	3.460E-02
miR-146a-5p	-2.117	-1.626	3.872E-02	0.240	-2.288	-1.569	2.590E-02
miR-320d	-2.106	-1.259	3.969E-02	0.240	-2.288	-1.242	2.590E-02

FC is fold change; FDR is false discovery rate.

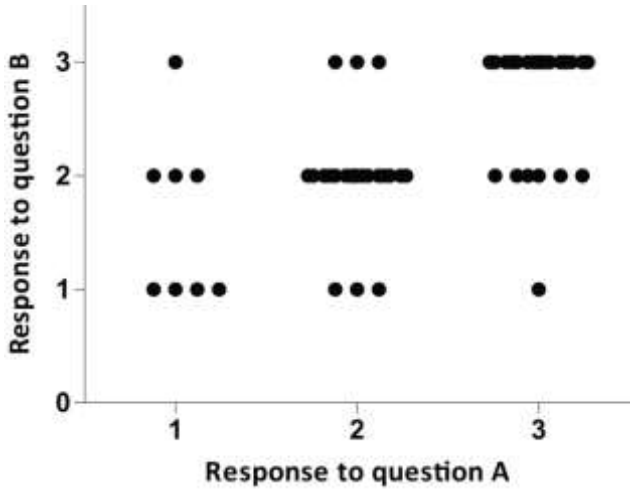

Figure S1. Responses of patients (n=63) to two questions used for defining CMH were strongly correlated. Dot plot showing the patients' answers to question A) "How often did you cough up sputum during the last three months?" and question B) "How often did you cough up sputum during the last week?". 1 represents "never cough" or "only during a cold"; 2 represents "sometimes" or "once in a while"; 3 represents "almost daily" for question A) or "often", "most of the times", "regularly", or "always" for question B). Spearman's rank correlation coefficient was performed. $r=0.651$; $p<0.0001$.

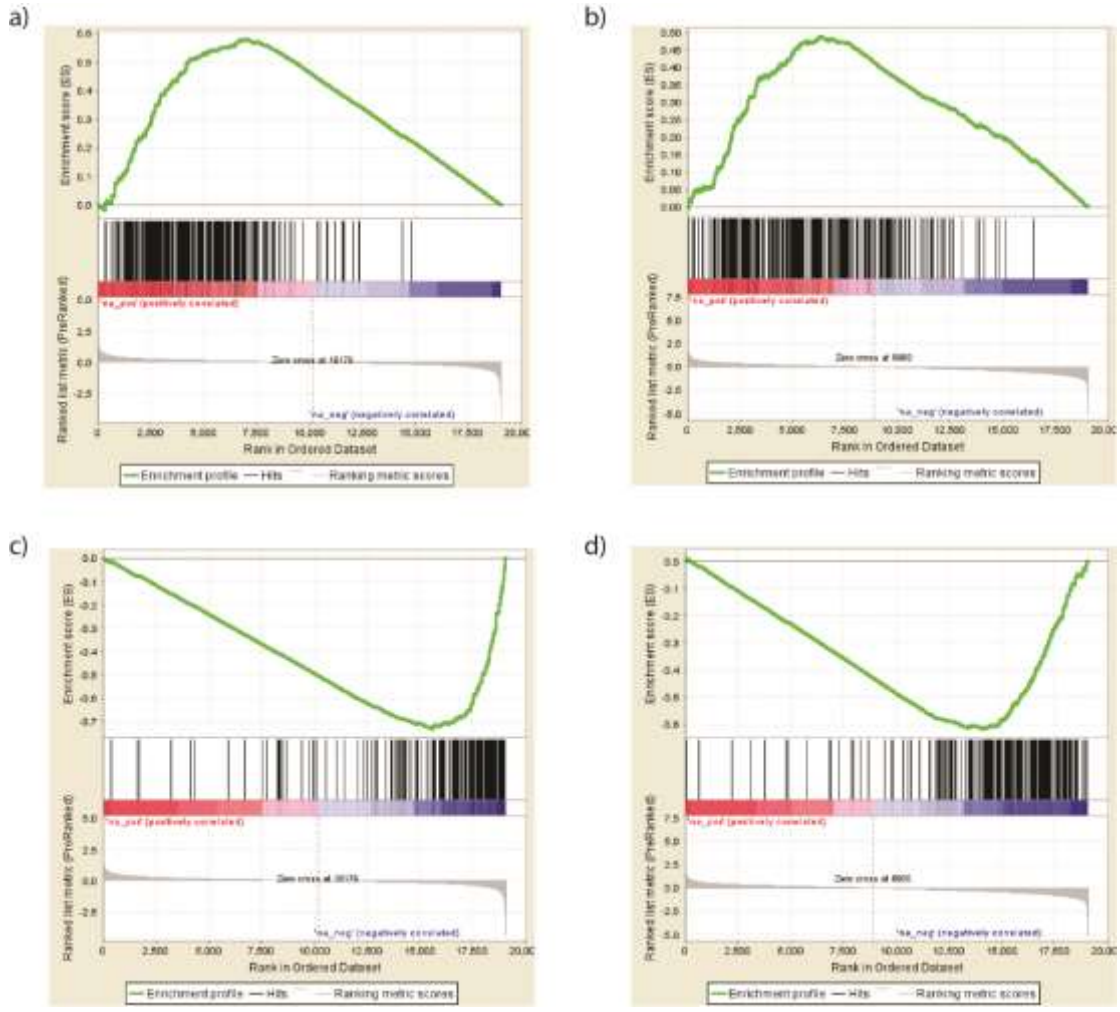

Figure S2. Gene set enrichment analysis (GSEA) on the RNA-seq dataset. Genes higher expressed with CMH in the microarray dataset were positively enriched based on a) CMH definition A and b) CMH definition B. Genes lower expressed with CMH in the microarray dataset were negatively enriched based on c) CMH definition A and d) CMH definition B. All enrichment was significant ($p<0.001$). Vertical black lines (Hits) represent position of 539 CMH-associated genes identified in the microarray dataset in the ranked gene list. From left to right, 19,075 genes were ranked by t-statistics for association with CMH in the RNA-seq dataset. Red and blue represent positive and negative association, respectively.

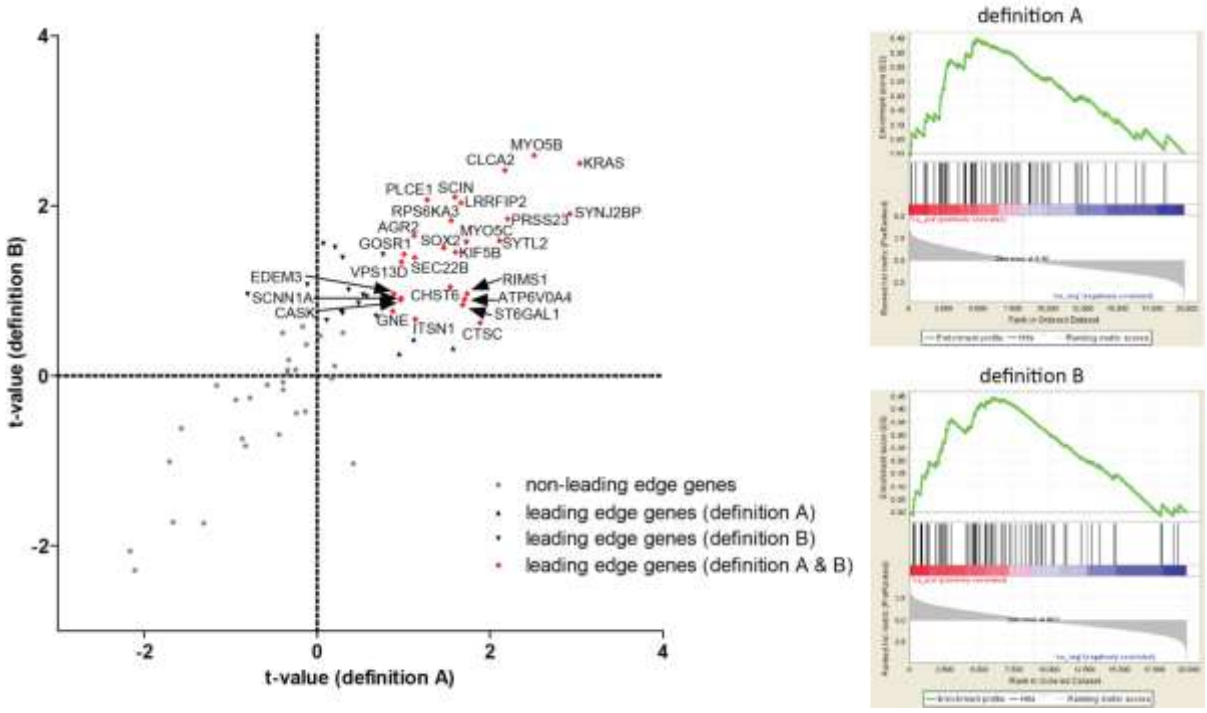

Figure S3. Overrepresentation of MUC5AC-associated genes among CMH-associated mRNAs. Enrichment analysis was performed using the ranked list of CMH-associated mRNAs, based on definition A and B, and the list of 73 MUC5AC- associated genes [8]. Significant enrichment was observed for both definition A (enrichment score (ES)=0.40, $p<0.001$) and B (ES=0.44, $p<0.001$). ES was calculated by walking down the ranked list of genes – a running-sum statistic increases when a gene is in the gene set and decreases when the gene is not. On the left graph, leading edge genes associated with CMH in both definitions are shown in red.

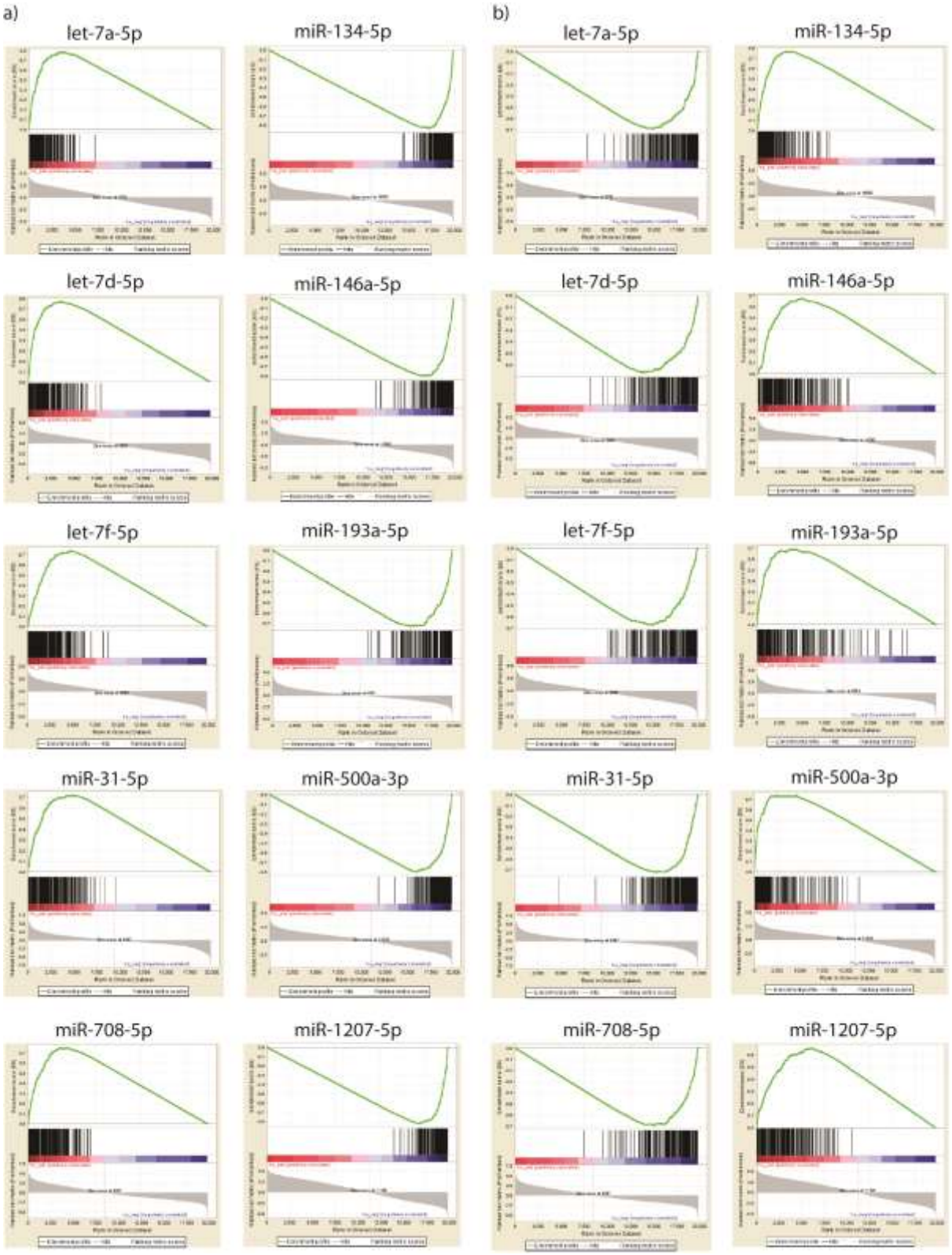

Figure S4. Enrichment of CMH-associated mRNAs among miRNA-correlated mRNAs.

a) Enrichment of mRNAs higher expressed with CMH. b) Enrichment of mRNAs lower expressed with CMH. All enrichments were significant ($p<0.001$).

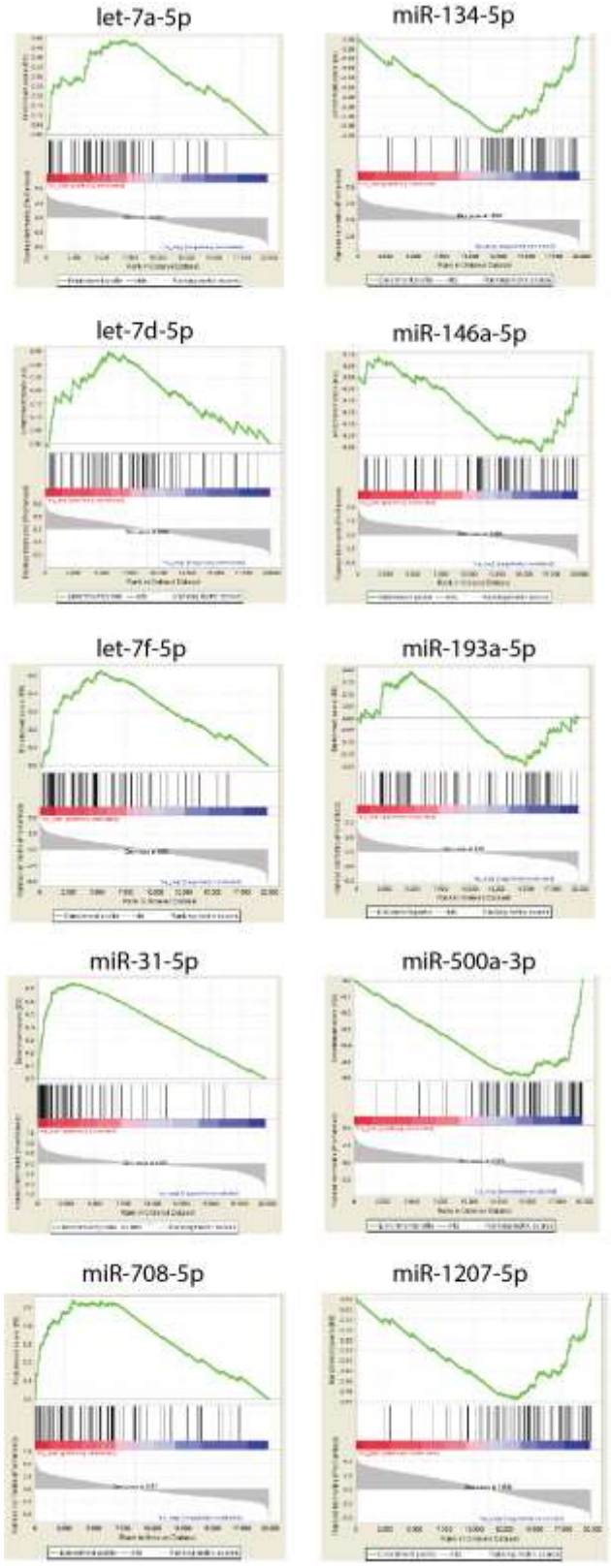

Figure S5. Enrichment of MUC5AC-core genes among miRNA-correlated mRNAs. All enrichments were significant ($p<0.005$) except for miR-193-5p.

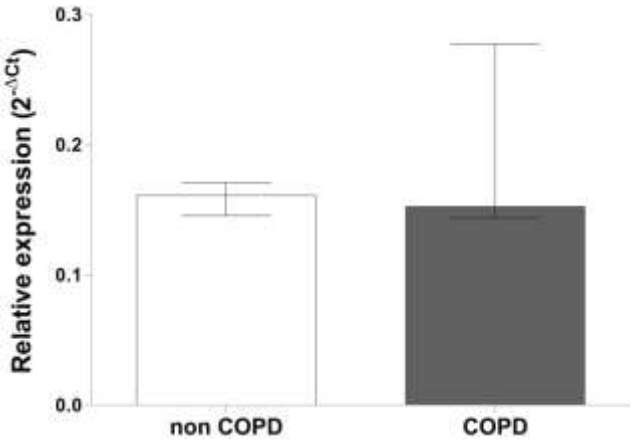

Figure S6. Expression of *MUC5AC* in primary airway epithelial cells (PAECs). PAECs were derived from COPD patients (n=6) and healthy controls (n=6) and grown at air-liquid interface for 14 days. Cells were hormonally deprived overnight and RNA was isolated. The MUC5AC expression was normalized to *B2M* and *PPIA*. Relative expression levels are shown (median±interquartile range).

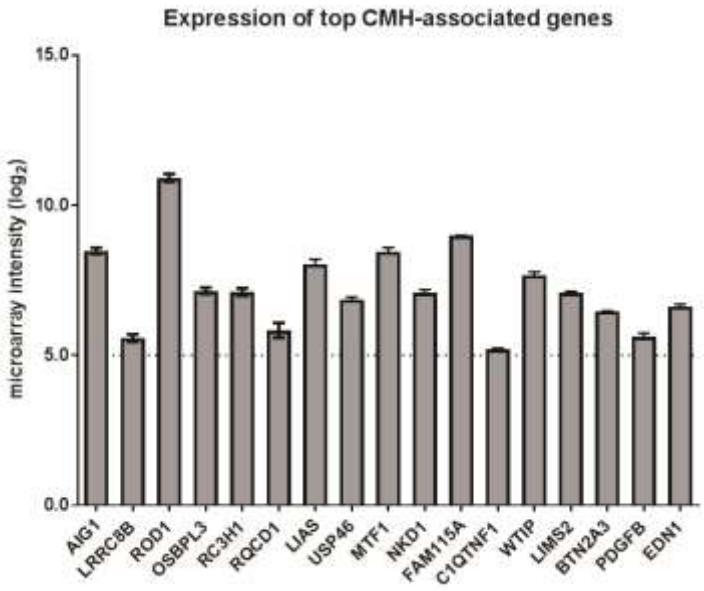

Figure S7. Expression of top CMH-associated genes in ALI-cultured primary airway epithelial cells (day 14) [12]**.** The microarray dataset is publicly available on www.ncbi.nlm.nih.gov/geo/ (series accession number GSE5264). Genes with microarray intensity above 5.0 were considered expressed.

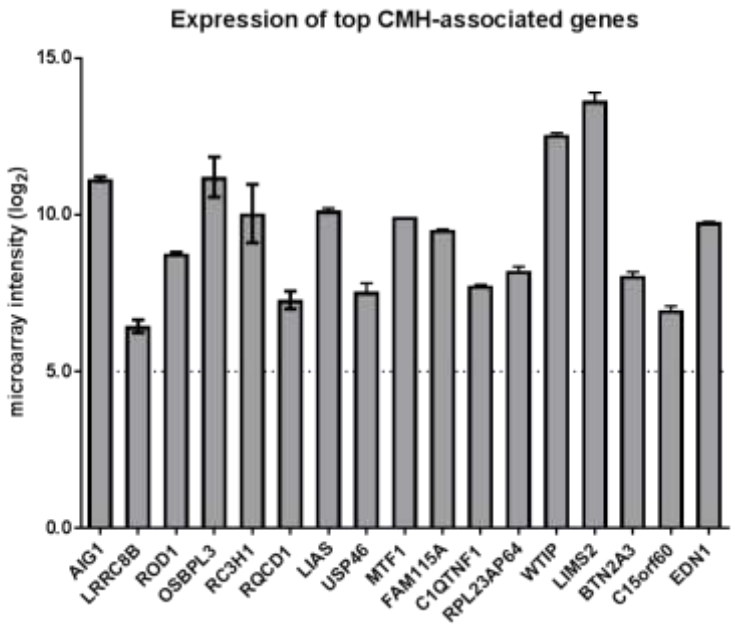

Figure S8. Expression of top CMH-associated genes in primary lung fibroblasts [13]**.**

The microarray dataset is publicly available on www.ncbi.nlm.nih.gov/geo/ (series accession number GSE86183). Genes with microarray intensity above 5.0 were considered expressed.

References

1. Molen T Van Der, Willemse BWM, Schokker S, Ht N, Postma DS, Juniper EF. Development , validity and responsiveness of the Clinical COPD Questionnaire. *Health Qual. Life Outcomes* 2003; 1: 1–10.
2. Lapperre TS, Postma DS, Gosman MME, Snoeck-Stroband JB, ten Hacken NHT, Hiemstra PS, Timens W, Sterk PJ, Mauad T. Relation between duration of smoking cessation and bronchial inflammation in COPD. *Thorax* 2006; 61: 115–121.
3. Lapperre TS, Sont JK, Schadewijk A Van, Gosman MME, Postma DS, Bajema IM, Timens W, Mauad T, Hiemstra PS, Group S. Smoking cessation and bronchial epithelial remodelling in COPD : a cross-sectional study. *Respir. Res.* 2007; 9: 1–9.
4. van den Berge M, Steiling K, Timens W, Hiemstra PS, Sterk PJ, Heijink IH, Liu G, Alekseyev YO, Lenburg ME, Spira A, Postma DS. Airway gene expression in COPD is dynamic with inhaled corticosteroid treatment and reflects biological pathways associated with disease activity. *Thorax* 2014; 69: 14–23.
5. Dai M, Wang P, Boyd AD, Kostov G, Athey B, Jones EG, Bunney WE, Myers RM, Speed TP, Akil H, Watson SJ, Meng F. Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. *Nucleic Acids Res.* 2005; 33: 1–9.
6. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 2014; 30: 2114–2120.
7. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* 2013; 29: 15–21.
8. Wang G, Xu Z, Wang R, Al-Hijji M, Salit J, Strulovici-Barel Y, Tilley AE, Mezey JG,

Crystal RG. Genes associated with MUC5AC expression in small airway epithelium of human smokers and non-smokers. *BMC Med. Genomics* 2012; 5: 21.

9. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, Kuhn M, Bork P, Jensen LJ, Von Mering C. STRING v10: Protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* 2015; 43: D447–D452.

10. Heijink IH, De Bruin HG, Dennebos R, Jonker MR, Noordhoek JA, Brandsma CA, Van Den Berge M, Postma DS. Cigarette smoke-induced epithelial expression of WNT-5B: Implications for COPD. *Eur. Respir. J.* 2016; 48: 504–515.

11. Brandsma C-A, Timens W, Jonker MR, Rutgers B, Noordhoek JA, Postma DS. Differential effects of fluticasone on extracellular matrix production by airway and parenchymal fibroblasts in severe COPD. *Am. J. Physiol. Lung Cell. Mol. Physiol.* 2013; 305: L582-9.

12. Ross AJ, Dailey LA, Brighton LE, Devlin RB. Transcriptional profiling of mucociliary differentiation in human airway epithelial cells. *Am. J. Respir. Cell Mol. Biol.* 2007; 37: 169–185.

13. Ong J, Timens W, Rajendran V, Algra A, Spira A, Lenburg ME, Campbell JD, Van Den Berge M, Postma DS, Van Den Berg A, Kluiver J, Brandsma CA. Identification of transforming growth factor-beta-regulated microRNAs and the microRNAtargetomes in primary lung fibroblasts. *PLoS One* 2017; 12: 1–24.