

RESEARCH ARTICLE

Open Access

Genes associated with MUC5AC expression in small airway epithelium of human smokers and non-smokers

Guoqing Wang^{1,5*}, Zhibo Xu^{1,2}, Rui Wang¹, Mohammed Al-Hijji¹, Jacqueline Salit¹, Yael Strulovici-Barel¹, Ann E Tilley^{1,3}, Jason G Mezey^{1,4} and Ronald G Crystal^{1,3}

Abstract

Background: Mucus hypersecretion contributes to the morbidity and mortality of smoking-related lung diseases, especially chronic obstructive pulmonary disease (COPD), which starts in the small airways. Despite progress in animal studies, the genes and their expression pattern involved in mucus production and secretion in human airway epithelium are not well understood. We hypothesized that comparison of the transcriptomes of the small airway epithelium of individuals that express high vs low levels of MUC5AC, the major macromolecular component of airway mucus, could be used as a probe to identify the genes related to human small airway mucus production/secretion.

Methods: Flexible bronchoscopy and brushing were used to obtain small airway epithelium (10th to 12th order bronchi) from healthy nonsmokers (n=60) and healthy smokers (n=72). Affymetrix HG-U133 plus 2.0 microarrays were used to assess gene expression. Massive parallel sequencing (RNA-Seq) was used to verify gene expression of small airway epithelium from 5 nonsmokers and 6 smokers.

Results: MUC5AC expression varied 31-fold among the healthy nonsmokers. Genome-wide comparison between healthy nonsmokers (n=60) grouped as "high MUC5AC expressors" vs "low MUC5AC expressors" identified 528 genes significantly up-regulated and 15 genes significantly down-regulated in the high vs low expressors. This strategy identified both mucus production and secretion related genes under control of a network composed of multiple transcription factors. Based on the literature, genes in the up-regulated list were used to identify a 73 "MUC5AC-associated core gene" list with 9 categories: mucus component; mucus-producing cell differentiation-related transcription factor; mucus-producing cell differentiation-related pathway or mediator; post-translational modification of mucin; vesicle transport; endoplasmic reticulum stress-related; secretory granule-associated; mucus secretion-related regulator and mucus hypersecretory-related ion channel. As a validation cohort, we assessed the MUC5AC-associated core gene list in the small airway epithelium of an independent set of healthy smokers (n=72). There was up-regulation of MUC5AC in the small airway epithelium of smokers (2.3-fold, $p < 10^{-8}$) associated with a coordinated up-regulation of MUC5AC-associated core gene expression pattern in the small airway epithelium of smokers ($p < 0.01$). Deep sequencing confirmed these observations.

Conclusion: The identification of the genes associated with increased airway mucin production in humans should be useful in understanding the pathogenesis of airway mucus hypersecretion and identifying therapeutic targets.

* Correspondence: geneticmedicine@med.cornell.edu

¹Department of Genetic Medicine, Weill Cornell Medical College, New York, NY, USA

⁵Department of Genetic Medicine, Weill Cornell Medical College, 1300 York Avenue, Box 9610065, New York, NY, USA

Full list of author information is available at the end of the article

Author summary

Mucus hypersecretion contributes to the morbidity and mortality of smoking-related lung diseases, especially chronic obstructive pulmonary disease (COPD), which starts in the small airways. Little is known about the gene networks associated with the synthesis and secretion of mucins in the human small airway epithelium. Taking advantage of the knowledge that MUC5AC is a major mucin secreted by the small airway epithelium, the expression of MUC5AC in small airway epithelium is highly regulated at the transcriptional level and our observation that healthy nonsmokers have variable numbers of MUC5AC⁺ secretory cells in the human small airway epithelium, we compared genome-wide gene expression of the small airway epithelium of high vs low MUC5AC expressors from 60 nonsmokers to identify the genes associated with MUC5AC expression. This novel strategy enabled identification of a 73 "MUC5AC-associated core gene" list with 9 categories, which control a series of processes from mucin biosynthesis to mucus secretion. The coordinated gene expression pattern of MUC5AC-associated core genes were corroborated in an independent cohort of 72 healthy smokers. Deep sequencing of small airway epithelium RNA confirmed these observations. This finding will be useful in identifying therapeutic targets to treat small airway mucus hypersecretion.

Background

The process of mucus production and secretion is central to the normal defense of the lung, with mucus an important component of the airway mucociliary escalator that continuously cleanses the lung of inhaled particulates, pathogens and xenobiotics [1]. Mucus is comprised of several components, but is dominated by the secretory mucins: high molecular weight, disulfide linked, heavily glycosylated proteins [2]. There has been considerable progress in defining the mucin composition of airway mucus in health and disease [3], and regulation of mucin synthesis and secretion in the mouse lung [4-9]. However, little is known about the gene networks associated with the synthesis and secretion of mucins in the human small airway epithelium, the cell population exhibiting the first abnormalities associated with smoking [10-14], and the major site of pathology in chronic obstructive pulmonary disease (COPD) [13,15].

One successful approach to understanding mucin synthesis/secretion in the human lung has been to use information from gene manipulation in murine models to identify genes associated with mucin synthesis/secretion, and to evaluate these genes and their associated genes in samples of human airways or in air-liquid interface cultures of human airway epithelium [3-9,16-22]. In order to expand this knowledge base on a genome wide level, we have developed a novel strategy to define the network of genes linked to MUC5AC expression in the human small airway epithelium. Our approach is based on the understanding that there are no mucus glands in the small airways of humans, and the secretory mucins are produced solely by surface secretory cells [1]. Taking advantage of the knowledge that MUC5AC is a major mucin secreted by the small airway epithelium, the expression of MUC5AC in small airway epithelium is highly regulated at the transcriptional level [17] and our observation that healthy nonsmokers have variable numbers of MUC5AC⁺ secretory cells in the human small airway epithelium, we

hypothesized that we could group healthy nonsmokers into small airway epithelium "high" and "low".

MUC5AC expressors. With this strategy, we compared genome-wide gene expression of the small airway epithelium of high vs low MUC5AC expressors to identify the genes associated with MUC5AC expression. Annotating this list of genes with the literature relating to mucus production/secretion enabled identification of a small airway epithelium "MUC5AC-associated core gene" list. As a validation strategy for the MUC5AC-associated core gene list, based on the knowledge that cigarette smoking is an environmental stressor associated with increased mucus production [1-3,17,18, 22,23], we assessed MUC5AC and MUC5AC-associated core gene expression in an independent cohort of healthy smokers. Not only was MUC5AC up-regulated in the smoker small airway epithelium, but many of the MUC5AC-associated core genes were as well, confirming the validity of the core gene list as being co-regulated with MUC5AC expression.

Results

MUC5AC expression in the small airway epithelium of healthy nonsmokers

Consistent with prior studies of airway mucin gene expression in humans [2,3], only 3 types of mucins (secretion-polymeric, tethered-shed in mucus and tethered) were expressed (P call % >80%) in the small airway epithelium (Table 1). Among these 8 mucin genes, we focused on MUC5AC (214385_s_at), a secreted-polymeric mucin, as it is highly expressed by airway surface mucus-producing cells [2,3,18].

To assess the proportion of MUC5AC positive cells in the small airway epithelium, MUC5AC staining was carried out on cytopsin slides prepared from 10-12th order brushed airway epithelial cells from healthy nonsmokers (Figure 1). Consistent with the literature [24,25], there

Table 1 Mucin gene expression in small airway epithelium of healthy nonsmokers

Category	Mucin gene	Probeset ID ¹	%P call	Normalized expression ²
Secreted, nonpolymeric	MUC7	217059_at	0	N/A ³
	MUC8	217295_at	0	N/A
Secreted, polymeric	MUC6	214133_at	8	N/A
	MUC19	1553436_at	0	N/A
	MUC2	204673_at	25	N/A
	MUC5AC	214385_s_at	100	54 ± 44
Tethered, shed in mucus	MUC5B	213432_at	100	48 ± 29
	MUC1	213693_s_at	100	70 ± 18
	MUC16	220196_at	100	46 ± 18
	MUC4	217109_at	100	81 ± 36
Tethered	MUC12	226654_at	6	N/A
	MUC13	218687_s_at	100	5 ± 4
	MUC15	227238_at	100	65 ± 18
	MUC17	232321_at	0	N/A
	MUC20	226622_at	100	46 ± 23
	MUC3A	217117_x_at	50	N/A
	MUC3B	214898_x_at	2	N/A

¹ If one gene had multiple probe sets, the one with highest % P was used.

² Gene expression data was normalized "per chip" in Genespring 7.0, which is normalized to the median expression level of all the genes on the chip. Data is expressed as mean ± standard deviation.

³ N/A, the expression is not shown because % P < 80%.

were MUC5AC positive cells in the small airway epithelium of healthy nonsmokers. Importantly, relevant to the present study, there was significant variability from individual to individual in the number of MUC5AC positive cells. In agreement with the histologic findings, healthy nonsmokers showed significant variation in the degree of MUC5AC gene expression. Compared to the house-keeping genes ACTB (CV = 23.9%), GAPDH (CV = 38.1%), B2M (CV = 15.2%), RPLP0 (CV = 23.0%) and PPIA (CV = 31.4%), the variability of MUC5AC was >2-fold higher (CV = 81.6%; Figure 2; Ansari-Bradley test, $p < 10^{-8}$, all comparisons).

Identification of MUC5AC-associated core genes

In order to identify genes associated with MUC5AC biosynthesis and secretion, we compared the gene expression differences between healthy nonsmoker high MUC5AC expressors vs low MUC5AC expressors. To avoid confusion of the dominant transcriptome response to oxidative stress in smokers, only nonsmokers were used to identify genes associated with MUC5AC expression. After controlling false discovery rate by Benjamini-Hochberg correction ($p < 0.05$), microarray analysis revealed 528 genes were up-regulated and 15 genes were

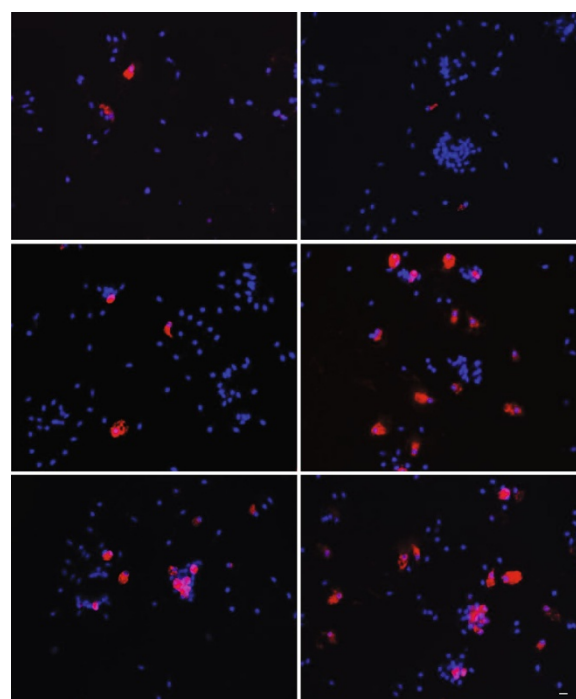


Figure 1 Variability of proportions of surface MUC5AC⁺ cells in the small airway epithelium of healthy nonsmokers.

Immunofluorescence of MUC5AC staining (red) was processed on cytospin slides prepared from 10th-12th order brushed airway epithelial cells. Shown are examples from 6 individuals. Not shown, IgG irrelevant control, negative for MUC5AC staining. Blue - DAPI. Bar = 20 μ m.

down-regulated in MUC5AC high expressors compared to the MUC5AC low expressors (Figure 3; Additional file 1: Table S2).

To explore the function of the up-regulated genes, the online annotation term enrichment tool, DAVID, was used. Genes with functional annotation of Golgi ($p < 10^{-9}$), endoplasmic reticulum (ER; $p < 10^{-5}$), and ER-Golgi transport ($p < 2 \times 10^{-4}$) were enriched in these genes (Additional file 1: Table S3). This is consistent with the knowledge that Golgi, ER apparatus, and transport between ER and Golgi are essential to the secretion pathway [26]. Based on these associations, further literature mining was carried out to identify which genes were known to be associated with or potentially involved in mucus production or secretion.

From 528 up-regulated genes, we found 73 genes with literature supported roles or potential roles in mucus production/secretion. The putative roles of these 73 "MUC5AC-core genes" were grouped into 9 categories: (1) mucus components; (2) mucus-producing cell differentiation-related transcription factors; (3) mucus-producing cell differentiation-related pathways or mediators; (4) post-translational modification of mucin; (5)

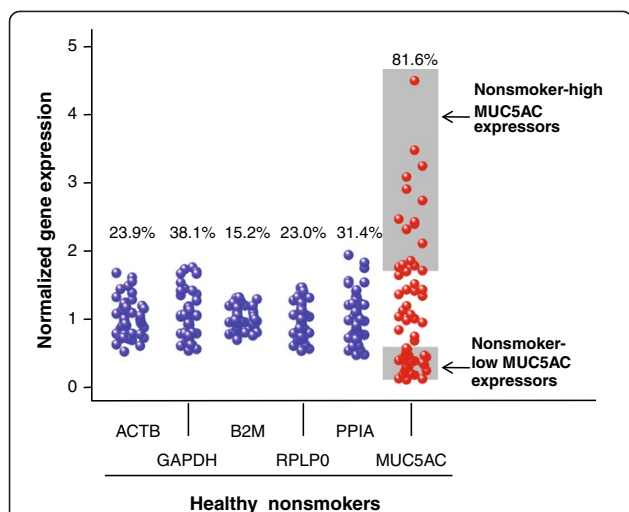


Figure 2 Healthy nonsmokers were grouped based on MUC5AC gene expression in 10-12th order brushed airway epithelial cells. Shown is data for n=60 nonsmokers divided into nonsmoker “high MUC5AC expressors” (MUC5AC expression, highest quartile of all healthy nonsmokers, n= 15, upper shaded region) and nonsmoker “low MUC5AC expressors” (MUC5AC expression, lowest quartile of all healthy non-smokers, n= 15, lower shaded region). Expression of 5 endogenous control genes (ACTB, GAPDH, B2M, RPLP0 and PPIA) in the same samples are used as controls to assess the expression variability in nonsmokers of housekeeping genes compared to MUC5AC. The coefficients of variation (CV) of MUC5AC and endogenous control gene expression are shown.

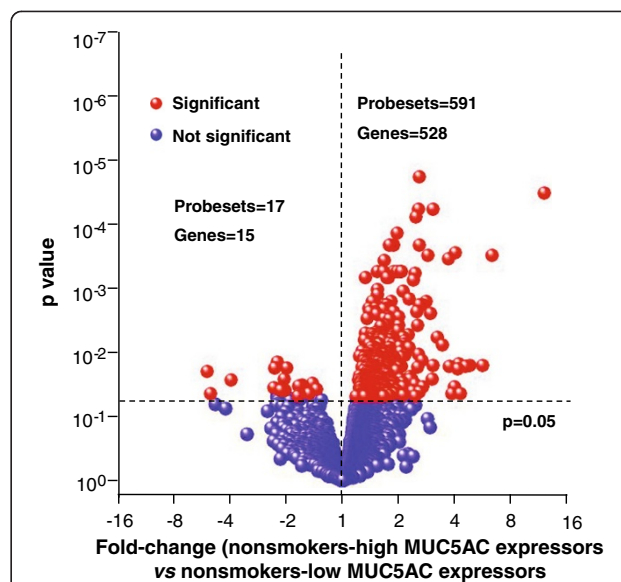


Figure 3 Identification of MUC5AC-associated core genes. Shown is a volcano plot comparing the normalized expression of gene probe sets in the 10-12th order brushed airway epithelial cells of nonsmoker-high MUC5AC expressors vs low MUC5AC expressors; y-axis - negative log of p value; x-axis - log₂-transformed fold-change; red dots - probe sets with significant p value; blue dots - probe sets with non-significant p value. Differentially expressed genes were determined by unequal variance Student’s *t*-test followed by Benjamini-Hochberg correction ($p < 0.05$). There were 528 genes identified genome-wide that were significantly correlated with MUC5AC high expressors. Of these, 73 genes were identified as MUC5AC-associated core genes based on literature mining (Table 1, Additional file 1: Table S1, S2).

vesicle transport; (6) endoplasmic reticulum stress-related; (7) secretory granule-associated; (8) mucus secretion-related regulators and (9) mucus hypersecretory-related ion channels (Table 2; see Additional file 1: Table S1 for all related references; see Additional file 1: Figure S1, S2 for the likely roles of the products of these genes in airway mucus production and secretion). Within each category, the literature regarding the relevance of each gene to small airway epithelial MUC5AC production and secretion were catalogued as to links in the literature based on non-human models, human cells *in vitro*, human airway epithelium *in vivo* and “functional association”, i.e., likely function in the airway epithelium based on the known function of the gene product.

From the assignments based on the literature, 27 of the 73 MUC5AC-core genes have been previously identified as being associated with mucus production or secretion, including those associated with mucus components (TFF3, TFF1); mucus-producing cell differentiation related-transcription factors (SPDEF, FOXA3, SOX2 and KLF4); mucus-producing cell differentiation-related pathways or mediators (HES1 and TSTA3, Notch pathway; RPS6KA3 and KRAS, MAPK pathway; PLA2G4A; CTSC; SERPINB4); post-translational modification of mucin (AGR2, GNE, GALNT4, GALNT7, GALNT12); secretory

granule-associated (SYTL2, RAB27B, RAB3D, SCIN, MYO5B, MYO5C); mucus secretion-related regulators (ITPR3, PRKCD); and a mucus hypersecretory-related ion channel (SLC12A2). Of these 27 MUC5AC-core genes known to be strongly linked to mucus production and secretion, most have been linked to airway mucus production and secretion by murine or *in vitro* studies; only 9 have been previously linked to MUC5AC production/secretion in human *in vivo* studies (Additional file 1: Table S1).

Of the remaining 46 MUC5AC-core genes, although they have not been strongly linked to airway mucus production/secretion, there is literature supporting their roles in mucus production/secretion in other cell types, including mucus-producing cell differentiation-related pathways or mediators (Wnt pathway gene, LRRFIP2 and MAPK pathway gene, MAPK13); post-translational modification of mucin (FUT3, FUT6, ST6GAL1, ST8SIA1, CHST6, GALNT5, GALNT6); vesicle transport (MIA3, SURF4, KDELR2, KDELR3, ITSN1, ERGIC1, CKAP4, GOSR1, SYNJ2BP, MPPE1, SEC31A, ARF4, VPS13D, SEC22B and TPD52); ER stress-related (EIF2AK3, EDEM3, XBP1 and CREB3L1); secretory

Table 2 Human small airway epithelium MUC5AC-associated core genes^{1,2}

Mucus components	Mucin producing cell differentiation-related		Post-translational modification of mucin	Vesicle transport	Endoplasmic reticulum stress-associated	Secretory granule-associated	Mucus secretion-related regulators	Mucus hypersecretory-related ion channels
	Transcription factors	Pathways or mediators						
TFF3	SPDEF	HES1	AGR2	MIA3	CREB3L1	SYTL2	PRSS23	SLC12A2
TFF1	FOXA3	TSTA3	GENE	SURF4	EDEM3	RAB3D	PLCE1	CLCA2
	SOX2	LRRFIP2	GALNT4	KDELRL2	XBP1	SCIN	DGKA	SCNN1A
	KLF4	KRAS	GALNT7	KDELRL3	EIF2AK3	STXBP6	ITPR3	GABRP
		MAPK13	GALNT12	ITSN1		RAB27B	PRKCD	
		RPS6KA3	PDIA5	ERGIC1		SYTL4		
		CTSC	FUT3	CKAP4		SYTL5		
		SERPINB4	FUT6	GOSR1		GSN		
		PLA2G4A	ST6GAL1	SYNJ2BP		RIMS1		
			ST8SIA1	MPPE1		CASK		
			CHST6	SEC31A		MYO5B		
			GALNT5	ARF4		MYO5C		
			GALNT6	VPS13D		PCLO		
				SEC22B		PAM		
				TPD52		ATP6V0A4 KIF5B CDC42EP5		

¹ Genes with a role in mucus production or mucus secretion (based on the literature) were selected from the significantly up-regulated genes in nonsmoker high-MUC5AC expressors compared to nonsmoker low-MUC5AC expressors (see Supplemental Methods for complete criteria).

² References for the functions of each gene are in Additional file 1: Table S1. See Additional file 1: Figure S1, S2 for the schematic illustration of the functions of these genes relevant to MUC5AC-production/secretion.

granule-associated (RIMS, PCLO, PAM, ATP6V0A4, KIF5B, CDC42EP5); mucus secretion-related regulators (PRSS23, DGKA, PLCE1); and mucus hypersecretory-related ion channels (CLCA2, SCNN1B and GABAP) (Table 2, Additional file 1: Table S1).

Together, the MUC5AC-core genes suggest that multiple molecular events involving the nucleus, endoplasmic reticulum, Golgi, vesicles and plasma membrane work coordinately in the human small airway epithelium to promote mucus production and secretion in MUC5AC-producing cells.

MUC5AC-core genes in smokers

Chronic cigarette smoking, with its 4000 xenobiotics and $>10^{14}$ oxidants per puff, is a complicated stress to the airway epithelium of smokers [22,27]. Based on this, and the knowledge that smoking induces metaplasia of MUC5AC positive cells in the airway epithelium [1-3,22], and prior literature detailing increased expression of MUC5AC in the airway epithelium of smokers [3,28,29], we used small airway epithelium of chronic smokers as a validation set, i.e., if the human MUC5AC-core genes identified in the nonsmoker small airway epithelium are truly associated with MUC5AC synthesis/secretion, then the MUC5AC-core genes should be up-regulated in the small airway epithelium of smokers. As

expected, healthy smokers had 2.3-fold higher MUC5AC gene expression than nonsmokers (Figure 4A). Like the nonsmokers, there was considerable variability in MUC5AC expression among smokers [MUC5AC coefficient of variation among smokers 47.1%, compared to $25.3 \pm 3.1\%$ for the 5 housekeeping genes, ACTB, GAPDH, B2M, RPLP0 and PPIA (Ansari-Bradley test, all $p < 10^{-11}$ for comparison between MUC5AC and each housekeeping gene)]. Comparison of the expression patterns of MUC5AC-core genes between healthy smokers and healthy nonsmokers demonstrated that 28 of the 73 (38%) MUC5AC-core genes were up-regulated in smokers (Figure 4B, Additional file 1: Table S4). Interestingly, 6 out of the 73 MUC5AC-core genes were down-regulated by smoking, suggesting smoking has a variable effect on some components of the orchestrated MUC5AC-production/secretion system in the small airway epithelium.

To assess these global smoking-related modifications of the MUC5AC core genes, we created an overall index of the MUC5AC-core genes. The index analysis supported the concept that healthy smokers had more active MUC5AC-core gene expression (up- or down-regulation) compared to the nonsmokers ($p < 0.01$, Figure 4C). Clustering analysis comparing the high and low MUC5AC expression in nonsmokers and smokers globally

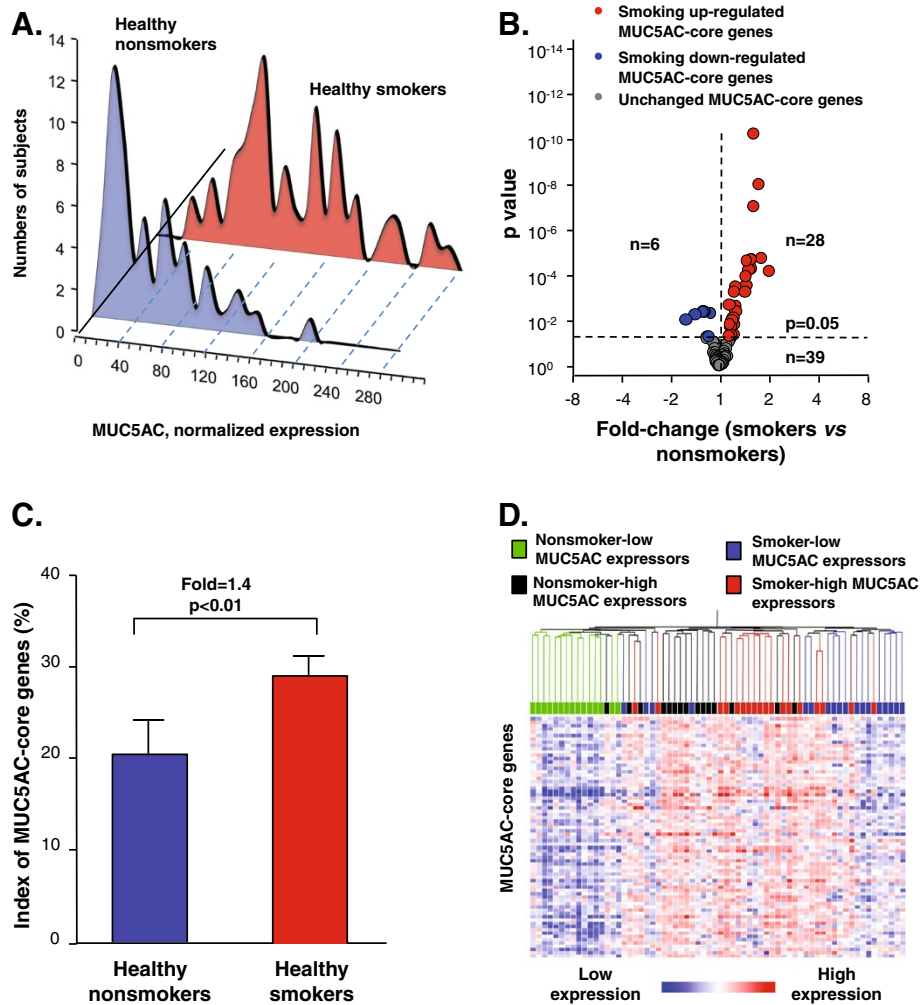


Figure 4 cMUC5AC expression and MUC5AC-core genes in the small airway epithelium of healthy smokers. **A.** MUC5AC gene expression distribution in healthy smokers (blue, n=60) compared to healthy nonsmokers (red, n=72). Y-axis, number of subjects; X-axis, normalized MUC5AC expression. **B.** Volcano plot of MUC5AC-core genes comparing healthy smokers and healthy nonsmokers. Differentially expressed genes between smokers (n=72) and nonsmokers (n=60) were determined by unequal variance Student's *t*-test followed by Benjamini-Hochberg correction ($p < 0.05$). Only MUC5AC-core genes were plotted. y-axis - negative log of *p* value; x-axis - log₂-transformed fold-change; red dots - smoking up-regulated MUC5AC-core genes; blue dots - smoking down-regulated MUC5AC-core genes; grey dots - smoking unchanged MUC5AC-core genes. Numbers of each group are shown. **C.** Index of expression of the MUC5AC-core genes (% of abnormally expressed MUC5AC-core genes beyond that of healthy nonsmoker MUC5AC low-expressors in healthy nonsmokers (blue) compared to healthy smokers (red). Fold-change represents the average index difference between the groups, *p* value indicates significant differences between the groups. **D.** Unsupervised hierarchical clustering analysis of expression of MUC5AC-core genes in nonsmoker-high MUC5AC expressors (black, highest quartile of all healthy non-smokers, n=15), nonsmoker-low MUC5AC expressors (green, lowest quartile of all healthy non-smokers, n=15), smoker-high MUC5AC expressors (red, highest quartile of all healthy smokers, n=18) and smoker-low MUC5AC expressors (blue, lowest quartile of all healthy smokers, n=18). Genes expressed above the average are represented in red, below average in blue, and average in white.

demonstrated that there were 2 major groups among smokers and nonsmokers, based on expression of the MUC5AC-core genes, with nonsmoker-high MUC5AC expressors having a similar expression pattern of MUC5AC-core genes to smoker-high MUC5AC expressors (Figure 4D). Together, these results suggest that most MUC5AC-core genes are shared by smokers and nonsmokers in association with high MUC5AC expression.

To confirm the microarray findings, RNA sequencing was used to assess the gene expression profile from airway epithelium of 5 nonsmokers and 6 smokers. Consistent with microarray data, RNA-Seq data showed that MUC5AC-core genes had a coordinated expression pattern (Figure 5). As with the microarray data, in general, smokers had more active MUC5AC-core gene expression, although similar to the microarray data, there was overlap, with some

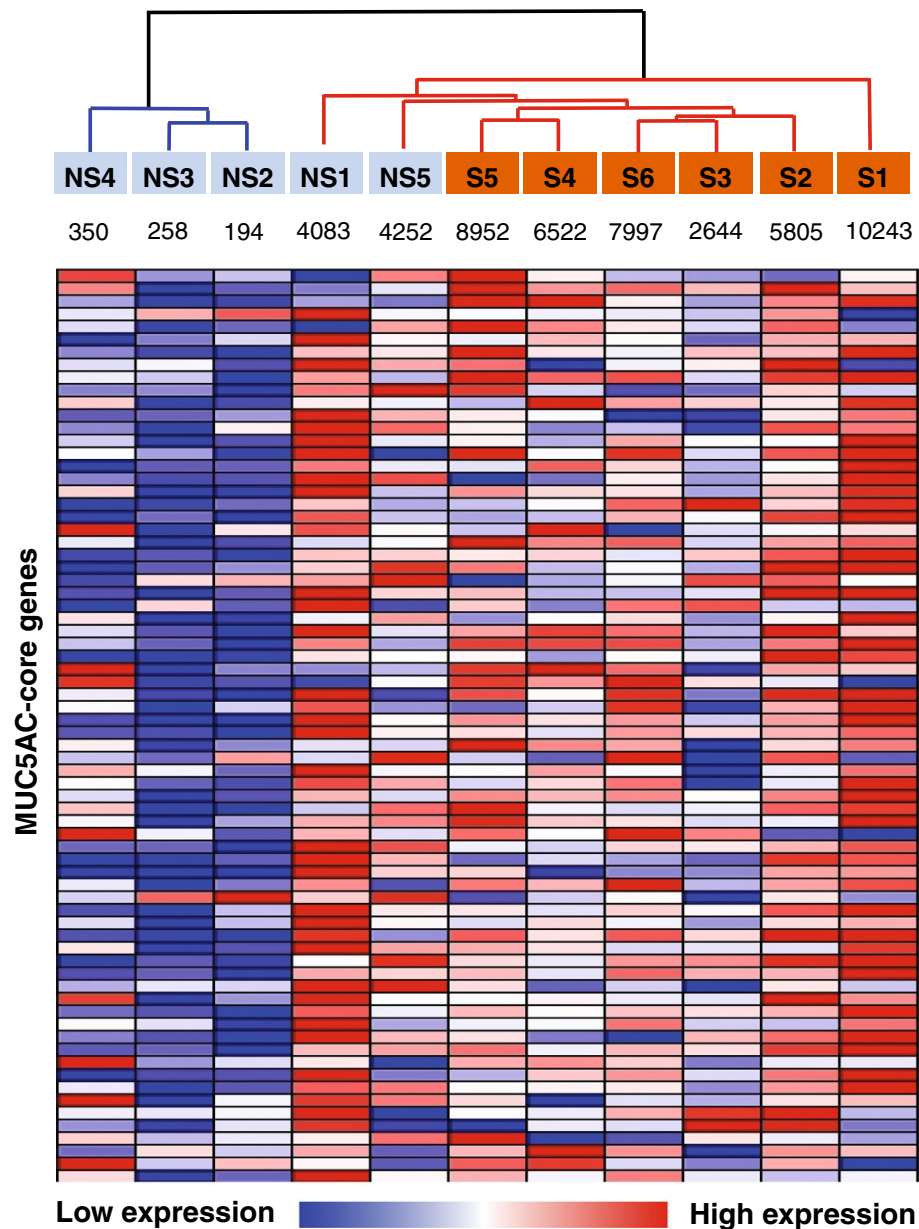


Figure 5 MUC5AC-core genes assessed by RNA-Seq analysis in healthy smokers and healthy nonsmokers. Shown is a hierarchical cluster analysis of RNA-Seq data of expression of the MUC5AC-associated core genes from 5 healthy nonsmokers and 6 healthy smokers. Smoking status was represented by different colors (nonsmoker, blue; smoker, brown). In the heat map, gene expression above the average is represented in red, below average in blue, and average in white. The color schemes represent the relative expression level across each row (each gene). Dendrogram color denotes there are two different groups independent of smoking status. The expression level (RPKM) of MUC5AC is shown under the label of each individual (NS = nonsmoker; S = smoker).

nonsmokers having high levels of MUC5AC-core gene expression.

MUC5AC-core genes in asthma

Asthma is known to be associated with airway mucus overproduction. To assess the effect of asthma on MUC5AC-core gene expression, the data set from a study by Woodruff et al. [30] was used. Comparison of

the expression patterns of MUC5AC-core genes between asthma patients at baseline and healthy nonsmokers demonstrated that 6 of the 73 (8%) MUC5AC-core genes were up-regulated and 1 MUC5AC-core gene was down-regulated in asthma (Additional file 1: Table S6). To assess whether expression of MUC5AC-core genes were correlated with MUC5AC in this asthma study, we performed genome-wide correlation analysis to

MUC5AC gene expression. Interestingly, 10 of 73 MUC5AC-core genes are among the top 300 genes correlated with MUC5AC ($p < 10^{-4}$, Additional file 1: Table S7).

Gene networks linking the MUC5AC-core genes

To help understand the underlying mechanism of the coordinate expression of the MUC5AC-core genes, a gene network was built with the focus on transcriptional control. Based on the knowledge from murine studies or *in vitro* studies that inhibition of the transcription factors Spdef, Sox2 and Klf4 or perturbation of Wnt, Notch and MAPK pathway affects differentiation

of mucus-producing cells or expression of MUC5AC [4-9,31], these genes were used as the backbone of the network (Figure 6). We manually screened each of 73 human small airway MUC5AC-core genes for their transcriptional connections with other MUC5AC-core genes. Only those genes with connections with other MUC5AC-core genes were integrated into the network. These known interactions among MUC5AC-core genes suggest that the findings based on the comparison between high and low MUC5AC expressors were not just by chance. The network was centered on MUC5AC, but, interestingly, involved multiple transcription factors and regulatory pathways. Downstream genes of these

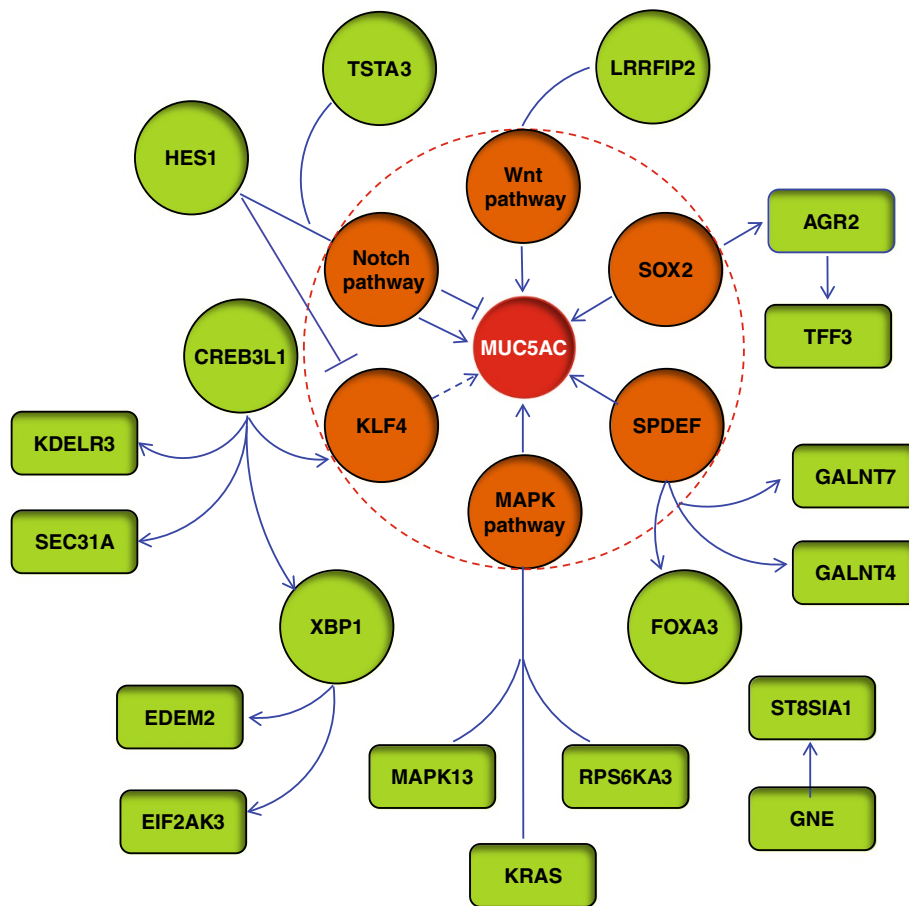


Figure 6 Proposed gene network of the MUC5AC-core genes. From the MUC5AC-core gene list, genes with at least one literature supported connection to other MUC5AC-core genes were selected (see Methods for detail). The connection denotes that the expression level of one gene is affected by the other gene; the data is derived from literature with *in vitro* gene overexpression/activation/knockdown experiments or transgenic mice data (see Additional file 1: Table S5 for references). Protein to protein interactions are not included. Genes in the brown solid circle include, transcription factors (SPDEF, SOX2, KLF4) and the Wnt, Notch, MAPK pathways, which can affect MUC5AC-producing cell differentiation. A red dashed circle was used to emphasize their important roles. Genes in green solid circle represent transcription factors in the MUC5AC-core gene list with literature supported potential role in MUC5AC-producing cells. Genes in green rectangles represent non-transcription factors in the MUC5AC-core gene list. Arrows indicate upstream gene can regulate downstream gene expression. Dashed arrow, based on mouse data [31], KLF4 can affect mucus-producing cell differentiation, but there is no direct evidence showing KLF4 can affect MUC5AC expression. The solid line linked to pathway indicates that MAPK13, KRAS and RPS6KA3 belong to "MAPK pathway"; HES1 and TSTA3 belong to "Notch pathway"; LRRFIP2 belong to "Wnt pathway". The "T" like symbol indicates a potential blocking effect.

transcription factors or pathways are linked to different steps of mucus production, which imply the coordinated expression patterns of MUC5AC-core genes are orchestrated by these transcription factors or pathways. Notably, SPDEF and CREB3L1 seem to be two major “driving forces” in this network. The gene network also provides clues to understand the function of other MUC5AC-associated genes in mucus production.

The EGF receptor family related signaling pathway plays an important role in mucus production [32]. However, our approach to identify MUC5AC-core genes did not identify EGF receptor family genes as MUC5AC-associated genes. Interestingly, overexpression of EGF receptor 2 (ERBB2, HER2) has been demonstrated in primary lung mucinous adenocarcinomas, which were characterized by KRAS activating mutations as well as very high levels of MUC5AC production [33,34]. To investigate the possible relation between ERBB2 and MUC5AC at the transcriptome level, we performed genome-wide correlation analysis for ERBB2 in both healthy nonsmokers and healthy smokers. Although EGF receptor-associated pathways were enriched in ERBB2 correlated genes (Additional file 1: Table S8), no close relation between ERBB2 and mucus production-related genes were found.

Discussion

Mucus hypersecretion is a cardinal feature of chronic airway diseases and mucins are the major component of airway mucus [1-4,17,18,23]. Despite the significant progress in understanding the transcriptional control, biosynthesis and secretion of mucins using murine models, advances in understanding the coordinated gene expression relevant to producing and secreting mucins in the human airway have been limited due to the lack of methods to isolate and culture pure populations of primary human airway mucin-producing cells. Taking advantage of the knowledge that MUC5AC is the major secreted mucin of the small airway epithelium, and our observation that there is variable small airway epithelium MUC5AC expression among healthy nonsmokers, we hypothesized that genome-wide comparison of small airway gene expression of healthy individuals with “high” MUC5AC expression to those with “low” MUC5AC expression would identify genes whose regulation paralleled that of MUC5AC.

From a list of 528 genes up-regulated in healthy high MUC5AC expressors compared to healthy low MUC5AC expressors, we identified 73 MUC5AC core genes that had literature support for a role in mucin production/secretion, from murine studies of the airways, immunohistochemical or *in situ* hybridization studies of human airway epithelium, studies from mucin-secreting cell lines, or studies relating to mucin

production/secretion in organs other than lung. These genes included those categorized as mucus components, mucus-producing cell differentiation-related transcription factors, mucus-producing cell differentiation-related pathways or mediators, post-translational modification of mucin, vesicle transport, ER stress-associated, secretory granule-associated, mucus secretion-related regulators and mucus hypersecretory related ion channels. While some of these 73 human small airway epithelium core MUC5AC genes have been previously associated with human airway MUC5AC expression, for most, this is the first demonstration that, in the human airways, these genes are coordinately up-regulated along with MUC5AC expression. As an independent validation set, we assessed the expression of these 73 core MUC5AC genes in the small airway epithelium of healthy smokers, where there is a stress on the airway epithelium associated with up-regulation of MUC5AC expression [1-4,17,18,22,23]. The data demonstrates that, as in the “high” vs “low” healthy nonsmokers, the healthy smokers up-regulate 38% of the MUC5AC-core genes. Finally, and consistent with studies in knockout and transgenic mice, MUC5AC expression in the human small airway epithelium was linked to several mucus production/secretion-related transcription factors (SOX2, KLF4, SPDEF, XBP1 and CREB3L1), as well as the Notch, Wnt and MAPK pathways, that, in turn, control coordinated expression of many of the MUC5AC-core genes identified in the present study. These findings expand the knowledge regarding molecular events underlying human airway MUC5AC-related production/secretion and provide several potential targets for medical intervention for mucus hypersecretion.

MUC5AC expression in the human small airway epithelium

MUC5AC is one of the major secretory mucins expressed by surface airway epithelial cells, and represents a marker for the airway surface mucus-producing cells [1-4,17,18,22,23]. As demonstrated in the present study and in several other studies [25,28,29], there is up-regulation of MUC5AC expression in the airway epithelium of smokers compared to nonsmokers. This observation is consistent with the knowledge that, on the average, smoking is associated with increased airway mucus production [1-4,17,18,22,23]. However, novel to the present study is the observation that MUC5AC expression in the small airway epithelium, assessed at both the mRNA and protein levels, is highly variable among healthy nonsmokers, implying that there are some individuals more prone to airway mucus production than others. We capitalized on this observation to categorize healthy nonsmokers into “high” and “low” MUC5AC expressors. Then, using genome wide

assessment of gene expression, we compared “high” vs “low” MUC5AC expressors to identify genes, the expression of which correlates with MUC5AC expression. Using the literature to identify genes already known to be associated with the airway epithelium (mostly from murine studies), as well as to identify genes associated with mucin expression in other organs, we were able to define “MUC5AC-core” genes i.e., the genes plausibly linked to MUC5AC gene expression in the human small airway epithelium. To test the validity of the MUC5AC-core genes, we hypothesized that many of these genes would be up-regulated further in the small airway epithelium of smokers, since smokers have, on the average, up-regulation of MUC5AC. This proved to be correct.

Mucus-producing cell differentiation-related transcription factors, pathways and mediators

Studies of MUC5AC expression in experimental non-human models have identified several transcription factors that modulate mucus-producing cell differentiation, including Stat6, Spdef, Klf4, Sox2 and Foxa2 [4,5,20,31,35]. In addition to these transcription factors, murine studies have identified several signaling pathways, including Wnt [6], Notch [7,8] and MAPK [9] pathways, which have been implicated in differentiation of MUC5AC-producing cells.

Consistent with these observations, the MUC5AC-core genes linked to MUC5AC expression in the human small airway epithelium in the present study include the mucus-producing cell differentiation-related transcription factors, SPDEF, KLF4, SOX2 and FOXA3. Up-regulation of FOXA3 and SPDEF have been reported during mucus overproduction in human airways cells *in vitro* and *in vivo* [4]. SPDEF can also drive differentiation of MUC5AC-producing cells in primary human airway epithelial cells *in vitro* [4]. The demonstration that SOX2 and KLF4 are linked to MUC5AC in the human airway epithelium is novel to the present study. The MUC5AC-core genes also included the Wnt pathway gene LRRFIP2, Notch pathway genes HES1 and TSTA3 [19], and the MAPK pathways genes MAPK13, KRAS [36], and RPS6KA3 (RSK2) [37] as linked to MUC5AC expression in the human small airway epithelium. Together, this suggests that a gene network, instead of single gene or pathway, controls differentiation of MUC5AC-producing cells.

The EGF receptor family mediated signaling pathway is involved in mucus production [32]. One of the EGF receptor family members, ERBB2, has been shown to be overexpressed in lung mucinous adenocarcinomas [33,34]. However, no correlation between ERBB2 and MUC5AC at transcription level was found in the current study using normal airway epithelium. These results

suggest that MUC5AC expression is not dependent on the expression level of EGF receptors in the normal airway epithelium. Instead, activation status of EGF receptors might be more important, while in mucinous adenocarcinomas, the constitutive activation of the ERBB2 pathway, which is caused by persistent up-regulation of ERBB2 and ERBB2 mutations, might contribute to the MUC5AC overexpression.

In addition to these transcription factors and pathways, it is known that several enzyme-related genes can mediate mucus-producing cell differentiation [38-40]. Consistent with these observations, we found that SERPINB4 (serpin peptidase inhibitor), CTSC (peptidase) and PLA2G4A (phospholipase) are up-regulated in high MUC5AC expressors. Serpinb3a (SERPINB4 homolog) deletion markedly attenuates mucus-producing cell hyperplasia in mouse airway after allergen challenge [38]; ozone and Staphylococcal enterotoxin B exposure is associated with CTSC-driven mucus production in the mouse [39]; and PLA2G4A activation is associated with mucus overproduction in murine airway [40]. Both SERPINB4 and CTSC are up-regulated in IL13 treated human primary bronchial epithelial cells on air liquid interface culture [16], and PLA2G4A is more active in bronchial explants from individuals with cystic fibrosis [40]. It is a novel observation that these genes are linked to small airway MUC5AC production/secretion in healthy humans.

MUC5AC biosynthesis

MUC5AC is a polymeric mucin which is heavily glycosylated and thus the ER and Golgi apparatus play a central role in biosynthesis of MUC5AC. To form the mature MUC5AC protein, a stepwise posttranslational modification is needed, including N-glycosylation, dimer formation, O-glycosylation, fucose/sialic acid/sulfate modification and polymer formation, processes involving several specific enzymes [2,3]. Among these, several were identified among the MUC5AC core genes.

Consistent with the findings from mouse or human *in vitro* studies, we found that AGR2 (protein disulfide isomerases) [41], GALNT4 and GALNT7 (GalNAc-T members) [4] and GNE (required for normal sialylation) [42] were up-regulated in high-MUC5AC expressors. Besides these genes, we found some novel enzyme genes that are likely involved in human MUC5AC posttranslational modifications, including PDIA5 (disulfide isomerase), GALNT5, GALNT6 and GALNT12 (GalNAc-T members), FUT3 and FUT6 (fucosyltransferase), ST6GAL1 and ST8SIA1 (sialic acid transferase). Interestingly, PDIA5 is also up-regulated in Barrett's esophagus, a disorder in which mucus-producing cells are a basic pathologic change [43].

Along the secretory pathway, vesicle-mediated transportation is important to carry newly synthesized

proteins from the ER to the Golgi [17]. To maintain the function of the ER, reverse transportation is also active, which balances the loss of protein embedded in the vesicles from the ER [44]. Consistent with this concept, we found 15 vesicle transport related genes associated with high MUC5AC expression, including MIA3, SURF4, KDELR2, KDELR3, ITSN1, ERGIC1, CKAP4, GOSR1, SYNJ2BP, MPPE1, SEC31A, ARF4, VPS13D, SEC22B, and TPD52. None have been associated with mucus production in the literature.

Mucin biosynthesis places a large metabolic burden on the cell, with potential ER and Golgi apparatus stress [45]. In agreement with the concept that mucus deregulation might induce ER stress, EIF2AK3 and XBP1 (classical unfolded protein response genes), EDEM3 (endoplasmic reticulum-associated degradation gene), and CREB3L1 (non-canonical unfolded protein response gene) were up-regulated in high-MUC5AC expressors. EIF2AK3 can repress global protein synthesis [46], XBP1 can induce chaperone synthesis [47], and EDEM3 can promote the degradation of poor-quality proteins [48]. Finally, the *Drosophila* ortholog of CREB3L1, CrebA, has been shown to be the major and direct regulator of secretory capacity in the salivary gland [49]. Our findings raise the possibility that the human airway surface MUC5AC-producing cells likely use an ER stress-associated signal as feedback to finely control mucus production.

MUC5AC secretion

MUC5AC secretion is highly regulated [17]. As shown in the “secretory granule-associated” and “mucus secretion-related regulator” categories, activation of the mechanisms underlying mucin secretion was reflected in the function of MUC5AC-core genes. For example, we found PRSS23, PLCE, PRKCD, DKGA, ITPR3, RAB3D, RAB27B, MYO5B, MYO5C, SYTL2, SYTL4, SYTL5, SCIN, CDC42EP5, ATP6V0A4, STXBP6, PAM, RIMS1, PLCO and CASK were up-regulated in high-MUC5AC expressors. PRSS23 is a serine protease, which can stimulate mucus glycoprotein release from hamster tracheal ring organ culture [50]. PLCE is a PLC family member that might be involved in generating the intracellular second messengers, IP3 and DAG, in mucus-producing cells [1]. ITPR3 is the receptor of IP3 [1]. PRKCD is a protein kinase C family member. RAB3D and RAB27B belong to RAB3 and RAB27 family, respectively [17]. ATP6V0A4 and PAM are associated with maturation of secretion granules [51]. Both MYO5B and MYO5C are type V myosins, and SYTL2 (SLP2A) [52], SYTL4 and SYTL5 are granulophilins [17]. KIF5B is associated with granule movement [53]. SCIN (scinderin) is an enzyme that is needed for actin disruption [17] and gelsolin can affect actin remodeling [54].

RIMS1 (RIM), PLCO (Piccolo), and CASK can help secretory granule tethering and docking [17]. STXBP6 belongs to Munc18 family. Among this group of MUC5AC-core genes, SYTL4, SYTL5, RIMS1, PLCO, CASK, KIF5B, CDC42EP5, ATP6V0A4, STXBP6, PAM, DKGA, and PLCE have not been previously identified in association with mucus-producing cells.

Mucus hypersecretion-related ion channels

Several studies have suggested that ion channels are involved in mucus production, including CFTR [55], SLC26A4 (pendrin) [56], SCN1B (ENaC subunit) [21], CLCA family [57], GABAergic system receptors [58] and SLC12A2 [59]. In the current study, we found SCN1A (ENaC alpha subunit), CLCA2 (CLCA family member), GABAP (a subunit of the GABAergic system receptor), SLC12A2 (Slc12a2 homolog) were up-regulated in high MUC5AC expressors. Among them, SLC12A2 expression is known to be restricted to human airway mucus-producing cells [59]. The links between SCN1A, CLCA2, GABAP and MUC5AC expression in human airway are novel. Since multiple ion channel genes are found to be associated with MUC5AC expression, these observations suggest an active water/ion exchange in the airway epithelial cell membrane during human airway mucus production.

Several limitations of this study should be pointed out. First, all MUC5AC-core genes were identified based on the literature, and thus, important genes might be ignored because we are lacking clues from current knowledge. Second, our strategy was based on the transcriptome level, so regulation mechanisms depending on protein level modulation are underestimated.

Methods

Nomenclature

The terminology associated with mucus-producing cells in the airway epithelium varies in the literature, including the terms “goblet cells” and “mucous cells” [1,23]. Since we are focusing on genes correlating with the expression of MUC5AC, to avoid confusion, we use the term “MUC5AC-producing cells” as the surface airway epithelium cells (the cells that are sampled with bronchoscopy and brushing) expressing MUC5AC. When referring to the literature, if the mucus-related gene/protein is not clear, the term “mucus-producing cells” will be used.

Ethics statement

All individuals were evaluated at the Weill Cornell NIH Clinical and Translational Science Center and Department of Genetic Medicine Clinical Research Facility, using Institutional Review Board-approved clinical protocols.

Study population

Healthy nonsmokers and healthy smokers were recruited from the general population in New York City. The nonsmokers were used as the “test” set to identify the MUC5AC-core genes; the smokers were used as a “validation” set to determine if the MUC5AC-core genes correlate with MUC5AC gene expression under conditions (smoking) where MUC5AC gene expression is known to be up-regulated.

The criteria for “healthy” was based on a history, physical exam, complete blood count, coagulation studies, liver function tests, urine studies, chest X-ray, EKG and pulmonary function tests (Table 3; see Supplemental Methods for inclusion/exclusion criteria). All subjects were negative for HIV1 and had normal α 1-antitrypsin levels. Urine nicotine and cotinine levels were measured using liquid chromatography-tandem mass spectrometry (ARUP

laboratories, Salt Lake City, UT). “Nonsmokers” (n = 60) were defined as self-reported life-long nonsmokers, with non-detectable urine nicotine (<2 ng/ml) and cotinine (<5 ng/ml); “Smokers” (n = 72) were defined as self-reported current smokers with urine nicotine >2 ng/ml and/or urine cotinine >5 ng/ml. No individuals smoked within 12 hr prior to the bronchoscopy procedure to collect the airway epithelium.

Epithelial sampling, cDNA preparation and microarray processing

Airway epithelium (10th to 12th order) was collected by fiberoptic bronchoscopy by brushing as previously described [14]. The expression of genes encoding surfactant and Clara cell secretory proteins confirmed the samples were derived from the small airway epithelium [14]. Cytopreparations were prepared by centrifugation (Cytospin 11, Shandon Instruments, Pittsburgh, PA) and stained with Diff-Quik (Dade Behring, Newark, NJ). For MUC5AC staining in small airway epithelium, cytopspin slides of brushed cells were stained with mouse anti-human MUC5AC antibody (Vector, Burlingame, CA) and detected by Cy3 labeled goat anti-mouse antibody (Jackson, West Grove, PA). Nuclei were counterstained with DAPI (Invitrogen, Carlsbad, CA).

RNA in the airway epithelium samples was processed for microarray analysis as previously described [14,60]. Total RNA was extracted from the small airway epithelium using a modification of the TRIzol method (Invitrogen, Carlsbad, CA) and processed to generate cDNA. Genome-wide gene expression analysis was performed using HG-U133 Plus 2.0 array (Affymetrix, Santa Clara, CA) according to Affymetrix protocols, hardware and software. Overall microarray quality was verified by the criteria: (1) 3'/5' ratio for GAPDH \leq 3; and (2) scaling factor \leq 10.0 [61]. The captured image data from the HG-U133 Plus 2.0 arrays was processed using the MAS5 algorithm. The raw data are available at the Gene Expression Omnibus (GEO) site (<http://www.ncbi.nlm.nih.gov/geo/>); accession number for this dataset is GSE34450.

Mucin gene expression in the small airway epithelium

The mucin genes represented on the HG-U133 Plus 2.0 array were grouped as: secretion-nonpolymeric; secretion-polymeric; tethered-shed in mucus; and tethered [2,3]. For genes with multiple probe sets, the one with highest expression level in nonsmokers was chosen. The percentage of Affymetrix “P call” was calculated for each probeset. To compare expression levels among mucin genes, Genespring data normalized “per chip” (normalized to the median of all the genes on the same chip) was used.

To evaluate the expression variability of MUC5AC, endogenous controls were used as references. Five endogenous control genes were randomly selected from

Table 3 Study population and biologic samples¹

Parameter	Healthy nonsmokers	Healthy smokers
n	60	72
Gender (male/female)	38/22	51/21
Age	41 \pm 12	43 \pm 8
Race (B/W/O) ²	27/23/10	45/16/11
Smoking history (pack-yr)	0	27 \pm 16
Urine nicotine (ng/ml)	negative	1098 \pm 1312
Urine cotinine (ng/ml)	negative	1228 \pm 939
Pulmonary function parameters ⁴		
FVC (% predicted)	106 \pm 13	110 \pm 14
FEV1 (% predicted)	106 \pm 14	110 \pm 15
FEV1/FVC (% observed)	82 \pm 6	81 \pm 4
TLC (% predicted)	99 \pm 13	100 \pm 12
DLCO (% predicted)	98 \pm 14	94 \pm 11
Epithelial cells ⁴		
Number recovered x10 ⁶	6.5 \pm 2.9	7.4 \pm 3.3
% epithelial cells ⁵	99.1 \pm 1.2	99.0 \pm 1.3
% inflammatory cells	0.9 \pm 1.2	1.0 \pm 1.3
Differential cell count ⁶		
Ciliated (%)	72.0 \pm 8.9	64.0 \pm 12
Secretory (%)	6.7 \pm 3.7	8.4 \pm 4.4
Basal (%)	12.6 \pm 6.6	14.5 \pm 8.0
Intermediate (%)	7.8 \pm 3.4	12.2 \pm 6.8

¹ Data are presented as mean \pm standard deviation.

² B = Black, W = White, O = Other.

³ Pulmonary function testing parameters are given as % of predicted value with the exception of FEV1/FVC, which is reported as % observed; FVC - forced vital capacity; FEV1 - forced expiratory volume in 1 sec; TLC - total lung capacity; DLCO - diffusing capacity.

⁴ To quantify the percentage of epithelial and inflammatory cells and the proportions of ciliated, basal, secretory, and intermediate epithelial cells, aliquots of 2x10⁴ cells were prepared by centrifugation and stained with DiffQuik.

⁵ Of total cells recovered.

⁶ Of total epithelial cells recovered.

literature (ACTB, GAPDH, B2M, RPLP0 and PPIA) [62] with “per chip” normalized gene expression level >99% genes (probesets) on the HG-U133 Plus 2.0 array. Coefficient of variation (CV) of expression of MUC5AC and the control genes were calculated. For control genes with multiple probesets, the one with the highest CV was chosen. Genespring “per chip” and “per gene” (normalized to the median of the same gene across all samples) normalizations were used. Comparison of the variance of MUC5AC vs control gene expression (“per chip” and “per gene” normalized data) was done using the Ansari-Bradley test in R language (<http://www.r-project.org/>).

Identification of MUC5AC-associated core genes

To identify the MUC5AC-associated core genes, MUC5AC expression of the healthy nonsmokers was quantified using probeset 214385_s_at, the probeset generating the strongest signal among the 3 of MUC5AC probesets on the microarray. Healthy nonsmokers were categorized as high MUC5AC expressors (MUC5AC expression highest quartile, $n=15$), medium MUC5AC expressors (the 2nd and 3rd quartiles; $n=15$ and $n=15$, respectively) and low MUC5AC expressors (MUC5AC expression lowest quartile, $n=15$). Genome-wide analysis of the genes associated with high and low MUC5AC expressors were filtered by the criteria: (1) Affymetrix “P call” >80% in nonsmokers; and (2) genes with known annotations. Significant differences of gene expression between high MUC5AC expressors and low MUC5AC expressors were determined by an unequal variance Student’s *t*-test followed by Benjamini-Hochberg correction (B-H correction, $p < 0.05$). The probe sets of MUC5AC were not included. For genes represented by more than one probeset, the probeset with the lowest *p* value was used. The Database for Annotation, Visualization and Integrated Discovery (DAVID) was used to do gene function annotation term enrichment analysis of the significant up-regulated genes [63]. The role of each up-regulated gene in mucus production and secretion was assessed by manual literature mining. Genes with literature supporting roles in mucus production or secretion were defined as the “MUC5AC-core genes” (see Supplemental methods for complete criteria).

Effects of smoking on expression of MUC5AC and MUC5AC-core genes

As a validation cohort to assess the role of smoking on MUC5AC expression and the MUC5AC-core genes, MUC5AC expression (probeset 214385_s_at) was compared in the small airway epithelium of healthy nonsmokers ($n=60$) and healthy smokers ($n=72$). The comparison of variability of MUC5AC

to the 5 housekeeping genes in smokers was analyzed with the Ansari-Bradley test. To determine if the MUC5AC-core genes identified in nonsmokers were also up-regulated in smokers, genome-wide transcriptome comparison between healthy smokers ($n=72$) and healthy nonsmokers ($n=60$) was based on the filtered probesets (using same filter criteria used to identify the MUC5AC-associated core genes). Significant differences of gene expression between smokers and nonsmokers were determined by an unequal variance Student’s *t*-test followed by Benjamini-Hochberg correction ($p < 0.05$). The results of MUC5AC-core genes (fold-change and Benjamini-Hochberg corrected *p* value) were then extracted.

Cluster analysis was performed on microarray data (“per chip” and “per gene” normalization) of the MUC5AC-core genes in nonsmokers and smokers using settings with Spearman correlation as similarity measure and the average linkage clustering algorithm. To reduce the variability associated with medium MUC5AC expressors, only nonsmoker-high MUC5AC expressors (highest quartile of all healthy nonsmokers, $n=15$), nonsmoker-low MUC5AC expressors (lowest quartile of all healthy nonsmokers, $n=15$), smoker-high MUC5AC expressors (highest quartile of all healthy smokers, $n=18$) and smoker-low MUC5AC expressors (lowest quartile of all healthy smokers, $n=18$) were used.

Asthma study

GEO4302 datasets [30] were used to evaluate the effect of asthma on MUC5AC-core genes, using data from 42 asthmatics and 28 healthy subjects in this data set. Microarray data were processed by RMA algorithm in Genespring 7.0. Significant differences of gene expression between asthma and healthy subjects were determined by an unequal variance Student’s *t*-test followed by Benjamini-Hochberg correction ($p < 0.05$). Genome-wide correlations to MUC5AC (214385_s_at) were assessed by Spearman’s rank correlation coefficient.

MUC5AC-core gene index

A “MUC5AC-core gene index” was created as a global measure of the coordinate MUC5AC-enhanced expression-associated changes by integrating information on expression levels of all MUC5AC-core genes using the conceptual framework described by Tilley et al. [64]. For each gene, the mean and standard deviation were calculated from the values in nonsmoker-low MUC5AC expressors. The “normal range” was defined as within 2 standard deviations (SD) of the mean. The MUC5AC-core gene index for each nonsmoker and smoker was defined as a percentage of MUC5AC-core genes with

expression levels outside the normal range, using the formula:

$$\text{MUC5AC}(\%) = \sum_{n=1}^n cEn$$

where E1 has a value of 1 if the expression level for gene 1 was >2 SD beyond that of nonsmoker-low MUC5AC expressors or a value of 0 if the expression level was ≤2 SD below that of nonsmoker-low MUC5AC expressors; E2 is the index for gene 2, etc., and the constant c [c = 100 divided by the number of MUC5AC-core genes (n) normalizes the index to the percent of MUC5AC-core genes that are outside of the range of healthy nonsmoker-low expressors]. The statistical significance of differences in index between the groups was determined using the Mann Whitney *U*-test.

RNA-seq data analysis

Analysis of the MUC5AC-core genes was also carried out using a database of massive parallel sequencing (RNA-Seq) of the transcriptome of healthy nonsmokers (n = 5) and healthy smokers (n = 6) [65]. Because there is no MUC5AC sequence in the reference genome build UCSC hg19, the resultant reads were aligned to *Homo sapiens* high coverage assembly GRCh37 using Bowtie v 0.12. Reads per kilobase of exon model per million mapped reads (RPKM) was used to quantify transcript levels of each gene. The sequence read data have been submitted to the NCBI SRA database (SRA accession #SRP005411). Of the 73 MUC5AC-core genes, 72 were used for clustering analysis (GALNT4 does not have an ensembl sequence). Log2 transformed gene level RNA-Seq data was used for hierarchical clustering analysis, which was performed in “HierarchicalClustering” of Genepattern (<http://genepattern.broadinstitute.org>) with the default setting (Distance measure: Pearson correlation; Clustering method: pairwise complete-linkage). To visualize the heatmap, “HierarchicalClusteringViewer” of Genepattern was used with the default setting (View Options are “relative color schemes” and “gradient color”).

Gene network analysis

A gene network analysis was built on the human small airway epithelium MUC5AC-core genes by manually data mining the literature. The network focused on transcription control, using transcription factors/pathways known to be associated with mucin biosynthesis (SPDEF, SOX2 and KLF4 and Wnt, Notch and MAPK pathways) as the network backbone. MUC5AC-core genes with transcriptional/pathway level connections with other MUC5AC-core genes were selected to expand the network. The interconnections among genes were based on

literature with gene overexpression/activation/knock-down experiments *in vitro* or data from genetically engineered mice showing that the expression level of one gene is affected by another gene. Protein-protein interactions were not included (See Additional file 1: Table S5 for details).

Additional file

Additional file 1: Inclusion and Exclusion Criteria for Healthy Nonsmokers and Healthy Smokers [1,2,4-6,8,9,16,17,19,21,30,31, 36-43,46-54,57-59,66-76].

Competing interests

The authors have no competing interests.

Authors' contributions

GW, ZX, RW and MAH conceptualized and designed the study. YSB, AET and RGC oversaw patient recruitment, sample acquisition and experimental protocols. JGM contributed to the design of the analytic strategy. JS and GW performed the statistical and computational analyses. GW and RW interpreted the results, and wrote the manuscript. RGC supervised the analyses and edited the manuscript. All authors read and approved the final manuscript.

Acknowledgments

We thank B. Ferris and B. Witover for help with immunohistochemistry analysis; N. Mohamed and D.N. McCarthy for help in preparing this manuscript. These studies were supported, in part, by P50 HL084936 and UL1-RR024996. GW and RW are supported, in part, by T32 HL094284. AET is supported, in part, by 1 K23 HL103837.

Author details

¹Department of Genetic Medicine, Weill Cornell Medical College, New York, NY, USA. ²Chengdu Second People's Hospital, Chengdu, China. ³Department of Medicine, Division of Pulmonary and Critical Care Medicine, Weill Cornell Medical College, New York, NY, USA. ⁴Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY, USA. ⁵Department of Genetic Medicine, Weill Cornell Medical College, 1300 York Avenue, Box 9610065, New York, NY, USA.

Received: 16 February 2012 Accepted: 1 May 2012

Published: 7 June 2012

References

1. Fahy JV, Dickey BF: Airway mucus function and dysfunction. *N Engl J Med* 2010, **363**:2233–2247.
2. Thornton DJ, Rousseau K, McGuckin MA: Structure and function of the polymeric mucins in airways mucus. *Annu Rev Physiol* 2008, **70**:459–486.
3. Rose MC, Voynow JA: Respiratory tract mucin genes and mucin glycoproteins in health and disease. *Physiol Rev* 2006, **86**:245–278.
4. Chen G, Korfhagen TR, Xu Y, Kitzmiller J, Wert SE, Maeda Y, et al: SPDEF is required for mouse pulmonary goblet cell differentiation and regulates a network of genes associated with mucus production. *J Clin Invest* 2009, **119**:2914–2924.
5. Tompkins DH, Besnard V, Lange AW, Wert SE, Keiser AR, Smith AN, et al: Sox2 is required for maintenance and differentiation of bronchiolar Clara, ciliated, and goblet cells. *PLoS One* 2009, **4**:e8248.
6. Mucenski ML, Nation JM, Thitoff AR, Besnard V, Xu Y, Wert SE, et al: Beta-catenin regulates differentiation of respiratory epithelial cells in vivo. *Am J Physiol Lung Cell Mol Physiol* 2005, **289**:L971–L979.
7. Tsao PN, Vasconcelos M, Izvolsky KI, Qian J, Lu J, Cardoso WV: Notch signaling controls the balance of ciliated and secretory cell fates in developing airways. *Development* 2009, **136**:2297–2307.
8. Guseh JS, Bores SA, Stanger BZ, Zhou Q, Anderson WJ, Melton DA, et al: Notch signaling promotes airway mucous metaplasia and inhibits alveolar development. *Development* 2009, **136**:1751–1759.

9. Fujisawa T, Ide K, Holtzman MJ, Suda T, Suzuki K, Kuroishi S, et al: **Involvement of the p38 MAPK pathway in IL-13-induced mucous cell metaplasia in mouse tracheal epithelial cells.** *Respirology* 2008, **13**:191–202.
10. Strulovici-Barel Y, Omberg L, O'Mahony M, Gordon C, Hollmann C, Tilley AE, et al: **Threshold of biologic responses of the small airway epithelium to low levels of tobacco smoke.** *Am J Respir Crit Care Med* 2010, **182**:1524–1532.
11. Cosio M, Ghezzi H, Hogg JC, Corbin R, Loveland M, Dosman J, et al: **The relations between structural changes in small airways and pulmonary-function tests.** *N Engl J Med* 1978, **298**:1277–1281.
12. Cosio MG, Hale KA, Niewoehner DE: **Morphologic and morphometric effects of prolonged cigarette smoking on the small airways.** *Am Rev Respir Dis* 1980, **122**:265–271.
13. Hogg JC, Chu F, Utokaparch S, Woods R, Elliott WM, Buzatu L, et al: **The nature of small-airway obstruction in chronic obstructive pulmonary disease.** *N Engl J Med* 2004, **350**:2645–2653.
14. Harvey BG, Heguy A, Leopold PL, Carolan BJ, Ferris B, Crystal RG: **Modification of gene expression of the small airway epithelium in response to cigarette smoking.** *J Mol Med (Berl)* 2007, **85**:39–53.
15. Barnes PJ: **Small airways in COPD.** *N Engl J Med* 2004, **350**:2635–2637.
16. Zhen G, Park SW, Nguyenvu LT, Rodriguez MW, Barbeau R, Paquet AC, et al: **IL-13 and epidermal growth factor receptor have critical but distinct roles in epithelial cell mucin production.** *Am J Respir Cell Mol Biol* 2007, **36**:244–253.
17. Davis CW, Dickey BF: **Regulated airway goblet cell mucin secretion.** *Annu Rev Physiol* 2008, **70**:487–512.
18. Thai P, Loukoianov A, Wachi S, Wu R: **Regulation of airway mucin gene expression.** *Annu Rev Physiol* 2008, **70**:405–429.
19. Waterhouse CC, Johnson S, Phillipson M, Zbytnuik L, Petri B, Kelly M, et al: **Secretory cell hyperplasia and defects in Notch activity in a mouse model of leukocyte adhesion deficiency type II.** *Gastroenterology* 2010, **138**:1079–1090.
20. Blease K, Schuh JM, Jakubzick C, Lukacs NW, Kunkel SL, Joshi BH, et al: **Stat6-deficient mice develop airway hyperresponsiveness and peribronchial fibrosis during chronic fungal asthma.** *Am J Pathol* 2002, **160**:481–490.
21. Mall M, Grubb BR, Harkema JR, O'Neal WK, Boucher RC: **Increased airway epithelial Na⁺ absorption produces cystic fibrosis-like lung disease in mice.** *Nat Med* 2004, **10**:487–493.
22. Yoshida T, Tuder RM: **Pathobiology of cigarette smoke-induced chronic obstructive pulmonary disease.** *Physiol Rev* 2007, **87**:1047–1082.
23. Curran DR, Cohn L: **Advances in mucous cell metaplasia: a plug for mucus as a therapeutic focus in chronic airway disease.** *Am J Respir Cell Mol Biol* 2010, **42**:268–275.
24. Lumsden AB, McLean A, Lamb D: **Goblet and Clara cells of human distal airways: evidence for smoking induced changes in their numbers.** *Thorax* 1984, **39**:844–849.
25. O'Donnell RA, Richter A, Ward J, Angco G, Mehta A, Rousseau K, et al: **Expression of ErbB receptors and mucins in the airways of long term current smokers.** *Thorax* 2004, **59**:1032–1040.
26. Lodish H, Berk A, Zipursky SL: **Overview of the secretory pathway.** In *Molecular Cell Biology*. Edited by. New York: W.H. Freeman; 2000.
27. Church DF, Pryor WA: **Free-radical chemistry of cigarette smoke and its toxicological implications.** *Environ Health Perspect* 1985, **64**:111–126.
28. Pierrou S, Broberg P, O'Donnell RA, Pawlowski K, Virtala R, Lindqvist E, et al: **Expression of genes involved in oxidative stress responses in airway epithelial cells of smokers with chronic obstructive pulmonary disease.** *Am J Respir Crit Care Med* 2007, **175**:577–586.
29. Chari R, Lonergan KM, Ng RT, Macaulay C, Lam WL, Lam S: **Effect of active smoking on the human bronchial epithelium transcriptome.** *BMC Genomics* 2007, **8**:297.
30. Woodruff PG, Boushey HA, Dolganov GM, Barker CS, Yang YH, Donnelly S, et al: **Genome-wide profiling identifies epithelial cell genes associated with asthma and with treatment response to corticosteroids.** *Proc Natl Acad Sci USA* 2007, **104**:15858–15863.
31. Katz JP, Perreault N, Goldstein BG, Lee CS, Labosky PA, Yang WW, et al: **The zinc-finger transcription factor Klf4 is required for terminal differentiation of goblet cells in the colon.** *Development* 2002, **129**:2619–2628.
32. Takeyama K, Fahy JV, Nadel JA: **Relationship of epidermal growth factor receptors to goblet cell production in human bronchi.** *Am J Respir Crit Care Med* 2001, **163**:511–516.
33. Zhang Y, Sun Y, Pan Y, Li C, Shen L, Li Y, et al: **Frequency of driver mutations in lung adenocarcinoma from female never-smokers varies with histologic subtypes and age at diagnosis.** *Clin Cancer Res* 2012, **18**:1947–1953.
34. Rossi G, Gasser B, Sartori G, Migaldi M, Costantini M, Mengoli MC, et al: **MUC5AC, cytokeratin 20 and HER2 expression and K-RAS mutations within mucinogenic growth in congenital pulmonary airway malformations.** *Histopathology* 2012, .
35. Park SW, Verhaeghe C, Nguyenvu LT, Barbeau R, Easley CJ, Nakagami Y, et al: **Distinct roles of FOXA2 and FOXA3 in allergic airway disease and asthma.** *Am J Respir Crit Care Med* 2009, **180**:603–610.
36. Marchetti A, Buttitta F, Pellegrini S, Chella A, Bertacca G, Filardo A, et al: **Bronchioloalveolar lung carcinomas: K-ras mutations are constant events in the mucinous subtype.** *J Pathol* 1996, **179**:254–259.
37. Kim SW, Hong JS, Ryu SH, Chung WC, Yoon JH, Koo JS: **Regulation of mucin gene expression by CREB via a nonclassical retinoic acid signaling pathway.** *Mol Cell Biol* 2007, **27**:6933–6947.
38. Sivaprasad U, Askew DJ, Ericksen MB, Gibson AM, Stier MT, Brandt EB, et al: **A nonredundant role for mouse Serpinb3a in the induction of mucus production in asthma.** *J Allergy Clin Immunol* 2011, **127**:254–261.
39. Foley JP, Bolognese BS, Long E, Salmon M, Podolin PL: **Ozone and staphylococcal enterotoxin B exposure induces emphysema and mucus production that is dependent upon cathepsin C expression [Abstract].** *Am J Respir Crit Care Med* 2009, .
40. Dif F, Wu YZ, Burgel PR, Ollero M, Leduc D, Aarbiou J, et al: **Critical role of cytosolic phospholipase A2[alpha] in bronchial mucus hypersecretion in CFTR-deficient mice.** *Eur Respir J* 2010, **36**:1120–1130.
41. Park SW, Zhen G, Verhaeghe C, Nakagami Y, Nguyenvu LT, Barczak AJ, et al: **The protein disulfide isomerase AGR2 is essential for production of intestinal mucus.** *Proc Natl Acad Sci USA* 2009, **106**:6950–6955.
42. Gagiannis D, Orthmann A, Danssmann I, Schwarzkopf M, Weidemann W, Horstkorte R: **Reduced sialylation status in UDP-N-acetylglucosamine-2-epimerase/N-acetylmannosamine kinase (GNE)-deficient mice.** *Glycoconj J* 2007, **24**:125–130.
43. Nancarrow DJ, Clouston AD, Smithers BM, Gotley DC, Drew PA, Watson DI, et al: **Whole genome expression array profiling highlights differences in mucosal defense genes in Barrett's esophagus and esophageal adenocarcinoma.** *PLoS One* 2011, **6**:e22513.
44. Lee MC, Miller EA, Goldberg J, Orci L, Schekman R: **Bi-directional protein transport between the ER and Golgi.** *Annu Rev Cell Dev Biol* 2004, **20**:87–123.
45. McGuckin MA, Eri RD, Das I, Lourie R, Florin TH: **ER stress and the unfolded protein response in intestinal inflammation.** *Am J Physiol Gastrointest Liver Physiol* 2010, **298**:G820–G832.
46. Yamaguchi Y, Larkin D, Lara-Lemus R, Ramos-Castaneda J, Liu M, Arvan P: **Endoplasmic reticulum (ER) chaperone regulation and survival of cells compensating for deficiency in the ER stress response kinase, PERK.** *J Biol Chem* 2008, **283**:17020–17029.
47. Lee AH, Iwakoshi NN, Glimcher LH: **XBP-1 regulates a subset of endoplasmic reticulum resident chaperone genes in the unfolded protein response.** *Mol Cell Biol* 2003, **23**:7448–7459.
48. Hirao K, Natsuka Y, Tamura T, Wada I, Morito D, Natsuka S, et al: **EDEM3, a soluble EDEM homolog, enhances glycoprotein endoplasmic reticulum-associated degradation and mannose trimming.** *J Biol Chem* 2006, **281**:9650–9658.
49. Fox RM, Hanlon CD, Andrew DJ: **The Creba/Creb3-like transcription factors are major and direct regulators of secretory capacity.** *J Cell Biol* 2010, **191**:479–492.
50. Niles RM, Christensen TG, Breuer R, Stone PJ, Snider GL: **Serine proteases stimulate mucous glycoprotein release from hamster tracheal ring organ culture.** *J Lab Clin Med* 1986, **108**:489–497.
51. Kogel T, Gerdes HH: **Maturation of secretory granules.** *Results Probl Cell Differ* 2010, **50**:1–20.
52. Saegusa C, Tanaka T, Tani S, Itohara S, Mikoshiba K, Fukuda M: **Decreased basal mucus secretion by Slp2-a-deficient gastric surface mucous cells.** *Genes Cells* 2006, **11**:623–631.
53. Varadi A, Ainscow EK, Allan VJ, Rutter GA: **Involvement of conventional kinesin in glucose-stimulated secretory granule movements and exocytosis in clonal pancreatic beta-cells.** *J Cell Sci* 2002, **115**:4177–4189.
54. Tomas A, Yermen B, Min L, Pessin JE, Halban PA: **Regulation of pancreatic beta-cell insulin secretion by actin cytoskeleton remodelling: role of gelsolin and cooperation with the MAPK signalling pathway.** *J Cell Sci* 2006, **119**:2156–2167.

55. Henke MO, John G, Germann M, Lindemann H, Rubin BK: **MUC5AC and MUC5B mucins increase in cystic fibrosis airway secretions during pulmonary exacerbation.** *Am J Respir Crit Care Med* 2007, **175**:816–821.
56. Nakao I, Kanaji S, Ohta S, Matsushita H, Arima K, Yuyama N, et al: **Identification of pendrin as a common mediator for mucus production in bronchial asthma and chronic obstructive pulmonary disease.** *J Immunol* 2008, **180**:6262–6269.
57. Loewen ME, Forsyth GW: **Structure and function of CLCA proteins.** *Physiol Rev* 2005, **85**:1061–1092.
58. Xiang YY, Wang S, Liu M, Hirota JA, Li J, Ju W, et al: **A GABAergic system in airway epithelium is essential for mucus overproduction in asthma.** *Nat Med* 2007, **13**:862–867.
59. Dolganov GM, Woodruff PG, Novikov AA, Zhang Y, Ferrando RE, Szubin R, et al: **A novel method of gene transcript profiling in airway biopsy homogenates reveals increased expression of a Na⁺-K⁺-ATPase cotransporter (NKCC1) in asthmatic subjects.** *Genome Res* 2001, **11**:1473–1483.
60. Wang R, Wang G, Ricard MJ, Ferris B, Strulovici-Barel Y, Salit J, et al: **Smoking-induced upregulation of AKR1B10 expression in the airway epithelium of healthy individuals.** *Chest* 2010, **138**:1402–1410.
61. Raman T, O'Connor TP, Hackett NR, Wang W, Harvey BG, Attiyeh MA, et al: **Quality control in microarray assessment of gene expression in human airway epithelium.** *BMC Genomics* 2009, **10**:493.
62. Kulkarni B, Mohammed I, Hopkinson A, Dua HS: **Validation of endogenous control genes for gene expression studies on human ocular surface epithelium.** *PLoS One* 2011, **6**:e22301.
63. Huang D, Sherman B, Lempicki R: **Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.** *Nat Protoc* 2009, **4**:44–57.
64. Tilley AE, O'Connor TP, Hackett NR, Strulovici-Barel Y, Salit J, Amoroso N, et al: **Biologic phenotyping of the human small airway epithelial response to cigarette smoking.** *PLoS One* 2011, **6**:e22798.
65. Hackett NR, Salit J, Crystal RG: **Novel smoking-related genes in the small airway epithelium identified by massively parallel RNA sequencing.** *Am J Respir Crit Care Med* 2010, .
66. Wiede A, Jagla W, Welte T, Kohnlein T, Busk H, Hoffmann W: **Localization of TFF3, a new mucus-associated peptide of the human respiratory tract.** *Am J Respir Crit Care Med* 1999, **159**:1330–1335.
67. Madsen J, Nielsen O, Tornøe I, Thim L, Holmskov U: **Tissue localization of human trefoil factors 1, 2, and 3.** *J Histochem Cytochem* 2007, **55**:505–513.
68. Zheng X, Tsuchiya K, Okamoto R, Iwasaki M, Kano Y, Sakamoto N, Nakamura T, Wata-nabe M: **Suppression of hath1 gene expression directly regulated by hes1 via notch signaling is associated with goblet cell depletion in ulcerative colitis.** *Inflamm Bowel Dis* 2011, **17**:2251–2260.
69. Lopez-Ferrer A, de BC, Barranco C, Garrido M, Isern J, Carlstedt I, Reis CA, Torrado J, Real Fx: **Role of fucosyltransferases in the association between apomucin and Lewis an-tigen expression in normal and malignant gastric epithelium.** *Gut* 2000, **47**:349–356.
70. Rebhan M, Chalifa-Caspi V, Prilusky J, Lancet D: **GeneCards: a novel functional genom-ics compendium with automated data mining and query reformulation support.** *Bioinformatics* 1998, **14**:656–664.
71. Ehre C, Rossi AH, Abdullah LH, De PK, Hill S, Olsen JC, Davis CW: **Barrier role of actin filaments in regulated mucin secretion from airway goblet cells.** *Am J Physiol Cell Physiol* 2005, **288**:C46–C56.
72. Castermans D, Volders K, Crepel A, Backx L, De VR, Freson K, Meulemans S, Vermeesch JR, Schrandt-Stumpel CT, De RP, et al: **SCAMP5, NBEA and AMISYN: three candi-date genes for autism involved in secretion of large dense-core vesicles.** *Hum Mol Genet* 2010, **19**:1368–1378.
73. Ghaleb AM, Aggarwal G, Bialkowska AB, Nandan MO, Yang VW: **Notch inhibits expres-sion of the Kruppel-like factor 4 tumor suppressor in the intestinal epithelium.** *Mol Cancer Res* 2008, **6**:1920–1927.
74. Olivari S, Galli C, Alanen H, Ruddock L, Molinari M: **A novel stress-induced EDEM va-riant regulating endoplasmic reticulum-associated glycoprotein degradation.** *J Biol Chem* 2005, **280**:2424–2428.
75. Sriburi R, Bommiasamy H, Buldak GL, Robbins GR, Frank M, Jackowski S, Brewer JW: **Coordinate regulation of phospholipid biosynthesis and secretory pathway gene ex-pression in xBP-1(S)-induced endoplasmic reticulum biogenesis.** *J Biol Chem* 2007, **282**:7024–7034.
76. Wang Z, Sun Z, Li AV, Yarema KJ: **Roles for UDP-GlcNAc 2-epimerase /ManNAc 6-kinase outside of sialic acid biosynthesis: modulation of sialyltransferase and BiP ex-pression, GM3 and GD3 biosynthesis, proliferation, and apoptosis, and ERK1/2 phos-phorylation.** *J Biol Chem* 2006, **281**:27016–27028.

doi:10.1186/1755-8794-5-21

Cite this article as: Wang et al.: Genes associated with MUC5AC expression in small airway epithelium of human smokers and non-smokers. *BMC Medical Genomics* 2012 5:21.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

