Cancer Research

Modulation of Cystatin A Expression in Human Airway Epithelium Related to Genotype, Smoking, COPD, and Lung Cancer

Marcus W. Butler[1], Tomoya Fukui[1], Jacqueline Salit[1], Renat Shaykhiev[1], Jason G. Mezey[1,2], Neil R. Hackett[1], and Ronald G. Crystal[1]

Abstract

The cathepsin inhibitor Cystatin A (*CSTA*) has antiapoptotic properties linked with neoplastic changes in squamous cell epithelium, where it has been proposed as a diagnostic and prognostic marker of lung cancer. Notably, cystatin A is upregulated in dysplastic epithelium, prompting us to hypothesize that it might be modulated in chronic obstructive pulmonary disease (COPD), a small airway epithelial (SAE) disorder that is a risk factor for non–small cell lung cancer (NSCLC) in a subset of smokers. Here we report that genetic variation, smoking, and COPD can all elevate levels of *CSTA* expression in lung small airway epithelia, with still further upregulation in squamous cell carcinoma (SCC), an NSCLC subtype. We examined SAE gene expression in 178 individuals, including healthy nonsmokers ($n = 60$), healthy smokers ($n = 82$), and COPD smokers ($n = 36$), with corresponding large airway epithelium (LAE) data included in a subset of subjects ($n = 52$). Blood DNA was genotyped by SNP microarray. Twelve SNPs upstream of the *CSTA* gene were found to associate with its expression in SAE. Levels were higher in COPD smokers than in healthy smokers, who, in turn, had higher levels than nonsmokers. *CSTA* gene expression in LAE was also smoking-responsive. Using publicly available NSCLC expression data we also found that *CSTA* was upregulated in SCC versus LAE and downregulated in adenocarcinoma versus smoke-exposed SAE. All phenotypes were associated with different proportional expression of *CSTA* to cathepsins. Our findings establish that genetic variability, smoking, and COPD all influence *CSTA* expression, as does SCC, supporting the concept that *CSTA* may make pivotal contributions to NSCLC pathogenesis in both early and late stages of disease development. *Cancer Res; 71(7); 2572–81. ©2011 AACR.*

Introduction

Cystatin A (*CSTA*; also referred to as stefin A, and acid cysteine proteinase inhibitor), a member of the class I or stefin subgroup of the 3 described cystatin families, is a 11 kDa single chain intracellular cysteine protease inhibitor capable of inhibiting papain and cathepsins B, H, and L, as well as the cysteine protease activity of the major house dust mite allergen Der p1 (1–3). *CSTA* has antiapoptotic properties (4–6) and cystatin A expression has been linked with neoplastic changes in squamous cell epithelium with cystatin A levels paralleling the response to antitumor therapy (7–11). As with other members of the stefin family, cystatin A has been proposed as a marker and prognostic indicator of lung cancer (11, 12). In the lung, cystatin A expression is upregulated in dysplastic epithelium, and highly expressed in many lung cancers, especially squamous cell carcinoma (SCC; refs 7, 10, 12).

With this background, and based on the knowledge that the majority of lung cancers are derived from the small airway epithelium (SAE) of cigarette smokers (13–15), but that only a subset of smokers develop lung cancer (16, 17), we hypothesized that the expression of *CSTA* in the SAE is modulated by genetic variation and upregulated by smoking. Further, in the context that chronic obstructive pulmonary disease (COPD), a disease that starts in the small airways (13, 18–21), is a risk factor for lung cancer independent of smoking (22–26), we asked whether *CSTA* is upregulated in the SAE of smokers with COPD beyond that of smokers *per se*, and whether this upregulation is also dependent on genetic variation. In view of the fact that both smoking and COPD are stronger risk factors for SCC than for adenocarcinoma of the lung, and that most SCCs are derived from large airways (15, 27–30), we also examined *CSTA* expression in large airway epithelium (LAE) samples obtained in a subset of the study subjects, and compared our *CSTA* gene expression data to publicly available lung cancer microarray

Authors' Affiliation: [1]Department of Genetic Medicine, Weill Cornell Medical College, New York, New York, and [2]Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, New York

Note: Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/).

Corresponding Author: Ronald G. Crystal, Department of Genetic Medicine, Weill Cornell Medical College, 1300 York Avenue, Box 96, New York, NY 10065. Phone: 646-962-4363; Fax: 646-962-0220; E-mail: geneticmedicine@med.cornell.edu

doi: 10.1158/0008-5472.CAN-10-2046

data derived using comparable methodologies (31), to see if *CSTA* is discordantly expressed in SCC and adenocarcinoma of the lung, relative to the SAE and LAE.

Materials and Methods

Study subjects

Healthy nonsmokers, healthy smokers, and smokers with COPD were assessed in the Weill Cornell National Institutes of Health Clinical and Translational Sciences Center and Department of Genetic Medicine Clinical Research Facility using protocols approved by the Weill Cornell Medical College Institutional Review Board. Prior to inclusion in the study, each individual provided written informed consent. No abnormality was found in the nonsmokers ($n = 60$) or the healthy smokers ($n = 82$) following a standardized screening evaluation composed of a history, physical examination, complete blood count, serum chemistries, coagulation profile, liver function tests, urine studies, chest X-ray, EKG, and full lung function studies. The COPD smokers ($n = 36$) were diagnosed in accordance with standard GOLD criteria (32). Blood was collected for genotyping on a random subset of individuals (nonsmokers, $n = 44$; healthy smokers, $n = 48$; COPD smokers, $n = 20$). Fewer subjects with genotyped samples were available than subjects with SAE gene expression data because of a significantly longer temporal interest of our laboratory in the study of airway epithelium gene expression than in genomic studies. LAE samples were not given as high a priority as SAE samples when individuals were bronchoscoped, reflecting the focus of our lab on SAE gene expression profiles, and as a consequence, LAE samples were available for a random subset of individuals who underwent bronchoscopy (healthy nonsmokers, $n = 21$; healthy smokers, $n = 31$). Urine nicotine and cotinine were measured, in conjunction with serum carboxyhemoglobin, to verify the self-reported smoking status of each group. Detailed inclusion/exclusion criteria and characteristics of the study populations can be found in the Supplementary Materials and Methods section.

Collection of airway epithelium and assessment of gene expression

Fiberoptic bronchoscopy was used to obtain small (and for a random subset, large) airway epithelial cells as previously described (33, 34). Further details of the airway epithelial sampling, preparation of cDNA, hybridization of labeled cRNA to Affymetrix HG-U133 Plus 2.0 arrays and confirmation by TaqMan (Applied Biosystems) real-time RT-PCR (reverse transcriptase PCR) can be found in the Supplementary Materials and Methods section. No cell lines were used in the conduct of this research; all samples used were freshly collected and immediately processed for RNA or protein.

Correlation of genotypes and haplotypes with gene expression

Genomic DNA was extracted from stored blood samples using the Autogen FX robotic system in accordance with Autogen's protocols (Autogen). Preprinted bar-coded labels were affixed to sample containers to minimize sample mix-ups, and critical steps in sample processing were only undertaken when 2 technicians were present. The Affymetrix Human SNP Array 5.0 platform was used to assess genotype using the manufacturer's protocols. The focus was on all 48 SNPs found on the SNP array whose chromosomal location was within 100,000 base pairs either side of the *CSTA* gene. To avoid artifactual association of genotype with expression caused by SNPs with known minor allele frequency more than 5% situated within the target sequence of Affymetrix expression probe sets, the sequences of individual probes (obtained from NetAffx, Affymetrix) were cross-checked against the National Center for Biotechnology Information (NCBI) dbSNP build 129. Potential effects of copy number variation on associations of genotype with gene expression were assessed as outlined in the Supplementary Materials and Methods section.

Data analysis and statistics for SAE expression and genomic data

Gene expression analyses on all samples were performed using the Microarray Suite 5.0 (MAS5) software (Affymetrix). Publicly available data from Kuner and colleagues (31) for 40 subjects with lung adenocarcinoma and 18 subjects with SCC of the lung were downloaded from the NCBI Web site (accession number GSE10245), using the gene expression data from the only *CSTA* probe set 204971_at on the Affymetrix HG-U133 Plus 2.0 chip, the same platform that was used on the SAE and LAE samples. The proportional relationship among *CSTA* and cathepsin gene expression was examined in all samples using the additional probe sets 200838_at (*cathepsin B*), 202295_s_at (*cathepsin H*), and 202087_s_at (*cathepsin L*). Gene expression data were normalized using GeneSpring version 7.2 software (Agilent Technologies) per chip and per gene across all samples. All microarray data have been deposited at the Gene Expression Omnibus (GEO) site (ref. 35; accession number GSE22047). Genomic data from the Affymetrix Human SNP 5.0 were assessed using the BRLMM-P Analysis Tool (BAT) 2.0 software (Affymetrix) to determine genotype for *cis*-SNPs in the vicinity of *CSTA* on chromosome 3. Associations between *CSTA* gene expression levels and genotype for the 48 SNPs located within 100 kb of the gene were performed using PLINK, with permutation testing used to control for the effect of genetic ancestry. Haplotype associations with expression were performed using PLINK and Haploview, with linear regression modeling (see the Supplementary Materials and Methods section).

Results

Association of *CSTA* small airway epithelial gene expression with genomic variation

The study population demographic findings and characteristics of brushed SAE samples and microarray expression probe performance are detailed in the Supplementary Results section. Based on the hypothesis that genetic variability might modulate the level of *CSTA* expression in the SAE, we evaluated the correlation of *CSTA* gene expression with genotypes of *cis*-SNPs within 100 kbp either side of the *CSTA* gene using

Table 1. Association of *CSTA* genotypes with *CSTA* gene expression in the small airway epithelium

SNP identity[b]	Chr. 3 location[c]	Minor allele frequency[d]	*P* value for association of genotype with expression level[a]			
			Assessed by microarray			
			All subjects[e] ($n = 112$)	Healthy nonsmokers ($n = 44$)	Healthy smokers ($n = 48$)	COPD smokers ($n = 20$)
rs1354162	123,436,767	0.040	0.02	NS	NS	4.6×10^{-3}
rs7652858	123,480,795	0.125	3.6×10^{-3}	NS	NS	NS
rs2134225	123,484,943	0.125	3.6×10^{-3}	NS	NS	NS
rs16832956	123,500,198	0.161	4.5×10^{-4}	NS	3.4×10^{-3}	0.010
rs1402200	123,505,107	0.411	7.3×10^{-3}	NS	NS	NS
rs5008830	123,513,152	0.089	1.0×10^{-3}	0.025	1.1×10^{-5}	0.038
rs2001548	123,515,479	0.101	1.7×10^{-3}	NS	3.4×10^{-5}	0.038
rs4678180	123,520,487	0.455	5.2×10^{-3}	NS	NS	NS
rs9864290	123,522,752	0.451	7.8×10^{-3}	NS	NS	NS
rs6803098	123,523,300	0.453	4.1×10^{-3}	NS	NS	NS
rs9817571	123,538,670	0.099	0.03	NS	0.018	NS
rs9842752	123,539,553	0.076	0.03	NS	NS	NS

Note: *CSTA* gene expression levels in small airway epithelium as assessed by Affymetrix HGU133 Plus 2.0 microarray expression probe set 204971_at were correlated with Affymetrix Human SNP Array 5.0 *cis*-SNPs situated within 100 kbp of the *CSTA* gene location on chromosome 3.3q21.

[a]For analysis with the group of all subjects, *P* values represent Wald statistic following permutation analysis: after initial associations were identified, each association was further tested by performing 10^3 permutations within clusters of similar genetic ancestry, as defined by STRUCTURE, using PLINK software. For the subgroup analyses, *t* tests are shown.

[b]Only those SNPs with *P* values for association of genotype with expression of <0.05 are shown. SNPs are listed in ascending order of location on chromosome 3. 5′ to 3′, the first 4 SNPs listed are within introns of the upstream *CASR* gene, the next 6 are intergenic, and the last 2 listed SNPs are located within introns of *CSTA*.

[c]Chr, chromosome. Locations refer to NCBI Human Genome Reference Genome Assembly, Build 36.3

[d]Minor allele frequencies observed in the study population.

[e]All subjects with SAE gene expression data and genotype data. Healthy nonsmokers ($n = 44$), healthy smokers ($n = 48$), and COPD smokers ($n = 20$) combined. NS = not significant ($P > 0.05$).

paired airway gene expression data and blood DNA SNP data from the subset of the study population for whom both gene expression and SNP data were available (healthy nonsmokers, $n = 44$; healthy smokers, $n = 48$; and COPD smokers, $n = 20$). Ten SNPs located upstream of the *CSTA* gene, and 2 more within introns of the *CSTA* gene, displayed significant correlation of genotype with SAE expression levels of *CSTA* (Table 1, Fig. 1A). For the total combined group of subjects (healthy nonsmokers, healthy smokers and smokers with COPD), the most significant correlation of the levels of *CSTA* small airway gene expression was with the SNP *rs16832956* ($P < 5 \times 10^{-4}$). Nine of the 12 SNPs identified had $P < 10^{-2}$ in reference to the strength of their association with *CSTA* SAE gene expression. The 4 SNPs most upstream of *CSTA* were located within introns of the calcium-sensing receptor (*CASR*) gene, but none had any relationship of genotype with *CASR* gene expression levels ($P > 0.8$, not shown).

These associations persisted at this significance level after assessing 10^3 permutations within clusters of similar genetic ancestry suggesting that genotype rather than genetic ancestry was the cause of the observed effect. For example, *G*_ genotypes of *rs16832956*, the SNP with the strongest genetic association, had low *CSTA* expression, with a >1.4-fold ± 0.1-fold increase seen in *CC* genotypes (Fig. 1B, $P < 1.6 \times 10^{-4}$). The SAE *CSTA* gene expression levels were confirmed by TaqMan RT-PCR (Fig. 1C, $r^2 = 0.67$, $P < 10^{-4}$).

Influence of ancestral background on genetic associations with *CSTA* small airway epithelium gene expression

Genetic ancestry could be a potential confounding variable limiting the generalizability of observations made in relation to the modulation of *CSTA* gene expression by genetic variability. However, the associations of *rs16832956* genotypes with *CSTA* expression levels were similar among the 2 major ancestral groupings ($P < 0.03$, both ancestral groups), with no significant difference in allele frequency among those of African American ancestry versus subjects of European ancestry (Fig. 2A, *G* allele frequency 0.19 in African American ancestral group, *G* allele frequency 0.09 in European ancestral group, $P > 0.08$). Haplotypes, comprising SNPs surrounding the *CSTA* genomic location, were significantly associated with

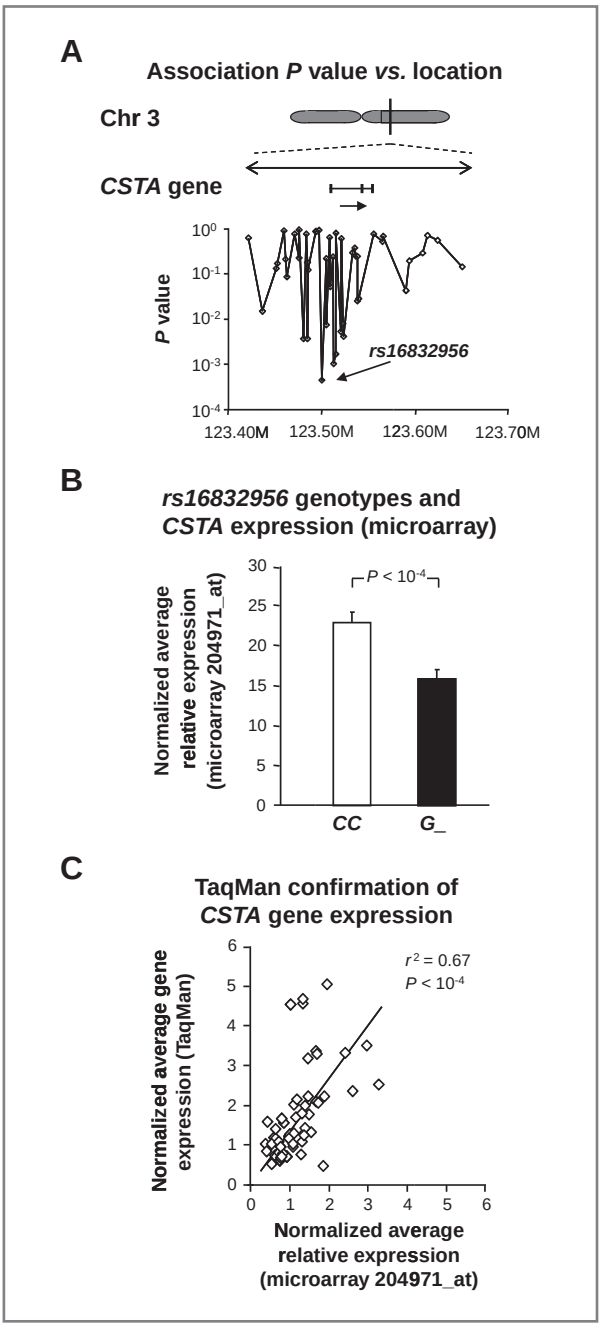

Figure 1. Genetic modulation of *CSTA* SAE gene expression. A, magnified view of the *CSTA* gene region, with direction of transcription (arrow). Below the gene, same scale, are the corresponding chromosomal locations of 48 *cis*-SNPs on the Affymetrix Human SNP Array 5.0 found within 100 kb either side of the gene. Shown are the correlations (Wald statistic) with microarray-assessed SAE *CSTA* gene expression in 112 healthy nonsmokers, healthy smokers, and smokers with COPD. B, microarray normalized average expression values of *CSTA* for genotypes of SNP *rs16832956*. Shown are data for all 112 genotyped subjects with *CSTA* SAE gene expression on the ordinate. *G_* = data for the combined genotypes *GG* and *CG*. C, TaqMan RT-PCR confirmation of microarray *CSTA* gene expression levels in SAE in a random subset of healthy nonsmokers ($n = 23$), healthy smokers ($n = 28$), and COPD smokers ($n = 13$).

CSTA gene expression for specific ancestral subgroups examined separately (Fig. 2B and C), with the strongest association seen in the group of individuals of European ancestry for haplotype *GAGGGACCCGCT* (Fig. 2C, $P < 0.008$; see the Supplementary Results section and Supplementary Table II).

Effect of smoking status and COPD on *CSTA* gene expression in the SAE

Gene expression levels of *CSTA* in SAE were significantly higher in the group of healthy smokers ($n = 82$) compared with healthy nonsmokers ($n = 60$, $P < 0.04$, pairwise Student's t test), but were even more upregulated in smokers with COPD ($n = 36$) compared with the healthy smokers (Fig. 3A, $P < 10^{-3}$, pairwise Student's t test; $P < 10^{-4}$ by analysis of variance for all 3 groups). The association of *CSTA* expression with phenotype was confirmed by RT-PCR, and was explained neither by genetic ancestry nor by age, though there was evidence of a possible dose–response relationship (Fig. 3B–D; see the Supplementary Results section). The effect of genotype on SAE *CSTA* gene expression seen in the total genotyped study population, was also observed when healthy nonsmokers, healthy smokers, and COPD smokers were examined separately (see the Supplementary Results section, Supplementary Fig. 1A and B).

Western analysis of cystatin A protein expression

Whole-cell lysates of SAE from healthy nonsmokers, healthy smokers, and smokers with COPD were quantitatively assessed for cystatin A expression using Western analysis. Western analysis revealed increased *CSTA* protein expression in healthy smokers compared with nonsmokers, with even higher expression in smokers with COPD compared with the healthy smokers (Fig. 4A), which was confirmed quantitatively by densitometry (Fig. 4B, $P < 0.04$ by pairwise comparisons by Student's t test; $P < 10^{-2}$ by analysis of variance for all 3 phenotypic groups).

CSTA gene expression in LAE, SCC, and adenocarcinoma

Most lung cancers, especially adenocarcinoma, are derived from the SAE, and yet it is well established that smoking and COPD are more strongly linked to SCC than to adenocarcinoma (14, 28, 30). SCCs tend to arise in the more central airways (15). A comparison of *CSTA* expression was made among SAE, LAE, and non–small cell lung cancer (NSCLC) specimens (see the Supplementary Results section for details of the study population demographics and characteristics of brushed LAE samples). *CSTA* expression levels in LAE of healthy nonsmokers were similar to the upregulated levels seen in the SAE of COPD smokers. LAE *CSTA* expression was smoke-responsive (Fig. 5A, $P < 10^{-3}$) and but further upregulated in SCC (Fig. 5A, $P < 10^{-2}$, all comparisons). In contrast, adenocarcinoma *CSTA* gene expression levels were significantly downregulated compared with all but the SAE of healthy nonsmokers (Fig. 5A; see the Supplementary Results section).

In view of the marked upregulation of *CSTA* observed in SCC subjects, a comparison was made of the ratio of *CSTA*

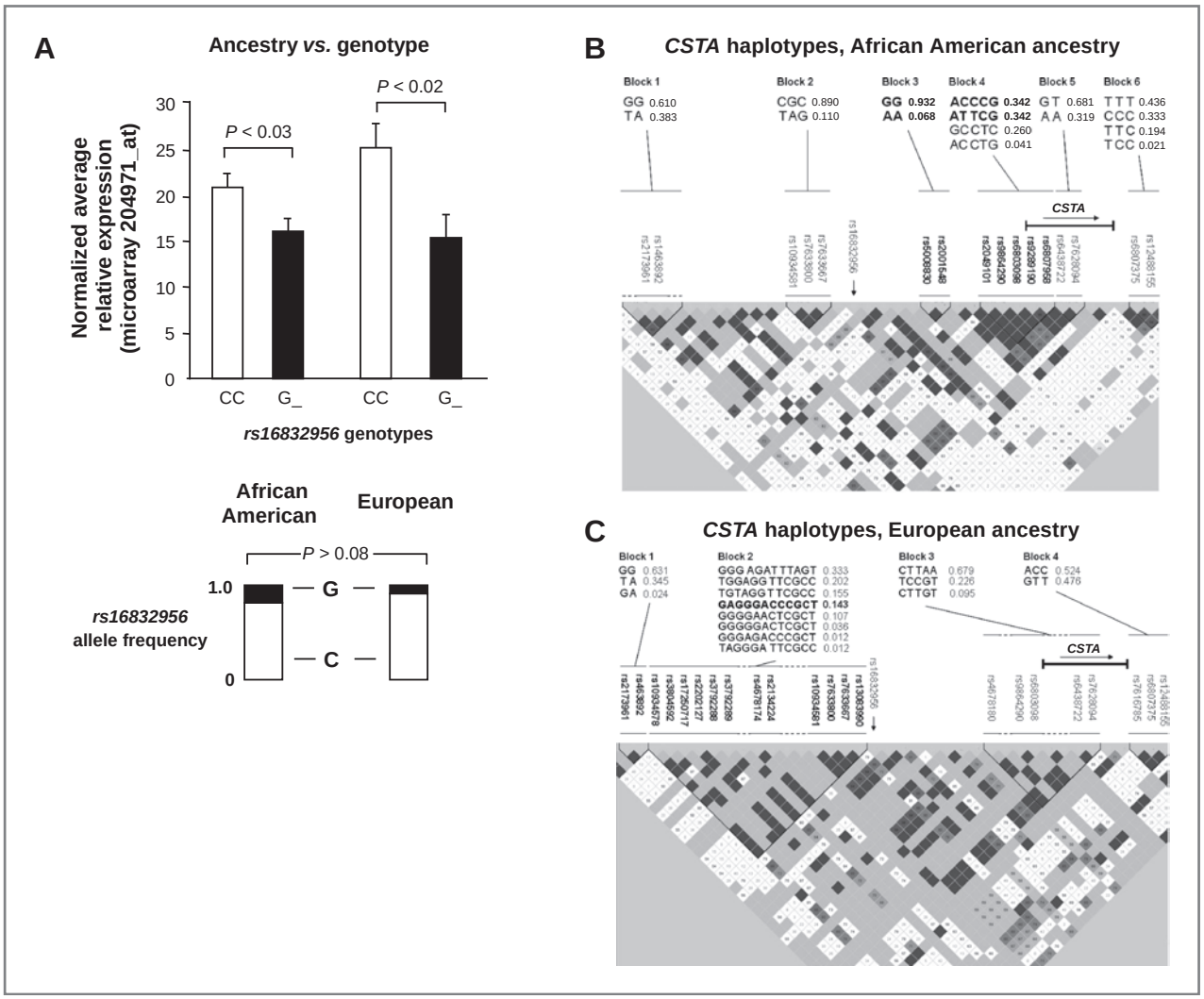

Figure 2. Effect of ancestry on genetic modulation of SAE *CSTA* expression. A, *CSTA* expression based on genotypes of *rs16832956* parsed by the 2 major ancestral groups, African American and European. Ordinate—normalized average relative *CSTA* gene expression. Allele frequencies as shown. B, analysis of haplotypes derived from local *CSTA cis*-genotypes in subjects of African American genetic ancestry. Analyses performed on the subset of 56 subjects of African American ancestry with available genotype data. The genomic sequence and frequencies of the 6 haplotypes blocks are shown in the upper portion of the figure, corresponding to the named SNPs indicated below. In the haplotype map, the darker the shading, the stronger the linkage disequilibrium (LD) between adjacent SNPs, and vice versa. The haplotypes in blocks 3 and 4 (highlighted in bold type, with the highlighted constituent SNPs shown beneath) were significantly associated with *CSTA* gene expression (see Supplementary Table II). The relative location and direction of the *CSTA* gene is indicated. The relative location of the SNP most strongly associated with *CSTA* gene expression, *rs16832956*, is indicated (vertical arrow). C, *CSTA* haplotype association data for 35 subjects of European genetic ancestry. The indicated haplotype in block 2, composed of the 12 SNPs shown in bold, was the only haplotype significantly associated with *CSTA* gene expression (see Supplementary Table II).

gene expression to expression levels of the 3 known cathepsin targets (B, H, and L) of *CSTA* within individuals. These expression ratios differed significantly among most phenotypic and tissue groupings, except for adenocarcinoma versus nonsmoker SAE (see the Supplementary Results section, Fig. 5B–D).

Discussion

The cytosolic cysteine protease inhibitor cystatin A, coded for by the *CSTA* gene, has attracted interest with a number of reports relating expression of *CSTA* protein levels to neoplastic states in a variety of tissues including those of epithelial origin (7–12, 36). Immunohistochemical data have suggested that *CSTA* is expressed at higher levels in dysplastic human bronchial epithelium, in SCC in particular, and less often in adenocarcinoma of the lung, compared with normal bronchial epithelium (11, 12). In light of the fact that COPD, a disease that arises in the small airways, is a risk factor for lung cancer independent of the risk attributable to smoking, but only occurs in a minority of smokers (16–20, 22–27, 29), we asked: does genetic variation modulate *CSTA* gene expression levels in the human SAE; is the expression of

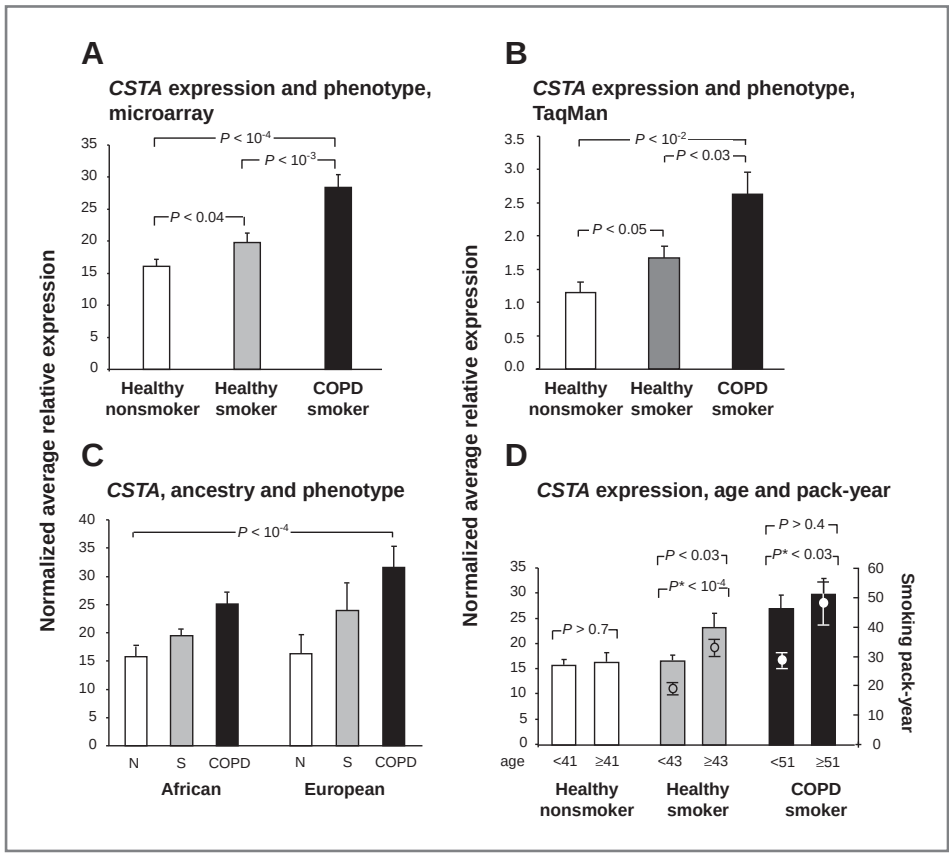

Figure 3. *CSTA* gene expression levels in small airway epithelium (SAE) of healthy nonsmokers, healthy smokers, and smokers with COPD. A, average relative mRNA expression values of *CSTA* (healthy nonsmokers, $n = 60$; healthy smokers, $n = 82$; and COPD smokers, $n = 36$) as detected by microarray. B, TaqMan RT-PCR confirmation of the association of SAE *CSTA* gene expression levels with phenotype in a subset of healthy nonsmokers ($n = 23$), healthy smokers ($n = 28$), and smokers with COPD ($n = 13$). C, average relative expression level of *CSTA* for the same subjects as in A, with genetic ancestry on the abscissa. Two-factor ANOVA: $P < 10^{-4}$ for phenotype; $P > 0.09$ for the factor "ancestry"; interaction among phenotype and ancestry $P > 0.5$. D, average relative expression level of *CSTA* in SAE as a function of age and smoking history. Left ordinate: *CSTA* gene expression, bar plots; right ordinate: smoking history of the 2 groups of smokers (in pack-years), dot and whisker plots and asterisk *P* values. Each of the phenotypic groups is divided by mean age (of that group). Error bars, standard error. *P* values are from pairwise student *t* tests, except where otherwise indicated.

CSTA in SAE influenced by both smoking and by COPD; and since smoking and COPD each have a propensity toward SCC rather than adenocarcinoma (28–30), is *CSTA* expression differentially influenced by SCC versus adenocarcinoma of the lung relative to SAE and LAE expression levels in the noncancerous samples?

Using fiberoptic bronchoscopy, pure populations of SAE were obtained from 60 carefully phenotyped healthy nonsmokers, 82 healthy smokers, and 36 smokers with COPD, and for the majority, both SAE *CSTA* gene expression and corresponding blood *CSTA* genotypes were assessed by microarray. The data demonstrate that *CSTA* gene expression levels in the human SAE are modulated by 12 SNPs within the vicinity of the gene, a genetic association that is not confounded by genetic ancestry. There is an association of local *cis*-haplotypes with *CSTA* gene expression for 4 haplotypes in the case of subjects of African American ancestry and 1 haplotype in those of European ancestry. In addition, healthy smokers have higher *CSTA* gene expression levels than healthy nonsmokers, with even higher levels observed in smokers with COPD compared with the healthy smokers independent of pack-years, albeit with a suggestive evidence of a dose–response relationship for smokers. The microarray expression was confirmed at the transcript level by quantitative TaqMan RT-PCR and at the protein level by Western analysis. The genetic modulation of SAE *CSTA* gene expression was a consistent finding when the 3 phenotypic groups were examined separately. Finally, tumor tissues from individuals with SCC of the lung have the highest relative levels of *CSTA* gene expression, but tumor tissue from adenocarcinoma subjects have substantially lower levels of *CSTA* than in any of the other tissues and phenotypes examined (except for being similar to healthy nonsmokers). Interestingly, the phenotypic states examined (healthy, COPD, lung cancer) are each associated with significantly different proportional gene expression of *CSTA* to cathepsins B, H, and L. These observations are in keeping with the concept that the *CSTA* gene plays a role in the evolution of the bronchial epithelium of peripheral and central airways from normal to disease, and suggests a

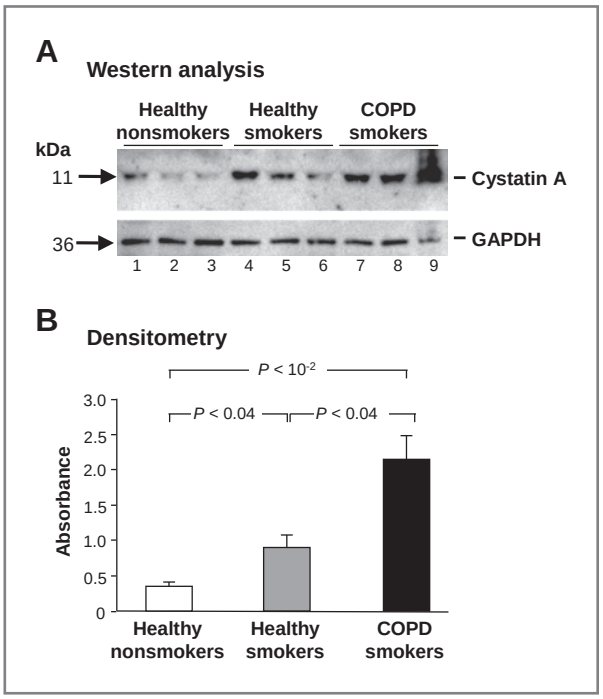

Figure 4. Western analysis of SAE cystatin A protein expression in nonsmokers, healthy smokers, and smokers with COPD. A, Western analysis. Top, SAE cystatin A protein expression in healthy nonsmokers (lanes 1–3), healthy smokers (lanes 4–6), and smokers with COPD (lanes 7–9). Bottom, protein loading control antibody (GAPDH). B, quantification of the data in A by densitometry normalized to GAPDH. *P* values, pairwise student *t* test. Error bars, standard error.

complex interplay among genetics, smoking, COPD, and lung cancer histologic subtype in relation to *CSTA* gene expression. *CSTA* appears to have a more plausible connection to the specific evolution of healthy smoker and COPD smoker airway epithelium into SCC rather than into adenocarcinoma, in keeping with the known stronger relationship of the separate risk factors of smoking and COPD with SCC than with adenocarcinoma of the lung (27–30).

Small airway epithelium, smoke-induced lung disease, and genetics

There is now a large body of evidence pointing to the SAE (defined as bronchi of $\geq$6 generations, <2 mm in diameter) as the earliest site of pathologic involvement in COPD, the primary site of airflow limitation in this disorder, and the site of development of most NSCLC (13–15, 18–21). Consistent with these observations in the major smoke-induced lung diseases, morphologic changes are found in small airways of asymptomatic smokers with normal lung function (37, 38). COPD is a relevant phenotype to study in the progression from healthy airway epithelium to lung cancer, an airway epithelial-derived disease, because COPD is itself an independent risk factor for the development of lung cancer, with a contribution to lung cancer risk that is separate from the risk attributable to cigarette smoking (22–26). Since only 15% to 20% of smokers will develop COPD, together with evidence of familial clustering in COPD and NSCLC and twin concordance for the risk for COPD (15–17, 22, 39–41), it is likely that genetic variation influences disease risk. The fact that SCC is more common than adenocarcinoma in smokers and COPD subjects also raises the question of the role of the LAE in the pathogenesis of these tumor subtypes. Although the current study was not designed to focus on LAE gene expression/genotype relationships, it is interesting to note that the smoke-responsive nature of *CSTA* expression in SAE was also observed in LAE. However, the levels of *CSTA* in healthy smoker LAE are still considerably less than in SCC subjects and taken together with the smoke-responsiveness, are therefore less likely to reflect mere constitutive expression of *CSTA* in a LAE-derived tumor. Although the observations of genetic modulation of *CSTA* gene expression and the upregulation of such expression in SAE from COPD smokers compared with healthy smokers is intriguing, the present study was not designed to address the question of whether or not *CSTA* expression is genetically associated with COPD. Testing this hypothesis would require the genotyping of a large number of subjects to have sufficient statistical power to reach a definitive conclusion.

Although the data show some evidence for a dose–response relationship between intensity of smoke exposure and *CSTA* gene expression, disease also influences the expression levels, as evidenced by the greater relative increase in *CSTA* levels in COPD and SCC versus healthy smokers compared with the smaller increase in healthy smokers versus nonsmokers, even though the latter comparison would have the greater difference in smoking exposure. Also, reciprocal effects of disease state and pack-years on *CSTA* gene expression in SAE were observed in the healthy smoker and COPD smoker age-defined subgroups.

Analysis of a published gene expression data set that used microarray technology to study lung cancer tissues did not show a significant effect of smoking status on *CSTA* gene expression (42). Others have noted that *CSTA* levels are higher in early-detected lung cancers compared with lung cancers presenting later in the disease course, despite higher pack-year values in the latter (43). These disparate observations highlight the need for further study to clarify if there is an effect of smoking-intensity on *CSTA* gene expression in lung airway epithelium and cancers derived from the airway epithelium. The observations suggest that *CSTA* expression is not uniformly smoke-responsive and is likely to be specific to cell populations. Controlling for the influence of age while examining pack-year consumption is challenging and would likely require large data sets, because more often than not, individuals with higher pack-year histories will also be older. From the data in the present study, it seems more likely that it is the intensity of smoking rather than age that exerts the greater influence on *CSTA* expression levels, as evidenced by the concordance among the magnitudes of pack-year difference and the difference in SAE *CSTA* levels in comparing the 2 smoking phenotypic groups (healthy smokers and COPD smokers) parsed by median age.

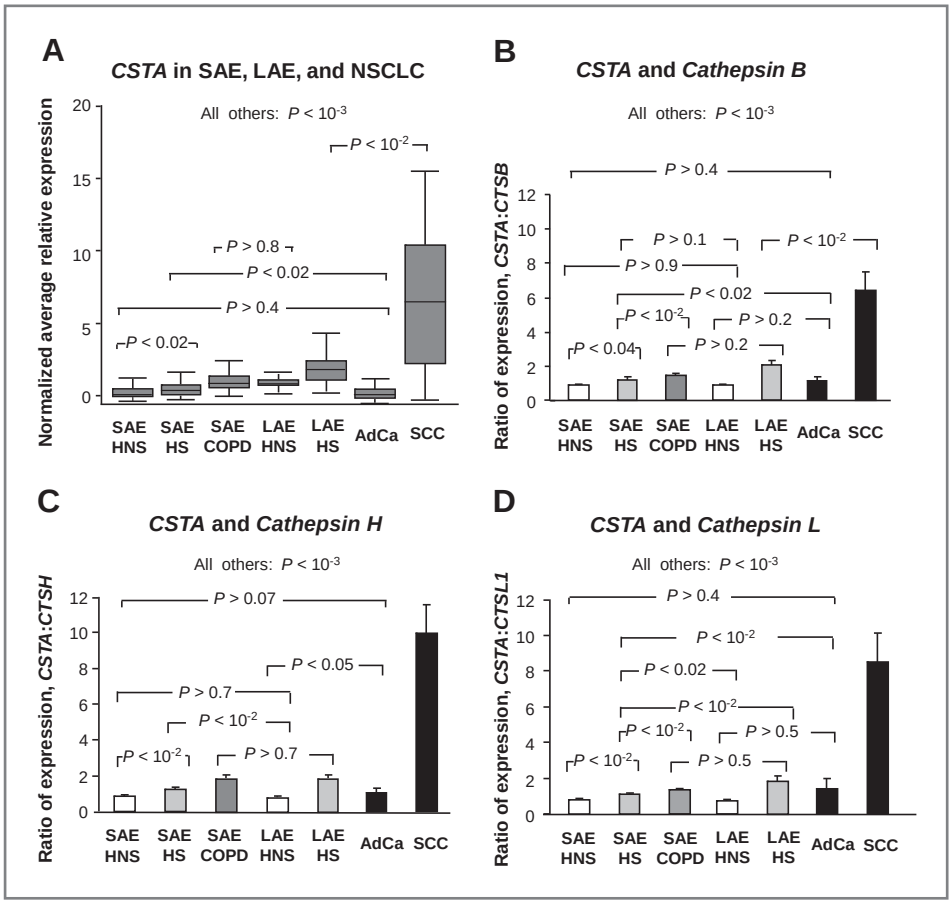

Figure 5. Comparison of *CSTA* gene expression in SAE, LAE, and in lung cancer. A, normalized average microarray *CSTA* gene expression levels are shown on the ordinate as box and whisker plots with median, interquartile range, and range for each of the groups depicted on the abscissa. B, ratio of microarray-determined gene expression level of *CSTA* to *cathepsin B* within individual subjects, for each of the 7 indicated phenotypic groups, including data from NSCLC. C, the ratio of microarray-derived SAE, LAE, and tumor *CSTA* levels to corresponding *cathepsin H* gene expression levels. D, shown on the ordinate is the same analysis as in B and C but with *cathepsin L* as the denominator. Error bars in B, C, and D represent standard error. *P* values are from ANOVA. HNS, healthy nonsmoker; HS, healthy smoker; COPD, smokers with chronic obstructive pulmonary disease; SAE, small airway epithelium; AdCa, adenocarcinoma of the lung; SCC, squamous cell carcinoma of the lung.

Cystatin A in lung disease

In the normal human airway epithelium, cystatin A is detected in the basal cells by immunohistochemistry, with more extensive staining in preneoplastic bronchial epithelium (12). Strong staining for cystatin A has also been observed in lung cancer tissue, especially SCC but also in adenocarcinoma and large cell carcinoma (11, 12). For SCC, cystatin A expression was less marked in less well-differentiated tumors, in which case lower cystatin A expression was also associated with tumor recurrence (12). Quantitative upregulation of *CSTA* protein has been noted in human lung cancer tissues versus control lung using ELISA, and univariate survival analyses showed that individuals with higher levels of cystatin A had a better survival probability, suggesting that *CSTA* is upregulated in lung tumors to counteract potentially harmful tumor-associated activity (11, 12). The present study confirms the previously published immunohistochemical data showing an upregulation of cystatin A in SCC versus normal airway epithelium, but also reveals a downregulation of *CSTA* in adenocarcinoma relative to levels in the healthy state, challenging data from another cohort that suggested an upregulation of cystatin A in some (but not in all) bronchogenic adenocarcinomas compared with normal airway epithelium (12).

Whereas all of the functions of *CSTA* are not clear, other than its known cysteine protease inhibitory effects, there is some evidence that it inhibits apoptosis in the presence of stimuli such as UVB radiation and viruses (4–6). Although deficiency states have been described arising from mutations in the genes that code for cystatins B and C, resulting in phenotypes of progressive myoclonus epilepsy and hereditary cystatin C amyloid angiopathy respectively, to date no functional mutations have been reported for *CSTA* in humans (44, 45).

The putative role of *CSTA* as a tumor suppressor is supported by observations in other tissues, including breast, prostate, and esophageal tumors, with evidence that cystatin A can inhibit tumor cell growth, angiogenesis, invasion, and metastasis (9, 36, 46). The observations in the present study that *CSTA* is downregulated in adenocarcinoma of the lung with evidence of less-well–opposed cysteine protease activity versus other smoke-exposed phenotypes are compatible with a tumor suppressor role in adenocarcinoma, but with a different function in the progression of some smokers to COPD and SCC, perhaps reflecting an "excess" of cystatin relative to its cognate proteases, the cathepsins. In contrast with these observations, a homozygous mutant mouse model involving chromosomal deletion of *csta*, the murine homolog of human *CSTA*, together with 3 other genes, showed no

phenotypic abnormality, the animals were not overtly susceptible to spontaneous or irradiation-induced tumor formation, and had evidence of compensatory gene expression of genes phylogenetically related to *Csta* (47).

Based on reports that cystatin A is upregulated in dysplastic airway epithelium and in lung cancer, the findings of the present study, which shows that the regulation of *CSTA* expression in SAE is influenced by genetic variability, smoking, and COPD, and that *CSTA* is differentially expressed by lung cancer histologic subtypes, suggests that each of these factors should be controlled for when considering the use of *CSTA* as a marker related to the pathogenesis of lung cancer. The progressive rise in *CSTA* expression observed in disease states (i.e., COPD and SCC) does not appear to be offset by a corresponding rise in cognate cathepsin levels in the same tissue from any given individual, an observation that may have implications for disease pathogenesis if intracellular cysteine protease-antiprotease homeostasis is relevant, particularly in genetically predisposed smokers.

Disclosure of Potential Conflicts of Interest

There are no potential conflicts of interest to disclose.

Acknowledgments

We thank N. Mohamed and T. Benios for help in preparing the manuscript.

Grant Support

These studies were supported, in part, by R01 HL074326, P50 HL084936, and UL1-RR024996.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received June 8, 2010; revised January 4, 2011; accepted February 2, 2011; published OnlineFirst February 16, 2011.

References

1. Barrett AJ. The cystatins: a diverse superfamily of cysteine peptidase inhibitors. Biomed Biochim Acta 1986;45:1363–74.
2. Takai T, Kato T, Hatanaka H, Inui K, Nakazawa T, Ichikawa S, et al. Modulation of allergenicity of major house dust mite allergens Der f 1 and Der p1 by interaction with an endogenous ligand. J Immunol 2009;183:7958–65.
3. Vray B, Hartmann S, Hoebeke J. Immunomodulatory properties of cystatins. Cell Mol Life Sci 2002;59:1503–12.
4. Bjorklund HV, Johansson TR, Rinne A. Rhabdovirus-induced apoptosis in a fish cell line is inhibited by a human endogenous acid cysteine proteinase inhibitor. J Virol 1997;71:5658–62.
5. Jones B, Roberts PJ, Faubion WA, Kominami E, Gores GJ. Cystatin A expression reduces bile salt-induced apoptosis in a rat hepatoma cell line. Am J Physiol 1998;275:G723-G730.
6. Takahashi H, Komatsu N, Ibe M, Ishida-Yamamoto A, Hashimoto Y, Iizuka H. Cystatin A suppresses ultraviolet B-induced apoptosis of keratinocytes. J Dermatol Sci 2007;46:179–87.
7. Calkins CC, Sloane BF. Mammalian cysteine protease inhibitors: biochemical properties and possible roles in tumor progression. Biol Chem Hoppe Seyler 1995;376:71–80.
8. Korolenko TA, Poteryaeva ON, Falameeva OV, Levina OA. Cystein proteinase inhibitor stefin A as an indicator of efficiency of tumor treatment in mice. Bull Exp Biol Med 2003;136:46–8.
9. Li W, Ding F, Zhang L, Liu Z, Wu Y, Luo A, et al. Overexpression of stefin A in human esophageal squamous cell carcinoma cells inhibits tumor cell growth, angiogenesis, invasion, and metastasis. Clin Cancer Res 2005;11:8753–62.
10. Rinne A, Jarvinen M, Rasanen O, Dorn A. [Demonstration of an epidermal SH-protease inhibitor in normal epithelium and in some human neoplasms—an immunological study (author's transl)]. Acta Histochem Suppl 1980;22:325–9.
11. Werle B, Schanzenbacher U, Lah TT, Ebert E, Jülke B, Ebert W, et al. Cystatins in non-small cell lung cancer: tissue levels, localization and relation to prognosis. Oncol Rep 2006;16:647–55.
12. Leinonen T, Pirinen R, Bohm J, Johansson R, Rinne A, Weber E, et al. Biological and prognostic role of acid cysteine proteinase inhibitor (ACPI, cystatin A) in non-small-cell lung cancer. J Clin Pathol 2007;60:515–9.
13. Petty RD, Nicolson MC, Kerr KM, Collie-Duguid E, Murray GI. Gene expression profiling in non-small cell lung cancer: from molecular mechanisms to clinical application. Clin Cancer Res 2004;10:3237–48.
14. Rosado-de-Christenson ML, Templeton PA, Moran CA. Bronchogenic carcinoma: radiologic-pathologic correlation. Radiographics 1994;14:429–46.
15. Toh CK. The changing epidemiology of lung cancer. Methods Mol Biol 2009;472:397–411.
16. Lokke A, Lange P, Scharling H, Fabricius P, Vestbo J. Developing COPD: a 25 year follow up study of the general population. Thorax 2006;61:935–9.
17. Rennard SI, Vestbo J. COPD: the dangerous underestimate of 15%. Lancet 2006;367:1216–9.
18. Barnes PJ, Shapiro SD, Pauwels RA. Chronic obstructive pulmonary disease: molecular and cellular mechanisms. Eur Respir J 2003;22:672–88.
19. Celli BR, MacNee W. Standards for the diagnosis and treatment of patients with COPD: a summary of the ATS/ERS position paper. Eur Respir J 2004;23:932–46.
20. 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;350:2645–53.
21. Kim V, Rogers TJ, Criner GJ. New concepts in the pathobiology of chronic obstructive pulmonary disease. Proc Am Thorac Soc 2008;5:478–85.
22. Cohen BH, Diamond EL, Graves CG, Kreiss P, Levy DA, Menkes HA, et al. A common familial component in lung cancer and chronic obstructive pulmonary disease. Lancet 1977;2:523–6.
23. Kassem J, Crystal RG. Chronic obstructive pulmonary disease and lung cancer. In:Schwab M, editor. Encyclopedia of Cancer. Germany: Springer-Verlag; 2008. p. 10.1007/978–3-540–47648-1_1157.
24. Koshiol J, Rotunno M, Consonni D, Pesatori AC, De Matteis S, Goldstein AM, et al. Chronic obstructive pulmonary disease and altered risk of lung cancer in a population-based case-control study. PLoS One 2009;4:e7380.
25. Tockman MS. Other host factors and lung cancer susceptibility. In: Samet JM, editor. Epidemiology of Lung Cancer (Lung Biology in Health and Disease).New York: Informa Healthcare; 1994. p. 397–412.
26. Turner MC, Chen Y, Krewski D, Calle EE, Thun MJ. Chronic obstructive pulmonary disease is associated with lung cancer mortality in a prospective study of never smokers. Am J Respir Crit Care Med 2007;176:285–90.
27. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Tobacco smoke and involuntary smoking. 2004;83:1–1438.
28. Khuder SA. Effect of cigarette smoking on major histological types of lung cancer: a meta-analysis. Lung Cancer 2001;31:139–48.
29. Purdue MP, Gold L, Jarvholm B, Alavanja MC, Ward MH, Vermeulen R. Impaired lung function and lung cancer incidence in a cohort of Swedish construction workers. Thorax 2007;62:51–6.

30. Wasswa-Kintu S, Gan WQ, Man SF, Pare PD, Sin DD. Relationship between reduced forced expiratory volume in one second and the risk of lung cancer: a systematic review and meta-analysis. Thorax 2005;60:570–5.
31. Kuner R, Muley T, Meister M, Ruschhaupt M, Buness A, Xu EC, et al. Global gene expression analysis reveals specific patterns of cell junctions in non-small cell lung cancer subtypes. Lung Cancer 2009;63:32–8.
32. Rabe KF, Hurd S, Anzueto A, Ruschhaupt M, Buness A, Xu EC, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. Am J Respir Crit Care Med 2007;176:532–55.
33. 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 2007;85:39–53.
34. Hackett NR, Heguy A, Harvey BG, O'Connor TP, Luettich K, Flieder DB, et al. Variability of antioxidant-related gene expression in the airway epithelium of cigarette smokers. Am J Respir Cell Mol Biol 2003;29:331–43.
35. Gene Expression Omnibus (GEO) site [last cited 2011 Jan 4]. Available from:http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=jjqvhwyeyegqmty&acc=GSE22047.
36. Parker BS, Ciocca DR, Bidwell BN, Gago FE, Fanelli MA, George J, et al. Primary tumour expression of the cysteine cathepsin inhibitor Stefin A inhibits distant metastasis in breast cancer. J Pathol 2008;214:337–46.
37. Niewoehner DE, Kleinerman J, Rice DB. Pathologic changes in the peripheral airways of young cigarette smokers. N Engl J Med 1974;291:755–8.
38. Willemse BW, ten Hacken NH, Rutgers B, Postma DS, Timens W. Association of current smoking with airway inflammation in chronic obstructive pulmonary disease and asymptomatic smokers. Respir Res 2005;6:38.
39. Brody JS, Spira A. State of the art. Chronic obstructive pulmonary disease, inflammation, and lung cancer. Proc Am Thorac Soc 2006;3:535–7.
40. Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, et al. Environmental and heritable factors in the causation of cancer–analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med 2000;343:78–85.
41. Webster PM, Lorimer EG, Man SF, Woolf CR, Zamel N. Pulmonary function in identical twins: comparison of nonsmokers and smokers. Am Rev Respir Dis 1979;119:223–8.
42. Chitale D, Gong Y, Taylor BS, Broderick S, Brennan C, Somwar R, et al. An integrated genomic analysis of lung cancer reveals loss of DUSP4 in EGFR-mutant tumors. Oncogene 2009;28:2773–83.
43. Bianchi F, Hu J, Pelosi G, Cirincione R, Ferguson M, Ratcliffe C, et al. Lung cancers detected by screening with spiral computed tomography have a malignant phenotype when analyzed by cDNA microarray. Clin Cancer Res 2004;10:6023–8.
44. Palsdottir A, Abrahamson M, Thorsteinsson L, Arnason A, Olafsson I, Grubb A, et al. Mutation in cystatin C gene causes hereditary brain haemorrhage. Lancet 1988;2:603–4.
45. Pennacchio LA, Lehesjoki AE, Stone NE, Willour VL, Virtaneva K, Miao J, et al. Mutations in the gene encoding cystatin B in progressive myoclonus epilepsy (EPM1). Science 1996;271:1731–4.
46. Mirtti T, Alanen K, Kallajoki M, Rinne A, Soderstrom KO. Expression of cystatins, high molecular weight cytokeratin, and proliferation markers in prostatic adenocarcinoma and hyperplasia. Prostate 2003;54:290–8.
47. Bilodeau M, MacRae T, Gaboury L, Laverdure JP, Hardy MP, Mayotte N, et al. Analysis of blood stem cell activity and cystatin gene expression in a mouse model presenting a chromosomal deletion encompassing *Csta* and Stfa2l1. PLoS One 2009;4:e7500.