Genotypes in matrix metalloproteinase 9 are a risk factor for COPD

Yohannes Tesfaigzi[1]
Orrin B Myers[2]
Christine A Stidley[2]
Kurt Schwalm[1]
Maria Picchi[1]
Richard E Crowell[3]
Frank D Gilliland[4]
Steven A Belinsky[1]

[1]Lovelace Respiratory Research Institute, Albuquerque, NM, USA; [2]Department of Internal Medicine, New Mexico Center for Environmental Health Sciences, University of New Mexico, Albuquerque, USA; [3]Department of Internal Medicine, University of New Mexico, and Pulmonary Section, New Mexico VA Health Care System, Albuquerque, NM, USA; [4]Department of Preventive Medicine, University of Southern California, Los Angeles, CA, USA

Correspondence: Yohannes Tesfaigzi
Lovelace Respiratory Research Institute, 2425 Ridgecrest Drive, SE, Albuquerque, NM 87108, USA
Tel +1 505 348 9495
Fax: +1 505 348 8567
Email ytesfaig@lrri.org

Abstract: A growing body of evidence indicates that matrix metalloproteinases (MMPs) play a role in the pathogenesis of COPD. Therefore, we conducted a candidate gene association study of 4 promoter polymorphisms that are known to modify expression levels of the MMP-1, MMP-2, and MMP-9 genes and a Gln279Arg polymorphism in exon 6 of MMP-9 that modifies the substrate-binding region. We examined the association of each variant and haplotypes in 385 male veterans with greater than 20 pack-years of cigarette smoking whose COPD status was characterized using spirometry. The association of these polymorphisms was also examined with decline of pulmonary function in a subset of participants. Only the 279Arg variant was more common in participants with COPD and the homozygous variant was associated with a 3-fold increased risk for COPD. In the haplotype analysis, the haplotype comprising the 249Arg and the CA promoter polymorphism within the MMP-9 gene was associated with risk, suggesting that either 279Arg or a linked variant on this haplotype underlies the association. No association of this polymorphism was found with decline in pulmonary function. These studies show that variants of the MMP-9 gene are associated with COPD in this cohort of veterans.

Keywords: metalloproteinase, smoking, pulmonary function, single nucleotide polymorphism, molecular epidemiology

Introduction

Genetic factors likely play a critical role in the development of COPD, because only a minority of heavy smokers develops this disease. The first genetic defect associated with COPD was a deficiency in α1-antitrypsin (Ganrot et al 1967). However, other genetic variations that affect the proteinase–antiproteinase imbalance (Evans and Pryor 1994) also likely affect the onset and development of COPD, because only <1% of COPD cases are associated with α1-antitrypsin deficiency (Sandford et al 2002).

A role for the matrix metalloproteinases (MMPs) has been suggested in the pathogenesis of pulmonary emphysema (Barnes 2000). Immunoreactivity for MMP-1, MMP-2, and MMP-9 is more intense in emphysematous samples compared with normal lung tissue (Ohnishi et al 1998; Segura-Valdez et al 2000). Alveolar macrophages obtained from smokers release more MMP-9 than those from non-smokers (Lim et al 2000).

There is an insertion–deletion polymorphism in the promoter region of the MMP-1 gene (1G-1607/2G). The addition of G in the MMP-1 promoter creates an Ets binding site, and leads to higher transcription of the protein compared with the common 1G polymorphism (Rutter et al 1998). The incidence of the 2G allele rises to 62% in tumor cells in ovarian and endometrial carcinomas (Rutter et al 1998). A single nucleotide polymorphism (SNP) in the MMP-2 promoter sequence (-1306 C→T) disrupts a SP1 site and results in lower promoter activity with the T allele (Price et al 2001), and participants homozygous for C have an overall 2-fold increased

risk for developing lung cancer compared with those with the CT or TT genotype (Yu et al 2002). We chose to study the association of these SNPs with COPD, because participants with COPD are at higher risk of developing lung cancer (Cohen et al 1977; Wu et al 1995).

Sequence analysis of the MMP-9 gene revealed a total of 10 variable sites, 4 in the promoter, 5 in the coding region (3 of which alter the amino acid encoded), and 1 in the 3′ untranslated region (Zhang et al 1999a). The variable length CA repeat located at −90 bp within the promoter region of MMP-9 may affect promoter activity and creates a sequence-specific DNA-binding protein site (Peters et al 1999). A C→T polymorphism at −1562 bp results in higher transcription rates for the T allele than the C allele promoter (Zhang et al 1999b) and dysregulation of MMP-9 function could increase susceptibility for smoking-induced airway obstruction. A polymorphism in exon 6 of MMP-9 is located in the substrate-binding domain of this enzyme and leads to a substitution of an uncharged amino acid (glutamine) by a positively charged amino acid (arginine). This polymorphism likely alters protein conformation, leading to a change in substrate-binding and enzyme activity, while the other two polymorphisms in the coding region that lead to altered amino acids are not likely to have such a profound change in enzyme activity (Zhang et al 1999a).

The major ethnic groups in New Mexico, the area from which the study population was recruited, consist of non-hispanic whites and hispanics. Therefore, the present study analyzed the prevalence of COPD in these ethnic groups and determined whether potentially functional variants that include the promoter regions of MMP-1, MMP -2, and MMP -9 and the Glu279Arg polymorphism within MMP-9 gene are associated with COPD in a cross-sectional study of veteran men and within the major ethnic groups.

Material and methods
Study population
Study participants were drawn from an ongoing cohort study of veterans (n=385). Participants were recruited from the New Mexico Veterans' Health Care System, which provides medical care for approximately 35 000 veterans in New Mexico, west Texas, and southern Colorado. Inclusion criteria included age 40–75 years and a minimum smoking history of ≥20 pack-years. Informed consent was obtained from each participant prior to completing a standardized clinical examination including spirometry and a standardized questionnaire that provided detailed information on demographics, smoking history, and respiratory health. Phlegm production was assessed by evaluating the St. George's Respiratory Disease Questionnaire. The same information was obtained at a follow-up visit at 18±3 months for 176 participants. This study was approved by the Institutional Review Boards of the Lovelace Respiratory Research Institute, the New Mexico Veterans Health Care System, and the University of New Mexico Health Sciences Center.

COPD definition
COPD was defined by criteria similar to those suggested as part of the GOLD criteria (Pauwels et al 2001) and were used in the Lung Health Study (Anthonisen et al 1994) and recently described by the ATS/ERS Task Force (Celli and MacNee 2004). We did not break down the cohort into various GOLD categories, because of small cell size for GOLD 1-4. However, three definitions for COPD were used: Definition I included all participants with $FEV_1/FVC<70\%$ as measured by spirometry. Definition II included only participants with $FEV_1/FVC<70\%$ and who had $FEV_1<80\%$ as cases and all other participants who were defined as COPD participants in Definition I were excluded from the analyses. Definition III included only participants with $FEV_1/FVC<70\%$ and $FEV_1<80\%$, excluding all who had reversibility of >15% post-bronchodialator exposure. Reversibility was defined as the percent change in percent predicted FEV_1 between pre- and post-bronchodialator exposures. For all three definitions, participants were considered to be non-COPD if $FEV_1/FVC≥70\%$. Participants who met the criteria for diagnosis of COPD at cohort entry were classified as prevalent cases. Participants with mild COPD were excluded in Definition II to reduce the potential for including participants who have "borderline" airflow obstruction and participants who have reversible airflow obstruction were excluded in Definition III to exclude the majority of participants who have asthma.

Genotyping of polymorphisms in MMP-9
Genomic DNA was isolated from peripheral blood lymphocytes by digestion with proteinase K in sodium dodecyl sulfate (1%), followed by standard phenol–chloroform extraction and ethanol precipitation as described (Sambrook et al 1989). The sequence flanking the polymorphic region of each SNP was amplified by PCR using 10 ng of DNA. Negative controls without DNA template were included with each set of reactions. Several

methods of genotyping were used to verify the obtained results. Restriction fragment length polymorphism (RFLP) was performed on 50–100 samples to verify the results obtained by the allelic discrimination assay. This assay integrates PCR and detection using allelic discrimination, which applies allele-specific TaqMan probes, a two-dye reporter system, and a reference dye. Wild-type and variant reporter sequences were discriminated using dual labeled reporters and by using mismatched sequences to validate the SNP-genotype. To further ensure quality control, DNA samples from 20 participants whose allelotypes were known were included at random within each of the 96-well plates of the allelic discrimination assay. In addition, approximately 10% of the samples were re-examined by a different person. All analyses were performed blinded with respect to participants' characteristics. These controls showed 100% concordance with expected results.

The methods of genotyping, including the primers and probes for allelic discrimination of SNPs in MMP-1, MMP-2, and MMP-9 or enzyme used for RFLP, are listed in Table 1. The C→T polymorphism was analyzed using a RFLP assay, because several attempts to design primers and probes for this polymorphism failed. The presence of the rare T allele introduces a restriction site for SphI, which is absent when the common C allele is present.

The number of CA repeats in the promoter region of MMP-9 was determined by PCR using a fluorescently labeled 6-FAM forward primer and an unlabeled reverse primer. The alleles were separated by a laser-based automated DNA sequencer, ABI PRISM 3100. The product peaks were visualized using GenescanView software (Bio Molecular Research, Padova, Italy). The size of the PCR products was calculated from a standard curve made from internal standards. DNA standards containing 14, 21, 22, and 23 CA repeats were identified by direct sequencing and

were included in each experiment as positive controls. For unknown reasons, the polymorphisms of certain samples (1%–5%) could not be assessed even upon repeated attempts, resulting in slight sample size variability for the various polymorphisms.

Statistical analysis

We examined summary information by COPD status for demographic variables (age, gender, ethnicity); clinical risk factors (current smoking status, duration and pack-years of cigarette smoking); and polymorphism status (homozygous-rare, heterozygous, homozygous-common). The chi-square test and Fisher's exact test were used to compare categorical variables, including gene polymorphisms between participants with and without COPD. Continuous variables such as age and smoking duration were compared using either the two-sample t-test or the Wilcoxon rank sum test. The small number of women (17 in total) enrolled in the larger cohort were excluded from the analysis cohort of 385 veteran men, but this exclusion and inclusion of these women did not change any of the conclusions.

Distributions of genotypes were assessed for Hardy-Weinberg equilibrium using a generalization of Fisher's exact test (Guo and Thompson 1992). All pairs of MMP-9 loci (T-1562, 279Arg, and CA repeats) were tested for linkage disequilibrium using likelihood tests and permutation procedures to estimate significance levels (Goudet et al 1996; Slatkin and Excoffier 1996) as implemented in PowerMarker v3.12 (Liu and Muse). The degree of disequilibrium was described using D′ (Lewontin 1964; Zapata et al 2001).

Logistic regression models were fit to estimate odds ratios (OR) for COPD and to allow adjustment for multiple covariates. We estimated effects for dominant, co-dominant, and recessive models for the variant alleles. Covariates in

Table 1 PCR primers and probes for detecting polymorphisms

Polymorphism	Primers	Probes
MMP-1 G/GG-1607	5′-GTTATGCCACTTAGATGAGGAAATTG (S)	5′-AGATAAGTCATAT**CC**TTTCT
	5′-CATAAACAATACTTCAGTATATCTTGGATTGA (AS)	5′-AGTCATATCTTT**C**TAATTAT
MMP-2 C/T -1306	5′-CCCTAAAACATTCCCCATA TTCC (S)	5′-CACTC**C**ACCTCTTTA
	5′-GTGACTTCTGAGC TGAGACCTGAA (AS)	5′-AGCACTC**T**ACCTCTT
MMP-9 C/T -1562	5′-GCCTGGCACATAGTAGGCCC (S)	No SphI site (C)
	5′-CTTCCTAGCCAGCCGGCATC (AS)	SphI site 300+260 bp (T)
MMP-9 Gln279Arg	5′-TCTCCCCCTTTCCCACATC (S)	5′-CTCTACACCC**G**GGACGGCAATG
	5′-TGTGCTGTCTCCGCCTTCT (AS)	5′-ACTCTACACCC**A**GGACGGCAATGC
MMP-9 CA repeat	5′-6FAM-AAGCCCTTTCTCATGCTGGTGC	Fragment analysis
	5′-TGACAGGCAAGTGCTGACTCAG	

NOTE: The bold, underlined letters are the single nucleotide polymorphisms designed in the probes for allelic discrimination.
Abbreviations: AS, antisense primer; S, sense primer.

all models included age, ethnicity, and smoking information (former or current, or the continuous forms of smoking duration and pack-years). Logistic regression models for each polymorphism were fitted and the indicators were included as predictors for only those polymorphisms at one region plus the covariates. Since there was linkage disequilibrium between some of the polymorphisms, the polymorphisms were examined separately. Interactions between the genotypes and each covariate, especially cigarette smoking (duration and pack-years), and interactions between the covariates also were investigated. To assess differences by ethnicity, models were fitted separately by ethnic group. Likelihood ratio tests were used to test models and Wald's chi-square tests were used to evaluate whether individual variables contributed to the model. Linear regression models were used to assess the association between the genetic predictor variables and the predicted FEV_1. Covariates and interactions were included in the modeling as done for the logistic regression models above. All standard statistical testing and modeling were conducted using Statistical Analysis System SAS v8.02. (SAS Institute, Inc., Cary, NC, USA). These analyses were done for each of the COPD definitions. Therefore, the analyses ranged from mild to moderate and severe cases as the outcome.

The association between estimated haplotype frequencies and COPD was evaluated with a two-stage process. First, MMP-9 haplotype frequencies based on T-1562, 279Arg, and CA repeats were estimated using Excoffier and Slatkin's maximum likelihood algorithm (Excoffier and Slatkin 1995). For each possible haplotype, h, an estimate is made of the number of copies of h contained in the true pair of haplotypes carried by that individual conditional on the observed genotype and assuming Hardy-Weinberg equilibrium. The expected haplotype frequencies are analogous to haplotype dosages and were used in the second step as independent variables in logistic regression models that also included adjustment for age, smoking duration, and smoking status. Haplotype dosages for each participant were calculated separately for non-hispanic whites and hispanics. For other and unknown ethnicity, haplotype doses were calculated based on the overall cohort. Second, logistic regression analyses were conducted as outlined previously.

Results

Description of study population

The demographics of the cohort are summarized by COPD status as per definition I, which includes mild to severe cases (Table 2). The prevalence of COPD was 32% and, as

Table 2 Summary of demographic and baseline variables plus information on baseline PFT for all participants and those with two visits

	All participants		Participants with 2 visits	
	COPD[a]	Non-COPD	COPD[a]	Non-COPD
N (%)	123(31.9)	262(68.1)	48 (27.3)	128 (72.7)
Age class (years) N (%)				
40–49	6 (4.9)	42 (16.0)	4 (8.3)	15 (11.7)
50–59	30 (24.4)	112 (42.7)	11 (22.9)	54 (42.2)
60–69	56 (45.5)	74 (28.2)	19 (40.0)	41 (32.0)
>=70	31 (25.2)	34 (13.0)	14 (29.2)	18 (14.1)
Ethnicity N (%)				
Non-hispanic white	90 (73.2)	162 (61.8)	38 (79.2)	83 (64.8)
Hispanic	20 (16.3)	61 (23.3)	5 (10.4)	28 (21.9)
Other	13 (10.6)	39 (14.9)	5 (10.4)	17 (13.3)
Current smoker yes/no	54/69	109/153	20/28	42/86
(%)	(43.9)	(41.6)	(41.7)	(32.8)
Smoking				
Duration (years)	40.4	33.7	40.2	33.0
average (min, max)	(5.0–60.0)	(10.0–60.0)	(6.0–60.0)	(10.0–60.0)
Smoking, pack-years	60.1	49.1	60.0	51.6
average (min, max)	(20.0–220.0)	(20.0–140.0)	(24.0–205.0)	(20.0–140.0)
Baseline %FEV_1 (average (min, max))	58.6 (19.0–99.0)	84.1 (40.0–120.0)	61.5 (33.0–87.0)	84.6 (57.0–113.0)

[a]Using COPD definition I.
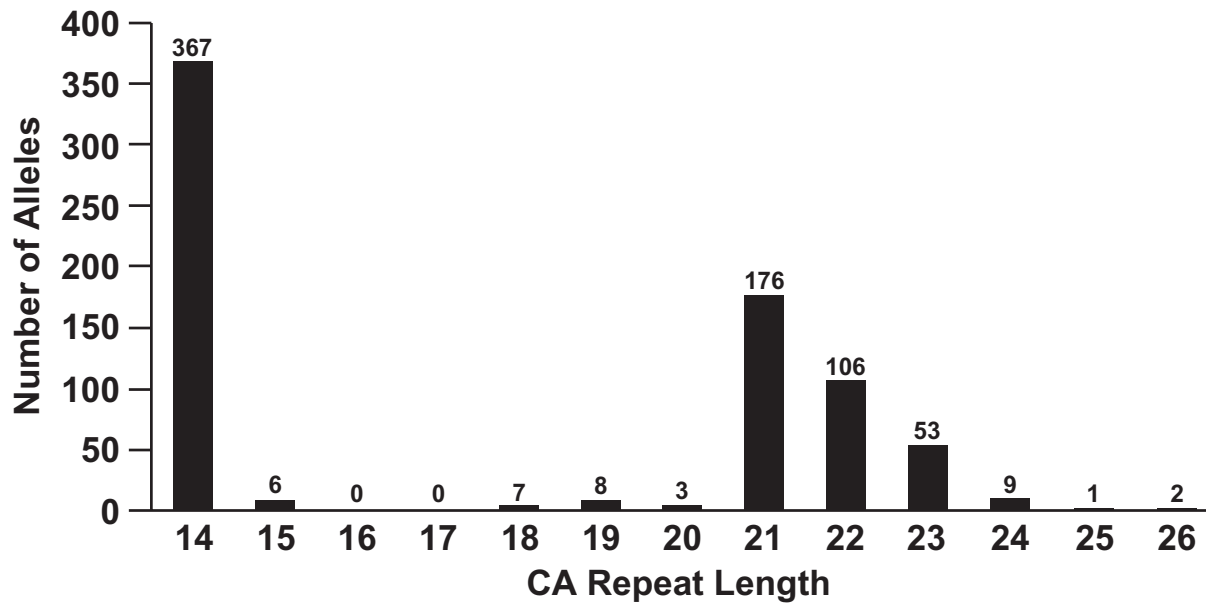
5587-4a

Figure I Distribution of CA repeat lengths.

expected for a disease that increases in frequency with age, overall the age of participants with COPD was older than for non-COPD participants. The median age for participants with and without COPD was 64 and 57 years (p<0.001), respectively. Interestingly, a higher proportion of non-hispanic white participants (36%) had COPD than hispanic participants (25%) (p=0.07). The number of participants included in analyses for COPD definitions II and III decreased to 366 and 339, respectively.

Average smoking duration and pack-years were greater for participants with COPD than for participants without COPD (p<0.001). Furthermore, the unadjusted OR for COPD was 2.1 and 2.0 for every 10 additional years of smoking duration and age, respectively (p<0.001). Finally, COPD was significantly associated with phlegm production when adjusted for age and smoking duration (OR=2.0, 95% CI=1.4 to 3.0; p<0.001). These basic demographic and clinical risk factors are similar to those reported in other studies of COPD (Zayas et al 1990) and reinforce the utility of our cohort for evaluating the role of polymorphisms within the MMP-9 gene as risk modifiers for COPD.

In general the percentages for COPD participants, ethnic distribution, smoking duration, and pack-years, and the baseline $\%FEV_1$ for the participants with a second visit remained identical to those for the overall cohort who came for the first visit (Table 2). Repeat testing was performed on these participants with 90% concordance of those initially classified with chronic airflow obstruction. These tests established the chronicity of airflow limitation through demonstration of $FEV_1/FVC<70\%$ by spirometry that was not reversed into the normal range after a bronchodilation, but had an overall tendency of decline over time (Pauwels et al 2001). No mean linear change in pulmonary function test was found in non-COPD participants within the participants with repeat measurements. However, there was a significant difference in yearly decline of $\%FEV_1$ for the COPD participants (2.6 (95% CI: 1.4–3.8)) compared with that of non-COPD participants (0.81 (95% CI: –0.01 to 1.62)) (p<0.02).

Distribution of genotypes

The SNPs were determined in DNA samples from 385 study participants. CA repeats of 14, 21, 22, and 23 were most frequently found while the frequency of other repeat lengths was below 2% (Figure 1). The CA repeat lengths that were most frequently found as genotypes were 14/14 (27%), 14/21 (20%), 14/22 (16%), 14/23 (8%), 21/21 (11%), and 22/22 (5%). The frequency of all other alleles was below 2% (Table 3). Because the frequency distribution of the CA repeat polymorphism showed a bimodal distribution of alleles, the alleles were divided into two subclasses: CA repeats ≤16 and ≥17 were designated as Short and Long, respectively, and treated similarly to genotype data.

The distributions of homozygotes and heterozygotes for each polymorphism in each ethnic group generally conformed to expectations under the assumption of Hardy-Weinberg equilibrium, with the exception of the CA-repeat among hispanic participants (p=0.02). As a measure for

Table 3 Distribution of CA repeat alleles

Allele I	Allele 2	Frequency	Percent
14	14	97	26.5
18	14	1	0.3
19	14	4	1.1
20	14	2	0.5
21	14	74	20.2
21	15	2	0.5
21	19	1	0.3
21	21	39	10.8
22	14	57	15.6
22	15	3	0.8
22	19	2	0.5
22	20	1	0.3
22	21	7	1.9
22	22	17	4.6
23	14	28	7.6
23	15	1	0.3
23	19	1	0.3
23	21	9	2.6
23	22	2	0.5
23	23	6	1.6
24	14	6	1.6
24	21	3	0.8
25	14	1	0.3
26	21	2	0.5

linkage disequilibrium, D′ was determined. The 279Arg was in linkage disequilibrium with both T-1562 (p<0.001) and CA repeat (p<0.001) within the ethnic groups. Of particular note, all participants with the homozygote rare T-1562 (N=20) had rare 279Arg. However, many participants who were homozygous for 279Arg had the C-1562 (Table 4). No differences among cases and controls were observed for any of the genotypes in MMP-1 and MMP-2.

A difference in prevalence of the 279Gln (35.0 vs 52.1, p<0.01) and 279Arg (18.7% vs 7.7%, p<0.01) homozygote genotypes was seen in participants with COPD compared with non-COPD participants (Table 4) and among the COPD and non-COPD participants at the second visit (see Appendix Table 4). Significantly different distributions of the homozygote rare 279Arg were also observed within the non-hispanic white participants (p<0.01). This trend in difference was recapitulated within the hispanic participants, but was not significant due to the lower number of hispanic participants with COPD. The rare T-1562 allele, the CA repeat lengths, and the SNPs in MMP-1 and MMP -2 were distributed similarly among the participants with and without COPD in the veterans' cohort (Table 4). The results for COPD definitions II and III remained identical, in that the distribution for the Gln279 was 52% compared with 7.7%

for the Arg279. For both COPD definitions II and III, the distribution for Gln279 remained at 47% in non-hispanic white participants compared with 9% for the rare Arg279.

Association of allelotype with COPD

The rare 279Arg variant was significantly associated with COPD (Table 5). This table gives the odds ratios which is the odds for COPD relative to the odds for those without COPD. After adjustment for age, smoking duration, and current smoking status, the OR for COPD participants who were homozygous for 279Arg variant was 3.8 relative to those who were homozygous for the common allele (p<0.01). In addition, participants who were heterozygous also showed a significant increased risk for the disease (OR=1.8; p<0.05), indicating a co-dominant association between 279Arg and COPD. This association was significant within the non-hispanic white but not within the hispanic participants, likely because of limited sample size for this ethnic group. Among the non-hispanic white participants, significant associations between COPD and the homozygous rare T-1562 (OR=4.5; p<0.05) and the Long/Long form of CA repeat (OR=2.3, p<0.05) were observed. Similar results were obtained for COPD definitions II and III (see Appendix Table 5). For example, for COPD definition II for the Gln279Arg the OR for the homozygote rare genotype remained significant at 3.13 (95% CI: 1.4–6.9) and in non-hispanic white participants at 2.8 (95% CI: 1.1–6.8). For COPD definition III, the OR was 3.1 (95% CI: 1.4–7.3) and 3.3 (95% CI: 1.3–8.4) overall and for non-hispanic whites, respectively. No association with COPD was found for the variants in MMP-1 and MMP -2. The similarity of results between the ethnic groups and among the definitions of COPD indicates that multiple comparison adjustments are not needed for these factors. All models showed that smoking variables, including duration, pack-years, and current status, were significantly associated with COPD, but none of the variables was a co-founder with the genetic variables.

Gene–gene and gene–environment interaction effects were evaluated by adding interaction terms in logistic regression models that included adjustments for age and smoking duration. To assess differences by ethnicity, analyses were done separately for non-hispanic white and hispanic participants. The gene–environment interactions used cigarette smoking measures as the environmental factors. We found no significant gene–gene or gene–environment interaction effects.

Association of SNPs with the decline in pulmonary function–longitudinal analyses

While we did not see a decline in pulmonary function over time when analyzing the data of the entire population, non-hispanic white participants and those who smoked more than 40 pack-years showed a statistically significant decrease per year when the raw or percentage difference of the baseline value in FEV_1 was analyzed. Non-hispanic whites showed an average decrease of 1.2 (95% CI: 0.5–1.9) or 1.7% (95% CI: 0.7–2.6), respectively, per year. Those who smoked ≥40 pack-years showed an average decrease of 1.6 (95% CI: 0.6–2.7) or 2.1% (95% CI: 0.8– 3.4%), respectively per year. However, we did not see any association of the polymorphisms tested with decline in pulmonary function assessed over 18 months, while the distribution of polymorphisms by COPD status was identical to that observed for all participants.

Haplotype analyses

Haplotype (common [C], rare [R], long [L], short [S]) frequencies formed from MMP-9 C/T-1562, MMP-9 Gln279Arg, and CA repeats were estimated for each cohort and ethnic group and evaluated for differences with respect to COPD status (Table 6). ORs adjusted for age, smoking duration, and current smoking status were calculated relative to the common C-1562, Gln279, and short CA repeats (C-C-S) for haplotypes with frequencies >0.005 (Table 6). The haplotypes with significant association with COPD were C-R-L (1.8, 95% CI, 1.1–2.9) and R-R-L (1.7, 95% CI, 1.1–2.7) compared with the C-C-S haplotype. Participants with the haplotype including the T-1562 and the rare 279Arg showed the same OR as participants with the haplotype, including the C-1562 and the rare 279Arg, indicating that either the 279Arg polymorphism is the main determinant or another variant on this RL haplotype may underlie the increased association by MMP-9 and risk for COPD. Models with and without covariates were compared using likelihood ratio tests. Models with haplotype dosages were no better

Table 4 Distribution of genotypes in COPD and non-COPD participants (%)

		All		Non-hispanic white		Hispanic	
		COPD	Non-COPD	COPD	Non-COPD	COPD	Non-COPD
MMP-9[a] Gln279Arg	C	35.0	52.1[**]	34.4	47.8[*]	40.0	67.2[*]
	H	46.3	40.2	43.3	42.9	55.0	31.2
	R	18.7	7.7[**]	22.2	9.3[**]	5.0	1.6
	n	123	261	90	161	20	61
MMP-9[a]C/T-1562	C	68.9	75.9	65.1	73.7	80.0	87.9
	H	26.0	21.7	27.9	23.7	20.0	12.1
	R	5.0	2.4	7.0	2.6	0.0	0.0
	n	119	253	86	156	20	58
MMP-9[b]CA repeat	SS	24.2	27.6	29.9	36.2	5.0	13.6
	LS	46.7	50.0	42.5	49.0	65.0	54.2
	LL	29.2	22.4	27.6	14.8[*]	30.0	32.2
	n	120	246	87	149	20	59
MMP-1[a]	C	26.2	27.2	23.6	19.8	45.0	42.6
	H	46.7	46.3	41.6	49.0	55.0	42.6
	R	27.1	26.5	34.8	31.2	0.0	14.8
	n	122	257	89	157	20	61
[a]MMP-2	C	62.3	55.0	62.9	53.2	60.0	57.4
	H	32.0	38.0	32.6	40.5	30.0	36.1
	R	5.7	7.0	4.5	6.3	10.0	6.6
	n	122	258	89	158	20	61

[a]For MMP-9 Gln279Arg, C-1562T, MMP-1, and MMP-2: C=common, H=heterozygote, R=rare.
[b]For MMP-9 CA repeat length: S=Short, L=Long.
Difference between COPD (by definition I) and non-COPD participants: ** p<0.01, *p<0.05.

Table 5 Association of genotypes with COPD[a]

		All	Non-hispanic white	Hispanic
MMP-9Gln279Arg	n	384	251	81
	Heterozyg.	1.8 (1.1, 3.0) *	1.7 (0.9, 3.1)	2.9 (0.9, 9.5)
	Homozyg. rare	3.8 (1.8, 7.9) **	3.5 (1.5, 8.1) *	6.9 (0.4, 137.5)
MMP-9C/T-1562	n	372	242	78
	Heterozyg.	1.3 (0.8, 2.3)	1.4 (0.7, 2.8)	1.7 (0.4, 7.7)
	Homozyg. rare	3.0 (0.8, 10.8)	4.5 (1.1, 10.4) *	NA[b]
MMP-9[c] CA repeat	n	366	236	79
	Long/Short	1.1 (0.6, 1.9)	1.2 (0.6, 2.2)	2.4 (0.2, 26.1)
	Long/Long	1.5 (0.8, 3.0)	2.3 (1.0, 5.1) *	2.4 (0.2, 29.0)
MMP-1	n	379	246	81
	Heterozyg.	1.2 (0.7, 2.2)	0.7 (0.3, 1.4)	2.09 (0.61, 7.15)
	Homozyg. rare	1.2 (0.7, 2.4)	0.9 (0.4, 2.0)	NA[b]
MMP-2	n	380	247	81
	Heterozyg.	0.8 (0.5, 1.4)	0.8 (0.4, 1.5)	0.7 (0.2, 2.4)
	Homozyg. rare	0.7 (0.3, 1.8)	0.6 (0.2, 2.4)	2.3 (0.3, 17.4)

[a]COPD definition I.
[b]not applicable due to zero participants with homozygous rare.
[c]OR (95% CI) relative to homozygous common and Short/Short CA repeats.
*p<0.05, ** p<0.001.
NOTE: Models include an adjustment for current smoking status and linear adjustments for age and smoking duration.

than covariate-only models for predicting COPD status of hispanic veterans. Haplotype information was useful for predicting COPD in non-hispanic white veterans; however, the only adjusted OR different from the reference haplotype, C-C-S, was R-R-L (p=0.02, OR=2.0, 95% CI=1.1–3.4).

Discussion

This cross-sectional study provides evidence that the tested polymorphisms in the MMP-9 gene are associated with susceptibility for onset of COPD in men. The increased risk for COPD may stem from the rare allele in exon 6, which replaces the non-charged glutamine (MMP-9Q) with the charged amino acid, arginine (MMP-9R).

MMP-9R could increase degradation of the ECM leading to emphysematous lungs. MMP-9 is expressed and secreted from cells as inactive zymogen through multiple mechanisms. This inactive form of MMP-9 is maintained by a cysteine and additional domains within the highly conserved region of the pro-peptide and the essential zinc atom in the catalytic domain (Park et al 1991). This tight regulation may be diminished in MMP-9R, especially in the presence of inflammatory responses that are induced by the presence of cigarette smoke. Alternatively, the affinity of MMP-9R to tissue inhibitors of metalloproteinases (TIMP)-1, the major inhibitor of MMP-9, may be lower

compared with that of MMP-9Q, because TIMP-1 forms a complex at the carboxy-terminus of the pro-enzyme and the catalytic domain of the active form (Goldberg et al 1992).

A previous study comparing smokers with COPD showed an association with a 5-year rate of decline with MMP-1 G-1607GG and haplotypes consisting of alleles from the MMP-1 G-1607GG and MMP-12 Asn357Ser polymorphisms (Joos et al 2002). In contrast, the CA repeat in MMP-9 was not associated with the rate of decline in

Table 6 MMP-9 C/T-1562, Gln279Arg, and CA repeat haplotype frequencies

Haplotype[b]	Relative frequency		Adjusted OR[a]	
	COPD[c]	Non-COPD	Estimate 95% CI	P value
C-C-S	0.47	0.53	1.0	
C-C-L	0.11	0.19	0.8 (0.5–1.2)	0.26
C-R-S	0.01	0.00		
C-R-L	0.23	0.15	1.8 (1.1–2.9)	0.02
R-C-S	0.00	0.00		
R-C-L	0.00	0.00		
R-R-S	0.00	0.03		
R-R-L	0.18	0.13	1.7 (1.1–2.7)	0.03

[a]Logistic model adjustments included age, smoking duration, and current smoking status.
[b]C-1562T (C=common, R=rare), Exon 6 (C=common, R=rare), CA repeat length (S=Short, L=Long).
[c]COPD definition I.

lung function. The differences in findings from our study may be because the Joos et al study was designed to focus on the rate of decline in lung function rather than overall risk for developing COPD. That study also did not include the Gln279Arg SNP of MMP-9. As expected, the frequencies of the A (65%) and G (35%) alleles in our study were very similar to those reported previously (Zhang et al 1999a). In general the frequencies of the alleles for each of the SNPs in this study were very similar to those reported in the NCBI website or in manuscripts (Zhang et al 1999a; Hinoda et al 2002).

The CA repeat length that was observed with high frequency in this cohort was 14 repeats. The second allele in these individuals had a CA repeat of 21, 22, or 23 repeats. The next most frequent alleles were between 21 and 22 CA repeats. While a similar distribution in CA repeats was observed in a Caucasian population (Joos et al 2002), this distribution of CA repeats is vastly different from that reported for 223 Japanese who had primarily MMP-9 alleles that contained 20 or more CA repeats. Only 2.8% of the Japanese population studied had 1 or 2 alleles with fewer than 20 CA repeats (Shimajiri et al 1999). This difference in the distribution of CA repeat length may be significant, because promoters with 21 CA repeats appeared to have a higher promoter activity compared with those with shorter repeats (Shimajiri et al 1999). In contrast, MMP-9 promoter containing 14 repeats had 53% promoter activity seen with 24 CA repeats (Huang et al 2003). Similarly, Peters et al (1999) found that the promoter with 23 CA repeats showed a 1.5-fold higher promoter activity compared with those containing 21 or 22 repeats. Consistent with these findings, non-hispanic white participants in particular with the Long/Long genotype had an increased risk for COPD compared with those with a Short/Short genotype in our study.

The -1562T genotype was significantly associated with increased risk for COPD in non-hispanic whites only. This finding supports the finding by Minematsu et al (2001) that the allelic frequency of the MMP-9 C-1562T promoter polymorphism was significantly increased in 45 smokers with emphysema compared with 65 matched smokers without emphysema, when emphysematous changes were scored from chest computer tomography scans. However, they did not find significant differences when COPD was characterized by FEV_1. Differences may be due to the association being specific for a specific ethnic group.

Non-hispanic whites were more likely to have COPD compared with hispanics and the decline in FEV_1 was more pronounced in these participants compared with hispanics.

In addition, all polymorphisms within MMP-9 that were tested were significantly associated with COPD in the non-hispanic whites. The -1562T genotype was associated with COPD among the non-hispanic whites, but the genotype is rare and was not observed in the hispanics. The association for 279Arg and Long/Long form of CA repeat was similar for both ethnic groups, but the result was not statistically significant for the hispanics, which may be a result of the smaller sample size. Collectively, these findings suggest that, while the distribution of the genotypes may vary by ethnic group, the association between genotype and COPD is not affected by ethnicity. Thus, confounding by population admixture is unlikely to account for the observed associations. Whether this association is also found for women with COPD will be addressed in the future.

Many studies have suggested the involvement of the MMPs in COPD (Ohnishi et al 1998; Vignola et al 1998; Segura-Valdez et al 2000; Imai et al 2001). Overexpression of MMP-1 results in emphysema in transgenic mice (D'Armiento et al 1992) and MMP-12 deficient mice are less susceptible to cigarette smoke-induced emphysema compared with wild-type mice (Hautamaki et al 1997). However, the relevance of these animal models to human disease has not been proven. Our study supports the hypothesis that MMP-9 may be a protease that is related to the pathogenesis of COPD and polymorphisms in MMP-9 may act as intrinsic factors to affect sensitivity to cigarette smoke-induced onset of COPD.

Acknowledgments

References

Anthonisen NR, Connett JE, Kiley JP, et al. 1994. Effects of smoking on intervention and use of an inhaled anticholinergic bronchodilator on the rate of decline of FEV1. *JAMA*, 272:1497–505.

Barnes PJ 2000. Chronic obstructive pulmonary disease. *N Engl J Med*, 343:269–80.

Celli BR, MacNee W. 2004. Standards for the diagnosis and treatment of patients with COPD: a summary of the ATS/ERS position paper. *Eur Respir J*, 23:932–46.

Cohen BH, Diamond EL, Graves CG, et al. 1977. A common familial component in lung cancer and chronic obstructive pulmonary disease. *Lancet*, 2:523-6.

D'Armiento J, Dalal SS, Okada Y, et al. 1992. Collagenase expression in the lungs of transgenic mice causes pulmonary emphysema. *Cell*, 71:955–61.

Evans MD, Pryor WA. 1994. Cigarette smoking, emphysema, and damage to alpha 1-proteinase inhibitor. *Am J Physiol*, 266: L593–611.

Excoffier L, Slatkin M. 1995. Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol Biol and Evolution*, 12:921–7.

Ganrot PO, Laurell CB, Eriksson S. 1967. Obstructive lung disease and trypsin inhibitors in alpha-1-antitrypsin deficiency. *Scand J Clin Lab Invest*, 19:205–8.

Goldberg GI, Strongin A, Collier IE, et al. 1992. Interaction of 92-kDa type IV collagenase with the tissue inhibitor of metalloproteinases prevents dimerization, complex formation with interstitial collagenase, and activation of the proenzyme with stromelysin. *J Biol Chem*, 267:4583–91.

Goudet J, Raymond M, Meeüs D, et al. 1996. Testing differentiation in diploid populations. *Genetics*, 144:1933–40.

Guo S, Thompson E. 1992. Performing exact tests of Hardy-Weinberg proportion for multiple alleles. *Biometrics*, 48:361–72.

Hautamaki RD, Kobayashi DK, Senior RM, et al. 1997. Requirement for macrophage elastase for cigarette smoke-induced emphysema in mice. *Science*, 277:2002–4.

Hinoda Y, Okayama N, Takano N, et al. 2002. Association of functional polymorphisms of matrix metalloproteinase (MMP)-1 and MMP-3 genes with colorectal cancer. *Int J Cancer*, 102:526–9.

Huang TS, Lee CC, Chang AC, et al. 2003. Shortening of microsatellite deoxy(CA) repeats involved in GL331-induced down-regulation of matrix metalloproteinase-9 gene expression. *Biochem Biophys Res Commun*, 300:901–7.

Imai K, Dalal SS, Chen ES, et al. 2001. Human collagenase (matrix metalloproteinase-1) expression in the lungs of patients with emphysema. *Am J Respir Crit Care Med*, 163:786–91.

Joos L, He JQ, Shepherdson MB, et al. 2002. The role of matrix metalloproteinase polymorphisms in the rate of decline in lung function. *Hum Mol Genet*, 11:569–76.

Lewontin RC. 1964. The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics*, 49: 49–67.

Lim S, Roche N, Oliver BG, et al. 2000. Balance of matrix metalloprotease-9 and tissue inhibitor of metalloprotease-1 from alveolar macrophages in cigarette smokers. Regulation by interleukin-10. *Am J Respir Crit Care Med*, 162:1355–60.

Liu K, Muse S. Free program distributed by the author over the internet. URL: http://www.powermarker.net.

Minematsu N, Nakamura H, Tateno H, et al. 2001. Genetic polymorphism in matrix metalloproteinase-9 and pulmonary emphysema. *Biochem Biophys Res Commun*, 289:116–19.

Ohnishi K, Takagi M, Kurokawa Y, et al. 1998. Matrix metalloproteinase-mediated extracellular matrix protein degradation in human pulmonary emphysema. *Lab Invest*, 78: 1077–87.

Park AJ, Matrisian LM, Kells AF, et al. 1991. Mutational analysis of the transin (rat stromelysin) autoinhibitor region demonstrates a role for residues surrounding the "cysteine switch". *J Biol Chem*, 266:1584–90.

Pauwels RA, Buist AS, Calverley PM, et al. 2001. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease. NHLBI/WHO Global Initiative for Chronic Obstructive Lung Disease (GOLD) Workshop summary. *Am J Respir Crit Care Med*, 163:1256–76.

Peters DG, Kassam A, St Jean PL, et al. 1999. Functional polymorphism in the matrix metalloproteinase-9 promoter as a potential risk factor for intracranial aneurysm. *Stroke*, 30:2612–16.

Price SJ, Greaves DR, Watkins H. 2001. Identification of novel, functional genetic variants in the human matrix metalloproteinase-2 gene: role of Sp1 in allele-specific transcriptional regulation. *J Biol Chem*, 276:7549–58.

Rutter JL, Mitchell TI, Buttice G, et al. 1998. A single nucleotide polymorphism in the matrix metalloproteinase-1 promoter creates an Ets binding site and augments transcription. *Cancer Res*, 58:5321–5.

Sambrook J, Fritsch EF, Maniatis T. 1989. Molecular Cloning. A Laboratory Manual. 2nd ed. New York: Cold Spring Harbor Press.

Sandford AJ, Joos L, Pare PD. 2002. Genetic risk factors for chronic obstructive pulmonary disease. *Curr Opin Pulm Med*, 8:87–94.

Segura-Valdez L, Pardo A, Gaxiola M, et al. 2000. Upregulation of gelatinases A and B, collagenases 1 and 2, and increased parenchymal cell death in COPD. *Chest*, 117:684–94.

Shimajiri S, Arima N, Tanimoto A, et al. 1999. Shortened microsatellite d(CA)21 sequence down-regulates promoter activity of matrix metalloproteinase 9 gene. *FEBS Letts*, 455:70–4.

Slatkin M, Excoffier L. 1996. Testing for linkage disequilibrium using the EM algorithm. *Heredity*, 76:377–83.

Vignola AM, Riccobono L, Mirabella A, et al. 1998. Sputum metalloproteinase-9/tissue inhibitor of metalloproteinase-1 ratio correlates with airflow obstruction in asthma and chronic bronchitis. *Am J Respir Crit Care Med*, 158:1945–50.

Wu AH, Fontham ET, Reynolds P, et al. 1995. Previous lung disease and risk of lung cancer among lifetime nonsmoking women in the United States. *Am J Epidemiol*, 141:1023–32.

Yu C, Pan K, Xing D, et al. 2002. Correlation between a single nucleotide polymorphism in the matrix metalloproteinase-2 promoter and risk of lung cancer. *Cancer Res*, 62:6430–3.

Zapata C, Carollo C, Rodriguez S. 2001. Sampling variance and distribution of the D' measure of overall gametic disequilibrium between multiallelic loci. *Ann Human Gen*, 65:395–406.

Zayas JG, Man GC, King M. 1990. Tracheal mucus rheology in patients undergoing diagnostic bronchoscopy. Interrelations with smoking and cancer. *Am Rev Respir Dis*, 141:1107–13.

Zhang B, Henney A, Eriksson P, et al. 1999a. Genetic variation at the matrix metalloproteinase-9 locus on chromosome 20q12.2-13.1. *Hum Genet*, 105:418–23.

Zhang B, Ye S, Herrmann SM, et al. 1999b. Functional polymorphism in the regulatory region of gelatinase B gene in relation to severity of coronary atherosclerosis. *Circulation*, 99:1788–94.

Appendix

Table 4 Distribution of polymorphisms by COPD status (definition I)

		All participants		Participants with 2 visits	
		COPD	**Non-COPD**	**COPD**	**Non-COPD**
MMP-9[a] Gln279Arg	C	35.0	52.1[**]	35.4	52.3
	H	46.3	40.2	37.5	41.4
	R	18.7	7.7 [**]	27.1	6.3
	n	123	261	48	128
MMP-9[a]C/T-1562	C	68.9	75.9	60.9	75.8
	H	26.0	21.7	32.6	21.8
	R	5.0	2.4	6.5	2.4
	n	119	253	46	124
MMP-9[b]CA repeat	SS	24.2	27.6	29.8	26.7
	LS	46.7	50.0	34.0	52.5
	LL	29.2	22.4	36.2	20.8
	n	120	246	47	120
MMP-1[a]	C	26.2	27.2	22.9	25.0
	H	46.7	46.3	47.9	46.9
	R	27.1	26.5	29.2	28.1
	n	122	257	48	128
MMP-2[a]	C	62.3	55.0	58.3	50.0
	H	32.0	38.0	33.3	42.2
	R	5.7	7.0	8.3	7.8
	n	122	258	48	128

[a]For MMP-9 Gln279Arg, C-1562T, MMP-1, and MMP-2: C=common, H=heterozygote, R=rare.
[b]For MMP-9 CA repeat length: S=Short, L=Long.
Difference between COPD (by definition I) and non-COPD participants: ** $p<0.01$, *$p<0.05$.

Table 5 Association of genotypes with COPD
COPD definition II

		All	Non-hispanic white	Hispanic
Gln279Arg	n	366	238	76
	Heterozyg.	2.04 (1.19, 3.49)*	1.75 (0.92, 3.35)	4.93 (1.26, 19.28)
	Homozyg. rare	3.13 (1.42, 6.94)**	2.77 (1.13, 6.81)*	12.61 (0.58, 275.15)
C/T-1562	n	354	229	73
	Heterozyg.	1.37 (0.77, 2.43)	1.48 (0.74, 2.96)	2.47 (0.55, 11.07)
	Homozyg. rare	2.24 (0.54, 9.30)	3.58 (0.72, 17.69)	NA[a]
CA repeat	n	348	223	74
	Long-Short	1.35 (0.72, 2.52)	1.23 (0.62, 2.45)	NA[a]
	Long-Long	2.00 (0.97, 4.12)	1.98 (0.85, 4.63)	NA[a]
MMP1	n	361	233	76
	Heterozyg.	1.28 (0.69, 2.36)	0.69 (0.32, 1.50)	3.19 (0.78, 12.30)
	Homozyg. rare	1.10 (0.54, 2.24)	0.85 (0.37, 1.93)	NA[a]
MMP2	n	362	234	76
	Heterozyg.	0.85 (0.50, 1.43)	0.86 (0.46, 1.60)	0.61 (0.15, 2.41)
	Homozyg. rare	0.70 (0.25, 1.99)	0.78 (0.21, 2.90)	1.15 (0.10, 13.96)
				(Continued over)

[a]not applicable due to zero participants with homozygous rare.
[b]OR (95% CI) relative to homozygous common and Short/Short CA repeats.
* $p<0.05$, ** $p<0.001$.
NOTE: Models include an adjustment for current smoking status and linear adjustments for age and smoking duration.

Table 5 (Continued)

COPD definition III

		All	Non-hispanic white	Hispanic
Gln279Arg	n	339	217	74
	Heterozyg.	1.78 (0.98, 3.23)	1.36 (0.65, 2.81)	4.08 (1.02, 16.27)
	Homozyg. rare	3.14 (1.35, 7.30)**	3.32 (1.31, 8.39)*	10.75 (0.50, 230.22)
C/T-1562	n	327	208	71
	Heterozyg.	1.31 (0.68, 2.53)	1.38 (0.63, 3.04)	2.22 (0.45, 10.89)
	Homozyg. rare	3.25 (0.78, 13.62)	5.08 (1.02, 25.18)*	NA[b]
CA repeat	n	321	202	72
	Long-Short	1.29 (0.64, 2.61)	1.01 (0.46, 2.20)	NA[a]
	Long-Long	2.08 (0.94, 4.57)	2.16 (0.87, 5.36)	NA[a]
MMP1	n	334	212	74
	Heterozyg.	1.31 (0.67, 2.54)	0.65 (0.29, 1.49)	3.17 (0.75, 13.41)
	Homozyg. rare	0.75 (0.33, 1.71)	0.55 (0.22, 1.39)	NA[a]
MMP2	n	335	213	74
	Heterozyg.	0.83 (0.46, 1.51)	0.84 (0.42, 1.70)	0.77 (0.19, 3.12)
	Homozyg. rare	1.03 (0.36, 2.94)	1.16 (0.31, 4.39)	1.38 (0.11, 16.86)

[a]not applicable due to zero participants with homozygous rare.

[b]OR (95% CI) relative to homozygous common and Short/Short CA repeats.

* p<0.05, ** p<0.001.

NOTE: Models include an adjustment for current smoking status and linear adjustments for age and smoking duration.