Matrix Metalloproteinase 1 (MMP1) Is Associated with Early-Onset Lung Cancer

Wiebke Sauter,1,9 Albert Rosenberger,2 Lars Beckmann,3 Silke Kropp,3 Kirstin Mittelstrass,1 Maria Timofeeva,4 Gabi Wölke,1 Angelika Steinwachs,1 Daniela Scheiner,1 Eckart Meese,5 Gerhard Sybrecht,6 Florian Kronenberg,7 Hendrik Dienemann,8 The LUCY-Consortium, Jenny Chang-Claude,3 Thomas Illig,1 Heinz-Erich Wichmann,1,9 Heike Bickeböller,2 and Angela Risch4

1Institute of Epidemiology, GSF-National Research Center for Environment and Health, Neuherberg, Germany; 2Department of Genetic Epidemiology, Georg-August University of Göttingen, Medical School, Göttingen, Germany; 3Division of Clinical Epidemiology, German Cancer Research Centre (DKFZ), Heidelberg, Germany; 4Division of Toxicology and Cancer Risk Factors, German Cancer Research Centre (DKFZ), Heidelberg, Germany; 5Institute of Human Genetics, University of the Saarland, Saarbrücken, Germany; 6Institute of Internal Medicine V, Medical School, University of the Saarland, Saarbrücken, Germany; 7Division of Genetic Epidemiology, Department of Medical Genetics, Molecular, and Clinical Pharmacology, Innsbruck Medical University, Innsbruck, Austria; 8Thoraxklinik, Heidelberg, Germany; 9Department of Medical Informatics, Biometrics, and Epidemiology (IBE), University of Munich, Munich, Germany

Abstract

Matrix metalloproteinases (MMP) play a key role in the breakdown of extracellular matrix and in inflammatory processes. MMP1 is the most highly expressed interstitial collagenase degrading fibrillar collagens. Over-expression of MMP1 has been shown in tumor tissues and has been suggested to be associated with tumor invasion and metastasis. Nine haplotype tagging and additional two intronic single nucleotide polymorphisms (SNP) of MMP1 were genotyped in a case control sample, consisting of 635 lung cancer cases with onset of disease below 51 years of age and 1,300 age- and sex-matched cancer-free controls. Two regions of linkage disequilibrium (LD) of MMP1 could be observed: a region of low LD comprising the 5′ region including the promoter and a region of high LD starting from exon 1 to the end of the gene and including the 3′ flanking region. Several SNPs were identified to be individually significantly associated with risk of early-onset lung cancer. The most significant effect was seen for rs1938901 (P = 0.0089), rs193008 (P = 0.0108), and rs996999 (P = 0.0439). For rs996999, significance vanished after correction for multiple testing. For each of these SNPs, the major allele was associated with an increase in risk with an odds ratio between 1.2 and 1.3 (95% confidence interval, 1.0-1.5). The haplotype analysis supported these findings, especially for subgroups with high smoking intensity. In summary, we identified MMP1 to be associated with an increased risk for lung cancer, which was modified by smoking. (Cancer Epidemiol Biomarkers Prev 2008;17(5):1127–35)

Introduction

Lung cancer is the most common cancer with 1,350,000 new cases per year worldwide. Lung cancer mostly occurs in individuals of ages ≥60 years and constitutes a highly lethal disease with an average 5-year survival rate of 10% (1). The most important risk factor for lung cancer is smoking, with evidence of a strong dose-response relationship between smoking and lung cancer risk (2). In tobacco smoke, there are ~7,000 substances including a multitude of carcinogens that induce a variety of DNA damages. Although smoking is considered to be the predominant risk factor, only ~10% of heavy smokers develop lung cancer. This suggests that genetic variation in sensitivity to carcinogen exposure may play an important role in the etiology of lung cancer (3).

Family aggregation and increased familial risk for lung cancer have been reported in several studies, providing indirect evidence that genetic factors contribute to susceptibility to lung cancer (4-6). Bailey-Wilson et al. (7) reported a region on chromosome 6q23-25 linked to familial lung cancer. The risk significantly increases with having an affected relative and further with earlier age of onset of the disease (4, 8-10).

Candidate susceptibility genes for lung cancer have been extensively studied, with most of the work focusing on mechanistically plausible variants in genes coding for enzymes involved in the activation and detoxification of carcinogens and repair of damage caused by tobacco smoke. Alterations in these pathways are hypothesized to affect an individual’s processing of tobacco carcinogens and, therefore, the risk of developing lung cancer. Inflammation has been thought to play a role in carcinogenesis of a variety of cancers, including lung cancer, and may be due to infectious agents or other...
environmental exposures such as tobacco smoke (11). Genes involved in the regulation of the inflammatory response were recently added to the lung cancer candidate genes in literature, and some of these focused genes belong to the family of matrix metalloproteinase (MMP).

MMPs comprise a structurally and functionally related family of zinc metalloproteinases degrading extracellular matrix and basement membrane barriers, and thus are thought to play a key role in angiogenesis, inflammatory processes, cancer development, cell proliferation, and apoptosis (12). The activity of MMPs is regulated at the level of transcription, activation, and inhibition by tissue inhibitors of metalloproteinases (TIMP; ref. 13). In this context, MMPs have been focused on as targets for therapeutic strategies. Because of their role in the degradation of the extracellular matrix leading to tumor invasion and metastasis, they may also serve as prognostic markers (14, 15).

Members of the MMP family have different substrate specificities and expression patterns. Among the MMPs, MMP1 is the most highly expressed interstitial collagenase degrading fibrillar collagens, which are major constituents of the extracellular matrix.

MMP1 is up-regulated in a wide variety of advanced cancers and has been suggested to be associated with tumor invasion and metastasis (16). A significant negative correlation between its expression and cancer survival has been found (17). Overexpression of MMP1 has been reported in lung cancer cells (18, 19) and several association studies between DNA variants of MMP1 and lung cancer have previously been conducted (20-22).

The level of MMP1 expression can be influenced by single nucleotide polymorphisms (SNP), especially, but not exclusively, if they are located within the promoter region of the MMP1 gene.

The aim of this case-control study of early-onset lung cancer was a detailed investigation of the genetic variability within the MMP1 gene in relation to early-onset lung cancer using individual SNP and haplotype analysis and considering potential effect modifications by the amount of smoke exposure.

Materials and Methods

Study Population. The present case-control study in Caucasians included 635 primary lung cancer patients being diagnosed before the age of 51 y, and 1,300 cancer-free individuals matched for sex and age. The control sample was collected within the population-based KORA study. Cases were selected either from the LUCY study or the Heidelberg lung cancer study.

The LUCY study (Lung Cancer in the Young) is a multicenter study with 31 participating clinics all over Germany. Only newly diagnosed patients with histologically or cytologically confirmed primary lung cancer entered the study. Detailed epidemiologic data on family history, tobacco and smoking exposure, education, and occupational exposure have been collected and blood samples were taken.

The Heidelberg lung cancer study is an ongoing hospital-based case-control study (23). The German Cancer Research Center (DKFZ) has recruited more than 1,000 lung cancer cases at and in collaboration with the Thoraxklinik Heidelberg. Thereof, 163 lung cancer cases with onset of disease before age 51 y, recruited between 01/1997 and 12/2003, were included in the analysis. Data on occupational exposure, tobacco smoking, and educational status, and for a subgroup also data on family history of lung cancer, were assessed with a self-administered questionnaire.

The KORA study (Cooperative Health Research in the Augsburg Region) is a population-based epidemiologic survey of individuals living in or near the city of Augsburg, Southern Germany (24, 25). With the genotyped polymorphisms to date, a major population stratification between KORA (Southwest Germany) and two other cohorts from Northern Germany could not be detected in a genomic control approach (26).

Informed consent was obtained from all study participants and the studies were approved by the ethics committee of the Bayerische Landesärztekammer, the corresponding local ethics committees of the participating clinics, and the ethics committee of the University of Heidelberg.

Concerning tobacco exposure, cases and controls were classified as never smokers, former smokers if they had quit for at least 1 y at the time of diagnosis/interview, and current smokers. Individuals who smoked <1 pack-year were considered as never smokers (Table 1). Smoking status was quantified in pack-years and subjects were grouped into four categories: very light smokers (0-10 pack-years), light smokers (11-20 pack-years), moderate smokers (21-30 pack-years), and heavy smokers.

Table 1. Demographic characteristics of cases and controls

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Cases (n = 635)</th>
<th>Controls (n = 1,300)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (y), mean ± SD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Males</td>
<td>45.4 ± 4.1</td>
<td>45.2 ± 4.3</td>
</tr>
<tr>
<td>Females</td>
<td>44.7 ± 4.6</td>
<td>44.7 ± 4.7</td>
</tr>
<tr>
<td>Gender, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>406 (64)</td>
<td>819 (63)</td>
</tr>
<tr>
<td>Female</td>
<td>229 (36)</td>
<td>481 (37)</td>
</tr>
<tr>
<td>Smoking status, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never smoker</td>
<td>12 (3)</td>
<td>243 (30)</td>
</tr>
<tr>
<td>Former smoker</td>
<td>69 (17)</td>
<td>249 (30)</td>
</tr>
<tr>
<td>Current smoker</td>
<td>322 (79)</td>
<td>325 (40)</td>
</tr>
<tr>
<td>Unknown</td>
<td>3 (1)</td>
<td>1 (0)</td>
</tr>
<tr>
<td>Female</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never smoker</td>
<td>29 (13)</td>
<td>207 (43)</td>
</tr>
<tr>
<td>Former smoker</td>
<td>41 (18)</td>
<td>128 (27)</td>
</tr>
<tr>
<td>Current smoker</td>
<td>158 (69)</td>
<td>147 (30)</td>
</tr>
<tr>
<td>Unknown</td>
<td>1 (0)</td>
<td></td>
</tr>
<tr>
<td>Smoking categories (male + female), n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weak smokers (≤20 pack-years)</td>
<td>168 (26)</td>
<td>885 (68)</td>
</tr>
<tr>
<td>Strong smokers (≥20 pack-years)</td>
<td>452 (71)</td>
<td>385 (30)</td>
</tr>
<tr>
<td>Unknown</td>
<td>15 (2)</td>
<td>30 (2)</td>
</tr>
<tr>
<td>Histology, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AC</td>
<td>222 (35)</td>
<td></td>
</tr>
<tr>
<td>SCLC</td>
<td>153 (24)</td>
<td></td>
</tr>
<tr>
<td>Other NSCLC</td>
<td>218 (34)</td>
<td></td>
</tr>
<tr>
<td>Other LC</td>
<td>42 (7)</td>
<td></td>
</tr>
</tbody>
</table>

NOTE: Categorical and quantitative variables are reported in absolute and relative frequencies. Metric variables are reported as mean ± SD. Other NSCLC, other non–small-cell lung cancers, excluding adenocarcinoma.

Abbreviations: AC, adenocarcinoma; SCLC, small-cell lung cancer; NSCLC, non–small-cell lung cancer; LC, lung cancer.
smokers (≥31 pack-years). The subjects were also grouped into two sets defined by weak smokers (very light and light) and strong smokers (moderate and intensive).

Histologic subtypes of cancer were grouped into small-cell carcinoma, adenocarcinoma, other non–small-cell lung carcinoma, and other types of lung cancer (Table 1).

Genotyping Methods and SNP Selection. DNA was extracted from EDTA anticoagulated blood using a commercially available kit (Gentra) according to the manufacturer’s protocol. For the Heidelberg lung cancer study, buffy coats from 5 mL of venous blood in EDTA were stored at −80°C, and human genomic DNA was isolated with the QIAamp DNA blood midi kit according to the manufacturer’s instructions (Qiagen GmbH).

Based on the data of HapMap phase II (release #20; ref. 27), a genomic region of 18.239 kb on chromosome 11q22-23 containing the entire MMP1 gene plus 5,000 bp each upstream (3’ untranslated region) and downstream (5’ untranslated region) was selected. Nine haplotype tagging SNPs were identified using the following tagging criteria: pairwise tagging of the HapMap CEU Population with r² ≥ 0.8 and a minor allele frequency ≥20%. To increase SNP density, two intronic SNPs (rs5031036 and rs193890) were selected from the literature and public databases (Fig. 1).

SNPs were genotyped using iPLEX single base primer extension and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (Sequenom) as recently described (28). Genotyping was done by laboratory personnel blinded to case-control status. Standard genotyping quality control included 10% duplicate samples and negative samples. Hardy-Weinberg equilibrium was tested in cases and controls separately using a log-likelihood ratio test. Pairwise linkage disequilibrium (LD) between the 11 SNPs in MMP1 was measured by D’ and r², and was calculated based on the genotypes of 1,300 control subjects.

Statistical Analysis. Single SNP association tests were calculated using conditional logistic regression. In the general model, two binary dummy variables, indicating heterozygous and homozygous predisposing alleles, were included into the logistic regression. Three models, dominant, recessive, and multiplicative, indicating the increase of risk for one additional predisposing allele, were fitted as well.

Cases and controls were frequency matched for gender and age at diagnosis or interview (≥40, 41-45, and 46-50 y). All analyses were done conditional on these age-gender strata and adjusted for smoking. We also fitted unconditional logistic models with age and sex as covariates and found the results almost identical to those of the conditional models (data not shown). These models were also carried out for cases of specific histologic subtypes. We carried out a Monte Carlo permutation, often considered as the gold standard for analysis (29), with 1,500 replications assigning the case-control status by permutation to evaluate whether single SNP results could be accountable to multiple testing while considering the underlying LD structure. Adjusted P values (P_adj) are derived as the P-quantiles of the empirical distribution of the 1,500 permutations where P is the unadjusted P value (30). The SNPs, chosen for this study, were carefully selected tagging SNPs. Thus, we applied two statistical methods using haplotype
information. First, we applied Mantel statistics based on haplotype sharing (31). In the vicinity of a predisposing mutation, haplotypes carrying a disease-relevant mutation (case haplotypes) should be more related than haplotypes not carrying such a mutation (random haplotypes). Therefore, case haplotypes are expected to share longer stretches of conserved DNA identical by descent around this mutation than random haplotypes. The pointwise test statistic correlates genetic similarity and phenotypic similarity across pairs of haplotypes to assess associations between a genetic marker x and phenotypes. The genetic similarity at x is measured as the shared length between haplotypes around the genetic marker as the number of markers surrounding x with the same alleles (i.e., identical by state). The phenotypic similarity for two haplotype copies is defined as the mean-corrected product, where the mean is estimated from the sample. Statistical significance was derived by Monte Carlo permutation with 10,000 replications. A step-down min-P algorithm was applied to control the family-wise error rate. The algorithm is based on the permutations and thus takes into account the underlying LD structure (32).

The second haplotype-based approach was to test for differences in haplotype frequencies between cases and controls using the software Chaplin9 (33, 34). Haplotype frequencies were derived by an extension of the standard Expectation-Maximization algorithm, called the Expectation-Conditional-Maximization algorithm. The approach allows for violations of Hardy-Weinberg equilibrium in haplotype frequencies in control and case samples that are typically required otherwise. Chaplin provides estimates of haplotype effects on disease and also tests whether such effects are significantly different from zero. We analyzed the multiplicative model of single haplotype effects and a general model. For both the haplotype frequency and the haplotype sharing analysis, the SNPs were separated into two LD sets. Due to the analysis of pairwise LD, the first set included SNP1 to SNP4 whereas the second set included SNP6 to SNP11 (Fig. 1). SNPs was not included because of deviation from Hardy-Weinberg equilibrium in controls.

Single point and haplotype analyses were done for the whole data set and within groups defined by smoking status. The level of significance was set at \( z = 5\% \) for each single test. All three statistical approaches used (i.e., single point analysis with conditional logistic regression, the Mantel statistics using haplotype sharing, and the haplotype frequency association test) account for multiple testing considering the correlation between the SNPs. The methods also provide P values for the null hypothesis that none of the SNPs is associated with the outcome (i.e., the family wise error rate). Pairwise LD was estimated using the software Haploview.10 Single point association analysis was done using SAS 9.1. For the haplotype frequency association analysis, we used Chaplin version 2.1. The software package Tomcat11 was used for Mantel statistics using haplotype sharing. Haplotypes were estimated with the software fast-PHASE12 (35). The genotyping success rates ranged from 98% to 100%. For 12 subjects, < of the 11 SNPs were genotyped successfully: the genotype distribution of SNP5 (rs7125062) showed a slight deviation from Hardy-Weinberg equilibrium (\( \chi^2 = 0.0229 \)) in controls, and this SNP was therefore excluded from haplotype analyses. Most polymorphisms chosen represent tagging SNPs and the minor allele frequencies ranged from 19% to 45%. The two additional intronic SNPs without tagging information are the intronic SNP7 (rs5031036) and SNP9 (rs1938901) where the minor allele frequency in controls was 5% and 31%, respectively (Supplementary Table S3).

Descriptive Analysis of the LD Structure of the Observed Region. We investigated the LD structure within controls around MMP1 covered by the genotyped SNPs (Fig. 1). The region was found to consist of two parts: a subregion of almost no LD covering the 5′ flanking region [between SNP1 (rs484915) and SNP4 (rs51492)], with pairwise \( D' = 0.6 \) and \( r^2 = 0.21 \), and a block of high LD between SNP5 (rs470358) in intron 1 and SNP11 (rs1939008) in the 3′ untranslated region. All pairwise LDs exceeded \( D' = 0.99 \) and \( r^2 = 0.22 \) except SNP7 (rs5031036). Between the adjacent SNPs (rs7125062) and SNP9 (rs1938901), \( D' = 0.57 \) (\( r^2 = 0.31 \)). The most pronounced multilocus LD was found in the 3′ flanking region containing SNP9, SNP10, and SNP11 (\( r^2 = 0.62 \)).

Tests for Association with Lung Cancer. Table 2 presents the main results for single SNP association analysis with lung cancer using conditional logistic regression analysis. In the general model, SNP6 (rs986999), SNP9 (rs1938901), and SNP11 (rs1939008) showed significant associations. The last two remained significant even after adjusting for multiple testing.

No significant association was observed for the four SNPs of the 5′ flanking region (Supplementary Table S1). The lowest P value [\( P = 0.0027 \); odds ratio (OR), 1.3; 95% confidence interval (95% CI), 1.1-1.5] was obtained for rs1938901 (SNP9, 7229 C>T, intron 8) assuming a multiplicative model of inheritance. For carriers of the C-allele (major allele), an increase in the risk for lung cancer was estimated with an OR of 1.5 (95% CI, 1.0-2.2; \( P = 0.0534 \)) for heterozygotes (CT) and an OR of 1.8 (95%
CI, 1.2-2.7; \( P = 0.0033 \)) for homozygotes (CC). The nearby located polymorphism rs193008 (SNP11, 12471 G>A) showed an almost identically increased lung cancer risk for carriers of the G allele. No association (\( P = 0.9181 \)) was found for the intermediate polymorphism rs5854 (SNP10, 8020 G>A; Supplementary Table S1). A significant association with higher risk of lung cancer for carriers of the C allele was also observed for rs996999 in intron 4 (SNP6, 1831 C>T) with \( P = 0.0152 \) and an OR of 1.3 (95% CI, 1.0-1.5) for each allele. For the same three SNPs, we further observed a stronger association with risk of lung cancer for those who smoked >20 pack-years whereas we observed no association for those who smoked <20 pack-years (Table 2).

The associations of polymorphisms with respect to histologic subtypes of lung cancer are shown in Table 3. The estimated effects for the polymorphisms rs996999 (SNP6), rs1938901 (SNP9), and rs193008 (SNP11) with respect to adenocarcinoma were found to be larger than those for the other non–small-cell lung carcinoma and small-cell lung carcinoma. The lowest \( P \) value was found for rs193008 (SNP11), assuming an additive mode of inheritance. For heterozygotes, we estimated an OR of 1.8 (95% CI, 1.0-3.2; \( P = 0.0708 \)), and for homozygotes, an OR

### Table 2. Conditional logistic regression analysis of lung cancer with SNPs in the MMP1 gene

<table>
<thead>
<tr>
<th>Model</th>
<th>MAF (%)</th>
<th>All cases and controls</th>
<th>Strong smokers (≥20 pack-years)</th>
<th>Never and weak smokers (&lt;20 pack-years)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Controls</td>
<td>Cases</td>
<td>( P )</td>
</tr>
<tr>
<td>SNP6: rs996999 (1831 C&gt;T)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>General</td>
<td>0.1654</td>
<td>1.5 (0.8-2.6)</td>
<td>0.195</td>
<td>1.6 (0.8-3.1)</td>
</tr>
<tr>
<td>( F )-statistics</td>
<td>0.0459</td>
<td>(( P_{\text{adj}} = 0.1182 ))</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dominant</td>
<td>0.0542</td>
<td>1.7 (1.0-2.9)</td>
<td>0.0751</td>
<td>1.8 (0.9-3.4)</td>
</tr>
<tr>
<td>Recessive</td>
<td>0.0376*</td>
<td>1.3 (1.0-1.6)</td>
<td>0.0702</td>
<td>1.3 (1.0-1.7)</td>
</tr>
<tr>
<td>Multiplicative</td>
<td>0.0152*</td>
<td>1.3 (1.0-1.5)</td>
<td>0.0314*</td>
<td>1.3 (1.0-1.6)</td>
</tr>
<tr>
<td>SNP9: rs1938901 (7229 C&gt;T)</td>
<td>31</td>
<td>28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>General</td>
<td>0.0534</td>
<td>1.5 (1.0-2.2)</td>
<td>0.0238*</td>
<td>1.79 (1.1-3.0)</td>
</tr>
<tr>
<td>( F )-statistics</td>
<td>0.0089</td>
<td>(( P_{\text{adj}} = 0.01821 ))</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dominant</td>
<td>0.0096</td>
<td>1.7 (1.1-2.4)</td>
<td>0.0039*</td>
<td>2.02 (1.3-3.3)</td>
</tr>
<tr>
<td>Recessive</td>
<td>0.0156*</td>
<td>1.3 (1.1-1.6)</td>
<td>0.0208*</td>
<td>1.39 (1.1-1.8)</td>
</tr>
<tr>
<td>Multiplicative</td>
<td>0.0027*</td>
<td>1.3 (1.1-1.5)</td>
<td>0.0023*</td>
<td>1.39 (1.1-1.7)</td>
</tr>
<tr>
<td>SNP11: rs193008 (12471 G&gt;A)</td>
<td>31</td>
<td>28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>General</td>
<td>0.0603</td>
<td>1.5 (1.0-2.2)</td>
<td>0.0409*</td>
<td>1.7 (1.0-2.8)</td>
</tr>
<tr>
<td>( F )-statistics</td>
<td>0.0108*</td>
<td>(( P_{\text{adj}} = 0.02334 ))</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dominant</td>
<td>0.0117*</td>
<td>1.6 (1.1-2.4)</td>
<td>0.0075*</td>
<td>1.93 (1.2-3.1)</td>
</tr>
<tr>
<td>Recessive</td>
<td>0.0173*</td>
<td>1.3 (1.0-1.6)</td>
<td>0.0208*</td>
<td>1.39 (1.1-1.8)</td>
</tr>
<tr>
<td>Multiplicative</td>
<td>0.0022</td>
<td>1.3 (1.0-1.5)</td>
<td>0.0031*</td>
<td>1.38 (1.1-1.7)</td>
</tr>
</tbody>
</table>

**NOTE:** Data are matched for age and gender. All ORs are adjusted for smoking. \( P_{\text{adj}} \), \( P \) values adjusted for multiple testing. SNP position is counted in basepairs to/from start (AUG); risk allele is in boldface. \( F \)-statistics was used for the full model versus the model without any genetic effect. The table shows only the polymorphisms with significant association results; the entire results can be seen in Supplementary Table S3.

**Abbreviation:** MAF, minor allele frequency.

* \( P < 0.05 \).

**c** \( P < 0.01 \).

### Table 3. Association of polymorphisms and different histologic subtypes of lung cancer for the general model

<table>
<thead>
<tr>
<th>Polymorphism</th>
<th>Risk allele</th>
<th>MAF (%)</th>
<th>Histologic subtype</th>
<th>One risk allele</th>
<th>Two risk alleles</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Controls</td>
<td>Cases</td>
<td>OR (95% CI)</td>
<td>( P )</td>
<td>OR (95% CI)</td>
</tr>
<tr>
<td>SNP6: rs996999</td>
<td>C</td>
<td>20</td>
<td>19</td>
<td>2.3 (0.9-5.6)</td>
<td>0.0809</td>
</tr>
<tr>
<td>SNP9: rs1938901</td>
<td>C</td>
<td>31</td>
<td>28</td>
<td>1.9 (0.6-5.6)</td>
<td>0.2619</td>
</tr>
<tr>
<td>SNP11: rs193008</td>
<td>G</td>
<td>31</td>
<td>28</td>
<td>1.8 (0.6-2.9)</td>
<td>0.2619</td>
</tr>
</tbody>
</table>

**NOTE:** The table shows only the polymorphisms with significant association results. The entire results are shown in Supplementary Table S4.

*SNP with significant association to histologic subtype.
SNP11, which showed significant results in the single characterized by the major alleles of SNP6, SNP9, and increase in risk). Interestingly, the four haplotypes were showed positive effect estimates (i.e., indicating an = 0.049 (Table 5). Four of the six estimated haplotypes significant results were found for strong smokers with haplotypes consisting of SNP6 to SNP11, globally effects in the respective data sets. However, for the did not find significant results for single haplotype sets: first for SNP1 to SNP4, and second for SNP6 to SNP11. Testing for a multiplicative disease model, we haplotype frequencies between cases and controls. (data not shown).

Table 4 presents the results for the Mantel statistics using haplotype sharing for the region of high LD (SNP6 to SNP11). No significant results were found in the data set consisting of all subjects and the subset of never and weak smokers (<20 pack-years). In the subset of strong smokers, however, significant results were found across the entire region. The lowest P value was found for SNP9 (P < 0.003). After correction for multiple testing, the entire region of high LD was found to be significantly associated with the risk of lung cancer (global P value P = 0.007). A haplotype sharing analysis of the 5' flanking region (SNP1-SNP4) did not yield significant results (data not shown).

We used the software Chaplin to test for differences in haplotype frequencies between cases and controls. Statistical tests were calculated in two separated SNP sets: first for SNP1 to SNP4, and second for SNP6 to SNP11. Testing for a multiplicative disease model, we did not find significant results for single haplotype effects in the respective data sets. However, for the haplotypes consisting of SNP6 to SNP11, globally significant results were found for strong smokers with P = 0.049 (Table 5). Four of the six estimated haplotypes showed positive effect estimates (i.e., indicating an increase in risk). Interestingly, the four haplotypes were characterized by the major alleles of SNP6, SNP9, and SNP11, which showed significant results in the single
glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and inhibitors of metalloproteinases. Additionally, environmental factors affect MMP1 expression: Cigarette smoke increases (54). The known effect of tobacco smoke on MMP1 expression is not affected (51-56). As a result, the collagenolytic activity of human airway cells increases (54).

High expression levels of MMP1 are responsible for loosening cell adhesion. Carcinogenic substances easily penetrate the epidermis and infiltrate and damage cells. If cancer cells are generated, metastasis cells can spread more easily in the surrounding area (57, 58). Supporting the known effect of tobacco smoke on MMP1 expression, we could observe a stronger association with risk of early-onset lung cancer for those who smoked ≥20 pack-years compared with those who smoked <20 pack-years.

Summarizing, it can be stated that the effect of MMP polymorphisms depends on the balance between MMPs and TIMPs, which is strongly influenced by tobacco exposure and hormone status and therefore seems to be a highly complex system. Recapitulating our results, we predict a disease locus either in the 3′ flanking region or near the end of the gene (from intron 6 to the gene’s end) especially increasing the risk for early-onset lung cancer in strong smokers. Due to the fact that the 3′ untranslated region and the flanking sequences are responsible for stabilization of mRNA, it could be hypothesized that the major alleles causes a stable mRNA with normal protein expression. Exposure to tobacco smoke increases the MMP1 level, resulting in increased collagenase activity, decreased cell adhesion, and easier penetration of particles and carcinogens. Carriers of the protective alleles may benefit from a less stable mRNA, developing lower MMP1 levels and lower collagenase activity, even if they are exposed to tobacco smoke.

Lung cancer at young age is rare (<10% of all lung cancer cases), and this study is based on one of the largest groups of young patients worldwide. For the investigation of genetic associations, these are of specific value because they show a much stronger aggregation with familial lung cancer and other cancer entities in comparison with lung cancer cases at a higher age (9). Our study was population based and we had sufficient SNP coverage across the MMP1 gene. It might be considered as a limitation that our study did not include the one frequently discussed promoter SNP (rs1799750) and that the haplotype analysis of the 5′ flanking region is hampered by low LD.

However, to our knowledge, this is the first investigation in Caucasians on association of SNPs and haplotypes on the entire MMP1-gene. In conclusion, individuals carrying the major allele of rs996999, rs1938901, or rs193008 and/or belonging to a haplotype carrying the same alleles may harbor a DNA variant that affects lung cancer risk, and this association was strongly modified by the amount of tobacco smoke exposure. This relationship between smoking, MMP1 polymorphisms, and lung cancer risk needs to be confirmed by other independent studies.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

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.

We thank all the participating hospitals (LUCY-Consortium): Aurich (Dr. Heidi Kleen); Bad Berka (Dr. med. R. Bonnet, Klinik für Pneumologie, Zentralklinik Bad Berka GmbH); Bonn (Prof. Ko, Dr. Geisen, Innere Medizin I, Johanniter Krankenhaus); Bonn (Dr. Stier, Medizinische Poliklinik, Universität Bonn); Bremen (Prof. Dr. D. Ukena, Dr. Penzl, Zentralkrankenhaus Bremen Ost, Pneumologische Klinik); Chemnitz (PD Dr. Schmidt, OA Dr. Jelje, Klinikum Chemnitz, Abteilung Innere Medizin); Coswig (Prof. Höfken, Dr. Schmidt, Fachklinik für Pneumologie Coswig); Diekholzen (Dr. Ham, Kreiskrankenhaus Diekholzen, Klinik für Pneumologie); Donaustauf (Prof. Pfeifer, Dr. v. Bülltingslöwen, Fr. Schneider, Fachklinik für Atemwegserkrankungen).
References


Matrix Metalloproteinase 1 (MMP1) Is Associated with Early-Onset Lung Cancer

Wiebke Sauter, Albert Rosenberger, Lars Beckmann, et al.

Cancer Epidemiol Biomarkers Prev 2008;17:1127-1135.

Updated version
Access the most recent version of this article at:
http://cebp.aacrjournals.org/content/17/5/1127

Cited articles
This article cites 56 articles, 22 of which you can access for free at:
http://cebp.aacrjournals.org/content/17/5/1127.full.html#ref-list-1

Citing articles
This article has been cited by 24 HighWire-hosted articles. Access the articles at:
/content/17/5/1127.full.html#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.