Multiple Genetic Loci Modulate Lung Adenocarcinoma Clinical Staging

Elisa Frullanti1, Antonella Galvan1, Felicia S. Falvella1, Giacomo Manenti1, Francesca Colombo1, Alberto Vannelli1, Matteo Incarbone3, Marco Alloisio3, Mario Nosotti2, Luigi Santambrogio2, Anna Gonzalez-Neira4, Ugo Pastorino1, and Tommaso A. Dragani1

Abstract

Purpose: The main prognostic factor of lung cancer patient outcome is clinical stage, a parameter of tumor aggressiveness. Our study was conducted to test whether germ line variations modulate individual differences in clinical stage.

Experimental Design: We conducted a case-only genome-wide association study (GWAS) using a 620,901 single-nucleotide polymorphism (SNP) array in a first series of 600 lung adenocarcinoma (ADCA) patients and in a replication series of 317 lung ADCA patients.

Results: GWAS identified 54 putatively associated SNPs, 3 of which were confirmed in the replication series. Joint analysis of the two series pointed to 22 statistically associated (P < 0.01) genetic variants that together explained about 20% of the phenotypic variation in clinical staging (P = 2 × 10−16) and showed a statistically significant difference in overall survival (P = 8.0 × 10−5). The strongest statistical association was observed at rs10278557 (P = 1.1 × 10−5), located in the mesenchyme homeobox 2 (MEOX2) gene.

Conclusion: These data point to the role of germ line variations involving multiple loci in modulating clinical stage and, therefore, prognosis in lung ADCA patients. Clin Cancer Res; 17(8); 2410–6. ©2011 AACR.

Introduction

Lung cancer prognosis is affected by clinical stage, tumor histologic subtype, and the possibility of surgical resection. However, clinical stage is the parameter with the greatest impact on lung cancer patient survival; indeed, patients with early-stage cancer have an excellent prognosis, whereas prognosis is poorer with increasing stage (reviewed in ref. 1). Variations in cancer aggressiveness and malignancy have been associated mainly with the accumulation of multiple somatic alterations and epigenetic changes in the neoplastic cells (2). While most studies aimed at identifying factors that affect cancer patient outcome/survival have focused on genetic alterations or transcriptional changes in the cancer tissue, a recent study suggests a role for germ line variations in the control of lung cancer patient survival (3). Those findings are supported by results in mouse models of lung tumorigenesis showing that specific genetic loci modulate tumor progression (4, 5).

In the present study, we conducted a genome-wide association study (GWAS) of lung adenocarcinoma (ADCA) patients to test the hypothesis that clinical stage can be genetically modulated. Comparison of DNA from patients at clinical stage I with DNA of patients at higher clinical stages revealed multiple and unlinked single-nucleotide polymorphisms (SNPs) that may genetically modulate clinical stage and that together are associated with patient overall survival.

Materials and Methods

Study population and DNAs

All patients were enrolled in the authors’ institutes in Milan, Italy (Table 1). Study protocols were approved by all ethics committees, and each subject gave informed consent to the use of their biological samples for research purposes.

Genomic DNA was extracted from peripheral blood using the DNeasy Blood & Tissue Kit (QIAGEN) and quantified by fluorimetry using the Picogreen dsDNA Quantitation Kit (Invitrogen). Patients in the first series were divided into 2 groups according to clinical stage (I or >I), and the same amounts of each DNA sample were used to create a DNA pool of 300 stage I patients and a DNA pool of 300 patients at higher clinical stages. Since the accuracy of analyses using a DNA pooling strategy depends heavily on the estimates of DNA concentration (6), we performed serial dilutions of each DNA sample.

SNP array and genotyping

A genome-wide DNA pooling strategy was used for initial screening to minimize interindividual sample variability and to reduce costs and time as compared with analyses of
Translational Relevance

Clinical stage is the main clinical parameter affecting prognosis in lung cancer patients. We tested whether a genetic component might be involved in modulating patient clinical stage. In 917 Italian lung adenocarcinoma patients, we detected 22 genetic variants that together explained a large individual variation in clinical stage and that were also associated with overall survival. This demonstration that individual genetic constitution can affect clinical stage represents a step toward understanding the role of genetic factors in a clinically relevant parameter and opens the possibility of identifying new genetic targets for lung cancer therapy based on individual genetic constitution.

GWAS identifies multiple SNPs associated with clinical stage

Genome-wide SNP array analysis conducted in 12 replicas of DNA pools from lung ADCA cases at clinical stage I or at higher clinical stages, respectively, allowed the screening of 620,901 SNPs. After discarding SNPs whose minor allele frequency was less than 0.10 in the pools and whose statistically significant imbalance of allelic frequencies between the 2 DNA pools was below the genome-wide threshold of \( P \leq 2.0 \times 10^{-7} \), analysis of the reconstructed number of chromosomes of the remaining SNPs in the 2 groups using a \( 2 \times 2 \) contingency table revealed 80 most statistically associated SNPs (\( P < 1.0 \times 10^{-8} \)). These were selected for individual genotyping to validate the SNP array findings in the DNA pool. Of the 80 SNPs, 2 mitochondrial SNPs, 1 SNP on chromosome Y, and 9 redundant SNPs in tight LD with close-by SNPs (<58-kb distance) in the same locus in the HapMap Caucasian (CEU) population were excluded; 1 SNP failed PCR or MassEXTEND primer design and 4 additional SNPs failed genotyping, reducing the number of markers to 63 SNPs.

A good correlation was observed in the minor allele frequencies obtained either by MassARRAY genotyping in single individuals or by SNP array analysis in DNA pools.
Among the 63 SNPs tested in the replication series of 317 lung ADCA samples (Table 1), 3 SNPs showed replication in the independent smaller ADCA series (Table 3). Joint analysis of the GWA and replication series, increasing the statistical power of association analyses (11) and bringing the total sample size to 917 lung ADCA patients, revealed no statistically significant deviation (P < 0.01) from the HWE for any of the 63 SNPs and identified 22 SNPs significantly associated with clinical stage at the statistical threshold of P < 0.01 by logistic analysis adjusted for age at diagnosis and smoking status (Table 4). The strongest
association was again observed at SNP rs10278557 ($P = 1.1 \times 10^{-3}$) mapping in the MEOX2 gene (Table 4).

### Differences in lung ADCA outcome are associated with patients’ genetic profile

We used a polygenic model (8) to evaluate additive effects of these 22 SNPs in modulating individual clinical stage, after removing 81 of 917 patients with more than 30% missing genotypes from the data set. For each patient, the allele-based OR (Table 4) was attributed to the carrier status of an allele of each SNP associated with clinical stage status, on the basis of its association with the probability of carrying a stage $>1$ lung ADCA.

The average genetic estimator was $-7.9 \times 10^{-3} \pm 5.4 \times 10^{-4}$ units (mean $\pm$ standard error) for patients with clinical stage I ($n = 418$) and $3.2 \times 10^{-3} \pm 5.3 \times 10^{-4}$ for patients with higher clinical stage ($n = 403; P < 2.2 \times 10^{-16}$, ANOVA analysis). The 22 SNPs explained 20.7% of the phenotypic variance in clinical staging. Although with a lower effect than in the first series and in the whole series, the genetic estimator was also statistically associated to clinical stage in the second ADCA series alone ($P = 0.0006$, ANOVA analysis). To verify the robustness of the model in our series, we carried out an empirical replication using bootstrap samples ($B = 2,000$ resamplings) and found that the difference in the genetic estimator between stage I and stage $>1$ patients was $-11.1 \times 10^{-3}$ units, 95% CI $= -12.7 \times 10^{-3}$ to $-9.7 \times 10^{-3}$, $P_{\text{diff}} = 0.0005$.

Subjects were divided into quartiles on the basis of the genetic risk score. Application of the generalized linear model to the quartile groups, with the lowest quartile as the reference, revealed a significant association between the genetic estimator and increased probability of developing a more aggressive lung ADCA (OR $= 2.9$, 95% CI $= 1.9$–$4.6$, $P = 2.7 \times 10^{-8}$ for the second quartile, OR $= 6.8$, 95% CI $= 4.4$–$10.7$, $P < 2 \times 10^{-18}$ for the third quartile; and OR

### Table 3. SNPs associated with lung ADCA clinical stage in the GWAS and replication studies

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chromosome</th>
<th>Position, Mb</th>
<th>Gene</th>
<th>Rare allele</th>
<th>GWAS OR</th>
<th>95% CI</th>
<th>$P^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs3823111</td>
<td>6</td>
<td>53.62</td>
<td>KLHL31</td>
<td>T</td>
<td>2.4</td>
<td>1.3–4.5</td>
<td>$5.0 \times 10^{-3}$</td>
</tr>
<tr>
<td>rs9927531</td>
<td>16</td>
<td>26.44</td>
<td></td>
<td>A</td>
<td>1.8</td>
<td>1.2–2.6</td>
<td>$1.6 \times 10^{-3}$</td>
</tr>
<tr>
<td>rs16950191</td>
<td>17</td>
<td>47.07</td>
<td>CA10</td>
<td>A</td>
<td>1.4</td>
<td>1.0–1.8</td>
<td>$3.2 \times 10^{-2}$</td>
</tr>
</tbody>
</table>

--

*SNPs sorted by chromosome and position.

$^b$Logistic regression procedure in PLINK toolset, based on allelic test for association, adjusted for age at cancer diagnosis and smoking status. Selection of SNPs based on $P < 0.05$ threshold in the replication study.
Table 4. SNPs associated with lung ADCA clinical stage in the joint analysis of the GWAS and replication studies and used to build up the polygenic model with additive effects of SNP rare alleles on risk of clinical stage >I

<table>
<thead>
<tr>
<th>SNPa</th>
<th>Chromosome</th>
<th>Position, Mb</th>
<th>Gene</th>
<th>Rare allele</th>
<th>OR</th>
<th>95% CI</th>
<th>Pb</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs951774</td>
<td>2</td>
<td>102.28</td>
<td>TTN</td>
<td>A</td>
<td>0.7</td>
<td>0.5–0.9</td>
<td>6.9 × 10⁻³</td>
</tr>
<tr>
<td>rs13909491</td>
<td>2</td>
<td>179.29</td>
<td>COL4A4</td>
<td>T</td>
<td>1.4</td>
<td>1.1–1.8</td>
<td>6.3 × 10⁻³</td>
</tr>
<tr>
<td>rs10498217</td>
<td>2</td>
<td>227.72</td>
<td>T</td>
<td>0.7</td>
<td>0.5–0.9</td>
<td>9.4 × 10⁻³</td>
<td></td>
</tr>
<tr>
<td>rs1904854</td>
<td>4</td>
<td>77.99</td>
<td>A</td>
<td>0.7</td>
<td>0.5–0.9</td>
<td>2.8 × 10⁻³</td>
<td></td>
</tr>
<tr>
<td>rs4505911</td>
<td>5</td>
<td>68.28</td>
<td>A</td>
<td>0.5</td>
<td>0.3–0.8</td>
<td>4.0 × 10⁻³</td>
<td></td>
</tr>
<tr>
<td>rs1000886</td>
<td>5</td>
<td>105.28</td>
<td>A</td>
<td>0.7</td>
<td>0.5–0.9</td>
<td>7.7 × 10⁻³</td>
<td></td>
</tr>
<tr>
<td>rs3823111</td>
<td>6</td>
<td>53.62</td>
<td>KLHL31</td>
<td>T</td>
<td>2.6</td>
<td>1.6–4.3</td>
<td>2.1 × 10⁻⁴</td>
</tr>
<tr>
<td>rs806435</td>
<td>6</td>
<td>88.83</td>
<td>SPAC41</td>
<td>T</td>
<td>1.8</td>
<td>1.3–2.4</td>
<td>2.5 × 10⁻⁴</td>
</tr>
<tr>
<td>rs10278557</td>
<td>7</td>
<td>15.67</td>
<td>MEOX2</td>
<td>A</td>
<td>0.6</td>
<td>0.4–0.7</td>
<td>1.1 × 10⁻⁵</td>
</tr>
<tr>
<td>rs2299297</td>
<td>7</td>
<td>104.53</td>
<td>MLL5</td>
<td>T</td>
<td>1.6</td>
<td>1.3–2.0</td>
<td>7.5 × 10⁻⁵</td>
</tr>
<tr>
<td>rs824249</td>
<td>9</td>
<td>28.76</td>
<td>A</td>
<td>0.6</td>
<td>0.5–0.9</td>
<td>9.5 × 10⁻³</td>
<td></td>
</tr>
<tr>
<td>rs10987191</td>
<td>9</td>
<td>127.99</td>
<td>A</td>
<td>0.5</td>
<td>0.4–0.8</td>
<td>2.5 × 10⁻³</td>
<td></td>
</tr>
<tr>
<td>rs10832757</td>
<td>11</td>
<td>17.29</td>
<td>NUCB2</td>
<td>A</td>
<td>0.6</td>
<td>0.5–0.8</td>
<td>5.8 × 10⁻⁵</td>
</tr>
<tr>
<td>rs9596742</td>
<td>13</td>
<td>52.46</td>
<td>A</td>
<td>0.6</td>
<td>0.5–0.9</td>
<td>7.7 × 10⁻³</td>
<td></td>
</tr>
<tr>
<td>rs2391875</td>
<td>13</td>
<td>110.29</td>
<td>A</td>
<td>0.6</td>
<td>0.4–0.8</td>
<td>8.8 × 10⁻⁵</td>
<td></td>
</tr>
<tr>
<td>rs8020076</td>
<td>14</td>
<td>27.53</td>
<td>C</td>
<td>1.3</td>
<td>1.1–1.6</td>
<td>9.8 × 10⁻³</td>
<td></td>
</tr>
<tr>
<td>rs10520058</td>
<td>15</td>
<td>36.34</td>
<td>SPRED1</td>
<td>A</td>
<td>0.5</td>
<td>0.3–0.8</td>
<td>5.8 × 10⁻³</td>
</tr>
<tr>
<td>rs9927531</td>
<td>16</td>
<td>26.44</td>
<td>A</td>
<td>1.7</td>
<td>1.3–2.3</td>
<td>2.5 × 10⁻⁴</td>
<td></td>
</tr>
<tr>
<td>rs10514440</td>
<td>16</td>
<td>77.24</td>
<td>WWOX</td>
<td>T</td>
<td>2.4</td>
<td>1.4–4.1</td>
<td>1.1 × 10⁻³</td>
</tr>
<tr>
<td>rs16950191</td>
<td>17</td>
<td>47.97</td>
<td>CA10</td>
<td>A</td>
<td>1.5</td>
<td>1.2–1.8</td>
<td>1.6 × 10⁻³</td>
</tr>
<tr>
<td>rs7887846</td>
<td>X</td>
<td>22.58</td>
<td>A</td>
<td>0.7</td>
<td>0.5–0.9</td>
<td>7.8 × 10⁻³</td>
<td></td>
</tr>
<tr>
<td>rs2207031</td>
<td>X</td>
<td>127.84</td>
<td>A</td>
<td>1.8</td>
<td>1.2–2.5</td>
<td>1.4 × 10⁻³</td>
<td></td>
</tr>
</tbody>
</table>

aSNPs sorted by chromosome and position.
bLogistic regression procedure in PLINK toolset, based on allelic test for association with clinical stage, adjusted for age at cancer diagnosis and smoking status. SNPs selected on the basis of P < 0.01 threshold for association.

= 14.5, 95% CI = 9.1–23.6, P < 2 × 10⁻¹⁶ for the fourth quartile group, Fig. 1A).

Finally, Kaplan–Meier curves showed a statistically significant association between the genetic risk score, in quartiles, and overall survival (P = 8.0 × 10⁻⁶, log-rank test; Fig. 1B). Use of multivariate Cox proportional hazard models for survival (adjusted for age and smoking habit) to evaluate the association between the genetic risk score and overall survival showed that the risk of death for quartiles 3 and 4 (HR = 1.5, 95% CI = 1.1–2.0, P = 0.016; HR = 2.3, 95% CI = 1.7–3.0, P = 8.7 × 10⁻⁵, respectively) was significantly higher than that of the lowest quartile.

Discussion

In recent years, several GWAS have focused on genetic risk for lung cancer, but none has examined the possible genetic modulation of the most powerful prognostic factor in lung cancer patients, that is, clinical stage (1). Our present study in a patient series of the same lung cancer histotype and of the same ethnicity identified 54 SNPs, putatively associated (P < 0.05) with clinical stage (Table 2) in a relatively large first series of 600 patients, and 3 SNPs that maintained their statistical association with clinical stage in the smaller replication series (Table 3). Joint analysis of the GWAS and replication series to increase the statistical power of the study and to obtain an overall unbiased estimate (11) identified 22 SNPs that, at nominal statistical value of P < 0.01, showed statistical association with clinical stage (Table 4). Analysis of additive effects of risk associated to the minor alleles of these 22 SNPs using a polygenic model (8) revealed a statistically significant association between the genetic estimator and an increased risk of higher clinical stage (Fig. 1A) and higher risk of death (Fig. 1B), suggesting the complex genetic control of lung ADCA patient clinical prognosis.

Of the 22 candidate SNPs, 6 mapped within genes. The most significantly associated SNP in the joint analysis (rs10278557, P = 1.1 × 10⁻⁵, Table 4) maps on chromosome 7 in the intronic region of the MEOX2 gene, also known as growth arrest–specific homeobox (CAX) gene, a member of a subfamily of nonclustered, divergent, antenapedia-like homeobox-containing genes. MEOX2, a key regulator of vascular cell function, has been proposed as a candidate tumor suppressor gene in Wilm’s tumor and
shows upregulation and aberrant methylation in lung cancer (12, 13).

The myeloid/lymphoid or mixed-lineage leukemia 5 (trithorax homolog, Drosophila; MLL5, rs2299297), the sprouty related, EVH1 domain containing 1 (SPRED1, rs10520058), and the WW domain containing oxidoreductase (WWOX, rs10514440; Table 4) candidacies are also of interest. Indeed, the MLL5 gene belongs to a gene family that activates and regulates homeobox (HOX) genes that are important in oncogenesis and tumor suppression (14, 15). MLL5 is located on chromosome 7q22, which is frequently deleted in myeloid leukemias and is a key regulator of normal hematopoiesis (16). SPRED1 negatively regulates the Ras-ERK (extracellular signal–regulated kinases) signaling pathway, cell motility, and metastasis, and its germ line loss-of-function mutations cause a neurofibromatosis 1–like syndrome (17, 18). WWOX acts as a tumor suppressor gene in different tumor types and plays a regulatory role in protein degradation, transcription, and RNA splicing (reviewed in ref. 19).

At present, it is unknown whether the observed associations between SNPs and lung cancer clinical stage underlie effects of nonsynonymous or regulatory variants in LD with these SNPs.

Empirical replication using bootstrap samples from the original data, rather than replication in independent samples, has been proposed in association studies since bootstrap samples likely share the same population structure of original data, whereas an independent series may be characterized by a different population structure and, thus, lead to false-negative results on analysis (20). Our empirical replication using bootstrap samples confirmed the statistically significant difference between stage I and stage >I patients in their genetic estimator on the basis of 22 SNPs. However, the lack of a replica in a truly independent population remains a potential weakness of the present study that may be overcome in future, large studies carried out by international consortia.

Together, our results indicate for the first time that clinical staging of lung ADCA can be under genetic control, with each patient displaying a tendency toward a low or high clinical stage depending on individual genetic variations. The significant association of the 22 SNPs with lung ADCA clinical stage and survival raises the possibility that the functional products of the genes linked to these SNPs use novel biochemical pathways associated with such patient outcome, and that identification of these pathways might provide gene targets for therapies to counter disease progression. Further clarification of the role of genetic mechanisms in lung ADCA patient outcomes may hold the promise of improved therapy and disease outcome.

**Disclosure of Potential Conflicts of Interest**

The funders had no role in the design and conduct of the study, in the collection, analysis, and interpretation of the data, and in the preparation, review, or approval of the manuscript.

**Acknowledgments**

The authors thank Harvard-Partners Center for Genetics and Genomics Genotyping Facility, Cambridge, MA, for custom genotyping by MassARRAY.

**Grant Support**

This work was funded in part by grants from Associazione and Fondazione Italiana Ricerca Cancro (AIRC and FIRC). E. Frullanti and F. Colombo were funded by the AIRC “Antonietta Andreoli” and Associazione Marta Nurizzo fellowships, respectively.

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 September 6, 2010; revised December 17, 2010; accepted December 18, 2010; published OnlineFirst January 17, 2011.
References
