

## Association of the *PDCD5* Locus With Lung Cancer Risk and Prognosis in Smokers

AQ: A

Monica Spinola, Peter Meyer, Stefan Kammerer, F. Stefania Falvella, Melanie B. Boettger, Carolyn R. Hoyal, Carmen Pignatiello, Reiner Fischer, Richard B. Roth, Ugo Pastorino, Karl Haeussinger, Matthew R. Nelson, Rainer Dierkesmann, Tommaso A. Dragani, and Andreas Braun

### ABSTRACT

#### Purpose

Whole-genome scan association analysis was carried out to identify genetic variants predictive of lung cancer risk in smokers and to confirm the identified variants in an independent sample.

#### Patients and Methods

A case-control study was performed using two pools consisting of DNA from 321 German smoking lung cancer patients and 273 healthy smoking controls, respectively. A replication study was carried out using 254 Italian lung adenocarcinoma (ADCA) patients and 235 healthy controls.

#### Results

Patients with genotypes GG or CG for the rs1862214 single nucleotide polymorphism, 5' upstream of the programmed cell death 5 (*PDCD5*) gene, compared with those with the common genotype CC showed an increased risk of lung cancer (odds ratio [OR], 1.6; 95% CI, 1.2 to 2.1) and a higher incidence of poor clinical stage disease (hazard ratio [HR], 1.9; 95% CI, 1.1 to 3.4;  $P = .023$ ), nodal involvement (HR, 1.9; 95% CI, 1.1 to 3.6;  $P = .033$ ), and short-term survivorship (HR, 1.8; 95% CI, 1.2 to 2.6,  $P = .003$ ). *PDCD5* mRNA expression levels were ~2.4-fold lower in lung ADCA as compared to normal lung tissue. Human NCI-H520 cancer cells transfected with *PDCD5* cDNA showed decreased colony-forming ability.

#### Conclusion

These results suggest that the rs1862214 polymorphism in *PDCD5* is predictive for lung cancer risk and prognosis, and that *PDCD5* may represent a novel tumor suppressor gene influencing lung cancer.

*J Clin Oncol* 24. © 2006 by American Society of Clinical Oncology

### INTRODUCTION

Lung cancer is the leading cause of cancer mortality in the Western world. Indeed, this malignancy has a poor prognosis and, so far, therapeutic strategies have shown only a limited effect. Lung cancer risk is strongly associated with exposure to environmental carcinogens, in particular to cigarette smoke. Nevertheless, only approximately 10% of smokers develop lung cancer, and the disease also occurs in the absence of exposure to cigarette smoke. Therefore, the potential role of hereditary components in determining the risk of lung cancer in smokers should not be underestimated. Although this hypothesis has been questioned in some studies,<sup>1</sup> epidemiologic data support the presence of genetic risk factors in lung cancer.<sup>2-7</sup>

Whole-genome analyses in mouse models have identified genetic loci that affect lung cancer risk. Indeed, genetic analysis of these models provided evidence of a complex inheritance of genetic predis-

position to lung tumorigenesis<sup>8</sup> and suggested that polygenic inheritance of predisposition to cancer might also occur in humans.<sup>9</sup>

In mice, whole-genome analyses require only a few hundred genetic markers because of the genetic homogeneity of inbred strains, the limited number of genetic recombinations in crosses, and the consequent tight linkage of genetic markers over wide chromosomal regions.<sup>8</sup> Because of the genetic heterogeneity of the human population, complete analysis of the human genome by association studies may require a huge number of genetic markers and large sample collections. However, the presence of haplotype blocks in the human genome<sup>10</sup> suggests that even a less dense marker coverage might be sufficient for a whole-genome analysis. Although the characterization of these haplotype blocks awaits completion,<sup>11</sup> population-based association studies are complicated by complex [linkage disequilibrium \(LD\)](#) patterns, possible population admixture in the

AQ: B

From the Department of Experimental Oncology and Laboratories, Thoracic Surgery, Istituto Nazionale Tumori, Milan, Italy; Asklepios Clinics, Pneumology Clinic, Munich-Gauting; Genefinder Technologies Ltd, Munich; Institute of Human Genetics, Molecular Oncogenetics Division, University Hospital, Tuebingen; Clinic Schillerhoehe, Pneumology Clinic, Stuttgart-Gerlingen, Germany; and Sequenom Inc, San Diego, CA.

Submitted September 30, 2005; accepted December 16, 2005.

Supported in part by grants from Associazione and Fondazione Italiana Ricerca Cancro (AIRC and FIR) and FIRB to T.A.D.

M.S., P.M., and S.K. contributed equally to this manuscript. R.D. and A.B. contributed equally to this manuscript.

Terms in [blue](#) are defined in the glossary, found at the end of this article and online at [www.jco.org](http://www.jco.org).

Authors' disclosures of potential conflicts of interest and author contributions are found at the end of this article.

Address reprint requests to Tommaso A. Dragani, MD, Department of Experimental Oncology, Istituto Nazionale Tumori, Via G. Venezian 1, 20133 Milan, Italy; e-mail: [tommaso.dragani@istitutotumori.mi.it](mailto:tommaso.dragani@istitutotumori.mi.it).

© 2006 by American Society of Clinical Oncology

0732-183X/06/2411-1/\$20.00

DOI: 10.1200/JCO.2005.04.4339

selection of cases or controls, differences in environmental factors, and other factors.<sup>12</sup>

In an effort to identify genetic variants that may predict individual risk of lung cancer in smokers, we conducted a large-scale case-control study using more than 80,000 single nucleotide polymorphisms (SNPs) located primarily in gene-based regions. We report the identification of an 80-kb genomic region on chromosome 19q12-q13.1 including the programmed cell death 5 gene (*PDCD5*, also called *TFAR19* [TF1 cell apoptosis-related gene 19]) that conferred increased lung cancer risk in two independent sample collections and that may predict disease outcome.

**PATIENTS AND METHODS**

**Patients and Tissue Samples**

The initial study included 321 male smokers with lung cancer who were recruited at the Pneumology and Thoracic Surgery Clinics in Stuttgart-Gerlingen and the Pneumology Clinic in Munich-Gauting (both in Germany), and 273 healthy smokers recruited at the same sites and at the Dermatology Department of the University Hospital in Tuebingen, Germany (Table 1). The Italian replication sample included 254 cases of pathologically documented lung adenocarcinoma (ADCA) patients who underwent surgery at Istituto Nazionale Tumori (Milan, Italy), and 235 healthy controls (blood donors) enrolled at the same institute. Personal files were recorded to obtain clinical data (Tables 1 and 2). All subjects involved in our studies signed a written informed consent, and the institutional ethics committees of participating institutions approved the study protocol.

For the German discovery samples, DNA was obtained from blood samples using a desalting method (Gentra Systems, Minneapolis, MN) and quantitated using Pico green reagents and a Fluorometer (Fluoroskan Ascent CF; Labsystems, Chicago, IL). Genomic DNAs of the Italian patients and controls were extracted from a small piece of nontumor tissue excised during surgery or from peripheral blood samples using the automatic DNA extractor Extragen 8C (Talent, Trieste, Italy).

Total RNA was extracted from 20 paired nontumor lung parenchyma and lung tumor tissue and additional 26 nontumor lung parenchyma of Italian lung ADCA patients using RNeasy Midi kit (Qiagen, Valencia, CA). cDNA was synthesized from RNA of each sample using the SuperScript First-Strand Synthesis System (Invitrogen, Carlsbad, CA).

**SNP Markers and Genotyping**

A set of 83,715 SNP markers was selected from a collection of 125,799 experimentally validated polymorphic variations.<sup>13</sup> This set was limited to SNPs located within gene coding regions, with minor allele frequencies greater than 0.02 (95% have frequencies greater than 0.05) and a median intermarker spacing of 15 kb. SNP annotation is based on the National Center for Biotechnology Information (NCBI) dbSNP database, refSNP, Build 118. Genomic

**Table 1.** Characteristics of German and Italian Lung Cancer Patients and Controls

Characteristic	German		Italian	
	Controls (n = 273)	Patients (n = 321)	Controls (n = 235)	Patients (n = 254)
Median age, years	63.0	65.3	49.6	61.1*
Sex				
Female	87	0	62	65
Male	183	321*	173	189
Smoker status				
Never	0	0	96	46
Ever	270	322	92	196*

\*P < .01, patients compared with controls.

**Table 2.** Clinicopathologic Characteristics of German and Italian Lung Cancer Patients

Characteristic	German Patients	Italian Patients
<b>Histology</b>		
Adenocarcinoma	78	253
Non-small-cell lung cancer	46	
Squamous cell carcinoma	104	
Small-cell lung cancer	72	
Other	12	
<b>Lymph node status*</b>		
N0	56	146
N1	208	74
<b>Clinical stage</b>		
I	39	125
II	21	46
III	149	36
IV	90	21
Follow-up (months)†	NA	56.5

Abbreviation: NA, not available.  
 \*N0, absence of nodal metastasis; N1, presence of nodal metastasis.  
 †Median follow-up of patients alive at the end of the follow-up (n = 104).

annotation is based on NCBI Genome Build 34. Gene annotation is based on Entrez Gene entries for which NCBI provided positions on the Mapview FTP site. AQ: D

DNA pools were formed by combining equimolar amounts of each sample as described.<sup>14</sup> For pooled DNA assays, 25 ng of case and control DNA pools was used for amplification at each site. All polymerase chain reaction (PCR) and homogeneous MassEXTEND reactions were conducted using standard conditions.<sup>15</sup> Relative allele frequency was estimated on the basis of the area under the peak calculated from mass spectrometry measurements of four aliquots, as described.<sup>15</sup> For individual genotyping, the same procedure was applied with the following differences: only 2.5 ng DNA was used and only one mass spectrometry measurement was taken. AQ: E

The following gene-specific primers were used to genotype the marker SNP rs1862214: PCR primer 1—5'-ACGTTGGATGCCTGTTTCCTGAT-CACACC-3'; PCR primer 2—5'-ACGTTGGATGTGCAGTCTGAAAG-GAAACC-3'; extension primer—5'-CACACCACTCACTGCAC-3'.

**Cloning of PDCD5 From Human Lung Tissue**

Primers 5'-GAGCCATGGCGGACGAGGAGC-3' and 5'-ATAATCGT-CATCTTCATCAGAG-3' were used for amplification of the human *PDCD5* gene from lung tissue. Total RNA from nontumor lung parenchyma of two lung cancer patients was reverse-transcribed using ThermoScript (Invitrogen). PCR amplification was carried out in 50-μL volume containing aliquots of healthy lung cDNA, 100 μmol/L deoxynucleotide triphosphates, 15 pmol gene-specific primers, 1.5 mmol/L MgCl<sub>2</sub>, and 0.6 U Ampliqaq Gold DNA polymerase (Applied Biosystems, Foster City, CA). Cycling conditions were 20 seconds at 94°C, 20 seconds at 55°C, and 60 seconds at 72°C for 40 cycles. Amplified products in aliquots were subcloned into pEF6/V5-His TOPO vector in-frame with the V5 epitope and polyhistidine tag. Clones were screened and sequenced using vector-specific primers. AQ: F

**Real-Time Quantitative PCR**

Expression of *PDCD5* mRNA was quantitated by kinetic reverse-transcriptase PCR using Hs00270435\_m1 TaqMan Gene Expression Assay-on-Demand (Applied Biosystems). The real-time PCR amplification mixture contained 1 μL cDNA template; 10 μL 2× TaqMan Universal PCR Master Mix, No AmpErase UNG (Applied Biosystems); and 1 μL of 20× Assay-on-Demand Gene Expression Assay Mix (Applied Biosystems). The final volume was adjusted to 20 μL with RNase-free water. Reactions were run in duplicate on an ABI PRISM 7900HT Sequence Detection System (Applied Biosystems) for 40 cycles (95°C for 15 seconds, 60°C for 1 minute) after initial activation steps at 50°C for 2 minutes and 95°C for 10 minutes. The experiment was carried out in duplicate. The human hydroxymethylbilane synthase (*HMBS*) gene (Hs00609297\_m1 TaqMan Gene Expression Assay-on-Demand) was

PDCD5 Locus in Lung Cancer

used as control for possible differences in cDNA amounts. Relative differences (-fold) were calculated according to the comparative Ct method.

**Transfection and In Vitro Clonogenic Assay**

Human cell lines A549 and NCI-NCI-NCI-H520, derived from lung carcinoma and lung squamous cell carcinoma, respectively, were transfected with 2.5 µg DNA of recombinant or control vector (pEF6/V5-His-TOPO) using SuperFect Transfection Reagent (Qiagen) and Nupherin-neuron (Biomol, Plymouth Meeting, PA). Transfected clones were selected with blasticidin 5 µg/mL (Invitrogen) for 3 (A549) or 4 (NCI-H520) weeks, methanol-fixed and stained with 10% giemsa. The clonogenic assay was carried out in selective medium containing 10% serum; additionally, A549 cells were tested in low-serum (0.5% fetal bovine serum) medium.

**Statistical Analysis**

Association between disease status and each SNP using pooled DNA was tested as described.<sup>16</sup> Sources of measurement variation included pool formation, PCR/mass extension, and chip measurement. When three or more replicate measurements of an SNP were available within a model level, the corresponding variance component was estimated from the data. Otherwise, the following historical laboratory averages were used: pool formation =  $5.0 \times 10^{-5}$ , PCR/mass extension =  $1.7 \times 10^{-4}$ , and chip measurement =  $1.0 \times 10^{-4}$ . The Hardy-Weinberg equilibrium was tested by the  $\chi^2$  method.<sup>17</sup> Logistic regression was used to compute odds ratios (ORs) and 95% CIs. The Kaplan-Meier product-limit method, the log-rank test,<sup>18,19</sup> and the Cox regression model were used to evaluate the effect of the PDCD5 genotype on overall survival. The Kruskal-Wallis test was used to assess the association between genotype and age at cancer onset and to analyze mRNA levels of the PDCD5 gene, as well as cell colony formation data.

genotyping identified 160 SNPs that differed significantly ( $P < .05$ ) between cases and controls. To verify true genetic effects by independent replication, we also studied a group of 254 Italian lung cancer patients and 235 unaffected controls.

One of the loci identified was located on chromosome 19q12-q13.1. A C/G polymorphism (rs1862214) located 35 kb upstream of the PDCD5 gene was associated with lung cancer susceptibility. Analysis of the German discovery sample revealed an increased frequency of the rare G allele in lung cancer patients (0.29) as compared with controls (0.22; Table 3). Frequencies of the CG and GG genotypes were higher among patients (41% and 9%, respectively) than controls (37% and 3%, respectively). Hardy-Weinberg genotypic proportions were taken into account for in both groups (Table 3). A significant association with lung cancer risk was found in the discovery set of cases, with an OR for GG homozygous versus CC homozygous patients of 3.5 (95% CI, 1.3 to 9.5; Table 3). Whereas comparison of the common CC genotype with CG heterozygous or with the G allele carrier status revealed no statistically significant associations, the relative ORs were more than 1 (Table 3). In this sample, the presence of 32% female controls and the lack of matched female cases might have decreased the statistical association. Indeed, logistic analysis without sex adjustment indicated a significant association for the G allele carrier status (OR, 1.4; 95% CI, 1.03 to 2.01), and the homozygous comparison (GG v CC) showed a stronger association in the sex-unadjusted (data not shown) than in the sex-adjusted analysis.

Frequencies of the susceptibility G allele were similar in the control groups of both discovery and replication samples (0.22; Table 3). The frequency of the G allele was also increased in Italian lung cancer patients (0.27), with a detected OR of 2.1 (95% CI, 1.3 to 3.3) for patients carrying the G allele (GG or CG) as compared with CC homozygous patients (Table 3). A combined analysis of the two samples resulted in an OR of 1.6 (95% CI, 1.2 to 2.1;  $P = .0004$ ) for individuals carrying the G allele (Table 3). A trend toward an increased risk was observed from heterozygous to homozygous patients carrying the rare

RESULTS

**Identification of a Lung Cancer Susceptibility Locus on 19q12-q13.1**

Using a screening strategy based on DNA pooling,<sup>20,21</sup> we examined DNA samples from 321 German smokers clinically diagnosed with lung cancer and 273 age-matched healthy smokers (Tables 1 and 2) in a genome-wide study consisting of 83,715 SNPs. Individual

**Table 3.** Association of PDCD5 Polymorphism rs1862214 With Lung Cancer Risk

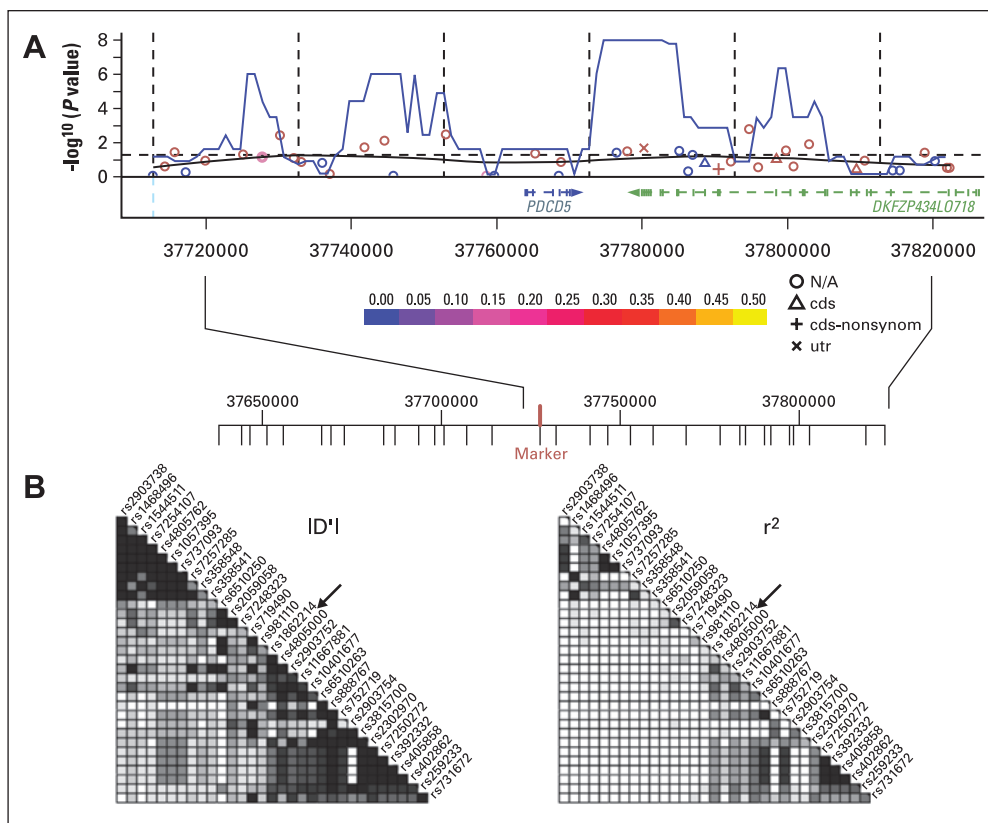
	No.		Odds Ratio	95% CI*
	Controls	Patients		
German sample				
CC	161	162	1.0	
CG	101	132	1.2	0.8 to 1.8
GG	8	28	3.5	1.3 to 9.5†
CG or GG	109	160	1.4	0.95 to 2.00
Frequency of the rare (G) allele	0.22‡	0.29‡		
Italian sample				
CC	138	124	1.0	
CG	86	115	2.1	1.3 to 3.3†
GG	9	11	1.9	0.6 to 6.1
CG or GG	97	126	2.1	1.3 to 3.3†
Frequency of the rare (G) allele	0.22‡	0.27§		
All samples				
CC	299	286	1.0	
CG	187	247	1.5	1.1 to 2.0†
GG	17	39	2.9	1.5 to 5.7†
CG or GG	204	286	1.6	1.2 to 2.1†
Frequency of the rare (G) allele	0.22‡	0.28‡		

\*Adjusted for sex, smoking habit, and age (in decades).

† $P < .05$ .

‡Nonsignificant deviation from the Hardy-Weinberg equilibrium.

§Significant ( $P = .01$ ) deviation from the Hardy-Weinberg equilibrium.



**Fig 1.** Association fine-mapping of lung cancer susceptibility region on chromosome 19q12-q13.1. (A) Forty-five public domain single nucleotide polymorphisms (SNPs) in a 110-kb window around the initial marker SNP (red arrow) were compared between discovery case and control pools. Sixteen SNPs were significant at  $P < .05$  (horizontal dashed line). The x-axis corresponds to their chromosomal position; the y-axis, to the test  $P$  values ( $-\log_{10}$  scale). The continuous blue line presents the results of a goodness-of-fit test for an excess of significance (compared with 0.05) in a 10-kb sliding window assessed at 1-kb increments. The continuous light gray line is the result of a nonlinear smoothing function showing a weighted average of the  $P$  values across the region. The Entrez Gene annotations for National Center for Biotechnology Information Genome Build 34 are included. (B) Estimates of linkage disequilibrium (LD) from HapMap CEPH30 data in a 180-kb region encompassing *PDCD5*. Estimates of LD ( $|D'|$  and  $r^2$ ) are represented as gray-scale ranging from white (LD = 0) to black (LD = 1) at increments of 0.1. SNP locations are indicated as downward tick marks in the ruler above. Arrow indicates the marker SNP rs1862214 location.

allele, with ORs of 1.5 (95% CI, 1.1 to 2.0) and 2.9 (95% CI, 1.5 to 5.7), respectively (Table 3).

**LD in the PDCD5 Gene Region**

To fine-map the region of association, we tested 44 additional SNPs located within 50 kb to either side of the *PDCD5* gene using the discovery pools (Fig 1). Sixteen of the 45 SNPs tested were significantly associated with lung cancer risk ( $P < .05$ ). The region of highest significance spanned approximately 80 kb and included the *PDCD5* gene and the 3' end of the ankyrin repeat domain 27 (VPS9 domain) *ANKRD27* gene (Fig 1A). This is in agreement with data from the CEPH30 sample from the HapMap project<sup>11</sup> and confirms the determined LD block (Fig 1B). Although little is known about the biologic functions of the *ANKRD27* gene, the *PDCD5* gene is known to be involved in apoptosis,<sup>22</sup> and is therefore a likely candidate for the observed effect on lung cancer risk.

**Association of rs1862214 SNP With Lung3 Cancer Prognosis**

Analysis of the rs1862214 polymorphism with respect to prognostic factors in Italian lung cancer patients revealed a significant positive association between the rare G allele and advanced clinical stage, nodal status, or overall survival. Indeed, the allele G carrier status showed a significant association with higher clinical stages (hazard ratio [HR], 1.9; 95% CI, 1.1 to 3.4;  $P = .023$ ). The association was statistically significant also for the heterozygous genotype group (HR, 1.9; 95% CI, 1.1 to 3.4;  $P = .028$ ) but not for the GG homozygous group (HR, 3.3; 95% CI, 0.7 to 15.3). Although the HR was higher in the latter group, the lack of association might be due to the low number of patients with the GG genotype and the small number of events detected in this group (Table 4).

Presence of lymph node metastases was more frequent in patients carrying the G allele than in CC homozygous patients (HR, 1.9; 95% CI, 1.1 to 3.6;  $P = .033$ ), and the CG and GG groups maintained an excess value over CC homozygous patients, although the

**Table 4.** Association Between *PDCD5* Polymorphism rs1862214 and Clinicopathologic Features of Italian Lung Cancer Patients

Variable	No.		Hazard Ratio	95% CI†
	Patients	Events*		
Overall survival				
CC	108	47	1.0	
CG	98	61	1.7	1.2 to 2.5‡
GG	10	8	2.4	1.1 to 5.6‡
CG or GG	108	69	1.8	1.2 to 2.6‡
Clinical stage§				
CC	106	66	1.0	
CG	108	53	1.9	1.1 to 3.4‡
GG	11	4	3.3	0.7 to 15.3
CG or GG	119	57	1.9	1.1 to 3.4‡
Nodal status¶				
CC	106	77	1.0	
CG	101	63	1.9	1.0 to 3.5
GG	10	4	5.2	1.0 to 26.7
CG or GG	111	67	1.9	1.1 to 3.6‡

\*Number of deaths, or number of patients with clinical stage I, or number of patients without nodal involvement (NO), respectively.

†Adjusted for sex, smoking, age at diagnosis (in decades).

‡ $P < .05$ .

§Stage I v higher stages.

¶NO versus  $N \geq 1$ .

AQ: M

AQ: M

FI

T4

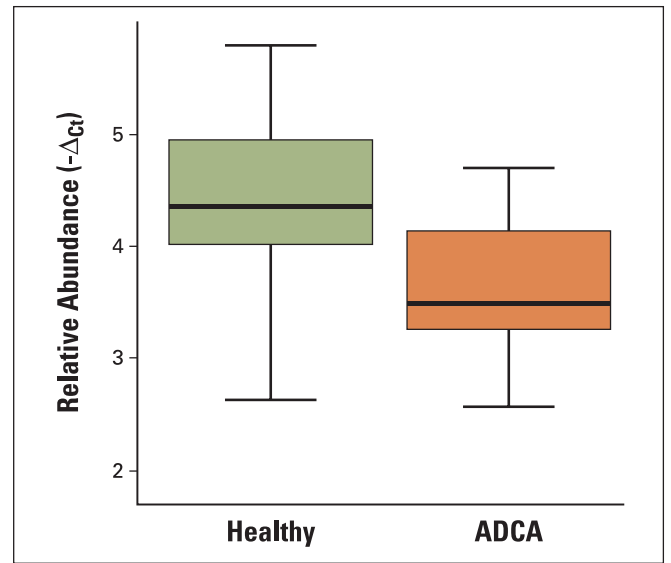
PDCD5 Locus in Lung Cancer

statistical significance was borderline ( $P = .054$  and  $P = .051$ , respectively; Table 4).

Median follow-up of patients alive at the end of the follow-up period was 56.5 months (Table 2). Kaplan-Meier survival curves analysis showed a significant difference among the three genotype groups (log-rank  $P = .0013$ ), with median survival times of 79.8, 40.2, and 19.0 months associated with the CC, CG, and GG genotypes, respectively (Fig 2). Cox proportional hazard analysis of survival, adjusted for sex, smoking habits, and age at diagnosis (in decades) showed that patients carrying the CG and GG genotype had a poorer survival than patients with the CC genotype, with a trend in excess HR associated with copy number of the G allele (HR, 1.7; 95% CI, 1.2 to 2.5;  $P = .007$  for the CG genotype and HR, 2.4; 95% CI, 1.1 to 5.6;  $P = .034$  for the GG genotype; Table 4). Overall comparison of patients carrying the G allele (at homozygosity or heterozygosity) versus CC homozygous patients revealed a significant excess of deaths associated with the G allele carrier status (HR, 1.8; 95% CI, 1.2 to 2.6;  $P = .003$ ; Table 4), with a median follow-up at death of 40.2 months for the G allele carriers and of 79.8 months for CC homozygous ADCA patients.

**PDCD5 mRNA Expression Is Downregulated in Lung Cancer**

Because genomic DNA and RNA were available from nontumor and tumor lung tissue of 46 Italian lung adenocarcinoma patients, samples with different rs1862214 genotypes (24 CC, 17 CG, and five GG) were analyzed for *PDCD5* expression to identify a possible correlation between SNP genotypes and mRNA levels. However, the mRNA levels in healthy lung tissue were similar in the common CC genotype as compared with G-allele carriers (data not shown). Comparison of *PDCD5* mRNA levels in nontumor normal and tumor lung tissue from the same patient revealed approximately 2.4-fold lower expression in tumor tissue (median  $-\Delta Ct$  values 4.4 [SE, 0.2] and 3.5 [SE, 0.2] in nontumor and tumor tissue, respectively; Fig 3;  $P = .006$ ).

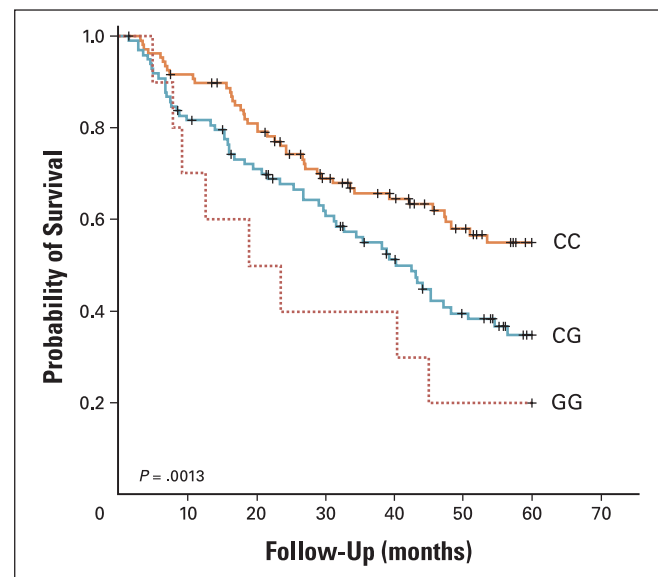


**Fig 3.** *PDCD5* mRNA expression levels in normal and tumor lung tissue of lung adenocarcinoma (ADCA) patients. *PDCD5* mRNA levels were detected by kinetic reverse-transcriptase polymerase chain reaction and normalized against those of the housekeeping gene *HMBS*. The horizontal line within the box represents the median  $-\Delta Ct$  value; the upper and lower boundaries of the box represent 75th and 25th percentile, respectively; the upper and lower bars indicate the largest and smallest observed values, respectively.

Differences between nontumor and tumor tissue in *PDCD5* mRNA levels did not correlate with the genotype at the rs1862214 polymorphism (data not shown).

**PDCD5 Overexpression Inhibits Colony Formation In Vitro**

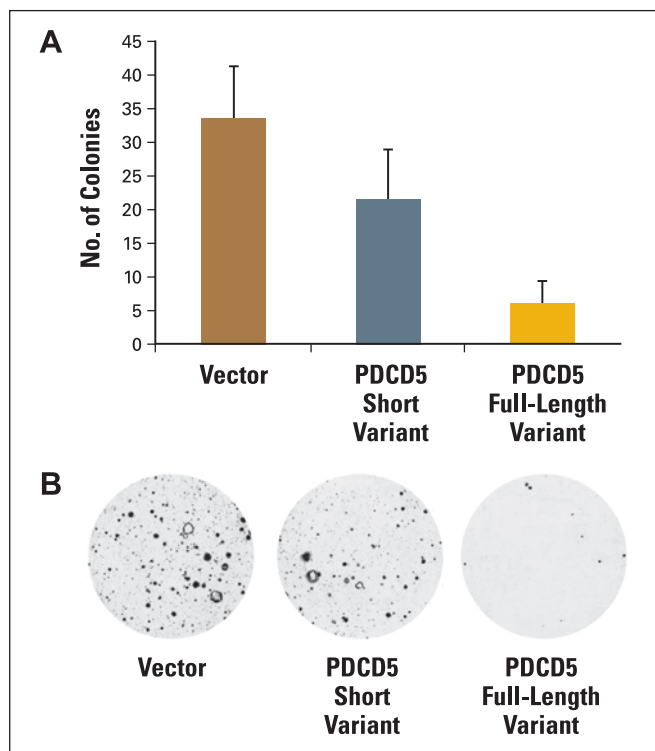
Amplification of the *PDCD5* gene from human lung cDNA produced two different transcripts: the annotated variant NM\_004708, consisting of six exons and encoding a protein of 125 amino acids, and an additional alternatively spliced variant that lacks exons 3 to 5 and is translated into a truncated protein product of 40 residues (GenBank Access No. DQ208400; <http://www.ncbi.nlm.nih.gov/Genbank/>). Mammalian expression vectors containing *PDCD5* isoforms were prepared and transfected into A549 and NCI-H520 human lung cancer cell lines, whose endogenous levels of *PDCD5* mRNA were approximately 0.8- and 0.3-fold lower, respectively, as compared with a pool of healthy lung tissue from 20 patients (data not shown). Overexpression of either *PDCD5* transcript in the human lung cancer cell line A549 did not lead to significant variation in number or size of colonies grown in both high- and low-serum conditions (data not shown). By contrast, a significant inhibitory effect on colony formation was observed in NCI-H520 cells, with the number of colonies more than 0.5 mm in diameter approximately 6-fold lower ( $P = .006$ ) in cells transfected with the full-length *PDCD5* transcript than in control cells (Fig 4A and B). Cells overexpressing the short *PDCD5* variant showed an approximately 1.5-fold reduction in clonogenicity, but the difference from the vector-transfected cells was not statistically significant (Fig 4A and B).



**Fig 2.** Kaplan-Meier survival curves in Italian lung adenocarcinoma patients. The curves of patients with *PDCD5* genotype CC ( $n = 108$ ), CG ( $n = 98$ ), or GG ( $n = 10$ ) are shown as orange, blue, and pink lines, respectively. Follow-up is shown truncated at 60 months. Small crosses indicate censored observations. Log-rank analysis indicates significant differences between curves ( $P = .0013$ ).

**DISCUSSION**

Genetic loci affecting lung cancer risk have been identified in mouse models, but the relationship with the risk of human lung cancer has



**Fig 4.** Effect of the full-length *PDCD5* variant on growth of NCI-H520 human lung cancer cell transfectants. (A) Mean  $\pm$  SE (n = 7) number of colonies more than 0.5 mm in diameter in a cell colony formation assay of cells transfected with the control vector, the short ( $P = .18$ ) or the full-length *PDCD5* transcript, respectively ( $P = .006$ ). (B) Clonogenic assay of the respective transfectants.

not yet been established. One reason for the gap between the characterization of genetic factors involved in this disease in humans and experimental models is the difficulty in carrying out family-based studies for lung cancer because this disease rarely clusters in families and has a poor prognosis, making a sib-pair study design difficult. The strong association of lung cancer risk with smoking habits further complicates family-based linkage studies.

Despite these difficulties, a recent genetic linkage study in lung cancer families reported a significant linkage of a single locus on chromosome 6q23-25 to lung cancer risk.<sup>23</sup> This result supports the idea that genetic factors influence lung cancer risk in humans, but does not rule out a potential polygenic mode of inheritance of such risk, in analogy with the mouse model. Indeed, the observation of a single locus linked to lung cancer risk may simply reflect the relatively low power of that study to detect additional loci. Because the study in pedigrees included rare families with multiple lung cancer cases, whereas the large majority of lung cancer cases occur without familial clustering, the role of this locus in nonfamilial lung cancer by population-based association studies remains to be established.

In our present genome-wide association study, we identified several chromosomal regions putatively associated with lung cancer

risk in smokers. Replication of these findings in an independent sample for which follow-up data were available verified the validity of the observed associations and their importance in different populations. We identified a significant association of a SNP on chromosome 19, 35 kb 5' of the *PDCD5* gene, with lung cancer risk. The association was first observed in a collection of German lung cancer cases consisting only of male smokers with a range of histologic subtypes and later confirmed in an independent population of Italian lung ADCA patients of both sexes, with mixed smoking habits. Despite the heterogeneity of the discovery and replication samples, the significant association between the rs1862214 SNP and lung cancer risk in both populations points to an important role for this chromosomal region in lung cancer predisposition. Moreover, in the replication sample, the same allele (G) associated with an increased lung cancer risk (Table 3) was also associated with poor survival, higher clinical stage, and nodal involvement (Table 4). This finding further supports the involvement of the *PDCD5* region in the etiology of lung cancer.

The rs1862214 variation is part of a haplotype block that contains the *PDCD5* gene. This gene is a strong candidate for the modulation of lung cancer risk because of its described biologic functions and its involvement in the regulation of cell apoptosis. Indeed, *PDCD5* levels are significantly increased in cells undergoing apoptosis, and the rapid translocation of the protein from the cytoplasm to the nucleus of apoptotic cells precedes the externalization of phosphatidylserine and the fragmentation of chromosome DNA.<sup>24</sup> Moreover, the introduction of anti-*PDCD5* antibody suppresses apoptosis,<sup>25</sup> supporting the critical role of *PDCD5* in this pathway.

Since transcription may also be affected by long-range regulatory sequences, we investigated whether the rs1862214 SNP, which is 35 kb upstream of the *PDCD5* gene, has a functional role in *PDCD5* mRNA expression in normal and tumor lung tissue. However, in healthy lung parenchyma *PDCD5* mRNA levels were similar in individuals with the common CC genotype as compared with G-allele carriers.

Preliminary functional studies in vitro using human lung cancer cell lines transfected with recombinant mammalian expression vectors for the *PDCD5* gene indicated that overexpression of this gene significantly inhibits colony growth in NCI-H520 cells but does not modulate colony formation in A549 cells. The reasons for such cell-specific effect are not known. Perhaps the ability of *PDCD5* in modulating clonogenicity may be related to specific somatic alterations of cancer cells.

Analysis of a shorter mRNA variant encoding only the 40 N-terminal amino acid residues of *PDCD5* in NCI-H520 cells showed that the inhibitory activity of *PDCD5* is almost completely lost, indicating that important functional domains are present in the truncated portion. However, mRNA levels of this shorter transcript were approximately 10-fold lower than those of the full-length variant in nontumor lung tissue, and the relative quantity of the two isoforms did not differ in lung tumor tissue (data not shown). Additional studies, including silencing of the *PDCD5* gene and studies in animal models, may provide further insight into the role of *PDCD5* in lung cancer risk and development.

**REFERENCES**

1. Braun MM, Caporaso NE, Page WF, et al: Genetic component of lung cancer: Cohort study of twins. *Lancet* 344:440-443, 1994
2. Ahlborn A, Lichtenstein P, Malmstrom H, et al: Cancer in twins: Genetic and nongenetic familial risk factors. *J Natl Cancer Inst* 89:287-293, 1997
3. Sellers TA, Potter JD, Bailey-Wilson JE, et al: Lung cancer detection and prevention: Evidence for an interaction between smoking and genetic predisposition. *Cancer Res* 52:2694s-2697s, 1992
4. Sellers TA, Bailey-Wilson JE, Elston RC, et al: Evidence for Mendelian inheritance in the pathogenesis of lung cancer. *J Natl Cancer Inst* 82:1272-1279, 1990

### PDCD5 Locus in Lung Cancer

5. Yang P, Schwartz AG, McAllister AE, et al: Lung cancer risk in families of nonsmoking probands: Heterogeneity by age at diagnosis. *Genet Epidemiol* 17:253-273, 1999
6. Gauderman WJ, Morrison JL: Evidence for age-specific genetic relative risks in lung cancer. *Am J Epidemiol* 151:41-49, 2000
7. Wu PF, Lee CH, Wang MJ, et al: Cancer aggregation and complex segregation analysis of families with female non-smoking lung cancer probands in Taiwan. *Eur J Cancer* 40:260-266, 2004
8. Dragani TA: 10 Years of mouse cancer modifier loci: Human relevance. *Cancer Res* 63:3011-3018, 2003
9. Dragani TA, Canzian F, Pierotti MA: A polygenic model of inherited predisposition to cancer. *FASEB J* 10:865-870, 1996
10. Wall JD, Pritchard JK: Haplotype blocks and linkage disequilibrium in the human genome. *Nat Rev Genet* 4:587-597, 2003
11. The International HapMap Consortium: The International HapMap Project. *Nature* 426:789-796, 2003
12. Vieland VJ: The replication requirement. *Nat Genet* 29:244-245, 2001
13. Nelson MR, Marnellos G, Kammerer S, et al: Large-scale validation of single nucleotide polymorphisms in gene regions. *Genome Res* 14:1664-1668, 2004
14. Buetow KH, Edmonson M, MacDonald R, et al: High-throughput development and characterization of a genomewide collection of gene-based single nucleotide polymorphism markers by chip-based matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Proc Natl Acad Sci U S A* 98:581-584, 2001
15. Bansal A, van den Boom D, Kammerer S, et al: Association testing by DNA pooling: An effective initial screen. *Proc Natl Acad Sci U S A* 99:16871-16874, 2002
16. Barratt BJ, Payne F, Rance HE, et al: Identification of the sources of error in allele frequency estimations from pooled DNA indicates an optimal experimental design. *Ann Hum Genet* 66:393-405, 2002
17. Weir BS: *Genetic Data Analysis 2: Methods for Discrete Population Genetic Data*. Sunderland, MA, Sinauer Associates Inc, 1996
18. Kaplan EL, Meier P: Non-parametric estimation from incomplete observations. *J Am Stat Assoc* 3:457-481, 1958
19. Peto R, Peto J: Asymptotically efficient rank invariant test procedures. *J R Stat Soc* 135:185-206, 1972
20. Kammerer S, Roth RB, Hoyal CR, et al: Association of the NuMA region on chromosome 11q13 with breast cancer susceptibility. *Proc Natl Acad Sci U S A* 102:2004-2009, 2005
21. Kammerer S, Roth RB, Reneland R, et al: Large-scale association study identifies ICAM gene region as breast and prostate cancer susceptibility locus. *Cancer Res* 64:8906-8910, 2004
22. Liu H, Wang Y, Zhang Y, et al: TFAR19, a novel apoptosis-related gene cloned from human leukemia cell line TF-1, could enhance apoptosis of some tumor cells induced by growth factor withdrawal. *Biochem Biophys Res Commun* 254:203-210, 1999
23. Bailey-Wilson JE, Amos CI, Pinney SM, et al: A major lung cancer susceptibility locus maps to chromosome 6q23-25. *Am J Hum Genet* 75:460-474, 2004
24. Chen Y, Sun R, Han W, et al: Nuclear translocation of PDCD5 (TFAR19): An early signal for apoptosis? *FEBS Lett* 509:191-196, 2001
25. Rui M, Chen Y, Zhang Y, et al: Transfer of anti-TFAR19 monoclonal antibody into HeLa cells by in situ electroporation can inhibit the apoptosis. *Life Sci* 71:1771-1778, 2002

### Acknowledgment

Some of the nonsmoking patients were recruited with the help of Associazione Marta Nurizzo, Brugherio, Italy.

### Authors' Disclosures of Potential Conflicts of Interest

The authors indicated no potential conflicts of interest.

### Author Contributions

**Conception and design:** Monica Spinola, Peter Meyer, Tommaso A. Dragani, Andreas Braun  
**Provision of study materials or patients:** Peter Meyer, Ugo Pastorino, Karl Haeussinger, Rainer Dierkesmann, Tommaso A. Dragani  
**Collection and assembly of data:** Monica Spinola, Stefan Kammerer, Felicia S. Falvella, Melanie B. Boettger, Carolyn Hoyal, Carmen Pignatiello, Reiner Fischer, Richard Roth, Matthew Nelson, Tommaso A. Dragani, Andreas Braun  
**Data analysis and interpretation:** Monica Spinola, Peter Meyer, Stefan Kammerer, Felicia S. Falvella, Melanie B. Boettger, Carolyn Hoyal, Reiner Fischer, Richard Roth, Matthew Nelson, Rainer Dierkesmann, Tommaso A. Dragani, Andreas Braun  
**Manuscript writing:** Tommaso A. Dragani, Andreas Braun  
**Final approval of manuscript:** Monica Spinola, Peter Meyer, Stefan Kammerer, Felicia S. Falvella, Melanie B. Boettger, Carolyn Hoyal, Carmen Pignatiello, Reiner Fischer, Richard Roth, Ugo Pastorino, Karl Haeussinger, Matthew Nelson, Rainer Dierkesmann, Tommaso A. Dragani, Andreas Braun

### GLOSSARY

**Linkage disequilibrium:** Nonrandom association of linked genes. This is the tendency of the alleles of two separate but already linked loci to be found together more frequently than would be expected by chance alone.

**PDCD5 (programmed cell death 5):** This gene encodes a protein expressed in tumor cells during apoptosis independent of the apoptosis-inducing stimuli. Before apoptosis induction, this gene product is distributed in both the nucleus and cytoplasm. Once apoptosis is induced, the level of this protein increases and by relocation from the cytoplasm, it accumulates in the nucleus. Although its exact function is not defined, this protein is thought to play an early and universal role in apoptosis. *PDCD5* is also called *TFAR19* (TF1 cell apoptosis-related gene 19).

**SNP (single nucleotide polymorphism):** Genetic polymorphisms are natural variations in the genomic DNA sequence present in greater than 1% of the population, with SNP representing DNA variations in a single nucleotide. SNPs are being widely used to better understand disease processes, thereby paving the way for genetic-based diagnostics and therapeutics.

**TFAR19 (TF1 cell apoptosis-related gene 19):** This gene encodes a protein expressed in tumor cells during apoptosis independent of the apoptosis-inducing stimuli. Before apoptosis induction, this gene product is distributed in both the nucleus and cytoplasm. Once apoptosis is induced, the level of this protein increases and by relocation from the cytoplasm, it accumulates in the nucleus. Although its exact function is not defined, this protein is thought to play an early and universal role in apoptosis. *TFAR19* is also called *PDCD5* (programmed cell death 5).