

# Germline Polymorphisms Associated with Overall Survival in Lung Adenocarcinoma: Genome-Wide Analysis

Francesca Minnai , [Sara Noci](#) , [Martina Esposito](#) , [Marc A. Schneider](#) , Sonja Kobinger , Martin Eichhorn , Hauke Winter , Hans Hoffmann , Mark Kriegsmann , [Matteo Incarbone](#) , [Giovanni Mattioni](#) , [Davide Tosi](#) , [Thomas Muley](#) , [Tommaso A. Dragani](#) <sup>\*</sup> , [Francesca Colombo](#)

Posted Date: 22 August 2024

doi: 10.20944/preprints202408.1575.v1

Keywords: GWAS; SNPs; Cox; eQTL; NT5DC2; NAGK



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## Article

# Germline Polymorphisms Associated with Overall Survival in Lung Adenocarcinoma: Genome-Wide Analysis

Francesca Minnai <sup>1,2</sup>, Sara Noci <sup>3</sup>, Martina Esposito <sup>1</sup>, Marc A. Schneider <sup>4,5</sup>, Sonja Kobinger <sup>6</sup>, Martin Eichhorn <sup>6</sup>, Hauke Winter <sup>5,6</sup>, Hans Hoffmann <sup>7</sup>, Mark Kriegsmann <sup>5,8,§</sup>, Matteo Incarbone <sup>9</sup>, Giovanni Mattioni <sup>10</sup>, Davide Tosi <sup>10</sup>, Thomas Muley <sup>4,5</sup>, Tommaso A. Dragani <sup>3,\*^</sup> and Francesca Colombo <sup>1</sup>

<sup>1</sup> Institute for Biomedical Technologies, National Research Council, Segrate (MI), Italy

<sup>2</sup> Department of Medical Biotechnology and Translational Medicine (BioMeTra), Università degli Studi di Milano, Milan, Italy

<sup>3</sup> Department of Experimental Oncology, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy

<sup>4</sup> Translational Research Unit (STF), Thoraxklinik, University Hospital Heidelberg, Heidelberg, Germany

<sup>5</sup> Translational Lung Research Center (TLRC), German Center for Lung Research (DZL), Heidelberg, Germany

<sup>6</sup> Department of Thoracic Surgery, Thoraxklinik, University Hospital Heidelberg, Heidelberg, Germany

<sup>7</sup> Department of Thoracic Surgery, Klinikum rechts der Isar, Technische Universität München, Germany

<sup>8</sup> Institute of Pathology, University Hospital Heidelberg, Heidelberg, Germany

<sup>9</sup> Department of Surgery, Ospedale San Giuseppe, IRCCS Multimedica, Milan, Italy

<sup>10</sup> Thoracic Surgery and Lung Transplantation Unit, Foundation IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy

\* Correspondence: tom.dragani@aspidia.com

§ Present address: Pathology Wiesbaden, Ludwig-Erhard-Str. 100, 65199 Wiesbaden, Germany.

^ Present address: Aspidia Srl, Milan, Italy

**Simple Summary:** Our study aims to understand why some people with lung adenocarcinoma, a type of lung cancer, survive longer than others, even when their conditions seem similar. By studying the DNA of 1,464 patients, we found six specific genetic differences that appear to affect survival. These differences are located in parts of the DNA that don't directly make proteins but might control how certain genes are turned on or off. The goal is to use this information to predict which patients might have better outcomes and to develop more personalized treatment options. Further research is needed to confirm these results and to understand the underlying biological mechanisms, but our findings could eventually improve how we treat lung cancer and understand its outcomes better.

**Abstract:** Background/Objectives: Lung cancer remains a global health concern, with substantial variation in patient survival. Despite advances in detection and treatment, the genetic basis for the divergent outcomes is not understood. We investigated germline polymorphisms that modulate overall survival in 1,464 surgically resected lung adenocarcinoma patients. Methods: A multivariable Cox proportional hazard model was used to assess the association of more than 7 million polymorphisms with overall survival at the 60-month follow-up, considering age, sex, pathological stage, decade of surgery and principal components as covariates. Genes in which variants were identified were studied in silico to investigate functional roles. Results: Six germline variants passed the genome-wide significance threshold. These single nucleotide polymorphisms mapped to non-coding (intronic) regions on chromosomes 2, 3, and 5. The minor alleles of rs13000315, rs151212827, and rs190923216 (chr. 2, 3 and 5, respectively) were found to be independent negative prognostic factors. All six variants have been reported to regulate the expression of nine genes, seven of which are protein-coding, in different tissues. Survival-associated variants on chromosomes 2 and 3 were already reported to regulate the expression of *NT5DC2* and *NAGK*, with high expression associated with the minor alleles. High *NT5DC2* and *NAGK* expression in lung adenocarcinoma tissue were already shown to correlate with poor overall survival. Conclusions: This study highlights a potential regulatory role of the identified polymorphisms in influencing outcome and suggests a mechanistic link between these variants, gene expression regulation, and lung

adenocarcinoma prognosis. Validation and functional studies are warranted to elucidate the mechanisms underlying these associations.

**Keywords:** GWAS; SNPs; Cox; eQTL; *NT5DC2*; *NAGK*

---

## 1. Introduction

Lung cancer is the leading cause of cancer death worldwide [1]. Nonetheless, lung cancer survival has progressively improved, with the 3-year survival rate increasing from 22% in cases diagnosed in 2004-2006 to 33% for patients diagnosed in 2016-2018 [2]. The histotype with the greatest improvement in survival (from 25% to 38%) is lung adenocarcinoma. Gains in survival are mainly due to earlier detection [3], advanced surgical procedures [4], better staging [5] and, particularly for non-small-cell lung cancer, the advent of targeted therapy and immunotherapy [6]. Despite these improvements, there is large variability in overall survival, for unknown reasons. The variability in outcomes is even observed among patients with the same tumor histotype and stage, and it is already known that age, sex and pathological stage are independent prognostic factors [7]. However, they do not fully explain the individual variability in prognosis due to the multifaceted nature of the disease [8].

It has been hypothesized that other prognostic factors affect survival, including a patient's genetic background. One possible explanation for the different outcomes is that individual germline polymorphisms modulate still unknown genetic mechanisms affecting cancer growth and metastasis. To identify such polymorphisms, two analytical approaches can be used. In the candidate-gene approach, knowledge of pathological mechanisms prompts the search for polymorphisms in genes whose products are believed to be involved in cancer survival. This approach has already identified polymorphisms in genes that influence overall survival of lung cancer patients [9–16], but replication of the findings is often lacking. The other approach to investigate the role of genetics in lung cancer survival uses unsupervised, genome- or exome-wide methods. So far, two studies of this type have identified polymorphisms and low frequency variants associated with survival, although at a low statistical significance level [17,18].

Since survival of lung cancer patients is a complex phenotype, the genome-wide approach is preferable to the candidate-gene approach, because it allows the exploration of many variants in almost all genes and also in non-coding regions of the genome. The exome-wide approach is good for studying rare, low frequency variants, but it does not provide information on non-coding regulatory variants. Genome- and exome-wide analyses require a large sample to achieve sufficient statistical power. As a result, studies that investigate survival in a homogeneous group of patients (e.g., those with the same histotype and genetic background) are often limited by an inadequate sample size. Thus, to get statistically robust results, we combined two European case series for a large genome-wide association study (GWAS) of 1,464 lung adenocarcinoma patients, and explored, using a Cox model, the association of 7,265,396 imputed germline polymorphisms with overall survival at 60 months.

## 2. Materials and Methods

### 2.1. Case Series and Research Ethics

The study investigated two cases series of surgically resected lung adenocarcinoma patients from hospitals in the area around Milan, Italy, and in Heidelberg, Germany. Patients in Italy were enrolled between 1992 and 2022 at the Fondazione IRCCS Istituto Nazionale dei Tumori, San Giuseppe Hospital, and Fondazione IRCCS Cà Granda Ospedale Maggiore Policlinico. Patients in Germany were recruited at the Thoraxklinik between 2006 and 2015. These case series are a subset of those analyzed in our previous study on lung adenocarcinoma prognostic factors in 3,078 patients [19].

Patients provided written informed consent to the use of their biological samples and data for research purposes, according to the European General Data Protection Regulation. The study was conducted in accordance with the Declaration of Helsinki and approved by the ethics committees of Fondazione IRCCS Istituto Nazionale dei Tumori (INT 224-17, on 19 December 2017), Ospedale San Giuseppe, IRCCS Multimedica (346.2018, on 01 October 2018), Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico (202\_2019bis; on 12 March 2019) and University of Heidelberg (S-270/2001).

### 2.2. Clinical Data and Biological Samples

Clinical data were collected from all patients about sex, age at lung resection for adenocarcinoma, year of surgery, smoking habit, survival status 60 months after surgery, and pathological stage, based on the 6th to 8th editions of TNM staging criteria for lung cancer [20–22]. More in detail, the German patients were staged (or re-staged, for patients who had surgery before 2009) using criteria of the 7th edition, while the Italian patients were staged according to the edition that was valid at the moment of surgery. Smoking habit was reported as either “never smoker” or “ever smoker” (current or former smoker), because information on smoking cessation was not available for many patients.

Genomic DNA samples from patients recruited in Italy were already available at Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy, as they had been prepared from non-involved lung tissue using a DNeasy Blood and Tissue kit (Qiagen), as previously described [17]. This DNA was fluorimetrically quantified with the Quant-iT PicoGreen dsDNA assay kit on an M1000 multiplate reader (Tecan). For patients in Germany, buffy coats were available at the Thoraxklinik biobank and used to extract genomic DNA using a FlexiGene DNA kit (Qiagen); this DNA was quantified using a Nanodrop ND-1000 spectrophotometer.

### 2.3. Genome-Wide Genotyping

Genome-wide genotype data for the entire sample of 1,592 patients were collected separately on three subgroups and then merged for this study. Data were already available for 582 of the patients recruited in Italy, at Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy. These data had been obtained with Infinium Omni2.5-8 BeadChip microarrays (Illumina) on an Illumina HiScan System and using Illumina's BeadStudio software, as described [23,24]. DNA from the remaining 530 patients from Italy and the 480 patients from Germany was genotyped using Axiom Precision Medicine Research Arrays (PMRA; Thermo Fisher Scientific) on GeneTitan instruments at two genotyping service providers, i.e., Thermo Fisher Scientific (Santa Clara, United States) and the Functional Genomics facility of the Instituto de Investigaciones Biomédicas August Pi i Sunyer (IDIBAPS, Barcelona, Spain), for Italian and German samples, respectively. Axiom Analysis Suite software (Thermo Fisher Scientific) was used to call genotypes on these samples using the “best practice” workflow (except for the average call rate threshold  $\geq 97$ ).

Genotype data for the three subgroups were separately subject to preliminary genotype quality control (QC) using PLINK v.1.9 software [25] (Supplementary Figure 1). For per-sample QC, we used an identity-by-descent test (as described in [26]) to identify and exclude related patients and



duplicates. We also excluded samples with a call rate < 98%, with sex discrepancies, or with excess heterozygosity (heterozygosity rate outside the range  $\pm 0.20$ ). In per-marker QC, we removed variants with a genotyping call rate < 98%, a minor allele frequency (MAF) < 1%, or a Hardy-Weinberg equilibrium test  $P < 1.0 \times 10^{-6}$ .

Genotype imputation to whole-genome sequence (for autosomal variants) was carried out separately for the three subgroups using the Minimac4 algorithm on the TOPMed Imputation Server. Data were phased with Eagle v.2.4 software, GRCh38/hg38 was set as the array build, and TOPMed-r2 was set as the reference panel [27–30]. Genotypes imputed with an  $R^2 \geq 0.3$  were considered of low-quality imputation [31] and thus filtered out together with those having a MAF < 0.01. Finally, the three datasets were merged, retaining only biallelic variants with an imputed genotyping rate < 98%.

PLINK 2 software [32] was used to perform a principal components analysis (PCA) on the entire dataset. The first 10 principal components (PCs) were used as covariates in survival analyses. The first four PCs were compared with those of 2,504 samples from five populations (Africans, Americans, South-East Asians, East Asians, and Europeans) in the 1000 Genomes Project [33].

#### 2.4. Statistical Analyses

The Italian and German series were compared using the chi-squared test (categorical variables) and Kolmogorov-Smirnov test (quantitative variables). Survival analyses were done in univariable and multivariable Cox proportional hazard models [34], using the `coxph` function of the survival package [35] in R environment. The following variables were considered: age (as both a quantitative and categorical variable), sex, pathological stage, country of enrollment, decade of surgery, smoking habit, and genotyping array. Data were censored at 60 months of follow-up.

The GWAS survival analyses tested, in an additive model, the association between variants and patients' overall survival limited at 60 months of follow-up. The GenABEL package in R environment [36] was used to test the multivariable Cox proportional hazard model with genotypes, using the first 10 PCs, sex, age, decade of surgery and pathological stage as covariates. The genome-wide statistical threshold was set at  $P < 5.0 \times 10^{-8}$ . A suggestive threshold was considered at  $P < 1.0 \times 10^{-5}$ . The Benjamini-Hochberg method of false discovery rate (FDR) [37] was used to correct for multiple testing.

Kaplan-Meier curves were plotted using the `survfit` function of the same survival package, and the log-rank test was used to assess significance. These analyses were performed using genotypes coded as in a dominant model. A two-sided  $P$ -value < 0.05 was set as the statistical significance threshold for these analyses.

A multivariable Cox proportional hazard model with the significant clinical variables and the top-significant SNPs was also tested. A backward stepwise model selection, based on the Akaike information criterion (AIC), was performed using the `stepAIC` function of MASS package, in R, to identify independent prognostic factors. The significance threshold for this analysis was  $P < 0.05$ .

#### 2.5. In Silico Functional Analyses

The identified germline polymorphisms (associated with overall survival at  $P < 1.0 \times 10^{-5}$ ) were investigated for a possible regulatory role, by searching for them in two public expression quantitative-trait locus (eQTL) databases (accessed on 5 February 2024): GTEx (Analysis V8 release, GTEx\_Analysis\_v8\_eQTL\_EUR.tar) and eQTLGen [38] (<https://www.eqtlgen.org/cis-eqtls.html>).

Genes reported as being regulated by the variants that were here found to associate with survival (at  $P < 5.0 \times 10^{-8}$ ) were selected for further analysis as possible prognostic factors. The Kaplan-Meier Plotter [39] online tool was used to look for associations between these genes' expression levels and survival according to published gene expression data from lung adenocarcinoma tumor tissue (accessed on 12 February 2024). The following parameters (different from default settings) were used: follow-up threshold of 60 months, adenocarcinoma histology, and multivariable Cox regression with stage and sex as covariates. With these settings, the analyses were performed using data from 534 lung adenocarcinoma patients. High and low expression groups were defined by dichotomizing at

the median value of log<sub>2</sub>-transformed probe intensities. A two-sided *P* < 0.005 was set as the significance threshold for this analysis.

3. Results

Genomic DNA from 1,592 surgically resected lung adenocarcinoma patients was genotyped, but in the QC steps 113 samples were excluded (Supplementary Figure 1A). Thus, data for 1,479 patients were used in survival analyses. This cohort included 1,049 patients from Italy (71%) and 430 from Germany (29%) (Table 1). The patients had a median age at surgery of 65 years, and there was a slight abundance of males (62.4%). Only 15.3% had never smoked, while the remaining 80.4% were classified as ever smokers (current or former smokers). More than half of patients (52.6%) had an early-stage tumor (pathological stage I). Regarding the period in which lung resection had been done, less than 15% of cases were operated between 1992 and 2000, 37.8% were treated between 2001 and 2010, and 47.2% underwent surgery after 2010. The median follow-up period was 54 months, and by 60 months only 62.7% were still alive.

**Table 1.** Characteristics of patients with lung adenocarcinoma, by country of enrollment.

Characteristic	Total (n = 1479)	Italian series (n = 1049)	German series (n = 430)	<i>P</i>
Age at surgery, years, median (range)	65 (30–90)	66 (30–85)	63 (37–88)	< 0.001 <sup>a</sup>
Age group, years, n (%)				< 0.001 <sup>b</sup>
< 55	227 (15.3)	139 (13.3)	88 (20.5)	
55–64	496 (33.5)	347 (33.1)	149 (34.7)	
65–74	563 (38.1)	413 (39.3)	150 (34.9)	
<sup>3</sup> 75	193 (13.0)	150 (14.3)	43 (10.0)	
Sex, n (%)				0.004 <sup>b</sup>
Male	922 (62.4)	679 (64.7)	243 (56.5)	
Female	557 (37.6)	370 (35.3)	187 (43.5)	
Smoking habit, n (%)				0.67 <sup>b</sup>
Never	225 (15.3)	160 (15.3)	65 (15.1)	
Ever	1190 (80.4)	826 (78.7)	364 (84.7)	
Missing	64 (4.3)	63 (6.0)	1 (0.2)	
Pathological stage, n (%)				< 0.001 <sup>b</sup>
I	778 (52.6)	612 (58.7)	166 (38.6)	
II	257 (17.4)	174 (16.3)	83 (19.3)	
III	348 (23.6)	195 (18.5)	153 (35.6)	
IV	82 (5.5)	54 (5.2)	28 (6.5)	
Missing	14 (0.95)	14 (1.3)	0 (0)	
Decade of surgery, n (%)				0.82 <sup>b</sup>
Before 2000	221 (14.9)	221 (21.1)	0 (0)	
2001–2010	558 (37.8)	370 (35.3)	188 (43.7)	
After 2010	699 (47.2)	458 (43.6)	241 (56.0)	
Missing	1 (0.1)	0 (0)	1 (0.2)	
Follow-up, months, median (IQR)	54 (21–60 )	50 (20–60)	60 (23–60)	0.011 <sup>a</sup>
Survival status at 60 months, n (%)				0.002 <sup>b</sup>
Alive	928 (62.7)	685 (65.3)	243 (56.5)	
Dead	551 (37.3)	364 (34.7)	187 (43.5)	
Genotyping array, n (%)				< 0.001 <sup>b</sup>
Infinium Omni2.5-8				
All tumors	559 (37.7)	559 (53.3)	0 (0)	
Stage I tumors *	366 (65.5)	366 (65.5)	0 (0)	
Axiom PMRA				
All tumors	920 (62.3)	490 (46.7)	430 (100)	

Stage I tumors *	412 (44.8)	246 (50.2)	166 (38.6)
------------------	------------	------------	------------

<sup>a</sup> Kolmogorov-Smirnov test; <sup>b</sup> chi-squared test; IQR, interquartile range; PMRA, Precision Medicine Research Array; \* Percentage calculated on the subset of samples genotyped with the Infinium Omni2.5-8 array or Axiom PMRA.

The subgroups of patients enrolled in Italy and Germany were different (Table 1). Patients from Italy were older than those from Germany (median, 66 vs. 63 years; Kolmogorov-Smirnov test,  $P < 0.001$ ) and more likely to be male (64.7% vs. 56.5%, chi-squared test,  $P = 0.004$ ). This latter observation may be due to the selection period (later for German series): indeed, lung cancer in women has been increasing in recent years [40]. The proportions of never- and ever-smokers were similar between the groups. Most patients enrolled in Italy (58.7%) had a stage I tumor, while most of those in the German series had tumors of stages II-IV (61.4%, chi-squared  $P < 0.001$ ). Regarding the period of surgery, the German series did not comprise patients operated before 2006, but there was no significant difference in the proportion of patients enrolled in the 2000's or after 2010 between the two series (chi-squared  $P = 0.82$ ). The median follow-up period was shorter for patients in Italy than Germany (50 vs. 60 months; Kolmogorov-Smirnov test  $P = 0.011$ ); this difference was mainly due to the incomplete 60-month follow-up for the most recently enrolled patients from Italy. A greater percentage of patients from Italy were alive at the 60-month follow-up (65.3% vs. 56.5%; chi-squared test  $P = 0.002$ ); this might be due to the longer median follow-up period or the higher percentage of more advanced stages in the German series. Finally, all patients from Germany were genotyped using the Axiom PMRA array, while the genotyping of patients from Italy was done with either the Infinium Omni2.5-8 or Axiom PRMA array.

To identify factors that affect the survival of lung adenocarcinoma patients, we first did a survival analysis using univariable proportional hazard Cox regression (Table 2). Sex, pathological stage, and the decade of surgery were the most significant prognostic factors. Survival probability also depended on the type of genotyping array, with a lower risk of death for cases genotyped on the Infinium Omni2.5-8 array. This difference may be attributed to the fact that approximately two thirds of patients genotyped with these arrays had stage I tumors (Table 1). To identify independent prognostic factors, we also did multivariable proportional hazard Cox regression. Due to missing data for 75 patients, this model was run on 1,404 patients. In this analysis, age, sex, pathological stage and the decade of surgery were independent prognostic factors, whereas the country of enrollment and genotyping array did not associate with survival. In detail, the mortality risk increased with age (hazard ratio [HR] = 1.02,  $P < 0.001$ ). The prognostic effect of age was more evident when this variable was treated as categorical: indeed, patients in the age groups 65-74 and  $\geq 75$  had higher mortality risks than younger patients (HR = 1.33 and HR = 1.82, respectively). Prognosis was better for females than males (HR = 0.66,  $P < 0.001$ ). Increasing pathological stage was associated with the highest mortality risk, with 2-, 4-, and about 6-fold higher HRs for stage II, III and IV tumors, respectively, than patients with stage I tumors ( $P < 0.001$ ). Overall survival was longer for patients who underwent resection more recently, with a consequential drop in HR from 1.0 for patients treated before 2000 to 0.68 for those treated between 2001 and 2010 and 0.48 for those treated after 2010 ( $P < 0.001$ ).

**Table 2.** Factors associated with overall survival, according to univariable and multivariable Cox regression.

Characteristic	Univariable analyses <sup>a</sup>		Multivariable analyses <sup>b</sup>	
	HR (95% CI)	Cox P	HR (95% CI)	Cox P
Age, years	1.00 (1.00–1.01)	0.40	1.02 (1.01–1.03)	2.93 × 10 <sup>-5</sup>
Age group, years				
< 55	1.00		1.00	
55-64	0.84 (0.66–1.08)	0.18	1.06 (0.82-1.37)*	0.62*
65-74	0.91 (0.71–1.16)	0.45	1.33 (1.03-1.71)*	0.028*
<sup>3</sup> 75	1.08 (0.80–1.46)	0.62	1.82 (1.31-2.52)*	2.98 × 10 <sup>-4</sup> *
Sex				

Male	1.00		1.00	
Female	0.61 (0.51–0.74)	1.88 x 10 <sup>-7</sup>	0.66 (0.54–0.81)	6.53 x 10 <sup>-5</sup>
Smoking habit				
Never	1.00		1.00	
Ever	1.25 (0.98–1.60)	0.078	1.05 (0.81–1.37)	0.72
Pathological stage				
I	1.00		1.00	
II	2.06 (1.61–2.64)	9.01 x 10 <sup>-9</sup>	2.05 (1.60–2.65)	2.38 x 10 <sup>-8</sup>
III	4.04 (3.30–4.94)	< 2.00 x 10 <sup>-16</sup>	4.14 (3.35–5.12)	< 2.00 x 10 <sup>-16</sup>
IV	5.95 (4.45–7.97)	< 2.00 x 10 <sup>-16</sup>	5.59 (4.12–7.57)	< 2.00 x 10 <sup>-16</sup>
Decade of surgery				
Before 2000	1.00		1.00	
2001–2010	0.77 (0.62–0.95)	0.017	0.68 (0.53–0.86)	0.0014
After 2010	0.45 (0.36–0.57)	2.5 x 10 <sup>-11</sup>	0.48 (0.37–0.63)	8.62 x 10 <sup>-8</sup>
Country				
Italy	0.85 (0.71–1.01)	0.063	0.96 (0.77–1.21)	0.73
Germany	1.00		1.00	
Genotyping array				
Infinium Omni2.5-8	0.79 (0.66–0.94)	0.0075	0.84 (0.68 –1.04)	0.11
Axiom PRMA	1.00		1.00	

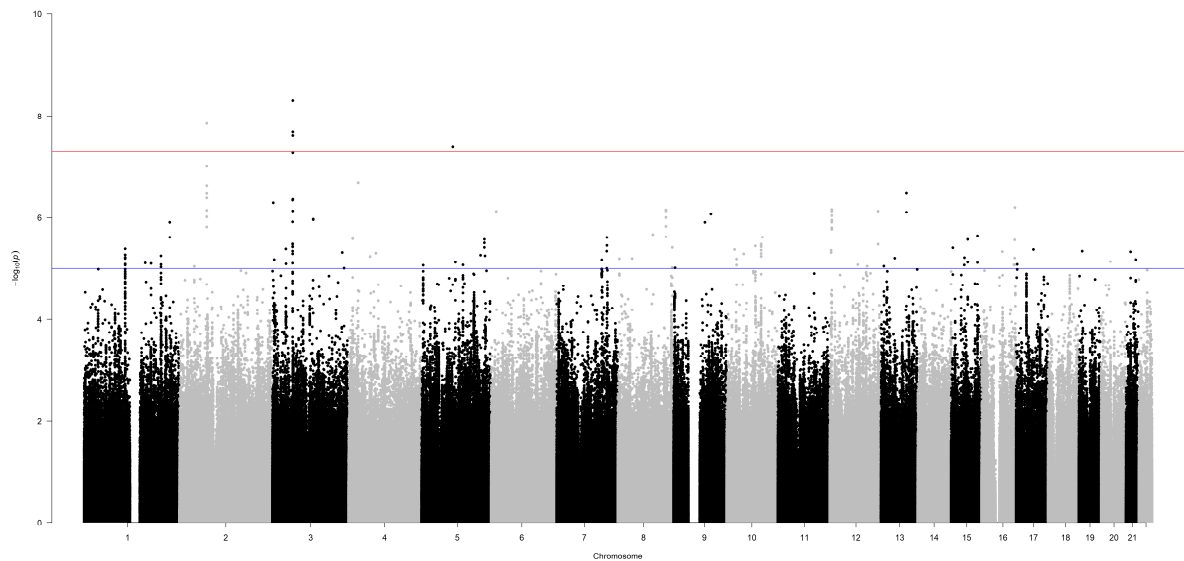
CI, confidential interval; HR, hazard ratio; <sup>a</sup>For 1,479 patients; <sup>b</sup>For 1,404 patients.; \* Results from a multivariable model with age coded as a categorical variable. The summary statistics of this model, for the other listed variables, were very similar to those shown in the table, which instead referred to the multivariable model with age as quantitative variable.

Based on these analyses, in the GWAS, Cox regression with genotypes was carried out using age, sex, pathological stage, and the decade of surgery as covariates, and the first 10 PCs to correct for population stratification. A plot of the first four PCs of our series, along those of 2,504 samples from five populations (Africans, Americans, South-East Asians, East Asians, and Europeans) in the 1000 Genomes Project [33] is given in Supplementary Figure 2, to visualize which ancestral group our patients belonged to.

3.1. Germline Variants Associated with Overall Survival

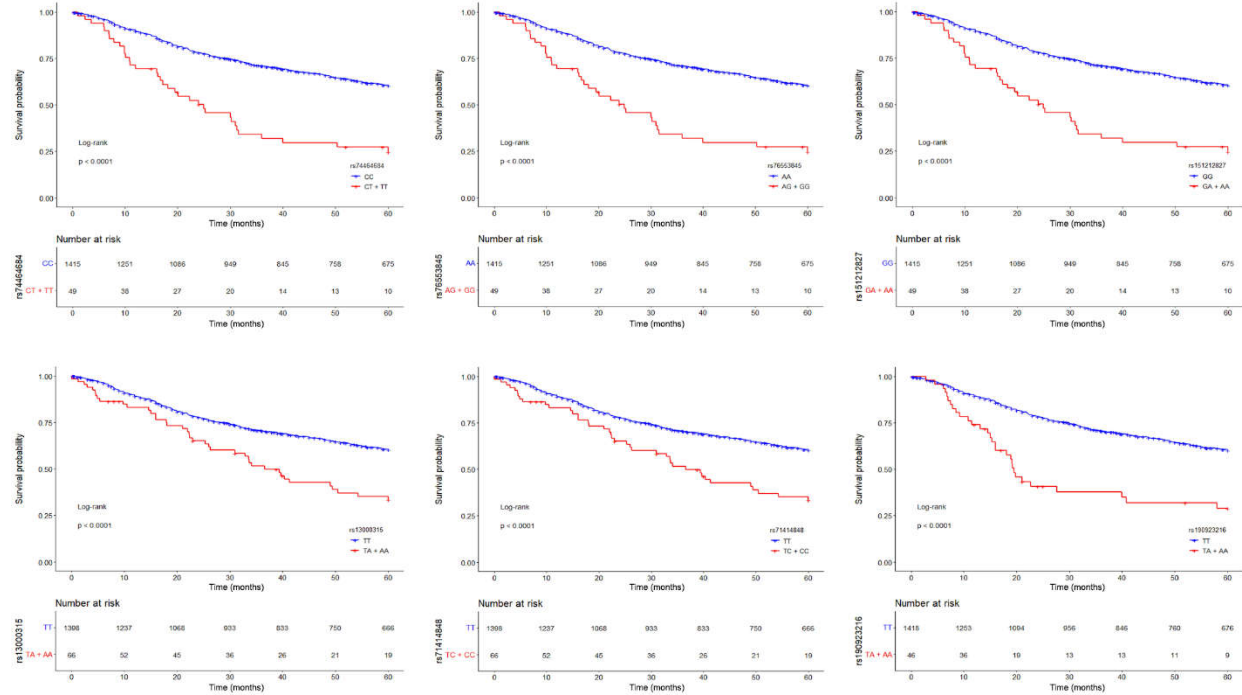
Genotype data of patients from the Italian and German series were used in the genome-wide survival analysis. After preliminary QC (Supplementary Figure 1B) and data imputation, the whole dataset comprised information on 7,265,396 germline polymorphisms and 1,464 patients (15 patients were excluded due to missing data on pathological stage or decade of surgery). Each variant was independently tested in an additive multivariable Cox model, and 224 single nucleotide polymorphisms (SNPs) were found to associate with overall survival at  $P < 1.0 \times 10^{-5}$  (Figure 1, Supplementary Table 1). Among them, six SNPs, on chromosomes 2, 3 and 5, passed the genome-wide statistical significance threshold ( $P < 5.0 \times 10^{-8}$ ). These SNPs are (in order of increasing  $P$ -value): rs74464684 (HR = 2.8), rs13000315 (HR = 2.5), rs71414848 (HR = 2.5), rs76553845 (HR = 2.7), rs151212827 (HR = 2.6), and rs190923216 (HR = 2.9). Because these SNPs have HRs greater than 1, their minor alleles are negative prognostic factors. An increasing number of their minor alleles associated with an at least 2-fold higher risk of death (HR > 2).





**Figure 1.** Manhattan plot of the results of the GWAS for overall survival, at 60 months of follow-up, of lung adenocarcinoma patients. Germline polymorphisms are plotted on the x-axis according to their genomic position (GChr 38, hg38 release) and on the y-axis according to their association with survival probability ( $-\log_{10}(P\text{-values})$ ). The horizontal red and blue lines represent the threshold of genome-wide significance ( $P < 5.0 \times 10^{-8}$ ) and a suggestive threshold at  $P < 1.0 \times 10^{-5}$ , respectively.

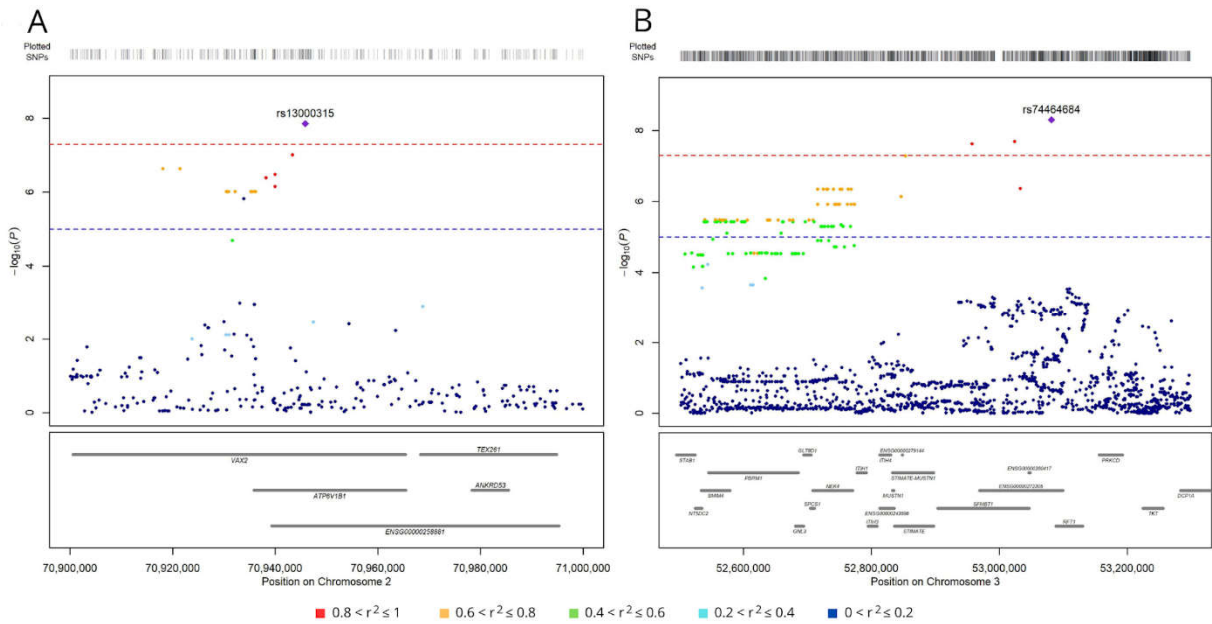
Since the MAF of these six SNPs was  $< 5\%$  (Supplementary Table 1), we grouped heterozygous patients with patients who were homozygous for the minor allele and drew Kaplan-Meier survival curves (Figure 2). Patients carrying at least one minor allele of each SNP had worse prognosis than patients homozygous for the major allele (log-rank test,  $P < 0.001$ ). Thus, these low-frequency variants have a dominant effect on the risk of death from lung adenocarcinoma.



**Figure 2.** Kaplan-Meier survival curves (at 60 months of follow-up) for lung adenocarcinoma patients according to the genotype of the six top-significant variants (rs74464684, rs76553845, rs151212827, rs13000315, rs71414848, and rs190923216, from top left to bottom right), coded as in a dominant model. Red lines represent patients homozygous for the minor alleles and heterozygotes, while blue lines represent patients homozygous for the major alleles. Crosses denote censored samples. Below

each plot are indicated the numbers of patients at risk in the genotype groups. Log-rank P-values are shown.

Five of the six top-ranking SNPs (excluding rs190923216 on chromosome 5) map near other variants associated with survival, although at a lower significance level ( $P < 1.0 \times 10^{-5}$ ). Indeed, the analysis identified 19 variants on chromosome 2 in a region  $< 30$  kbp (Figure 3A). The top-ranking variant, rs13000315, was in linkage disequilibrium (LD) with all the other variants ( $r^2 > 0.6$  and  $D' > 0.7$ ) except rs56354394. On chromosome 3, the analysis identified 279 variants in a region  $< 600$  kbp (Figure 3B), and the top-ranking variant, rs74464684, was in LD with the other variants ( $r^2 > 0.5$  and  $D' > 0.7$ ).



**Figure 3.** Regional association plots for SNPs associated with survival (at  $P < 1.0 \times 10^{-5}$ ) on chromosome 2 (A) and 3 (B). SNPs are plotted on the x-axis according to their genomic position (GChr 38, hg38 release), and  $-\log_{10}[P\text{-values}]$  for their association with overall survival are plotted on the y-axis. Horizontal dashed red and blue lines represent the genome-wide threshold of significance ( $P < 5.0 \times 10^{-8}$ ) and the suggestive threshold ( $P < 1.0 \times 10^{-5}$ ), respectively. Dot color represents the level of linkage disequilibrium, expressed as  $r^2$ , between each SNP and the top variants (purple diamonds): (A) rs74464684 and (B) rs13000315.

Finally, we tested, in a stepwise multivariable Cox model, the association of survival with age, sex, stage, decade of surgery and genotype of the six top-ranking SNPs, to understand if they were independent prognostic factors. The four clinical variables were all independent prognostic factors. Of the six top-ranking SNPs, three (rs13000315, rs151212827, and rs190923216) were independently associated with overall survival (Table 3).

**Table 3.** Hazard ratios for overall survival in 1,464 lung adenocarcinoma patients (15 patients were excluded due to missing data).

Variable	HR (95% CI)	Cox P
Age	1.02 (1.01 – 1.03)	2.8 x 10 <sup>-6</sup>
Sex		
Male	1.00	
Female	0.68 (0.56 – 0.82)	8.0 x 10 <sup>-5</sup>
Pathological stage		
I	1.00	

II	2.00 (1.56 – 2.57)	4.8 × 10 <sup>-8</sup>
III	4.34 (3.53 – 5.34)	< 2.0 × 10 <sup>-16</sup>
IV	6.24 (4.62 – 8.41)	< 2.0 × 10 <sup>-16</sup>
Decade of surgery		
Before 2000	1.00	
2001 – 2010	0.70 (0.56 – 0.88)	2.4 × 10 <sup>-3</sup>
After 2010	0.49 (0.39 – 0.63)	2.4 × 10 <sup>-8</sup>
Genomic variant <sup>a</sup>		
rs13000315	2.62 (1.92 – 3.56)	9.6 × 10 <sup>-10</sup>
rs151212827	2.32 (1.63 – 3.29)	2.8 × 10 <sup>-6</sup>
rs190923216	2.58 (1.75 – 3.79)	1.5 × 10 <sup>-6</sup>

HR, hazard ratio; CI, confidential interval; <sup>a</sup> SNPs were entered under an additive model, with the minor allele being the effect allele.

3.2. SNPs Associated with Lung Adenocarcinoma Survival Have Regulatory Roles

The 224 significant SNPs mapped in non-coding regions of the genome (mostly intronic; Supplementary Table 1), so we hypothesized that they participate in the regulation of gene expression. Therefore, we searched in two eQTL databases and found that 73 and 128 of them were already identified as eQTLs in GTEx and eQTLGen, respectively.

According to GTEx, the 73 SNPs regulate the expression of 34 mRNAs in 48 tissues, for a total of 1,125 eQTLs (Supplementary Table 2). Limiting the analysis to lung tissue, 16 SNPs (one on chromosome 5, six on chromosome 7, and nine on chromosome 10) had been reported as lung eQTLs of four genes: the minor allele of the SNP on chromosome 5 (rs190923216) was associated with higher expression levels of two mRNAs (the antisense RNA, *CKMT2-AS1*, and the pseudogene *RPS12*), the six SNPs on chromosome 7 affect the expression of *COPG2*, and the nine SNPs on chromosome 10 influence the expression of a long non-coding RNA, *LINC00865*.

According to eQTLGen, the 128 SNPs regulate the expression of 43 unique mRNAs in blood, for a total of 543 eQTLs (Supplementary Table 3). Of note, the top five ranking SNPs from our genome-wide analysis (on chromosomes 2 and 3) had already been reported in eQTLGen as regulating the mRNA levels of seven coding genes (Table 4). In this database, the minor alleles of rs13000315 and rs71414848 on chromosome 2 correlate positively with the levels of *CLEC4F*, *NAGK*, *MCEE*, and *CD207*. Moreover, the three SNPs on chromosome 3 (rs74464684, rs76553845, and rs151212827) associate with the expression of *NT5DC2*, *TKT*, and *UQCC5*: increasing numbers of the minor allele of these three SNPs correlate positively with the expression of *NT5DC2* and negatively with the expression of *TKT* and *UQCC5*.

**Table 4.** Evidence from eQTLGen that the top five ranking SNPs are *cis*-eQTLs in blood.

SNP	Chr.	Minor allele	Major allele	Regulated gene	P-value	Z-score	FDR
rs13000315	2	A	T	CLEC4F	5.08 × 10 <sup>-53</sup>	15.3	0
		A	T	NAGK	4.30 × 10 <sup>-10</sup>	6.24	0
		A	T	MCEE	3.74 × 10 <sup>-6</sup>	4.63	0.0100
		A	T	CD207	1.50 × 10 <sup>-5</sup>	4.33	0.0387
rs71414848	2	C	T	CLEC4F	3.16 × 10 <sup>-53</sup>	15.4	0
		C	T	NAGK	3.74 × 10 <sup>-10</sup>	6.26	0
		C	T	MCEE	4.81 × 10 <sup>-6</sup>	4.57	0.0130
		C	T	CD207	1.88 × 10 <sup>-5</sup>	4.28	0.0470
rs74464684	3	T	C	NT5DC2	6.54 × 10 <sup>-49</sup>	14.7	0
		T	C	TKT	8.97 × 10 <sup>-16</sup>	-8.04	0
		T	C	UQCC5	4.80 × 10 <sup>-7</sup>	-5.03	0.00147
rs76553845	3	G	A	NT5DC2	4.40 × 10 <sup>-45</sup>	14.1	0
		G	A	TKT	2.18 × 10 <sup>-12</sup>	-7.02	0
		G	A	UQCC5	4.12 × 10 <sup>-7</sup>	-5.06	0.00122

rs151212827	3	A	G	NT5DC2	1.57 × 10 <sup>-47</sup>	14.5	0
		A	G	TKT	1.08 × 10 <sup>-11</sup>	-6.80	0
		A	G	UQCC5	2.76 × 10 <sup>-8</sup>	-5.56	0.000121

Chr, chromosome; FDR, false discovery rate.

The combined results from GTEx and eQTLGen databases identified a total of 11 genes, including eight coding genes that were regulated by the six top-ranking SNPs in our genome-wide survival analysis. Analyzing the expression levels of these coding genes in the tumor tissue of lung adenocarcinoma patients (in publicly available databases), we observed that four were associated with overall survival. High expression levels of *NT5DC2*, *TKT*, *UQCC5*, and *NAGK* genes were negative prognostic factors (multivariable Cox and log-rank *P* < 0.005; Supplementary Figure 3). Thus, there was an agreement between the direction of effect of gene expression on prognosis and that of the minor allele on gene expression for *NAGK* and *NT5DC2*, whereas this was not the case for *TKT* and *UQCC5*.

4. Discussion

This study explored the association between germline polymorphisms and survival, 60 months after surgery for lung adenocarcinoma, in a series of 1,464 patients. A multivariable Cox proportional hazard model identified six SNPs whose genotype associated with overall survival at the genome-wide significance threshold and whose minor allele was a negative prognostic factor (HR > 1). Three of these SNPs (rs13000315, rs151212827, rs190923216) were independent prognostic factors, together with the patients’ age, sex, pathological stage, and decade of surgery. All six top-ranking SNPs had already been reported as regulators of gene expression.

Before doing the genome-wide survival analysis, we examined the patients’ clinical data and confirmed that age, sex and pathological stage were independent prognostic factors for lung adenocarcinoma [7]. In addition, as expected for a series of patients who were enrolled over a large time interval, the probability of survival associated with the decade in which they had surgery. In our analysis, smoking habit did not affect survival, in contrast to previous reports (e.g., [41,42]). Yet because we did not have complete data about the patients’ smoking history (e.g., pack years and smoking cessation data for former smokers), we may have underestimated the effect of smoking on survival.

The study identified six SNPs associated with survival at the genome-wide significance level. Because they map to non-coding regions of the genome, we hypothesized that they have regulatory roles in gene expression. Indeed, five SNPs were previously reported as eQTLs targeting coding genes: the two variants on chromosome 2 associate with the expression levels of *CLEC4F*, *NAGK*, *MCEE*, and *CD207* genes, and the three variants on chromosome 3 are eQTLs of *NT5DC2*, *TKT*, and *UQCC5* genes. Expression levels of these genes correlate directly with the number of minor alleles of the regulatory SNPs, except for *TKT* and *UQCC5* which correlate inversely.

It has already been demonstrated that tumor expression of *NT5DC2* is a prognostic marker of lung adenocarcinoma, with high levels of expression associated with poor survival [43]. It has also been observed that this gene has a role in non-small-cell lung cancer progression: indeed, its overexpression promoted the proliferative, migratory, and invasive capacities of NSCLC cells, whereas its down-regulation induced cell cycle arrest and apoptosis [44]. We found three SNPs (rs74464684, rs76553845, and rs151212827) whose minor allele associated with worse prognosis. Already published data indicated that these minor alleles associated with high levels of *NT5DC2*, and that, in tumor tissue, high levels of *NT5DC2* associated with poor prognosis. These associations suggest that individuals with a minor allele of these germline polymorphisms are more susceptible to a negative lung adenocarcinoma outcome than patients homozygous for the major allele, possibly due to a higher expression of this gene. The minor alleles of these same SNPs, associated with poor survival, have also been reported to negatively regulate the expression of *TKT* and *UQCC5* genes. However, high expression levels of these genes in lung adenocarcinoma have been associated with poor overall survival: *TKT* has been suggested to be a negative prognostic marker in lung

adenocarcinoma [45], and *UQCC5* (alias *SMIM4*) has been proposed, together with another six polymorphisms, as a survival prediction tool [46]. Therefore, it is more plausible that the SNPs on chromosome 3 affect lung adenocarcinoma survival through the regulation of expression of *NT5DC2* gene.

In the literature, we did not find evidence of a prognostic role of the four genes (*CLEC4F*, *NAGK*, *MCEE*, and *CD207*) regulated by the top-significant SNPs on chromosome 2 (rs13000315 and rs71414848). Nonetheless, we found that, in lung adenocarcinoma tissue, *NAGK* expression levels associate with overall survival. Indeed, patients expressing high levels (above the median) of *NAGK* had a higher death probability than those expressing low levels. As our patients with at least one copy of the minor allele of rs13000315 and rs71414848 (associated with higher levels of *NAGK*) had a negative lung adenocarcinoma outcome, we speculate that this was due to a genetic predisposition to higher *NAGK* expression than patients homozygous for the major alleles of these SNPs.

A limitation of our study is the lack of information about other possible prognostic factors, such as the somatic mutational status and therapies administered to the patients in addition to surgical resection. Unfortunately, these data were missing for a rather large number of patients, thus we preferred not to reduce the sample size of our GWAS. Nevertheless, we believe that our results are promising.

Functional studies are needed to test the hypothesized biological mechanism of action of the identified SNPs in modulating survival. It would be interesting to test whether the already reported eQTLs act in lung adenocarcinoma tissue. It would also be useful to understand, for instance, the role of rs190923216 (on chromosome 5), which has been reported to be an eQTL for an anti-sense gene (*CKMT2-AS1*) in normal lung. Studies are needed to understand whether this gene plays a role in lung adenocarcinoma prognosis. Finally, validation in an independent, but homogeneous series is needed: to do this it will be important to consider that our findings were obtained from the analysis of prevalently European patients and that the allele frequencies of the identified variants were quite low.

## 5. Conclusions

Our study identified germline variants affecting lung adenocarcinoma patient survival, possibly due to a regulatory role on gene expression, in particular of *NT5DC2* and *NAGK* genes. Indeed, their expression in lung adenocarcinoma tissue was previously reported to associate with poor prognosis. Overall, our results underscore the significant role of genetic factors in predisposing lung adenocarcinoma patients to different outcomes.

**Supplementary Materials:** The following supporting information can be downloaded at: [www.mdpi.com/xxx/s1](http://www.mdpi.com/xxx/s1). **Supplementary Table 1.** Genetic variants associated with overall survival (60 months after surgery) at  $P < 1.0 \times 10^{-5}$ , sorted by  $P$ -value. **Supplementary Table 2.** eQTLs, in the GTEx database, among the 224 variants associated with survival at  $P < 1.0 \times 10^{-5}$ . **Supplementary Table 3.** eQTLs, in the eQTLGen database, among the 224 SNPs associated with patient survival at  $P < 1.0 \times 10^{-5}$ . **Supplementary Figure 1.** Per-sample and per-marker quality control (QC) of genotyping data. **Supplementary Figure 2.** Plots of the first four principal components. **Supplementary Figure 3.** Kaplan–Meier survival curves (truncated at 60 months) for lung adenocarcinoma patients according to tumor expression levels of *NT5DC2*, *TKT*, *UQCC5*, and *NAGK* genes (from top to bottom; [www.kmplot.com/lung](http://www.kmplot.com/lung)).

**Author Contributions:** Conceptualization, T.A.D. and F.C.; formal analysis, F.M., M.Es., and F.C.; resources, T.M., M.A.S., M.K., M.Ei., H.H., S.K., H.W., M.I., G.M., and D.T.; data curation and sample preparation, S.N.; writing—original draft preparation, T.A.D., F.C. and F.M.; writing—review and editing, T.A.D., F.C., F.M., and T.M.; funding acquisition, T.A.D. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Fondazione AIRC per la ricerca sul cancro (AIRC 2017 IG 20226 to T.A.D.).

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki, and approved by the ethics committees of Fondazione IRCCS Istituto Nazionale dei Tumori (INT 224-17, on 19 December 2017), Ospedale San Giuseppe, IRCCS Multimedica (346.2018, on 01 October 2018),



Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico (202\_2019bis; on 12 March 2019) and University of Heidelberg (S-270/2001).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author due to privacy restrictions.

**Acknowledgments:** The authors acknowledge the contributions of Valerie Matarese, PhD, who provided scientific editing.

**Conflicts of Interest:** The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## References

1. Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. *Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries*. *CA Cancer J Clin* **2021**, *71*, 209–249. <https://doi.org/10.3322/caac.21660>.
2. Siegel, R.L.; Miller, K.D.; Wagle, N.S.; Jemal, A. *Cancer Statistics, 2023*. *CA Cancer J Clin* **2023**, *73*, 17–48. <https://doi.org/10.3322/caac.21763>.
3. Potter, A.L.; Rosenstein, A.L.; Kiang, M. V; Shah, S.A.; Gaissert, H.A.; Chang, D.C.; Fintelmann, F.J.; Yang, C.-F.J. Association of Computed Tomography Screening with Lung Cancer Stage Shift and Survival in the United States: Quasi-Experimental Study. *BMJ* **2022**, e069008. <https://doi.org/10.1136/bmj-2021-069008>.
4. Whitson, B.A.; Groth, S.S.; Duval, S.J.; Swanson, S.J.; Maddaus, M.A. Surgery for Early-Stage Non-Small Cell Lung Cancer: A Systematic Review of the Video-Assisted Thoracoscopic Surgery Versus Thoracotomy Approaches to Lobectomy. *Ann Thorac Surg* **2008**, *86*, 2008–2018. <https://doi.org/10.1016/j.athoracsur.2008.07.009>.
5. Rami-Porta, R.; Call, S.; Doms, C.; Obiols, C.; Sánchez, M.; Travis, W.D.; Vollmer, I. Lung Cancer Staging: A Concise Update. *European Respiratory Journal* **2018**, *51*, 1800190. <https://doi.org/10.1183/13993003.00190-2018>.
6. Howlader, N.; Forjaz, G.; Mooradian, M.J.; Meza, R.; Kong, C.Y.; Cronin, K.A.; Mariotto, A.B.; Lowy, D.R.; Feuer, E.J. The Effect of Advances in Lung-Cancer Treatment on Population Mortality. *New England Journal of Medicine* **2020**, *383*, 640–649. <https://doi.org/10.1056/NEJMoa1916623>.
7. Garinet, S.; Wang, P.; Mansuet-Lupo, A.; Fournel, L.; Wislez, M.; Blons, H. Updated Prognostic Factors in Localized NSCLC. *Cancers (Basel)* **2022**, *14*. <https://doi.org/10.3390/cancers14061400>.
8. Chen, Z.; Fillmore, C.M.; Hammerman, P.S.; Kim, C.F.; Wong, K.-K. Non-Small-Cell Lung Cancers: A Heterogeneous Set of Diseases. *Nat Rev Cancer* **2014**, *14*, 535–546. <https://doi.org/10.1038/nrc3775>.
9. Chen, K.; Liu, H.; Liu, Z.; Luo, S.; Patz, E.F.; Moorman, P.G.; Su, L.; Shen, S.; Christiani, D.C.; Wei, Q. Genetic Variants in RUNX3, AMD1 and MSRA in the Methionine Metabolic Pathway and Survival in Non-small Cell Lung Cancer Patients. *Int J Cancer* **2019**, *145*, 621–631. <https://doi.org/10.1002/ijc.32128>.
10. Du, H.; Liu, L.; Liu, H.; Luo, S.; Patz, E.F.; Glass, C.; Su, L.; Du, M.; Christiani, D.C.; Wei, Q. Genetic Variants of DOCK2, EPHB1 and VAV2 in the Natural Killer Cell-Related Pathway Are Associated with Non-Small Cell Lung Cancer Survival. *Am J Cancer Res* **2021**, *11*, 2264–2277.
11. Qian, D.; Liu, H.; Zhao, L.; Wang, X.; Luo, S.; Moorman, P.G.; Patz Jr, E.F.; Su, L.; Shen, S.; Christiani, D.C.; et al. Novel Genetic Variants in Genes of the Fc Gamma Receptor-Mediated Phagocytosis Pathway Predict Non-Small Cell Lung Cancer Survival. *Transl Lung Cancer Res* **2020**, *9*, 575–586. <https://doi.org/10.21037/tlcr-19-318>.
12. Zhang, H.; Li, Y.; Guo, S.; Wang, Y.; Wang, H.; Lu, D.; Wang, J.; Jin, L.; Jiang, G.; Wu, J.; et al. Effect of ERCC2 Rs13181 and Rs1799793 Polymorphisms and Environmental Factors on the Prognosis of Patients with Lung Cancer. *Am J Transl Res* **2020**, *12*, 6941–6953.
13. Pintarelli, G.; Cotroneo, C.E.; Noci, S.; Dugo, M.; Galvan, A.; Delli Carpini, S.; Citterio, L.; Manunta, P.; Incarbone, M.; Tosi, D.; et al. Genetic Susceptibility Variants for Lung Cancer: Replication Study and Assessment as Expression Quantitative Trait Loci. *Sci Rep* **2017**, *7*. <https://doi.org/10.1038/srep42185>.
14. Du, H.; Mu, R.; Liu, L.; Liu, H.; Luo, S.; Patz, E.F.; Glass, C.; Su, L.; Du, M.; Christiani, D.C.; et al. Single Nucleotide Polymorphisms in FOXP1 and RORA of the Lymphocyte Activation-Related Pathway Affect Survival of Lung Cancer Patients. *Transl Lung Cancer Res* **2022**, *11*, 890–901. <https://doi.org/10.21037/TLCR-22-104/COIF>.
15. Chen, A.S.; Liu, H.; Wu, Y.; Luo, S.; Patz, E.F.; Glass, C.; Su, L.; Du, M.; Christiani, D.C.; Wei, Q. Genetic Variants in DDO and PEX5L in Peroxisome-Related Pathways Predict Non-Small Cell Lung Cancer Survival. *Mol Carcinog* **2022**, *61*, 619–628. <https://doi.org/10.1002/MC.23400>.

16. Yang, S.; Tang, D.; Zhao, Y.C.; Liu, H.; Luo, S.; Stinchcombe, T.E.; Glass, C.; Su, L.; Shen, S.; Christiani, D.C.; et al. Potentially Functional Variants of ERAP1, PSMF1 and NCF2 in the MHC-I-Related Pathway Predict Non-Small Cell Lung Cancer Survival. *Cancer Immunol Immunother* **2021**, *70*, 2819–2833. <https://doi.org/10.1007/S00262-021-02877-9>.
17. Galvan, A.; Colombo, F.; Frullanti, E.; Dassano, A.; Noci, S.; Wang, Y.; Eisen, T.; Matakidou, A.; Tomasello, L.; Vezzalini, M.; et al. Germline Polymorphisms and Survival of Lung Adenocarcinoma Patients: A Genome-Wide Study in Two European Patient Series. *Int J Cancer* **2015**, *136*. <https://doi.org/10.1002/ijc.29195>.
18. Zhu, M.; Geng, L.; Shen, W.; Wang, Y.; Liu, J.; Cheng, Y.; Wang, C.; Dai, J.; Jin, G.; Hu, Z.; et al. Exome-Wide Association Study Identifies Low-Frequency Coding Variants in 2p23.2 and 7p11.2 Associated with Survival of Non-Small Cell Lung Cancer Patients. *Journal of Thoracic Oncology* **2017**, *12*, 644–656. <https://doi.org/10.1016/j.jtho.2016.12.025>.
19. Dragani, T.A.; Muley, T.; Schneider, M.A.; Kobinger, S.; Eichhorn, M.; Winter, H.; Hoffmann, H.; Kriegsmann, M.; Noci, S.; Incarbone, M.; et al. Lung Adenocarcinoma Diagnosed at a Younger Age Is Associated with Advanced Stage, Female Sex, and Ever-Smoker Status, in Patients Treated with Lung Resection. *Cancers (Basel)* **2023**, *15*, 2395. <https://doi.org/10.3390/cancers15082395>.
20. TNM Classification of Malignant Tumours.; Sobin, L., Christian, W., Eds.; UICC International Union Against Cancer, 2002;.
21. TNM Classification of Malignant Tumours; Sobin LH, Gospodarowicz Mary K, Wittekind Christian, Eds.; 7th ed.; Wiley-Blackwell, 2009; ISBN 978-1-4443-3241-4.
22. TNM Classification of Malignant Tumours, 8th Edition; Brierley James D., Gospodarowicz Mary K., Wittekind Christian, Eds.; Wiley-Blackwell, 2016; ISBN 978-1-119-26357-9.
23. Maspero, D.; Dassano, A.; Pintarelli, G.; Noci, S.; de Cecco, L.; Incarbone, M.; Tosi, D.; Santambrogio, L.; Dragani, T.A.; Colombo, F. Read-through Transcripts in Lung: Germline Genetic Regulation and Correlation with the Expression of Other Genes. *Carcinogenesis* **2020**, *41*. <https://doi.org/10.1093/CARCIN/BGAA020>.
24. Cotroneo, C.E.; Mangano, N.; Dragani, T.A.; Colombo, F. Lung Expression of Genes Putatively Involved in SARS-CoV-2 Infection Is Modulated in Cis by Germline Variants. *European Journal of Human Genetics* **2021**, *29*, s41431–s021.
25. Purcell, S.; Neale, B.; Todd-Brown, K.; Thomas, L.; Ferreira, M.A.R.; Bender, D.; Maller, J.; Sklar, P.; De Bakker, P.I.W.; Daly, M.J.; et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am J Hum Genet* **2007**, *81*, 559–575. <https://doi.org/10.1086/519795>.
26. Anderson, C.A.; Pettersson, F.H.; Clarke, G.M.; Cardon, L.R.; Morris, A.P.; Zondervan, K.T. Data Quality Control in Genetic Case-Control Association Studies. *Nat Protoc* **2010**, *5*, 1564–1573. <https://doi.org/10.1038/nprot.2010.116>.
27. Das, S.; Forer, L.; Schönherr, S.; Sidore, C.; Locke, A.E.; Kwong, A.; Vrieze, S.I.; Chew, E.Y.; Levy, S.; McGue, M.; et al. Next-Generation Genotype Imputation Service and Methods. *Nature Genetics* **2016**, *48*, 1284–1287. <https://doi.org/10.1038/ng.3656>.
28. Loh, P.R.; Danecek, P.; Palamara, P.F.; Fuchsberger, C.; Reshef, Y.A.; Finucane, H.K.; Schoenherr, S.; Forer, L.; McCarthy, S.; Abecasis, G.R.; et al. Reference-Based Phasing Using the Haplotype Reference Consortium Panel. *Nature Genetics* **2016**, *48*, 1443–1448. <https://doi.org/10.1038/ng.3679>.
29. Fuchsberger, C.; Abecasis, G.R.; Hinds, D.A. Minimac2: Faster Genotype Imputation. *Bioinformatics* **2015**, *31*, 782–784. <https://doi.org/10.1093/BIOINFORMATICS/BTU704>.
30. Taliun, D.; Harris, D.N.; Kessler, M.D.; Carlson, J.; Szpiech, Z.A.; Torres, R.; Taliun, S.A.G.; Corvelo, A.; Gogarten, S.M.; Kang, H.M.; et al. Sequencing of 53,831 Diverse Genomes from the NHLBI TOPMed Program. *Nature* **2021**, *590*, 290–299. <https://doi.org/10.1038/s41586-021-03205-y>.
31. Verlouw, J.A.M.; Clemens, E.; de Vries, J.H.; Zolk, O.; Verkerk, A.J.M.H.; am Zehnhoff-Dinnesen, A.; Medina-Gomez, C.; Lanvers-Kaminsky, C.; Rivadeneira, F.; Langer, T.; et al. A Comparison of Genotyping Arrays. *European Journal of Human Genetics* **2021**, *29*, 1611–1624. <https://doi.org/10.1038/s41431-021-00917-7>.
32. Chang, C.C.; Chow, C.C.; Tellier, L.C.A.M.; Vattikuti, S.; Purcell, S.M.; Lee, J.J. Second-Generation PLINK: Rising to the Challenge of Larger and Richer Datasets. *Gigascience* **2015**, *4*, 7. <https://doi.org/10.1186/S13742-015-0047-8/2707533>.
33. Delaneau, O.; Marchini, J.; McVean, G.A.; Donnelly, P.; Lunter, G.; Marchini, J.L.; Myers, S.; Gupta-Hinch, A.; Iqbal, Z.; Mathieson, I.; et al. Integrating Sequence and Array Data to Create an Improved 1000 Genomes Project Haplotype Reference Panel. *Nat Commun* **2014**, *5*, 3934. <https://doi.org/10.1038/ncomms4934>.
34. Clark, T.G.; Bradburn, M.J.; Love, S.B.; Altman, D.G. Survival Analysis Part I: Basic Concepts and First Analyses. *British Journal of Cancer* **2003**, *89*, 232–238. <https://doi.org/10.1038/sj.bjc.6601118>.
35. Terry M. Therneau and Patricia M. Grambsch Modeling Survival Data: Extending the Cox Model. , Springer-Verlag, New York, 2000. ; Springer-Verlag: New York, 2000;.

36. Aulchenko, Y.S.; Ripke, S.; Isaacs, A.; van Duijn, C.M. GenABEL: An R Library for Genome-Wide Association Analysis. *Bioinformatics* **2007**, *23*, 1294–1296. <https://doi.org/10.1093/bioinformatics/btm108>.
37. Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)* **1995**, *57*, 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>.
38. Vösa, U.; Claringbould, A.; Westra, H.-J.; Bonder, M.J.; Deelen, P.; Zeng, B.; Kirsten, H.; Saha, A.; Kreuzhuber, R.; Yazar, S.; et al. Large-Scale Cis- and Trans-EQTL Analyses Identify Thousands of Genetic Loci and Polygenic Scores That Regulate Blood Gene Expression. *Nat Genet* **2021**, *53*, 1300–1310. <https://doi.org/10.1038/s41588-021-00913-z>.
39. Györfy, B.; Surowiak, P.; Budczies, J.; Lánczky, A. Online Survival Analysis Software to Assess the Prognostic Value of Biomarkers Using Transcriptomic Data in Non-Small-Cell Lung Cancer. *PLoS ONE* **2013**, *8*. <https://doi.org/10.1371/JOURNAL.PONE.0082241>.
40. Kratzer, T.B.; Bandi, P.; Freedman, N.D.; Smith, R.A.; Travis, W.D.; Jemal, A.; Siegel, R.L. Lung Cancer Statistics, 2023. *Cancer* **2024**, *130*, 1330–1348. <https://doi.org/10.1002/cncr.35128>.
41. Kawaguchi, T.; Takada, M.; Kubo, A.; Matsumura, A.; Fukai, S.; Tamura, A.; Saito, R.; Maruyama, Y.; Kawahara, M.; Ignatius Ou, S.-H. Performance Status and Smoking Status Are Independent Favorable Prognostic Factors for Survival in Non-Small Cell Lung Cancer: A Comprehensive Analysis of 26,957 Patients with NSCLC. *J Thorac Oncol* **2010**, *5*, 620–630. <https://doi.org/10.1097/JTO.0b013e3181d2dcd9>.
42. Sheikh, M.; Mukeriyah, A.; Shangina, O.; Brennan, P.; Zaridze, D. Postdiagnosis Smoking Cessation and Reduced Risk for Lung Cancer Progression and Mortality: A Prospective Cohort Study. *Ann Intern Med* **2021**, *174*, 1232–1239. <https://doi.org/10.7326/M21-0252>.
43. Schulze, A.B.; Kuntze, A.; Schmidt, L.H.; Mohr, M.; Marra, A.; Hillejan, L.; Schulz, C.; Görlich, D.; Hartmann, W.; Bleckmann, A.; et al. High Expression of NT5DC2 Is a Negative Prognostic Marker in Pulmonary Adenocarcinoma. *Cancers (Basel)* **2022**, *14*. <https://doi.org/10.3390/CANCERS14061395>.
44. Jin, X.; Liu, X.; Zhang, Z.; Xu, L. NT5DC2 Suppression Restrains Progression towards Metastasis of Non-Small-Cell Lung Cancer through Regulation P53 Signaling. *Biochem Biophys Res Commun* **2020**, *533*, 354–361. <https://doi.org/10.1016/J.BBRC.2020.06.139>.
45. Niu, C.; Qiu, W.; Li, X.; Li, H.; Zhou, J.; Zhu, H. Transketolase Serves as a Biomarker for Poor Prognosis in Human Lung Adenocarcinoma. *J Cancer* **2022**, *13*, 2584–2593. <https://doi.org/10.7150/JCA.69583>.
46. Zhang, S.; Zeng, X.; Lin, S.; Liang, M.; Huang, H. Identification of Seven-Gene Marker to Predict the Survival of Patients with Lung Adenocarcinoma Using Integrated Multi-Omics Data Analysis. *J Clin Lab Anal* **2022**, *36*, e24190. <https://doi.org/10.1002/jcla.24190>.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.