Genes, Chromosomes and Cancer

WILEY

RESEARCH ARTICLE OPEN ACCESS

Germline Whole-Exome Sequencing in Non-Smoker Lung Cancer Patients Reveals Pathogenic Variants in Lung Cancer Driver Genes

Giovanni Carapezza[1] | Simone Paolo Minardi[1] | Sara Noci[2] | Giulia Pintarelli[2] | Susanna Zanutto[2] | Matteo Incarbone[3] | Davide Tosi[4] | Tommaso Antonio Dragani[2] | Francesca Colombo[5] | Marco Alessandro Pierotti[1] | Manuela Gariboldi[2]

[1]Cogentech S.R.L.Benefit C. With Only Stakeholder Fondazione IFOM ETS, Milano, Italy | [2]Fondazione IRCCS Istituto Nazionale Dei Tumori, Milano, Italy | [3]Ospedale San Giuseppe, IRCCS Multimedica, Milano, Italy | [4]Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milano, Italy | [5]National Research Council, Institute for Biomedical Technologies, Segrate, Italy

Correspondence: Manuela Gariboldi (manuela.gariboldi@istitutotumori.mi.it)

Received: 23 December 2024 | **Revised:** 27 February 2025 | **Accepted:** 28 February 2025

Funding: This research was supported by the Associazione Italiana Ricerca sul Cancro (AIRC 2017 IG 20226 to T.A.D.) and by Ministry of Health "Ricerca Corrente".

Keywords: DNA repair-related genes | genetic risk factors | germline molecular testing | non-smoker lung cancer | pathogenic germline variants

ABSTRACT

Approximately 10%–15% of all lung cancers arise in non-smokers. Although there are no established aetiological factors, non-smokers with a family history of cancer have an increased risk of lung cancer, implying host genetic factors in lung cancer susceptibility. We sought to identify, in a cohort of 75 patients recruited before lung lobectomy, germline alterations with a strong association with lung cancer. Whole-exome sequencing was performed on genomic DNA from peripheral blood. Six resources were used to select pathogenic germline variants with strong clinical significance. In total, 33 pathogenic or likely pathogenic variants in 31 genes were identified. Of these, 13 were located in cancer-predisposing genes (nine were lung cancer drivers), most of which were involved in DNA repair mechanisms and diseases of metabolism. Among DNA repair-related genes, *BRCA1* and *BRCA2*, and *ATM* have also been identified in other studies on non-smokers. Our results strongly support the hypothesis that a number of non-smoker lung cancer patients carry germline variants in cancer-predisposing genes, suggesting that lung cancer patients, particularly non-smokers, should be considered for germline molecular testing.

1 | Introduction

Lung cancer is the most common cancer and the leading cause of cancer-related mortality worldwide [1]. It is generally regarded as a smoker's cancer; however, about 10% to 15% of lung cancers arise in non-smokers. The rate of lung cancer in non-smokers is increasing [2] and tends to be more common in females [3]. Most of the non-smokers are asymptomatic or have non-specific features, and cancer is diagnosed during cardiological follow-up or surveillance imaging for other previous

Abbreviations: *ACADVL*, acyl-CoA dehydrogenase very long chain; *ADA*, adenosine deaminase; *ATM*, ATM serine/threonine kinase; *BRAF*, B-Raf proto-oncogene, serine/threonine kinase; *BRCA1*, BReast CAncer gene 1 DNA repair associated; *BRCA2*, BReast CAncer gene 2 DNA repair associated; *BTD*, biotinidase; *COL1A1*, collagen type I alpha 1 chain; CPG, cancer predisposing gene; *DNMT3A*, DNA methyltransferase 3 alpha; *FH*, fumarate hydratase (FH); *GAA*, alpha glucosidase; *GALNS*, galactosamine (*N*-acetyl)-6-sulfatase; *GJB2*, gap junction protein beta 2; *GYG1*, glycogenin 1; *HERC2*, HECT and RLD domain-containing E3 ubiquitin protein ligase 2; *IVD*, isovaleryl-CoA dehydrogenase; *KEAP1*, kelch like ECH associated protein 1; LCDG, lung cancer driver gene; *MLH1*, mutL homolog 1; *MMAB*, metabolism of cobalamin associated with B; *MUTYH*, mutY DNA glycosylase; NSCLC, Non-small-cell lung cancer; *NUP93*, nucleoporin 93; *PEX1*, peroxisomal biogenesis factor 1; PGV, pathogenic germline variants; *POLG*, DNA polymerase gamma, catalytic subunit; *PYGM*, glycogen phosphorylase, muscle associated; SCLC, small cell lung cancer; *SDHA*, succinate dehydrogenase complex flavoprotein subunit A; *SLC3A1*, solute carrier family 3 member 1; *TNFRSF13B*, TNF receptor superfamily member 13B; *TP53*, tumor protein p53; *TSHR*, thyroid stimulating hormone receptor; *TYR*, tyrosinase; WES, Whole-exome sequencing.

Giovanni Carapezza and Simone Paolo Minardi contributed equally to this article.

cancers. Although small cell lung cancer (SCLC) is the histological type tightly linked with tobacco consumption [4], the most frequent type of cancer in non-smokers is non-small cell lung cancer (NSCLC), predominantly adenocarcinoma [5]. There are no established aetiological factors associated with the development of lung cancer in non-smokers; radon exposure, second-hand smoke, and other environmental risks, most of which are linked to pollution, may explain some cases, but the causes of most of these cancers are still unknown (as reviewed in [5]). Non-smokers with a family history of cancer also have an increased risk of lung cancer, implicating host genetic factors in lung cancer susceptibility [5]. Genome-wide association studies have identified common genetic variations that modify lung cancer risk, but they explain less than 10% of the risk contribution to lung cancer [6].

Next-generation sequencing applied to cancer research has made available a wealth of genome-wide data with which it has been possible to highlight the presence of clinically relevant germline variants that confer significant risk to different types of cancer. Indeed, a study of more than 10,000 individuals in 33 types of cancer in the TCGA cohort found that 8% of cases carried pathogenic or likely pathogenic germline variants [7]. Concerning lung cancer, few studies have examined the presence of pathogenic germline variants (PGVs) and have mainly focused on variants in *TP53* and *EGFR* (T790M in particular) genes, which have been associated with hereditary predisposition to this tumor type (as reviewed in [8]). Recent studies, however, suggest that germline pathogenic variants are present in a high percentage of lung cancer patients, in particular, if non-smokers [9, 10]. Identifying genetic alterations with a strong association with lung cancer can help to define population risk and also find patient subpopulations that can benefit from personalized follow-up programs (e.g., CT scan) and therapy.

We studied the landscape of PGVs in lung cancer drivers and in cancer predisposing genes, which most likely could act as the genetic counterpart of the cigarette smoke-induced genotoxic damage, in 75 non-smoker lung cancer patients.

2 | Materials and Methods

2.1 | Patient Series and DNA Samples

The study included 75 non-smoker lung adenocarcinoma (NSCLC) patients selected from a series of 823 patients with NSCLC recruited prior to lung lobectomy at hospitals in the Milan, Italy, area [11]. The main characteristics of the patients are summarized in Table 1. For this study, we used frozen samples of genomic DNA previously extracted from the patients' peripheral blood [12].

Written informed consent was obtained from patients to use their biological material and data for research purposes, according to the European General Data Protection Regulation. The protocol of this study was approved by the Committee for Ethics of the Fondazione IRCCS Istituto Nazionale dei Tumori, Milan (INT 224–17), Fondazione IRCCS Ca′ Granda Ospedale Maggiore Policlinico, Milan (2022019bis) and Ospedale San Giuseppe, IRCCS Multimedica, Milan (346.2018). The research was conducted in accordance with the tenets of the Declaration of Helsinki.

TABLE 1 | Numbers and clinical characteristics of the patients in the study.

Patient characteristic	Number of patients (%)
Sex	
Male	24 (32)
Female	51 (68)
Age, median (range)	59 (26–79)
Pathological stage	
I	34 (45.33)
II	13 (17.33)
III	16 (21.33)
IV	10 (13.33)
Not available	2 (2.66)
Status	
Alive[a]	40 (53.3)
Dead[b]	35 (46.7)

[a] > 3 years from surgery [median 563 days, range (87–2572)].
[b] < 7 years from surgery [median 2934 days, range (1305–7335)].

2.2 | Whole-Exome Sequencing (WES)

Two different WES analyses were conducted. A first group of 32 cases (cohort 1, C1) was analyzed at the CRIBI Sequencing Core of the University of Padua (Padua, Italy). Whole-exome libraries were prepared from 100 ng of genomic DNA using Ion Ampliseq Exome RDY kits (ThermoFisher Scientific). Libraries were sequenced on Ion Torrent P1 Chips on an Ion Proton Sequencer. To identify variants, the AmpliSeq Exome Panel module with parameters Proton, Germ Line, and High Stringency, under the Torrent Variant Caller software, was used.

A second group of 60 cases (cohort 2, C2), including 17 samples from the first analysis, was analyzed at the Genomic Unit of Cogentech Ltd. Benefit Corporation with IFOM as a Sole Shareholder, in Milan, Italy. Indexed fragment libraries were prepared from 150 ng of genomic DNA using Illumina DNA Prep (S) Tagmentation kit and processed with xGen Exome Research Panel v2 (IDT) for enrichment of exonic probes to generate final WES libraries. Sequencing of the libraries was performed using an Illumina NextSeq550DX instrument in 2 × 150 nt readmode, to generate 50 M paired-end reads per sample. To identify variants, we used a custom pipeline which includes: (i) Quality checks on fastq [13] and bam [14] files; (ii) Pre-processing steps according to the "GATK best practices" (https://gatk.broadinstitute.org/hc/en-us/articles/360035535912-Data-pre-processing-for-variant-discovery); (iii) Two germinal variant callers: Haplotype Caller (https://gatk.broadinstitute.org/hc/en-us/articles/360037225632-HaplotypeC

aller) and Strelka2 [15]; (iv) Union of the good quality variants (labeled as PASS by each caller); (v) Filtering of variants called with a coverage (DP) less than 10 or with an allelic fraction (AF) less than 0.1.

WES generated a median of 33 million paired-end reads per patient (range, 24 to 51 million reads) resulting in a median coverage of 97 (range: 72–149) for the 32 samples included in C1, and a median of 120 million paired-end reads per patient (range, 80 to 156 million reads) resulting in a median coverage of 276 for the 60 samples of C2 (range: 184–367). The median percentage of target covered at 20X was 89.8% (range: 85.4%–94.7%) for C1 and 99.3% for C2 (range: 99%–99.6%) (Figure S1). No samples were discarded for low coverage (Table S1).

2.3 | Pathogenic Variant Selection (PGV)

Six resources were used (two tools, two scores, two reference databases) to select PGVs with strong clinical significance (Tier1) [16]: (i) Cancer Genome Interpreter [17], a classifier based on heuristic rules that depend on variant consequence, CADD score, gene category (driver for the tumor or for other tumors), gene mechanism of action, variant position in a cluster, position within a domain or within the protein, and is integrated with BoostDM [18], a gradient boosting classifier trained on more than 28,000 samples but available only for certain genes. The tool classifies the variants into Driver/Pathogenic or Passenger/Benign; (ii) Charger [19], which annotates and prioritizes variants by adopting the AMP-ACMG guidelines and defines 12 levels of pathological evidence and four levels of benign evidence using a variety of data sets, including ExAC and ClinVar, and computational tools (SIFT, PolyPhen). Variants in predisposition genes receive a level of evidence based on variant type and mode of inheritance. Variants are referred to as: P (Pathogenic), LP (Likely Pathogenic), VUS (Variant of Uncertain Significance), LB (Likely Benign), and B (Benign); (iii) Ai-Driver [20], a score combining 23 pathogenicity scores, including VEST 3.0, Polyphen2, SIFT, FATHMM, and CADD into a single classifier that predicts the possibility of the variant being driver/pathogenic; (iv) Vest [21], a score using a random forest classifier that returns a p-value; variants with a p-value < 0.05 are considered pathogenic; (v, vi) Two public archives of reports of human variant-phenotype relationships, Clinvar [22], and another developed on TCGA data by the TCGA PanCanAtlas Germline Working Group [7].

Each variant was assigned a "Pathogenic Number" based on the number of resources classifying it as pathogenic (Range 0–6). Variants were defined as Tier1 if one of the following four rules was valid: (1) Defined by Clinvar as pathogenic and reviewed by experts; (2) Defined by Cancer Genome Interpreter as known tumor drivers; (3) Defined by Charger as pathogenic; (4) Variants with pathogenic number ≥ 3. Variants that were not already classified as Tier 1 were defined as Tier2 and Tier3, with pathogenic numbers of 2 and 1, respectively. Finally, all the other variants were classified as Tier4. In addition, we distinguished variants affecting lung cancer driver genes (LCDGs) from those in genes defined as cancer predisposing (CPGs) [17, 23], and manually examined these variants.

3 | Results

3.1 | Patients' Characteristics

Genomic DNA for WES analysis was obtained from peripheral blood of 75 never-smoking patients with NSCLC with unknown exposures to lung cancer risk factors. Patients had a median age of 51 years; 68% were women. There was a wide distribution of pathological stage, with a prevalence of stage I (45% of cases). Of them, 53% were alive 3 years after surgery (Table 1).

3.2 | Pathogenic Variants

The median number of variants identified was 36 K (range: 35–38 K) for cohort 1 and 24 K for cohort 2 (range: 23–25 K). The percentage of variant overlap between samples was 60% (within each cohort), whereas the percentage of variant overlap between samples from different cohorts was 30% and 50% for C1 and C2, respectively, although some samples had a higher percentage (50% and 75% for C1 and C2, respectively). The 17 duplicated samples had the highest overlap, ranging from 50% for C1 to 75% for C2 (Figure S2). During analyses, variants identified in patients profiled in both cohorts were merged into a unique profile.

PGVs were selected as described in the Methods section, by applying very stringent criteria making very likely a true pathogenetic effect of such variants. A total of 33 pathogenic or likely pathogenic variants were identified in 31 genes (Tables 2, 3, S2 and S3). Twenty-seven patients carried at least one of these variants. Eleven of them were present in duplicate samples in the two cohorts: 10 were found in both experiments and one only in cohort 2. No variants were identified in nine of the duplicate samples (Table S4). Thirteen variants were located in genes classified as LCDG or CPG [17, 23], and mainly involved in DNA repair mechanisms (Tables 2 and S2). Six of them were found in shared samples from the two cohorts: five were detected in both experiments (in *BRCA1*, *FH*, *KEAP1*, *SDHA*, and *TP53* genes), while the *GJB2* mutation was found only in cohort 2. The 13 genes were mutated only in one patient. The other 20 identified variants belonged to 18 genes defined as non-LCDG or -CPG (Tables 3 and S3), three of which (*HERC2*, *NUP93*, and *TSHR*) are known to be associated with breast cancer, endometrioid carcinoma, and thyroid carcinoma [26]. Three genes (*SLC3A1, IVD, and PYGM*) were found mutated in two samples. *SLC3A1* carried the same pathogenic variant in two samples, while two different variants were identified in *IVD* (both pathogenic) and *PYGM* (one pathogenic, the other likely pathogenic). This set of genes is mainly involved in diseases of metabolism. One of the two *SLC3A1* variants, the two variants found in *IVD*, and those in *GALNS, SLC3A1, and TNFRSF13B were* detected in both duplicate samples from the two cohorts (Table S4).

3.3 | Gene Comparison With Literature

Comparison with data on non-smokers from previously published papers [27, 28], highlighted a prevalence of germline

TABLE 2 | List of the lung cancer driver genes (LCDG) and cancer predisposing genes (CPG) carrying the 13 identified pathogenic variants.

Gene	Driver in cancer type	Cancer predisposing gene	Cohort	Total number of mutated patients	Variant ID [24]	Manual ACMG [25]	Mutation class	Type	Protein change
ATM	YES	YES	C1	1	chr11 108 330 233 C T	Pathogenic	Nonsense	SNP	R2443*
BRAF	YES	NO	C1	1	chr7 140 753 354 T C	Pathogenic	Missense	SNP	D594G
BRCA1	OTHER	YES	C1, C2	1	chr17 43 057 062 T TG	Pathogenic	Frame Shift	INS	Q1777fs
BRCA2	YES	YES	C1	1	chr13 32 339 138 C T	Pathogenic	Nonsense	SNP	Q1595*
COL1A1	YES	NO	C2	1	chr17 50 186 465 G A	Likely Pathogenic	Missense	SNP	A1286V
DNMT3A	YES	NO	C2	1	chr2 25 234 374 G A	Pathogenic	Missense	SNP	R882C
FH	YES	YES	C1, C2	1	chr1 241 497 946 G A	Likely Pathogenic	Missense	SNP	A472V
GJB2	OTHER	YES	C2[a]	1	chr13 20 189 313 A G	Likely Pathogenic	Missense	SNP	L90P
KEAP1	YES	NO	C1, C2	1	chr19 10 489 811 C G	Likely Pathogenic	Missense	SNP	M456I
MLH1	YES	YES	C2	1	chr3 36 993 639 C G	Pathogenic	Missense	SNP	A31G
MUTYH	OTHER	YES	C1	1	chr1 45 331 699 AG A	Pathogenic	Intron	DEL	—
SDHA	OTHER	YES	C1, C2	1	chr5 251 415 G A	Likely Pathogenic	Missense	SNP	G581R
TP53	YES	YES	C1, C2	1	chr17 7 675 140 G A	Pathogenic	Missense	SNP	R158C

[a]Not found in its duplicate on C1.
*stop codon.

TABLE 3 | List of genes, classified as non-LCDG or non-CPG, carrying the 20 identified pathogenic variants.

Gene	Cohort	Total number of mutated patients	Variant ID [24]	Manual ACMG [25]	Mutation class	Type	Protein change
ACADVL	C1	1	chr17 7 223 152 G A	Likely pathogenic	Missense	SNP	R366H
ADA	C2	1	chr20 44 636 279 G C	Pathogenic	Missense	SNP	H15D
BTD	C2	1	chr3 15 644 590 G A	Pathogenic	Missense	SNP	C225Y
GAA	C2	1	chr17 80 110 754 G A	Pathogenic	Missense	SNP	D489N
GALNS	C1, C2	1	chr16 88 841 948 G A	Pathogenic	Missense	SNP	R90W
GYG1	C1	1	chr3 148 996 462 G C	Likely pathogenic	Missense	SNP	D102H
HERC2	C2	1	chr15 28 142 240 C T	Likely pathogenic	Missense	SNP	E3900K
IVD	C1, C2	2 (different variants)	chr15 40 418 190 A G	Pathogenic	Missense	SNP	Y400C
	C1, C2		chr15 40 415 454 C T	Pathogenic	Missense	SNP	A311V
MMAB	C2	1	chr12 109 561 055 C T	Pathogenic	Missense	SNP	R190H
NUP93	C1	1	chr16 56 834 768 G T	Pathogenic	Missense	SNP	G591V
PEX1	C2	1	chr7 92 501 562 C T	Pathogenic	Missense	SNP	G786D
POLG	C2	1	chr15 89 325 639 G A	Pathogenic	Missense	SNP	P587L
PYGM	C2	2 (different variants)	chr11 64 755 483 A G	Likely pathogenic	Missense	SNP	S246P
	C2		chr11 64 753 126 G A	Pathogenic	Missense	SNP	P489S
RYR1	C2	1	chr19 38 543 551 A G	Likely pathogenic	Missense	SNP	Y3933C
SLC3A1	C1, C2	2	chr2 44 312 653 T C	Pathogenic	Missense	SNP	M467T
	C2						
TNFRSF13B	C1, C2	1	chr17 16 948 873 A G	Pathogenic	Missense	SNP	C104R
TSHR	C2	1	chr14 81 143 715 G A	Likely pathogenic	Missense	SNP	A553T
TYR	C2	1	chr11 89 178 342 A T	Likely pathogenic	Missense	SNP	E130V

variants in genes involved in DNA repair, cell cycle regulation, and tumor suppression. The most frequently mutated genes identified in these studies, including ours, were involved in DNA repair mechanisms, and many of them were common to multiple studies (Table 4). For example, *BRCA1* and *BRCA2* were identified in four studies, including ours, and *ATM* was found in three studies. Finally, in our study, no association was found between the presence of lung-driver or cancer-predisposing variants and the available clinical characteristics (sex, age, and survival, data not shown).

TABLE 4 | Comparison of previously published studies on pathogenic variant carrier genes identified in non-smokers.

Study	Type of data	Sample analyzed	Genes with germline variants in non-smokers	Genes in common with our study
Zhang et al. 2021 [27]	WES	232 non-smokers	83	*ATM* *BRCA1* *GJB2*
Liu et al. 2020 [9]	NGS of 58 known hereditary cancer-related genes	1026 (442 non-smokers)	22	*ATM* *BRCA1* *BRCA2*
Tlemsani et al. 2021 [29]	NGS of 607 cancer-related genes	87 (8 non-smokers)	4	*BRCA2*
Devarakonda et al. 2021 [28]	WES	160 non-smokers	10	BRCA1 BRCA2

4 | Discussion

It is highly plausible that the genetic background has a significant impact on the etiology of lung cancer, in particular in non-smoker subjects. While showing a significant heterogeneity both in type and function of the underlying potential PGVs, several studies support this conclusion. The WES analysis we conducted on 75 non-smoker NSCLC patients showed that 27 of them (36%) carried PGVs in 33 genes, 13 of which were classified as LCDG or CPG. PGVs identified in these genes were present in 13 patients (17.3%). This percentage is higher than that found in a study on 160 lung adenocarcinomas from non-smokers [28] and in a study on 7788 lung cancer cases who underwent germline genetic testing, mainly for a personal or a family history of cancer, and regardless of smoking status [10]; those studies identified PGVs in 6.9% and 14.9% of the cases, respectively. Zhang et al. [27] analyzed 232 NSCLC from non-smokers, mostly of European ancestry, and found a higher percentage of cases (36.6%) carrying pathogenic variants, but the threshold they set for pathogenic variant detection was lower. They identified 83 genes harboring PGVs in at least one sample and, among them, three were in common with our study (*BRCA1*, *GJB2*, and *ATM*). Regarding the study from Devarakonda and colleagues [28] on 160 lung adenocarcinomas from non-smoker patients, pathogenic variants were found in 10 genes, seven of which were involved in DNA repair: *BRCA1*, *BRCA2*, *FANCG, FANCM, HMBS, MSH6, NF1, POLD1, TMEM127*, and *WRN* genes. Two other genes, one of which, *RAD51C1*, was also involved in DNA repair, and *MPL*, were shared with smoker cases. Among these genes, *BRCA1* and *BRCA2* were in common with our study. Another study, dealing with an analysis focused on 58 known hereditary cancer-related genes, in 1026 NSCLCs collected at several hospitals in China (including 442 non-smoker cases), identified germline variants, eight pathogenic and 14 likely pathogenic, in 22 non-smoker patients [9]. No significant differences among smokers and non-smokers were identified. Interestingly, in non-smoker cases, genes overlapping those identified in our study were *BRCA2, ATM*, and *BRCA1*. *BRCA2* was the most common mutated gene (6/22 cases). In addition, a WES analysis of germline DNA of 87 patients (77 SCLC and 10 extrapulmonary small cell cancers) considering a total of 607 genes (including genes associated with hereditary cancers, cancer-associated genes, those with recurrent somatic mutation in lung cancers or curated from previously published studies on cancer predisposition genes) identified 35 genes mutated in at least one sample. Seven of them were also identified in our study (*MUTYH*, *BRCA2*, *BRCA1*, *DNMT3A*, *GJB2*, *MLH1*, and *TP53*). The cohort analyzed included eight never-smoking patients, four of whom were carrying PGVs, one in the *BRCA2* gene [29] Finally, in a very recent study that analyzed rare PGVs in 67 DDR genes in 3.040 cases of lung cancer and controls, variants in *MUTYH*, a gene historically associated with colorectal cancer and other gastrointestinal cancers, were also significantly associated with lung adenocarcinoma (the most common type in non-smokers) and the absence of smoking history [30].

Seven of the genes we identified (54%) are involved in DNA repair (*ATM, BRCA1, BRCA2, FH, MLH1, MUTYH*, *and TP53*) [26] suggesting that variants in this pathway may cause deregulation of DNA repair and cell cycle and increase cancer risk. The fact that even among the genes classified as LCDG or CPG, there were genes involved in DNA repair (*ADA*, *HERC2*, and *POLG*) highlights the importance of this pathway in lung cancer development in non-smokers. Indeed, also the other studies have identified pathogenic variants in DNA repair genes. Moreover, in the study on 7788 lung cancers, 61.3% of the identified PGVs were related to DNA damage-repair (DDR)/homologous recombination-repair (HRR) genes [10]. This retrospective study was conducted on data derived from germline genetic testing of lung cancer patients, performed mostly for a personal or family history of cancer; this may have increased the presence of germline pathogenic variants. However, the prevalence of germline pathogenic variants in DNA repair genes (mainly *BRCA2*, *ATM*, *CHEK2*, and *TP53*) was confirmed in patients with a personal history of lung cancer (2247 subjects) and also in the subgroup with only a physician-reported personal history of lung cancer and no reported family history. The association with lung cancer risk of PGVs in *ATM* and *BRCA2* genes was also observed in genes mapping at loci found in GWAS studies on selected high-risk LC cases, for example, with an early or family history of LC [31–33].

Environmental and occupational insults that can induce DNA damage have been identified as potential causes of lung cancer

[34]. In the context of defective DNA repair, environmental insults like cigarette smoking are highly mutagenic and, not surprisingly, PGVs in genes involved in DNA repair pathways have also been identified in smoking patients [35]. Germline variants in DNA repair genes are common also to other types of cancers [36] and, most importantly, are clinically actionable. A further consideration relies on the concept that these genes belong to the class of tumor suppressor genes (TSGs) and behave, in principle, as recessive genes. Indeed, in some instances, it has been possible to sequence also the tumor sample derived from a patient with a pathogenic heterozygous germline mutation in one TSG, finding evidence of a second inactivating hit [29]. However, a role in the pathogenic process of germline mutations, even in heterozygosity, in the TSG cannot be ruled out on the basis of a gene dosage effect.

We have further identified other genes carrying PGVs known to be associated with NSCLC, such as *BRAF*, *SDHA*, and *KEAP1* [26]. In this case, their role in carcinogenesis is more complicated to define. Regarding the 18 genes carrying PGVs classified as non-LCDG or CPG, many are related to metabolism [26]. Altered metabolism is a hallmark of cancer [37] and variants in genes involved in metabolic pathways have been found to play a role in tumor development [38]. Other relevant biological processes identified are DNA repair and base excision repair (three genes: *ADA*, *HERC2*, and *POLG*), confirming the importance of this mechanism in lung cancer development. Expression of *ADA* (adenosine deaminase), which converts deoxyadenosine to deoxyinosine generated by DNA damage, is altered in several cancer types, including lung cancer [39], while mutations in the *ADA* gene cause severe combined immunodeficiency disease (SCID) [40]. *HERC2* modulates p53-dependent gene expression during the *ATM* and *ATR* induced DNA double-strand break repair response [41]. *POLG* encodes the catalytic subunit of DNA polymerase γ, responsible for mitochondrial DNA replication, and has been related to cancer initiation, development, and prognosis [42]. Missense mutations in *POLG* have been correlated with breast cancer risk in patients with an indication of hereditary disease susceptibility [43]. Three genes, *SLC3A1*, *IVD*, and *PYGM*, have been found mutated in two patients. The pathogenic variant in *SLC3A1*, whose product is a dibasic amino acid transporter protein involved in the pathogenesis of cystinuria, was the same in the two patients and is the most common in the Italian population. The region including *SLC3A1* (chromosome 2p21) has been found gained in lung adenocarcinomas of non-smokers PEVuZE5vdGU [44]. The product of *IVD* (Isovaleryl-CoA dehydrogenase) is involved in leucine catabolism. Mutations in *IVD* are associated with isovaleric acidemia. Interestingly, a genetic variant in this gene has been associated with idiopathic pulmonary fibrosis, a progressive lung fibrotic condition that increases a patient's risk of developing lung cancer [45]. *PYGM* is a muscle glycogen phosphorylase implicated in several pathological states, including cancer. An analysis of genetic alterations in 60 NSCLC patients found the *PYGM* locus altered in 14.8% of the samples [46]. In addition, PYGM protein has been found up-regulated in lung cancer [47].

Reduced DNA repair capacity has been linked to an increased risk of lung cancer, as epithelial cells may be more susceptible to genotoxic damage [48]. However, the lack of information on exposure to environmental carcinogens and on the family history of patients prevented us from proposing a direct role of the identified PGVs in tumor development. Furthermore, we could not verify the inactivation of the normal allele at the tumor tissue level, confirming the second hit on the involved tumor suppressor gene, which could demonstrate a causal link between the germline mutation and the carcinogenic process.

Overall, our results strongly support the concept that non-smokers lung cancer patients carry PGVs in cancer predisposing genes, especially in lung cancer driver genes, and are directly and/or indirectly involved in cancer development. PGVs are, therefore, a genetic risk factor for lung cancer, suggesting that patients diagnosed with lung cancer, especially non-smokers, should be considered for germline molecular testing. In addition to patients, their family members may also benefit from recommended measures to prevent and diagnose early cancers for which these variants confer increased risk.

Author Contributions

Giovanni Carapezza: contributed to study conception and design, conducted data analyses and interpretation, and wrote the manuscript. **Simone Paolo Minardi:** provided technologies, supervised WES, acquired data, and wrote the manuscript. **Sara Noci:** contributed to study conception and design, formal analysis, and developed methodology; **Giulia Pintarelli:** contributed to formal analysis and developed methodology. **Susanna Zanutto:** contributed to study conception and design and developed methodology. **Matteo Incarbone:** contributed to study conception and design and provided material. **Davide Tosi:** contributed to study conception and design and provided material. **Tommaso Antonio Dragani:** contributed to study conception and design, supervised the study, interpreted data, provided material, and funding. **Francesca Colombo:** supervised the study, interpreted data, and provided material. **Marco Alessandro Pierotti:** conceived, designed the study, interpreted data, and wrote the manuscript. **Manuela Gariboldi:** conceived, designed, supervised the study, interpreted the data, and wrote the manuscript. All authors contributed to the review of the manuscript.

Acknowledgments

The authors thank the patients for participating in the study. Open access funding provided by BIBLIOSAN.

Ethics Statement

Approval of the research protocol by an Institutional Review Board: the protocol of this study was approved by the Committee for Ethics of the Fondazione IRCCS Istituto Nazionale dei Tumori, Milan (INT 224–17), Fondazione IRCCS Ca′ Granda Ospedale Maggiore Policlinico, Milan (2022019bis) and Ospedale San Giuseppe, IRCCS Multimedica, Milan (346.2018). Informed consent: written informed consent was obtained from patients to use their biological material and data for research purposes, according to the European General Data Protection. Regulation Registry and the registration No. of the study/trial: N/A. Animal studies: N/A.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Raw data (BAM files) of the whole genome sequencing datasets are available in the ArrayExpress database (http://www.ebi.ac.uk/arrayexpress) under accession numbers: E-MTAB-14468 and E-MTAB-14520.

References

1. J. Ferlay, I. Soerjomataram, R. Dikshit, et al., "Cancer Incidence and Mortality Worldwide: Sources, Methods and Major Patterns in GLOBOCAN 2012," *International Journal of Cancer* 136 (2015): E359–E386.

2. L. Pelosof, C. Ahn, A. Gao, et al., "Proportion of Never-Smoker Non-Small Cell Lung Cancer Patients at Three Diverse Institutions," *Journal of the National Cancer Institute* 109 (2017): djw295.

3. S. Dubin and D. Griffin, "Lung Cancer in Non-Smokers," *Missouri Medicine* 117 (2020): 375–379.

4. M. Fruh, D. De Ruysscher, S. Popat, et al., "Small-Cell Lung Cancer (SCLC): ESMO Clinical Practice Guidelines for Diagnosis, Treatment and Follow-Up," *Annals of Oncology* 24 Suppl 6 (2013): vi99–vi105.

5. J. LoPiccolo, A. Gusev, D. C. Christiani, and P. A. Janne, "Lung Cancer in Patients Who Have Never Smoked–An Emerging Disease," *Nature Reviews Clinical Oncology* 21 (2024): 121–146.

6. Y. Bosse and C. I. Amos, "A Decade of GWAS Results in Lung Cancer," *Cancer Epidemiology, Biomarkers & Prevention* 27 (2018): 363–379.

7. K. L. Huang, R. J. Mashl, Y. Wu, et al., "Pathogenic Germline Variants in 10,389 Adult Cancers," *Cell* 173 (2018): 355–370.e14.

8. P. R. Benusiglio, V. Fallet, M. Sanchis-Borja, F. Coulet, and J. Cadranel, "Lung Cancer Is Also a Hereditary Disease," *European Respiratory Review* 30, no. 162 (2021): 210045, https://doi.org/10.1183/16000617.0045-2021.

9. M. Liu, X. Liu, P. Suo, et al., "The Contribution of Hereditary Cancer-Related Germline Mutations to Lung Cancer Susceptibility," *Translational Lung Cancer Research* 9 (2020): 646–658.

10. S. Sorscher, J. LoPiccolo, B. Heald, et al., "Rate of Pathogenic Germline Variants in Patients With Lung Cancer," *JCO Precision Oncology* 7 (2023): e2300190, https://doi.org/10.1200/PO.23.00190.

11. G. Pintarelli, C. E. Cotroneo, S. Noci, et al., "Genetic Susceptibility Variants for Lung Cancer: Replication Study and Assessment as Expression Quantitative Trait Loci," *Scientific Reports* 7 (2017): 42185.

12. A. Galvan, F. Colombo, E. Frullanti, et al., "Germline Polymorphisms and Survival of Lung Adenocarcinoma Patients: A Genome-Wide Study in Two European Patient Series," *International Journal of Cancer* 136 (2015): E262–E271.

13. S. Chen, Y. Zhou, Y. Chen, and J. Gu, "Fastp: An Ultra-Fast All-In-One FASTQ Preprocessor," *Bioinformatics* 34 (2018): i884–i890.

14. K. Okonechnikov, A. Conesa, and F. Garcia-Alcalde, "Qualimap 2: Advanced Multi-Sample Quality Control for High-Throughput Sequencing Data," *Bioinformatics* 32 (2016): 292–294.

15. S. Kim, K. Scheffler, A. L. Halpern, et al., "Strelka2: Fast and Accurate Calling of Germline and Somatic Variants," *Nature Methods* 15 (2018): 591–594.

16. M. M. Li, M. Datto, E. J. Duncavage, et al., "Standards and Guidelines for the Interpretation and Reporting of Sequence Variants in Cancer: A Joint Consensus Recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists," *Journal of Molecular Diagnostics* 19 (2017): 4–23.

17. D. Tamborero, C. Rubio-Perez, J. Deu-Pons, et al., "Cancer Genome Interpreter Annotates the Biological and Clinical Relevance of Tumor Alterations," *Genome Medicine* 10 (2018): 25.

18. F. Muinos, F. Martinez-Jimenez, O. Pich, A. Gonzalez-Perez, and N. Lopez-Bigas, "In Silico Saturation Mutagenesis of Cancer Genes," *Nature* 596 (2021): 428–432.

19. A. D. Scott, K. L. Huang, A. Weerasinghe, et al., "CharGer: Clinical Characterization of Germline Variants," *Bioinformatics* 35 (2019): 865–867.

20. H. Wang, T. Wang, X. Zhao, et al., "AI-Driver: An Ensemble Method for Identifying Driver Mutations in Personal Cancer Genomes," *NAR Genomics and Bioinformatics* 2 (2020): lqaa084.

21. D. L. Masica, C. Douville, C. Tokheim, et al., "CRAVAT 4: Cancer-Related Analysis of Variants Toolkit," *Cancer Research* 77 (2017): e35–e38.

22. M. J. Landrum, S. Chitipiralla, G. R. Brown, et al., "ClinVar: Improvements to Accessing Data," *Nucleic Acids Research* 48 (2020): D835–D844.

23. F. Martinez-Jimenez, F. Muinos, I. Sentis, et al., "A Compendium of Mutational Cancer Driver Genes," *Nature Reviews. Cancer* 20 (2020): 555–572.

24. "Varsome: The Human Genomic Variants Search Engine" (2023), https://varsome.com/.

25. S. Richards, N. Aziz, S. Bale, et al., "Standards and Guidelines for the Interpretation of Sequence Variants: A Joint Consensus Recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology," *Genetics in Medicine* 17, no. 5 (2015): 405–424, https://doi.org/10.1038/gim.2015.30.

26. Y. Zhou, B. Zhou, L. Pache, et al., "Metascape Provides a Biologist-Oriented Resource for the Analysis of Systems-Level Datasets," *Nature Communications* 10 (2019): 1523.

27. T. Zhang, P. Joubert, N. Ansari-Pour, et al., "Genomic and Evolutionary Classification of Lung Cancer in Never Smokers," *Nature Genetics* 53, no. 9 (2021): 1348–1359, https://doi.org/10.1038/s41588-021-00920-0.

28. S. Devarakonda, Y. Li, F. Martins Rodrigues, et al., "Genomic Profiling of Lung Adenocarcinoma in Never-Smokers," *Journal of Clinical Oncology* 39, no. 33 (2021): 3747–3758, https://doi.org/10.1200/JCO.21.01691.

29. C. Tlemsani, N. Takahashi, L. Pongor, et al., "Whole-Exome Sequencing Reveals Germline-Mutated Small Cell Lung Cancer Subtype With Favorable Response to DNA Repair-Targeted Therapies," *Science Translational Medicine* 13 (2021): 13.

30. M. R. Trendowski, C. M. Lusk, A. S. Wenzlaff, et al., "Association of Germline Pathogenic Variants in MUTYH and Other DNA Damage Response Genes With Lung Cancer Risk Among Non-Hispanic Whites and African Americans," *JCO Precision Oncology* 9 (2025): e2400558, https://doi.org/10.1200/PO-24-00558.

31. Y. Liu, J. Xia, J. McKay, et al., "Rare Deleterious Germline Variants and Risk of Lung Cancer," *NPJ Precision Oncology* 5 (2021): 12.

32. Y. Liu, C. M. Lusk, M. H. Cho, et al., "Rare Variants in Known Susceptibility Loci and Their Contribution to Risk of Lung Cancer," *Journal of Thoracic Oncology* 13 (2018): 1483–1495.

33. Y. Liu, F. Kheradmand, C. F. Davis, et al., "Focused Analysis of Exome Sequencing Data for Rare Germline Mutations in Familial and Sporadic Lung Cancer," *Journal of Thoracic Oncology* 11 (2016): 52–61.

34. B. C. Bade and C. S. Dela Cruz, "Lung Cancer 2020: Epidemiology, Etiology, and Prevention," *Clinics in Chest Medicine* 41 (2020): 1–24.

35. E. M. Parry, D. L. Gable, S. E. Stanley, et al., "Germline Mutations in DNA Repair Genes in Lung Adenocarcinoma," *Journal of Thoracic Oncology* 12 (2017): 1673–1678.

36. C. J. Lord and A. Ashworth, "BRCAness Revisited," *Nature Reviews. Cancer* 16 (2016): 110–120.

37. D. Hanahan and R. A. Weinberg, "Hallmarks of Cancer: The Next Generation," *Cell* 144 (2011): 646–674.

38. A. Erez and R. J. DeBerardinis, "Metabolic Dysregulation in Monogenic Disorders and Cancer - Finding Method in Madness," *Nature Reviews. Cancer* 15 (2015): 440–448.

39. G. Zhulai, E. Oleinik, M. Shibaev, and K. Ignatev, "Adenosine-Metabolizing Enzymes, Adenosine Kinase and Adenosine Deaminase, in Cancer," *Biomolecules* 12, no. 3 (2022): 418, https://doi.org/10.3390/biom12030418.

40. D. B. Kohn, M. S. Hershfield, D. Carbonaro, et al., "T Lymphocytes With a Normal ADA Gene Accumulate After Transplantation of

Transduced Autologous Umbilical Cord Blood CD34+ Cells in ADA-Deficient SCID Neonates," *Nature Medicine* 4, no. 7 (1998): 775–780, https://doi.org/10.1038/nm0798-775.

41. M. Cubillos-Rojas, F. Amair-Pinedo, R. Peiro-Jordan, R. Bartrons, F. Ventura, and J. L. Rosa, "The E3 Ubiquitin Protein Ligase HERC2 Modulates the Activity of Tumor Protein p53 by Regulating Its Oligomerization," *Journal of Biological Chemistry* 289 (2014): 14782–14795.

42. A. R. Maiuri, H. Li, B. D. Stein, J. M. Tennessen, and H. M. O'Hagan, "Inflammation-Induced DNA Methylation of DNA Polymerase Gamma Alters the Metabolic Profile of Colon Tumors," *Cancer & Metabolism* 6 (2018): 9.

43. A. Tervasmaki, T. Mantere, J. M. Hartikainen, et al., "Rare Missense Mutations in RECQL and POLG Associate With Inherited Predisposition to Breast Cancer," *International Journal of Cancer* 142 (2018): 2286–2292.

44. B. Job, A. Bernheim, M. Beau-Faller, et al., "Genomic Aberrations in Lung Adenocarcinoma in Never Smokers," *PLoS One* 5 (2010): e15145.

45. H. Furusawa, A. L. Peljto, A. D. Walts, et al., "Common Idiopathic Pulmonary Fibrosis Risk Variants Are Associated With Hypersensitivity Pneumonitis," *Thorax* 77 (2022): 508–510.

46. A. S. Chan, W. K. Lam, M. P. Wong, et al., "Chromosomal 11 Alterations in Non-Small-Cell Lung Carcinomas in Hong Kong," *Lung Cancer* 15 (1996): 51–65.

47. M. D. Pastor, A. Nogal, S. Molina-Pinelo, et al., "Identification of Proteomic Signatures Associated With Lung Cancer and COPD," *Journal of Proteomics* 89 (2013): 227–237.

48. H. C. Wu, R. Kehm, R. M. Santella, D. J. Brenner, and M. B. Terry, "DNA Repair Phenotype and Cancer Risk: A Systematic Review and Meta-Analysis of 55 Case-Control Studies," *Scientific Reports* 12 (2022): 3405.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.