Original Article

The contribution of hereditary cancer-related germline mutations to lung cancer susceptibility

Mengyuan Liu[1#], Xinyi Liu[1#], Peisu Suo[1#], Yuan Gong[2#], Baolin Qu[2], Xiumei Peng[3], Wenhua Xiao[3], Yuemin Li[4], Yan Chen[5], Zhen Zeng[5], Yinying Lu[5], Tanxiao Huang[1], Yingshen Zhao[1], Ming Liu[1], Lifeng Li[1], Yaru Chen[1], Yanqing Zhou[1], Guifeng Liu[1], Jianfei Yao[1], Shifu Chen[1], Lele Song[1,4#]

[1]HaploX Biotechnology, Co., Ltd., Shenzhen 518057, China; [2]The Second Medical Center of the Chinese PLA General Hospital, Beijing 100853, China; [3]The Fourth Medical Center of the Chinese PLA General Hospital, Beijing 100037, China; [4]The Eighth Medical Center of the Chinese PLA General Hospital, Beijing 100091, China; [5]The Fifth Medical Center of the Chinese PLA General Hospital, Beijing 100039, China

Contributions: (I) Conception and design: L Song, M Liu; (II) Administrative support: G Liu, J Yao; (III) Provision of study materials or patients: Y Gong, B Qu, X Peng, W Xiao, Y Li, Y Chen, Z Zeng, Y Lu; (IV) Collection and assembly of data: M Liu, X Liu, P Suo, T Huang, Y Zhao, M Liu, L Li, Y Chen, Y Zhou, G Liu, J Yao, S Chen; (V) Data analysis and interpretation: M Liu, X Liu, P Suo, L Song; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Lele Song. Department of Radiotherapy, the Eighth Medical Center of the Chinese PLA General Hospital, No. 17, Heishanhu Road, Haidian District, Beijing 100091, China. Email: songlele@sina.com.

Background: Germline variations may contribute to lung cancer susceptibility besides environmental factors. The influence of germline mutations on lung cancer susceptibility and their correlation with somatic mutations has not been systematically investigated.

Methods: In this study, germline mutations from 1,026 non-small cell lung cancer (NSCLC) patients were analyzed with a 58-gene next-generation sequencing (NGS) panel containing known hereditary cancer-related genes, and were categorized based on American College of Medical Genetics and Genomics (ACMG) guidelines in pathogenicity, and the corresponding somatic mutations were analyzed using a 605-gene NGS panel containing known cancer-related genes.

Results: Plausible genetic susceptibility was found in 4.7% of lung cancer patients, in which 14 patients with pathogenic mutations (P group) and 34 patients with likely-pathogenic mutations (LP group) were identified. The ratio of the first degree relatives with lung cancer history of the P groups was significantly higher than the Non-P group (P=0.009). The ratio of lung cancer patients with history of other cancers was higher in P (P=0.0007) or LP (P=0.017) group than the Non-P group. Pathogenic mutations fell most commonly in *BRCA2*, followed by *CHEK2* and *ATM*. Likely-pathogenic mutations fell most commonly in *NTRK1* and *EXT2*, followed by *BRIP1* and *PALB2*. These genes are involved in DNA repair, cell cycle regulation and tumor suppression. By comparing the germline mutation frequency from this study with that from the whole population or East Asian population (gnomAD database), we found that the overall odds ratio (OR) for P or LP group was 17.93 and 15.86, respectively, when compared with the whole population, and was 2.88 and 3.80, respectively, when compared with the East Asian population, suggesting the germline mutations of the P and LP groups were risk factors for lung cancer. Somatic mutation analysis revealed no significant difference in tumor mutation burden (TMB) among the groups, although a trend of lower TMB in the pathogenic group was found. The SNV/INDEL mutation frequency of TP53 in the P group was significantly lower than the other two groups, and the copy number variation (CNV) mutation frequency of *PIK3CA* and *MET* was significantly higher than the Non-P group. Pathway enrichment analysis found no significant difference in aberrant pathways among the three groups.

Conclusions: A proportion of 4.7% of patients carrying germline variants may be potentially linked to increased susceptibility to lung cancer. Patients with pathogenic germline mutations exhibited stronger family history and higher lung cancer risk.

Keywords: Lung cancer; germline; susceptibility; pathogenic; BRCA2; EGFR

Submitted Sep 06, 2019. Accepted for publication Mar 13, 2020.
doi: 10.21037/tlcr-19-403
View this article at: http://dx.doi.org/10.21037/tlcr-19-403

Introduction

The germline mutations in multiple genes confer significant risks to several cancers, including breast, ovarian, colorectal cancer and melanoma. In contrast, the genetic predisposition of lung cancer has not yet been elucidated. Although most lung cancers develop sporadically and cigarette smoking is considered to be the predominant risk factor (1), many lung cancer patients present a family clustered pattern. It was reported that a family history confer a substantial risk to lung cancer, especially for those with two or more affected individuals in a family (2).

Since the incidence of definite pathogenic germline mutations are very low, most studies on germline mutations in lung cancer were case report studies, and only a couple of population-based studies so far reporting the prevalence of germline mutations in lung cancer (3-5). Germline *EGFR* mutations are by far the most frequently reported genetic variations in lung cancer (6), among which *EGFR* T790M was the most reported germline mutation. It was reported that the prevalence of *EGFR* T790M germline mutations in East Asian was much lower than that in the Western population (7-9). Therefore, the germline mutation spectrum in lung cancer in different ethnics may be distinct. Other *EGFR* germline mutations, including V843I, R776G/H, P848L, K757R, D1014N, I646S, G724S, V786M, L792F, R831H, and L844V were also reported with very low incidence (7-9). Apart from EGFR, germline mutations of other genes, including *HER2*, *RET*, *BRCA1*, *BRCA2* (9), *PARK2* (10), *YAP1* (11), *CHEK2* (12), *TERT* (13), *TP53*, *CDKN2A*, *MET*, *NBN* (14), were also reported and linked with lung cancer risk.

Although some germline mutations, such as those in *EGFR* and *HER2*, have been identified in lung cancer in previous observations (3-14), the susceptibility of lung cancer with known hereditary cancer-related germline mutations has not been investigated, and the correlation between germline mutations and somatic mutations has not been studied in detail. The information is sorely lacking among the Chinese population. In this study, we studied the potential susceptibility of lung cancer by categorizing the germline mutations of individual lung cancer patients into three groups based on pathogenicity. Germline and somatic mutation spectrum for each group were obtained by next-generation sequencing (NGS) with a 58-gene panel and a 605-gene panel, respectively. Potential risk factors, such as age, sex, family history, and cancer characteristics, such as cancer type, mutation frequency, tumor mutation burden (TMB) and aberrant pathways, were investigated and compared.

Methods

Ethic approval by participating hospitals

All experiment plans and protocols for the study were submitted to the ethics/licensing committees of the named participating hospitals for review and approval before the start of the clinical study, and were approved by the corresponding committees of hospitals, including the Chinese PLA General Hospital, the Fourth Medical Center of the Chinese PLA General Hospital, the Fifth Medical Center of the Chinese PLA General Hospital and the Eighth Medical Center of the Chinese PLA General Hospital. Confirmation of approval for clinical studies was received from the ethics board of the Chinese PLA General Hospital (approval number: S2018-081-02) before the start of the clinical study. Since the study was designed as a retrospectively study and used retrospective samples collected by the above hospitals, no informed consent was required. Patients with pathogenic or likely pathogenic germline mutations were informed the test results. All experiments, methods, procedures and personnel training were carried out in accordance with relevant guidelines and regulations of participating hospitals and laboratories.

Study design, patients and samples

The study was designed and implemented in four Chinese hospitals, and both cancer tissue and blood samples were collected retrospectively. The study was designed to include as many non-small cell lung cancer (NSCLC) patients as possible, as long as the tissue or blood samples

were available for next generation sequencing (NGS). As a result, samples collected between June, 2018 and June, 2019 from 1,026 NSCLC patients were obtained based on the availability of samples for NGS test in the participating hospitals, including 792 patients with adenocarcinoma (ADC), 222 patients with squamous cell carcinoma (SCC), 6 patients with large cell carcinoma (LCC) and 6 patients with adenosquamous carcinoma (ASC) (*Table 1*). Information on clinicopathological status of all patients was collected (*Table 1*). Family history here is defined as: the confirmed lung cancer patient has at least one immediate family member (first degree relatives) who had a history of lung cancer diagnosis. The immediate family member includes father, mother, brother(s), sister(s), son(s), daughter(s). The collected samples involved tissue samples, including formalin-fix paraffin-embedded (FFPE) samples or frozen samples from surgery or needle biopsy, and blood samples obtained at the time of confirmed lung cancer diagnosis. All technicians were blinded to the clinical information of subjects. The classification of all conditions was based on diagnosis from imaging examinations and subsequent pathological examinations. None of the subjects received chemotherapy, radiotherapy, targeted therapy or immunotherapy before tissue or blood samples were collected. The somatic sequencing data presented in this study were from FFPE samples or frozen tissue samples. Germline sequencing data was obtained from the corresponding genomic DNA of white blood cells.

Sample preparation, targeted NGS and data processing

For the FFPE samples, ten 5 μm tumor slices were used for DNA extraction using the QIAamp DNA FFPE Kit (QIAGEN, Valencia, CA, USA) following the manufacturer's instructions. For blood samples, 2 mL blood were collected in tubes containing EDTA and centrifuged at 1,600 ×g for 10 min at 4 ℃ within 2 h of collection. The peripheral blood lymphocyte (PBL) debris was stored at −20 ℃ until further use. DNA from PBLs was extracted using the RelaxGene Blood DNA system (Tiangen Biotech Co., Ltd., Beijing, China) according to the manufacturers' instructions. Both cancer tissue and white blood cell genomic DNA was quantified with the Qubit 2.0 Fluorometer and the Qubit dsDNA HS assay kit (Thermo Fisher Scientific, Inc., Waltham, MA, USA) according to manufacturer's instructions. Fragmented genomic DNA underwent end-repairing, A-tailing and ligation with indexed adapters sequentially, followed by size selection using Agencourt AMPure XP beads (Beckman Coulter Inc., Brea, CA, USA), and DNA fragments were used for library construction using the KAPA Library Preparation kit (Kapa Biosystems, Inc., Wilmington, MA, USA) according to the manufacturer's protocol. Hybridization-based target enrichment was carried out with HaploX germline gene panel (58 known hereditary cancer-related genes, HaploX Biotechnology, gene list is provided in *Table S1*) for white blood cell genomic DNA or HaploX pan-cancer gene panel (605 cancer-relevant genes, HaploX Biotechnology, gene list is provided in *Table S2*) for cancer tissue sequencing. Seven to eight polymerase chain reaction (PCR) cycles, depending on the amount of DNA used, were performed by pre-capture ligation-mediated PCR (Pre−LM−PCR) Oligos (Kapa Biosystems, Inc.) in 50 μL reactions. DNA sequencing was then performed on the Illumina Novaseq 6000 system according to the manufacturer's recommendations at an average depth of 2,200×.

Data which meet the following criteria were chosen for subsequent analysis: the ratio of remaining data filtered by fastq in raw data is ≥85%; the proportion of Q30 bases is ≥85%; the ratio of reads on the reference genome is ≥85%; target region coverage ≥98%; average sequencing depth in tissues is ≥2,200×. The called somatic variants need to meet the following criteria: the read depth at a position is ≥20×; the variant allele fraction (VAF) is ≥2% for tissue and PBL genomic DNA; somatic-P value ≤0.01; strand filter ≥1. VAF were calculated for Q30 bases. The copy number variation (CNV) was detected by CNVkit version 0.9.3 (https://github.com/etal/cnvkit). Further analyses of genomic alterations were also performed, including single nucleotide variants (SNVs), CNVs, insertion/deletion (Indels), fusions and structural variation.

Interpretation of pathogenicity of germline mutations and calculation of somatic TMB

Pathogenicity of germline mutations was defined and predicted based on the five-grade classification system according to the American College of Medical Genetics and Genomics (ACMG) Guidelines for the Interpretation of Sequence (15). The VUS, benign and likely benign mutations were defined as the non-pathogenic group (Non-P) in this study. As a result, all germline mutations were categorized into pathogenic (P), likely pathogenic (LP) or non-pathogenic group (Non-P) in this study. TMB was calculated by dividing the total number of tissue non-synonymous SNP and INDEL variations (VAF >2%) by

Table 1 The summary of clinicopathological and history information for NSCLC patients with distinct germline mutation pathogenicity

Clinicopathological factors	Subgroups	Total (N=1,026)		Pathogenic (N=14)		Likely pathogenic (N=34)		Non-pathogenic (N=978)		P
		n	%	n	%	n	%	n	%	
NSCLC	Adenocarcinoma	792	77.19	12	85.71	26	76.47	754	77.10	0.45
	Squamous	222	21.64	1	7.14	8	23.53	213	21.78	
	Large cell	6	0.58	1	7.14	0	0.00	5	0.51	
	Adenosquamous	6	0.58	0	0.00	0	0.00	6	0.61	
Age, year	<40	47	4.58	1	7.14	1	2.94	45	4.60	0.81
	≥40	979	95.42	13	92.86	33	97.06	933	95.40	
	<50	181	17.64	4	28.57	5	14.71	172	17.59	0.51
	≥50	845	82.36	10	71.43	29	85.29	806	82.41	
	<60	473	46.10	10	71.43	12	35.29	451	46.11	0.074
	≥60	553	53.90	4	28.57	22	64.71	527	53.89	
	<70	820	79.92	13	92.86	28	82.35	779	79.65	0.44
	≥70	206	20.08	1	7.14	6	17.65	199	20.35	
Sex	Male	594	57.89	8	57.14	22	64.71	564	57.67	0.72
	Female	432	42.11	6	42.86	12	35.29	414	42.33	
Stage	I–IIIA	568	55.36	5	35.71	15	44.12	548	56.03	0.12
	IIIB–IV	458	44.64	9	64.29	19	55.88	430	43.97	
Smoking history	Yes	584	56.92	6	42.86	20	58.82	558	57.06	0.55
	No	442	43.08	8	57.14	14	41.18	420	42.94	
History of prior malignancy	Yes	40	3.90	3	21.43	4	11.76	36	3.68	0.0004
	No	986	96.10	11	78.57	30	88.24	942	96.32	
Family history*	Yes	275	26.80	8	57.14	11	32.35	256	26.18	0.026
	No	751	73.20	6	42.86	23	67.65	722	73.82	

*, family history: the confirmed lung cancer patient has at least one immediate family member (first degree relatives) who had a history of lung cancer diagnosis.

the full length of the exome region of the 605-gene NGS panel (*Table S2*). Genomic sequence from the DNA of PBLs was used for genomic alignment when calling the somatic mutations.

Statistics and data analysis

Statistical analysis was performed and figures were plotted with GraphPad Prism 5.0 software (GraphPad Software, Inc, La Jolla, CA 92037, USA). Student *t*-test was performed when two groups were compared, and ANOVA and *post hoc* tests were performed when three or more groups were compared. Chi-square test and Fisher test were performed when rate or percentage was compared for significance. Figures for mutation spectrum were made with the R software (https://www.r-project.org/). Data for pathway enrichment analysis was analyzed using the method described by DAVID Bioinformatics Resources 6.8 (https://david.ncifcrf.gov/) and visualized by corresponding packages of the R software. The odds ratio was calculated based on the frequency of a certain germline mutation from the Genome Aggregation Database (gnomAD) in general population or East Asian population and the corresponding frequency of mutation obtained from this study. The odds

ratio and 95% confidence interval (CI) for each germline mutation was calculated using the calculation module from the SPSS 17.0 software (IBM China Company Limited, Beijing 100101, China). P<0.05 is statistically significant.

Results

Characteristics of pathogenic and likely pathogenic germline mutations in Chinese lung cancer patients and their impact on lung cancer risk

Fourteen patients were found to carry 13 pathogenic (P) germline mutations, and 34 patients carried 36 likely pathogenic (LP) germline mutations, and the remaining 978 patients all carried non-pathogenic (Non-P) mutations (*Table 1*, *Figure 1A*,*B*). No significant difference among the three groups were found with pathological subtypes (P=0.45), age (P values was shown for various age groups in *Table 1*), stage (P=0.12), sex (P=0.72) or smoking history (P=0.55) (*Table 1*). This was also true when P and LP groups were combined (*Table S3*). Interestingly, the ratio of lung cancer patients with at least one immediate family member (first degree relatives) with lung cancer history was significantly higher in the P group than the Non-P group (P=0.009), indicating that pathogenic cancer-predisposing variants predisposed to lung cancer and resulted in familial clustering. Furthermore, the ratio of lung cancer patients with history of other cancers (history of prior malignancy) was higher in P (P=0.0007) or LP (P=0.017) group than the Non-P group (*Table 1*), suggesting that the presence of pathogenic germline mutations also increased the incidence of other cancers. This was also true when P and LP groups were combined and compared with the Non-P group (*Table S3*), in which significant differences were also found regarding family history (P=0.041) and history of prior malignancy (P=0.0002).

Detailed study identified 6 out of 14 patients in the P group carried *BRCA2* pathogenic mutations (6/14), followed by *CHEK2* (3/14) and *ATM* (2/14) (*Table 2*, *Figure 1A*). In the LP group, 4 out of 34 patients carried *NTRK1* mutations (4/34), 4 carried *EXT2* mutations (4/34), followed by *BRIP1*(3/34) and *PALB2* (3/34) (*Table 2*, *Figure 1B*). The functions of genes with pathogenic and likely pathogenic mutations mainly involved DNA repair (*BRCA1* and *BRCA2*, *BLM*, *RAD50*, *BRIP1*, *MLH3*), cell cycle regulation (such as *CHEK2*, *ATM*, *NTRK1* and *EPCAM*) and tumor suppressor (such as *PALB2* and *BRCA1*). Most of these fragmental mutations were located within or close to known important protein functional domains (*Figure 1C*,*D*) and may have great impacts on protein function.

In order to study the risk of lung cancer in individuals carrying pathogenic or likely pathogenic germline mutations, we searched the mutation prevalence of all germline mutations in total population and the East Asian population from the Genome Aggregation Database (gnomAD) (*Table 2*). By comparing the germline mutation frequency found in this study with the variant prevalence in total population and East Asian population, we calculated the overall odds ratio (OR) for the germline mutations in our study. The overall OR value of the P and LP groups was 17.93 (95% CI: 9.74 to 33.01) and 15.86 (95% CI: 5.999 to 133.2), respectively, when compared with the total population, and was 2.88 (95% CI: 0.32 to 25.79) and 3.80 (95% CI: 0.47 to 30.96), respectively, when compared with the East Asian population, suggesting that the pathogenic and likely pathogenic germline mutations were risk factors for lung cancer (*Table 2*).

Characteristics of somatic mutations of lung cancer patients carrying germline pathogenic or likely pathogenic mutations

The relationship between germline variations and somatic mutations in lung cancer has not been investigated in detail. We therefore mapped the somatic SNV/INDEL mutation spectrum (*Figure S1*) and CNV mutation spectrum (*Figure S2*) categorized by pathogenicity of germline mutations of all lung cancer patients in this study, and investigated the involved genes and somatic mutation characteristics (*Figure 2*). No statistically significant difference in TMB among the three groups was identified (*Figure 2A*), however, there was a trend that the TMB in the P group was lower than that of the LP group (P=0.13) and the Non-P group (P=0.09). The average TMB and Inter-Quartile Range (IQR) were 4.07 muts/MB (IQR: 6.74), 5.94 muts/MB (IQR: 5.22) and 6.56 muts/MB (IQR: 6.09) for the P, LP and Non-P group, respectively. The specific driver genes involved attracted our attention. The SNV/INDEL mutation rate (frequency) of *TP53* and *EGFR* was the highest among all genes (*Figure 2B*). The *TP53* mutation rate in the P group was significantly lower than that of the LP (P=0.018) and Non-P groups (P=0.003) (*Figure 2B*, *Figure S1*), while no such difference was found with *EGFR*. We also examined the mutation rate of CNVs in the three groups (*Figure 2C*). The most common genes with CNVs involved *TERT*, *EGFR*, *RICTOR* and *PIK3CA*. It appeared that the CNV mutation rate (frequency) of

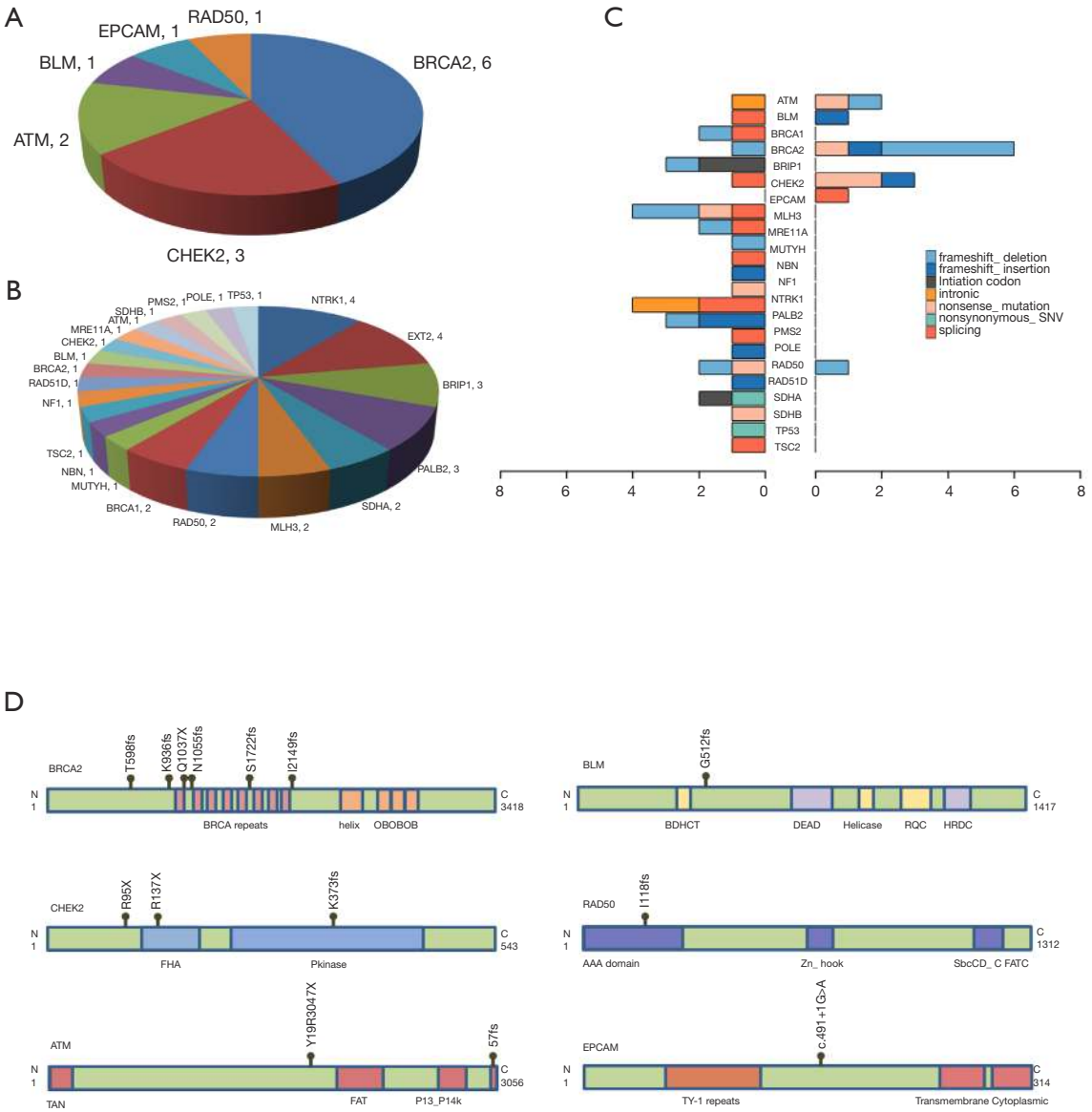

Figure 1 Gene names, variation types and number of variations of all pathogenic (P) and likely pathogenic (LP) germline mutations, and a scheme of the pathogenic germline variants and the position of individual mutations of the pathogenic mutations found in this study. Gene names, the number of mutations and the ratio of mutations of pathogenic germline variations and likely pathogenic variations are shown in (A,B), respectively. Mutation types and the corresponding number of mutations for P and LP groups are shown in (C). The scheme and key functional domains of *BRCA2*, *CHECK2*, *ATM*, *BLM*, *RAD50* and *EPCAM* are shown as individual panels in (D), and the position of 14 germline mutations are marked on each panel.

PIK3CA in the LP group was significantly higher than that of the Non-P group (P=0.013) but not the P group (P=0.35) (*Figure 2C*, *Figure S2*). Furthermore, the CNV mutation rate of the *MET* in the LP group was significantly higher than that of the Non-P group (P=0.011). Pathway enrichment analysis on P, LP and Non-P groups was performed, and both GO and KEGG enrichment revealed no significant differences in the functions or biological

Table 2 Summary of patient and mutation information and OR for lung cancer patients with pathogenic or likely pathogenic germline mutations in this study

Number	Age	Gender	Cancer type	Family history	Smoking history	Gene	Protein change	Annotation	Association with diseases	General population*			East Asian*		
										Allele frequency	OR	95% CI	Allele frequency	OR	95% CI
Pathogenic															
1	56	M	ADC	Yes	Yes	*BRCA2*	p.S1722fs	P	HBOC or PC	0.000032 (1/30,910)	28.26	6.00 to 133.17	0.00062 (1/1,614)	1.57	0.098 to 25.19
2	65	F	ADC	Yes	No					0.000032 (1/30,910)	28.26	6.00 to 133.17	0.00062 (1/1,614)	1.57	0.098 to 25.19
3	46	F	ADC	Yes	No	*BRCA2*	p.I2149fs	P	HBOC, PC, HCPS	N/A	N/A	N/A	N/A	N/A	N/A
4	65	M	ADC	No	Yes	*BRCA2*	p.K936fs	P	HBOC or PC	0.000012 (3/245,804)	37.65	3.92 to 362.3	N/A	N/A	N/A
5	56	F	ADC	Yes	No	*BRCA2*	p.T598fs	P	HBOC, PC, HCPS	0.0000042 (1/239,126)	113	7.07 to 1807	N/A	N/A	N/A
6	49	M	ADC	No	Yes	*BRCA2*	p.Q1037X	P	HBOC or PC	0.0000041 (1/224,307)	113	7.07 to 1807	0.000058 (1/17,218)	16.8	1.05 to 268.75
7	54	M	ADC	Yes	No	*CHEK2*	p.R95X	P	Hereditary or familial breast cancer, HCPS	0.0000081 (2/246,164)	56.48	5.12 to 623.4	N/A	N/A	N/A
8	75	M	LCC	No	Yes	*CHEK2*	p.R137X	P	Hereditary or familial breast cancer, HCPS	0.000024 (6/246,076)	18.83	2.27 to 156.5	N/A	N/A	N/A
9	66	F	ADC	Yes	No	*CHEK2*	p.K373fs	P	Hereditary or familial breast cancer, HCPS	N/A	N/A	N/A	N/A	N/A	N/A
10	60	F	ADC	No	No	*ATM*	p.Y1957fs	P	Ataxia-telangiectasia syndrome, HCPS	0.0000041 (1/245,874)	113	7.07 to 1,807	N/A	N/A	N/A
11	86	M	ADC	No	No	*ATM*	p.R3047X	P	Ataxia-telangiectasia syndrome, HCPS	0.000016 (4/246,234)	28.24	3.16 to 252.9	N/A	N/A	N/A
12	47	F	ADC	Yes	No	*BLM*	p.G512fs	P	Bloom syndrome	0.00011 (25/236,928)	4.34	0.59 to 32.04	0.00006 (1/16,610)	16.205	1.01 to 259.26
13	58	M	SCC	Yes	Yes	*RAD50*	p.I118fs	P	Hereditary or familial breast cancer, HCPS	0.000012 (3/245,582)	37.65	3.92 to 362.3	N/A	N/A	N/A
14	51	M	ADC	No	Yes	*EPCAM*	c.491+1G>A	P	Lynch syndrome; congenital tufting enteropathy	0.000053 (13/246,044)	8.69	1.14 to 66.48	N/A	N/A	N/A
Overall										0.00031	17.93	9.74 to 33.01	0.00136	2.88	0.32 to 25.79
Likely pathogenic															
1	70	M	ADC	No	Yes	*NTRK1*	IVS851-33T>A	LP	HCPS	0.0000345 (8/231,854)	28.26	5.999 to 133.2	0.00047 (8/16,924)	2.063	0.26 to 16.51
2	66	M	ADC	No	No	*NTRK1*	IVS851-33T>A	LP	HCPS	0.0000345 (8/231,854)	28.26	5.999 to 133.2	0.00047 (8/16,924)	2.063	0.26 to 16.51
3	63	M	ADC	Yes	Yes	*NTRK1*	IVS1806-2A>G	LP	Not reported	N/A	N/A	N/A	N/A	N/A	N/A
4	70	F	ADC	No	No	*NTRK1*	IVS1354+1G>T	LP	Only reported in normal individual	0.0000163 (4/246,148)	28.25	3.156 to 252.9	0.00023 (4/17,248)	4.21	0.47 to 37.66
5	45	M	SCC	No	Yes	*EXT2*	p.W606X	LP	Only reported in normal individual	0.0000323 (1/30,974)	14.13	1.766 to 113.0	N/A	N/A	N/A
6	37	M	ADC	Yes	Yes	*EXT2*	IVS1762-1G>A	LP	Not reported	N/A	N/A	N/A	N/A	N/A	N/A
7	62	M	ADC	Yes	Yes	*EXT2*	p.T507fs	LP	Not reported	N/A	N/A	N/A	N/A	N/A	N/A
8						*BRIP1 (homozygous)*	p.M1V	LP	Neoplasm of ovary; Fanconi anemia; HCPS	0.0000163 (4/245,960)	28.25	3.156 to 252.9	0.00023 (4/17,228)	4.2	0.47 to 37.62
9	94	M	ADC	Yes	Yes	*EXT2*	p.T642fs	LP	Not reported	N/A	N/A	N/A	N/A	N/A	N/A
10						*NBN*	p.N85fs	LP	Not reported	N/A	N/A	N/A	N/A	N/A	N/A
11	60	F	ADC	No	No	*PALB2*	p.N280fs	LP	Not reported	N/A	N/A	N/A	N/A	N/A	N/A
12	52	M	SCC	Yes	No	*PALB2*	p.P117fs	LP	Not reported	N/A	N/A	N/A	N/A	N/A	N/A
13	41	M	ADC	No	Yes	*PALB2*	p.Q921fs	LP	HCPS	N/A	N/A	N/A	N/A	N/A	N/A
14	60	M	SCC	No	Yes	*BRIP1*	p.T997fs	LP	Not reported	0.0000325 (8/245,824)	14.13	1.766 to 113.0	0.000058 (1/17,240)	16.82	1.05 to 269.08
15	46	F	ADC	No	No	*BRIP1*	p.M1V	LP	Not reported	N/A	N/A	N/A	0.00023 (4/17,228)	4.2	0.47 to 37.62
16	51	F	ADC	Yes	No	*SDHA*	p.R589W	LP	HCPS; paragangliomas	0.0000122 (3/245,836)	37.67	3.917 to 362.3	N/A	N/A	N/A

Table 2 (*continued*)

Table 2 (*continued*)

Number	Age	Gender	Cancer type	Family history	Smoking history	*Gene*	Protein change	Annotation	Association with diseases	General population*			East Asian*		
										Allele frequency	OR	95% CI	Allele frequency	OR	95% CI
17	54	F	ADC	No	Yes	*SDHA*	p.M1V	LP	Paragangliomas; Mitochondrial complex II deficiency; HCPS	0.00000857 (1/116,732)	56.5	5.122 to 623.4	N/A	N/A	N/A
18	66	M	ADC	No	Yes	*RAD50*	p.L719fs	LP	HCPS	0.000136 (32/235,016)	3.424	0.4681 to 25.05	0.00012 (2/16,510)	8.05	0.73 to 88.88
19	67	M	ADC	No	Yes	*RAD50*	p.E115X	LP	Not reported	N/A	N/A	N/A	N/A	N/A	N/A
20	28	M	ADC	Yes		*MLH3*	p.E931fs	LP	Only reported in normal individual	0.0000081 (2/246,100)	56.5	5.122 to 623.4	N/A	N/A	N/A
21	61	M	ADC	No	Yes	*MLH3*	IVS4243-1G>A	LP	Not reported	N/A	N/A	N/A	N/A	N/A	N/A
22	58	F	SCC	No	No	*BRCA1*	IVS5332+1G>-	LP	Familial cancer of breast	N/A	N/A	N/A	N/A	N/A	N/A
23	52	F	ADC	No	No	*BRCA1*	p.I1824fs	LP	HCPS; HBOC	N/A	N/A	N/A	N/A	N/A	N/A
24	48	F	ADC	Yes	Yes	*BRCA2*	p.N1055fs	LP	Not reported	N/A	N/A	N/A	N/A	N/A	N/A
25	64	M	ADC	No	Yes	*MUTYH*	IVS1477-1G>A	LP	MYH-associated polyposis	N/A	N/A	N/A	N/A	N/A	N/A
26	72	M	ADC	No	Yes	*TSC2*	IVS3815-1G>A	LP	Not reported	N/A	N/A	N/A	N/A	N/A	N/A
27	65	F	ADC	No	No	*NF1*	p.R1456_ F1457delinsRX	LP	Not reported	N/A	N/A	N/A	N/A	N/A	N/A
28	87	M	ADC	No	Yes	*RAD51D*	p.A210fs	LP	Not reported	N/A	N/A	N/A	N/A	N/A	N/A
29	70	M	ADC	No	Yes	*BLM*	IVS98+1->T	LP	Only reported in normal individual gnomAD exomes	0.00000444 (1/225,466)	113	7.066 to 1,807	N/A	N/A	N/A
30	77	F	ADC	Yes	No	*CHEK2*	IVS1096-1G>C	LP	HCPS; Familial cancer of breast	N/A	N/A	N/A	N/A	N/A	N/A
31	80	M	ADC	No	No	*MRE11A*	p.K105fs	LP	Not reported	N/A	N/A	N/A	N/A	N/A	N/A
32	60	M	ADC	No	Yes	*ATM*	IVS331+5G>A	LP	Ataxia-telangiectasia syndrome; HCPS	0.00000409 (1/244,414)	113	7.066 to 1,807	N/A	N/A	N/A
33	62	M	ADC	No	No	*SDHB*	p.L87X	LP	Hereditary Paraganglioma-Pheochromocytoma Syndromes	N/A	N/A	N/A	N/A	N/A	N/A
34	70	F	ADC	Yes	No	*PMS2*	IVS2175-2A>G	LP	Not reported	N/A	N/A	N/A	N/A	N/A	N/A
35	64	M	ADC	Yes	Yes	*POLE*	p.S1204fs	LP	Not reported	N/A	N/A	N/A	N/A	N/A	N/A
36	29	M	ADC	No	Yes	*TP53*	p.R181H	LP	LFS	0.0000122 (3/246,118)	37.67	3.917 to 362.3	N/A	N/A	N/A
Overall										0.0004954	15.86	9.529 to 26.38	0.00181	3.8	0.47 to 30.96

*, data from gnomAD database. OR, odds ratio; M, male; F, female; ADC, adenocarcinoma; SCC, squamous cell carcinoma; LC, large cell carcinoma; LP, likely pathogenic; MYH, MUTYH; HBOC, hereditary breast and ovarian cancer; PC, prostate cancer; HCPS, hereditary cancer predisposition syndrome; LFS, Li-Fraumeni Syndrome; CI, confidence interval.

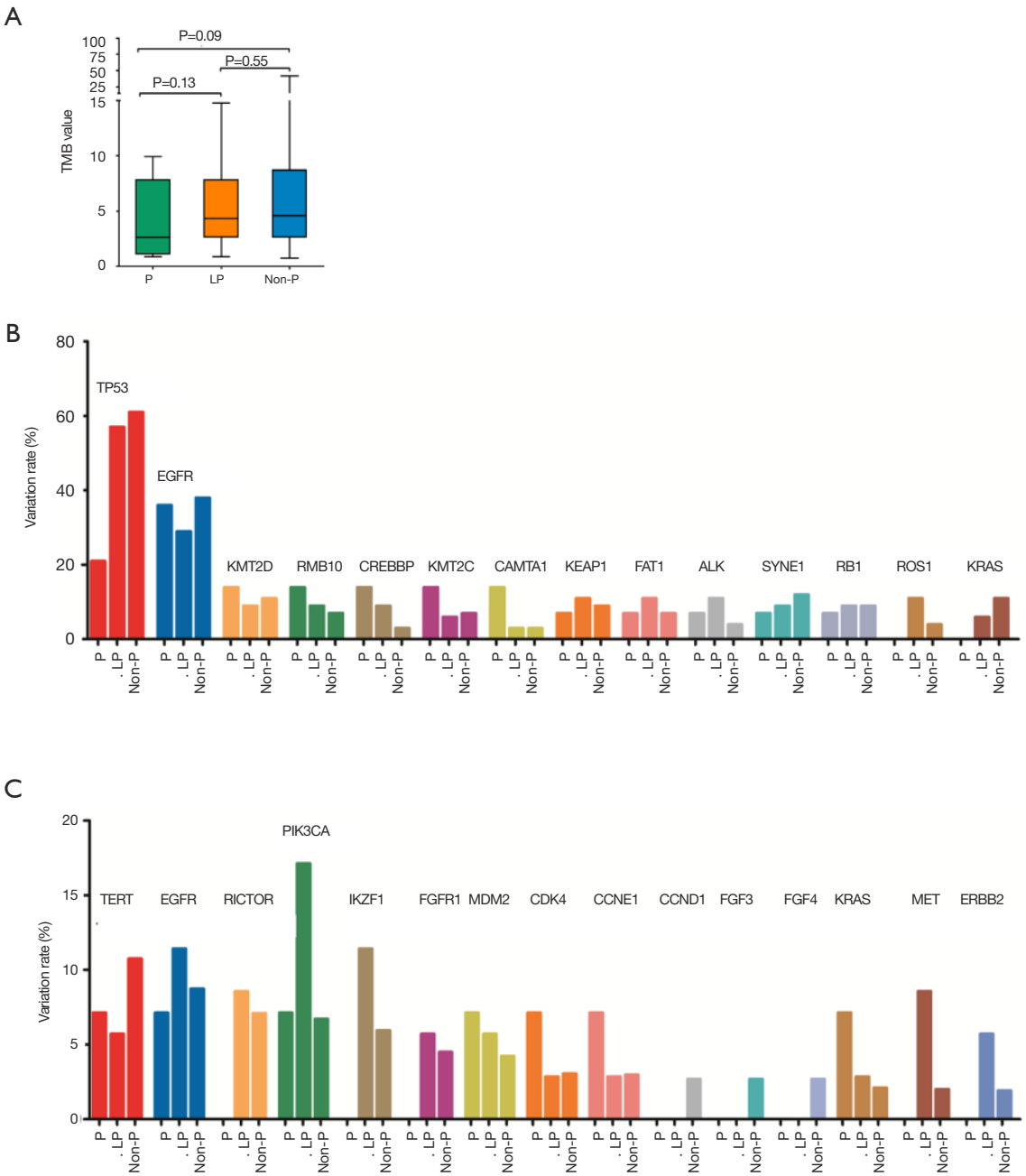

Figure 2 The TMB and the gene somatic variation rate for all patients in this study. (A) Comparison of the TMB from nonsynonymous somatic mutations of the P, LP and the Non-P groups. (B) Comparison of the variation rate (mutational frequency) for main genes with somatic SNV and INDEL mutations for P, LP and Non-P group. (C) Comparison of the variation rate (mutational frequency) for main genes with copy number variations (CNVs) for P, LP and Non-P group. TMB, tumor mutation burden; P, pathogenic; LP, likely pathogenic; SNV, single nucleotide variation; INDEL, insertion and deletion.

processes among the P, LP and Non-P groups (*Figure S3*).

Discussion

Our study provided the first set of evidence on the correlation between the hereditary tumor-related germline mutations and the risk of lung cancer in Chinese population. We found that *BRCA2* accounted for the top pathogenic mutations (6/14) in Chinese lung cancer patients, followed by *CHEK2* (3/14) and *ATM* (2/14). Pathogenic mutations were mainly frameshift and nonsense, indicating that germline mutations causing large fragment alterations were the main types in Chinese lung cancer patients. In addition, the functions of *BRCA2*, *CHEK2*, *ATM*, *BLM*, *EPCAM* and *RAD50* are mainly related to DNA repair and cell cycle regulation, suggesting that the germline mutations of these genes may cause dysregulation of DNA repair and cell cycle and be one genetic risk factor for the development of lung cancer. In the LP group, there were also many splicing mutations in addition to frameshift mutations, indicating that the influence of non-coding splicing sites on protein function cannot be ignored. In this study, the somatic mutations in patients with pathogenic or likely pathogenic germline mutations showed some interesting features. The trend of lower TMB in the pathogenic group indicated the somatic mutations in patients with pathogenic germline variations may be more focused on key driver genes and key pathways, while the somatic mutations in patients without pathogenic germline variations may be more sporadic. Therefore, patients with pathogenic germline mutations may be more likely to develop aberrancies in key driver genes and key pathways, leading to increased risk of lung cancer. It is interesting to find that the affected pathways in patients with or without pathogenic germline mutations were similar, suggesting that the carcinogenesis mechanism of pathogenic group would be consistent with that from the non-pathogenic groups, i.e., the sporadic lung cancer patients, in which cigarette smoke-induced genotoxic damage or other environmental hazards are main causes of malignant transformation (1,2). This indicates that the influence of pathogenic germline mutations mimics the effects of the smoke and environmental factors. One possible explanation for this phenomenon is that the affected germline mutations happen to be those mainly relating to DNA damage and repair. Another possibility is that the presence of pathogenic germline mutations possibly increased the susceptibility to these risk factors and individuals are more likely to develop mutations relating to these factors.

Germline mutations that have been reported in previous studies have focused primarily on *EGFR* mutations (9,14), mainly because the use of TKI is closely related to *EGFR* mutations. However, *EGFR* mutations are not conventional germline mutations related to hereditary cancers, and population studies have reported that *EGFR* germline mutations were not common in lung cancer [prevalence of 0.13% (12/9,091)] (9), although *EGFR* germline mutations at multiple sites have been reported (14). Its incidence is even lower in general population with no lung cancer. Therefore, the significance of large-scale screening for *EGFR* germline mutations in general population is not clear due to its low incidence. However, lung cancer patients and their relatives may benefit from the screening of *EGFR* germline mutations. In contrast, the *BRCA2* germline mutations in this study exhibited a higher overall incidence of 0.68% (7/1,026) than *EGFR* germline mutations, and therefore may be of more significance in clinical guidance and risk assessment for patients and their families. In addition to *EGFR*, previous studies have also found that germline susceptibility loci of multiple genes in lung cancer patients were associated with lung cancer risk, including *ATM*, *BRCA2*, *CHEK2*, *EGFR*, *PARK2*, *TERT*, *TP53* and *YAP1* (5), *BRCA1*, *BRCA2*, *ERCC4*, *EXT1*, *HNF1A*, *PTCH1*, *SMARCB1*, *TP53* (16), *BRCA2* p.Lys3326X, *CHEK2* p.Ile157Thr, *TP63*, rs13314271 (12), *ARHGEF5*, *ANKRD20A2*, *ZNF595*, *ZNF812*, *MYO18B* (17), and *BRCA2* K3326X, *LTB* p.Leu87Phe, *P3H2* p.Gln185His, *DAAM2* p.Asp762Gly (18). Among these studies, Parry and colleagues (5) performed a population-based study with TCGA database and found that the *ATM* gene accounted for 50% of lung cancer germline mutations, followed by *TP53*, *BRCA2*, *EGFR*, and *PARK2*. This was quite different from the prevalence of germline mutations found in this study, which may be due to the selection of different populations and different target genes. In another recent population-based study, *BRCA2* germline mutations ranked the highest in all germline mutations tested, with a detection rate of 0.38% (17/4,459) (3), which was similar to the finding of this study. It should be noted that the above two population-based studies included only 8 or 16 germline genes (3,5). In contrast, our study containing 58 germline genes is therefore more comprehensive and representative than the above studies in reflecting the profile of germline mutations in lung cancer patients.

We found that the somatic average mutation rate varied with different germline mutations. For example, the mutation rate of *TP53* in the P group was significantly lower than that of the other two groups, while no such difference in the mutation rate of *EGFR* was observed, which indicates differential effects of pathogenic germline mutations on somatic driver genes. Interestingly, the CNV mutation rate of *PIK3CA* and *MET* of the LP group were significantly higher than that of the Non-P group, suggesting that the somatic amplification of these two genes may be more prominent than other genes when likely-pathogenic germline mutations were present. These observations indicate that the activation of PI3K/AKT and *MET* pathways may be characteristic in CNV-related alterations. We therefore speculate that patients with DDR-related germline driver gene mutations (such as *BRCA2*) may be affected by both germline and somatic driver gene mutations, suggesting a different mechanism and a higher risk compared with those without germline driver gene mutations.

The frequency of mutations queried in the GnomAD database represents the frequency of a certain mutation site in the general population. Since most pathogenic or likely pathogenic germline mutations exhibited very low incidence in the general population, the frequency in the database may have certain randomness and may not accurately represent the true frequency in the population. Similarly, the frequency of pathogenic or likely pathogenic germline mutations found in this study was also affected by randomness, and the OR value for a single mutation site may not accurately represent the true frequency in lung cancer population. However, when we pooled all the germline mutations together, the overall mutation frequency was statistically significant, and the overall OR of the P or LP group was comparable with that from the gnomAD database. In this study, the OR of the P group and the LP group suggested that the germline mutations were risk factors for lung cancer. This was also observed in previous studies on lung cancer germline mutations. For example, Parry *et al.* reported that the overall OR was 66 from 14 germline mutations including *ATM* and *TP53* (5), and Wang *et al.* reported that the OR for *BRCA2* L3326X was 2.47 (12). It is not easy to define the OR value of a certain locus of a certain gene, as the sample size for lung cancer patients and general population need to be large enough for the value to be accurately calculated. Therefore, the report from Parry *et al.* and our study estimated the overall OR of pooled germline mutations to assess the risk of lung cancer in population (5). In any case, our study and previous studies have demonstrated that pathogenic germline mutations are a risk factor for lung cancer.

It is not uncommon to see lung cancer patients with a familial history. We identified 26.74% of lung cancer patients in this study who had at least one immediate family member with lung cancer. However, unlike other hereditary tumors, most of these lung cancer patients did not had clear pathogenic germline mutations, and the germline mutations or susceptibility loci of the families reported in the previous cases varied greatly, and no clear genetic abnormalities or aggregation has been identified (17,19,20). Therefore, it can be speculated that the occurrence of familial lung cancer may be due to a combination of multiple genetic factors and environmental factors. Elucidation of these factors may require comprehensive family study including typical familial lung cancer patients and their relatives to collect enough data for correlation analysis. In contrast, familial risk is relatively clear for lung cancer patients with clear pathogenic or likely pathogenic germline mutations, therefore, screening for germline mutations in lung cancer patients can help their relatives to understand the risk of the disease and prevent it in advance. Meanwhile, due to the high proportion of *BRCA2* pathogenic germline mutations in Chinese population, PARP inhibitors may be applied for this specific population in addition to traditional chemoradiotherapy, targeted therapy or immunotherapy, and relevant clinical trials have also shown positive results (21). Future studies on germline mutations in lung cancer patients should focus on the identification of genetic factors of familial lung cancer and the elucidation of pathogenicity of germline mutations, which will help more patients and their relatives with the prevention and treatment of lung cancer.

Acknowledgments

Funding: This study was supported by the Special Funds for Strategic Emerging Industry Development of Shenzhen (grant number 20170922151538732), and the Science and Technology Project of Shenzhen (grant number JSGG20180703164202084). All funders did not participate in the study design, study implementation, data collection, data analysis, data interpretation and manuscript writing of the study.

Footnote

Conflicts of Interest: All authors have completed the

ICMJE uniform disclosure form (available at http://dx.doi.org/10.21037/tlcr-19-403). Mengyuan L, XL, PS, TH, YZ, Ming L, LL, Yaru C, YZ, GL, JY and SC report non-financial support from HaploX Biotechnology outside the submitted work. LS reports grants from The Special Funds for Strategic Emerging Industry Development of Shenzhen and The Science and Technology Project of Shenzhen, non-financial support and other from HaploX Biotechnology Co., Ltd. outside the submitted work. The other authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the ethics board of the Chinese PLA General Hospital (approval number: S2018-081-02) and individual consent for this retrospective analysis was waived.

References

1. Alberg AJ, Brock MV, Ford JG, et al. Epidemiology of lung cancer: Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. Chest 2013;143:e1S-e29S.
2. Lissowska J, Foretova L, Dabek J, et al. Family history and lung cancer risk: international multicentre case-control study in Eastern and Central Europe and meta-analyses. Cancer Causes Control 2010;21:1091-104.
3. Slavin TP, Banks KC, Chudova D, et al. Identification of Incidental Germline Mutations in Patients With Advanced Solid Tumors Who Underwent Cell-Free Circulating Tumor DNA Sequencing. J Clin Oncol 2018. doi: 10.1200/JCO.18.00328
4. Hu Y, Alden RS, Odegaard JI, et al. Discrimination of Germline EGFR T790M Mutations in Plasma Cell-Free DNA Allows Study of Prevalence Across 31,414 Cancer Patients. Clin Cancer Res 2017;23:7351-9.
5. Parry EM, Gable DL, Stanley SE, et al. Germline Mutations in DNA Repair Genes in Lung Adenocarcinoma. J Thorac Oncol 2017;12:1673-8.
6. Yamamoto H, Yatabe Y, Toyooka S. Inherited lung cancer syndromes targeting never smokers. Transl Lung Cancer Res 2018;7:498-504.
7. Ikeda K, Nomori H, Mori T, et al. Novel germline mutation: EGFR V843I in patient with multiple lung adenocarcinomas and family members with lung cancer. Ann Thorac Surg 2008;85:1430-2.
8. Prim N, Legrain M, Guerin E, et al. Germline exon 21 EGFR mutations, V843I and P848L, in non-small cell lung cancer patients. Eur Respir Rev 2014;23:390-2.
9. Lu S, Yu Y, Li Z, et al. EGFR and ERBB2 Germline Mutations in Chinese Lung Cancer Patients and Their Roles in Genetic Susceptibility to Cancer. J Thorac Oncol 2019;14:732-6.
10. Xiong D, Wang Y, Kupert E, et al. A recurrent mutation in PARK2 is associated with familial lung cancer. Am J Hum Genet 2015;96:301-8.
11. Chen HY, Yu SL, Ho BC, et al. R331W Missense Mutation of Oncogene YAP1 Is a Germline Risk Allele for Lung Adenocarcinoma With Medical Actionability. J Clin Oncol 2015;33:2303-10.
12. Wang Y, McKay JD, Rafnar T, et al. Rare variants of large effect in BRCA2 and CHEK2 affect risk of lung cancer. Nat Genet 2014;46:736-41.
13. Landi MT, Chatterjee N, Yu K, et al. A genome-wide association study of lung cancer identifies a region of chromosome 5p15 associated with risk for adenocarcinoma. Am J Hum Genet 2009;85:679-91.
14. Shukuya T, Patel S, Shane-Carson K, et al. Lung Cancer Patients with Germline Mutations Detected by Next-Generation Sequencing and/or Liquid Biopsy. J Thorac Oncol 2018;13:e17-e19.
15. Richards S, Aziz N, Bale S, 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. Genet Med 2015;17:405-24.
16. Donner I, Katainen R, Sipilä LJ, et al. Germline mutations in young non-smoking women with lung adenocarcinoma. Lung Cancer 2018;122:76-82.
17. Kanwal M, Ding XJ, Ma ZH, et al. Characterization of germline mutations in familial lung cancer from the

Chinese population. Gene 2018;641:94-104.
18. Liu Y, Lusk CM, Cho MH, et al. Rare Variants in Known Susceptibility Loci and Their Contribution to Risk of Lung Cancer. J Thorac Oncol 2018;13:1483-95.
19. Hsu KH, Tseng JS, Wang CL, et al. Higher frequency but random distribution of EGFR mutation subtypes in familial lung cancer patients. Oncotarget 2016;7:53299-308.
20. Shukuya T, Takahashi K. Germline mutations in lung cancer. Respir Investig 2019;57:201-6.
21. Gadgeel SM. Targeted Therapy and Immune Therapy for Small Cell Lung Cancer. Curr Treat Options Oncol 2018;19:53.

Cite this article as: Liu M, Liu X, Suo P, Gong Y, Qu B, Peng X, Xiao W, Li Y, Chen Y, Zeng Z, Lu Y, Huang T, Zhao Y, Liu M, Li L, Chen Y, Zhou Y, Liu G, Yao J, Chen S, Song L. The contribution of hereditary cancer-related germline mutations to lung cancer susceptibility. Transl Lung Cancer Res 2020;9(3):646-658. doi: 10.21037/tlcr-19-403

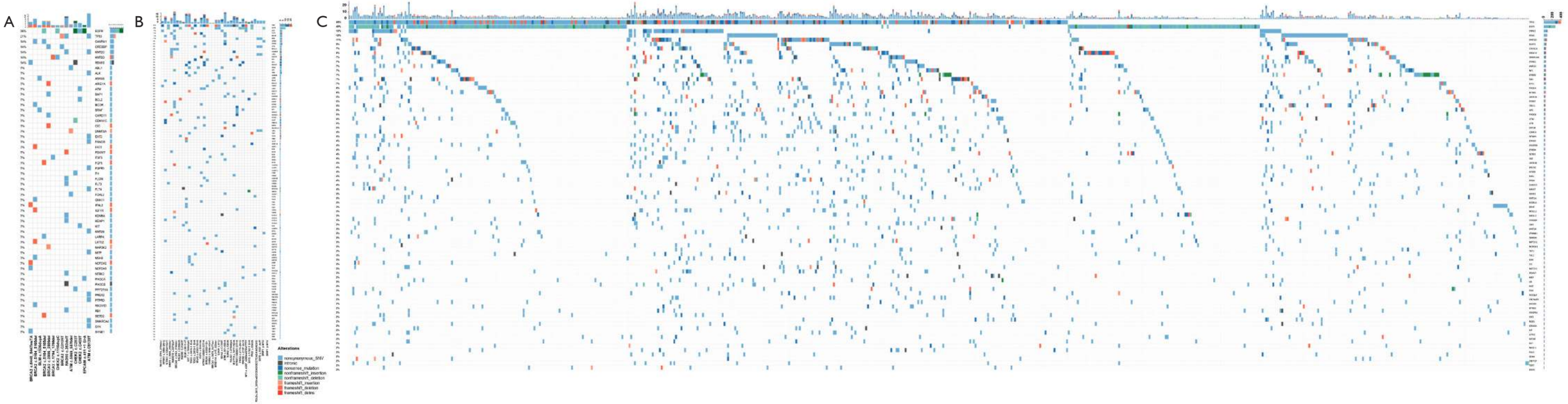

Figure S1 Full SNV and INDEL somatic mutation spectrum for patients with pathogenic (A), likely pathogenic (B) or non-pathogenic (C) germline mutations. Somatic mutation spectrum for 14 patients with pathogenic germline mutations is shown in (A). Somatic mutation spectrum for 35 patients with likely pathogenic germline mutations is shown in (B). Somatic mutation spectrum for 1041 patients with non-pathogenic germline mutations is shown in (C). Details of germline mutations are labeled beneath the figures for (A,B), and somatic mutated genes are listed in the order of variation rate to the right of the figures. The rightest bars represent the overall number of mutations for each gene. Percentage to the left of the figures represents variation rate for each gene. Y-axis above the figures represents the number of somatic mutations detected for each patient. Colors represent mutation types as indicated by the figure legend.

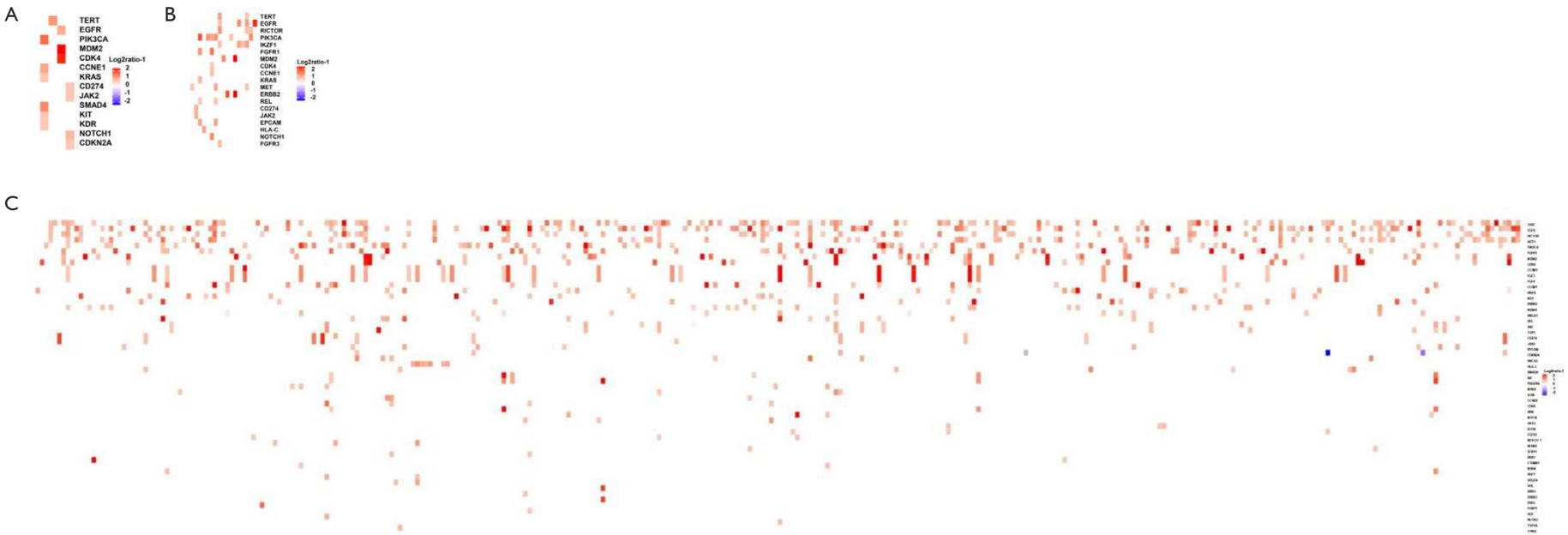

Figure S2 Full CNV somatic mutation spectrum for patients with pathogenic (A), likely pathogenic (B) or non-pathogenic (C) germline mutations. Gene names with CNVs are shown to the right of the figures. Each column represents one patient, and the corresponding germline mutations are labeled beneath the figures. Colors represent the copy number for each gene, which is visualized based on the calculation of log2ratio-1. Only those patients with CNVs are shown in this figure. CNV, copy number variation.

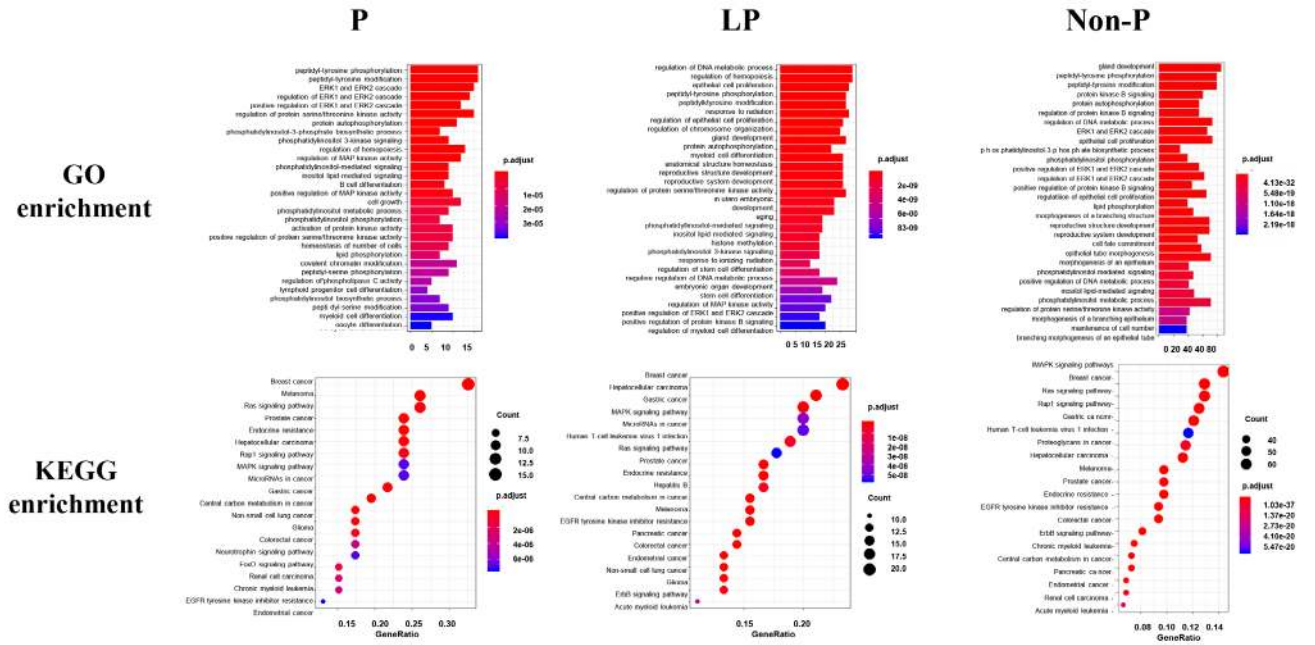

Figure S3 Results of GO and KEGG enrichment analysis for P, LP and the Non-P groups. The upper panel shows the results of GO enrichment and the lower panel shows the results of KEGG enrichment analysis, respectively. In GO enrichment panel, color represents the degree of significance (adjusted P value) as labeled, and bars represent the number of genes with mutations involved for each function or pathway. In KEGG enrichment panel, color represents the degree of significance (adjusted P value) as labeled, and the size of dots represents the ratio of genes in which the mutations were found for each function or pathway, and bigger dots represent higher ratio. GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; P, pathogenic; LP, likely pathogenic.

Table S1 The gene list for the 58-gene panel used for germline mutation detection in this study

APC	ATM	AXIN2	BRCA1	BRCA2	BARD1	BLM	BMPR1A	BRIP1	CDC73
CDH1	CDK4	CDKN1B	CDKN2A	CHEK2	EPCAM	EXT1	EXT2	FH	FLCN
GREM1	MAX	MEN1	MET	MITF	MLH1	MLH3	MRE11A	MSH2	MSH6
MUTYH	NBN	NF1	NF2	NTRK1	PALB2	PMS1	PMS2	POLD1	POLE
PTEN	RAD50	RAD51C	RAD51D	RB1	RET	SDHA	SDHAF2	SDHB	SDHC
SDHC	SMAD4	STK11	TMEM127	TP53	TSC1	TSC2	VHL		

Table S2 The gene list of the 605-gene panel used for somatic variation sequencing in this study

ABCB1	BCL2L11	CDKN2C	EGF	FGF23	GSR	KDM5C	MLH3	PARD3B	PTPN11	SEMA3C	TERC	ZBTB16
ABCC1	BCL6	CEBPA	EGFR	FGF3	GSTA1	KDM6A	MPL	PARK2	PTPRD	SETBP1	TERT	ZNF367
ABCC11	BCOR	CFD	EIF3A	FGF4	GSTM3	KDR	MRE11A	PARP1	PTPRT	SETD2	TET1	ZNF423
ABCC2	BCORL1	CFH	ELAC2	FGF5	GSTP1	KEAP1	MSH2	PAX3	PZP	SETD7	TET2	ZNF717
ABCC4	BLM	CHD4	ENOSF1	FGF6	H19	KIF1B	MSH3	PAX5	RAC1	SF3B1	TFE3	ZNF750
ABCC5	BMPR1A	CHEK1	EP300	FGF7	H3F3A	KIT	MSH6	PAX7	RAD21	SH2B3	TGFB1	
ABCG1	BRAF	CHEK2	EPCAM	FGF8	HBV	KLF4	MST1R	PAX8	RAD50	SHMT1	TGFBR2	
ABCG2	BRCA1	CIC	EPHA2	FGF9	HCV	KLLN	MTHFR	PBRM1	RAD51	SHOX	TMEM127	
ABL1	BRCA2	CMPK1	EPHA3	FGFR1	HDAC2	KMT2A	MTOR	PCBP1	RAD51B	SLC15A2	TMPRSS2	
ACSS2	BRD2	CNTNAP5	EPHA5	FGFR2	HFE2	KMT2B	MTUS1	PDCD1	RAD51C	SLC19A1	TNF	
ACTL6A	BRD4	CREBBP	EPHA7	FGFR3	HGF	KMT2C	MUTYH	PDCD1LG2	RAD51D	SLC22A1	TNFAIP3	
ACVR1	BRIP1	CRKL	EPHB1	FGFR4	HIF1A	KMT2D	MYC	PDGFB	RAD52	SLC22A16	TNFRSF11B	
ADCY2	BTK	CRLF2	EPHX1	FH	HLA-G	KRAS	MYCL	PDGFRA	RAD54L	SLC22A2	TNFRSF14	
ADH1B	BUB1	CSF1R	ERBB2	FLCN	HMGA2	KRT14	MYCN	PDGFRB	RAF1	SLC22A4	TNFRSF19	
ADH1C	C10orf11	CSF3R	ERBB3	FLT1	HMGCR	KRT15	MYD88	PDPK1	RARA	SLC22A5	TNFSF11	
AKR1C3	C18orf56,TYMS	CSMD3	ERBB4	FLT3	HNF1A	KRT5	MYOD1	PGR	RB1	SLC28A1	TNFSF8	
AKT1	C8orf34	CTCF	ERCC1	FLT4	HNF1B	LARP4	NAB2	PIGB	RBFOX1	SLC28A2	TOP1	
AKT2	CACNA1C	CTNNB1	ERCC2	FNTB	HOTAIR	LATS1	NAT2	PIK3CA	RBM10	SLC29A1	TP53	
AKT3	CADM2	CUL3	ERCC3	FOLR3	HOXB13	LATS2	NBN	PIK3CB	RECK	SLC31A1	TPMT	
ALDH2	CALR	CXCR4	ERCC4	FOXA1	HPV	LBR	NCOA1	PIK3CD	RECQL	SLCO1B1	TRAF1	
ALK	CAMTA1	CXXC4	ERCC5	FOXK2	HRAS	LGR5	NCOA3	PIK3CG	RECQL4	SLCO1B3	TSC1	
ALOX12	CAPN2	CYLD	EREG	FOXL2	HSD17B3	LIG3	NF1	PIK3R1	REL	SLX4	TSC2	
AMER1	CARD11	CYP19A1	ERG	FOXM1	HSD3B2	LMO1	NF2	PIK3R2	RET	SMAD2	TSHR	
ANXA5	CASP7	CYP1A1	ERRFI1	FOXP1	HSP90AA1	LRIG3	NFE2L2	PIM1	RGS5	SMAD3	TSPAN31	
APC	CASP8	CYP1A2	ESR1	FOXP2	HSPA5	LRP1B	NFKBIA	PLAUR	RHBDF2	SMAD4	TUBB1	
APLF	CBFB	CYP1B1	ESR2	FUBP1	HTRA1	LRP2	NKX2-1	PLCG2	RHEB	SMARCA4	TYMS	
AR	CBL	CYP2B6	ETV1	FUS	IDH1	LYN	NOS3	PLIN2	RHOA	SMARCB1	U2AF1	
ARAF	CBLB	CYP2C19	ETV4	GAB2	IDH2	MAD1L1	NOTCH1	PMS1	RICTOR	SMO	UBE2I	
AREG	CBR1	CYP2C8	ETV6	GALNT14	IFNL2	MALAT1	NOTCH2	PMS2	RIF1	SOCS1	UGT1A	
ARID1A	CBR3	CYP2D6	EWSR1	GATA1	IFNLR1	MAP2K1	NOTCH3	POLD1	RILP	SOCS6	UGT1A1	
ARID1B	CCL18	CYP2E1	EXT1	GATA2	IGF1R	MAP2K2	NOVA1	POLE	RIT1	SOD2	UGT1A4	
ARID2	CCND1	CYP3A4	EXT2	GATA3	IGF2	MAP2K4	NPM1	POR	RNASEL	SOX10	UGT1A6	
ARMS2	CCND2	CYP3A5	EZH2	GATA6	IGFBP3	MAP3K1	NQO1	PPIB	RNF43	SOX2	UGT1A9	
ASNS	CCND3	DAXX	FAM175A	GEMIN6	IKBKE	MAP4K4	NQO2	PPP2R1A	ROBO2	SOX9	VEGFA	
ASPH	CCNE1	DDIT3	FAM46C	GEN1	IKZF1	MAPK1	NR1I2	PPP2R2A	ROS1	SPEN	VEGFC	
ASXL1	CD274	DDR2	FANCA	GGH	IL13	MAPK3	NR4A3	PRDM1	RPS6KB1	SPINK1	VHL	
ATM	CD79A	DDX3X	FANCB	GK5	IL16	MAPKBP1	NRAS	PRDX4	RPTOR	SPOP	WAS	
ATP7B	CD79B	DDX51	FANCC	GLI1	IL1B	MAX	NRG1	PREX2	RRAS2	SRC	WIF1	
ATR	CDA	DHFR	FANCG	GLIPR1	IL23R	MCL1	NSD1	PRKACA	RRM1	SRD5A2	WNT5B	
ATRX	CDC73	DICER1	FANCI	GLRX	IL7R	MDC1	NT5C2	PRKACB	RSF1	SRSF2	WRN	
AURKA	CDH1	DNMT3A	FANCL	GMEB1	INHBA	MDM2	NTRK1	PRKAR1A	RUNX1	SS18	WT1	
AURKB	CDK12	DOT1L	FAT1	GNA11	INPP4B	MDM4	NTRK2	PRKCI	SBDS	STAG2	XBP1	
AXIN1	CDK4	DPYD	FBN3	GNAQ	IRF4	MED12	NTRK3	PRSS1	SCN10A	STAT3	XPA	
AXIN2	CDK6	DSCAM	FBXW7	GNAS	IRS2	MEF2B	NUP93	PSME2	SDHA	STK11	XPC	
AXL	CDK8	DYNC2H1	FCGR2A	GPER1	JAK1	MEN1	NUTM1	PTCH1	SDHAF2	SUFU	XPO1	
B2M	CDKN1A	E2F7	FCGR3A	GPRIN2	JAK2	MET	OPRM1	PTEN	SDHB	SULT1A1	XRCC1	
BAP1	CDKN1B	EBV	FGF1	GPX5	JAK3	MGAT4A	OTOS	PTGER4	SDHC	SUZ12	XRCC3	
BARD1	CDKN1C	ECT2L	FGF10	GREM1	JUN	MITF	PAK1	PTGES	SDHD	SYK	XRCC4	
BCL2	CDKN2A	EDN1	FGF19	GRIN2A	KCNJ5	MKI67	PALB2	PTGS2	SELE	SYNE1	YAP1	
BCL2L1	CDKN2B	EED	FGF2	GSK3B	KDM5A	MLH1	PALLD	PTN	SELL	TBX3	YES1	

Table S3 The summary of clinicopathological and history information for NSCLC patients with distinct germline mutation pathogenicity (P and LP groups combined)

Clinicopathological factors	Subgroups	Total (N=1,026)		P/LP (N=48)		Non-pathogenic (N=978)		P
		n	%	n	%	n	%	
NSCLC	Adenocarcinoma	792	77.19	38	79.17	754	77.10	0.48
	Squamous	222	21.64	9	18.75	213	21.78	
	Large cell	6	0.58	1	2.08	5	0.51	
	Adenosquamous	6	0.58	0	0.00	6	0.61	
Age, year	<40	47	4.58	2	4.17	45	4.60	0.89
	≥40	979	95.42	46	95.83	933	95.40	
	<50	181	17.64	9	18.75	172	17.59	0.84
	≥50	845	82.36	39	81.25	806	82.41	
	<60	473	46.10	22	45.83	451	46.11	0.97
	≥60	553	53.90	26	54.17	527	53.89	
	<70	820	79.92	41	85.42	779	79.65	0.33
	≥70	206	20.08	7	14.58	199	20.35	
Sex	Male	594	57.89	30	62.50	564	57.67	0.51
	Female	432	42.11	18	37.50	414	42.33	
Stage	I–IIIA	568	55.36	20	41.67	548	56.03	0.051
	IIIB–IV	458	44.64	28	58.33	430	43.97	
Smoking history	Yes	584	56.92	26	54.17	558	57.06	0.69
	No	442	43.08	22	45.83	420	42.94	
History of prior malignancy	Yes	40	3.90	7	14.58	36	3.68	0.0002
	No	986	96.10	41	85.42	942	96.32	
Family history*	Yes	275	26.80	19	39.58	256	26.18	0.041
	No	751	73.20	29	60.42	722	73.82	

*, family history: the confirmed lung cancer patient has at least one immediate family member (first degree relatives) who had a history of lung cancer diagnosis.