OncoTargets and Therapy

Dovepress
open access to scientific and medical research

Open Access Full Text Article

ORIGINAL RESEARCH

Genomic Alteration Spectrum of Non-Small Cell Lung Cancer Patients in East-China Characterized by Tumor Tissue DNA and Cell-Free DNA

Jie Li[1,*], Siwen Chen[1,*], Hui Xue[2,*], Haoyi Wang[3,*], Tianwei Huang[4], Hongya Xie[5], Jiang He[6], Cai Ke[3], Zhaonan Yu[3], Bin Ni[4]

[1]Department of General Medical Ward, First Affiliated Hospital of Soochow University, Suzhou, People's Republic of China; [2]Department of Oncology Medical Ward, Hanzhong Central Hospital, Hanzhong, People's Republic of China; [3]Hangzhou D.A. Medical Laboratory, Hangzhou, People's Republic of China; [4]Department of Thoracic Surgery, First Affiliated Hospital of Soochow University, Suzhou, People's Republic of China; [5]Department of Thoracic Surgery, Suzhou Municipal Hospital, Suzhou, People's Republic of China; [6]Department of Thoracic Surgery, Suzhou Wuzhong People's Hospital, Suzhou, People's Republic of China

*These authors contributed equally to this work

Correspondence: Bin Ni, Department of Thoracic Surgery, First Affiliated Hospital of Soochow University, Suzhou, People's Republic of China, Tel +86 17774015977, Email bin_ni@163.com; Zhaonan Yu, Hangzhou D.A. Medical Laboratory, Hangzhou, People's Republic of China, Tel +86 15558078770, Email yuzhaonan870213@163.com

Introduction: From an oncologic perspective, genetic detection is becoming a frontline clinical test, used to identify actionable alterations for targeted therapy, monitor molecular clonal tumor evolution, indicate disease progression and prognosis, and predict medication efficacy and resistance. From analysis of both tumor tissue and cell-free DNA from a large cohort of non-small cell lung cancer patients in East-China, we characterized the full spectrum of genomic alterations.

Methods: The study comprised 3000 unpaired samples including 1351 tumor tissue DNA (tDNA) and 1649 cell-free circulating tumor DNA (cfDNA) samples, from which 67 cancer-related genes were sequenced and the genetic alteration profiles were depicted. Integrative molecular analyses identified the frequently mutated genes, uncovered co-occurring somatic alterations, described the distribution of hotspot variants, analyzed the frequency of variant alleles, and notably distinguished actionable, novel variants.

Results: The most commonly affected genes were EGFR, TP53, KRAS, CDKN2A, and PIK3CA in both tDNA and cfDNA samples. EGFR and CTNNB1, PIK3CA and PTEN, ERBB2 and SMO were found to have frequent co-occurring alterations in tDNA samples, while EGFR and SMO, KRAS and PDGFRA, PIK3CA and KDR were in cfDNA samples. A large number of primary druggable variants were identified in tDNA samples, while numerous drug-resistance variants, rare actionable variants, and non-EGFR actionable variants were detected in cfDNA samples. Novel variants were enriched in KDR, KIT, TP53, ABL1, FGFR1 in tDNA samples while the majority of novel variants were distributed in PDGFRA, EGFR, KIT, ROS1, BRCA2 in cfDNA samples. Variant allele frequency in tDNA samples was significantly ($P < 0.001$) higher than that in cfDNA samples.

Conclusion: The results revealed considerable differences in the alteration characteristics between the two kinds of specimens. To date, this study represents the largest real-world investigation of its kind, derived from the largest number of patients in East-China. It reinforced and expanded the mechanism of molecular analysis of neoplastic genetic profiles.

Keywords: real world study, non-small cell lung cancer patients in East-China, tumor tissue DNA, cell-free circulating tumor DNA, next-generation-sequencing, genetic alteration

Introduction

Lung cancer is the leading cause of cancer deaths worldwide and is responsible for 30% of all cancer deaths in China.[1] With the development of precision medicine and next-generation sequencing (NGS) technology, cancer patients have increasingly benefited from gene detection based targeted therapy. In addition to identifying actionable alterations for targeted therapy, gene detection can also monitor molecular clonal tumor evolution, indicate disease progression and

Received: 24 November 2021
Accepted: 27 April 2022
Published: 19 May 2022

prognosis, and predict drug efficacy and resistance. Marked research interest has been focused on genetic alteration and targeted drugs, particularly directed to non-small cell lung cancer (NSCLC), such that gene detection has become a frontline clinical test.[2,3] However, previous studies have indicated that NSCLC showed genetic disparities based on geography and ethnicity, and most of the studies that characterized the tumor genotyping of NSCLC, have been conducted in clinical drug trials.[4] Therefore, there is a critical need for real-world genomic alteration profiles for East-China Chinese NSCLC patients.

Sampling methods for gene detection testing consist of tissue biopsy and liquid biopsy. Although tumor tissue for genotyping is considered the gold standard, it is plagued with certain disadvantages, such as the need for invasive tumor sampling and inter- and intratumoral heterogeneity.[5] In contrast, liquid biopsy has evolved into either a supplement or even a preferred substitute for tissue biopsy due to its minimal invasiveness and impressive technical advances.[5–7]

The aim of the present cross-sectional study was to characterize the genetic testing of 3000 East-China Chinese patients diagnosed with NSCLC.

Materials and Methods

Patients and specimens. The study comprised 3000 NSCLC patients from East-China who had undergone genetic testing between January 2018 and May 2021, from whom tumor tissue (1351 patients) and peripheral blood (PB; 1649 patients) were obtained (Table 1). Tumor tissue specimens included 460 FRE, 64 FNA, 729 FFPE, and 98 HYTC. Tumor tissue and blood specimens were unpaired and collected from independent patient groups, including the First Affiliated Hospital of Soochow University (Suzhou, Jiangsu, China), Hanzhong Central Hospital, Suzhou Municipal Hospital (Hanzhong, Shanxi, China), and Suzhou Wuzhong People's Hospital (Suzhou, Jiangsu, China). There were no statistically significant differences regarding gender, age, smoking history or tumor type between patient groups providing tissue or blood specimens (Table 1).

All patients provided written informed consent prior to enrollment in the study. All procedures in this study were conducted in accordance with the Helsinki Declaration and approved by the Ethics Committee of the First Affiliated Hospital of Soochow University (approved No. 2018069) and the methods were carried out in accordance with the approved guidelines.

Table 1 Details of Patients Diagnosed with Non-Small Cell Lung Cancer in This Study

Characteristics	Tissues (1351 Cases)	Blood (1649 Cases)	*P*
	No. (%)	No. (%)	
Age			
>60	855 (63.3)	1085 (65.8)	*
≤60	496 (36.7)	564 (34.2)	*
Sex			
Female	632 (46.8)	712 (43.2)	*
Male	719 (53.2)	937 (56.8)	*
Smoking			
Non-smoking	1266 (93.7)	1510 (91.6)	*
Occasionally	23 (1.7)	19 (1.1)	**
Often	62 (4.6)	120 (7.3)	*
Tumor type			
Adenocarcinoma	1289 (95.4)	1552 (94.1)	*
Squamous carcinoma	58 (4.3)	84 (5.1)	**
Unknown	4 (0.3)	13 (0.8)	**
Tumor stage			
I	1020 (75.5)	44 (2.7)	
II	163 (12.1)	78 (4.7)	
III	109 (8.1)	145 (8.8)	
IV	59 (4.3)	1382 (83.8)	

Note: *p < 0.05; **p < 0.01.

Sampling requirements and DNA extraction. As previously described,[8] streck tubes were used to collect no less than 10 mL of peripheral blood for gene detection. The tubes were then turned upside down gently for blending at least 10 times and stored at 6–25°C until use (within 3 days). For FRE specimens, the total mass of each specimen was no less than 60mg, with the proportion of tumor cells no less than 70%, and the proportion of necrotic cells no more than 10%. No less than 3 needles were required for FNA specimens. The preservation time of FFPE specimens was not longer than half a year. All tissue samples were stored in DNA preservation tubes and handled within 3 days.

DNeasy Blood & Tissue kits (Qiagen, Hilden, Germany) were used for the extraction of cell free and tumor tissue DNA from blood sourced and tissue sourced specimens, respectively. Then, agarose gel electrophoresis was performed to test the degradation and protein/RNA pollution of extracted DNA. Concentration of extracted DNA was determined by Qubit dsDNA HS assay kit on Qubit Fluorometer according to the manufacturing protocol (Life Technologies, Carlsbad, CA, USA). DNA quality (OD 260/280) was measured by Nanodrop (Thermo Fisher Scientific, USA). The extracted DNA sample that had an OD ratio between 1.8 and 2.0 and a mass more than 100 ng (tDNA) and 10ng (cfDNA) was considered qualified and would be used for the subsequent library construction.

Library preparation and target-hybrid capturing. Library preparation was performed as per the manufacturer's instructions of KAPA Hyper prep kit (Illumina Co., Ltd, San Diego, USA). In short, the fragment DNA was orderly subjected to terminal repair, A-tailing adenylation, and ligation to indexed adapters. The constructed library was subsequently amplified by polymerase chain reaction (PCR) with 10 cycles and 5 cycles for cfDNA and tDNA samples, respectively. The targeted sequencing was an improved assay designed by and based on conventional hybridization capture. The tDNA and cfDNA exome platforms were self-developed for clinical use and they included the detection of single-nucleotide variants (SNVs), small insertions/deletions (InDels), gene rearrangements/fusions, and copy number variants (CNVs) with a panel of 67 clinically relevant cancer genes (Table 2). After library construction and hybridization capture, the targeted tDNA and cfDNA fragments were amplified by PCR with 15 cycles for sequencing.

Sequencing and bioinformatic processing. Sequencing was performed on Illumina NextSeq CN500. The quality control of raw data was conducted, and clean reads were obtained after removing adaptor sequences and reads of low mapping quality. For alteration and variant analysis, sequences were aligned to the reference human genome (hg19) by BWA software. The linkage of sequencing fragments was optimized by GATK software and the information of SNV/InDel was acquired by VarScan2 software.[9,10] The sequencing coverage of targeted positions was ≥90% required, and the average sequencing depth for blood specimens and tissue specimens was, respectively, ≥1000× and ≥500×required. A variant with an alteration frequency ≥0.2% for blood specimens or ≥1.0% for tissue specimens was considered to be credible. Final alteration calling results were performed by non-synonymous variants filtering with 1000 genomes database for single nucleotide polymorphism (SNP) loci excluding. The R package, maftools,[11] was used to depict the alteration landscapes and conduct the mutation relation test (MRT).

Statistical analysis. All statistical analyses were conducted by statistical software SPSS 20.0. Measurement data was exhibited as mean±standard deviation (x±s). The comparison of alteration spectrum between tissue and blood specimens was measured by chi-square or fisher's exact tests as appropriate. The comparison of variants' frequency between

Table 2 Targeted Panel Used in This Study

Detection	Genes								
SNV & InDel	ABL1	AKT1	ALK	APC	ARID1A	ATM	BRAF	BRCA1	BRCA2
	CDKN2A	CTNNB1	DDR2	EGFR	ERBB2	ERBB3	ERBB4	ESR1	FBXW7
	FGFR2	FGFR3	FLT3	GNA11	GNAQ	HNF1A	HRAS	IDH1	IDH2
	JAK2	KDR	KIT	KRAS	MAP2K1	MET	MLH1	MPL	MTOR
	NOTCH1	NPM1	NRAS	NTRK3	PDGFRA	PIK3CA	PTEN	RAF1	RET
	ROS1	SMAD4	SMO	SRC	TP53	TSC2	GSTP1	CDA	UGT1A1
	MTHFR	CYP1B1	CYP2D6	NQO2	MTR	DPYD			
CNV Rearrangement	EGFR	ERBB2	FGFR1	FGFR2	FGFR3	KIT	MET	RICTOR	
	ALK	FGFR3	MET	RET	ROS1				

different specimens was also measured by chi-square or fisher's exact tests as appropriate. Two-sided $p < 0.05$ were considered statistically significant.

Results

Characteristics of Acquired Somatic DNA Alterations

Based on the target-captured sequencing data of 3000 patients, the spectra of somatic alteration of NSCLC in East-China via tissue and blood specimens were depicted respectively.

Of 1351 tissue specimens, 3080 alterations were detected in 67 genes in 1151 (85%) patients, including 2162 (70%) missense SNVs, 87 (3%) nonsense SNVs, 117 (4%) frameshift InDels, 323 (10%) non-frameshift InDels, 266 (9%) CNVs, and 125 (4%) rearrangements (Figure 1A). Among the alteration-positive tissue specimens, 428 (32%) specimens were detected with 1 variant, 341 (25%) had 2 variants, 180 (13%) had 3 variants, and 202 (15%) had more than 3 variants (Figure 1C).

In 1649 blood specimens, the number of gene alterations in aggregate totaled 3953, in which 2956 (75%) were missense SNVs, 121 (3%) were nonsense SNVs, 100 (3%) were frameshift InDels, 250 (6%) were non-frameshift InDels, 443 (11%) were CNVs, and 83 (2%) were rearrangements (Figure 1B). The 1147 (69%) blood specimens with detected alterations comprised 402 (24%) with 1 variant detected, 311 (19%) with 2 variants detected, 187 (11%) and 247 (15%) with 3 variants and more than 3 variants detected, respectively (Figure 1D).

Among all the mutated genes, EGFR showed the highest alteration frequency in both tissue (49.7%) and blood (32.1%) specimens. Aside from EGFR, tissue and blood specimens also shared a series of frequently mutated genes including TP53 (22.7% and 31.4%), KRAS (10.1% and 11.6%), CDKN2A (5.3% and 4.8%), PIK3CA (4.8% and 7.1%), ALK (4.2% and 4.8%), ERBB2 (3.6% and 4.4%) and MET (3.7% and 4.3%) (Figure 2). Gene amplification was frequently detected in EGFR (5.0%), MET (1.9%) in tissue specimens and in EGFR (1.8%), MET (1.8%), RICTOR (2.2%) in blood specimens, respectively. Gene fusions were detected in ALK (2.7% and 0.9%), RET (1.5% and 0.7%), ROS1 (0.6% and 0.2%) and FGFR3 (0.4% and 0.06%) in both tissue and blood specimens.

Subsequently, the alteration spectra between tissue and blood specimens were compared with each other (Figure 3A), and further with the data from the TCGA database (Figure 3B).[12,13] When compared with blood specimens, alterations in EGFR

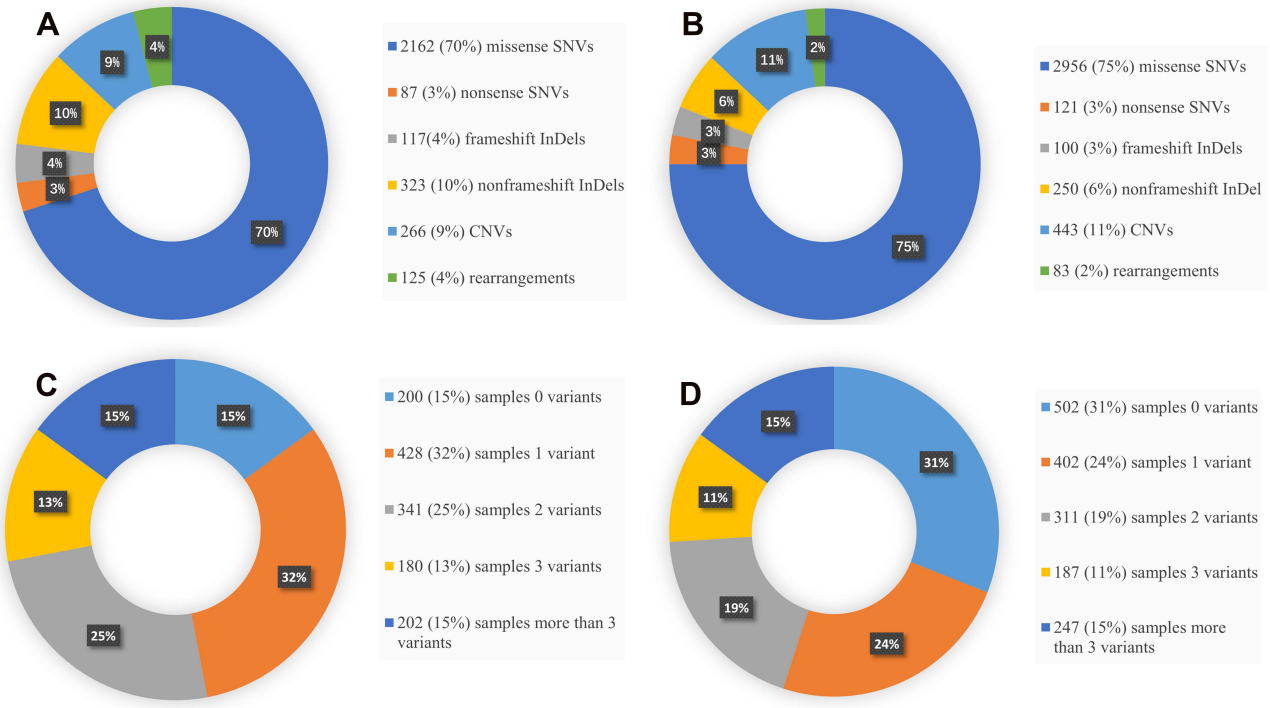

Figure 1 Alteration spectrum of Chinese NSCLC patients depicted by using tissue specimens (**A** and **C**) and by using blood specimens (**B** and **D**).

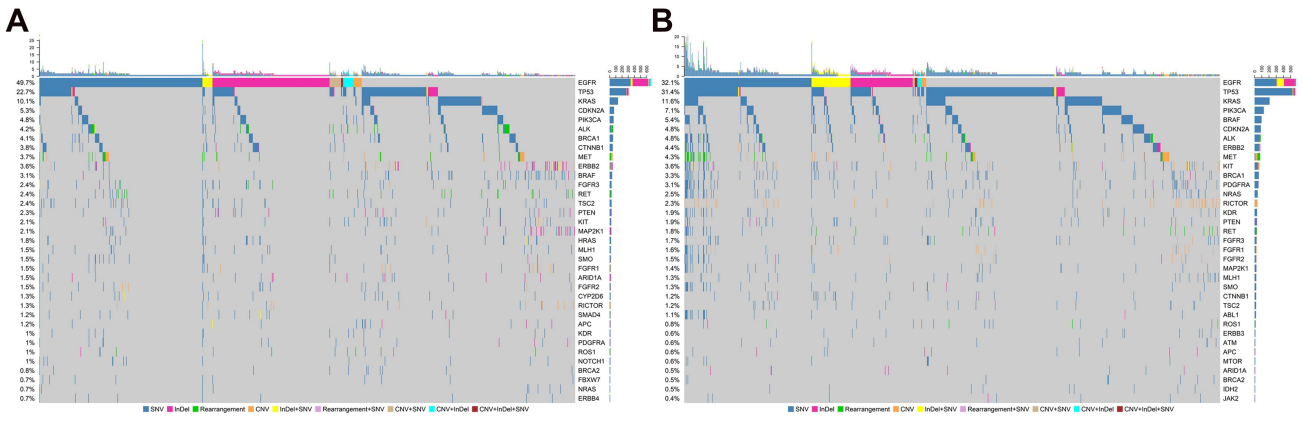

Figure 2 The landscapes of alterations in tissue (**A**) and blood (**B**) cohort.

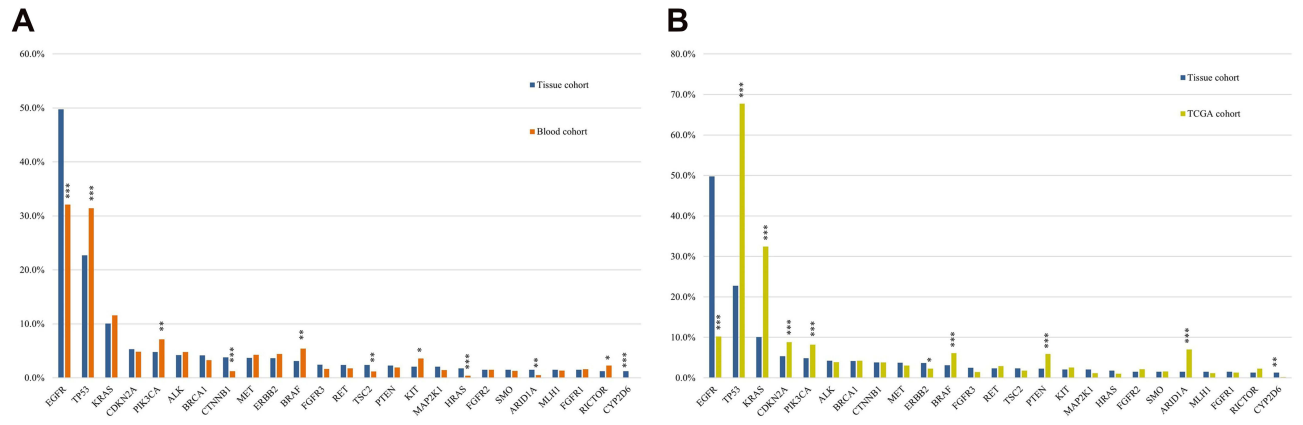

Figure 3 The comparison of spectrum between tissue and blood cohort (**A**), and between tissue and TCGA cohort (**B**). *p < 0.05; **p < 0.01; ***p < 0.001.

(P < 0.001), CTNNB1 (P < 0.001), TSC2 (P < 0.01), HRAS (P < 0.001), ARID1A (P < 0.01), CYP2D6 (P < 0.001) were found to be significantly more frequent in tissue specimens, whereas alterations in TP53 (P < 0.001), PIK3CA (P < 0.01), BRAF (P < 0.01), KIT (P < 0.05), RICTOR (P < 0.05) were notably less frequent. Prominently high alteration frequencies of EGFR (P < 0.001) and ERBB2 (P < 0.05) were exhibited in tissue specimens in East-China when compared with the most frequently mutated genes on record in the TCGA database, whereas TP53 (P < 0.001), KRAS (P < 0.001), CDKN2A (P < 0.001), PIK3CA (P < 0.001), BRAF (P < 0.001), PTEN (P < 0.001) were found to be significantly lower.

Hotspot Variants in NSCLCs in East-China

The most commonly affected genes found in the genomic landscape studies were EGFR, TP53, KRAS, CDKN2A, and PIK3CA, with an occurrence of 49.7% and 32.1%, 22.7% and 31.4%, 10.1% in tissue and 11.6%, 5.3% and 7.1%, 4.8% and 5.4% in blood specimens, respectively, and most of these variants resulted in poor prognosis in retrospective studies.[4,14,15] Consequently, we conducted a comprehensive description of the composition of hotspot variants by large-scale tissue and blood specimens (Supplementary Figure 1).

The most frequently mutated variants in EGFR were L858R and deletion of exon 19 in both tissue and blood specimens, and simultaneously unusual variants such as L858M and insertions of exon 19 (like 738>VKIPVAI, 740>IPVAIKG, 741>PVAIKVH, etc) were found. Among diverse deletions of exon 19, 745_750del, 746_750del, 747_753del, 746_751del, and 746_748del represented greater than 80% of all variants. The insertions of exon 20 in EGFR mainly occurred from codon 766 to codon 775, in which the variant in codon 766 and codon 770 comprised more than half of all insertions in both specimens. Some rare variants that did not appear in tissue specimens were detected in

blood specimens, such as C719D, S768R, L861R, etc. When comparing the composition of altered exons, exons 19, 20, and 21 showed significant differences in the two cohorts (Supplementary Figure 1A).

The prevalence of hotspot variants in TP53 proved to be complicated. Despite a lack of extremely common variants detected, in general, alterations of codon V173, R175, H179, G245, R248, R249, R273 and R282 were the major components, and R273 seemed to be the most frequently mutated in both tissue and blood specimens. Statistical differences were found in exons 4, 6, 7, and 8 with respect to the composition of altered exons comparing tissue and blood specimens (Supplementary Figure 1B).

The distribution of variants in KRAS was similar in tissue and blood specimens. Nearly 80% of variants were concentrated in exon 2, of which the alterations of codon G12 were the major components. Moreover, although codons G12, G13, Q61 and A146 had a relatively high constituent ratio in both specimens, more complex molecular variants were detected in blood specimens, such as G13R/S, Q61L/R, K147E, K5N (Supplementary Figure 1C). Most variants in PIK3CA were located in exons 10 and 21, of which the common mutations were E542, E545 and H1047. The comparison of composition of altered exons between the two kinds of specimen was not significantly different (Supplementary Figure 1D). Hotspot variants in CDKN2A showed a centralized distribution in exon 2, in which codons E69, D74, D84 were relatively prominent (Supplementary Figure 1E).

Co-Occurring Somatic Alterations in NSCLC in East-China

A large-scale analysis of somatic alterations was conducted for the co-occurring alterations in NSCLC in East-China. The analysis of 1151 alteration-positive tissue specimens and 1293 alteration-positive blood specimens revealed co-occurring genetic alterations were widely prevalent (48% in tissue and 43% in blood specimens), in addition to the most commonly known EGFR-related alterations – L858R and Exon 19 deletion.

MRT was conducted to reveal correlations and mutual-exclusion relationships among significantly and highly mutated genes in a pairwise fashion. In tissue specimens (Figure 4A), as a result, positive correlations were observed between EGFR and CTNNB1, between PIK3CA and PTEN, as well as between ERBB2 and SMO, which implied that the alterations of these gene pairs likely function synergistically to promote carcinogenesis. Whereas, negative correlations between EGFR and TP53, between EGFR and KRAS, as well as between EGFR and BRAF suggested that the alteration of a single component or pathway may be sufficient, wholly or in part, for carcinogenesis.[16] In blood specimens (Figure 4B), positive correlations were found between EGFR and SMO, between KRAS and PDGFRA, as well as between PIK3CA and KDR, while negative correlations were most likely to appear between EGFR and TP53, between EGFR and KRAS, as well as between TP53 and PIK3CA.

Although MRT showed negative correlations between EGFR and TP53 in both tissue and blood specimens (Figure 4), the results were different when analyzing one specific EGFR-related oncogenic alteration. In accordance with Figure 5A, 22.9% and 46.7% of patients with canonical EGFR-L858R alteration carried TP53 mutations in tissue and blood specimens,

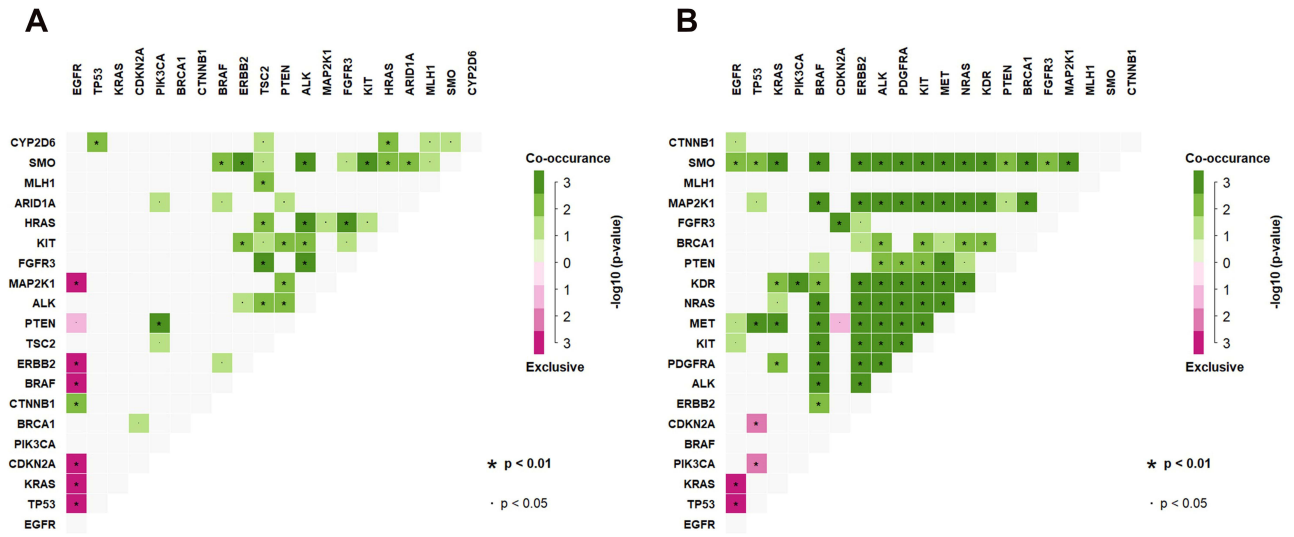

Figure 4 Mutation relation test (MRT) in tissue cohort (**A**) and in blood cohort (**B**). Color shade indicates the degree of co-occurrence or exclusivity. ˙p < 0.05; *p < 0.01.

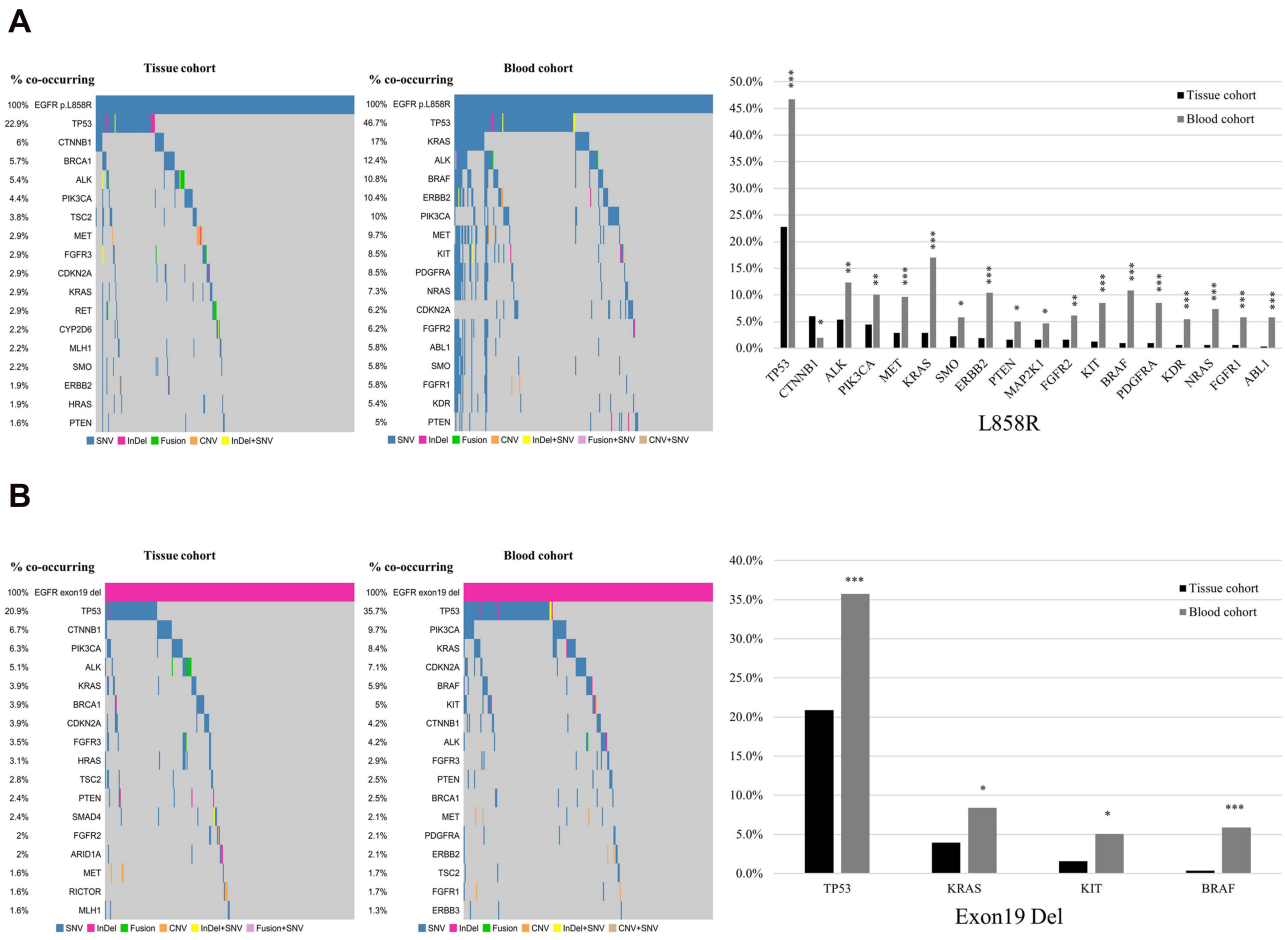

Figure 5 Comprehensive analysis of co-occurring genes in EGFR-L858R positive (**A**) and EGFR-exon19del positive (**B**) cohort. *p < 0.05; **p < 0.01; ***p < 0.001.

respectively. Likewise, for another hotspot EGFR mutation, exon19del, concurrent TP53 mutations were found in 20.9% and 35.7% of carriers in tissue and blood specimens, respectively (Figure 5B). Data also showed that canonical EGFR-L858R alteration co-occurred with alterations in several other genes besides TP53 (Figure 5A), including CTNNB1, BRCA1, PIK3CA, etc. in tissue specimens, and KRAS, ERBB2, BRAF, etc. in blood specimens. Similar to L858R, the major co-alteration genes in addition to TP53 presented in exon19del-positive specimens (Figure 5B) were CTNNB1, PIK3CA, KRAS, etc. in tissue specimens, and PIK3CA, KRAS, CDKN2A, etc. in blood specimens. Regardless of L858R or exon19del, moreover, the enrichment of co-alteration for certain genetic events was significantly higher in blood specimens than that in tissue specimens (Figure 5), which was putatively because blood specimens were mainly collected from advanced NSCLC patients with higher co-alteration rates.

Actionable and Novel Alterations Detected in Tissue and Blood Specimens

The main purpose of gene detection related to clinical therapy was to provide clinical target-drug medication guidance for NSCLC patients. At that point, the final variant-calling was annotated with the "druggable or potentially actionable database". The database was established, and alteration-related medication was variously categorized, including labels such as "FDA approved", "NMPA approved", or "NCCN recommended". Additionally, the database included gene targets of existing and novel therapeutics, prognostic markers, and other oncogenes and tumor suppressors that were frequently mutated in cancer. In total, 884 (65.4%) tissue specimens and 948 (52.8%) blood specimens were detected with actionable alterations, some of which were summarized and analyzed to investigate the differences of preference between tissue and blood specimens (Table 3).

Table 3 Frequencies of Actionable Variants in Tissue and Blood Cohorts

Gene	Variant	Tissue		Blood		P
		Amount	Frequency	Amount	Frequency	
EGFR	G719X	29	2.10%	40	2.20%	–
	S768I	18	1.30%	5	0.30%	***
	T790M	20	1.50%	110	6.10%	***
	L858R	322	23.80%	260	14.50%	***
	L861Q	12	0.90%	17	0.90%	–
	Exon18 del	6	0.40%	2	0.10%	–
	Exon19 del	275	20.40%	260	14.50%	***
	Exon19 ins	4	0.30%	5	0.30%	–
	Exon20 ins	29	2.10%	20	1.10%	*
	E709K	2	0.10%	5	0.30%	–
	L747S	1	0.10%	5	0.30%	–
	D761Y	0	0.00%	4	0.20%	–
	C797S	2	0.10%	18	1.00%	**
	T854A	0	0.00%	3	0.20%	–
	L861R	0	0.00%	2	0.10%	–
	Amp	67	5.00%	32	1.80%	***
KRAS	Codon 12	110	8.10%	137	7.60%	–
	Codon 13	13	1.00%	24	1.30%	–
	Exon 3 (A59/Q61, etc)	6	0.40%	35	1.90%	***
	Exon 4 (A146/K147, etc)	1	0.10%	12	0.70%	*
PIK3CA	Exon 2 (R88, etc)	3	0.20%	3	0.20%	–
	Exon 10 (E545/Q546, etc)	43	3.20%	91	5.10%	**
	Exon 21 (M1043/H1047, etc)	17	1.30%	26	1.40%	–
BRAF	V600	9	0.70%	34	1.90%	**
	Non-V600 (G469/D594/K601, etc)	24	1.80%	74	4.10%	***
ERBB2	Activation	6	0.40%	46	2.60%	***
	Exon20 ins	23	1.70%	12	0.70%	**
	Amp	8	0.60%	6	0.30%	-
PTEN	R130/R233/Q245, etc	13	1.00%	25	1.40%	-
MET	Exon14 skipping	22	1.60%	36	2.00%	-
	Amp	26	1.90%	32	1.80%	-
RICTOR	Amp	10	0.70%	39	2.20%	**
ALK	Fusion	36	2.70%	17	0.90%	***
RET	Fusion	20	1.50%	13	0.70%	*
ROS1	Fusion	8	0.60%	4	0.20%	-

Note: *p < 0.05; **p < 0.01; ***p < 0.001; -p ≥ 0.05.

The results of tissue specimens showed a large number of primary druggable variants (like EGFR-L858R, EGFR-exon19del, EGFR-S768I, etc), InDel of small fragments (like EGFR-exon19del, EGFR-exon20ins, ERBB2-exon20ins, etc), and chromosomal structural rearrangement (like ALK-Fusion, RET-Fusion, etc). All of the aforementioned variants were significantly different between tissue and blood specimens. However, results of

blood specimens showed numerous identified drug-resistance variants (like EGFR-T790M, EGFR-C797S), rare actionable variants (KRAS-exon 3/4, BRAF-nonV600, etc), and non-EGFR actionable variants (such as BRAF-V600, activated variants in ERBB2 and PIK3CA, etc). These variants also showed dissimilarity between tissue and blood. The differences and dissimilarity might be attributed to the sampling time: most tissue specimens were tested with the orthotopic resection, while blood was typically sampled following a series of therapeutic contacts.[17] Blood specimens might be more suitable for determining the extremely rare EGFR variants, such as E709K, L747S, D761Y, T854A, L861R, etc, relative to tissue specimens, although there were no statistical differences between the two specimens.

Although numerous lung cancer-related genes and variants have been recorded in a series of databases such as COSMIC, this study identified novel variants that had not been recorded in COSMIC or Reference SNP Report and had no specific target drugs reported in either tissue and blood specimens. Strikingly, 845 (27.4%) and 1252 (31.7%) novel alterations were detected in tissue and blood specimens, respectively. Moreover, the distribution patterns of novel variants were quite different between the two specimens: novel variants were enriched in KDR (45/845, 5.3%), KIT (5.0%), TP53 (4.9%), ABL1 (4.4%), FGFR1 (4.0%) in tissue specimens while the majority of novel variants were distributed in PDGFRA (89/1252, 7.1%), EGFR (7.0%), KIT (6.0%), ROS1 (5.0%), BRCA2 (4.8%) in blood specimens (Figure 6). Interestingly, the dominant proportion of novel variants were found in mutated gene TP53 in tissue specimens in contrast to EGFR in blood specimens, which reflects the great genetic heterogeneity of NSCLC.

Variant Allele Frequency in the Two Specimens

Variant allele frequency (VAF) can be used to infer tumor heterogeneity and tumor purity. In addition, the level of VAF may affect the prognosis of cancer.[18] Therefore, VAFs in tissue and blood specimens were analyzed. As we speculated, VAFs in tissue specimens (mean = 14.4%) were significantly ($P < 0.001$) higher than that in blood specimens (3.9%). Furthermore, predominantly higher VAFs of the five most commonly affected genes (EGFR, TP53, KRAS, PIK3CA and CDKN2A) were also observed in tissue specimens compared with blood specimens, respectively (Figure 7).

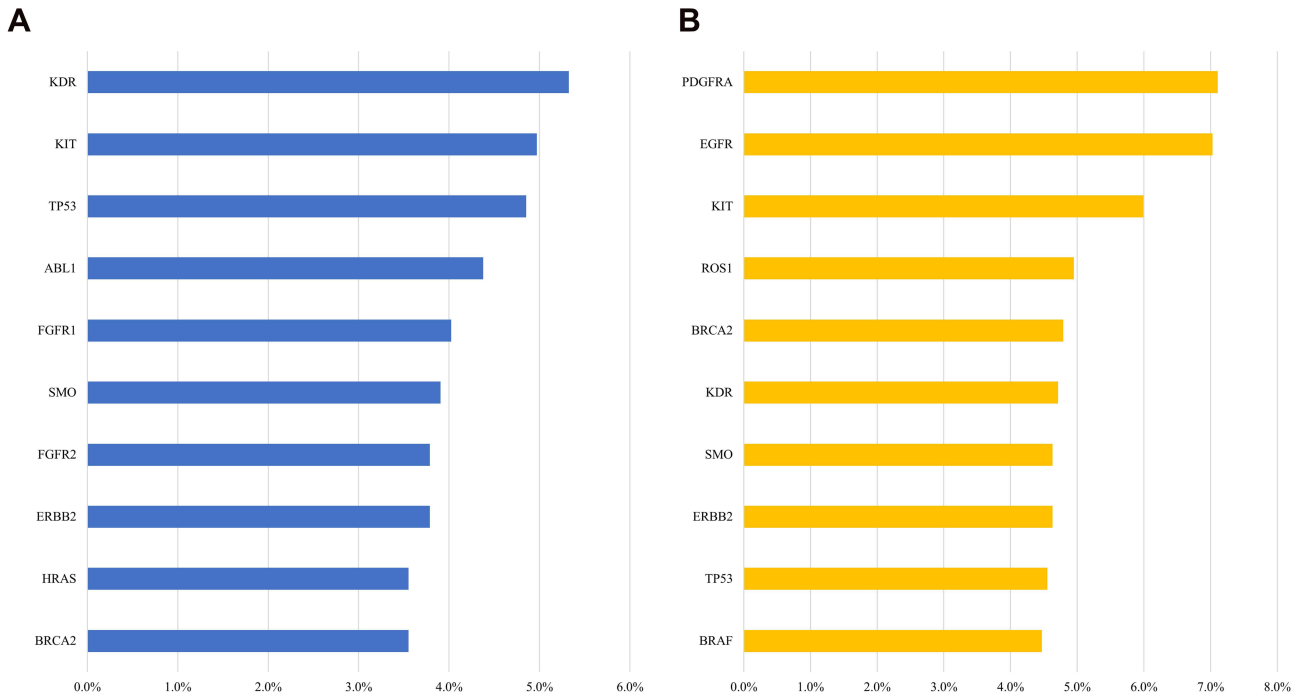

Figure 6 10 genes with the highest number of novel variants and their proportions respectively in tissue (**A**) and blood (**B**) cohort.

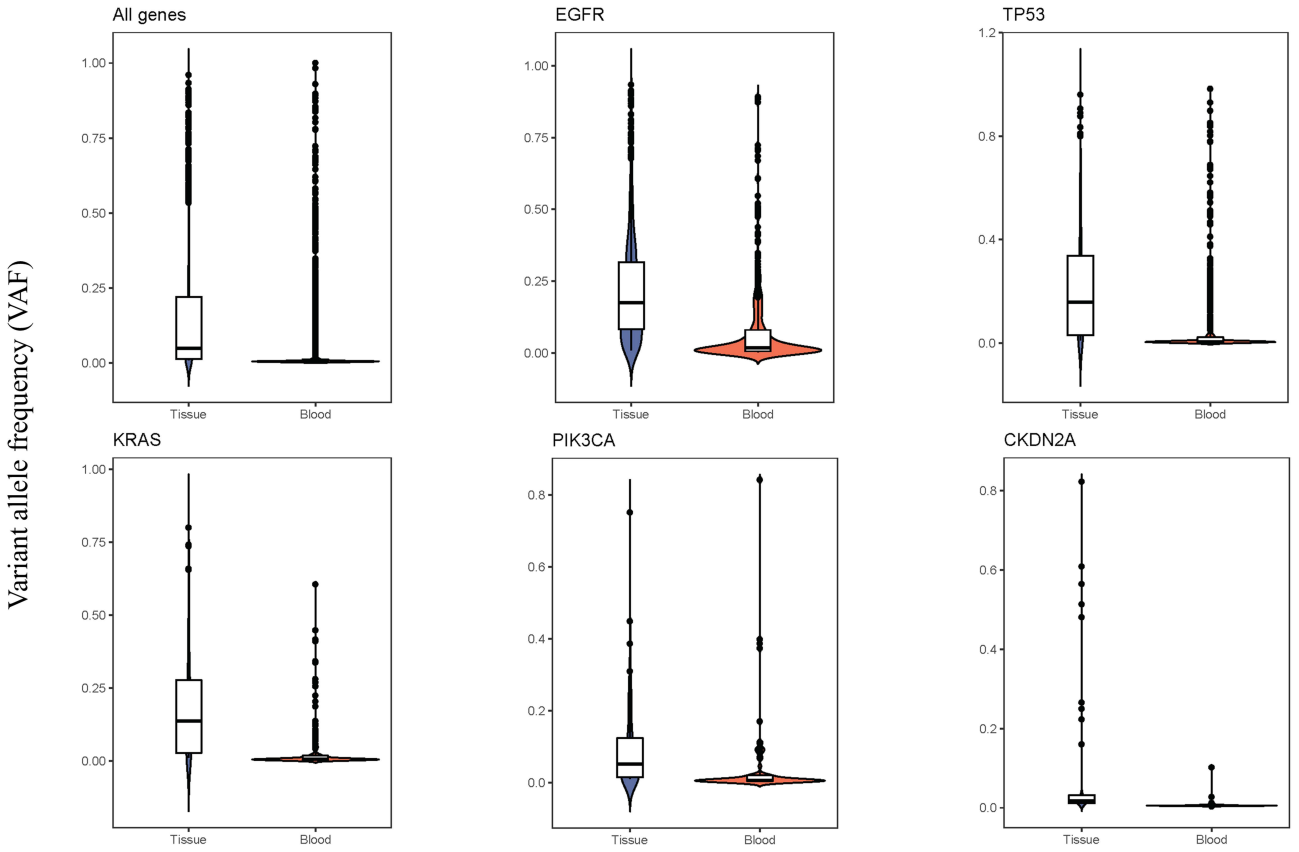

Figure 7 Comparison of variant allele frequency (VAF) of all genes, EGFR, TP53, KRAS, PIK3CA and CDKN2A between tissue and blood cohort.

Discussion

Concurrent with the development of targeted medicine, individual therapy based on molecular biomarker analysis has gradually evolved into standard of care in NSCLC.[19] Therefore, a comprehensive understanding of genetic alterations of the disease that remains persistently associated with high mortality and morbidity is crucial for treatment. Although numerous studies have been conducted investigating the genetic alterations of NSCLCs in recent years,[20–22] there is still a paucity of large-scale data sets to depict the detailed genomic profiling of Chinese NSCLC patients, especially cfDNA assays. In response, a large-scale investigation of 1351 tissue specimens and 1649 blood specimens was performed to analyze the genetic alterations of NSCLCs in East-China, to evaluate the similarities and differences between tissue and blood specimens, and to extend the clinical applications.

The comprehensive molecular analysis of genetic alterations (SNV, InDel, CNV, Rearrangement) of the 67 cancer-related genes depicted the alteration spectrum of NSCLCs in East-China. In concert with previously well-known studies,[20,23,24] this analysis identified the frequently mutated genes, such as oncogenes EGFR, KRAS, PIK3CA and tumor suppressor genes TP53 and CDKN2A, in both tissue and blood specimens (Figure 2). In general, the alteration frequencies of KRAS, CDKN2A, ALK, BRCA1, MET, ERBB2, and PTEN were comparable between two specimen types, while blood specimens exhibited significantly higher alteration frequencies of TP53, PIK3CA, BRAF and significantly lower alteration frequencies of EGFR, CTNNB1 (Figure 3A). The more mutated genes detected in blood specimens may derive from intra- and inter-tumor heterogeneity, which represent genetic alteration differences between the primary lesion and its metastases or between different regions of the same lesion.[25–27] With the single-site sampling method, it is difficult to avoid tumor heterogeneity.

To our knowledge, the present study represents the largest Chinese cohort (1351 tissue samples and 1649 blood samples) of genomic characteristics of Chinese NSCLC patients, reminiscent of a recent study (Zhou et al) that analyzed 1000 FFPE tissue samples and 500 blood samples using targeted exome sequencing.[28] Despite slight discrepancies, our

results strongly supported the genomic alteration spectrum depicted by that study, demonstrating similar patterns of frequently mutated genes EGFR, KRAS, TP53, CDKN2A and their hotspot variants as well as recurrent ALK fusions and EGFR co-alterations that were found in tissue and blood samples. One of the most prominent differences in the two studies, however, was that EGFR proved to be the most frequently mutated gene in our cohort, whereas TP53 was even more recurrently mutated than EGFR in Zhou's study. Possible explanations for these differences include: (i) compared with our data, the proportion of patients with lung squamous carcinoma (tissue: 14.2% vs 4.3%; blood: 17% vs 5.1%) and a smoking history (tissue: 5.5% vs 2.4%; blood: 5% vs 3.2%) was higher in Zhou's study, and it was reported that TP53 was more frequently mutated in squamous cell carcinomas and patients with a smoking history;[29] (ii) most of the patients in Zhou's study were from Jiangxi Province that is located in the interior of China, whereas in our cohort the majority of patients were from Jiangsu, a coastal province in eastern China. These possible reasons remind us of the importance of sampling and regional effects in the exploration of genomic characteristics in NSCLC patients. Compared to TCGA (mainly Caucasian population), some prominent mutated genes in tissue specimens showed a significantly lower somatic alteration rate, including TP53, KRAS, CDKN2A, PIK3CA, BRAF, PTEN, except for EGFR, whereas ALK, BRCA1, MET, and RET had consistent alteration rates (Figure 3B). Our results also indicated that NSCLC patients in East-China often had relatively simple alteration characteristics but extremely high alteration frequency of EGFR. It has been speculated that the inhomogeneous alteration frequencies might result principally from the ethnically innate differences and cancer subtypes among Chinese populations and European/American populations.[30–32]

Co-occurring oncogenic events are usually related to therapeutic style and prognostic prediction. For example, concurrent TP53 mutations were reported to be associated with the poor prognosis of Chinese EGFR-mutant NSCLC patients.[29,33] Our study sheds new light on the genetic basis of oncogenesis and cancer progression and indicated that multiple co-occurring oncogenic events were present in the majority of NSCLCs in East-China. Similar to previous studies,[34–36] our results indicated that the co-alteration rates between EGFR and TP53, EGFR and KRAS, and EGFR and BRAF were significantly low, but those patients detected with EGFR-L858R or EGFR-Exon19del were more likely to simultaneously carry variants of TP53, KRAS, or BRAF in both tissue and blood specimens (Figure 4). The data reemphasize that there were no absolute co-occurring or mutually exclusive events in the highly complex oncogenesis mechanisms of NSCLCs. Also worthy of note, the spectrum of co-alteration genes in cfDNA was vastly different than tDNA, which might also reflect tumor heterogeneity as previously specified. Overall, the co-occurring genetic alterations in this study may supplement the current view of NSCLC oncogenesis, and lay the foundation of future exploration for both basic and clinical investigations.

In clinical treatment, both tissue and liquid biopsy are frequently used to identify genetic alterations that facilitate specific molecular treatments, monitor tumor evolution and determine drug-resistance in real-time.[37] In our study, a comprehensive analysis of druggable genes was subsequently conducted to distinguish the alteration preference between the two kinds of specimens. The results revealed some significant differences: First, in tissue specimens obtained at initial diagnosis, EGFR-related druggable variants were detected, such as L858R and exon19del. In contrast, for blood specimens obtained after treatment, other alteration-related drug-resistance variants were detected, which seemed to reflect the molecular clonal evolution of the tumor in response to the treatment. Differences in the molecular results of tissue and blood likely reflected the circumstances of real-life clinical testing. Because genetic tests remain expensive, to obtain both tissue and blood would represent cost duplication. Therefore, tissue necessary for diagnosis was also used for genetic analysis. Repeat testing would only occur if the effectiveness of the treatment medicine indicated resistance. At this point, obtaining blood only would be logical to avoid the expense and invasiveness of obtaining tissue. Our results reflected this situation to a large extent. Second, tissue specimens were more suitable for identifying chromosomal structural variants, such as gene amplifications and fusions, and it might have something to do with DNA chain fragmentation and differences in detection methods. Third, blood specimens were more suitable for detecting extremely rare variants and non-EGFR driver gene alterations, which might be related to lower tumor heterogeneity in sampling methods.

Numerous previously published mutational studies established sound basis for targeted therapies to treat NSCLC, such as erlotinib, gefitinib, cetuximab (against EGFR) and most recently, crizotinib (against rearranged ALK).[38] Nevertheless, our data, in conjunction with other reports have verified the highly complex oncogenic mechanisms of

NSCLC. In our study, approximately one-third of variants detected were novel variants without specific targeted drug treatments. The identification of these novel variants was due in part to five issues: (i) the more recently introduced NGS technology was able to detect known and novel mutations, (ii) sufficient sequencing depth was performed and deeper sequencing led to discovery of a greater number of rare alterations, (iii) all the data were collected from real-world hospitals instead of clinical trial series, (iv) the number and variety of samples analyzed were large, (v) all the patients were Han Chinese from East China. The distribution patterns of novel variants between the two types of specimens were quite distinct (Figure 6), perhaps attributable to the samples being obtained from different patients at different stages of clinical treatment. To our knowledge, these unreported mutations showed significant genetic heterogeneity of NSCLC and suggested that the alteration spectra of NSCLC in previous studies are far from complete. We believe that the novel mutations are yet to be exploited for the development of new targeted drugs. Moreover, there might be great significance on clinical application of these novel mutations for judging malignant degree and prognosis, which needs to be further investigated.

In general, smoking history, gender, age and tumor type are all associated with the pathological type of NSCLC. In China, according to the existing NSCLC-related guidelines or consensus, clinicians prefer to recommend genetic testing for young NSCLC patients with non-smoking history and adenocarcinoma. In this real-world study, therefore, the case information in both tissue and blood specimens was severely biased between smokers and non-smokers and between adenocarcinoma and squamous carcinoma (Table 1). But it did not mean that there is a higher incidence of NSCLC in non-smokers, nor that there are more NSCLC patients with adenocarcinoma. Moreover, although tissue and blood specimens were unpaired, there was no significant difference for age, gender, smoking history and tumor type between the two groups of patients providing samples (Table 1). Due to the different features of tissue and liquid biopsy, there was significant difference for tumor stage between tissue and liquid specimens. Concretely, the tissue samples were mainly collected from NSCLC patients with early stage (stage I) while the majority of blood samples were collected from patients with advanced NSCLC (stage IV).

In summary, our study depicted the genetic alteration spectra of NSCLC in East-China by large-scale tissue and blood specimen, respectively, which also revealed the hotspot, actionable and novel variants of the two kinds of specimens. Results indicated that the genetic alteration spectrum of NSCLC in East-China was associated with innate ethnic differences. Moreover, detection of cfDNA allowed not only individualized targeted medicine, but also a more comprehensive analysis of the disease progression and monitoring, which has been a major challenge for personalized treatment of NSCLC.[39] Additionally, co-occurring somatic alterations should receive more attention, especially related to drug resistance and disease progression. Besides, cfDNA based liquid biopsy was also suitable for the rare variant detection, not just limited to the gene EGFR. Therefore, it would seem prudent to obtain liquid biopsy for patients who are undergoing or about to receive tailored targeted therapy. Finally, further studies on pathological classification, response to treatment, drug monitoring and evaluation of therapeutic efficiency should continue to be given high priority.

Article Highlights

- A real-world study was conducted utilizing the largest East-China Chinese cohort (3000 patients) diagnosed with NSCLC.
- Analysis of both tumor tissue and peripheral blood samples from the Chinese NSCLC cohort produced a wide spectrum of genomic alterations.
- EGFR, TP53, KRAS, CDKN2A, and PIK3CA were frequently mutated in both tissue and blood samples.
- Novel and druggable genomic variants were significantly different between tissue and blood samples.
- The frequency of variant alleles in tissue samples was significantly greater than that in blood samples.

Abbreviations

NSCLC, non-small cell lung cancer; NGS, next-generation sequencing; tDNA, tumor tissue DNA; cfDNA, cell free circulating tumor DNA; FRE, fresh surgical; FNA, fine-needle aspiration; FFPE, formalin-fixed paraffin-embedded; HYTC, hydrothorax cells; MRT, mutation relation test; PB, peripheral blood; SNV, single-nucleotide variant; InDel,

insertion and deletion; CNV, copy number variant; SNP, single-nucleotide polymorphism; TCGA, the cancer genome atlas; VAF, variant allele frequency.

Acknowledgments

The authors thank the patients and their families for agreeing to use their data for research purposes, and specifically, for publication of this paper. Thank Precision Diagnosis Center of Dian for providing technical help. The authors would like to express their gratitude to EditSprings (https://www.editsprings.cn/) for the expert linguistic services provided.

Disclosure

The authors declare no conflict of interest.

References

1. Yang D, Liu Y, Bai C, et al. Epidemiology of lung cancer and lung cancer screening programs in China and the United States. *Cancer Lett*. 2020;468:82–87. doi:10.1016/j.canlet.2019.10.009
2. Mayekar MK, Bivona TG. Current landscape of targeted therapy in lung cancer. *Clin Pharmacol Ther*. 2017;102(5):757–764. doi:10.1002/cpt.810
3. Zhou X, Xu X, Tian Z, et al. Mutational profiling of lung adenocarcinoma in China detected by next-generation sequencing. *J Cancer Res Clin Oncol*. 2020;146(9):2277–2287. doi:10.1007/s00432-020-03284-w
4. Zhang B, Zhang L, Yue D, et al. Genomic characteristics in Chinese non-small cell lung cancer patients and its value in prediction of postoperative prognosis. *Transl Lung Cancer Res*. 2020;9(4):1187. doi:10.21037/tlcr-19-664
5. Liu HE, Vuppalapaty M, Wilkerson C, et al. Detection of EGFR mutations in cfDNA and CTCs, and comparison to tumor tissue in non-small-cell-lung-cancer (NSCLC) patients. *Front Oncol*. 2020;10:2030.
6. Chabon JJ, Simmons AD, Lovejoy AF, et al. Circulating tumour DNA profiling reveals heterogeneity of EGFR inhibitor resistance mechanisms in lung cancer patients. *Nat Commun*. 2016;7(1):1–15.
7. Murtaza M, Dawson SJ, Tsui DWY, et al. Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. *Nature*. 2013;497(7447):108–112. doi:10.1038/nature12065
8. Lin X, Dong W, Lai X, et al. The clinical value of circulating tumor DNA detection in advanced non-small cell lung cancer. *Transl Cancer Res*. 2019;8(1):170–179. doi:10.21037/tcr.2019.01.20
9. McKenna A, Hanna M, Banks E, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297–1303. doi:10.1101/gr.107524.110
10. Koboldt DC, Zhang Q, Larson DE, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res*. 2012;22(3):568–576. doi:10.1101/gr.129684.111
11. Mayakonda A, Lin DC, Assenov Y, et al. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res*. 2018;28 (11):1747–1756. doi:10.1101/gr.239244.118
12. Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. *Nature*. 2012;489(7417):519. doi:10.1038/nature11404
13. Collisson E, Campbell J, Brooks A, et al. Comprehensive molecular profiling of lung adenocarcinoma: the cancer genome atlas research network. *Nature*. 2014;511(7511):543–550.
14. Canale M, Petracci E, Delmonte A, et al. Impact of TP53 mutations on outcome in EGFR-mutated patients treated with first-line tyrosine kinase inhibitors. *Clin Cancer Res*. 2017;23(9):2195–2202. doi:10.1158/1078-0432.CCR-16-0966
15. VanderLaan PA, Rangachari D, Mockus SM, et al. Mutations in TP53, PIK3CA, PTEN and other genes in EGFR mutated lung cancers: correlation with clinical outcomes. *Lung Cancer*. 2017;106:17–21. doi:10.1016/j.lungcan.2017.01.011
16. Dees ND, Zhang Q, Kandoth C, et al. Music: identifying mutational significance in cancer genomes. *Genome Res*. 2012;22(8):1589–1598. doi:10.1101/gr.134635.111
17. van Delft F, Koffijberg H, Retèl V, et al. The validity and predictive value of blood-based biomarkers in prediction of response in the treatment of metastatic non-small cell lung cancer: a systematic review. *Cancers*. 2020;12(5):1120. doi:10.3390/cancers12051120
18. Tran HT, Lam VK, Elamin YY, et al. Clinical outcomes in non–small-cell lung cancer patients treated with EGFR-tyrosine kinase inhibitors and other targeted therapies based on tumor versus plasma genomic profiling. *JCO Precis Oncol*. 2021;5(5):1241–1249. doi:10.1200/PO.20.00532
19. Hirsch FR, Scagliotti GV, Mulshine JL, et al. Lung cancer: current therapies and new targeted treatments. *Lancet*. 2017;389(10066):299–311. doi:10.1016/S0140-6736(16)30958-8
20. Schrock AB, Welsh A, Chung JH, et al. Hybrid capture–based genomic profiling of circulating tumor DNA from patients with advanced non–small cell lung cancer. *J Thorac Oncol*. 2019;14(2):255–264. doi:10.1016/j.jtho.2018.10.008
21. Mascarenhas E, Gelatti AC, Araújo LH, et al. Comprehensive genomic profiling of Brazilian non-small cell lung cancer patients (GBOT 0118/ LACOG0418). *Thorac Cancer*. 2021;12(5):580–587. doi:10.1111/1759-7714.13777
22. Passaro A, Attili I, Rappa A, et al. Genomic characterization of concurrent alterations in Non-Small Cell Lung Cancer (NSCLC) harboring actionable mutations. *Cancers*. 2021;13(9):2172. doi:10.3390/cancers13092172
23. Li T, Kung HJ, Mack PC, et al. Genotyping and genomic profiling of non–small-cell lung cancer: implications for current and future therapies. *J Clin Oncol*. 2013;31(8):1039. doi:10.1200/JCO.2012.45.3753
24. Zhang XC, Wang J, Shao GG, et al. Comprehensive genomic and immunological characterization of Chinese non-small cell lung cancer patients. *Nat Commun*. 2019;10(1):1–12. doi:10.1038/s41467-018-07882-8
25. de Bruin EC, McGranahan N, Mitter R, et al. Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. *Science*. 2014;346(6206):251–256. doi:10.1126/science.1253462

26. Zhang J, Fujimoto J, Zhang J, et al. Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. *Science*. 2014;346(6206):256–259. doi:10.1126/science.1256930
27. Popper HH. Commentary on tumor heterogeneity. *Transl Lung Cancer Res*. 2016;5(4):433. doi:10.21037/tlcr.2016.08.07
28. Zhou Y, Zheng W, Zeng Q, et al. Targeted exome sequencing identifies mutational landscape in a cohort of 1500 Chinese patients with non-small cell lung carcinoma (NSCLC). *Hum Genomics*. 2021;15(1):1–10. doi:10.1186/s40246-021-00320-9
29. Hou H, Qin K, Liang Y, et al. Concurrent TP53 mutations predict poor outcomes of EGFR-TKI treatments in Chinese patients with advanced NSCLC. *Cancer Manag Res*. 2019;11:5665–5675. doi:10.2147/CMAR.S201513
30. Shi Y, Au JSK, Thongprasert S, et al. A prospective, molecular epidemiology study of EGFR mutations in Asian patients with advanced non–small-cell lung cancer of adenocarcinoma histology (PIONEER). *J Thorac Oncol*. 2014;9(2):154–162. doi:10.1097/JTO.0000000000000033
31. Fang B, Mehran RJ, Heymach JV, et al. Predictive biomarkers in precision medicine and drug development against lung cancer. *Chin J Cancer*. 2015;34(3):1–15. doi:10.1186/s40880-015-0028-4
32. Tang CS, Dattani S, So M, et al. Actionable secondary findings from whole-genome sequencing of 954 East Asians. *Hum Genet*. 2018;137(1):31–37. doi:10.1007/s00439-017-1852-1
33. Zheng C, Li X, Ren Y, et al. Coexisting EGFR and TP53 mutations in lung adenocarcinoma patients are associated with COMP and ITGB8 upregulation and poor prognosis. *Front Mol Biosci*. 2020;7:30. doi:10.3389/fmolb.2020.00030
34. Li S, Li L, Zhu Y, et al. Coexistence of EGFR with KRAS, or BRAF, or PIK3CA somatic mutations in lung cancer: a comprehensive mutation profiling from 5125 Chinese cohorts. *Br J Cancer*. 2014;110(11):2812–2820. doi:10.1038/bjc.2014.210
35. Blakely CM, Watkins TBK, Wu W, et al. Evolution and clinical impact of co-occurring genetic alterations in advanced-stage EGFR-mutant lung cancers. *Nat Genet*. 2017;49(12):1693–1704. doi:10.1038/ng.3990
36. Gini B, Thomas N, Blakely CM. Impact of concurrent genomic alterations in epidermal growth factor receptor (EGFR)-mutated lung cancer. *J Thorac Dis*. 2020;12(5):2883. doi:10.21037/jtd.2020.03.78
37. Mascaux C, Tsao MS, Hirsch FR. Genomic testing in lung cancer: past, present, and future. *J Natl Compr Canc Netw*. 2018;16(3):323–334. doi:10.6004/jnccn.2017.7019
38. Hoang T, Myung SK, Pham TT, et al. Comparative efficacy of targeted therapies in patients with non-small cell lung cancer: a network meta-analysis of clinical trials. *J Clin Med*. 2020;9(4):1063. doi:10.3390/jcm9041063
39. Jung A, Kirchner T. Liquid biopsy in tumor genetic diagnosis. *Dtsch Arztebl Int*. 2018;115(10):169. doi:10.3238/arztebl.2018.0169