Original Research *Feature article*

The signature of pharmaceutical sensitivity based on ctDNA mutation in eleven cancers

Shumei Zhang[1], Mu Su[2], Zhongyi Sun[2], Haibo Lu[3] and Yan Zhang[4]

[1]College of Information and Computer Engineering, Northeast Forestry University, Harbin 150040, China; [2]School of Life Science and Technology, Computational Biology Research Center, Harbin Institute of Technology, Harbin 150001, China; [3]Department of Gastrointestinal Medical Oncology, Harbin Medical University Cancer Hospital, Harbin 150081, China; [4]State Key Laboratory of Respiratory Disease, Guangzhou Medical University, Guangzhou 51000, China
Corresponding authors: Yan Zhang. Email: zhangtyo@hit.edu.cn; Haibo Lu. Email: 13613657491@126.com

Impact statement

Gene mutations are closely related to cancers and drug sensitivity and noninvasive liquid biopsy was used to detect mutations of ctDNA in plasma. In this study, we performed exon sequencing of 416 cancer-related genes for cancer primary tissue and plasma samples of 20 patients in 11 cancers and obtained the comprehensive mutation landscape. We found that liquid biopsy is reliable in place of tissue biopsy. And 31 potential unique mutation prognostic markers were screened in 7 cancer types. Moreover, the drug-mutation network (DMN) was constructed and 9 gene mutations (B-Mut-9) were confirmed that can be served as drug biomarkers in blood. Our study showed that the variation in ctDNA can be used as the biomarkers for cancer prognosis and drug efficacy prediction. This can provide a reference for clinical noninvasive testing.

Abstract

Gene mutations are closely related to cancers and drug sensitivity. Noninvasive liquid biopsy was used to detect mutations of ctDNA in plasma, which is regarded as an indicator of chemotherapy reaction. In this study, we performed exon sequencing of 416 cancer-related genes for cancer primary tissue and plasma samples of 20 patients in 11 cancers. The comprehensive mutation landscape was obtained by bioinformatics tools. In all samples, a total of 0–135 genes involved somatic mutations, and 5–209 genes involved copy number variation. APC, KRAS, and TP53 were detected as frequently mutated genes. Nineteen genes with high-frequency copy-number amplification and 59 with frequent copy-number deletions were identified. By quantitatively assessing the degree of agreement, we found that liquid biopsy is reliable instead of tissues. Besides, 31 mutation prognostic markers in 7 cancers were screened by integrating the consistent mutations and enlarging samples in TCGA. Moreover, from drug-mutation network, 25 drugs connected with 9 mutations (B-Mut-9) were obtained which can be served as drug biomarkers in blood. This was proved by further integrating the mutation information of patients in TCGA into drug-mutation network. In summary, the variation in ctDNA can be used as the biomarkers for cancer prognosis and drug efficacy prediction.

Keywords: Pharmaceutical sensitivity, drug biomarkers, drug-mutation network, prognostic markers, ctDNA, mutation

***Experimental Biology and Medicine* 2020; 245: 720–732. DOI: 10.1177/1535370220906518**

Introduction

In recent years, cancer has become an important factor threatening human health. Through epidemiological studies, experimental and clinical observations, it was found that the environment and behavior have important impact on the occurrence of human malignant tumors. Various environmental and genetic carcinogenic factors may induce cellular non-fatal DNA damage in a synergistic or orderly manner, thereby activating oncogenes and/or inactivating tumor suppressor genes; in addition, with the change of apoptotic regulatory genes and/or DNA repair gene, cells are transformed into malignancies. Therefore, the tumor is essentially a genetic disease.

Gene mutations are closely related to the production and metastasis of various tumors. Studies have shown that the BAP1 gene germline mutation involves autosomal dominant tumor susceptibility syndrome, which is associated with the tendency of uveal melanomas, melanocytic tumors, and mesotheliomas.[1–3]Chinnaiyan and his coworkers[4] performed exome sequencing of 11 metastatic breast cancer patients who underwent anti-estrogen and estrogen deprivation therapy and found that ESR1 mutations were found in six tumor patients. In addition, gene

mutations also affect the patient's sensitivity to chemotherapy and prognosis. In a population-based study, COX proportional hazards regression was employed to demonstrate that KRAS-mutant colorectal cancer patients had a significantly lower survival rate than wild-type patients after diagnosis.[5] And Lochhead *et al.*[6] demonstrated that the combined BRAF/MSI status in colorectal cancer is a tumor molecular biomarker for prognostic risk stratification.

In recent years, the methods of tumor detection and diagnosis have made many important breakthroughs, but the patients may have lost the best time to treat when tumor or metastasis was detected by these means. Therefore, the detection of biomarkers in blood as a quick, noninvasive, and effective method for early detection of tumors or metastases has an important advantage. The cell-free circulating tumor DNA (ctDNA) in plasma from peripheral blood provides an opportunity for noninvasive sampling of tumor DNA. There are about 1% of ctDNA in cfDNA and it can be distinguished from non-tumor DNA by the presence of tumor-specific mutations and copy number variations. With the rapid development of molecular experiments, trace amounts of ctDNA can be detected. This "liquid biopsy" can be used as monitoring indicators of cancer diagnose, metastasize, recurrence, and treatment by detecting the DNA including the number, chromosome changes, sequence mutations, and epigenetic changes. Studies have shown that ctDNA sequencing can be used to determine tolerance and to guide the precise administration of anti-HER2-targeted therapy in the treatment of metastatic breast cancer and it has a distinguishing effect in the prognosis of HER2-positive metastatic breast cancer.[7] Takahashi *et al.*[8] performed an analysis of before and after neoadjuvant chemotherapy in 87 patients with primary breast cancer showed that methylated ctDNA was a more sensitive biomarker than CEA and CA15-3 and could be used to monitor clinical tumor response to neoadjuvant chemotherapy. Zheng *et al.*,[9] dynamically monitoring T790M mutations in ctDNA in 318 non-small cell lung cancer patients, indicate that the EGFR T790M mutation in plasma ctDNA is associated with poor prognosis in patients with advanced NSCLC with acquired EGFR-TKI resistance. Therefore, the changes in ctDNA in plasma as noninvasive biomarkers for early diagnosis, recurrence, and therapeutic effect of cancer have a greater clinical application value than tissue biopsy.

In this study, we performed exon sequencing of 416 cancer-related genes for cancer primary tissue and plasma samples of 20 patients in 11 cancers to detect gene mutations. These 416 genes are genes related to tumor signaling pathways reported in the literature. Through the detection of mutations, we obtained a comprehensive landscape of mutations in tissues and plasma of these patients. And through the comparison among the patients' tissues and their plasma, we found the common mutations among the same kind of cancers and the most recurrent variants among the different tumors. The consistency of tissue and plasma mutations, as well as the degree of substitution of plasma to tissue for copy number variations, was evaluated by comparing the tissue and plasma for each patient and this can be used as a basis for mutations detection using ctDNA instead of tissue biopsy. Besides, the molecular biomarkers for prognostic risk stratification in mutation level were screened by survival analysis and this provides important guidance for clinicians to monitor the patient's condition and to switch treatment protocols in time. In addition, our study also analyzes the efficacy of the mutations in the tumor to the drugs through a drug-mutation network. And the network was used to find the cancer mutations that cause effective or invalid to drugs, which provides a reference for cancer drug therapy.

Materials and methods

Data acquisition

In this study, 20 cases of cancer patients were obtained from the Tumor Hospital of Harbin Medical University, including six patients with COAD (colon adenocarcinoma), four patients with LUAD (lung adenocarcinoma), two patients with SARC (sarcoma), one case of LUSC (lung squamous cell carcinoma), one case of PAAD (pancreatic adenocarcinoma), one case of READ (rectal adenocarcinoma), one case of STAD (stomach adenocarcinoma), one case of ESCA (esophageal cancer), one case of CGSC (cecal gland squamous cell carcinoma), one case of BRCA (breast cancer), and one case of UCEC (uterine corpus endometrial cancer). Written informed consent was obtained from the patients for publication, and study was approved by the Harbin Medical University Cancer Hospital Ethics Committee. Besides, the Affymetrix Genome-Wide Human SNP 6.0 array data and somatic mutation (SNPs and small INDELs) data generated on IlluminaGA system of 10 cancers were downloaded from TCGA (The Cancer Genome Atlas) data portal[10] with their clinical follow-up information. The copy number amplification variations were defined as $\log^2$ ratio >0.25, and the copy number deletion variations were defined as $\log^2$ ratio <-0.25.

ctDNA extraction and library construction

Whole blood (5 mL) is collected by EDTA blood collection tubes then centrifuged within 1 h of collection at $1800 \times g$ for 10 min at $\geq 4^{\circ}$C or RT to remove blood cells. The supernatant containing the plasma is removed with special care taken as to not disturb the buffy coat. This is then centrifuged at $16{,}000 \times g$ for 10 min to remove any remaining cells. ctDNA was extracted from 2 mL plasma, by digestion in 100 μL proteinase K buffer for 10 min at 37°C followed by purification with the NucleoSpin Plasma XS kit with modified protocols. The purified ctDNA is quantified by a Picogreen fluorescence assay using the provided lambda DNA standards (Invitrogen). Then, library construction with the KAPA Hyper DNA Library Prep Kit, containing mixes for end repair, dA addition and ligation, was performed in 96-well plates (Eppendorf). Dual-indexed sequencing libraries are PCR amplified for four to seven cycles.

Hybrid selection and ultra-deep next generation sequencing of ctDNA

The 5′-biotinylated probe solution is provided as capture probes, the baits target 416 cancer-related genes (Supplementary Table S1); 1 μg of each ctDNA-fragment sequencing library is mixed with 5 μg of human Cot-1 DNA, 5 μg of salmon sperm DNA, and 1 unit adaptor-specific blocker DNA in hybridization buffer, heated for 10 min at 95°C, and held for 5 min at 65°C in the thermocycler. Within 5 min, the capture probes are added to the mixture, and the solution hybridization is performed for 16–18 h at 65°C. After hybridization is complete, the captured targets are selected by pulling down the biotinylated probe/target hybrids using streptavidin-coated magnetic beads, and off-target library is removed by washing with wash buffer. The PCR master mix is added to directly amplify (six to eight cycles) the captured library from the washed beads. After amplification, the samples are purified by AMPure XP beads, quantified by qPCR (Kapa), and sized on bioanalyzer 2100 (Agilent). Libraries are normalized to 2.5 nM and pooled. Deep sequencing is performed on Illumina HiSeq 4000 using PE75 V1 Kit. Cluster generation and sequencing are performed according to manufacturer's protocol. The materials used for tissue sequencing are paraffin-embedded tissue sections.

Sequence alignment and processing

Base calling was performed using bcl2fastq v2.16.0.10 (Illumina, Inc.) to generate sequence reads in FASTQ format (Illumina 1.8+ encoding). Quality control (QC) was applied with Trimmomatic.[11] High-quality reads were mapped to the human genome (hg19, GRCh37 Genome Reference Consortium Human Reference 37) using modified BWA aligner 0.7.12[12] with BWA-MEM algorithm and default parameters to create SAM files. Picard 1.119 (http://picard.sourceforge.net/) was used to convert SAM files to compressed BAM files which were then sorted according to chromosome coordinates. The Genome Analysis Toolkit[13] (GATK, version 3.4–0) was modified and used to locally realign the BAMs files at intervals with indel mismatches and recalibrate base quality scores of reads in BAM files.[14]

Variants detection and annotation

For paired tumor-normal libraries, we run the SAMtools[15] mpileup command to convert them to mpileup files. And then VarScan 2 software[16] was used to identify mutations, including single nucleotide variants (SNVs) and short insertions/deletions (indels) and copy number variants. For SNVs and indels, they were identified with minimum variant allele frequency threshold set at 0.01 and *P*-value threshold for calling variants set at 0.05 to generate variant call format (VCF) files. Called SNP variants and indels were annotated and classified using ANNOVAR.[17] Variants were filtered using data from cosmic70, clinvar, dbSNP 138 (http://www.ncbi.nlm.nih.gov/snp), and the 1000 Genomes Project (http://www.1000genomes.org/). Gene annotations were made against the UCSC KnownGenes database. PolyPhen2[18] and SIFT (sift.jcvi.org) were used to evaluate the possibly damaging effects of single amino-acid substitutions on the expression of the proteins of these genes. Finally, only the exonic mutations were used for further analyses.

For the copy number variants, we followed the recommended workflow provided in the VarScan project page in SourceForge website (http://varscan.sourceforge.net/copy-number-calling.html): (1) Run VarScan "copynumber" routine on normal and tumor mpileup files generated by SAMtools, (2) Run VarScan "copyCaller" to adjust for GC content and make preliminary calls on the results from step (1), (3) Apply circular binary segmentation (CBS) from DNAcopy package[19] to segment the raw regions and identify significant change-points, (4) Visualize the results and recenter if necessary. If baseline is adjusted, then repeat steps (3) and (4) again. In addition, the R package "CNTools" and "cghMCR" are used to integrate the output of DNAcopy into gene-level copy number variation profiles for further analysis. During this process, we dropped data on the X and Y chromosomes (there are samples from both male and female patients in the data set).

Evaluation of replacement degree of plasma to tissue mutation

The significance of the coincidence of mutant genes in tumor tissue and plasma was assessed based on hypergeometric test, and the formula is

$$p = 1 - \frac{\binom{T}{C}\binom{N-T}{P-C}}{\binom{N}{P}} \tag{1}$$

where N is the total number of genes, T is the number of mutated genes in tissue, P is the number of mutated genes in plasma, and C is the number of mutated genes that are common in tissue and plasma. In addition, the degree of plasma-to-tissue substitution was also assessed, using the enrichment imbalance strategy proposed by Jiang *et al.* in 2012,[20] which refers to the statistical method of linkage disequilibrium in genetics. It is a method to measure the degree of enrichment in gene enrichment analysis and can be used to describe the non-random association of two sets. If there is no association between the gene set T and the gene set P, we usually think that the two sets are independent of each other. Based on the principle of independence, we will get the following expression

$$p(T \times P) = p(T) \times p(P)$$

Then, if there is non-random association between T and P, we will get $p(T \times P) \neq p(T) \times p(P)$. Therefore, we define the coefficient sd as follows

$$sd = p(T \times P) - p(T) \times p(P) \tag{2}$$

where sd represents the degree of plasma-to-tissue substitution, $p(T \times P)$ is the probability of the number of mutated genes that are common in tissue and plasma, $p(T)$ is the probability of the number of mutated genes in tissue, and $p(P)$ is the probability of the number of mutated genes in plasma. The method also standardized the enrichment imbalance metric

$$SD = \frac{sd}{sd_{max}} \tag{3}$$

where

$$sd_{max} = \begin{cases} p(T) \times \Big(1 - p(P)\Big), \; sd > 0 \quad \text{and} \quad N_T < N_P \\ p(P) \times \Big(1 - p(T)\Big), \; sd > 0 \quad \text{and} \quad N_T > N_P \\ p(T) \times p(P), \; sd < 0 \end{cases} \tag{4}$$

SD range from -1 to $+1$, and the relationships between the gene set T and the gene set P under different SD intervals are shown in Table 1.

Table 1. The relationships between P and T under different SD intervals.

SD	Relationship between T and P
SD = 1	T ⊂ P or P ⊂ T, in the two cases, the gene set P and the gene set T have the strongest association.
0 < SD < 1	The number of gene set P annotated in gene set T is more than random. A larger SD indicates a higher degree of enrichment.
SD = 0	The gene set P and the gene set T are independent.
−1 < SD < 0	The number of gene set P annotated in gene set T is less than random. A smaller SD indicates a higher degree of depletion.
SD = −1	The gene set P and the gene set T are incompatible.

Drug-mutation background network construction, analysis, and visualization

The mutation-gene-drug relations were obtained from OncoKB database and VarDrugPub database. For the VarDrugPub database, only the mutation-gene-drug relations with clear gene and amino acid changes were used. Besides, we also collected 175 relationships between gene mutations and drug efficacy from the published literatures. These 175 relationships and the two public databases were integrated and unique 3233 relationships among gene mutations and drug efficacy were obtained. Here we only took the mutations occurred in the 416 genes in the present study and finally 1557 relationships were obtained. These 1557 relationships were used to construct a drug-mutation background network.

The network was constructed for analysis and visualization using cytoscape 2.8.2[21] (http://www.cytoscape.org/), an open source software platform for visualizing molecular interactions and biological pathways, and integrating these networks with annotation information, gene expression profiles, and other types data.

Survival analysis and statistical analysis

The prognostic of gene variants on overall survival (OS) was examined using the Kaplan–Meier estimates and Cox proportional hazard ratio (HR) regression analyses. Log-rank test was used to assess the statistical significance of the difference. All of these were performed using the R package "survival". The clustering diagram and heat maps were implemented using the heatmap.2 function in the R package "gplots", using the hierarchical clustering algorithm and the European distance similarity measure.

Functional analysis of gene sets

g: Profiler,[22] a web server for functional interpretation of gene lists was used to perform gene enrichment analysis of Gene Ontology, KEGG pathway, and Human Protein Atlas of genes that were mutated in tumors.

Data availability

The raw data are available upon request. The authors affirm that all data necessary for confirming the conclusions of the article are present within the article, figures, and tables.

Results

Overview of the studied samples

The tissues and plasma samples from 20 cancer patients were obtained from the Tumor Hospital of Harbin Medical University, including six patients with COAD, four patients with LUAD, two patients with SARC, one case of LUSC, one case of PAAD, one case of READ, one case of STAD, one case of ESCA, one case of CGSC, one case of BRCA, and one case of UCEC (Table 2). The ages of these patients are between 29 and 75 years old. Most of them occurred in cancer metastasis, and the metastatic sites include liver, bone, brain, lung, supraclavicular lymph node, pleural effusion, pelvic peritoneum, left groin, side of the lumbar spine as well as local recurrence. One of the patients was diagnosed as II TNM stage, nine were diagnosed as III TNM stage, and 10 were diagnosed as Ⅳ TNM stage. The patients above were treated with chemotherapy, endocrine therapy, radiotherapy, TKI target treatment, and surgery at least. However, they took place in recurrence or metastasis. In order to assess the next choice of treatment options, we collected blood from patients and detected abnormal changes in the 416 genes.

The consistent somatic point mutations detection in tissues and plasma

First, we called the somatic mutations which include SNVs and indels in the tissues and plasma, and the leukocytes in the matched blood were used as normal controls. The number of candidate variants per sample ranged from 0 to 364 corresponding to 0–135 genes (Figure 1).

Table 2. The clinical information of 20 cancer patients.

number	Abbreviated name	Gender	age	Diagnosis	Site of metastases	TNM stage	Treatment
1	BRCA1	Female	40	Breast cancer	Liver, bone	Ⅵ	Chemotherapy, endocrine therapy
2	CGSC1	Female	56	Cecal gland squamous cell carcinoma	Liver	III	Surgery and chemotherapy
3	COAD1	Male	50	Colon adenocarcinoma	lung	Ⅵ	Chemotherapy
4	COAD2	Female	48	Colon adenocarcinoma	NA	III	Surgery and chemotherapy
5	COAD3	Male	43	Colon adenocarcinoma	Liver	Ⅵ	Surgery and chemotherapy
6	COAD4	Male	75	Colon adenocarcinoma	Liver	Ⅵ	Chemotherapy
7	COAD5	Male	67	Colon adenocarcinoma	liver	Ⅵ	Surgery and chemotherapy
8	COAD6	Male	73	Colon adenocarcinoma	Liver	Ⅵ	Surgery and chemotherapy
9	ESCA1	Male	53	Esophageal cancer	Supraclavicular lymph node	III	Radiotherapy and chemotherapy
10	LUAD1	Male	29	Lung adenocarcinoma	Liver, brain, bone	Ⅵ	TKI target treatment
11	LUAD2	Female	48	Lung adenocarcinoma	NA	III	Surgery and chemotherapy
12	LUAD3	Female	60	Lung adenocarcinoma	NA	II	TKI target treatment
13	LUAD4	Male	64	Lung adenocarcinoma	Pleural effusion	III	TKI target treatment
14	LUSC1	Male	72	Lung squamous cell carcinoma	NA	Ⅵ	Radiotherapy and chemotherapy
15	PAAD1	Female	62	Pancreatic adenocarcinoma	Pelvic peritoneum	III	Chemotherapy
16	READ1	Male	56	Rectal adenocarcinoma	Local recurrence	III	Radiotherapy and chemotherapy
17	SARC1	Female	60	Clear cell carcinoma of left heel	Left groin, side of the lumbar spine	Ⅵ	Chemotherapy
18	SARC2	female	47	Leiomyosarcoma of the abdominal cavity	NA	III	Surgery and chemotherapy
19	STAD1	Male	65	Stomach adenocarcinoma	Bone	Ⅵ	Surgery and chemotherapy
20	UCEC1	Female	67	Uterine corpus endometrial carcinoma	NA	III	Surgery and chemotherapy

For indels, there are 5 frameshift insertion structural variations, 11 frameshift deletion variations, 1 nonframeshift insertion variations, 7 nonframeshift deletion variations, and 1 stopgain in total. For SNVs, there are 400 silent variations and 534 nonsilent variations in total (Figure 1(e)). All of the variants were mapped to 216 unique genes and the number of consistent somatic gene mutations detection in tissues and plasma for all of the patients is shown in Table 3. In Figure 1(a), we deleted the mutations that existed in only one sample and the mutations that existed only in the tissue of LUSC1 and SARC2 or only in the tissue of LUAD3 and SARC2 to ensure that the figure was fit for visualization. These complete mutations information is shown in the Supplementary Table S2.

For SNVs, the most recurrent mutations occurred in APC, KRAS, and TP53. They were all important genes in cancer occurrence and development which have been well documented. And they have been confirmed frequently mutated in the colorectal cancer.[23] The adenomatous polyposis coli (APC) tumor suppressor is the most commonly mutated in colorectal cancers[24] which were consistent with our results. KRAS type has been shown independently and significantly associated with survival in the multivariate analysis.[25] TP53 polymorphism was a common crucial gene in gastric cancer and colorectal cancer and plays important role in their development.[26,27] Four of the six colon cancer patients had TP53 mutations detected both in their tissues and plasma, and three of them had KRAS mutations. Among these three, there are two APC mutations detected both in their tissues and plasma (Figure 1(a)). Nearly all of the variants in these three genes were predicted pathogenic by cosmic or SIFT and PolyPhen2 scores. The mutations occurred in COAD2 patient were minimal among these six colon cancers, only two genes. This patient was a relatively young, low TNM staging patient compared to the other five colon cancer patients and there was no distant metastasis which represented a relatively good condition. Besides, it is worthwhile to note that most of the mutations occurred in plasma can also be detected in the tissue in the global landscape, in other words, the mutations occurred in pairs. This indicates that using plasma instead of tissue detection is achievable. For indels, it was a very sparse matrix (Figure 1(c)). The TP53 mutation occurred both in the tissue and plasma of esophageal cancer patient, and APC mutation occurred both in the tissue and plasma of a colon adenocarcinoma patient which was consistent with the previous studies. All of these indel mutations are nonsilent mutations (Figure 1(f)).

In addition, we also concern the functions of these somatic mutations. An enrichment analysis was conducted for these 216 mutated genes using the web server for functional interpretation of gene lists named g:Profiler. The result showed that they were significantly enriched in the processes and pathways associated with mutation and cancer, including production of small RNA involved in gene silencing by RNA, somatic diversification of immune receptors via somatic mutation, somatic hypermutation of immunoglobulin genes, colorectal cancer, lung cancer, pancreatic cancer, pathways in cancer, etc. (Figure S1(a)). Not surprisingly, their abnormalities can lead to the occurrence and development of cancer.

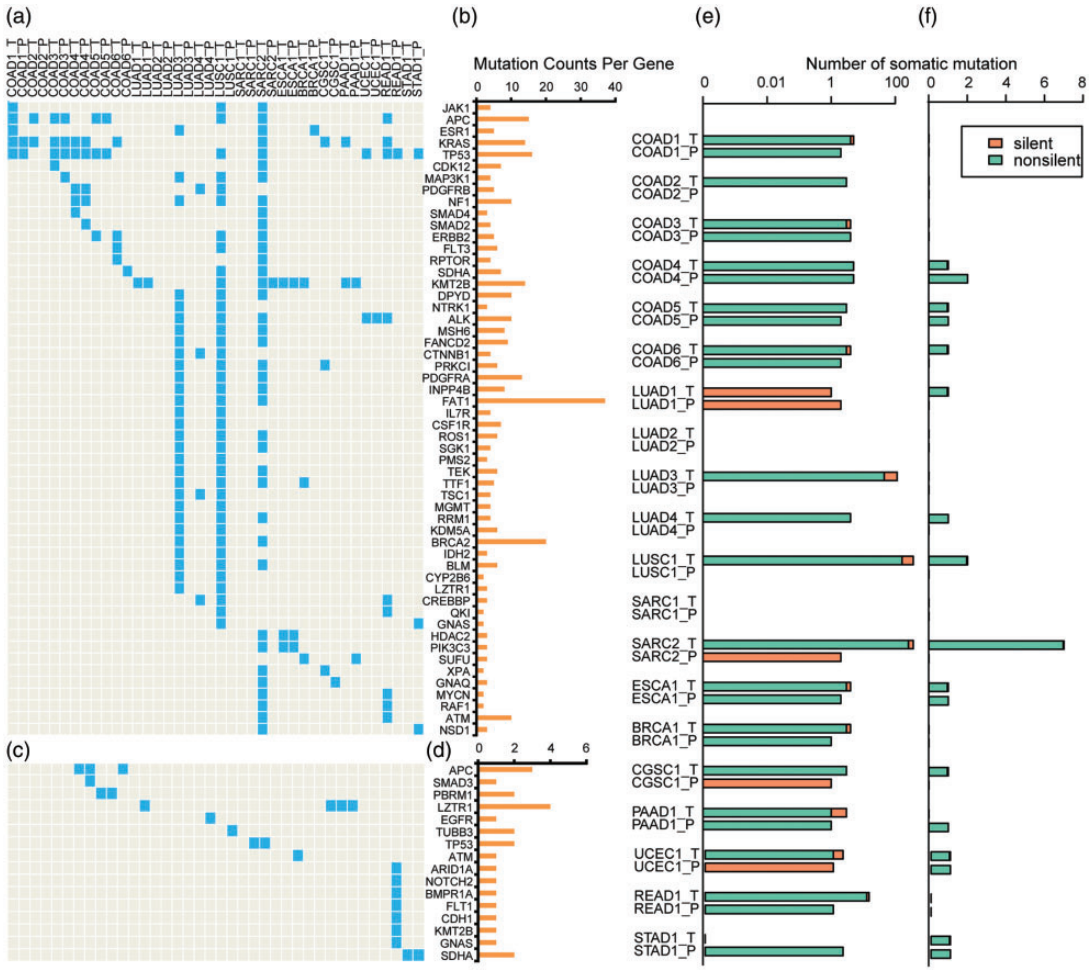

Figure 1. Somatic mutation landscapes in the tissues and plasma of all patients. (a) Mutation landscape map for SNVs, each blue block represents the occurrence of mutation in this gene in the sample. In this figure, we deleted the mutations that existed in only one sample and the mutations that existed only in the tissue of LUSC1 and SARC2 or only in the tissue of LUAD3 and SARC2 to ensure that the figure was fit for visualization. (b). The total number of mutations that occurred on each gene in panel A. (c) Mutation landscape map for indels, each blue block represents the occurrence of mutation in this gene in the sample. (d) The total number of mutations that occurred on each gene in panel C. (e) The counts of silent variations and nonsilent variations for SNVs in each sample. (f) The counts of silent variations and nonsilent variations for indels in each sample. (A color version of this figure is available in the online journal.)

The landscape of copy number variants among all patients in tissues and plasma

The copy number variation spectrum of 387 autosomal genes was obtained by applying VarScan software and "CNTools" and "cghMCR" R packages (Materials and Methods). The global heat map of the segment.mean is shown in Figure 2(a). The segment.mean refers to the logarithm of the ratio of the copy number of tumor sample to the normal control which is an indicator of the change in the copy number. In general, the copy number is considered to be amplification if the segment.mean is greater than 0.25 and the copy number is considered to be deletion if the segment.mean is less than −0.25 (the default threshold for most software). Here we refer to these two thresholds to identify the genes that own the copy number amplification and the copy number deletion variation in each sample. The number of genes with copy number variants per sample ranged from 5 to 209. Specifically, the number of genes that undergo copy number amplification variation ranged from 0 to 142 and the number of genes that own copy number deletion variation ranged from 2 to 134. The specific numbers of copy number variations in tissue, plasma and the consistence in tissue and plasma for each specific patient are shown in Supplementary Table S3. The consistent CNVs in tissue and plasma were seen as the candidate biomarkers in blood. It is noteworthy that even among individuals with the same cancer, the heterogeneity of copy number variation is greater than somatic mutations.

The functional analysis of genes with copy number variation showed that they were mainly involved in the GO terms of regulation of cell population proliferation, apoptotic process, cell death, programmed cell death, immune system development, response to chemical and regulation of cell differentiation, and many cancer-related pathways such as colorectal cancer, hepatocellular carcinoma, prostate cancer, gastric cancer, pancreatic cancer, breast cancer and pathways in cancer, etc. This suggests that variations in these genes can make a significant impact on the development of cancers.

Liquid biopsy has an important significance when used to discover biomarkers of tumor metastasis, but this must

Table 3. The number of consistent somatic gene mutations detection in tissues and plasma for 20 patients.

	SNV			Indel		
	Tissue	plasma	Consistent	Tissue	plasma	Consistent
COAD1	5	2	2	0	0	0
COAD2	3	0	0	0	0	0
COAD3	4	4	3	0	0	0
COAD4	5	5	4	1	2	1
COAD5	3	2	2	1	1	1
COAD6	4	1	0	1	0	0
LUAD1	1	2	1	1	0	0
LUAD2	0	0	0	0	0	0
LUAD3	71	0	0	0	0	0
LUAD4	4	0	0	1	0	0
LUSC1	89	0	0	1	0	0
SARC1	0	0	0	0	0	0
SARC2	135	1	1	7	0	0
ESCA1	4	3	3	1	1	1
BRCA1	4	1	0	0	0	0
CGSC1	3	1	0	1	0	0
PAAD1	2	2	1	0	1	0
UCEC1	2	1	1	1	1	1
READ1	11	1	1	0	0	0
STAD1	0	3	0	1	1	1

be based on the fact that ctDNA in plasma and primary tumor tissue has a certain consistency in tumor biomarkers. Therefore, we still focused on the reproducibility of the copy number variations in plasma for that in tumor tissues. By calculating SD of the gene set P in the plasma with copy number variation and the gene set T in the tissue with copy number variation (Materials and Methods), we found that 73% (24 out of 33) of the SDs that can be calculated imply that the association between tissue and plasma copy number amplification or deletion variations is non-random. And the correlations between almost all of the tissue and plasma gene pairs with an SD greater than 0.1 (19 out of 24) are significant (hypermetrological test, formula 1), with only one exception which is slightly greater than the 0.05 level (Supplementary Table S3). This suggests that the liquid biopsy used to speculate variations occur in tumor tissues is reliable.

Next, we concerned the recurrent copy number variation genes among all patients. Genes with frequent copy number variation among multiple cancers may have the potential as markers for same treatment for multiple cancers (homotherapy for heteropathy). The recurrent copy number amplification variation genes were defined in this study as the genes that have copy number amplification variation in more than 10 samples (including 10, 25%) and that have copy number deletion variation in no more than 10 samples. Similarly, we defined the recurrent copy number deletion variation genes as the genes that have copy number deletion variation in more than 10 samples and that have copy number amplification variation in no more than 10 samples. This strict definition standard guarantees the accuracy of marker identification. Finally, 19 recurrent copy number amplification variation genes and 59 recurrent copy number deletion variation genes were identified (Figure 2(b)).

The gene ontology and KEGG pathway enrichment analysis showed that the recurrent copy number amplification variation genes were mainly involved in the positive regulation of metabolic, biosynthetic, apoptotic, developmental and cellular processes, and many cancer-related pathways (Figure S1(b)). While the recurrent copy number deletion variation genes were mainly involved in the processes of response to stimulus, chemical stimulus and stress, regulation of cell proliferation, cell differentiation, cell death, and many cancer-related pathways (Figure S1(c)). The recurrent copy number amplification variation genes and the recurrent copy number deletion variation genes were involved in different aspects of biological processes but were all related to cancer. They may be the potential therapeutic marker for adjuvant chemotherapy for multiple cancers.

The prognostic biomarkers screening for cancers

The screening of cancer prognostic biomarkers provides important guidance for clinicians to understand the patient's condition and to switch treatment protocols in time. Therefore, we investigated whether the consistent mutations in patient tissue and plasma in our study could be used as markers for cancer prognosis. We obtained the mutation data for SNPs and small INDELs from cancer patients matching the cancer type in our study from the TCGA database, as well as segments data for copy number variation. Finally, we obtained patient data for nine cancers (except CGSC which was no data in TCGA and PAAD which no common mutation detected in our study). In addition, we also obtained survival data for these cancer patients from the TCGA database. For the copy number segments data, the R package "CNTools" and "cghMCR" were also used to integrate them into gene-level copy number variation profiles and the copy number is considered to be amplification if the segment.mean is greater than 0.25 and the copy number is considered to be deletion if the segment.mean is less than −0.25 as mentioned before.

For each consistent mutation in tissue and plasma detected in our study, we examined the difference in survival between patients with this mutation in the corresponding cancer patients in the TCGA database and those who did not. The statistical significance was tested using the log-rank test method. Finally, we obtained 31 unique gene markers as prognostic biomarkers for seven cancer types (Supplementary Table S4). Specifically, there are 10 prognostic markers for BRCA, 15 prognostic markers for COAD, 3 prognostic markers for ESCA, 1 prognostic marker for LUAD, 1 prognostic marker for READ, and 2 prognostic markers for SARC, and 2 prognostic markers for UCEC. Among these markers, the loss of GSTM1 gene serves as a common prognostic marker for BRCA, COAD, and READ, and the loss of NRAS gene serves as a common prognostic marker for BRCA and COAD (Figure 3).

GSTM1 genetic variants have been proved associated with the survival in breast cancer[28,29] and many other cancers.[30,31] Bansal *et al.*[28] have proposed that GSTM1 gene deletion was significantly correlated with breast cancer in

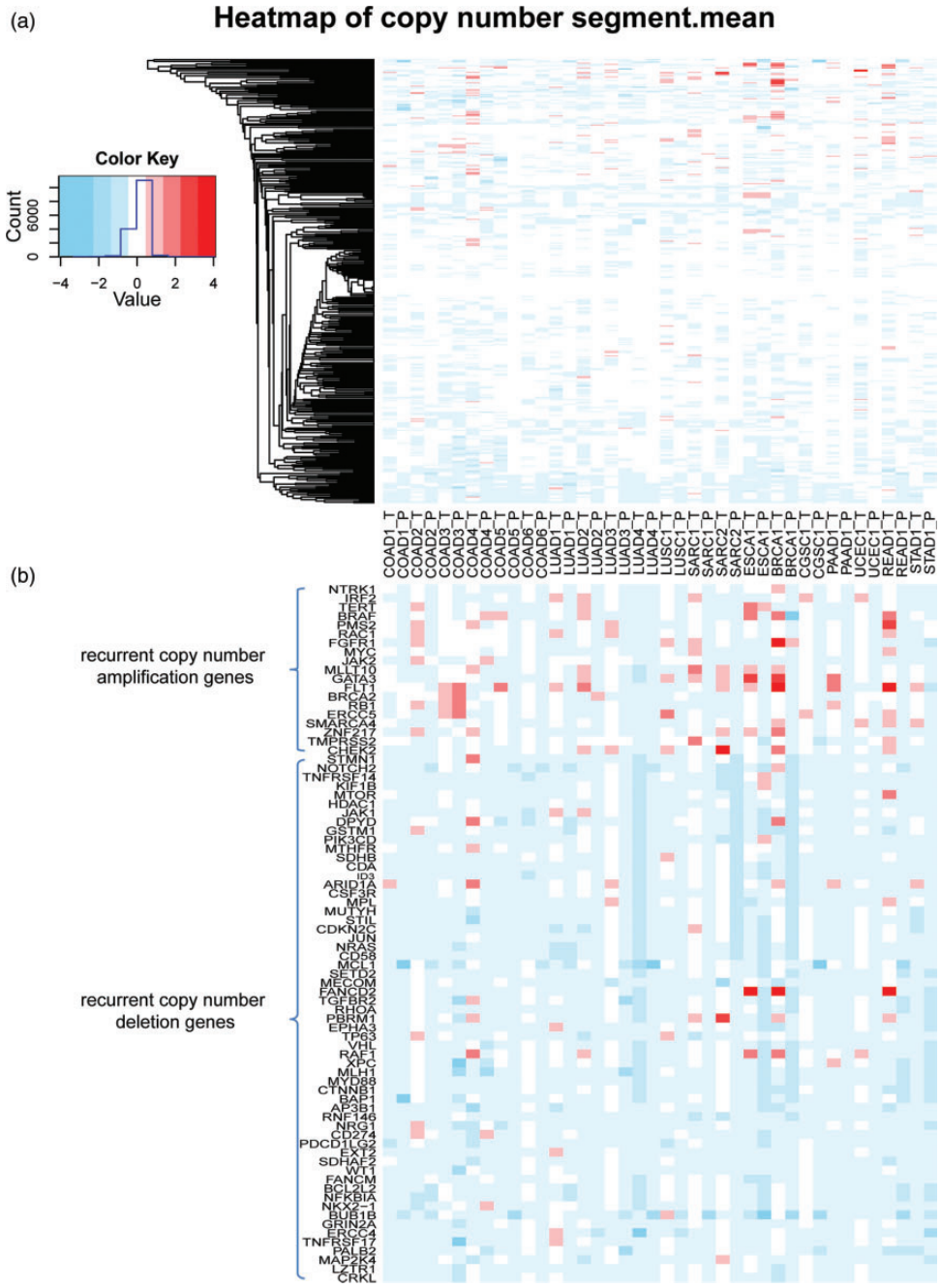

Figure 2. Heat maps of copy number segment.means. (a) The copy number variation landscape for all patients in global. (b) The copy number variation landscape for 19 recurrent copy number amplification variation genes and 59 recurrent copy number deletion variation genes. (A color version of this figure is available in the online journal.)

2015,[28] and our study further illustrated that GSTM1 gene deletion was associated with poor prognosis in breast cancer compared with those without this mutation. And NRAS mutation has been studied to have prognostic values in colorectal cancer patients.[32] This implies that these cancers share some similar patterns, consistent with previous studies. Our study further proves and complements the findings of previous studies. It is worth noting that almost all of these gene mutations are risk factors of survival (HR <1), only KIF1B gene deletion (Figure 3(f)) was exceptional which was a protective factor (HR $=1.9322$, 95% CI 1.0603–3.5210, $P=0.0287$)).

The construction of drug-mutation network and analysis

In order to predict drugs that correspond to gene mutations that occur in the plasma, the drug-mutation network (DMN) was constructed. First, a drug-mutation background network was constructed using the 1557 drug-mutation relations obtained from OncoKB database, VarDrugPub database, and published literatures as described in the section of materials and methods. In the network, there was an edge represents sensitive or resistant between the gene mutation and the corresponding drug.

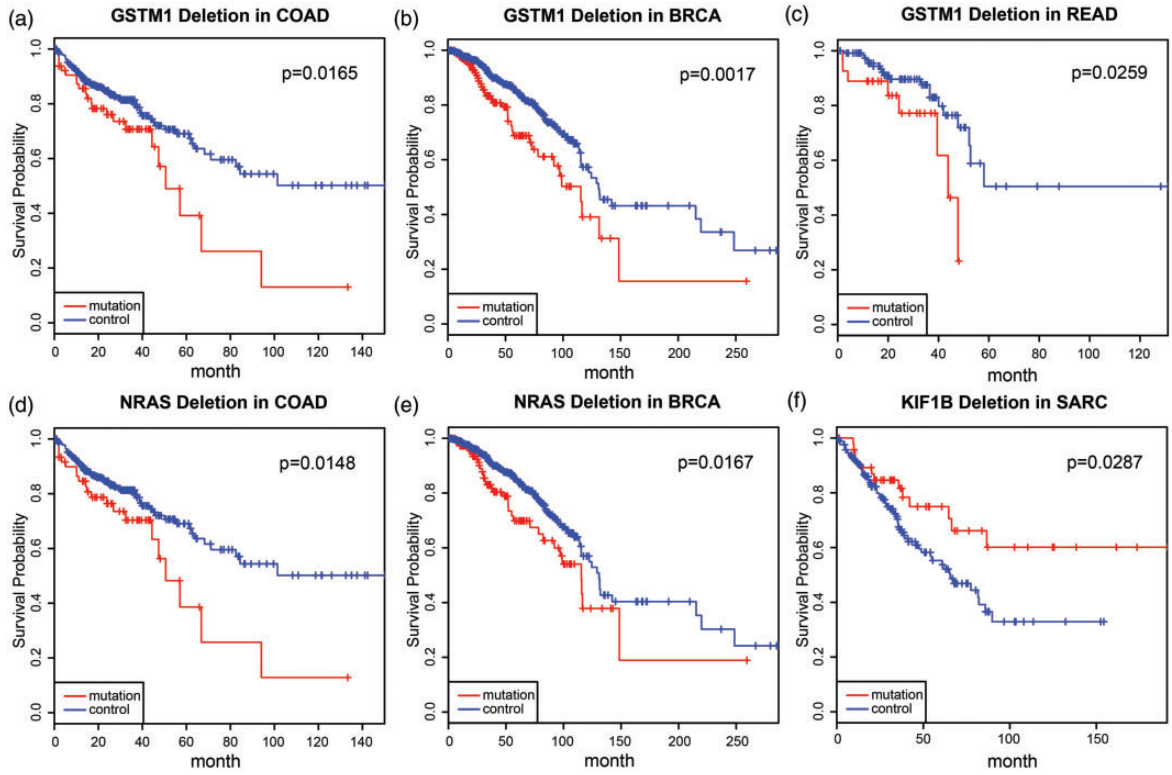

Figure 3. Kaplan–Meier plots for patients with gene mutation and who did not in cancers. (a) Kaplan–Meier plot for GSTM1 deletion group and the control group in COAD. (b) Kaplan–Meier plot for GSTM1 deletion group and the control group in BRCA. (c) Kaplan–Meier plot for GSTM1 deletion group and the control group in READ. (d) Kaplan–Meier plot for NRAS deletion group and the control group in COAD. (e) Kaplan–Meier plot for NRAS deletion group and the control group in BRCA. (f) Kaplan–Meier plot for KIF1B deletion group and the control group in SARC. (A color version of this figure is available in the online journal.)

Eventually, we got a network which includes 1004 nodes (658 gene mutations and 346 drugs) and 1557 edges (Figure 4(a)).

The most important key nodes (the nodes with the largest degree) in the network were the BRAF V600E and EGFR T790M mutation. They were related with 91 and 62 drugs, respectively. And they were famous cancer-derived oncogenic mutation in multiple cancer types. Oncogenic BRAF V600E substitutions are observed primarily in melanoma, colon cancer, and non-small cell lung cancer, but have been identified in multiple tumor types. Klempner *et al.*[33] described the recurrent BRAF mutations in advanced high-grade colorectal neuroendocrine tumors and identify a BRAF alteration frequency of 9% in 108 cases. Besides, EGFR mutation was also an important variation which exerts synergistic antitumor interactions with several cytotoxic drugs.[34]

For the 20 patients, we take all of the common mutations including SNVs, indels, and CNVs in tissue and plasma ctDNA in each of them as the candidate mutations. The common mutations can be regarded as stable mutations occurred in patients. There were 32 common SNVs, 11 common indels, and 258 common CNV genes in total for 20 patients. These mutations were mapped to 151 unique genes. Then these gene mutations were taken as seed nodes into the network and along with their first neighbors were extracted to build a subnet (Figure 4(b)). In the subnet, there were 9 gene mutations in our study connected by 25 drugs mapped into the mutations in the network. These nine gene mutations were corresponding to five cancer types in our study. In particular, TP53 R273C in READ, TSC2 Deletion in LUAD, CCND1 Amplification in ESCA, FGFR1 Amplification in BRCA, KRAS G12D, TP53 V216M, TP53 R248Q, FLT1 Amplification, and FLT3 Amplification in COAD. These nine mutations can be served as drug biomarkers in the blood (we defined B-Mut-9). When the liquid biopsy detects these mutations, they represent mutations in the corresponding tumor tissues. It is worth noting that the TP53 and KRAS genes were the most recurrent mutations found in the previous analysis. Three of the nine drug markers were mutations that occurred in TP53 gene. The TP53 gene is the most closely related gene to human cancers that has been discovered so far and is an important gene that prevents normal cells from becoming cancerous cells. The 25 drugs connected with them were the effective drugs related to these cancers. In addition, the drugs which connected with gene mutations in multiple cancer types may be used as spectral drugs for multiple cancers. For example, doxorubicin can be used to treat both READ and COAD. This implement the idea of different diseases treated with same treatment regimen. The method in our study can provide a reference for clinicians in selecting medication regimens.

The drug efficacy prediction for TCGA patients

In order to take advantage of the established network and the B-Mut-9, we have found to predict the efficacy of drugs for patients with certain gene mutations in plasma.

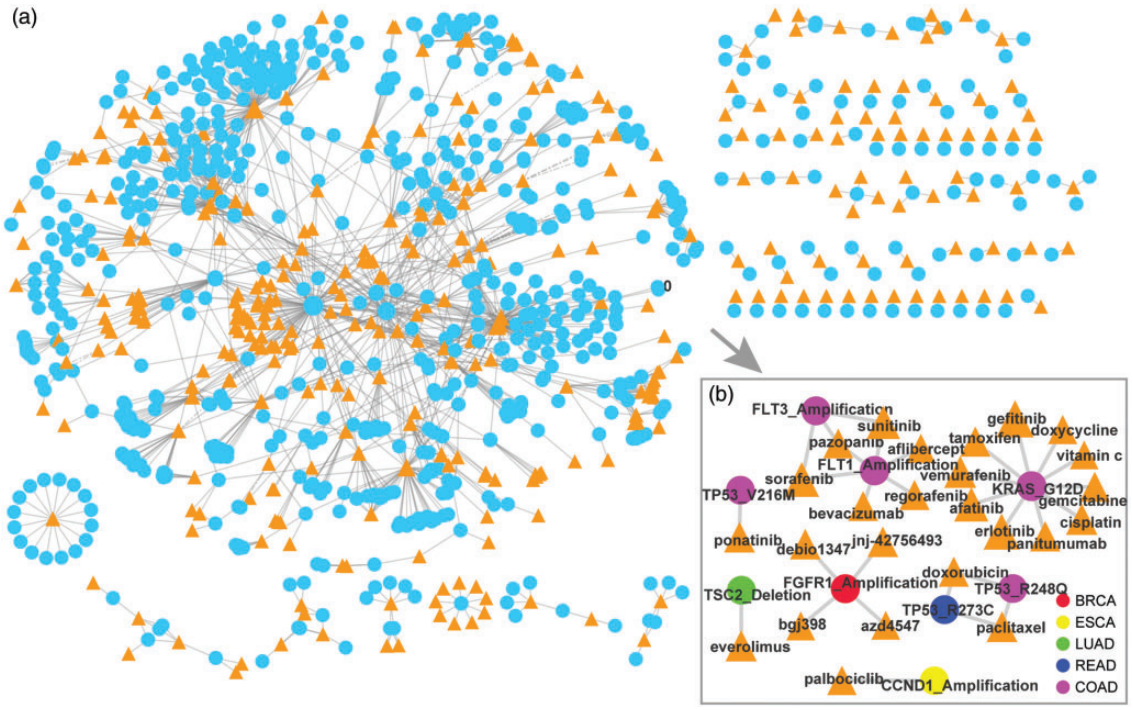

Figure 4. The drug-mutation network (DMN). (a) Drug–mutation network which includes 1004 nodes and 1557 edges. Blue circles represent mutations, orange triangles represent drugs. The sizes of the nodes are represented by the sizes of the nodes' degrees. Solid lines represent drug sensitivity and dashed lines represent tolerance. (b) The DMN subnet which was extracted from DMN through the common mutations including 32 common SNVs, 11 common indels, and 258 common CNV genes in tissue and plasma ctDNA as the seed nodes mapped into the network and extracted along with their first neighbors. (A color version of this figure is available in the online journal.)

We obtained the large sample dataset from TCGA database. The mutation data for SNPs and small INDELs and the segments data for copy number variation of nine cancers from the TCGA database used before were also employed for this analysis. Here we use the mutations detected in these patient tissues to hypothesize their plasma mutations. An edge was generated between one patient and one gene mutation in the drug-mutation subnetwork (Figure 4(b)) if this mutation occurred in this patient to construct a drug-mutation-patients network (DMPN) (Figure 5). To be more specific, the DMPN contained the B-Mut-9 we found above and the drugs as well as the patients who have been detected for these mutations. The numbers of mutations of B-Mut-9 detected in these nine cancers were 8 for BRCA, 8 for COAD, 8 for ESCA, 9 for LUAD, 7 for LUSC, 8 for READ, 6 for SARC, 8 for STAD, and 8 for UCEC. The sensitive drugs that were originally linked to the gene mutation were considered to be effective drugs for this mutation in the specific patient. In other words, the drugs those are effective for patients were obtained by two step neighborhood through the gene mutations.

Through the drug-mutation-patients network, we obtained drugs that are suitable for patients with these nine cancer types through the drug biomarkers B-Mut-9 (Supplementary Table S5). This provides a certain reference for clinicians in choosing drug plans. If a new tumor patient arrived and the biomarkers in B-Mut-9 are detected in their plasma, then clinicians can consider using these drugs to treat them. The network in our research can also be transplanted to other studies to screen for drugs suitable for cancer patients.

In addition, in order to verify the accuracy of the drugs for treating cancers obtained from the network, we compared the predicted drugs in the network for patients of each cancer with the patient's medication information in the clinical information tables of TCGA patients. We found a total of 92 pairs of patient medication information from 9 cancers that corresponded to the network. Specifically, there are 14 pairs in BRCA, 8 pairs in COAD, 9 pairs in ESCA, 20 pairs in LUAD, 4 pairs in LUSC, 4 pairs in READ, 8 pairs in SARC, 5 pairs in STAD, and 20 pairs in UCEC. Among these 92 pairs of relationships, there were 30 pairs whose therapeutic effects were absent. In the remaining 62 pairs of relationships, there were 44 pairs (71%) whose treatment effects showed effective. This result indicates that the suitable prediction of our network of drugs for cancer patients through genetic mutations is effective.

Discussion

Liquid biopsy as a non-invasive monitoring indicators of cancer diagnose, metastasize, and recurrence has been attracting more and more attention. The detection of biomarkers in blood has an important advantage for early detection of tumors or metastases and relapse.[35,36] It can also be used for dynamic monitoring of early recurrence of cancer, and can even detect signs of cancer recurrence more than a year before imaging methods.[37] The cell-free circulating tumor DNA (ctDNA) in plasma from peripheral

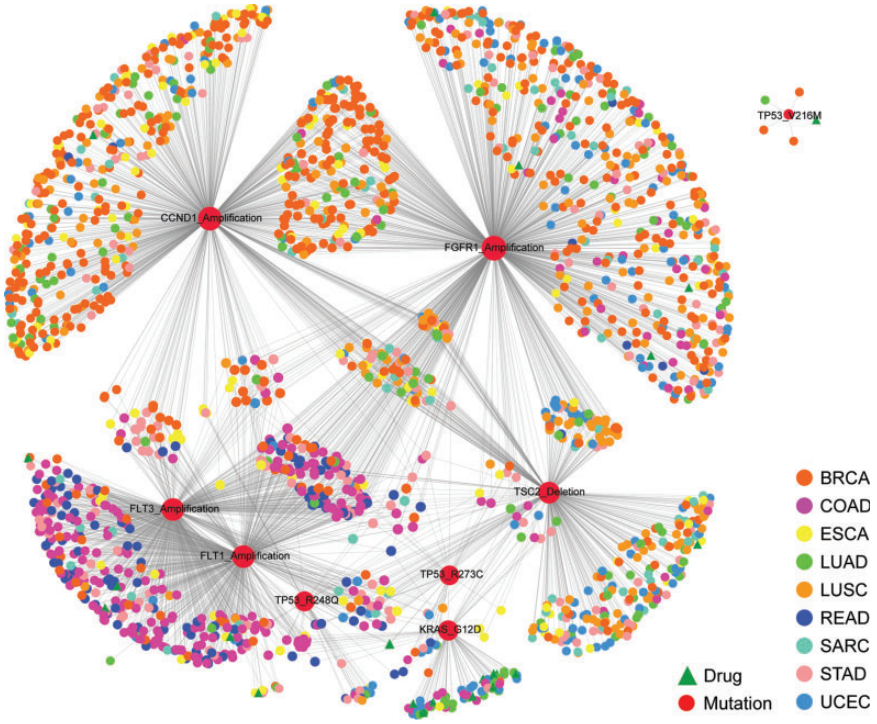

Figure 5. The drug-mutation-patients network. The sizes of the nodes are represented by the sizes of the nodes' degrees. Different colors of nodes represent patients with different cancer types and drugs as well as gene mutations. (A color version of this figure is available in the online journal.)

blood provides an opportunity for noninvasive sampling of tumor DNA. It has been an important biomarker for tumor in blood.

CtDNA retains the characteristics of primary tissue to a certain extent, in other words, it has a high consistency with the primary tissue, including DNA methylation,[38,39] mutation,[40,41] and so on. The characteristics of tumor in the primary tissue can be reflected by ctDNA.[42] Simultaneous detection of these two types of data (primary tissue and plasma ctDNA) of same patient allows us to clearly understand the degree of tissue and plasma consistency and the types of mutation in plasma that can be used instead of tissue detection.

Here we present the 416 tumor-related driver genes. And the exon sequencing of these 416 genes was performed on 20 tumor patients in 11 cancers to obtain their mutant landscape. By comparing the coincidence degree of inter-cancer mutations, it was found that the degree of overlap between the tissues and plasma of different patients was consistent, i.e. the coincidence degree of the mutant genes between the same cancers or the same tissue-derived cancers was higher, which provided the initial foundation for plasma detection instead of tissue. Furthermore, we accurately quantify the degree of substitution of plasma to tissue to prove that is feasible for plasma detection instead of tissue biopsy. Our results can provide a basis for the reliability of whether a mutation that detected by liquid biopsy can reveal a real mutation in cancer tissue. Besides, the screening of cancer prognostic biomarkers of mutation level provides important guidance for clinicians to monitor the patient's condition and to switch treatment protocols in time.

Chemical drug therapy and targeted drug therapy are the the most universal means of current cancer treatment, and the drug selection is critical to the efficacy of the patients. And when some gene mutations occur in the tumors, these mutations will have an impact on the efficacy to the drugs. Therefore, it is necessary to know the relation between gene mutations and the efficacy to drugs. In this study, the networks were used to analyze the current susceptibility of drugs that have been reported to be affected by gene mutations in cancers. Effective drugs related to gene mutations in the cancers can be obtained, providing a reference for clinicians in the selection of drug therapy for specific cancer type. The consistent mutations in tissue and plasma of our study mapping to the network were nine gene mutations (B-Mut-9) which was proved as the drug biomarkers in blood. From the network, we also obtained the same treatment regimens for different diseases and understand the mechanisms of them.

The total number of patients in this study as well as the number of patients with each type of cancer is not very adequate to study which is one of the limitations of this study. But most of our research results are based on the networks, rather than using mathematical statistics to identify the differences, which greatly reduces the impact of the sample number. As the sample size increases, our collection of blood markers will also increase, which can provide more liquid biopsy markers that can replace tissue biopsies for clinicians. In addition, due to the monotony of our sampling, there is a lack of research on cancer metastasis, so further analysis of larger sample sizes and time series sampling of multiple processes is necessary.

Authors' contributions: YZ and HBL designed the study and collected the data; SMZ analyzed, wrote and reviewed the manuscript; MS and ZYS participated in the discussion of the algorithm. All authors have read and approved the final manuscript.

ACKNOWLEDGEMENTS

We thank Yang Shao in Geneseeq Technology Inc. for providing the experimental process of data sequencing.

DECLARATION OF CONFLICTING INTERESTS

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS APPROVAL

Written informed consent was obtained from the patients for publication, and study was approved by the Harbin Medical University Cancer Hospital Ethics Committee.

FUNDING

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: We thank the support of National Natural Science Foundation of China (grant number 61972116) and the Study Abroad Returned Science Foundation of Heilongjiang Province (grand number LC2013C27).

ORCID iD

Yan Zhang https://orcid.org/0000-0002-5307-2484

SUPPLEMENTAL MATERIAL

Supplemental material for this article is available online.

REFERENCES

1. Testa JR, Cheung M, Pei J, Below JE, Tan Y, Sementino E, Cox NJ, Dogan AU, Pass HI, Trusa S, Hesdorffer M, Nasu M, Powers A, Rivera Z, Comertpay S, Tanji M, Gaudino G, Yang H, Carbone M. Germline BAP1 mutations predispose to malignant mesothelioma. *Nat Genet* 2011;**43**:1022–5
2. Harbour JW, Onken MD, Roberson ED, Duan S, Cao L, Worley LA, Council ML, Matatall KA, Helms C, Bowcock AM. Frequent mutation of BAP1 in metastasizing uveal melanomas. *Science* 2010;**330**:1410–3
3. Wiesner T, Obenauf AC, Murali R, Fried I, Griewank KG, Ulz P, Windpassinger C, Wackernagel W, Loy S, Wolf I, Viale A, Lash AE, Pirun M, Socci ND, Rutten A, Palmedo G, Abramson D, Offit K, Ott A, Becker JC, Cerroni L, Kutzner H, Bastian BC, Speicher MR. Germline mutations in BAP1 predispose to melanocytic tumors. *Nat Genet* 2011;**43**:1018–21
4. Robinson DR, Wu YM, Vats P, Su F, Lonigro RJ, Cao X, Kalyana-Sundaram S, Wang R, Ning Y, Hodges L, Gursky A, Siddiqui J, Tomlins SA, Roychowdhury S, Pienta KJ, Kim SY, Roberts JS, Rae JM, Van Poznak CH, Hayes DF, Chugh R, Kunju LP, Talpaz M, Schott AF, Chinnaiyan AM. Activating ESR1 mutations in hormone-resistant metastatic breast cancer. *Nat Genet* 2013;**45**:1446–51
5. Phipps AI, Buchanan DD, Makar KW, Win AK, Baron JA, Lindor NM, Potter JD, Newcomb PA. KRAS-mutation status in relation to colorectal cancer survival: the joint impact of correlated tumour markers. *Br J Cancer* 2013;**108**:1757–64
6. Lochhead P, Kuchiba A, Imamura Y, Liao X, Yamauchi M, Nishihara R, Qian ZR, Morikawa T, Shen J, Meyerhardt JA, Fuchs CS, Ogino S. Microsatellite instability and BRAF mutation testing in colorectal cancer prognostication. *J Natl Cancer Inst* 2013;**105**:1151–6
7. Ma F, Zhu W, Guan Y, Yang L, Xia X, Chen S, Li Q, Guan X, Yi Z, Qian H, Yi X, Xu B. ctDNA dynamics: a novel indicator to track resistance in metastatic breast cancer treated with anti-HER2 therapy. *Oncotarget* 2016;**7**:66020–31
8. Takahashi H, Kagara N, Tanei T, Naoi Y, Shimoda M, Shimomura A, Shimazu K, Kim SJ, Noguchi S. Correlation of methylated circulating tumor DNA with response to neoadjuvant chemotherapy in breast cancer patients. *Clinical Breast Cancer* 2017;**17**:61–9.e3
9. Zheng D, Ye X, Zhang MZ, Sun Y, Wang JY, Ni J, Zhang HP, Zhang L, Luo J, Zhang J, Tang L, Su B, Chen G, Zhu G, Gu Y, Xu JF. Plasma EGFR T790M ctDNA status is associated with clinical outcome in advanced NSCLC patients with acquired EGFR-TKI resistance. *Sci Rep* 2016;**6**:20913
10. Cancer Genome Atlas Research N, Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, Ellrott K, Shmulevich I, Sander C, Stuart JM. The cancer genome atlas Pan-Cancer analysis project. *Nat Genet* 2013;**45**:1113–20
11. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics* 2014;**30**:2114–20
12. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;**25**:1754–60
13. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010;**20**:1297–303
14. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, Banks E, Garimella KV, Altshuler D, Gabriel S, DePristo MA. From FastQ data to high confidence variant calls: the genome analysis toolkit best practices pipeline. *Curr Protoc Bioinform* 2013;**43**:11.10.1–33
15. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S The sequence alignment/map format and SAMtools. *Bioinformatics* 2009;**25**:2078–9
16. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res* 2012;**22**:568–76
17. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucl Acids Res* 2010;**38**:e164
18. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. *Nat Methods* 2010;**7**:248–9
19. Olshen AB, Venkatraman ES, Lucito R, Wigler M. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* 2004;**5**:557–72
20. Jiang Y, Zhang M, Guo X, Zhang R. Enrichment disequilibrium: a novel approach for measuring the degree of enrichment after gene enrichment test. *Biochem Biophys Res Commun* 2012;**424**:563–7
21. Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* 2011;**27**:431–2
22. Reimand J, Arak T, Adler P, Kolberg L, Reisberg S, Peterson H, Vilo J. g: Profiler-a web server for functional interpretation of gene lists (2016 update). *Nucl Acids Res* 2016;**44**:W83–9
23. Wang X, Fang H, Cheng Y, Li L, Sun X, Fu T, Huang P, Zhang A, Feng Z, Li C, Huang X, Li G, Du P, Yang H, Fang X, Li F, Gao Q, Liu B. The molecular landscape of synchronous colorectal cancer reveals genetic heterogeneity. *Carcinogenesis* 2018;**39**:708–18
24. Lesko AC, Goss KH, Prosperi JR. Exploiting APC function as a novel cancer therapy. *Curr Drug Targets* 2014;**15**:90–102

25. Kimura M, Iwai M, Usami E, Teramachi H, Yoshimura T. Prognostic factors in patients with advanced and recurrent colorectal cancer receiving last-line chemotherapy. *Die Pharmazie* 2018;**73**:115–9
26. Vaseghi Maghvan P, Rezaei-Tavirani M, Zali H, Nikzamir A, Abdi S, Khodadoostan M, Asadzadeh-Aghdaei H. Network analysis of common genes related to esophageal, gastric, and Colon cancers. *Gastroenterol Hepatol Bed Bench* 2017;**10**:295–302
27. Cavalcante GC, Amador MA, Ribeiro Dos Santos AM, Carvalho DC, Andrade RB, Pereira EE, Fernandes MR, Costa DF, Santos NP, Assumpcao PP, Ribeiro Dos Santos A, Santos S. Analysis of 12 variants in the development of gastric and colorectal cancers. *World J Gastroenterol* 2017;**23**:8533–43
28. Bansal VK, Rajan K, Sharma A, Paliwal P, Chaubal G, Jindal V, Misra MC, Kucheria K. Prospective Case-Control study to evaluate the role of glutathione S transferases (GSTT1 and GSTM1) gene deletion in breast carcinoma and its prognostic significance. *Ind J Surg* 2015;**77**:1067–72
29. Zhang J, Wu Y, Hu X, Wang B, Wang L, Zhang S, Cao J, Wang Z. GSTT1, GSTP1, and GSTM1 genetic variants are associated with survival in previously untreated metastatic breast cancer. *Oncotarget* 2017;**8**:105905–14
30. Coric VM, Simic TP, Pekmezovic TD, Basta-Jovanovic GM, Savic-Radojevic AR, Radojevic-Skodric SM, Matic MG, Suvakov SR, Dragicevic DP, Radic TM, Dzamic ZM, Pljesa-Ercegovac MS. GSTM1 genotype is an independent prognostic factor in clear cell renal cell carcinoma. *Urol Oncol* 2017;**35**:409–17
31. Nasr AS, Sami RM, Ibrahim NY, Darwish DO. Glutathione S transferase (GSTP 1, GSTM 1, and GSTT 1) gene polymorphisms in egyptian patients with acute myeloid leukemia. *Ind J Cancer* 2015;**52**:490–5
32. Guo F, Gong H, Zhao H, Chen J, Zhang Y, Zhang L, Shi X, Zhang A, Jin H, Zhang J, He Y. Mutation status and prognostic values of KRAS, NRAS, BRAF and PIK3CA in 353 Chinese colorectal cancer patients. *Sci Rep* 2018;**8**:6076
33. Klempner SJ, Gershenhorn B, Tran P, Lee TK, Erlander MG, Gowen K, Schrock AB, Morosini D, Ross JS, Miller VA, Stephens PJ, Ou SH, Ali SM. BRAFV600E mutations in high-grade colorectal neuroendocrine tumors may predict responsiveness to BRAF-MEK combination therapy. *Cancer Discov* 2016;**6**:594–600
34. Correale P, Marra M, Remondo C, Migali C, Misso G, Arcuri FP, Del Vecchio MT, Carducci A, Loiacono L, Tassone P, Abbruzzese A, Tagliaferri P, Caraglia M. Cytotoxic drugs up-regulate epidermal growth factor receptor (EGFR) expression in colon cancer cells and enhance their susceptibility to EGFR-targeted antibody-dependent cell-mediated-cytotoxicity (ADCC). *Eur J Cancer* 2010;**46**:1703–11
35. Ma M, Zhu H, Zhang C, Sun X, Gao X, Chen G. Liquid biopsy"-ctDNA detection with great potential and challenges. *Ann Transl Med* 2015;**3**:235
36. Tie J, Kinde I, Wang Y, Wong HL, Roebert J, Christie M, Tacey M, Wong R, Singh M, Karapetis CS, Desai J, Tran B, Strausberg RL, Diaz LA, Jr., Papadopoulos N, Kinzler KW, Vogelstein B, Gibbs P. Circulating tumor DNA as an early marker of therapeutic response in patients with metastatic colorectal cancer. *Ann Oncol* 2015;**26**:1715–22
37. Abbosh C, Birkbak NJ, Wilson GA, Jamal-Hanjani M, Constantin T, Salari R, Le Quesne J, Moore DA, Veeriah S, Rosenthal R, Marafioti T, Kirkizlar E, Watkins TBK, McGranahan N, Ward S, Martinson L, Riley J, Fraioli F, Al Bakir M, Gronroos E, Zambrana F, Endozo R, Bi WL, Fennessy FM, Sponer N, Johnson D, Laycock J, Shafi S, Czyzewska-Khan J, Rowan A, Chambers T, Matthews N, Turajlic S, Hiley C, Lee SM, Forster MD, Ahmad T, Falzon M, Borg E, Lawrence D, Hayward M, Kolvekar S, Panagiotopoulos N, Janes SM, Thakrar R, Ahmed A, Blackhall F, Summers Y, Hafez D, Naik A, Ganguly A, Kareht S, Shah R, Joseph L, Marie Quinn A, Crosbie PA, Naidu B, Middleton G, Langman G, Trotter S, Nicolson M, Remmen H, Kerr K, Chetty M, Gomersall L, Fennell DA, Nakas A, Rathinam S, Anand G, Khan S, Russell P, Ezhil V, Ismail B, Irvin-Sellers M, Prakash V, Lester JF, Kornaszewska M, Attanoos R, Adams H, Davies H, Oukrif D, Akarca AU, Hartley JA, Lowe HL, Lock S, Iles N, Bell H, Ngai Y, Elgar G, Szallasi Z, Schwarz RF, Herrero J, Stewart A, Quezada SA, Peggs KS, Van Loo P, Dive C, Lin CJ, Rabinowitz M, Aerts H, Hackshaw A, Shaw JA, Zimmermann BG, Consortium TR, Consortium P, Swanton C. Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. *Nature* 2017;**545**:446–51
38. Warton K, Mahon KL, Samimi G. Methylated circulating tumor DNA in blood: power in cancer prognosis and response. *Endocr Relat Cancer* 2016;**23**:R157–71
39. Balgkouranidou I, Chimonidou M, Milaki G, Tsaroucha E, Kakolyris S, Georgoulias V, Lianidou E. SOX17 promoter methylation in plasma circulating tumor DNA of patients with non-small cell lung cancer. *Clin Chem Lab Med* 2016;**54**:1385–93
40. Sundaresan TK, Sequist LV, Heymach JV, Riely GJ, Janne PA, Koch WH, Sullivan JP, Fox DB, Maher R, Muzikansky A, Webb A, Tran HT, Giri U, Fleisher M, Yu HA, Wei W, Johnson BE, Barber TA, Walsh JR, Engelman JA, Stott SL, Kapur R, Maheswaran S, Toner M, Haber DA. Detection of T790M, the acquired resistance EGFR mutation, by tumor biopsy versus noninvasive Blood-Based analyses. *Clin Cancer Res* 2016;**22**:1103–10
41. Kidess E, Heirich K, Wiggin M, Vysotskaia V, Visser BC, Marziali A, Wiedenmann B, Norton JA, Lee M, Jeffrey SS, Poultsides GA. Mutation profiling of tumor DNA from plasma and tumor tissue of colorectal cancer patients with a novel, high-sensitivity multiplexed mutation detection platform. *Oncotarget* 2015;**6**:2549–61
42. Lebofsky R, Decraene C, Bernard V, Kamal M, Blin A, Leroy Q, Rio Frio T, Pierron G, Callens C, Bieche I, Saliou A, Madic J, Rouleau E, Bidard FC, Lantz O, Stern MH, Le Tourneau C, Pierga JY. Circulating tumor DNA as a non-invasive substitute to metastasis biopsy for tumor genotyping and personalized medicine in a prospective trial across all tumor types. *Mol Oncol* 2015;**9**:783–90

(Received December 10, 2019, Accepted January 24, 2020)