Detection of somatic mutations in circulating cell-free DNA from patients with esophageal squamous cell carcinoma

Zuyang Yuan

Chinese Academy of Medical Sciences Cancer Institute and Hospital

Xinfeng Wang

Chinese Academy of Medical Sciences Cancer Institute and Hospital

Xiao Geng

Chinese Academy of Medical Sciences Cancer Institute and Hospital

Yin Li

Chinese Academy of Medical Sciences Cancer Institute and Hospital

Juwei Mu

Chinese Academy of Medical Sciences Cancer Institute and Hospital

Fengwei Tan

Chinese Academy of Medical Sciences Cancer Institute and Hospital

Qi Xue

Chinese Academy of Medical Sciences Cancer Institute and Hospital

Shugeng Gao

Chinese Academy of Medical Sciences Cancer Institute and Hospital

Jie He (✉ prof.jiehe@gmail.com)

Chinese Academy of Medical Sciences Cancer Institute and Hospital

Research

Keywords: esophageal squamous cell carcinoma; circulating cell-free DNA; somatic mutations; next generation sequencing

Posted Date: August 3rd, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-51232/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: The aim of this study was to assess whether both ubiquitous and heterogeneous somatic mutations could be detected in circulating cell-free DNA (cfDNA) from patients with esophageal squamous cell carcinoma (ESCC).

Methods: Paired multi-regional tumor tissues, cfDNA and white blood cells (WBCs) collected from five ESCC patients before treatment from a prospective study (NCT02395705). Of them, samples from Cohort 1 (E102 and E110) were sequenced by whole-exome sequencing (WES) and those from Cohort 2 (E104, E111 and E121) were sequenced by targeted captured sequencing with a panel of 560 cancer-related genes respectively. To call somatic single nucleotide variations (SNVs) by comparing the solid tumor or cfDNA with matched WBCs, the minimal variant allele frequency (VAF_{min}) as 0.1% and P value <0.05 were allowed.

Results: Genomic DNA (gDNA) and plasma-derived cfDNA from 26 samples were successfully sequenced. In Cohort 1, 596 (596/712, 83%) and 562 (562/796, 71%) were heterogeneous SNVs in E102 and E110 respectively. There was a statistically significant linear relationship between the VAFs for tumor and cfDNA ($R^2 = 0.78$, $P < 0.0001$). In Cohort 2, 296 (296/323, 92%), 384 (384/423, 91%) and 331 (331/357, 93%) were heterogeneous SNVs in E104, E111 and E121 respectively. cfDNA could recover an average of 60.7% (31/51; range, 35.7%-76.2%) of somatic mutations present in matched solid tumors. The correlation of VAFs between cfDNA and matched solid tumor was significantly positive ($r^2 = 0.92$, $P < 0.0001$).

Conclusions: Both sequencing approaches revealed the highly intratumoral heterogeneity in ESCC and enabled the detection of both ubiquitous and heterogeneous mutations in cfDNA. Further validation in cfDNA is required to define its potential utility for ESCC in clinical practice.

Trial registration

All patients selected in this study were from the registered clinical trial from *ClinicalTrials.gov* (NCT02395705). Date of registration: March 24, 2015.

Background

Esophageal cancer is a highly aggressive, lethal malignancy with over 400,000 deaths annually, representing a major public health concern worldwide. It is classified into two main histopathological subtypes: esophageal squamous cell carcinoma (ESCC) and esophageal adenocarcinoma (EAC). Although both subtypes share poor outcomes with a 5-year overall survival rate of approximately 15%, they are distinct diseases in terms of the cell of origin, incidence, epidemiology, and molecular signatures[1]. ESCC is the predominant subtype that usually originates from squamous epithelial cells of the esophagus [2], which is prevalent in the developing countries, especially China which contributes to nearly 50% of the global ESCC cases[3, 4]. The precise mechanisms underlying the pathogenesis of ESCC

are still unclear, although both environmental and genetic factors are suspected to play roles[5-7]. The management of ESCC is dependent on the characteristics of the patient and those of the tumor, mainly the TNM stage. Very early stage tumors may be suitable for endoscopic resection, whereas locally advanced cancers are treated with chemotherapy, chemoradiotherapy, surgical resection or combinations of these. Patients with ESCC that are not suitable for surgical management are treated with systemic chemotherapy[8].

In 1948, Mandel and Metais discovered the presence of circulating cell-free DNA (cfDNA) in human blood samples[9]. Thirty years later, it was demonstrated that serum and plasma from cancer patients contain higher concentrations of cfDNA than those from healthy individuals[10]. Later, it was found that the former harbor tumor-specific molecular alterations, suggesting that tumor-derived cfDNA, that is, circulating tumor DNA (ctDNA), can appear in the circulation[11, 12]. Despite these findings and the increased attention during the last decade, the exact origin and molecular release mechanism of cfDNA is still not fully understood. It is assumed that release of cfDNA occurs through apoptosis and necrosis of normal as well as malignant cells[13]. Moreover, some studies provided evidence for an active release via secretion of extracellular vesicles such as exosomes[14-16]. It is likely that mechanisms of compaction and release of cfDNA into circulation, which may differ depending on its origin, will be reflected by different fragment sizes[17]. The cell of origin and the mechanism of cfDNA release into blood can mark cfDNA with specific fragmentation signatures, potentially providing precise information about cell type, gene expression, cell physiology or pathology, or action of treatment[18-20]. cfDNA is typically found as double-stranded fragments of approximately 150 to 200 base pairs (bp) in length, corresponding to nucleosome-associated DNA[21]. The fraction of cfDNA that is released from primary tumors or metastases (i.e. ctDNA) represents genetic aberrations in cancer cells, which are a potential source for diagnostic, prognostic, and predictive biomarkers[11, 12].

Recent studies have demonstrated technical feasibility and clinical applications of cfDNA including detection of drug targets and resistance mutations as well as longitudinal monitoring of tumors under therapy. The potential to assess the genetic profile of a patient's tumor from a simple blood draw, without the need for an invasive biopsy, makes cfDNA analysis an attractive tool[22]. However, a key initial question is whether the mutational profile established through cfDNA testing reliably reproduces the mutational profile derived from a direct tumor biopsy, which remains the standard of care[23]. To date, the related studies on cfDNA for ESCC are highly limited. To determine the potential to be surrogate of tissue biopsy, we did this study to assess whether both ubiquitous and heterogeneous somatic mutations could be detected in cfDNA from patients with ESCC.

Methods

Patients and sample collection

All five ESCC patients (E102, E104, E110, E111, E121) from the ongoing clinical trial (*ClinicalTrials.gov* NCT02395705) in Cancer Hospital, Chinese Academy of Medical Sciences. All patients selected received

no treatments before surgery and gave informed consent. Blood samples were collected in K₂EDTA tubes (BD, USA) before surgery. Matched fresh tumor tissues were obtained during operation at the same day and stored at -80°C within 30 minutes after collection. Three sub-regional tumor tissues were analyzed in each patient, except patient E102 in whom four regions were analyzed. The study was conducted in accordance with the ethical guidelines of the Declaration of Helsinki and approved by the Independent Ethics Committee of the Cancer Hospital Chinese Academy of Medical Sciences (Approval No. NCC2020C-207).

DNA Extraction and quantification.

Peripheral blood samples were processed within 2 hours after collection. Plasma was separated from blood by centrifugation at 1,600 g for 10 min at 4°C followed by the second centrifugation at 16,000 g for 10 min at 4°C to remove debris. After the first centrifugation, the white blood cells (WBCs) were collected. Afterwards, Plasma and WBCs were stored at -80°C until DNA extraction. Germline DNA (gDNA) of tumor tissues and WBCs was extracted using the QIAamp DNA Mini Kit (QIAGEN, Germany). cfDNA was extracted from 4 mL plasma using the QIAamp Circulating Nucleic Acid Kit (QIAGEN, Germany) and eluted in 50 µL of AVE buffer. The quality of gDNA about degradation and contamination was monitored on 1% agarose gel, while the concentration was measured by Qubit DNA Assay Kit in Qubit 2.0 Fluorometer (Invitrogen, USA). The quality and quantity of cfDNA were assessed with Agilent Bioanalyzer 2100 (Agilent Technologies, USA) and Qubit dsDNA HS Assay Kit (Invitrogen, USA).

Library construction and sequencing of DNA.

Indexed Illumina NGS libraries were prepared from cfDNA and gDNA. Briefly, fragmentation was carried out on gDNA by hydrodynamic shearing system (Covaris, Massachusetts, USA) to generate 180-280 bp fragments. cfDNA was used for library construction without additional fragmentation. The NGS libraries were constructed using the Agilent SureSelect XT Custom Kit (Agilent Technologies, USA) following the manufacturer's instruments. Every sequencing library was prepared using a combination of the KAPA Hyper Prep Kit (Kapa Biosystems, USA) and the SureSelect Target Enrichment System (Agilent Technologies, USA). End repair and A-tailing reactions were performed in reaction of 60 µL volumes. The mixtures were incubated at 20°C for 30 minutes and 65°C for 30 minutes respectively. Adapter ligation was performed using 110 µL and samples were incubated at 16°C overnight using Agilent SureSelect Adapter. After post-ligation clean-up, the ligated fragments were amplified in 50 µL reaction containing KAPA HiFi HotStart ReadyMix and KAPA library Amplification Primer Mix. The following PCR amplification protocol was used: initial denaturation 98°C for 45 s; 14-16 cycles (minimum number required for optimal amplification depending on the input DNA amount) of denaturation 98°C for 15s, annealing 60°C for 30s, and extension 72°C for 30s; final extension 72°C for 1 min; and hold in 4°C. The quality and quantity of library were determined by Qubit 2.0 DNA Assays (Invitrogen, USA) and Agilent Bioanalyzer 2100 (Agilent, USA).

At first, the whole-exome sequencing (WES) was carried out on both cfDNA and gDNA in Cohort 1. The captured DNA was sequenced on the NovaSeq 6000 platform (Illumina, USA) with a mean sequencing depth of 300×. Afterwards, the targeted deep sequencing was performed on captured DNA of three patients in Cohort 2 by an established panel of 560 tumor-related genes (All-in-one Cancer Panel, Novogene, China) (a list of 560 genes was in Additional files 1) on the same sequencing platform with 150bp reads at a mean sequencing depth of 1000×.

Data processing and somatic mutation detection

The original fluorescence image files obtained from NovaSeq platform were transformed to short reads (raw data) by base calling and recorded in FASTQ format, which contained sequence information and corresponding sequencing quality information. After excluding reads containing adapter contamination and low-quality or unrecognizable nucleotides, clean data were applied for downstream bioinformatical analyses. At the same time, the total reads number, sequencing error rate, percentage of reads with average quality >20 and with average quality >30. Valid sequencing data was mapped the 150 bp paired-end reads to the reference hg38 by the BWA-MEM in default mode to get the original mapping results stored in BAM format [24], with the mapping efficiency achieved > 99%. The GATK was used to remove the duplicates for solid samples. We used SAMtools [25] to Picard MarkDuplicates [26] to sort BAM files and do duplicate marking, local realignment, and base quality recalibration to generate final files of solid tumor or cfDNA and matched WBCs for computing the sequence coverage and depth. The MuTect and Strelka [27, 28] were respectively applied to call somatic single nucleotide variations (SNVs) and small insertions and deletions (InDels) by comparing the solid tumor or cfDNA with matched WBCs sample, allowing the minimal variant allele frequency (VAF_{min}) as 0.1% and the P value <0.05. Then we filtered out those variants: (1) the number of supporting tumor reads in WBCs sample >2 or the VAF in WBCs > 1%; or (2) the read depth of solid tumor or cfDNA < 100; or (3) the supporting tumor reads in solid tumor or cfDNA appeared in only one strand; or (4) the number of supporting tumor reads in solid tumor or cfDNA $\leq 10 \times$ the number of supporting tumor reads in WBCs samples. Subsequently, the filtered variants were annotated by ANNOVAR[29]. All qualified mutations including silent and non-silent SNVs were used to construct the phylogenetic trees. At first, we used the appearance of all the SNVs in each sample to generate a binary table. Then Camin-Sokal parsimony method using PHYLIP [30] based on the binary table was used to construct the phylogenetic tree. The relationship of VAFs between cfDNA and matched solid tumor was analyzed by the Pearson's correlation analysis and the P-value < 0.05 was considered significant.

Results

Overall, all five patients (E102, E104, E110, E111, E121) with ESCC were analyzed in this study. The baseline characteristics of five patients in our study were showed in Table 1. The mean concentration of cfDNA from 4mL plasma was 0.74ng/uL (range, 0.35-1.28ng/uL). The predominant type of cfDNA extracted had a fragment size centered around 160 bp exhibited by Agilent Bioanalyzer 2100 (supplementary Figure S1A and B) (Additional files 3). The gDNA extracted was not degraded and not

contaminated with RNA or protein exhibited by agarose gel electrophoresis (supplementary Figure S1C) (Additional files 3).

Table 1. The baseline characteristics of five patients.

Patient ID	Sex	Age (years)	Differentiation	Location	T stage	N stage	M stage	Stage
E102	male	60	high	mid-thoracic	3	2	0	IIIB
E104	male	58	low	mid-thoracic	1	2	0	IIIA
E110	male	64	moderate	mid-thoracic	3	2	0	IIIB
E111	male	69	moderate	mid-thoracic	3	0	0	IIB
E121	male	70	moderate	mid-thoracic	2	0	0	IIA

Cohort 1: Whole-exome sequencing of multi-regions of primary tumor and matched cfDNA identified genomic heterogeneity

The WES was successfully performed in those two patients of 9 samples. The quality of WES data with mapping ratio of >99% and exome enrichment of >99% was acceptable (supplementary Table S1) (Additional files 2). Overall, the most common type of mutation was C>T transition and the major type of mutation was missense mutation (supplementary Figure S2) (Additional files 3), which had been observed in previous study of ESCC [31]. The total mutations (silent and non-silent) for solid tumor ranged from 290 to 483, with a median value 342 (221 for silent ones and 123 for non-silent ones). We found 915 and 863 mutations (376 silent and 340 non-silent ones) for two cfDNA samples respectively. Tumor mutation burden (TMB) for solid tumor samples ranged from 4.92 to 8.19 mutations /M, whereas cfDNA samples had higher TMB from 14.63 to 15.51 mutations /M. A total of 694 non-silent mutations were discovered in 9 samples. In E102, a total of 712 mutations (silent and non-silent mutations) were identified in tumor tissues, 119 of which were ubiquitous in T1, T2, T3 and T4 tumor regions. Meanwhile, 796 mutations (silent and non-silent mutations) were identified in E110, 234 of which were ubiquitous in T1, T2 and T4 tumor regions. To explore intratumoral heterogeneity and the genomic evolution of ESCC, phylogenetic trees were constructed on the basis of somatic mutations (both silent and non-silent mutations) identified in each region. The phylogenetic trees varied extensively in two cases, which showed evidence of spatial intratumoral heterogeneity, with an average of 76.6% (E102 of 83.2% and E110 of 70.6%) of somatic variants having spatial heterogeneity (Figure 1A).

13 genes previously reported in ESCC were confirmed in this cohort, including *TP53*, *MUC16*, *CTNND2*, *DNAH9*, *EP300*, *NOTCH3*, *SYNE1*, *RP1*, *TTN*, *ATM*, *KMT2D*, *NOTCH1* and *USH2A* [31, 32]. However, only *MUC16*, *CTNND2* and *NOTCH3* were detected in matched cfDNA (Figure 1B). In addition, 12 of 20 (60.0%) tumor-associated mutations identified were found in cfDNA, which were the important driver genes in four oncogenic pathways, including NOTCH (*NOTCH3*, *NCOR2*, *SPEN*), RTK-RAS (*ARHGAP35*, *KSR1*, *IRS1*), WNT (*DVL3*, *FRAT1*) and Hippo (*TAOK2*, *LLGL2*, *TEAD2*) signaling pathways (Figure 2). In addition, we found two novel cancer-related genes (*MUC4* and *MUC17*) in both of the two patients which were not previously reported in ESCC. The details of somatic mutations and prevalence in two ESCC patients were

showed in supplementary Table S2 and S3 (Additional files 2). The Pearson's correlation analysis of VAFs between cfDNA and matched solid tumor showed the positive linear relationship ($R^2 = 0.78$, $P < 0.0001$) (Figure 3). The VAFs of ubiquitous and heterogeneous somatic mutations in cfDNA and solid tumor were listed in supplementary Table S4 (Additional files 2).

Cohort 2: Targeted captured sequencing revealed correlation of genomic alterations between cfDNA and matched solid tumor

Since ESCC has low tumor fraction in cfDNA, in order to better identify mutations with low VAFs, we performed targeted captured sequencing by a panel of 560 cancer-associated genes with higher sequencing coverage ($\sim 1000\times$) for another cohort (E104, E111 and E121) (supplementary Table S5) (Additional files 2). The most common type of mutation was C>T transition and the majority of variant type was missense mutation, which were the same as Cohort 1 via WES (supplementary Figure S3) (Additional files 3). The total mutations (silent and non-silent) for solid tumor ranged from 107 to 194, with a median value of 142. We found 494, 769 and 262 mutations (245, 407 and 121 non-silent ones respectively) for three cfDNA samples respectively. A total of 333 genes were mutated (non-silent mutations) in 12 samples (supplementary Table S6 and S7) (Additional files 2). The phylogenetic trees were constructed on the basis of somatic mutations (both silent and non-silent mutations) identified in each region of those three cases. The phylogenetic trees varied extensively as well, which showed evidence of spatial intratumoral heterogeneity, with an average of 91.7% (1011/1103; range, 90.5%-92.7%) of heterogeneous somatic mutations (Figure 4A).

32 mutated genes previously reported in ESCC were confirmed in this cohort [31-33]. In patient E104, 323 variants were detected in three sub-regions, 27 out of which were ubiquitous variants. *KMT2D*, *TP53*, *UBR5* from the identified variants were previously confirmed as frequent mutations in ESCC [31, 32]. In patient E111, 423 mutations were detected totally, 39 out of which were ubiquitous variants in three regions. Among them, *CSMD3*, *PTCH1*, *TP53*, *PIK3CA* were important and associated with ESCC. In Patient E121, 357 mutations were detected totally, 26 of which were ubiquitous variants. Among them, *FBXW7*, *LRP1B*, *TP53*, *MTOR* and *EP300* were confirmed as frequent mutations in ESCC (Figure 4B). To determine the VAFs of tumor mutations in cfDNA. The VAF_{min} we setup in algorithm for detection of mutations was 0.1%. The first quantile and third quantile of VAFs in cfDNA were 0.49-0.88%, 0.44-0.77% and 0.41-0.875% for patient E104, E111 and E121 respectively (supplementary Figure S4) (Additional files 3). In total, there were 14 mutations with VAFs of >5% detected in cfDNA. The results showed ESCC had low ctDNA in the blood, which was similar to one previous study on EAC [34]. Moreover, the variants called from cfDNA were much more than the ones called from solid tumor tissues. Some true positive calls were missed in our solid tumor samples due to the intratumor heterogeneity and sampling biases. Some were potential false positive calls possibly due to clonal hematopoiesis (CH). Therefore, we focused on the ESCC recurrent genes to answer the following questions: how many mutated genes in solid tumor were identified in cfDNA? How many important mutations were found in cfDNA only? For patient E104, 10 out of 16 (62.5%) ESCC-associated mutations detected in solid tumor were recovered in paired cfDNA, including *CSMD3*, *KMT2D*, *LRP1B*, *SYNE1*, *ATR*, *EP300*, *PRDM1*, *UBR5*, *FSTL5*, *LIFR*. In E111,

16 out of 21 (76.2%) ESCC-associated mutations were recovered in cfDNA, including *KMT2D*, *LRP1B*, *SYNE1*, *BRCA2*, *ATR*, *EP300*, *ASXL1*, *MTOR*, *PKHD1*, *PTCH1*, *UBR5*, *BRIP1*, *KMT2A*, *NUP214*, *NOTCH2*, *CREBBP*. In E121, 35.7% (5/14) ESCC-associated mutations were found in both solid tumor and cfDNA, including *KMT2D*, *LRP1B*, *EP300*, *PKHD1*, *PRDM1*. Moreover, *KMT2D*, *SYNE1* and *UBR5* were shared by cfDNA and all regions of tumor in E104. *PTCH1* was the shared one in E111 and *LRP1B* was shared by three sub-regional tumors and cfDNA in E121 (Figure 5). Mutations in *TP53* were often detected in solid tumor only, whereas mutations in *CBL*, *POLE*, *PTCH1* and *NFE2L2* were only detected in cfDNA. Furthermore, the correlation of VAFs between cfDNA and matched solid tumor was strong ($R^2 = 0.92$) with a significant P-value of <0.0001 (Figure 6). The VAFs of somatic mutations in cfDNA and solid tumors were showed in supplementary Table S8 (Additional files 2).

Discussion

Analysis of somatic alterations in tumor tissue has become routine practice in clinical oncology. Although these alterations are highly informative, sampling tumor tissue has limitations as tissue biopsies are often difficult to obtain and are subjected to sampling bias resulting from temporal and spatial tumor heterogeneity[33]. Therefore, alternative strategies, such as liquid biopsies, are currently evaluated for applicability in different clinical settings. Liquid biopsies have a few advantages compared to tissue biopsies; they are less invasive, safe and may overcome difficulties of intratumoral heterogeneity[35]. Together with the possibility of multiple assessments over time, the use of cfDNA may be able to predict response to treatment and patterns of therapeutic resistance earlier and more accurately than radiological imaging[34, 36]. The ability to analyze tumor-derived DNA from a routine blood draw without the need for an invasive tumor biopsy represents a critical advance with potentially transformative clinical applications. In particular, the minimally invasive nature of cfDNA analysis provides a means of molecular profiling for tumors that are difficult or unsafe to biopsy and allows a practical means for monitoring tumor DNA serially over time without the risk and potential complications of standard tumor biopsy. In addition, cfDNA analysis may better capture the molecular heterogeneity harbored by multiple distinct clonal populations in a patient's tumor, as compared with a needle biopsy of a single tumor lesion. Finally, cfDNA analysis offers the potential for tumor detection or monitoring in patients without clinically evident disease[35].

Recent developments in NGS technologies, enable the testing of cfDNA with high sensitivity and specificity[37]. However, the low amount of cfDNA in the blood presents challenges in constructing sequencing libraries with high quality and complexity, especially in ESCC. Notably, to minimize the likelihood of detecting mutations due to CH, some mutations involved in CH that were also observed in the matched WBCs samples should be excluded[38]. Thus, we simultaneously sequenced the leukocytes fraction of an individual to be able to properly attribute the mutations in their cfDNA to tumors[39]. In this pilot study, we respectively used the WES and targeted sequencing with a panel of 560 cancer-associated genes to detect mutations in paired multi-regional solid tumors and cfDNA. Some ESCC-associated mutations identified previously were detected in cfDNA in two cohorts, including *MUC16*, *CTNND2*,

NOTCH3, CSMD3, KMT2D, LRP1B, SYNE1, BRCA2, NFE2L2, EP300, ATR, ATM, ASXL1, MTOR, PTCH1, PKHD1, PRDM1, UBR5, KMT2A, NOTCH2, ERCC3, URIP1, CBL, FSTL5, LIFR and POLE [31-33]. All the results showed that both platforms had the ability to detect ESCC-associated mutations in cfDNA. We found the targeted sequencing in cohort 2 identified more important genes than WES in cohort 1. Since ESCC has low tumor fraction in cfDNA, at a similar price, targeted sequencing with panel of higher sequencing depth is better for detecting variants of low VAFs than WES of compromised sequencing depth, which is why most groups used NGS panels encompassing a select set of genes commonly in studies on cfDNA. From our results, multi-regions of primary tumor and matched cfDNA WES was used to identify somatic mutations present in each of tumor regions and matched cfDNA, which demonstrated highly spatial intratumoral heterogeneity in all tumors. Moreover, 12 cancer-associated mutations identified were found in cfDNA, which were the important driver genes in four oncogenic signaling pathways, including NOTCH, RTK-RAS, WNT and Hippo signaling pathways which play important roles in the occurrence and development of tumor [40]. Although a few ubiquitous mutations were found in cfDNA due to limited number of cases, the VAFs of those common alterations between paired solid tumor and cfDNA were significantly correlated. The results suggested that genetic information from limited number of tumor tissues biopsied could not represent the whole genetic profile of ESCC and cfDNA could be an adjuvant tool for revealing the genetic characteristics of ESCC.

Afterwards, we performed target captured sequencing with a panel of 560 cancer-related genes of higher depth in another cohort. For the previously confirmed variants in ESCC, the mutation density in cfDNA was significantly higher than solid tumor, which suggested that cfDNA could present more information of mutations than solid tumor. On the one hand, cfDNA could recover an average of 60.7% (31/51; range, 35.7%-76.2%) of mutations present in solid tumor samples, which was similar to the rates of previous studies on other cancers[38, 41, 42]. The Pearson's correlation of VAFs between cfDNA and matched solid tumor was significantly positive, which indicated genomic profile of cfDNA could represent the information of ubiquitous and heterogeneous mutations in solid tumors. On the other hand, many somatic mutations which were appeared in three sub-regions were not identified in their matched cfDNA. For instance, we found that some important ESCC-related mutations identified in solid tumors of some patients, including *TP53 R273H* and *TP53 R175H* mutations, were not detected in matched cfDNA, and out of expectation those mutations had high VAFs in solid tumor. The majority of these variants of high VAFs in solid tumor had important functions in the development of ESCC. In addition, *TP53 R273H* mutation was shown in three sub-regions of patient E111 with VAF of 22.98%, 42.86% and 34.79% respectively, but only 0.93% in cfDNA. This observation agreed with a previous study that one ESCC patient had a *TP53 N131I* mutation with VAF of 52.9 % in tumor and 1.3% in cfDNA from preoperative plasma [43]. One possible explanation could be that the colonies of tumor cells with driver mutations are resistant to apoptosis, and rarely release mutated DNA fragments to plasma.

Successful monitoring of response to treatment carries several important implications. First, it provides physicians with a rapid assessment of the success of the treatment choice and may allow for modification of therapeutic regimens if inadequate response is noted. Second, poor clinical response to neoadjuvant treatment may be identified early during therapy, thereby allowing changes in the treatment

plan, including surgical management[34]. Therefore, as a repeat or serial testing tool, cfDNA testing for treatment selection and tracking therapeutic response is increasingly used as an alternative to repeat invasive biopsy and may reveal actionable mutations that guide treatment decisions in these patients[22]. Recently, data from one study with longitudinal EAC samples indicated that ctDNA levels correlate with and precede evidence of response to therapy, which demonstrated that the potential of ctDNA as a dynamic biomarker to monitor treatment response in patients with EAC. Moreover, the VAFs of some mutations were lower or equaled zero in postoperative plasma from ESCC patients[34]. Similarly, somatic mutations can be detected in preoperative cfDNA from patients with ESCC at stage IIA to IIIB , and at a lower frequency in postoperative cfDNA in another study[44]. Pectasides et al. recently explored cfDNA as a tool to identify therapeutic targets not detectable from standard tissue-based testing in gastric and esophageal adenocarcinomas (GEA) with metastatic lesions. cfDNA profiling may ultimately provide a more accurate representation of disseminated disease in GEA [45]. These results demonstrated that ctDNA is a valuable biomarker for tracking tumor status and evaluating treatment effect. One such study in our medical center is ongoing (*ClinicalTrials.gov* NCT02395705), which involves the evaluation of the survival benefit of neoadjuvant chemotherapy (cisplatin plus paclitaxel) versus surgery alone in ESCC patients. This study will also determine the feasibility of using circulating biomarkers, including cfDNA, to reliably predict the sensitivity of neoadjuvant chemotherapy and early screen patients with insensitive response to chemotherapy before treatment so as to reduce unnecessary chemotherapy and change therapeutic plan for those ESCC patients. In addition, longitudinal sampling of cfDNA from diagnosis to relapse will be done to monitor minimal residual disease after surgery and detect the recurrence. The targeted sequencing approach with a panel of 560 cancer-related genes presented here will be applied to this clinical trial to analyze cfDNA. The results from our pilot study and previously related studies demonstrated that integration of real-time cfDNA analysis into clinical trials and eventually into standard clinical management has the potential to become a valuable tool for revealing tumor heterogeneity and monitoring therapeutic response.

Inevitably, there were some limitations in our study. First, our study cohort size was limited. Although the number of patients enrolled was small, we collected multi-regional samples of each tumor, which could exactly indicate the intratumoral heterogeneity in ESCC. As far as we know, compared to most studies which sampled single region of tumor, it is the first time to sequence multi-regional tumors and matched cfDNA in ESCC that could exactly reveal the high intratumoral heterogeneity in ESCC and the ability of cfDNA to detect both ubiquitous and heterogeneous mutations and represent the genetic characteristics of solid tumor. Second, owing to lack of materials we could not do an independent validation of the mutations detected in cfDNA. To overcome this shortcoming, we carefully monitored sequencing error rates for all positions for which we identified mutations in cfDNA by our predefined algorithm and criteria of data processing. In addition, we used two NGS approaches to sequence two cohorts independently and to some extent, they could validate each other mutually. Technically, we found the targeted sequencing of cancer-specific gene panel with high depth is suitable for detecting variants of low VAFs in cfDNA in some tumors, such as ESCC. In the last decade, cfDNA has become a very hot topic in oncology and it is almost impossible to keep pace with the number of new papers published every day. However,

there are quite a few fundamental problems for which we do not have an answer. First, our basic knowledge on cfDNA is far from being complete. We still do not know exactly all of the mechanisms leading to the release of cell-free DNA into circulation. Second, there are some technical and methodological issues that have to be solved. So far, there has not been consensus on a “gold standard” for the isolation of cfDNA. Finally, computational approaches should be improved to perform the genetic analysis of cfDNA sensitively deconvoluting the cancer-specific signals from the mixture of cancer and normal signals in cfDNA.

Conclusions

Both WES and targeted sequencing approaches revealed the highly spatial intratumoral heterogeneity in ESCC and enabled the detection of both ubiquitous and heterogeneous mutations in cfDNA. The VAFs of somatic mutations identified between cfDNA and matched solid tumors were highly correlated. This implies that cfDNA has the potential to be a biomarker for the presence of tumor cells in ESCC patients and aid in monitoring and accurate prediction of therapeutic response during ESCC treatments and management. Further larger studies are needed to validate the clinical applicability of cfDNA as we only had five patients in this pilot study.

List Of Abbreviations

ESCC, esophageal squamous cell carcinoma; EAC, esophageal adenocarcinoma; cfDNA, circulating cell-free DNA; ctDNA, circulating tumor DNA; WBC, white blood cell; WES, whole-exome sequencing; SNV, single nucleotide variation; VAF, variant allele frequency; VAF_{min} , minimal variant allele frequency; NGS, next generation sequencing; InDel, insertion and deletion; TMB, Tumor mutation burden; CH, clonal hematopoiesis; GEA, gastric and esophageal adenocarcinomas.

Declarations

Ethics approval and consent to participate

The study was performed in accordance with the ethical guidelines of the Declaration of Helsinki and approved by the Independent Ethics Committee of the Cancer Hospital Chinese Academy of Medical Sciences (Approval No. NCC2020C-207).

Consent for publication

Not applicable.

Availability of data and materials

All data generated or analyzed during this study are included in this published article and its supplementary information files and are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

Not applicable.

Authors' contributions

ZY, YL and JH conceived this study. ZY, XW and XG collected the samples and conducted the experiments of the study. ZY and XW analyzed and interpreted the data. ZY was a major contributor in writing the manuscript. YL, JM, FT, QX, SG, JH revised the manuscript. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

References

1. Rustgi AK, El-Serag HB. Esophageal carcinoma. *N Engl J Med*. 2014; 371:2499-2509.
2. Pennathur A, Gibson MK, Jobe BA, Luketich JD. Oesophageal carcinoma. *Lancet*. 2013; 381:400-412.
3. Arnold M, Soerjomataram I, Ferlay J, Forman D. Global incidence of oesophageal cancer by histological subtype in 2012. *Gut*. 2015; 64:381-387.
4. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, Jemal A, Yu XQ, He J. Cancer statistics in China, 2015. *CA Cancer J Clin*. 2016; 66:115-132.
5. Clocchiatti A, Cora E, Zhang Y, Dotto GP. Sexual dimorphism in cancer. *Nat Rev Cancer*. 2016; 16:330-339.
6. Dorak MT, Karpuzoglu E. Gender differences in cancer susceptibility: an inadequately addressed issue. *Front Genet*. 2012; 3:268.
7. Pohl H, Wrobel K, Bojarski C, Voderholzer W, Sonnenberg A, Rosch T, et al. Risk factors in the development of esophageal adenocarcinoma. *Am J Gastroenterol*. 2013; 108:200-207.
8. Smyth EC, Lagergren J, Fitzgerald RC, Lordick F, Shah MA, Lagergren P, et al. Oesophageal cancer. *Nat Rev Dis Primers*. 2017; 3:17048.
9. Mandel P, Metais P. [Not Available]. *C R Seances Soc Biol Fil*. 1948; 142:241-243.
10. Leon SA, Shapiro B, Sklaroff DM, Yaros MJ. Free DNA in the serum of cancer patients and the effect of therapy. *Cancer Res*. 1977; 37:646-650.
11. Diehl F, Schmidt K, Choti MA, Romans K, Goodman S, Li M, et al. Circulating mutant DNA to assess tumor dynamics. *Nat Med*. 2008; 14:985-990.

12. Bettgowda C, Sausen M, Leary RJ, Kinde I, Wang Y, Agrawal N, et al. Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci Transl Med*. 2014; 6:224ra224.
13. Jahr S, Hentze H, Englisch S, Hardt D, Fackelmayer FO, Hesch RD, et al. DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells. *Cancer Res*. 2001; 61:1659-1665.
14. Gahan PB, Anker P, Stroun M. Metabolic DNA as the origin of spontaneously released DNA? *Ann N Y Acad Sci*. 2008; 1137:7-17.
15. Kahlert C, Melo SA, Protopopov A, Tang J, Seth S, Koch M, et al. Identification of double-stranded genomic DNA spanning all chromosomes with mutated KRAS and p53 DNA in the serum exosomes of patients with pancreatic cancer. *J Biol Chem*. 2014; 289:3869-3875.
16. Thakur BK, Zhang H, Becker A, Matei I, Huang Y, Costa-Silva B, Zheng Y, Hoshino A, Brazier H, Xiang J et al. Double-stranded DNA in exosomes: a novel biomarker in cancer detection. *Cell Res*. 2014; 24:766-769.
17. Thierry AR, El Messaoudi S, Gahan PB, Anker P, Stroun M. Origins, structures, and functions of circulating DNA in oncology. *Cancer Metastasis Rev*. 2016; 35:347-376.
18. Ulz P, Thallinger GG, Auer M, Graf R, Kashofer K, Jahn SW, et al. Inferring expressed genes by whole-genome sequencing of plasma DNA. *Nat Genet*. 2016; 48:1273-1278.
19. Snyder MW, Kircher M, Hill AJ, Daza RM, Shendure J. Cell-free DNA Comprises an In Vivo Nucleosome Footprint that Informs Its Tissues-Of-Origin. *Cell*. 2016; 164:57-68.
20. Bronkhorst AJ, Wentzel JF, Aucamp J, van Dyk E, du Plessis L, Pretorius PJ. Characterization of the cell-free DNA released by cultured cancer cells. *Biochim Biophys Acta*. 2016; 1863:157-165.
21. Fan HC, Blumenfeld YJ, Chitkara U, Hudgins L, Quake SR. Analysis of the size distributions of fetal and maternal cell-free DNA by paired-end sequencing. *Clin Chem*. 2010; 56:1279-1286.
22. Volckmar AL, Sultmann H, Riediger A, Fioretos T, Schirmacher P, Endris V, et al. A field guide for cancer diagnostics using cell-free DNA. From principles to practice and clinical applications. *Genes Chromosomes Cancer*. 2018; 57:123-139.
23. Chen XX, Zhong Q, Liu Y, Yan SM, Chen ZH, Jin SZ, et al. Genomic comparison of esophageal squamous cell carcinoma and its precursor lesions by multi-region whole-exome sequencing. *Nat Commun*. 2017; 8:524.
24. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2010; 26:589-595.
25. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009; 25:2078-2079.
26. Ebbert MT, Wadsworth ME, Staley LA, Hoyt KL, Pickett B, Miller J, et al. Evaluating the necessity of PCR duplicate removal from next-generation sequencing data and a comparison of approaches. *BMC Bioinformatics*. 2016; 17:239.

27. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol.* 2013; 31:213-219.
28. Saunders CT, Wong WS, Swamy S, Becq J, Murray LJ, Cheetham RK. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics.* 2012; 28:1811-1817.
29. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 2010; 38:e164.
30. Retief JD. Phylogenetic analysis using PHYLIP. *Methods Mol Biol.* 2000; 132:243-258.
31. Gao YB, Chen ZL, Li JG, Hu XD, Shi XJ, Sun ZM, et al. Genetic landscape of esophageal squamous cell carcinoma. *Nat Genet.* 2014; 46:1097-1102.
32. Song Y, Li L, Ou Y, Gao Z, Li E, Li X, et al. Identification of genomic alterations in oesophageal squamous cell cancer. *Nature.* 2014; 509:91-95.
33. Hao JJ, Lin DC, Dinh HQ, Mayakonda A, Jiang YY, Chang C, et al. Spatial intratumoral heterogeneity and temporal clonal evolution in esophageal squamous cell carcinoma. *Nat Genet.* 2016; 48:1500-1507.
34. Egyud M, Tejani M, Pennathur A, Luketich J, Sridhar P, Yamada E, et al. Detection of Circulating Tumor DNA in Plasma: A Potential Biomarker for Esophageal Adenocarcinoma. *Ann Thorac Surg.* 2019; 108:343-349.
35. Corcoran RB, Chabner BA. Application of Cell-free DNA Analysis to Cancer Treatment. *N Engl J Med.* 2018; 379:1754-1765.
36. Azad TD, Chaudhuri AA, Fang P, Qiao Y, Esfahani MS, Chabon JJ, et al. Circulating Tumor DNA Analysis for Detection of Minimal Residual Disease After Chemoradiotherapy for Localized Esophageal Cancer. *Gastroenterology.* 2020; 158:494-505 e496.
37. Volik S, Alcaide M, Morin RD, Collins C. Cell-free DNA (cfDNA): Clinical Significance and Utility in Cancer Shaped By Emerging Technologies. *Mol Cancer Res.* 2016; 14:898-908.
38. Razavi P, Li BT, Brown DN, Jung B, Hubbell E, Shen R, et al. High-intensity sequencing reveals the sources of plasma circulating cell-free DNA variants. *Nat Med.* 2019; 25:1928-1937.
39. Bellosillo B, Montagut C. High-accuracy liquid biopsies. *Nat Med.* 2019; 25:1820-1821.
40. Sanchez-Vega F, Mina M, Armenia J, Chatila WK, Luna A, La KC, et al. Oncogenic Signaling Pathways in The Cancer Genome Atlas. *Cell.* 2018; 173:321-337 e310.
41. Strickler JH, Loree JM, Ahronian LG, Parikh AR, Niedzwiecki D, Pereira AAL, et al. Genomic Landscape of Cell-Free DNA in Patients with Colorectal Cancer. *Cancer Discov.* 2018; 8:164-173.
42. Chicard M, Colmet-Daage L, Clement N, Danzon A, Bohec M, Bernard V, et al. Whole-Exome Sequencing of Cell-Free DNA Reveals Temporo-spatial Heterogeneity and Identifies Treatment-Resistant Clones in Neuroblastoma. *Clin Cancer Res.* 2018; 24:939-949.
43. Luo H, Li H, Hu Z, Wu H, Liu C, Li Y, et al. Noninvasive diagnosis and monitoring of mutations by deep sequencing of circulating tumor DNA in esophageal squamous cell carcinoma. *Biochem Biophys Res*

Commun. 2016; 471:596-602.

44. Meng P, Wei J, Geng Y, Chen S, Terpstra MM, Huang Q, et al. Targeted sequencing of circulating cell-free DNA in stage II-III resectable oesophageal squamous cell carcinoma patients. BMC Cancer. 2019; 19:818.
45. Pectasides E, Stachler MD, Derks S, Liu Y, Maron S, Islam M, et al. Genomic Heterogeneity as a Barrier to Precision Medicine in Gastroesophageal Adenocarcinoma. Cancer Discov. 2018; 8:37-48.

Figures

Figure 1

The occurrence of known ESCC-related mutations in Cohort 1. Phylogenetic trees were constructed from all somatic mutations detected in multi-regional tumors, with ubiquitous mutations on the trunk of the tree, and heterogeneous mutations on the branches. Underneath each phylogenetic tree is the number (n) and percentage of heterogeneous mutations (A). 13 genes previously reported in ESCC were confirmed in

this cohort, including TP53, MUC16, CTNND2, DNAH9, EP300, NOTCH3, SYNE1, RP1, TTN, ATM, KMT2D, NOTCH1 and USH2A. Only MUC16, CTNND2 and NOTCH3 are detected in cfDNA (B).

Figure 2

The occurrence of oncogenic pathway genes. 12 (12/20, 60.0%) cancer-associated mutations identified were found in cfDNA, which were the important driver genes in five oncogenic pathways. Each pathway corresponds to related driver genes in the heat map.

Figure 3

The correlation of VAFs of the concordant variants between cfDNA and solid tumor by WES.

Figure 4

The occurrence of confirmed ESCC-associated mutations in Cohort 2. Phylogenetic trees were constructed from all somatic mutations detected in multi-regional tumors, with ubiquitous mutations on the trunk of the tree, and heterogeneous mutations on the branches. Underneath each phylogenetic tree is the number (n) and percentage of heterogeneous mutations (A). 32 genes previously reported in ESCC were confirmed in this cohort, 28 of which were detected in cfDNA (B).

Figure 5

The distribution of the overlapped ESCC-related genes between multi-regional solid tumors and cfDNA. In E104, 62.5% (10/16) ESCC-associated mutations detected in solid tumor were recovered in paired cfDNA (A, left); 76.2% (16/21) and 35.7% (5/14) of mutations were found in E111 (B, left) and E121 (C, left) respectively. KMT2D, SYNE1 and UBR5 were shared by cfDNA and all regions of tumor in E104 (A, right). PTCH1 was the shared one in E111 (B, right) and LRP1B was shared in E121 (C, right).

Figure 6

The correlation of VAFs of the concordant variants between cfDNA and solid tumor by targeted captured sequencing.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [genelist.txt](#)
- [TablesInSupplementary.xlsx](#)
- [FiguresInSupplementary.pdf](#)