



Enrichment of short mutant cell-free DNA fragments enhanced detection of pancreatic cancer[☆]

Xiaoyu Liu^{a,b,c,1}, Lingxiao Liu^{a,b,1}, Yuan Ji^{d,1}, Changyu Li^{a,b,1}, Tao Wei^e, Xuerong Yang^f, Yuefang Zhang^c, Xuyu Cai^g, Yangbin Gao^h, Weihong Xuⁱ, Shengxiang Rao^j, Dayong Jin^{k,l}, Wenhui Lou^{k,l,**}, Zilong Qiu^{c,3,*}, Xiaolin Wang^{a,b,2,***}

^a Department of Interventional Radiology, Zhongshan Hospital Fudan University, Shanghai, China

^b Shanghai Institute of Medical Imaging, Shanghai, China

^c Institute of Neuroscience, State Key Laboratory of Neuroscience, CAS Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, Shanghai, China

^d Department of Pathology, Zhongshan Hospital Fudan University, Shanghai, China

^e Department of Hepatobiliary and Pancreatic Surgery, School of Medicine, The Second Affiliated Hospital of Zhejiang University, Hangzhou, China

^f Department of Radiology, Shuguang Hospital Shanghai University of Traditional Chinese Medicine, Shanghai, China

^g Precision Medicine Center, West China Hospital, Sichuan University, Chengdu, Sichuan, China

^h Qihan Bio Inc., Hangzhou, Zhejiang, China

ⁱ Stanford Genome Technology Center, Palo Alto, CA, USA

^j Department of Radiology, Zhongshan Hospital Fudan University, Shanghai, China

^k Department of Pancreatic Surgery, Zhongshan Hospital Fudan University, Shanghai, China

^l Department of General Surgery, Zhongshan Hospital Fudan University, Shanghai, China

ARTICLE INFO

Article history:

Received 12 December 2018

Received in revised form 30 January 2019

Accepted 6 February 2019

Available online 9 March 2019

Keywords:

Pancreatic cancer

Cell-free DNA

Fragmentation

Tissue biopsy

Diagnosis and prognosis

ABSTRACT

Background: Analysis of cell-free DNA (cfDNA) is promising for broad applications in clinical settings, but with significant bias towards late-stage cancers. Although recent studies have discussed the diverse and degraded nature of cfDNA molecules, little is known about its impact on the practice of cfDNA analysis.

Methods: We developed single-strand library preparation and hybrid-capture-based cfDNA sequencing (SLHC-seq) to analysis degraded cfDNA fragments. Next we used SLHC-seq to perform cfDNA profiling in 112 pancreatic cancer patients, and the results were compared with 13 previous reports. Extensive analysis was performed in terms of cfDNA fragments to explore the reasons for higher detection rate of *KRAS* mutations in the circulation of pancreatic cancers.

Findings: By applying the new approach, we achieved higher efficiency in analysis of mutations than previously reported using other detection assays. 791 cancer-specific mutations were detected in plasma of 88% patients with *KRAS* hotspots detected in 70% of all patients. Only 8 mutations were detected in 28 healthy controls without any known oncogenic or truncating alleles. cfDNA profiling by SLHC-seq was largely consistent with results of tissue-based sequencing. SLHC-seq rescued short or damaged cfDNA fragments along to increase the sensitivity and accuracy of circulating-tumour DNA detection.

Interpretation: We found that the small mutant fragments are prevalent in early-stage patients, which provides strong evidence for fragment size-based detection of pancreatic cancer. The new pipeline enhanced our understanding of cfDNA biology and provide new insights for liquid biopsy.

© 2019 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

[☆] This study was mainly funded by NSFC Grants (#31625013, #91732302).

* Corresponding author.

** Correspondence to: W. Lou, Department of Pancreatic Surgery, Zhongshan Hospital Fudan University, Shanghai, China.

*** Correspondence to: X. Wang, Department of Interventional Radiology, Zhongshan Hospital Fudan University, Shanghai, China.

E-mail addresses: lou.wenhui@zs-hospital.sh.cn (W. Lou), zqiu@ion.ac.cn (Z. Qiu), wang.xiaolin@zs-hospital.sh.cn (X. Wang).

¹ These authors contributed equally to this work.

² Xiaolin Wang is the leading corresponding author.

³ Zilong Qiu and Wenhui Lou are co-corresponding authors.

1. Introduction

Pancreatic cancer is among the most lethal malignancies, with a 5-year survival rate of approximately 8% [1], and is predicted to be the second leading cause of cancer-related death by the year 2030 [2]. Early detection and personalized treatment can greatly benefit patients with this disease [3,4]. Because targeted therapies have been reported to be highly effective for a subset of cancer patients with actionable mutations, characterizing pancreatic cancer at the genetic level holds great promise for precision medicine in the near future [5–7].

Research in context

Evidence before this study

Analysis of cell-free DNA (cfDNA) is promising for broad applications in clinical settings, but with significant bias towards late-stage cancers. Recent advances have uncovered that cfDNA is a group of molecules with high diversification and degradation. Conventional next generation sequencing (NGS) or polymerase chain reaction (PCR) approaches are not sensitive enough to capture the full diversity of these molecules, particularly degraded fragments with nicks in either strand. In 2012, Meyer et al introduced single-stranded library preparation for sequencing highly degraded DNA samples from an extinct archaic human. Several recent studies have also shown that single-strand library preparation improved the recovery of cfDNA from cancer patients, transplantation recipients and maternal plasma. It is of great significance to determine whether single-strand library preparation could indeed improve the performance of cfDNA analysis in clinical practice, especially in early-stage cancers.

Added value of this study

We introduced a unique application of single-strand library preparation to recover degraded or short cfDNA fragments in conception. Here we refined the genomic landscape of cell-free DNA in patients with pancreatic cancer, and found the shortened fragments bearing KRAS oncogenic hotspot alleles were more pronounced in patients with early pancreatic cancer. The ability of library preparation and analytical methods to enrich these short mutant fragments significantly enhanced the detection of KRAS mutations in plasma.

Implications of all the available evidence

Our observations found strong evidence that the fragmentation pattern of ctDNA may profoundly affect the detection of precancerous or very small cancer lesions. The biologic features of ctDNA may ultimately determine the ability to detect low abundant mutations in the circulation of patients with early-stage cancers. The single-strand approach could be a powerful tool for large-scale mutation measurements of cfDNA and has the potential to be generalized to cfDNA analysis for other common cancers.

However, genomic profiling of pancreatic cancer is challenging due to several limitations. First, pancreatic lesions often have a high stromal cell content, which may complicate the analysis of tissue sequencing [8–10]. Second, given the genetic instability and heterogeneity of pancreatic cancer, representative sampling is difficult for a comprehensive assessment of the genomic landscape of this disease [11–13]. Multiple sampling using fine-needle aspiration (FNA) biopsies for pancreatic cancer is both unsafe and unfeasible in clinical practice [14,15]. These challenges limit personalized medicine for pancreatic cancers. The possibility of detecting tumour-derived mutations in circulating cell-free DNA (cfDNA) provides a potential solution [16,17]. cfDNA profiling can provide a more comprehensive snapshot of the genomic landscape of pancreatic cancers than tissue-based sequencing [18,19]. Moreover, the minimally invasive nature of cfDNA analysis enables doctors to monitor disease progression [17,18,20]. However, cfDNA analysis using large-scale next-generation sequencing (NGS) has been largely confined to advanced or metastatic cancers, in which the ctDNA (circulating-tumour DNA) fraction is higher and is more readily detectable

[21,22]. Detection of ctDNA in early-stage cancers is still filled with cautionary tales that highlight the technical challenges in this field. Recently, Velculescu et al reported improved targeted error correction sequencing (TEC-Seq) as an efficient approach for the detection of ctDNA in early-stage colorectal, ovarian, breast, and lung cancers, with a detection rate ranging from 59% to 71% [23]. The massive parallel sequencing approach provides an applicable paradigm guiding cfDNA analysis in early-stage cancers.

However, the performance of existing approaches is limited in cfDNA analysis in early-stage pancreatic cancers. This limitation may be partially because of a lower ctDNA content in the circulation of pancreatic cancers [19]. Moreover, the biological features of ctDNA remain unclear, and this ambiguity may exacerbate the difficulty of detecting cancer-specific mutations in plasma of early-stage cancers. Recent advances have uncovered that cfDNA is a group of molecules with high diversification and degradation [24,25]. Conventional NGS or PCR approaches are not sensitive enough to capture the full diversity of these molecules, particularly degraded fragments with nicks in either strand. In 2012, Meyer et al introduced single-stranded library preparation for sequencing highly degraded DNA samples from an extinct archaic human [26]. The single-strand approach has shown higher efficiency in the recovery of degraded and shorter DNA fragments [27,28]. Several recent studies have also shown that single-strand library preparation improved the recovery of cfDNA from cancer patients [29], transplantation recipients [24] and maternal plasma [30]. It is of great significance to determine whether single-strand library preparation could indeed improve the performance of cfDNA analysis in clinical practice, especially in early-stage cancers.

Thus, we developed a single-strand library preparation and hybrid-capture-based cfDNA sequencing (SLHC-seq) approach and applied it to perform cfDNA profiling in a clinical cohort of 112 patients with pancreatic cancers and 28 healthy volunteers. SLHC-seq showed high sensitivity and specificity for detecting ctDNA in patients with pancreatic cancer across all stages. Using this new approach, the genomic landscape cfDNA was highly comparable to the results from direct tumour tissue sequencing. More importantly, single-strand library preparation efficiently enriched short cfDNA fragments (<100 bp), which may prominently originate from degraded ctDNA fragments with nicks in either strand. We demonstrated that the ability to recover these degraded fragments directly improved the sensitivity and accuracy of the detection of KRAS mutations in early-stage pancreatic cancer. Our observations highlight that the biologic features of ctDNA may ultimately determine the ability to detect precancerous or very small cancer lesions. The single-strand approach could be a powerful tool for large-scale mutation measurements of cfDNA and has the potential to be generalized to cfDNA analysis for other common cancers.

2. Methods and materials

2.1. Experimental design

The design schematic is presented in Fig. S1. A total of 112 patients with pancreatic lesions were recruited at Zhongshan Hospital Fudan University from July 2015 to May 2016. In addition, 32 age-matched healthy individuals were recruited as the control group. A total of 10 ml of treatment-naïve blood from each individual was collected using K₂-EDTA tubes (BD, Cat# 367863) and processed within 4 h. Formalin-fixed, paraffin-embedded (FFPE) tumour tissues (4 samples from FNA biopsy and 13 from surgical resection) were collected from the Department of Pathology at our center, and a single sample of resected tissue ($n = 13$) was collected from each patient without further dissection, representing the ‘mimic’ for biopsies. Patients were followed up every 3–5 months by a full-time clinical investigator as previously described [31,32].

2.2. Ethics approval and consent to participate

The study protocol was approved by the Ethics Committee of Zhongshan Hospital Fudan University (Approval No. B2014–098 and B2017–048). This included consent to publish from the participant to report individual patient data in our report.

2.3. Single-strand library preparation, targeted sequencing and bioinformatics analysis

The full details of the DNA extraction, library preparation, NGS and bioinformatics metrics are presented as below. Generally, cfDNA was first denatured into single-strand fragments after end repair. Then, these fragments were processed for library construction and hybrid-capture of targeted regions covering all exons of these 62 genes, which encompassed approximately a 221-kb region (Table S1). We achieved approximately 4000-fold coverage of sequencing. FreeBayes software (version 1.1.0–3) was used to call SNV/INDEL variants. The datasets of tissue-based sequencing were obtained from the The Cancer Genome Atlas (TCGA, cBioPortal datasets) and International Cancer Genome Consortium (ICGC) [33,34]. R packages ('maftools' version 3.5.0) [35] and custom scripts were used for further annotation, filtration and visualization of the dataset.

2.4. Extraction and preparation of cfDNA and genomic DNA

The clinical cohort and tissue bank was previous registered (<http://www.chictr.org.cn/>, ChiCTR-ROC-14005393). cfDNA was isolated using the QIAamp Circulating Nucleic Acid Kit (QIAGEN, Cat#55114) from 5 to 10 ml of plasma samples. Concentration of the total cfDNA was measured using the Qubit™ dsDNA HS Assay Kit (Thermo Fisher Scientific, Cat# Q32854), and the quality was analyzed using the Agilent High Sensitivity DNA Kit (Cat# 5067–4626) on an Agilent 2100 Bioanalyzer (Agilent Technologies). cfDNA samples with excessive high molecular weight DNA content were excluded from further processing. Genomic DNA of tissue samples was isolated using the Qiagen QIAamp DNA FFPE Tissue Kit (Qiagen, Cat# 56404) following the manufacturer's instructions. The concentration of the genomic DNA was measured using a Nanodrop 2000 (Thermo Scientific). Genomic DNA was fragmented to an average size of 200 bp using M220 Focused-ultrasonicator™ (Covaris, Inc.) following the manufacturer's instructions.

2.4.1. Pre-library construction

End repair of fragmented gDNA or intact cfDNA was performed using the VEE1 enzyme at 37 °C for 30 min. The DNA was then denatured at 95 °C for 5 min and snap chilled on ice. The VLE1 and VLE2 enzymes were used at 37 °C for 30 min. First, amplification was immediately performed to generate reverse complementary DNA molecules using the VAB1 PCR master mix with the following PCR program:

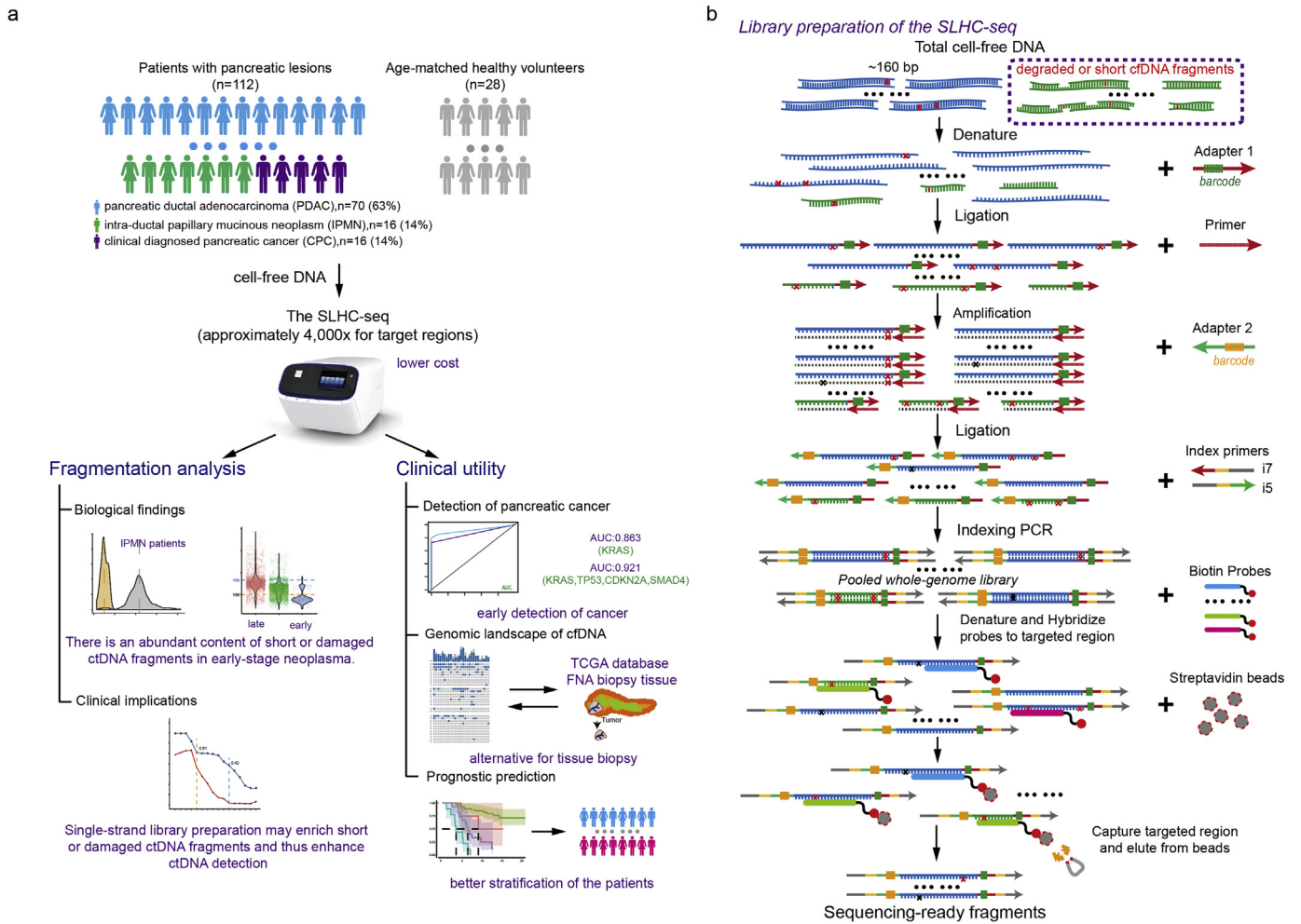


Fig. 1. The study design and single-strand library preparation of cfDNA in conception. cfDNA fragments were first denatured into single-strand DNA (ssDNA) fragments. Then, the ssDNA fragments were ligated with a unique molecular identifier. The pre-library was enriched using hybridization and magnetic-beads capture and then subjected to high-throughput sequencing.

1 cycle of 98 °C for 30 s; 4 cycles of 98 °C for 15 s, 60 °C for 30 s, and 72 °C for 30 s; and 1 cycle of 72 °C for 5 min. Amplified DNA was purified using the AMB1 Magnetic Beads and eluted in 20 µl of EB buffer. Another round of 3' end adaptor ligation was performed using the VSE1 and VSE2 enzymes at 37 °C for 30 min, followed by indexing PCR (i5 and i7) using the VIB1 PCR master mix with the following PCR program: 1 cycle of 95 °C for 3 min; 7 to 11 cycles (cycle number depending on the input DNA mass) of 98 °C for 15 s, 60 °C for 30 s, and 72 °C for 30 s; and 1 cycle of 72 °C for 5 min. The indexing PCR products were purified using IPB1 Magnetic Beads. The purified PCR products were whole genome libraries (pre-libraries), and their concentration was measured using the Qubit™ dsDNA HS Assay Kit. Pre-libraries <500 ng were disqualified for subsequent target enrichment.

2.4.2. Target enrichment

Target enrichment was performed using the AnchorDx SmartVisio™ Target Enrichment Kit (AnchorDx, Cat# AOUX00693). A total of 1000 ng of pre-libraries from up to 4 samples was pooled for one target enrichment using our custom-made panel focusing on 62 genes. Briefly, HE, HBA and HBB blocking reagents were added to the 1000 ng pooled pre-libraries and completely dried using a heated spin vacuum. The mixture was reconstituted in 7.5 µl of VHB1 hybridization buffer plus 3 µl of VHE1 hybridization enhancer. The mixture was next denatured at 95 °C for 10 min and immediately transferred to a hybridization oven preheated to 47 °C. The capture probes were added, and the mixture was quickly mixed and transferred to a thermocycler for hybridization incubation following the manufacturer's protocol. After hybridization, target DNA libraries captured by the biotinylated probes

were pulled down using M270 Dynabead streptavidin beads (Thermo Fisher Scientific, Cat# 65306). Briefly, 30 µl of M270 beads were washed twice with 1× Binding Wash Buffer, and the supernatant was removed. Pre-library pools were added and mixed thoroughly with the beads by repeated pipetting, and the mixture was incubated on a rotator at 47 °C for 45 min. After incubation for the probes to bind to the beads, 100 µl of 1× Transfer Buffer pre-warmed to 47 °C was added to the mixture, and the whole content was transferred to a magnetic stand. The supernatant was quickly removed and discarded as soon as it turned clear, and the beads were washed twice using 1× Stringent Wash Buffer pre-warmed to 47 °C. The beads were then washed once with 200 µl of room temperature 1× Wash Buffer I, once with 1× Wash Buffer II and then once with 1× Wash Buffer III. The enriched libraries bound on the beads were further amplified with P5 and P7 primers using the KAPA HiFi HotStart Ready Mix (KAPA Biosystems, Cat# KK2602) with the following PCR program: 1 cycle of 98 °C for 45 s; 12 cycles of 98 °C for 15 s, 60 °C for 30 s, and 72 °C for 30 s; and 1 cycle of 72 °C for 1 min. PCR products were purified with Agencourt AMPure XP Magnetic Beads (Beckman Coulter, Cat# A63882) and eluted in 40 µl of EB buffer. The concentration of the enriched library was measured using the Qubit dsDNA HS Assay.

2.4.3. Read mapping and SNV/INDEL analysis

Sequencing adapters and 3' low quality bases were trimmed from raw sequencing reads using Trim Galore, version 0.4.1 (<https://github.com/FelixKrueger/TrimGalore>). After adapter trimming and quality control, reads were mapped to the UCSC hg19 reference genome using BWA MEM (version 0.7.13) using the default settings.

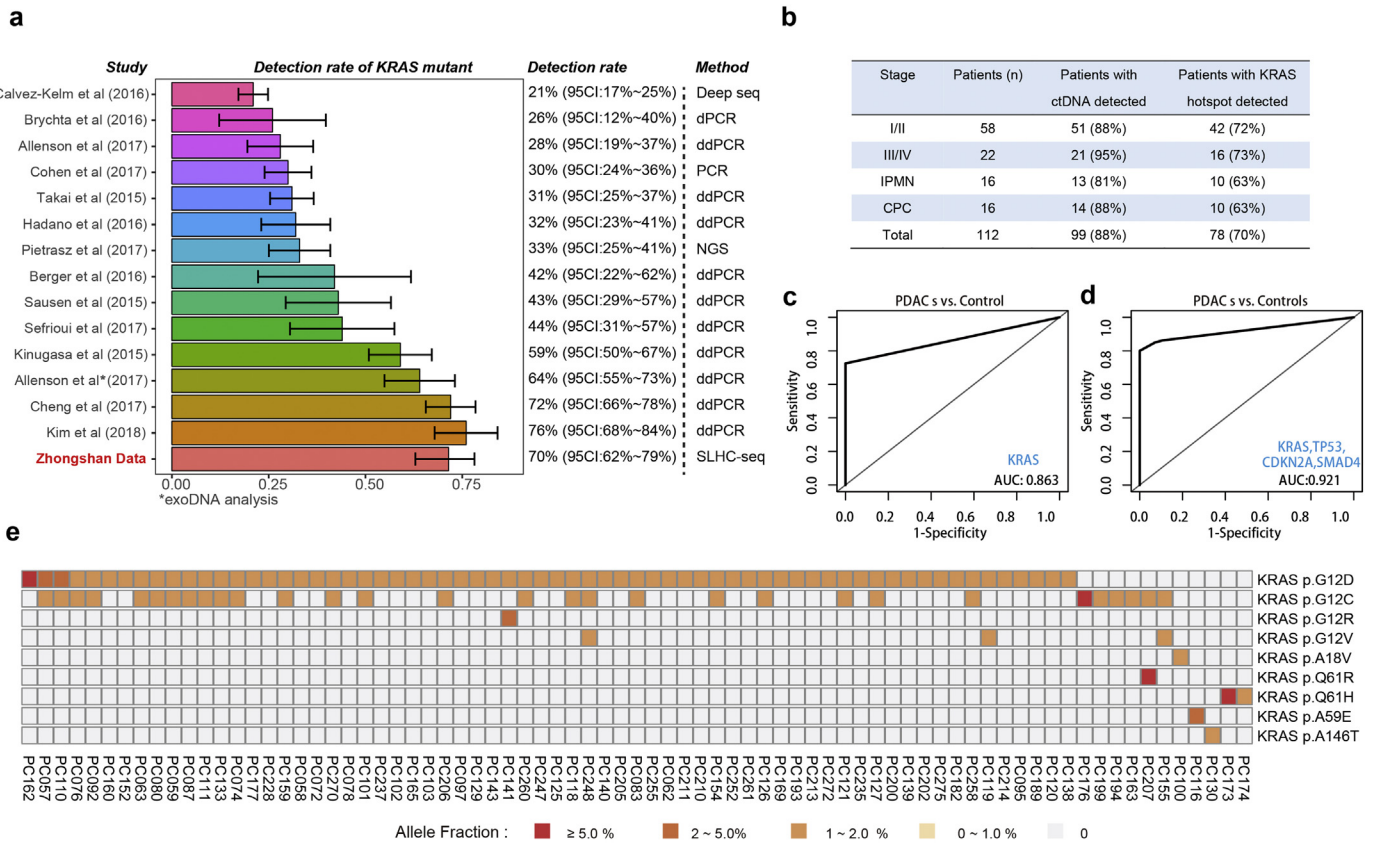


Fig. 2. Detection of KRAS mutations in plasma of patients with pancreatic cancers. (a) Comprehensive analysis of ctDNA-based measurements of KRAS mutations in 13 previous reports. (b) ctDNA and KRAS hotspot mutations detected in cancer patients using SLHC-Seq. Diagnosis of PDAC using detection of mutations in (c) KRAS gene or (d) a combination of KRAS, CDKN2A, TP53 and SMAD4. (e) Heatmap is presented to illustrate the KRAS mutational heterogeneity in plasma ctDNA. Colour scale of each square represents the allele frequency of mutations.

Next, custom scripts were applied to remove PCR duplications based on information regarding unique molecule identifiers (UMI). Finally, FreeBayes (version 1.1.0–3) was applied to call SNV/INDEL variants using the settings “-F 0.001–standard-filters–haplotype-length 0–min-coverage 100–C 5–report-all-haplotype-alleles”. Of note, the BAM was further checked using Integrative Genomics Viewer (IGV) to remove potential artifacts in *TP53* and *EGFR* genes. The R packages ‘maftools’ (version 3.4) ‘ggplot2’(version 2.2.1) and custom scripts were used for further analysis and visualization of the data (R version 3.4.0).

2.4.4. Identification the fragment length of cfDNA fragments bearing *KRAS* mutated or wild-type alleles

First, reads were cleaned by trimming adapters and low-quality bases. Next, cleaned reads were mapped to human reference genome (hg19) using BWA MEM, followed by unique molecule index (UMI)-based deduplication using our custom script. Finally, somatic variants were identified using Freebayes with the following parameters [-F 0.001 –standard-filters –min-coverage 100 –report-all-haplotype-alleles -C 5; depth ≥500, frequency ≥0.5%, AO ≥5 (for cfDNA analysis); depth ≥500, frequency ≥2%, AO ≥5 (for tissue analysis)]. For a variant site of interest, reads were grouped into REF and ALT by the genotype at the site, and insert size information was computed from the size field of the bam file after deduplication using the Picard tools (<https://github.com/broadinstitute/picard>). The estimation of number of copies of ctDNA bearing *KRAS* mutated alleles was mainly based on the assumption that there are 3.3 pg DNA per haploid copy of

genome, the formula was used for calculating mutant copies per ul of plasma through NGS method:

$$\text{Copies/}\mu\text{l of plasma} = \frac{\text{TM}}{\text{PV} \times 3.3 \text{ pg/copy}} \times \frac{\text{AF}(\%)}{100}$$

TM: Total mass of cfDNA (pg).

PV: Volume of plasma used for cfDNA extraction (ul).

AF(%): Mutated wild-type to allele fraction.

2.5. Statistical analyses

Chi-square test was used to analyze differences in categorical variables, while the Wilcoxon-Mann-Whitney test was used for continuous variables. Receiver operating characteristics (ROC) were used to evaluate the diagnostic power. In the univariate survival analysis, maximally selected rank statistics were used to determine optimal cutoffs for continuous variables, and the Kaplan-Meier method was next used to draft survival curves (log-rank test). Cox proportional hazards model was used to calculate the hazard ratio (HR) in univariate or multivariable survival analysis. Two-sided *p*-values <.05 were considered significant. All statistical analyses were performed using R statistical project (version 3.4.0).

2.5.1. Trial registration

ChiCTR-ROC-14005393 (<http://www.chictr.org.cn/>).

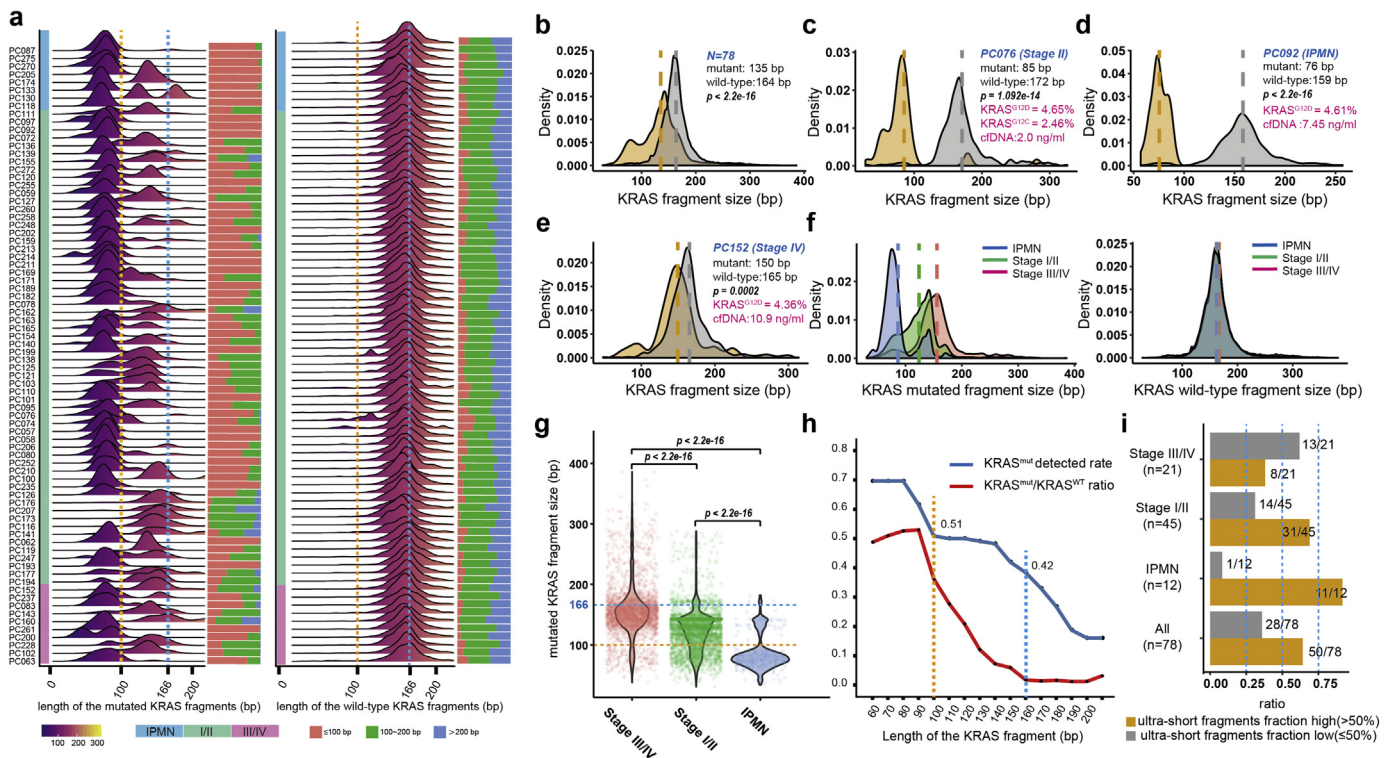


Fig. 3. Ultra-short fragments containing the *KRAS* hotspot alleles in patients with pancreatic cancer. (a) Comprehensive view of fragment lengths bearing *KRAS* mutated and wild-type alleles derived from cell-free DNA deep-sequencing in patients with pancreatic cancer (*n* = 78). (b) The fragment length of cfDNA bearing mutant *KRAS* alleles tended to be significantly shorter compared with DNA fragments bearing wild-type allele by densitometry in pancreatic cancer (*n* = 78). The fragments bearing *KRAS* mutant alleles tended to be enriched in the region below 100 bp in patients (c) PC076 (stage II) and (d) PC092 (IPMN). (e) The distribution of *KRAS* mutated fragments included a large proportion of overlapping fragment sizes with wild-type *KRAS* fragments in advanced patients (PC152). (f) There was considerable discrepancy across IPMN (*n* = 10), early-stage (*n* = 42) and advanced (*n* = 26) pancreatic cancer in terms of the length of *KRAS* mutated fragments by densitometry. (g) Violin plots representing the length of *KRAS* mutated fragments across different disease stages. (h) The ability of library preparation to enrich shorter fragments has a great impact on the detection of *KRAS* hotspot mutations in plasma (blue line), and the mutated-to-wild-type fraction of fragments bearing *KRAS* alleles reached the highest in the interval of 60–100 bp and declined sharply thereafter (red line). (i) Ultra-short fragments (<100 bp) accounted for a quite high proportion of total *KRAS* mutated fragments in a substantial group of patients with pancreatic cancer.

3. Results

3.1. Single-strand library preparation and hybrid-capture-based cfDNA sequencing

Based on the hypothesis that single-strand DNA library preparation is, in principle, more sensitive to recover the full spectrum of cfDNA fragments from plasma, we developed a single-strand library preparation and hybrid-capture-based cfDNA sequencing methodology (SLHC-seq, Fig. 1a, b). First, we selected the 62 most recurrently mutated genes in common cancer types according to the TCGA, ICGC database [36] and large-scale cohorts relevant to pancreatic cancer sequencing [5,9,10,37]. Then, we designed hybrid-capture probes covering all exons of these 62 genes, which encompassed an approximately 211-kb region (Table S1). Theoretically, the single-strand library preparation should improve the conversion efficiency of degraded DNA fragments (especially fragments with nicks in either strand) compared with that of conventional methods. The following steps were similar to those in the targeted sequencing of cancer tissue. Bioinformatics analyses were performed to identify tumour-specific mutations and eliminate potential artifacts (Fig. S1). These analyses mainly included the following: (i) variant calling using consensus sequences built from reads sharing identical unique molecular identifiers and mapped coordinates

[38,39]; (ii) removing germline mutations according to the 1000 Genome database [40], or an observed allele frequency of $\geq 40\%$ (p -value ≤ 0.1 for observing a germline variant at such frequency at coverage > 500); and (iii) enrichment of likely cancer-specific mutations by removal of alternations not identified in common cancer types according to the COSMIC database [41] (instances of overlapping mutations ≤ 2 were removed) (Fig. S1).

We next compared the performance of SLHC-seq with that of conventional NGS methods for detecting known cancer-specific mutations, including single nucleotide variants and small insertions and deletions, in reference genomic DNA. Generally, SLHC-seq and conventional approaches both showed high sensitivity and specificity in the analysis of mutations with higher fractions. Initially, with 1 ng of DNA input, SLHC-seq, compared with the conventional approach, showed a significantly higher efficiency for the detection of mutations in the reference genome (detection rate: 75% vs. 45%, respectively, $p < .01$, Fig. S2). Then, we evaluated the performance of SLHC-seq using a mixture of references at different dilutions. Initially, with 5 ng of DNA input, SLHC-seq identified over 50%, 70%, 92% and 92% mutations in 10, 5, 3, 2-fold dilutions, respectively. Further analysis of the variant alleles achieved high concordance to the expected fraction of mutations (adjusted $R^2 = 0.94$; $p < .0001$, Fig. S2, Table S2), as well as high sensitivity and specificity. Overall, the analytical sensitivity was 92%, 88% and 83% for

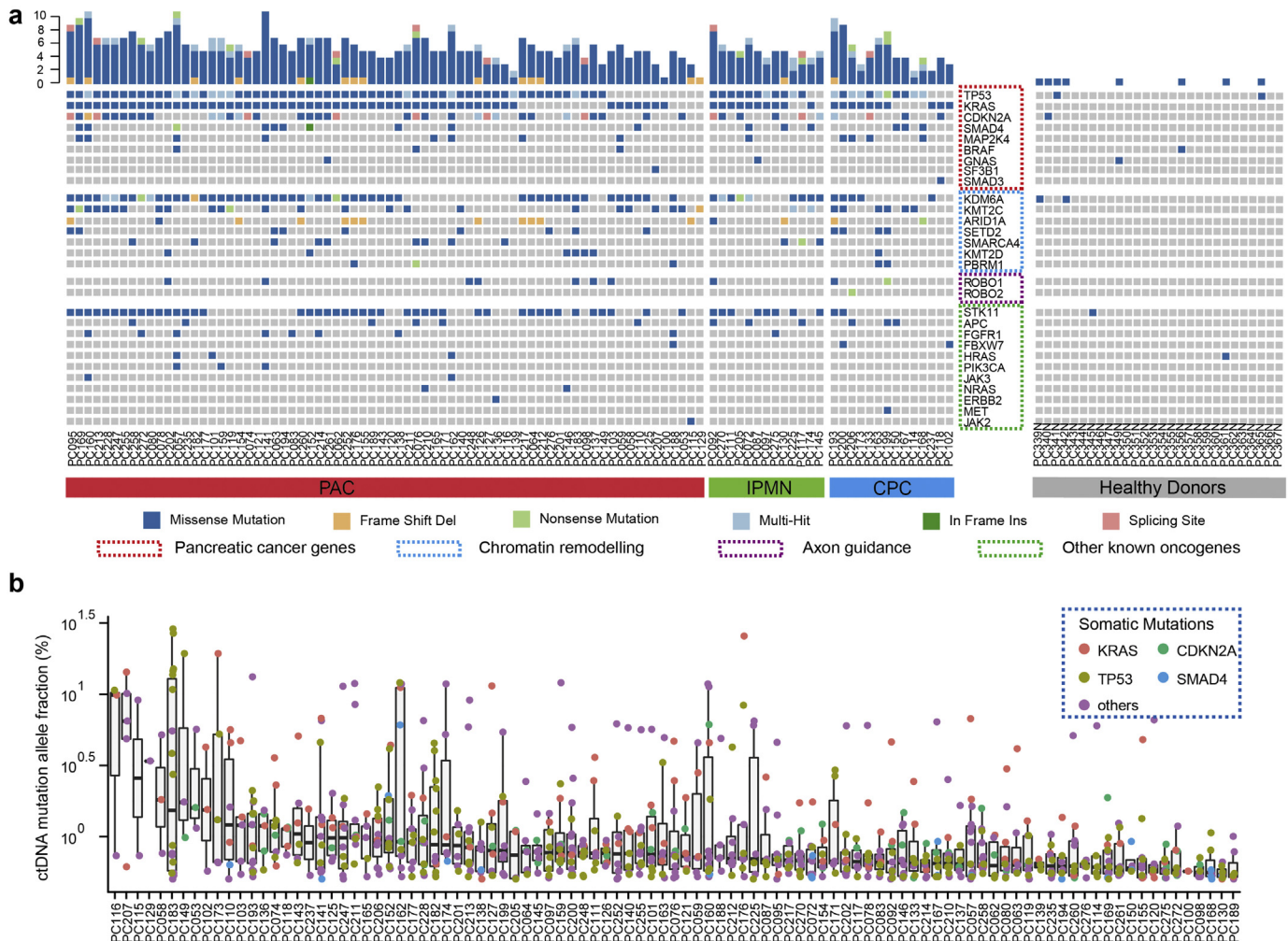


Fig. 4. Genomic landscape of cfDNA in patients with pancreatic cancers. (a) Genomic landscape of cfDNA in patients with pancreatic cancers. A total of 792 somatic mutations were identified in the plasma of 88% patients. The mutations were highly prevalent in common pancreatic cancer genes and in genes involved in chromatin remodeling, and axon guidance pathways. A total of 8 mutations were detected in the plasma of the healthy controls ($n = 28$). The upper bars indicate the total mutation count for each sample. (b) The allele fraction of cfDNA mutation in each patient is shown; the box was sorted according to the median value of the mutational allele fraction in each individual. The colored dots represent mutations on different genes.

detecting mutations using a threshold allele fraction of 0.5%, 0.2% and 0.1%, respectively. No false-positive mutations of the identified alleles were detected in the blank control using the SLHC-seq approach. According to the results, an allele fraction of 0.5% was determined as the threshold for identifying cancer-specific mutations in plasma cfDNA.

3.2. Detection of KRAS mutations using SLHC-seq in the circulation of pancreatic cancers

Of the 112 included patients, 70 were histologically diagnosed as having pancreatic ductal adenocarcinoma (PDAC), 16 were intra-ductal papillary mucinous neoplasm (IPMN), and 16 were clinically diagnosed pancreatic cancer (CPC) (Fig. S3, Table S3). The median cfDNA concentration across the pancreatic patient group was 16.2 ng/ml (IQR: 9.3–25.9 ng/ml), which was significantly higher than that in the healthy individual group (median 3.2, $p < .0001$, Wilcoxon-Mann-Whitney test). The total cfDNA concentration of the patients was independent of tumour stage and CA19–9 value, while was increased with radiological tumour sizes in patients with localized PDAC (spearman rho = 0.4021, $p = .0042$, Fig. S3).

Detection of KRAS hotspot mutations in the circulation is often used to detect pancreatic cancer. We compared the performance of our approach in measuring KRAS mutations with that in previous

reports using PCR- or NGS-based methods (Table S4) [42–54]. Across 13 high-quality studies, the detection rate of the KRAS mutations in plasma ranged from 21% to 76% (Fig. 2a). In our cohort, which included a substantial proportion of early-stage patients, tumour-specific mutations were detected in 88% (99/112) of patients, with KRAS hotspot mutations detected in 70% [95% confidence interval (CI): 62–79%] of patients (Fig. 2b). Of note, SLHC-seq largely outperformed most reports using the ddPCR method in the detection of KRAS mutations in early-stage diseases. Using the SLHC-seq approach, the detection of KRAS mutations served as an efficient marker to distinguish PDAC from healthy individuals in our cohort [AUC (area under curve): 0.863, 95% CI: 0.830–0.898, Fig. 2c]. By utilizing a combination of the KRAS, TP53, CDKN2A and SMAD4 genes, we achieved a higher performance in the diagnosis of PDAC (AUC: 0.921, 95% CI: 0.890–0.956, sensitivity: 80%, specificity: 100%, Fig. 2d). Diagnostic accuracy was highest when using the full targeted panel, with an AUC of 0.951 (95% CI: 0.932–0.983) with an optimum specificity of 100% and sensitivity of 89%. Additionally, the detection of mutations using the full targeted panel also had the ability to distinguish IPMN from PDAC patients (AUC: 0.837, 95% CI: 0.821–0.979, specificity: 87.5%, sensitivity: 66.2%).

Of note, multiple KRAS hotspot mutations, including G12D ($n = 67$), G12C ($n = 30$), were detected as well as other hotspot mutations at a

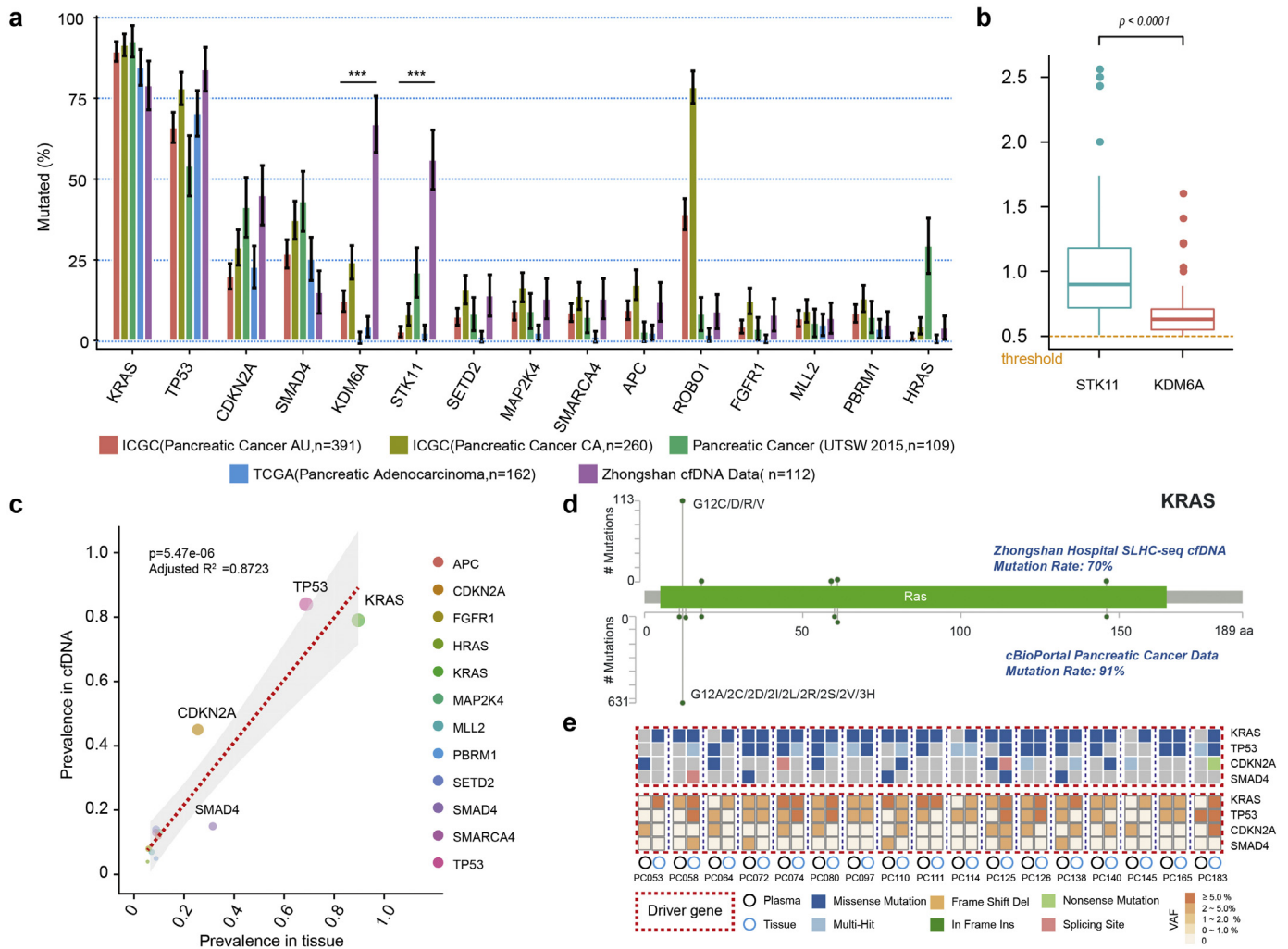
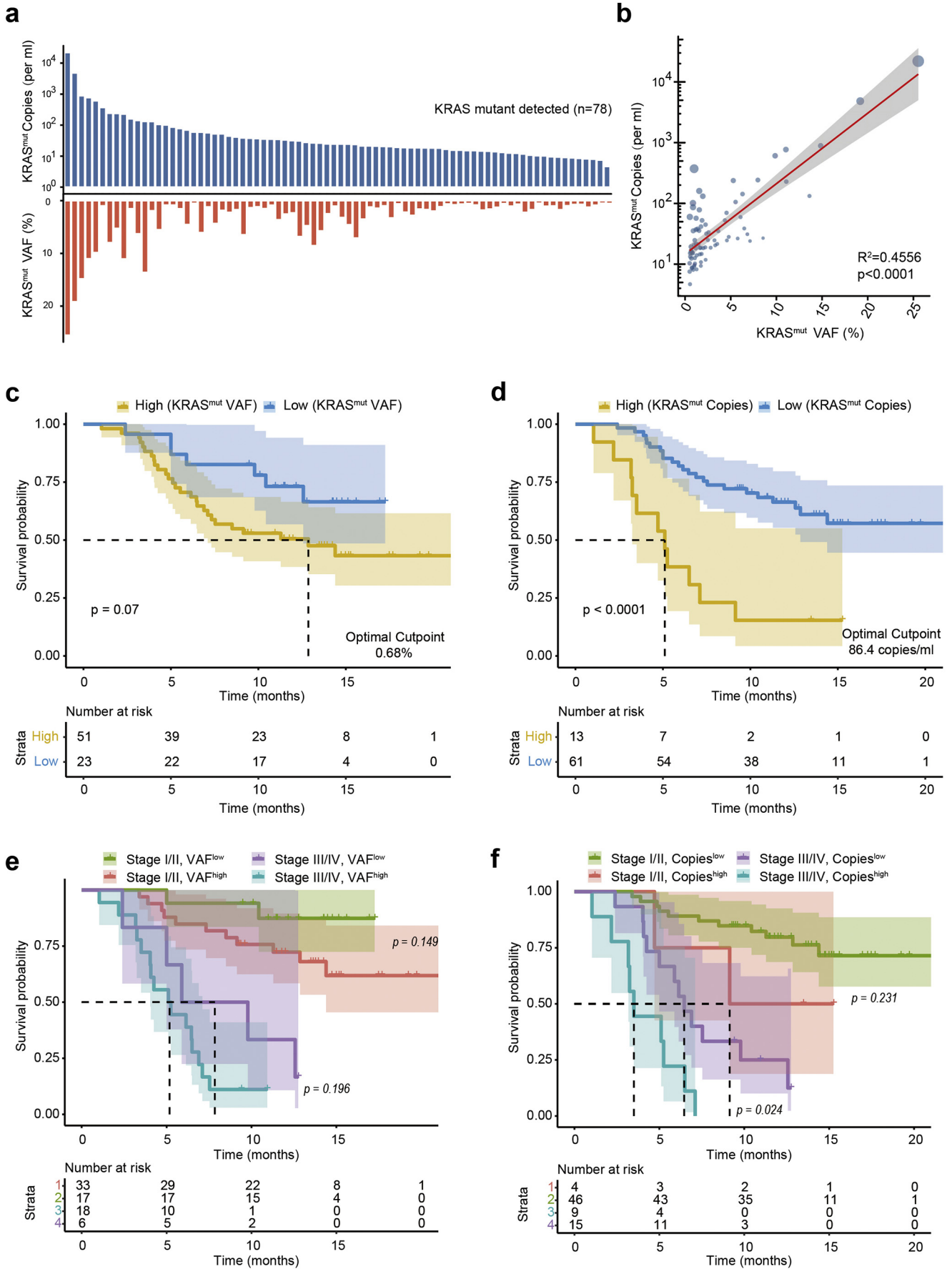


Fig. 5. Genomic landscape of cfDNA or tumour tissue sequencing in pancreatic cancer. (a) Comparison of mutational prevalence in cfDNA and tissue-based cohorts (SNV/InDel only). The top 15 mutated genes in cfDNA are listed. (b) Comparison of the mutational allele fraction of *STK11* and *KDM6A* in cfDNA. (c) Correlation between mutational prevalence in cfDNA versus tissue-based datasets (*STK11*, *KDM6A* and *ROBO1* were removed). (d) The amino acid changes of *KRAS* predicted by cfDNA profiling versus tissue (cBioPortal). (e) Mutation profiling in cfDNA and paired biopsy tissue. The heatmap represents the mutation types, and the lower heatmap indicates summarized variant allele fractions (including clonal and subclonal) for each gene.



lower prevalence (Fig. 2e). We identified 22 samples with a dominant KRAS clone concurrent with sub-clonal KRAS mutations with lower allele fraction, suggesting that these patients may harbor multiple KRAS mutations in tumour tissues.

3.3. SLHC-seq enhanced detection of pancreatic cancers via enrichment of ultra-short fragments bearing KRAS hotspot alleles

A pooled analysis of 96 patients and 28 healthy individuals show that the fragment length of total cfDNA between the two groups was similar (median: 161 vs. 162 bp). Utilizing single-strand library preparation, we explored the fragment length of cfDNA with a focus on the KRAS gene, given the extremely high prevalence of KRAS mutations in pancreatic cancers. Interestingly, a large fraction of patients with pancreatic tumours have mutated KRAS fragments below 100 bp, while the fragments containing wild-type KRAS alleles of these patients are homogeneously approximately 160 bp in length (Fig. 3a). A pooled analysis of 78 patients showed that the ctDNA fragments bearing the KRAS mutant alleles, which originated from the pancreatic cancer, were significantly shorter than their wild-type counterparts (median: 135 vs. 164 bp, $p < .0001$, Fig. 3b).

Strikingly, the shortened fragments were more pronounced in pre-cancerous IPMN and early-stage PDAC (stage I/II) patients. For example, in one case with IPMN (PC092), the fragments bearing the KRAS^{G12D} and KRAS^{G12C} allele were enriched in the region with a shorter length, and a dominant peak 75 bp in size (Fig. 3c). In another patient with stage II resectable PDAC (PC076), the distribution of fragments bearing the KRAS^{G12D} allele also showed significant deviation from the wild-type fragments with a dominant peak 85 bp in length (Fig. 3d). In contrast, in an advanced-stage patient (PC216) with a high mutated KRAS fraction (~25%), we observed a broad distribution of mutated fragments including a major proportion of fragments that overlapped with the wild-type fragments at a peak approximately 135–180 bp in length (Fig. 3e). The median length of the mutated fragment in IPMN ($n = 10$), early-stage ($n = 58$), and advanced-stage ($n = 10$) PDAC was approximately 80, 140, 160 bp, respectively ($p < .0001$, Fig. 3f, g), while the wild-type counterparts had a length of ~160 bp across the subgroups (Fig. 3f). Of note, the relative wide distribution of wild-type fragments also included a small proportion of overlapping short fragments with the KRAS mutants, possibly as a result of the heterozygous features of KRAS mutations.

Subsequently, the length of fragments was binned into subgroups (60 = 0–60 bp, 70 = 71–70 bp, etc.), and the mutated-to-wild-type fragment ratio was calculated for each interval. The proportion of the fragments bearing KRAS mutations reached the maximum in the interval of 60–100 bp and sharply declined thereafter. The proportion of the KRAS mutated fractions declined to <5% for regions of 150–200 bp in length. Importantly, we found that the ability of library preparation methods to enrich these shorter mutated fragments had a significant impact on the performance of detecting KRAS mutations in plasma. The detection rate of KRAS mutations significantly decreased when setting the threshold at 100 bp or 160 bp (100 bp: 0.70 vs. 0.51, respectively, $p < .05$; 160 bp: 0.70 vs. 0.42, respectively, $p < .05$, Fig. 3h). The ability to recover ultra-short fragments also affected the quantification of KRAS mutations in the plasma of patients with pancreatic cancer (Fig. 3i).

3.4. Genomic landscape of cell-free DNA in patients with pancreatic cancer

We next explored the genomic landscape of cfDNA in pancreatic cancer using SLHC-seq.

Generally, we identified 791 somatic mutations in the plasma of 88% (99/112) patients (approximately 4000-fold coverage, Fig. S4). In

contrast, 8 mutations were detected in the plasma of the healthy controls ($n = 28$) without any known oncogenic or truncating alleles. The average burden of cancer-related mutations was significantly higher in the plasma of patients with detectable ctDNA than that in healthy individuals (8.0 vs. 0.29 per individual, $p < .01$). Mutations in the TP53, KRAS, CDKN2A and SMAD4 genes were detected in 86%, 79%, 70%, and 58% of plasma cfDNA samples, respectively. In addition to these driver genes, we also observed a high prevalence of mutated genes in the chromatin-remodeling pathway (KDM6A, AIRD1A, and PBRM1) and axon guidance (ROBO1, BOBO2) (Fig. 4a, b). KRAS mutations (KRAS^{G12D} and KRAS^{G12V}) were the mostly recurrent mutations in the cohort, and we also observed R175H and R273H mutations of TP53, both frequently present across cancers. In addition, we observed that 13% (15/112) of patients harbored actionable mutations, including mutations of BRAF, ERBB2, PIK3CA, JAK2 and JAK3 genes [55]. The BRAF^{V600E} mutation was detected in 3 (2.6%) patients with allele fraction ranging from 0.7% to 0.9%. JAK3^{A573V}, recurrently mutated in hematopoietic malignancies, was observed in 2 (1.8%) patients. JAK2^{V617F} was detected in the plasma of a 67-year-old male patient independent of KRAS hotspots mutations (allele fraction: 9.1%, Fig. S4). While several JAK1/JAK2 inhibitors have been reported to have efficacy in cancer patients with JAK2-activating mutations, these observations indicated that large-scale cfDNA test might help to identify individuals who would benefit from genotype-driven therapy trials [56].

3.5. Comparison of cfDNA profiling and tissue-based sequencing of pancreatic cancers

We next compared the cfDNA profiling results in our cohort and tissue-based sequencing data with 4 publicly available cohorts comprising 922 pancreatic cancer patients. For consistency, the top 15 mutated genes with prevalence over 5% in cfDNA were included in the final analysis. Generally, the cfDNA profiling in patients with detectable ctDNA was largely consistent with the results of the tissue-based sequencing database. The prevalence of KRAS, TP53, CDKN2A and SMAD4 mutations was highly correlated in plasma cfDNA and tissue-based cohorts (Fig. 5a). The mutational prevalence of the most commonly mutated genes in cfDNA was significantly associated with the mutational prevalence in tumour tissues (adjusted $R^2 = 0.456$, $p = .0057$, Fig. S5). However, there was a significant discrepancy in the prevalence of KDM6A (67% vs. 13%, $p < .001$) and STK11 (56% vs. 6%, $p < .001$) between ctDNA profiling and tissue-based sequencing (Fig. 5a, b). The prevalence of ROBO1 mutations was also not uniform across the tissue-based sequencing dataset. After the removal of KDM6A, STK11 and ROBO1, we observed a stronger correlation of the genomic landscape of cfDNA and tumour tissue profiling (adjusted $R^2 = 0.8723$, $p = 5.47e-06$, Fig. 5c). The amino acid changes of KRAS predicted by cfDNA profiling were highly consistent with tissue-based sequencing (cBioPortal, Fig. 5d).

Additionally, we assessed the concordance of mutations in cfDNA and DNA from paired tumour tissues ($n = 17$, Fig. S5). Tumour cellularity ranged approximately from 1% to 15% in the biopsy tissue. The concordance of KRAS, TP53, and CDKN2A alternations between tissue DNA and cfDNA was 75.3%, 58.8%, and 41.2%, respectively (Fig. 5e, Table S2). SMAD4 mutations were sporadically detected in 3 plasma samples and 2 tissue samples.

3.6. ctDNA level and disease burden

In our study set, 74 patients were followed-up for a median time of approximately 2 years. The median overall survival of the patients was

Fig. 6. ctDNA level in plasma correlated with patient prognosis. (a) Relative and absolute levels of ctDNA bearing a KRAS mutation in each individual. (b) Quantitative correlation between the relative and absolute levels of KRAS mutations (the size of the blue dots represents the total cfDNA concentration in each case). The (c) relative and (d) absolute levels of KRAS mutation in plasma strongly correlated with the overall survival of the patients. (e–f) The absolute level of KRAS mutation in plasma was superior to the relative fraction for prognostic prediction together with AJCC tumour staging (edition 7).

15.9 months. Higher levels of CA19–9, CA125, CEA and total cfDNA concentration strongly correlated with poorer overall survival (Table S4).

Allele fraction of the mutations was previously reported as an indicator of plasma ctDNA burden. However, the concentration of total cfDNA is not specific to the specific cancer and is associated with conditions such as cancer-related inflammation. Thus, we estimated both the relative and absolute levels of *KRAS* mutations using the fraction of mutant-to-wild-type fragments and mutated ctDNA copies per ml in blood (Fig. 6a). The copies of *KRAS* mutated fragments (copies/ml) was associated with the mutant allele fraction and was also affected by total cfDNA quantity ($R^2 = 0.4456$, $p < .0001$, Fig. 6b). A higher *KRAS* mutant fraction was associated with a poor prognosis (HR = 2.1, 95% CI: 0.9–4.8, $p = .077$, Fig. 6c) but the difference did not reach significance. In contrast, the absolute quantification of *KRAS* mutations (copies/ml) directly marked patients with a worse prognosis to a good prognosis (HR = 4.3, 95% CI: 2.1–9.1, $p < .001$, Fig. 6d). After adjusting for AJCC stage, the absolute quantification of *KRAS* mutations still directly marked patients with a worse prognosis (HR = 2.9, 95% CI: 1.4–6.3, $p = .007$, Fig. 6e, f). In the multivariable analysis, the copy number was a significant marker of OS independent of AJCC Stage, CA19–9, CA125 and CEA level (HR: 3.3, 95% CI: 1.1–10.6; $p = .037$, cox proportional hazards model, Table S5). Together, our results suggest that accurate quantification of *KRAS* mutations in plasma may provide a useful marker for prognostic prediction in patients with pancreatic cancer across all stages.

4. Discussion

Here by application of single-strand library preparation, we refined the cfDNA profiling in a large cohort of patients with pancreatic cancer across all stages. The cfDNA profiling of pancreatic cancer was highly comparable to tissue-based sequencing in terms of the mutational landscape. Based on previous research, we further revealed the different fragmentation patterns of ctDNA between early and advanced pancreatic neoplasms. Together with several current reports, the results demonstrate that the fragmentation pattern of ctDNA may largely affect the detection and quantification of ctDNA in the circulation. These findings facilitate a deeper understanding of ctDNA biology and support the validity and feasibility of single-strand library preparation for ctDNA analysis in early-stage cancers.

cfDNA analysis could provide a non-invasive approach for cancer diagnosis and surveillance of drug resistance and may also encompass the heterogeneity of tumour genomics [16,18,23,51]. Recently, studies focusing on the Circulating Cell-free Genome Atlas (CCGA) have been performed to identify the genomic landscape of cfDNA in cancer patients and healthy participants [19]. This study will provide critically important knowledge regarding cfDNA profiling and enhance the detection of cancers. Despite this promise, the identification of ctDNA signals is technically challenging, especially in non-metastatic solid tumours [19]. Conceptually, a benefit of our new approach is the adoption of a single-strand library preparation, which can recover the full spectrum of cfDNA fragments in plasma and thus improve the applicability for the detection of tumour-specific mutations in early-stage cancers. Here, we achieved an effective sensitivity in ctDNA detection across all stages of pancreatic cancer with only approximately 4000-fold sequencing coverage. Of note, the detection of *KRAS* mutations in the plasma of pancreatic cancer patients has been studied for a long time, but with significant bias towards late-stage patients. Using this new approach, *KRAS* hotspot mutations were detected in over 70% of patients in our cohort, which comprised 66% of patients with precancerous or early-stage disease, while no hotspot mutations were detected in the circulation of healthy individuals. The performance of SLHC-seq in identifying ctDNA was highly comparable to that of ddPCR-based methods or NGS with ultra-deep sequencing depth (10,000-fold coverage or more). The high sensitivity and reliability of SLHC-seq make it highly suitable for early cancer detection and intervention, and its ability for large-scale parallel sequencing of multiple and broad genomic regions also

enables the quantitative surveillance of tumour-derived mutations that may evolve during early tumourigenesis, tumour progression, and drug resistance.

Next, we tried to analyze the fragmentation pattern of cfDNA molecules to find out why SLHC-seq may improve the detection of tumour-derived signals. In previous reports, the periodical shortening pattern of ctDNA fragments has been well described in xenograft animal models or in patients with late-stage cancers [25,57]. In two important studies published recently, researchers found that size-based analysis of cfDNA fragments improved the detection of tumour-derived DNA fragments [58,59]. Compared with their experimental pipelines, the advantages of the single-strand library preparation were to recover the full spectrum of cfDNA fragments. First, the different shortening pattern of cfDNA and ctDNA in our results was consistent with these previous reports. Then, emergence of *KRAS* hotspot mutations has been globally recognized as the initialization of pancreatic neoplasia, analyzing fragments bearing *KRAS* oncogenic mutations provides an ideal model for further exploring the biological features of ctDNA during early tumorigenesis. Here, we identified a substantial subset of *KRAS* mutated fragments with a length <100 bp, and this phenomenon was more pronounced in early-stage disease. Strikingly, almost all fragments bearing *KRAS* mutations were enriched at the region of <100 bp in length in quite a large portion of patients. The enrichment of ultra-short fragments (<100 bp) in ctDNA has hardly been reported before. Based on previous reports, we thought these observations could be more reasonably explained using a ctDNA model with widespread DNA damage (particularly with nicks in either strand), rather than a theory of efficient enrichment of ultra-short fragments [24,27,29]. While conventional library preparation worked poorly in the recovery of damaged fragments, the single-strand approach retains more biological information from those damaged ctDNA fragments [26,27,29]. The ability to enrich these damaged ctDNA fragments, which may recover quite a high proportion of ctDNA in early-stage cancers, directly affects the detection and quantification of ctDNA in circulation. As ctDNA is presumably shed from tumours either through necrosis or apoptosis, we propose a plausible hypothesis to explain the different fragmentation pattern of ctDNA between early- and late-stage cancers. ctDNA in early-stage cancers mainly originates from cancer cells that have disintegrated via apoptosis and immune clearance procedures, in which the endonuclease activation and uptake of macrophages may account for the widespread damage in DNA fragments. In advanced-stage cancer, hypoxia-related necrosis occurs due to increasing tumour size and deficient vascularization. Large amounts of tumour-derived DNA fragments may not be efficiently degraded through endogenous and exogenous pathways. These less degraded DNA fragments shedding into the circulation could be detected via conventional approaches while the highly degraded fragments could not. This hypothesis may, in principle, provide a plausible explanation for the low detection rate of *KRAS* mutations in early-stage pancreatic cancer even using extremely sensitive approaches. Collectively, our findings provide strong evidence that the biologic features of ctDNA may ultimately determine the ability to detect precancerous or very small cancer lesions [23].

Another important aspect of our study was to explore the utility of SLHC-seq in clinical practice. First, SLHC-seq improved the detection of *KRAS* hotspot mutations in early-stage pancreatic cancers. Then, by utilizing SLHC-seq, the genomic profiling of cfDNA showed high similarity to the genomic alternations identified in 4 tissue-based sequencing databases of pancreatic cancer. In a subset analysis of 17 paired plasma and biopsy tissue samples, cfDNA profiling accurately reproduced the landscape of driver genes (*KRAS*, *TP53* and *CDKN2A*) highly mutated in biopsy tissues. As biopsy tissues of pancreatic lesions are not routinely accessible, the high concordance of the genomic landscape between cfDNA and tumour tissues suggested that highly efficient cfDNA analysis could complement tissue-based sequencing in clinical practice. Further, cfDNA analysis theoretically provides more comprehensive information on tumour-derived mutations that may otherwise be missed due to the

selection bias of tissue sampling and thus provides new insights into cancer biology. In our cohort, we observed mutational heterogeneity, as shown by the presence of distinct KRAS hotspot mutations within a cfDNA sample. Though tumour heterogeneity has always been emphasized in oncology, sub-clonal mutations may be missed due to technical limitations and the selection bias of current approaches in tissue analysis. Through the application of multiple sampling and microdissection, recent studies have identified polyclonal KRAS mutations in the tissue of papillary mucinous neoplasms [60,61] and ductal adenocarcinoma of the pancreas [62]. Complementing this progress, our observations strongly indicated that polyclonal origin may be a common feature during the formation of pancreatic tumours, as polyclonal KRAS mutations in approximately 20% patients with pancreatic neoplasm have been observed. As previous researchers have identified that distinct KRAS mutations play different roles in the oncogenic pathway of lung cancer, further studies are needed to uncover the role of multiple RAS pathways in pancreatic cancer. At last, ctDNA fragments bearing KRAS hotspot alleles are directly released from the tumour. We demonstrated the accurate quantification of KRAS mutations in plasma may provide a useful marker for prognostic prediction in patients with pancreatic cancer across all stages.

Several limitations of this study should be emphasized. First, alterations originating from a hematopoietic clone may confound the tumour-specific mutations, which may lead to differences between the cfDNA profiling and tissue-based sequencing and over diagnosis. Going forward, investigations are needed to determine whether these mutations are clinically meaningful over time in healthy individuals. Second, though we identified a subgroup of patients with potentially actionable mutations, correlation between clinical drug response and detection of targeted mutations still need further investigation. Third, chronic pancreatitis should be used as a control ideally. However, it's extreme difficult to collect blood sample from age-matched visiting patients with chronic pancreatitis in most circumstance. Thus, like most of previous studies, we used samples from age-matched healthy volunteers as an alternative. Lastly, sensitivity of ctDNA detection may be further improved by sequencing with deeper depth or repeated sampling.

5. Conclusions

Profiling somatic alternations has become routine for cancer patients, but the non-invasive nature of cfDNA analysis makes it promising for broad application in terms of clinical use. We introduced a unique application of single-strand library preparation to recover degraded or short cfDNA fragments in conception. Here we refined the genomic landscape of cell-free DNA in patients with pancreatic cancer, and found strong evidence that fragmentation pattern of ctDNA may profoundly affect the detection of precancerous or very small cancer lesions. We further demonstrated the clinical significance of this approach that high efficient cfDNA analysis could be an alternative to tissue-based sequencing and efficient biomarkers for pancreatic cancer in clinic. Together with other recent advances, our results provide a deep understanding of cfDNA biology and may promote the development of new adaptive approaches for ctDNA-based testing in the future.

Funding source

This work is supported by National Natural Science Foundation of China (NSFC) Grants (#31625013, #91732302, Recipient: ZLQ), Rong-Chang Charity Fund for Pancreatic Cancer Research (Recipient: XLW), Zhongshan Hospital Fund for Young Scholars (#2015ZSQN31, Recipient: CYL), National High Technology Research and Development Program of China (No. S2015AA020405, Recipient: WHL).

Author contributions

Study concept and design: XYL, XLW, ZLQ, WHL; Acquisition of data: XYL, LXL, XRY, GYB, DYJ, SXR, YFZ, WHL; Analysis and interpretation of data: XYL, WHX, Drafting of the manuscript: XYL; Critical revision of the manuscript for important intellectual content: ZLQ, YBG, XYC, WT; Statistical and bioinformatics metrics: XYL, WHX; Technical, or material support: WHL, CYL, YFZ, YBG, CYL, LXL; Study supervision: ZLQ, XLW, WHL.

Declaration of interest

The authors declare that they have no conflict of interest.

Data and materials availability

Mutation calls from FreeBayes and annotation are listed in Table S2. All other relevant data are within the paper and its supplementary files. According to ethical statements, interested researchers may contact Dr. Zilong Qiu (zqiu@ion.ac.cn) to obtain raw human sequencing data.

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ebiom.2019.02.010>.

Acknowledgments

We thank Dr. Yang Zhou, Dr. Junhao Li, Dr. Zhuqin Cao, Dr. Hao Liu and Ting Yang for sample collection and follow-up of the patients.

References

- [1] Siegel RL, Miller KD, Jemal A. Cancer statistics, 2017. *CA Cancer J Clin* 2017;Vol. 67: 7–30.
- [2] Rahib L, Smith BD, Aizenberg R, Rosenzweig AB, Fleshman JM, Matrisian LM. Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States. *Cancer Res* 2014;74:2913–21.
- [3] Kamisawa T, Wood LD, Itoi T, Takaori K. Pancreatic cancer. *Lancet* 2016;388:73–85.
- [4] Lennon AM, Wolfgang CL, Canto MI, Klein AP, Herman JM, Goggins M, et al. The early detection of pancreatic cancer: what will it take to diagnose and treat curable pancreatic neoplasia? *Cancer Res* 2014;74:3381–9.
- [5] Witkiewicz AK, McMillan EA, Balaji U, Baek G, Lin WC, Mansour J, et al. Whole-exome sequencing of pancreatic cancer defines genetic diversity and therapeutic targets. *Nat Commun* 2015;6:6744.
- [6] Garrido-Laguna I, Hidalgo M. Pancreatic cancer: from state-of-the-art treatments to promising novel therapies. *Nat Rev Clin Oncol* 2015;12:319–34.
- [7] Mardis ER. Applying next-generation sequencing to pancreatic cancer treatment. *Nat Rev Gastroenterol Hepatol* 2012;9:477–86.
- [8] Neeße A, Algul H, Tuveson DA, Gress TM. Stromal biology and therapy in pancreatic cancer: a changing paradigm. *Gut* 2015;64:1476–84.
- [9] Waddell N, Pajic M, Patch AM, Chang DK, Kassahn KS, Bailey P, et al. Whole genomes redefine the mutational landscape of pancreatic cancer. *Nature* 2015;518:495–501.
- [10] Moffitt RA, Marayati R, Flate EL, Volmar KE, Loeza SG, Hoadley KA, et al. Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. *Nat Genet* 2015;47:1168–78.
- [11] Notta F, Chan-Seng-Yue M, Lemire M, Li Y, Wilson GW, Connor AA, et al. A renewed model of pancreatic cancer evolution based on genomic rearrangement patterns. *Nature* 2016;538:378–82.
- [12] Lee JK, Choi YL, Kwon M, Park PJ. Mechanisms and consequences of cancer genome instability: lessons from genome sequencing studies. *Annu Rev Pathol* 2016;11: 283–312.
- [13] Negrini S, Gorgoulis VG, Halazonetis TD. Genomic instability—an evolving hallmark of cancer. *Nat Rev Mol Cell Biol* 2010;11:220–8.
- [14] Fabbri C, Luigiano C, Maimone A, Tarantino I, Baccarini P, Fornelli A, et al. Endoscopic ultrasound-guided fine-needle biopsy of small solid pancreatic lesions using a 22-gauge needle with side fenestration. *Surg Endosc* 2015;29:1586–90.
- [15] Bourmet B, Selves J, Grand D, Danjoux M, Hanoun N, Cordelier P, et al. Endoscopic ultrasound-guided fine-needle aspiration biopsy coupled with a KRAS mutation assay using allelic discrimination improves the diagnosis of pancreatic cancer. *J Clin Gastroenterol* 2015;49:50–6.
- [16] 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.
- [17] Murtaza M, Dawson SJ, Tsui DW, Gale D, Forshew T, Piskorz AM, et al. Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. *Nature* 2013;497:108–12.
- [18] Zill OA, Greene C, Sebanovic D, Siew LM, Leng J, Vu M, et al. Cell-free DNA next-generation sequencing in pancreaticobiliary carcinomas. *Cancer Discov* 2015;5: 1040–8.

- [19] Aravanis AM, Lee M, Klausner RD. Next-generation sequencing of circulating tumor DNA for early cancer detection. *Cell* 2017;168:571–4.
- [20] Couraud S, Vaca-Paniagua F, Villar S, Oliver J, Schuster T, Blanche H, et al. Noninvasive diagnosis of actionable mutations by deep sequencing of circulating free DNA in lung cancer from never-smokers: a proof-of-concept study from BioCAST/IFCT-1002. *Clin Cancer Res* 2014;20:4613–24.
- [21] Haber DA, Velculescu VE. Blood-based analyses of cancer: circulating tumor cells and circulating tumor DNA. *Cancer Discov* 2014;4:650–61.
- [22] Gorganzhad L, Umer M, Islam MN, Nguyen NT, Shiddiky MJA. Circulating tumor DNA and liquid biopsy: opportunities, challenges, and recent advances in detection technologies. *Lab Chip* 2018;18:1174–96.
- [23] Phallen J, Sausen M, Adleff V, Leal A, Hruban C, White J, et al. Direct detection of early-stage cancers using circulating tumor DNA. *Sci Transl Med* 2017;9.
- [24] Burnham P, Kim MS, Agbor-Enoh S, Luikart H, Valentine HA, Khush KK, et al. Single-stranded DNA library preparation uncovers the origin and diversity of ultrashort cell-free DNA in plasma. *Sci Rep* 2016;6:27859.
- [25] Underhill HR, Kitzman JO, Hellwig S, Welker NC, Daza R, Baker DN, et al. Fragment length of circulating tumor DNA. *PLoS Genet* 2016;12:e1006162.
- [26] Meyer M, Kircher M, Gansauge MT, Li H, Racimo F, Mallick S, et al. A high-coverage genome sequence from an archaic Denisovan individual. *Science* 2012;338:222–6.
- [27] Gansauge MT, Gerber T, Glocke I, Korlevic P, Lippik L, Nagel S, et al. Single-stranded DNA library preparation from highly degraded DNA using T4 DNA ligase. *Nucleic Acids Res* 2017;45:e79.
- [28] Glocke I, Meyer M. Extending the spectrum of DNA sequences retrieved from ancient bones and teeth. *Genome Res* 2017;27:1230–7.
- [29] 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.
- [30] Vong JSL, Tsang JCH, Jiang P, Lee WS, Leung TY, Chan KCA, et al. Single-stranded DNA library preparation preferentially enriches short maternal DNA in maternal plasma. *Clin Chem* 2017;63:1031–7.
- [31] Zhou G, Liu X, Wang X, Jin D, Chen Y, Li G, et al. Combination of preoperative CEA and CA19-9 improves prediction outcomes in patients with resectable pancreatic adenocarcinoma: results from a large follow-up cohort. *Oncotargets Ther* 2017;10:1199–206.
- [32] Liu X, Yang X, Zhou G, Chen Y, Li C, Wang X. Gemcitabine-based regional intra-arterial infusion chemotherapy in patients with advanced pancreatic adenocarcinoma. *Medicine (Baltimore)* 2016;95:e3098.
- [33] Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* 2012;2:401–4.
- [34] Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* 2013;6:pl1.
- [35] Mayakonda A, Koefler HP. Maftools: efficient analysis, visualization and summarization of MAF files from large-scale cohort based cancer studies. *bioRxiv* 2016. <https://doi.org/10.1101/052662>.
- [36] International Cancer Genome Consortium, Hudson TJ, Anderson W, Artez A, Barker AD, Bell C, et al. International network of cancer genome projects. *Nature* 2010;464:993–8.
- [37] Bailey P, Chang DK, Nones K, Johns AL, Patch AM, Gingras MC, et al. Genomic analyses identify molecular subtypes of pancreatic cancer. *Nature* 2016;531:47–52.
- [38] Fu GK, Hu J, Wang PH, Fodor SP. Counting individual DNA molecules by the stochastic attachment of diverse labels. *Proc Natl Acad Sci U S A* 2011;108:9026–31.
- [39] Fu GK, Xu W, Wilhelmy J, Mindrinos MN, Davis RW, Xiao W, et al. Molecular indexing enables quantitative targeted RNA sequencing and reveals poor efficiencies in standard library preparations. *Proc Natl Acad Sci U S A* 2014;111:1891–6.
- [40] Clarke L, Zheng-Bradley X, Smith R, Kulesha E, Xiao C, Toneva I, et al. The 1000 Genomes Project: data management and community access. *Nat Methods* 2012;9:459–62.
- [41] Forbes SA, Bindal N, Bamford S, Cole C, Kok CY, Beare D, et al. COSMIC: mining complete cancer genomes in the catalogue of somatic mutations in cancer. *Nucleic Acids Res* 2011;39:D945–50.
- [42] Kim MK, Woo SM, Park B, Yoon KA, Kim YH, Joo J, et al. Prognostic implications of multiplex detection of KRAS mutations in Cell-Free DNA from patients with pancreatic ductal adenocarcinoma. *Clin Chem* 2018;64:726–34.
- [43] Allenson K, Castillo J, San Lucas FA, Scelo G, Kim DU, Bernard V, et al. High prevalence of mutant KRAS in circulating exosome-derived DNA from early-stage pancreatic cancer patients. *Ann Oncol* 2017;28:741–7.
- [44] Cheng H, Liu C, Jiang J, Luo G, Lu Y, Jin K, et al. Analysis of ctDNA to predict prognosis and monitor treatment responses in metastatic pancreatic cancer patients. *Int J Cancer* 2017;140:2344–50.
- [45] Sefrioui D, Blanchard F, Toure E, Basile P, Beaussire L, Dolfus C, et al. Diagnostic value of CA19.9, circulating tumour DNA and circulating tumour cells in patients with solid pancreatic tumours. *Br J Cancer* 2017;117:1017–25.
- [46] Cohen JD, Javed AA, Thoburn C, Wong F, Tie J, Gibbs P, et al. Combined circulating tumor DNA and protein biomarker-based liquid biopsy for the earlier detection of pancreatic cancers. *Proc Natl Acad Sci U S A* 2017;114:10202–7.
- [47] Berger AW, Schwerdel D, Costa IG, Hackert T, Strobel O, Lam S, et al. Detection of hot-spot mutations in circulating cell-free DNA from patients with intraductal papillary mucinous neoplasms of the pancreas. *Gastroenterology* 2016;151:267–70.
- [48] Hadano N, Murakami Y, Uemura K, Hashimoto Y, Kondo N, Nakagawa N, et al. Prognostic value of circulating tumour DNA in patients undergoing curative resection for pancreatic cancer. *Br J Cancer* 2016;115:59–65.
- [49] Brychta N, Krahn T, von Ahnen O. Detection of KRAS mutations in circulating tumor DNA by digital PCR in early stages of pancreatic cancer. *Clin Chem* 2016;62:1482–91.
- [50] Kinugasa H, Nouse K, Miyahara K, Morimoto Y, Dohi C, Tsutsumi K, et al. Detection of K-ras gene mutation by liquid biopsy in patients with pancreatic cancer. *Cancer* 2015;121:2271–80.
- [51] Sausen M, Phallen J, Adleff V, Jones S, Leary RJ, Barrett MT, et al. Clinical implications of genomic alterations in the tumour and circulation of pancreatic cancer patients. *Nat Commun* 2015;6:7686.
- [52] Le Calvez-Kelm F, Foll M, Wozniak MB, Delhomme TM, Durand G, Chopard P, et al. KRAS mutations in blood circulating cell-free DNA: a pancreatic cancer case-control. *Oncotarget* 2016;7:78827–40.
- [53] Pietrasz D, Pecuchet N, Garlan F, Didelot A, Dubreuil O, Doat S, et al. Plasma circulating tumor DNA in pancreatic cancer patients is a prognostic marker. *Clin Cancer Res* 2017;23:116–23.
- [54] Takai E, Totoki Y, Nakamura H, Morizane C, Nara S, Hama N, et al. Clinical utility of circulating tumor DNA for molecular assessment in pancreatic cancer. *Sci Rep* 2015;5:18425.
- [55] Landrum MJ, Lee JM, Benson M, Brown G, Chao C, Chitipiralla S, et al. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res* 2016;44:D862–8.
- [56] Hurwitz HI, Uppal N, Wagner SA, Bendell JC, Beck JT, Wade 3rd SM, et al. Randomized, double-blind, phase II Study of ruxitinib or placebo in combination with capecitabine in patients with metastatic pancreatic cancer for whom therapy with gemcitabine has failed. *J Clin Oncol* 2015;33:4039–47.
- [57] Jiang P, Chan CW, Chan KC, Cheng SH, Wong J, Wong VW, et al. Lengthening and shortening of plasma DNA in hepatocellular carcinoma patients. *Proc Natl Acad Sci U S A* 2015;112:E1317–25.
- [58] Mouliere F, Chandrananda D, Piskorz AM, Moore EK, Morris J, Ahlborn LB, et al. Enhanced detection of circulating tumor DNA by fragment size analysis. *Sci Transl Med* 2018;10.
- [59] Mouliere F, Mair R, Chandrananda D, Marass F, Smith CG, Su J, et al. Detection of cell-free DNA fragmentation and copy number alterations in cerebrospinal fluid from glioma patients. *EMBO Mol Med* 2018 Dec;10(12).
- [60] Tan MC, Basturk O, Brannon AR, Bhanot U, Scott SN, Bouvier N, et al. GNAS and KRAS mutations define separate progression pathways in intraductal papillary mucinous neoplasm-associated carcinoma. *J Am Coll Surg* 2015;220:845–54 [e841].
- [61] Wu J, Matthaei H, Maitra A, Dal Molin M, Wood LD, Eshleman JR, et al. Recurrent GNAS mutations define an unexpected pathway for pancreatic cyst development. *Sci Transl Med* 2011;3:92ra66.
- [62] Cancer Genome Atlas Research Network. Electronic address aadhe, Cancer Genome Atlas Research N. Integrated genomic characterization of pancreatic ductal adenocarcinoma. *Cancer Cell* 2017;32:185–203 e113.