

OPEN

Comparison of Target Enrichment Platforms for Circulating Tumor DNA Detection

So Ngo Lam^{1,3}, Ying Chun Zhou^{2,3}, Yee Man Chan¹, Ching Man Foo¹, Po Yi Lee¹, Wing Yeung Mok¹, Wing Sum Wong¹, Yan Yee Fung¹, Kit Yee Wong¹, Jun Yuan Huang² & Chun Kin Chow^{1*}

Cancer-related mortality of solid tumors remains the major cause of death worldwide. Circulating tumor DNA (ctDNA) released from cancer cells harbors specific somatic mutations. Sequencing ctDNA opens opportunities to non-invasive population screening and lays foundations for personalized therapy. In this study, two commercially available platforms, Roche's Avenio ctDNA Expanded panel and QIAgen's QIAseq Human Comprehensive Cancer panel were compared for (1) panel coverage of clinically relevant variants; (2) target enrichment specificity and sequencing performance; (3) the sensitivity; (4) concordance and (5) sequencing coverage using the same human blood sample with ultra-deep next-generation sequencing. Our finding suggests that Avenio detected somatic mutations in common cancers in over 70% of patients while QIAseq covered nearly 90% with a higher average number of variants per patient (Avenio: 3; QIAseq: 8 variants per patient). Both panels demonstrated similar on-target rate and percentage of reads mapped. However, Avenio had more uniform sequencing coverage across regions with different GC content. Avenio had a higher sensitivity and concordance compared with QIAseq at the same sequencing depth. This study identifies a unique niche for the application of each of the panel and allows the scientific community to make an informed decision on the technologies to meet research or application needs.

Cancer is the second leading cause of death worldwide. Cancer-related mortality of most solid tumors remains steady despite intense research on carcinogenesis and significant advancement in effective treatments. It is estimated that 1 in every 6 deaths is related to cancer and late-stage presentation is still very common¹. For cervical and colorectal cancers, population-wide screenings have contributed to the decreasing mortality². However, there is still an urgent need for accurate, effective and non-invasive paradigms to reduce cancer-related mortality for other common and deadly cancer types via early diagnosis, personalized therapeutic strategies and disease monitoring.

Circulating cell-free DNA (cfDNA) is released into the peripheral blood due to apoptosis, necrosis or active release^{3–5}. Higher cfDNA level is found in the plasma of cancer patients comparing to healthy controls^{6,7}. Circulating tumor DNA (ctDNA) from cancer cells harboring a unique set of genetic and epigenetic alterations creates a biological signature. This mutation signature is not only specific to cancer in general when compared to normal tissues but is also highly individual specific^{3,8,9}. The level of ctDNA, somatic mutations burden, the mutation signature and epigenetic marks are highly correlated with cancer pathophysiology and treatment response^{3,10}. Unlike traditional biopsy, profiling of somatic mutations via ctDNA is not only minimally invasive but also provides a less localized and biased sampling. Profiling of ctDNA, released from various cells in the tumor, allows a snapshot of somatic mutation burden to provide a more comprehensive overview of the highly heterogeneous tumor¹¹. Moreover, for tumors that are not easily accessible with a needle biopsy, ctDNA profiling provides a convenient way to make an informed treatment decision for an optimal outcome. Furthermore, due to the minimally invasive nature of ctDNA profiling, longitudinal examination of the ever-changing mutation signatures which reflect the properties of the evolving tumor, can be used not merely to monitor treatment response and relapse, but also helps clinicians to update therapeutic strategies accordingly^{10,12}.

¹Department of Research and Development, Medtimes Molecular Laboratory Ltd., Hong Kong, Hong Kong SAR.

²First Affiliated Hospital of Guangzhou University of Chinese Medicine, Guangzhou, Guangdong, China. ³These authors contributed equally: So Ngo Lam and Ying Chun Zhou. *email: isaac@medtimes.com.hk

	Avenio ctDNA Expanded panel	QIAseq Human Comprehensive Cancer panel
Size (kbp)*	162.10	836.67
Enrichment technology	Hybridization	Amplicon
Extraction method	Affinity column	NA
Fragmentation method	NA	Enzymatic
Automation	++	++
Throughput	++	+++
Flexibility	Custom unavailable; have 3 panel choices	Custom available
Used of UMI	Yes (4bp)	Yes (12bp)
Unique dual index compatibility	No	Yes
Cost	\$\$\$	\$\$

Table 1. Overview of characteristics of Avenio ctDNA Expanded panel and QIAseq Human Comprehensive Cancer panel. *Calculated from the .bed file provided by the manufacturers. [§]Representation of the relative cost of the kit per reaction.

Although the examination of somatic variations in ctDNA is challenging due to the high dilution by cfDNA of normal cell origin in the inherent background, various studies have harnessed the detection power of next-generation sequencing (NGS) via ultra-deep sequencing to uncover these somatic mutations for the development of early screening paradigm, monitoring treatment response and censoring residual diseases^{13–18}. Though the sequencing cost has plummeted significantly in the last decade, the need to sequence to high depth to identify rare and highly diluted mutations from the background makes the cost of genome-wide sequencing discouraging for clinical application^{19–21}. It is more cost-effective and time-efficient to capture and sequence only the genomic regions which are mutation hotspots or of high clinical importance. To address this need, several companies have developed cancer panels to enrich genomic regions of interest for a specific cancer type or in a pan-cancer manner.

Commercial platforms including Roche's Avenio ctDNA Expanded panel and QIAgen's QIAseq Human Comprehensive Cancer panel are currently available for minimal invasive ctDNA detection. These platforms fall into two categories based on their enrichment technologies, probe-based solution hybridization and amplicon-based enrichment. Both platforms are claimed to be applicable for the identification of ctDNA which often have very low allele fraction (AF). While making an educated choice of which commercial platforms and technologies to choose for detection of ctDNA to suit specific application or research needs, several factors including the price, size and design of the region of interest, sensitivity, accuracy and sequencing uniformity are worth careful considerations. However, most of these questions are still left unanswered. In this study, we compared the performance of these two commercially available cancer panels for efficacy in ctDNA detection. We evaluated several key parameters, including (1) panel coverage of clinically relevant variants; (2) target enrichment specificity and sequencing performance; (3) the sensitivity; (4) concordance and (5) sequencing coverage.

Results

Overview of platform difference. Although both panels aim for the detection of somatic mutations via deep sequencing, the development, design and technologies of the panels and kits are substantially different. Avenio is a commercialized kit developed from CAPP-seq which enriches recurrent mutations in driver genes via hybridization to DNA probes^{16,22,23}. On the other hand, QIAseq enriches target regions through PCR amplification with a proprietary single primer extension reaction (Table 1). The two panels were 162 kbp (Avenio) and 837 kbp (QIAseq) in size with around 136.7 kbp in common. 15.7% and 83.7% of the panels are unique to Avenio and QIAseq respectively (Fig. 1a).

The Avenio ctDNA Expanded kit but not the QIAseq Human Comprehensive Cancer panel includes a cfDNA extraction kit that extracts cfDNA with affinity columns (Table 1). For the fairness of comparison and to ensure differences in the performance of the panels are not due to the use of different cfDNA extraction protocols. We used the cfDNA extraction kit (TIANGEN Biotech Co. Ltd., Beijing, China) to extract all the samples in this study. Though the Avenio cfDNA extraction kit was not used, we compared its extraction yield and DNA size profile with TIANGEN's cfDNA extraction kit. Both extraction kits extracted cfDNA with comparable quantity (Avenio: 36.89 ± 33.44 ng; TIANGEN: 45.10 ± 28.85 ng; $p = 0.072$; data not shown). Furthermore, Avenio extracted DNA with a size of 173.8 ± 3.3 bp while TIANGEN extracted DNA of 175 ± 2.2 bp ($p = 0.412$; data not shown). The size distribution of DNA extracted by the kits is characteristic to cfDNA²⁴.

By far, fresh tumor biopsies and formalin-fixed paraffin-embedded (FFPE) sections are still the most common sample types for molecular mutation profiling. Kits that can process multiple sample types would provide a fairer evaluation of concordance and accuracy of ctDNA detection. The Avenio ctDNA panel is specialized for ctDNA application while the QIAseq panel is also applicable to freshly extracted DNA from tumor samples and FFPE sections. Though Roche also provides Avenio panel for FFPE samples, compatibility of the QIAseq Human Comprehensive Cancer panel with fresh biopsy and FFPE allows the study of concordance of variants detected in ctDNA with tumor biopsy or FFPE using the same kit.

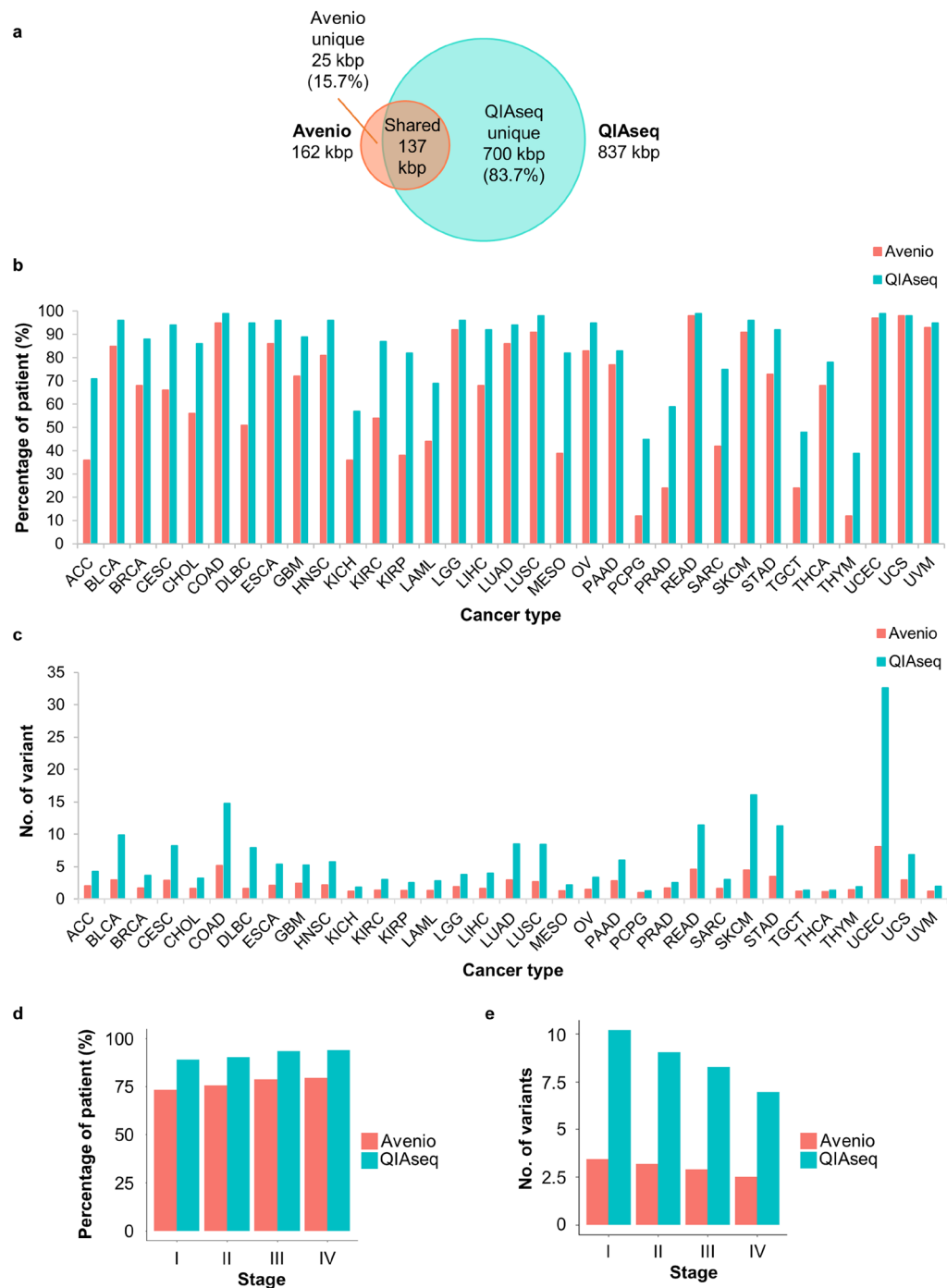


Figure 1. Panel design coverage of somatic mutation in pan-cancer. (a) Venn diagram demonstrating the panel size and overlapping targeted region. Panel design coverage on (b) the percentage of patient with at least 1 somatic mutation that is detected by Avenio and QIAseq for all cohorts from TCGA at all stages, (c) the number of variant per patient, (d) the percentage of patient that is covered by the panels at different stage and (e) the number of variants per patient in each stage. TCGA study abbreviations, **ACC**: adrenocortical carcinoma; **BLCA**: bladder urothelial carcinoma; **BRCA**: breast invasive carcinoma; **CESC**: cervical squamous cell carcinoma and endocervical adenocarcinoma; **CHOL**: cholangiocarcinoma; **COAD**: colon adenocarcinoma; **DLBC**: lymphoid neoplasm diffuse large B-cell lymphoma; **ESCA**: esophageal carcinoma; **GBM**: glioblastoma multiforme; **HNSC**: head and neck squamous cell carcinoma; **KICH**: kidney chromophobe; **KIRC**: kidney renal clear cell carcinoma; **KIRP**: kidney renal papillary cell carcinoma; **LAML**: acute myeloid leukemia; **LGG**: brain lower grade glioma; **LIHC**: liver hepatocellular carcinoma; **LUAD**: lung adenocarcinoma; **LUSC**: lung squamous cell carcinoma; **MESO**: mesothelioma; **OV**: ovarian serous cystadenocarcinoma; **PAAD**: pancreatic adenocarcinoma; **PCPG**: pheochromocytoma and paraganglioma; **PRAD**: prostate adenocarcinoma; **READ**: rectal adenocarcinoma; **SARC**: sarcoma; **SKCM**: skin cutaneous melanoma; **STAD**: stomach adenocarcinoma; **TGCT**: testicular germ cell tumors; **THCA**: thyroid carcinoma; **THYM**: thymoma; **UCEC**: uterine corpus endometrial carcinoma; **UCS**: uterine carcinosarcoma; **UVM**: uveal melanoma.

For clinical applications, whether the kit is compatible with robotic automation systems and its maximum number of multiplexing would highly impact the overall cost and economic incentive of marketing the mutation detection test. Both kits use magnetic beads for cleaning up and size selection and are compatible with most library preparation automation systems for NGS. However, since QIAseq consists of 96 indexing combinations for dual indexing, the throughput of QIAseq is much higher than that of Avenio (Avenio: 16 single-index adapters; QIAseq: 96 dual-index adapters). Due to the proprietary single-index adapter design of Avenio, the panel is incompatible with the unique dual index which has been found to mitigate index-hopping in pattern flow cells of Illumina. Furthermore, both panels made use of unique molecular index (UMI) to tag individual DNA fragments to provide a better quantification of copy number and *in silico* error suppression.

Panel coverage of clinically relevant variants. To test the degree of the panel coverage of recurrent mutations in cancers, we studied the number of mutations each panel can target enrich for detection using variants data of 33 cohorts profiled by The Cancer Genome Atlas (TCGA). We studied the number of patients with at least 1 somatic mutation which is targeted by Avenio and QIAseq. On average, Avenio covered 71% of patients with an average of 3 mutations per patient while 88% of patients have an average of 8 mutations covered by QIAseq (Fig. 1b, c). Since the number of variants detected depends greatly on panel size, we studied the variant detection efficiency by normalizing the number of variants per patient detected by the panel size of each platform. After normalization, Avenio detected 1.80 variants per 100 kbp of the panel. QIAseq was heavily penalized by its large panel size, the number variants targeted per 100 kbp of the panel was reduced to 0.95.

Despite its much smaller panel size, Avenio has comparable patient coverage with QIAseq for lung (LUAD: Avenio 86% vs QIAseq 94%; LUSC: Avenio 91% vs QIAseq 98%) and colorectal cancer (COAD: Avenio 95% vs QIAseq 99%; READ: Avenio 98% vs QIAseq 99%) (Fig. 1b), which could be due to the use of an iterative algorithm for the design of the selector²². On the other hand, QIAseq covered 40% more patients with rare cancers such as kidney cancer (KIRP), lymphoma (DLBC) and mesothelioma (MESO). Overall, QIAseq targeted more variants per patient (Fig. 1c).

Both panels were able to cover mutations frequently occurred in the early stage of cancers (Fig. 1d, e). Avenio targeted 73% and 76% of stage I and II patients respectively with at least 1 mutation. Consistently, QIAseq showed a higher coverage of these early-stage cancer patients. It covered around 90% of patients at stage I and II.

Although the sequencing cost has decreased substantially over the last decade, due to the need to sequence to ultra-high depth to discover mutations in ctDNA in plasma, a larger panel size would greatly increase the cost and prohibit its application in cancer detection and surveillance. Furthermore, a larger panel would decrease the sensitivity and accuracy via the identification of artifactual mutations which further reduces signal-to-noise ratio¹⁸. Since QIAseq is more than 5 times larger than Avenio, we investigated the unique region targeted by QIAseq to evaluate its add-on value on the detection of clinically significant variants. On average, this unique region alone would allow the QIAseq panel to target an average of 6.6 variants per patient (Fig. 2a). However, the number of variants targeted per 100 kbp decreased from 0.95 (whole panel region) to 0.79 (unique panel region), suggesting the density of recurring somatic mutations in these genomic regions within the unique panel region is relatively lower (Fig. 2a). Furthermore, the clinical significance of these variants targeted in the QIAseq unique region was analyzed and annotated with the ClinVar database. The ClinVar database annotates different clinical significance values of the variant, including 1) according to the American College of Medical Genetics and Genomics or the Association for Molecular Pathology guidelines; 2) drug response and 3) risk factor. Although over 70% of these variants in the unique region of QIAseq were having a moderate to high impact on the protein function, none has been proved to be correlated with drug response. For variants with moderate or high clinical significance, their impact on survival was further investigated, 41 out of the 204 genes have been shown to present significant impacts on patient survival in cohort having at least one moderate or high clinical variants compared to the wild-type cohort (Fig. 2b).

Target enrichment specificity and sequencing performance. To assess the extent of target enrichment, the on-target rate, which was defined as the percentage of reads mapped to the target region, was compared. Both panels had comparable on-target rate (Avenio: $76.8 \pm 0.9\%$; QIAseq: $77.4 \pm 0.4\%$) (Fig. 3a). Avenio demonstrated a higher percentage of reads mapped, 87% comparing to 75% of QIAseq (Fig. 3b). GC content of the fragment sequence has been shown to have a systemic effect on target enrichment and sequencing uniformity. We next examined the uniformity of sequencing across genomic regions of different GC content. For a fair comparison, only the shared region of both panels was studied. The number of reads of the fragments within the shared region was further normalized by the average read depth. The heatmaps showed that Avenio had more fragments having a normalized read depth close to 1, suggesting that Avenio had more uniform coverage across regions with different GC content (Fig. 3c, d).

The sensitivity of ctDNA detection. To assess the sensitivity of detecting frequently occurring somatic mutations in each platform, we sequenced libraries generated by spiking 0.1, 1, 10 or 50% of cfDNA reference which consists of 32 well-characterized mutations into the cfDNA of a healthy volunteer of Han Chinese ancestry. These cfDNA references are commonly enriched by both platforms. cfDNA from the healthy volunteer was pre-sequenced with both platforms to ensure the sample is free of the spike-in variants in the background. The library was constructed and enriched in accordance with the manufacturers' instructions. 150-bp paired-end reads were generated using $\frac{1}{4}$ lane of an Illumina NextSeq High Output kit. Read counts were normalized between the two platforms by randomly selecting 50–350 million reads with an increment of 50 million reads. For both platforms, the sensitivity of variant detection increased with the number of read pairs (Fig. 4a–d). The sensitivity of variant detection depends greatly on the AF and sequencing depth. With 15 ng of cfDNA input consisting of variants with expected AF of 0.075–2.96% (10% cfDNA reference spiked in), a clinically relevant AF of most

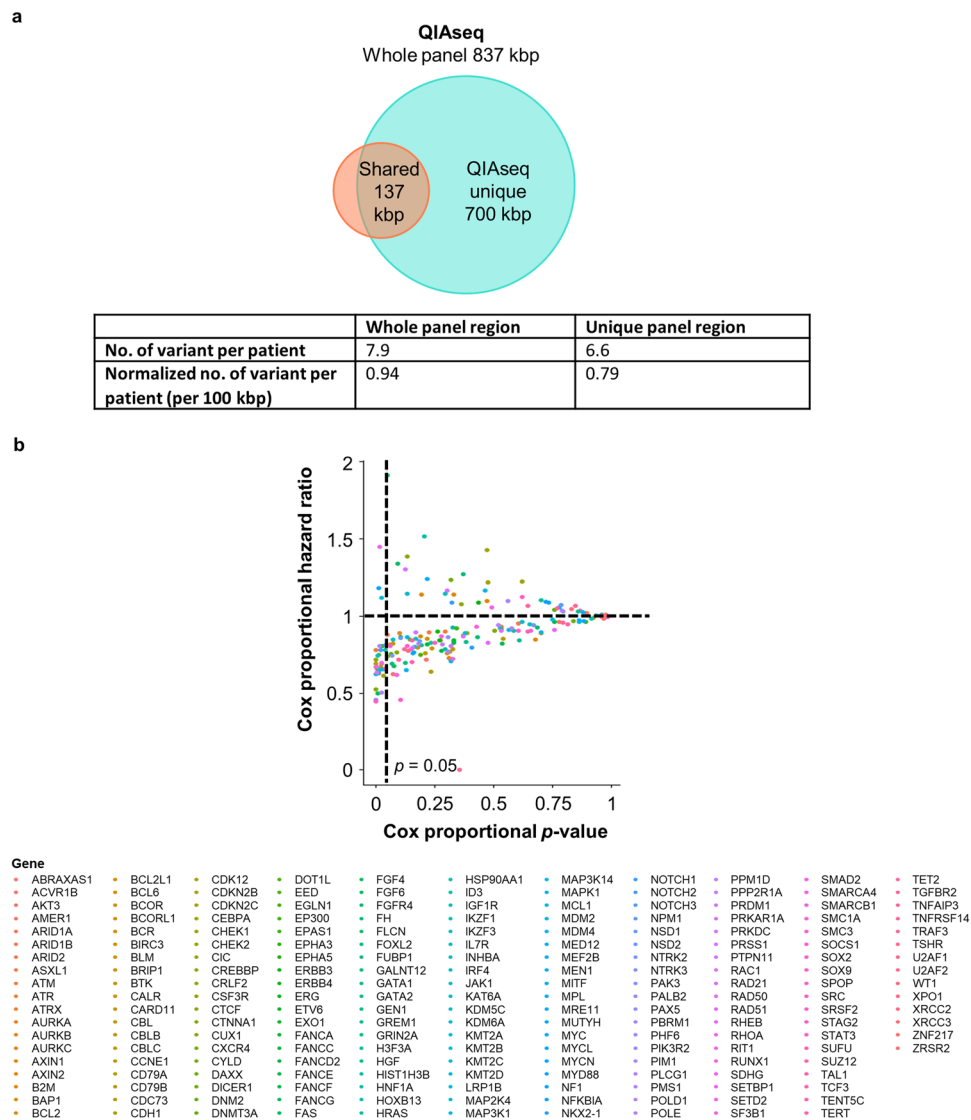


Figure 2. Evaluation of the add-on value of the QIaseq unique region in the detection of clinically significant variants. **(a)** No. of variants per patient detected by QIaseq whole panel region and unique panel region. **(b)** The clinical significance of genes that are detected by the unique panel region of QIaseq for all cohorts from TCGA at all stages.

variants detected in ctDNA^{25,26}, sensitivity of Avenio started to plateau at 85% with 35 million of reads while the sensitivity of QIaseq levelled off at 20% with 65 million of reads (Fig. 4c). With 50% cfDNA reference spike-in, Avenio demonstrated a higher sensitivity of ctDNA detection, 100% at all read depth, compared to around 60% with at least 60 million of reads in QIaseq (Fig. 4d).

Concordance and accuracy. To investigate the accuracy of Avenio and QIaseq in estimating AF of somatic mutations, we computed the concordance correlation coefficient (CCC) between AFs of spike-in references which are independently quantified with ddPCR (expected AF) and the Avenio or QIaseq detected AFs (observed AF) (Fig. 5a–d). ddPCR estimates the absolute copy number of the spike-in references. Thus, it is used as a “gold standard” to evaluate the concordance and accuracy of Avenio and QIaseq. CCC was estimated using U-statistics without making assumption on the AF distribution normality²⁷. Concordance between expected and observed AF with 350 million normalized read pairs was 0.923 (95% confidence interval (CI): 0.891–0.946) for Avenio and 0.864 (95% CI: 0.819–0.898) for QIaseq for AF ≤ 20% (Fig. 5a). For AF ≤ 5%, Avenio showed higher accuracy of AF estimation (Avenio: 0.750, 95% CI: 0.678–0.808; QIaseq: 0.538, 95% CI: 0.447–0.619) (Fig. 5b). Furthermore, at a clinically relevant AF, AF ≤ 1%, Avenio demonstrated substantially higher concordance (Avenio: 0.577, 95% CI: 0.541–0.610; QIaseq: 0.070, 95% CI: -0.108–0.244) (Fig. 5c).

Sequencing coverage. To assess the coverage of the targeted bases for each platform, the percentage of targeted bases in each panel that has been sequenced for at least 250, 500, 1000, 1500, 2000 and 2500X was quantified after removal of PCR duplication in the 10% cfDNA reference spike-in sample. Over 99% of targeted bases were

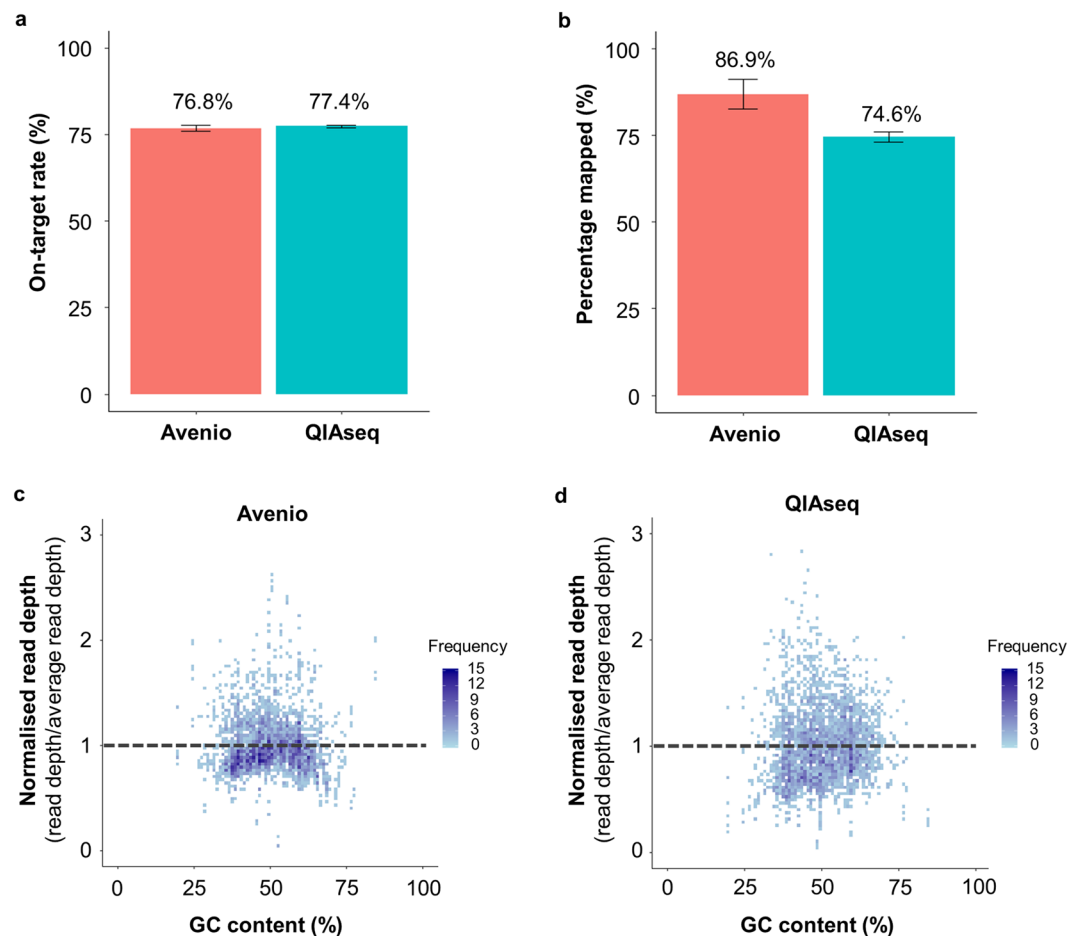


Figure 3. Sequencing performance of the panels. (a) On-target rate, (b) percentage mapped of the reads. Normalised read depth across reads of different GC content within the shared region of (c) Avenio and (d) QIAseq.

covered $\geq 250X$ in Avenio while around 85% of bases were covered $\geq 250X$ in QIAseq with 80 million pair reads per sample (Fig. 6a). At all read counts and depth cut-offs, Avenio gave a higher unique coverage of targeted bases which may be a benefit from the smaller size of the panel (Fig. 6a–f).

The panel size of QIAseq is about five times (5.16X) larger than that of Avenio. To evaluate the performance of Avenio as if it is of the same size as QIAseq. We examined the relationship of sequencing depth and unique coverage of the targeted bases (250X, 500X, 750X, 1000X) after normalizing the Avenio sequencing reads by panel size ratio (5.16). After normalization, Avenio presented a higher unique coverage at low read depth threshold, $\geq 250X$ (Supplementary Fig. 1a). However, the unique coverage of Avenio decreased significantly especially at a higher read depth threshold (Supplementary Fig. 1a–d).

Discussion

Users of commercialized cancer panels aim to detect clinically-relevant mutations for cancer profiling, response monitoring and therapy guiding. The work presented here is significant in helping the research, clinical and scientific community to make an informed decision about the platform best suit their applications. By comparing 2 commercialized cancer panels that use different target enrichment technologies, this study demonstrated the performance of both commercial kits for identifying highly diluted ctDNA in plasma. Using TCGA datasets consisting of 33 cohorts, both platforms were shown to enrich recurrent somatic mutations in common cancers. We have observed that the Avenio ctDNA Expanded panel was able to adequately cover most hotspots. Due to the use of an iterative algorithm to maximize the number of missense mutations per patient while minimizing the panel size during panel development. Thanks to its smaller panel size, it allows sequencing to a higher depth at a reasonable and affordable sequencing cost for the identification of diluted ctDNA in the plasma. On the other hand, QIAseq demonstrated excellent ability in targeting more mutations across all cancer types and has a higher patient coverage. The QIAseq panel showed higher patient coverage in some rare cancers and was designed to cover nearly 90% of patients at the early stages. Moreover, we have also demonstrated high sensitivity and concordance in the detection of ctDNA using both platforms. Therefore, the Avenio platform may be a better choice for applications to detect mutations in more common cancer types via sequencing at ultra-high depth while QIAseq suits applications aiming to have a broader spectrum of cancer types.

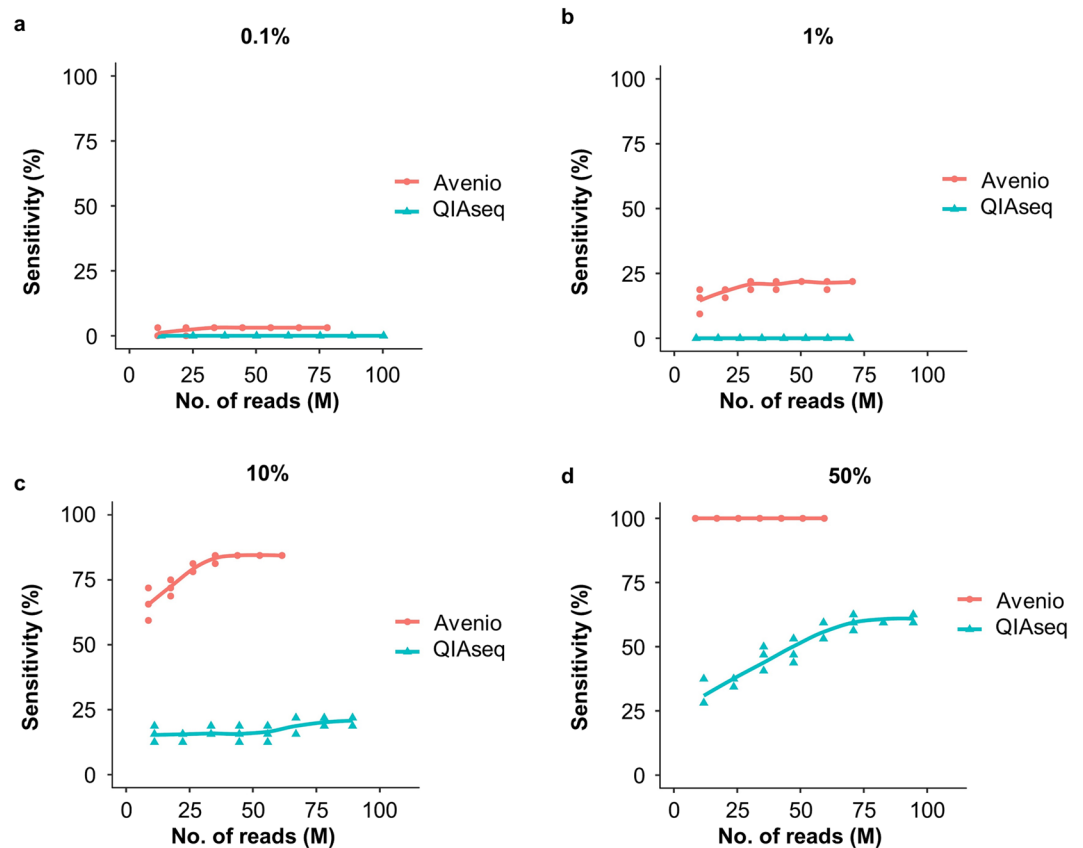


Figure 4. The sensitivity of detecting spike-in variants at different read depth. Libraries prepared with 15 ng of cfDNA spiked in with (a) 0.1%, (b) 1%, (c) 10% and (d) 50% of cfDNA reference were sequenced with $\frac{1}{4}$ lanes of NextSeq High Output kit. Read pairs were normalized by random sampling at 50–350 M read pairs with 50 M increment.

For the relative ease of application, Avenio requires 3 days for library construction while the library preparation with QIAseq takes only 1 working day. Although the turnaround time for hybridization-based enrichment workflow is often longer than its amplicon-based counterparts, the actual hands-on time for both panels is comparable. Furthermore, Avenio provides an end-to-end solution from cfDNA extraction to bioinformatic analysis. Therefore, Avenio may be more favorable in small labs or small scale of applications which do not have a bioinformatics team. In contrast, though the bioinformatic pipeline of QIAseq is open-sourced, it would require personnel with bioinformatics background to turn the out of sequencer raw data into annotated variants.

Both library preparation workflow tagged individual DNA fragment with UMI, 4 and 12 bp in length in Avenio and QIAseq respectively. Sequencing errors arisen from PCR amplification mis-incorporation, cluster amplification and sequencing account for a background level of 0.1–1% base mis-identification depending on sequencing platforms²⁸. This error rate is within the same order of magnitude as most clinical variants in the highly heterogeneous cfDNA mixture and therefore hinders the identification of true rare variants. By uniquely tagging individual fragments with a UMI before amplification, variants with AF below 0.1% could also be detected^{29–36}. The benefit of accurate quantification provided by UMI is evident with the high concordance between expected and observed AF of variants of higher AF. Our data showed that at AF less than 1%, the concordance correlation coefficient of Avenio is substantially higher than that of QIAseq. Together with a higher sensitivity of Avenio in detecting variants with low AF, Avenio outperforms QIAseq in the detection of ctDNA in plasma.

For the application on ctDNA profiling for early-stage asymptomatic screening, the tissue-of-origin would be needed to guide follow-ups. The Avenio bioinformatic pipeline annotates variants detected with TCGA data to provide information on the prevalence of certain mutations being detected in a specific cancer type. This may provide insight into the most probable tissue-of-origin based on the somatic mutation signature by a simple voting scheme.

The possibility of inferring tissue-of-origin by secondary analysis of genomic fragment ends, preserved by hybridization-enrichment library preparation but not its amplicon-based counterparts, is also an important factor for determining which enrichment technology and platform to be employed. The size distribution of cfDNA, which has a median of 166 bp corresponding to that of chromosomes (nucleosome + linker histone, ~167 bp)²⁴, has drawn speculations on the association of epigenetic landscape and DNA fragmentation pattern. Nucleosome occupancy protects DNA fragments from apoptotic nuclease digestion^{37–39}. Recent studies have demonstrated that the DNA fragment ends retain information on the nucleosome occupancy and chromatin structure which highly correlates with the tissue-of-origin^{40,41}. Moreover, the DNA fragment end pattern could be used not only

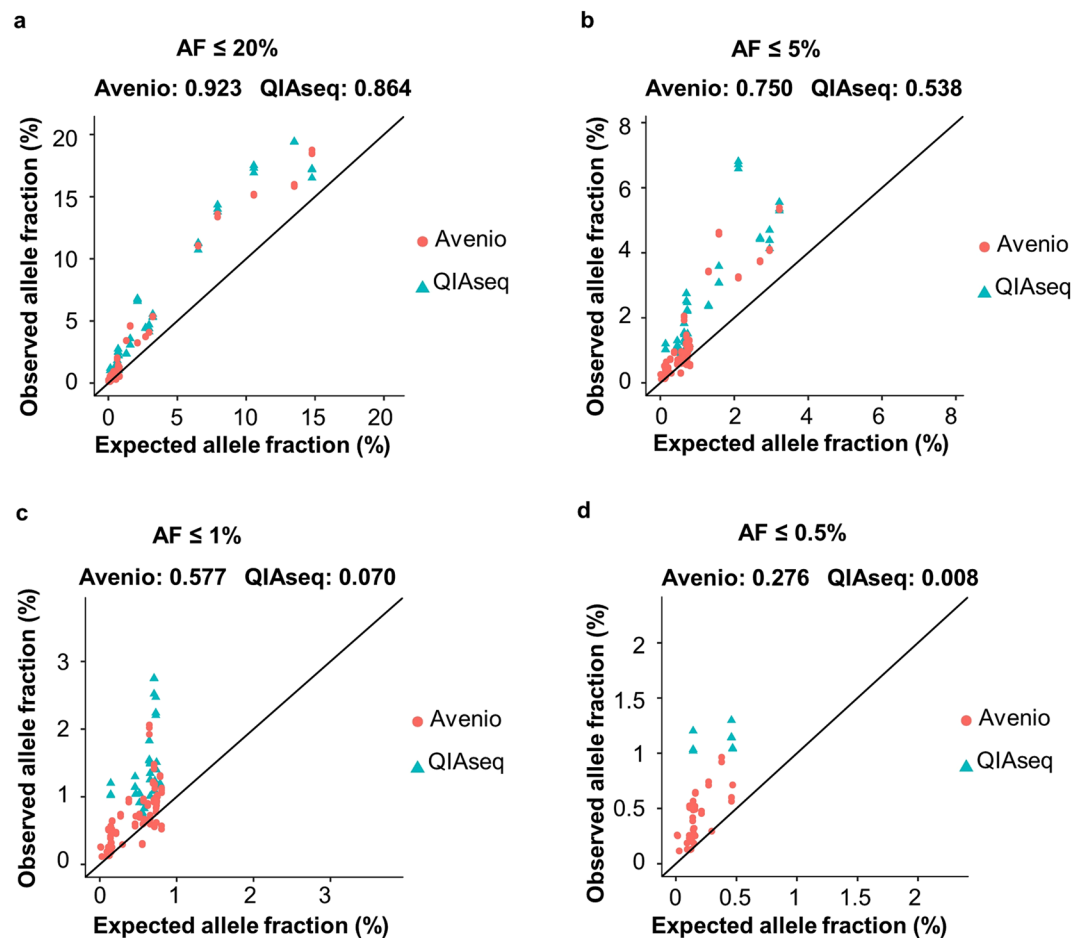


Figure 5. The concordance and linearity of expected AF and observed AF of spike-in variants at expected AF threshold of (a) $\leq 20\%$, (b) $\leq 5\%$, (c) $\leq 1\%$ and (d) $\leq 0.5\%$.

to predict the tissue-of-origin for ctDNA profiling but also help eliminate false-positive caused by hemopoietic clonal mutations which is a major source of false-positive in ctDNA profiling. Library preparation workflow which ligates adapter to both ends of the fragment and enriches via hybridization preserves the genomic ends of the DNA fragments and thus the fragmentation pattern. In the case of amplicon-based target enrichment workflow, the use of gene-specific primers for enrichment causes the loss of information on the genomic coordinate of the fragment endpoints. Though the classification of tissue-of-origin based on fragmentation pattern inferred from fragment endpoints is still a field of on-going research, deep sequencing data generated by Avenio, which target enriches region of interest via hybridization with DNA baits, can be used not only for mutation identification but might also be secondarily processed to infer the possible tissue-of-origin⁴¹.

In summary, after the first identification of ctDNA in the circulation, the pursuit of developing a non-invasive, effective and affordable pan-cancer screening test is relentless. The use of ultra-deep sequencing with carefully designed and optimized pan-cancer panels and library construction workflow would allow direct measurement of the somatic changes in cancers. This study made a systemic comparison of the performance of Avenio ctDNA Expanded panel and QIAseq Human Comprehensive Cancer panel in ctDNA profiling. Both platforms demonstrated high patient coverage, sensitivity and concordance in the detection of clinically relevant variants with minimal cfDNA input. With a smaller panel, Avenio offers an excellent performance in detecting mutations in the hotspots of most common cancers with higher sensitivity and concordance. On the other hand, QIAseq outperforms by enabling detection in some relatively less common cancer types, therefore offering a true pan-cancer screening. These findings identified a unique niche for each of the commercially available panels.

Methods

Subjects. The panels were tested with the same starting DNA material originated from a 30-year-old healthy Han Chinese female. The subject does not demonstrate any pathological, histological or molecular sign of cancer. The subject was neither pregnant nor had received a blood transfusion within a month of commencement of the blood sampling of the study or diagnosed with autoimmune disease. The use of the specimen for the study and all experiments were conducted in accordance with the relevant guidelines and regulations approved by the Medtimes Medical Group Ethics Review Board. Written informed consent for the use of peripheral blood was obtained.

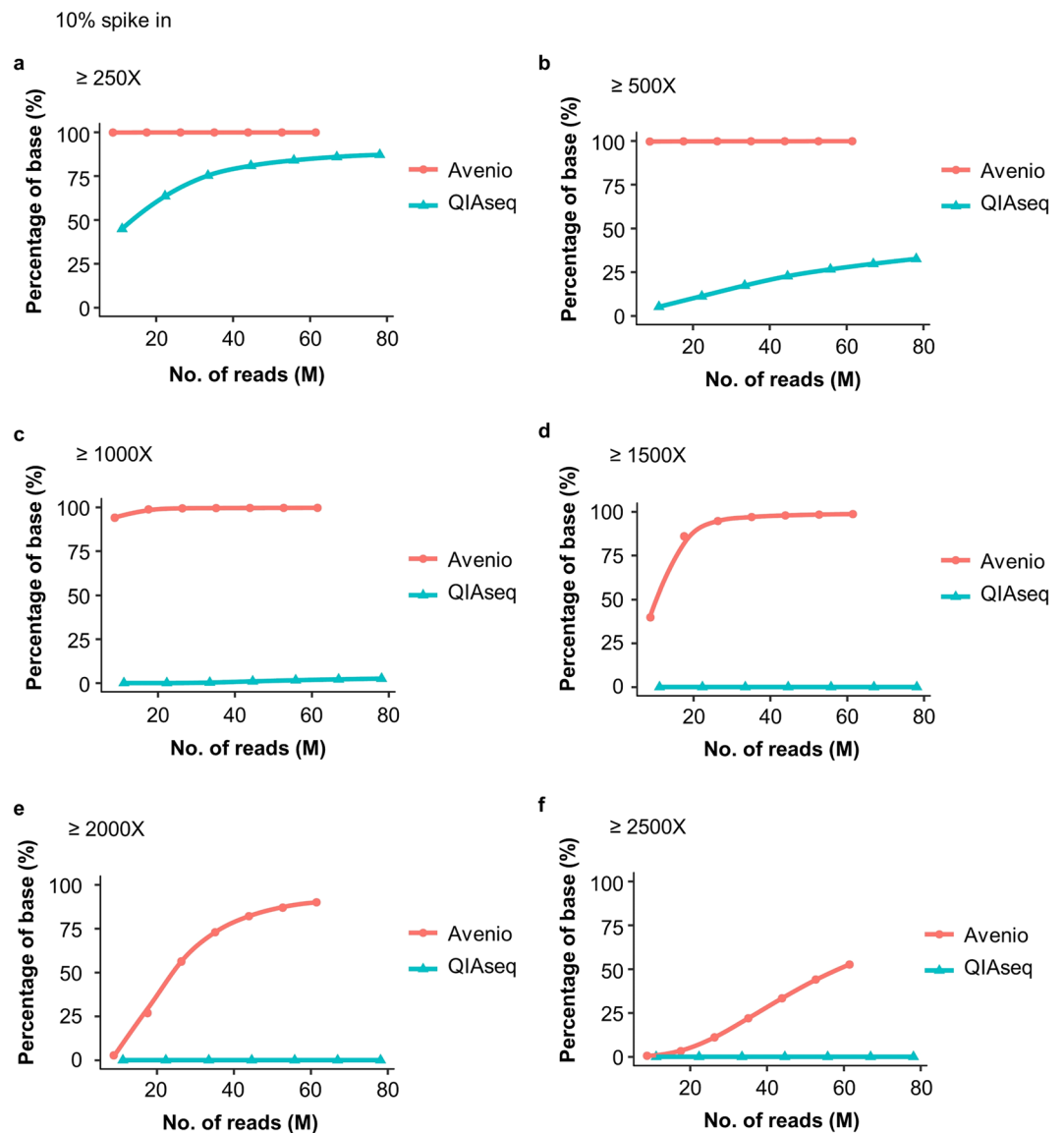


Figure 6. The percent of targeted bases covered at (a) ≥ 250 -fold, (b) ≥ 500 -fold, (c) ≥ 1000 -fold, (d) ≥ 1500 -fold, (e) ≥ 2000 -fold and (f) ≥ 2500 -fold read depth.

Blood sampling. Blood samples were collected using Cell-Free DNA Collection tubes (Roche, Mannheim, Germany) by venipuncture and were stored at 4°C for not more than 24 hr before preprocessing. Blood samples were drawn carefully to avoid any hemolysis. Whole blood samples were centrifuged at 3,000 rpm for 10 min at 4°C followed by second centrifugation at 14,000 rpm for 10 min at 4°C to obtain cell-free plasma. Plasma was collected and stored in aliquots in -30°C until further analysis.

Plasma DNA extraction and quantifications. DNA was extracted from 4 mL of plasma using cfDNA extraction kit (TIANGEN Biotech Co. Ltd., Beijing, China) with Kingfisher Duo Prime (Thermo Fisher Scientific, San Jose, CA) according to manufacturer's instructions except for the following modification: 10 μL of beads was used for each 1 mL plasma extraction. For 4 mL plasma input, DNA was eluted into 130 μL of elution buffer and stored at -30°C . cfDNA was quantified using QuBit dsDNA HS Assay kit with QuBit 3.0 fluorimeter (Thermo Fisher Scientific, San Jose, CA).

Library preparation with Avenio ctDNA Expanded kit. The kit was purchased from Roche (Mannheim, Germany). Sequencing libraries were prepared according to the manufacturer's instructions with the following modifications: plasma cfDNA was extracted using cfDNA extraction kit (TIANGEN Biotech Co. Ltd., Beijing, China) instead of the Avenio ctDNA isolation kit included. Briefly, 15 ng of plasma cfDNA fragments were end-repaired, A-tailed, ligated with UMI-barcoded adaptors and amplified with PCR (12 cycles). The adaptor-ligated libraries were hybridized for 18 hr with biotinylated oligo DNA baits and enriched with streptavidin-conjugated magnetic beads. The target enriched libraries were further amplified for 15 cycles with

PCR and were size-selected for an average fragment size of 350 bp. The library profile was analyzed with the Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA) and quantified using QuBit dsDNA HS Assay kit with QuBit 3.0 fluorimeter (Thermo Fisher Scientific, San Jose, CA) and qPCR with QIAseq Library Quant Assay Kit (QIAGEN, Hilden, Germany).

Library preparation with QIAseq Human Comprehensive Cancer panel. The kit was purchased from QIAGEN (Hilden, Germany). Sequencing libraries were prepared according to the manufacturer's instructions. Briefly, 15 ng of plasma cfDNA fragments were end-repaired, A-tailed, ligated with UMI-barcoded adaptors. The adaptor-ligated libraries were target enriched with PCR using a panel of loci specific primers (6 cycles). The target enriched libraries were further amplified for 21 cycles with PCR and were size selected for an average fragment size of 350 bp. The library profile was analyzed with the Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA) and quantified using QuBit dsDNA HS Assay kit with QuBit 3.0 fluorimeter (Thermo Fisher Scientific, San Jose, CA) and qPCR with QIAseq Library Quant Assay Kit (QIAGEN, Hilden, Germany).

Spike in dilution series. Dilution series were performed to assess the sensitivity and concordance of both panels in detecting and quantitating ctDNA. cfDNA references consist of frequent occurring mutations of AF quantified by ddPCR were spiked into circulating DNA of a healthy individual. A total of 15 ng of DNA containing 0.1, 1, 10 and 50% of the cfDNA reference, comprised of expected AF from 0.00075–15%, was then used for library construction.

Next-generation sequencing. Sequencing was performed on Illumina NextSeq 550. For a single sequencing run, a 4-multiplexed library was created by pooling the libraries, quantified using QuBit dsDNA HS Assay kit with QuBit 3.0 fluorimeter (Thermo Fisher Scientific, San Jose, CA) and qPCR with QIAseq Library Quant Assay Kit (QIAGEN, Hilden, Germany), at an equal molar ratio. The multiplexed library was denatured and sequenced with NextSeq High Output kit 2 × 150 cycles paired.

Read-count normalization. Total read count of all sequencing was normalized to 50–350 M reads with an increment of 50 M reads, by randomly drawing reads from the raw data. At each read depth, a triplicate of random drawings of reads was performed with a recorded random seed.

Panel evaluation with TCGA variant data. Variant data of 10,418 patients of 33 cancer types profiled by TCGA were downloaded from GDC Data Portal. Variant data called by 4 variant callers, Mutect2, VarScan2, MuSE and SomaticSniper were used to assess the average number of variants per patient covered by the cancer panels.

Bioinformatic analysis. Two analysis pipelines for the cancer panel were compared. Raw sequencing data were analyzed with Avenio ctDNA analysis software (version 1.1.0) with default parameter settings for Expanded panel and QIAseq open sources analysis pipeline (version 2, dated May 29, 2018)⁴² with default parameter settings respectively.

Data availability. Sequence data generated and analyzed during the current study are available in the Sequence Read Archive under accession number PRJNA577992 (<https://www.ncbi.nlm.nih.gov/sra/PRJNA577992>).

Received: 23 October 2019; Accepted: 4 February 2020;

Published online: 05 March 2020

References

- Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2019. *CA Cancer J. Clin.* **69**, 7–34, <https://doi.org/10.3322/caac.21551> (2019).
- Arnold, M. *et al.* Global patterns and trends in colorectal cancer incidence and mortality. *Gut* **66**, 683–691, <https://doi.org/10.1136/gutjnl-2015-310912> (2017).
- Schwarzenbach, H., Hoon, D. S. & Pantel, K. Cell-free nucleic acids as biomarkers in cancer patients. *Nat. Rev. Cancer* **11**, 426–437, <https://doi.org/10.1038/nrc3066> (2011).
- Stroun, M. *et al.* The origin and mechanism of circulating DNA. *Ann. N.Y. Acad. Sci.* **906**, 161–168, <https://doi.org/10.1111/j.1749-6632.2000.tb06608.x> (2000).
- Stroun, M., Lyautey, J., Lederrey, C., Olson-Sand, A. & Anker, P. About the possible origin and mechanism of circulating DNA apoptosis and active DNA release. *Clin. Chim. Acta* **313**, 139–142, [https://doi.org/10.1016/s0009-8981\(01\)00665-9](https://doi.org/10.1016/s0009-8981(01)00665-9) (2001).
- Szpechcinski, A. *et al.* Cell-free DNA levels in plasma of patients with non-small-cell lung cancer and inflammatory lung disease. *Br. J. Cancer* **113**, 476–483, <https://doi.org/10.1038/bjc.2015.225> (2015).
- Lapin, M. *et al.* Fragment size and level of cell-free DNA provide prognostic information in patients with advanced pancreatic cancer. *J. Transl. Med.* **16**, 300, <https://doi.org/10.1186/s12967-018-1677-2> (2018).
- Haber, D. A. & Velculescu, V. E. Blood-based analyses of cancer: circulating tumor cells and circulating tumor DNA. *Cancer Discov.* **4**, 650–661, <https://doi.org/10.1158/2159-8290.CD-13-1014> (2014).
- Vogelstein, B. *et al.* Cancer genome landscapes. *Sci.* **339**, 1546–1558, <https://doi.org/10.1126/science.1235122> (2013).
- Reinert, T. *et al.* Analysis of circulating tumour DNA to monitor disease burden following colorectal cancer surgery. *Gut* **65**, 625–634, <https://doi.org/10.1136/gutjnl-2014-308859> (2016).
- Stanta, G. & Bonin, S. Overview on Clinical Relevance of Intra-Tumor Heterogeneity. *Front. Med.* **5**, 85, <https://doi.org/10.3389/fmed.2018.00085> (2018).
- Tchekmedyan, N. *et al.* Longitudinal monitoring of ctDNA EGFR mutation burden from urine correlates with patient response to EGFR TKIs: A case series. *Lung Cancer* **108**, 22–28, <https://doi.org/10.1016/j.lungcan.2017.02.010> (2017).
- Board, R. E. *et al.* Detection of PIK3CA mutations in circulating free DNA in patients with breast cancer. *Breast Cancer Res. Treat.* **120**, 461–467, <https://doi.org/10.1007/s10549-010-0747-9> (2010).

14. Kim, K. *et al.* Circulating cell-free DNA as a promising biomarker in patients with gastric cancer: diagnostic validity and significant reduction of cfDNA after surgical resection. *Ann. Surg. Treat. Res.* **86**, 136–142, <https://doi.org/10.4174/astr.2014.86.3.136> (2014).
15. Chen, K. Z. *et al.* Circulating Tumor DNA Detection in Early-Stage Non-Small Cell Lung Cancer Patients by Targeted Sequencing. *Sci. Rep.* **6**, 31985, <https://doi.org/10.1038/srep31985> (2016).
16. Newman, A. M. *et al.* An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. *Nat. Med.* **20**, 548–554, <https://doi.org/10.1038/nm.3519> (2014).
17. Fiala, C. & Diamandis, E. P. Circulating tumor DNA for personalized lung cancer monitoring. *BMC Med.* **15**, 157, <https://doi.org/10.1186/s12916-017-0921-6> (2017).
18. Cohen, J. D. *et al.* Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Sci.* **359**, 926–930, <https://doi.org/10.1126/science.aar3247> (2018).
19. Diehl, F. *et al.* Detection and quantification of mutations in the plasma of patients with colorectal tumors. *Proc. Natl Acad. Sci. USA* **102**, 16368–16373, <https://doi.org/10.1073/pnas.0507904102> (2005).
20. Bettgowda, C. *et al.* Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci. Transl. Med.* **6**, 224ra224, <https://doi.org/10.1126/scitranslmed.3007094> (2014).
21. Cree, I. A. *et al.* The evidence base for circulating tumour DNA blood-based biomarkers for the early detection of cancer: a systematic mapping review. *BMC Cancer* **17**, 697, <https://doi.org/10.1186/s12885-017-3693-7> (2017).
22. Bratman, S. V., Newman, A. M., Alizadeh, A. A. & Diehn, M. Potential clinical utility of ultrasensitive circulating tumor DNA detection with CAPP-Seq. *Expert. Rev. Mol. Diagn.* **15**, 715–719, <https://doi.org/10.1586/14737159.2015.1019476> (2015).
23. Otsubo, K. *et al.* Genetic Profiling of Non-Small Cell Lung Cancer at Development of Resistance to First- or Second-Generation EGFR-TKIs by CAPP-Seq Analysis of Circulating Tumor DNA. *Oncologist* **24**, 1022–1026, <https://doi.org/10.1634/theoncologist.2019-0101> (2019).
24. Fan, H. C., Blumenfeld, Y. J., Chitkara, U., Hudgins, L. & Quake, S. R. Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood. *Proc. Natl Acad. Sci. USA* **105**, 16266–16271, <https://doi.org/10.1073/pnas.0808319105> (2008).
25. Fiala, C. & Diamandis, E. P. Utility of circulating tumor DNA in cancer diagnostics with emphasis on early detection. *BMC Med.* **16**, 166, <https://doi.org/10.1186/s12916-018-1157-9> (2018).
26. Diamandis, E. P. & Fiala, C. Can circulating tumor DNA be used for direct and early stage cancer detection? *F1000Res* **6**, 2129, <https://doi.org/10.12688/f1000research.13440.1> (2017).
27. Carrasco, J. L., Jover, L., King, T. S. & Chinchilli, V. M. Comparison of concordance correlation coefficient estimating approaches with skewed data. *J. Biopharm. Stat.* **17**, 673–684, <https://doi.org/10.1080/10543400701329463> (2007).
28. Goodwin, S., McPherson, J. D. & McCombie, W. R. Coming of age: ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* **17**, 333–351, <https://doi.org/10.1038/nrg.2016.49> (2016).
29. Miner, B. E., Stoger, R. J., Burden, A. F., Laird, C. D. & Hansen, R. S. Molecular barcodes detect redundancy and contamination in hairpin-bisulfite PCR. *Nucleic Acids Res.* **32**, e135, <https://doi.org/10.1093/nar/gnh132> (2004).
30. McCloskey, M. L., Stoger, R., Hansen, R. S. & Laird, C. D. Encoding PCR products with batch-stamps and barcodes. *Biochem. Genet.* **45**, 761–767, <https://doi.org/10.1007/s10528-007-9114-x> (2007).
31. Kinde, I., Wu, J., Papadopoulos, N., Kinzler, K. W. & Vogelstein, B. Detection and quantification of rare mutations with massively parallel sequencing. *Proc. Natl Acad. Sci. USA* **108**, 9530–9535, <https://doi.org/10.1073/pnas.1105422108> (2011).
32. Jabara, C. B., Jones, C. D., Roach, J., Anderson, J. A. & Swanson, R. Accurate sampling and deep sequencing of the HIV-1 protease gene using a Primer ID. *Proc. Natl Acad. Sci. USA* **108**, 20166–20171, <https://doi.org/10.1073/pnas.1110064108> (2011).
33. Kivioja, T. *et al.* Counting absolute numbers of molecules using unique molecular identifiers. *Nat. Methods* **9**, 72–74, <https://doi.org/10.1038/nmeth.1778> (2011).
34. Casbon, J. A., Osborne, R. J., Brenner, S. & Lichtenstein, C. P. A method for counting PCR template molecules with application to next-generation sequencing. *Nucleic Acids Res.* **39**, e81, <https://doi.org/10.1093/nar/gkr217> (2011).
35. Shiroguchi, K., Jia, T. Z., Sims, P. A. & Xie, X. S. Digital RNA sequencing minimizes sequence-dependent bias and amplification noise with optimized single-molecule barcodes. *Proc. Natl Acad. Sci. USA* **109**, 1347–1352, <https://doi.org/10.1073/pnas.1118018109> (2012).
36. Fu, G. K., Hu, J., Wang, P. H. & Fodor, S. P. Counting individual DNA molecules by the stochastic attachment of diverse labels. *Proc. Natl Acad. Sci. USA* **108**, 9026–9031, <https://doi.org/10.1073/pnas.1017621108> (2011).
37. Hewish, D. R. & Burgoyne, L. A. Chromatin sub-structure. The digestion of chromatin DNA at regularly spaced sites by a nuclear deoxyribonuclease. *Biochem. Biophys. Res. Commun.* **52**, 504–510, [https://doi.org/10.1016/0006-291x\(73\)90740-7](https://doi.org/10.1016/0006-291x(73)90740-7) (1973).
38. Hughes, F. M. Jr. & Cidlowski, J. A. Utilization of an *in vitro* assay to evaluate chromatin degradation by candidate apoptotic nucleases. *Cell Death Differ.* **4**, 200–208, <https://doi.org/10.1038/sj.cdd.4400221> (1997).
39. Matassov, D., Kagan, T., Leblanc, J., Sikorska, M. & Zakeri, Z. Measurement of apoptosis by DNA fragmentation. *Methods Mol. Biol.* **282**, 1–17, <https://doi.org/10.1385/1-59259-812-9:001> (2004).
40. Ivanov, M., Baranova, A., Butler, T., Spellman, P. & Mileyko, V. Non-random fragmentation patterns in circulating cell-free DNA reflect epigenetic regulation. *BMC Genomics* **16**(Suppl 13), S1, <https://doi.org/10.1186/1471-2164-16-S13-S1> (2015).
41. Snyder, M. W., Kircher, M., Hill, A. J., Daza, R. M. & Shendure, J. Cell-free DNA Comprises an *In Vivo* Nucleosome Footprint that Informs Its Tissues-Of-Origin. *Cell* **164**, 57–68, <https://doi.org/10.1016/j.cell.2015.11.050> (2016).
42. Xu, C. *et al.* smCounter2: an accurate low-frequency variant caller for targeted sequencing data with unique molecular identifiers. *Bioinformatics* **35**, 1299–1309, <https://doi.org/10.1093/bioinformatics/bty790> J Bioinformatics (2018).

Acknowledgements

We thank C. Hui for assistance in blood sampling. We express gratitude to representatives from Roche and QIAGEN for their technical support on the experiments.

Author contributions

C.K. Chow and S.N. Lam conceived and planned the study. C.M. Foo and P.Y. Lee performed the library preparations and sequencing. Y.M. Chan guided data analysis. W.Y. Mok, W.S. Wong and Y.Y. Fung contributed to the data analysis and visualization. S.N. Lam prepared the manuscript. Y.C. Zhou, K.Y. Wong and J.Y. Huang participated in discussions, assisted in the manuscript preparation and critically revised the manuscript. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-60375-x>.

Correspondence and requests for materials should be addressed to C.K.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020