Application Note

Detect ultra-rare variants in cfDNA from liquid biopsy samples using an end-to-end NGS workflow

Zhong Wu[1], Krishna Amin[1], Christa Haldrup[2], Emil Christensen[2] and Simon Hertlein[3]

[1] QIAGEN LLC, Germantown, USA; [2] QIAGEN Aarhus, Aarhus, Denmark; [3] QIAGEN GmbH, Hilden, Germany

Introduction

Liquid biopsy is becoming a powerful enabler of precision oncology. Instead of relying on invasive procedures, researchers can repeatedly sample and analyze cell-free DNA (cfDNA) and other liquid biopsy analytes from body fluids to provide real-time information about tumor dynamics and genetic changes, facilitating personalized treatments (1).

Next-generation sequencing (NGS) of cfDNA in cancer research stands out from other detection technologies due to its high discovery power. The high throughput, exceptional sensitivity, remarkable multiplexing capabilities and comprehensive genome-wide coverage of NGS, when combined with non-invasive sampling, enable fast discovery of novel variants, providing a comprehensive ▷

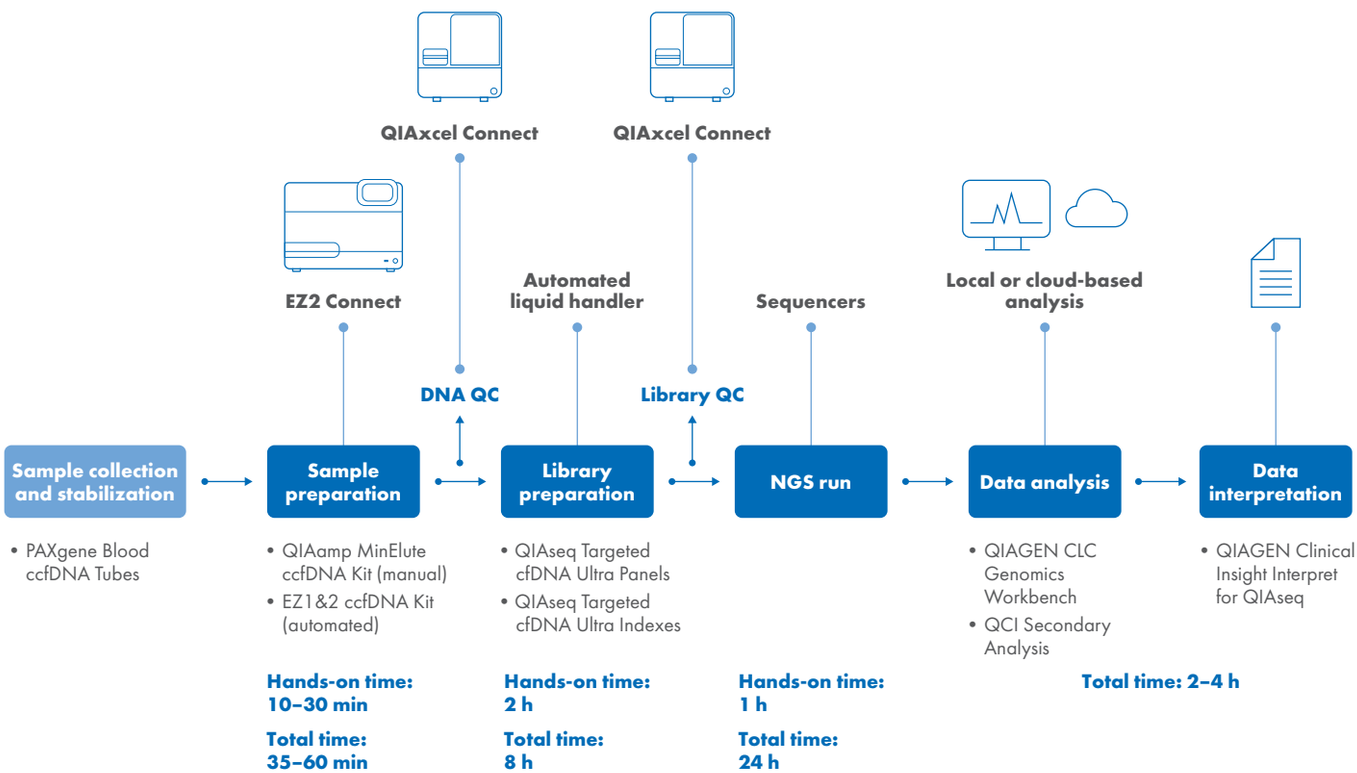

Figure 1. A time-efficient workflow for targeted cfDNA sequencing

view of the genetic and molecular landscape of tumors (2,3).

Yet, the development of NGS-based assays for cfDNA is limited by challenges. cfDNA is highly fragmented, present in low quantities in the blood and can be further diluted by genomic DNA (gDNA) from apoptotic or lysed cells. Therefore, pre-analytical stabilization is required before cfDNA isolation from blood and other biofluids (2,3).

Extremely low cfDNA concentrations, well below 10 ng/µL, particularly in non-advanced, non-metastatic cancers, limit the availability of cfDNA as a testing analyte. This warrants the need for large input volumes to detect rare mutations with high sensitivity (4). Besides, circulating tumor DNA (ctDNA) often represents just 1–2% of total cfDNA, thus requiring a limit of detection below 1% in many applications (5). Further down the workflow, sequencing reaction errors, enrichment biases and false positives drive the limit of detection (LoD) up to an estimated 0.1% variant allele frequency (VAF) (6). Hence, we need a solution to systematically lower the background errors and confidently call 0.1% VAF with high sensitivity.

Here, we present an end-to-end workflow that enables ultra-rare variant detection from liquid biopsy samples down to 0.1% VAF with >99% specificity (Figure 1). With both automated and manual options for sample preparation, this workflow aims to empower labs working on areas such as liquid biopsy research, diagnostics development, therapy response monitoring, early resistance mechanisms detection and minimal residual disease monitoring to go from sample to actionable insights with confidence.

Materials and Methods

Recommendation for standardized blood collection, stabilization, storage and transport

Use the PAXgene® Blood ccfDNA Tube to minimize the release of DNA into plasma and stabilize your sample during transport and storage. Additionally, benefit from a formaldehyde-free, non-crosslinking stabilization of the circulating, cfDNA in plasma and genomic DNA in the nuclear cellular fraction.

Sample

SeraCare reference material with 0.125 % mutation frequency [Seraseq® ctDNA Mutation Mix v2 (AF 0.125%)] was used.

Manual sample preparation

Samples were prepared manually from 10 mL input volumes using the QIAamp® MinElute® ccfDNA Midi Kit and eluted in 55 µL.

Automated sample preparation

Samples were prepared from 8 mL input volumes through automated sample preparation on the EZ2® Connect using EZ1&2™ ccfDNA Kit and eluted in 45 µL.

Quality control of extracted cfDNA and libraries

Sample yield was determined using the Qubit™ 1x dsDNA HS Assay (Thermo Fisher Scientific). cfDNA size distribution and quality were analyzed on the QIAxcel® Connect using the QIAxcel DNA High Sensitivity Kit.

NGS library preparation

Libraries were prepared with 50 ng cfDNA as input using the QIAseq® Targeted cfDNA Ultra Lung Cancer Panel workflow. cfDNA samples were first end-repaired and A-tailed within a single, controlled multienzyme reaction. The prepared DNA fragments were then ligated with a sequencing platform-specific adapter containing unique molecular index (UMI). For enrichment, ligated DNA

molecules were subjected to several cycles of targeted PCR using one region-specific primer and one universal primer complementary to the adapter. Lastly, a universal PCR was carried out to amplify the library and add platform-specific adapter sequences and sample indices.

NGS

The libraries were sequenced on an Illumina® NextSeq® 500/550 Mid Output (300 cycles) instrument – PE150. Demultiplexing and FASTQ file generation were performed on the sequencing instrument.

Data analysis

Variants were detected using the QIAseq DNA Ultra Somatic Variants (Illumina) workflow available in the Biomedical Genomics Analysis plugin of the QIAGEN CLC Genomics Workbench. The workflow is also available as a cloud-based service, QIAGEN Clinical Insights Secondary Analysis (QCI®-SA), which automates the tasks of FASTQ upload, workflow execution and results handling. Variant interpretation and report generation were then performed using QIAGEN Clinical Insight (QCI) Interpret for QIAseq.

Results and Discussion

Easy and reproducible extraction of high-quality cfDNA for sensitive NGS analysis

Both manual and automated cfDNA extraction methods yielded consistent and comparable cfDNA in terms of concentration and size. Results show that the cfDNA yields from both methods were more than 130 ng as determined using the Qubit 1x dsDNA HS Assay (Figure 2A).

The higher sample input volume used with the QIAamp MinElute ccfDNA Midi Kit clearly shows an increase in yield (Figure 2A). The extracted cfDNA shows the expected size distribution of the mononucleosomal (~160 bp) and dinucleosomal peaks (~300 bp) for both extraction methods on the QIAxcel Connect (DNA High-Sensitivity Cartridge). No cellular gDNA contamination was observed (Figure 2B).

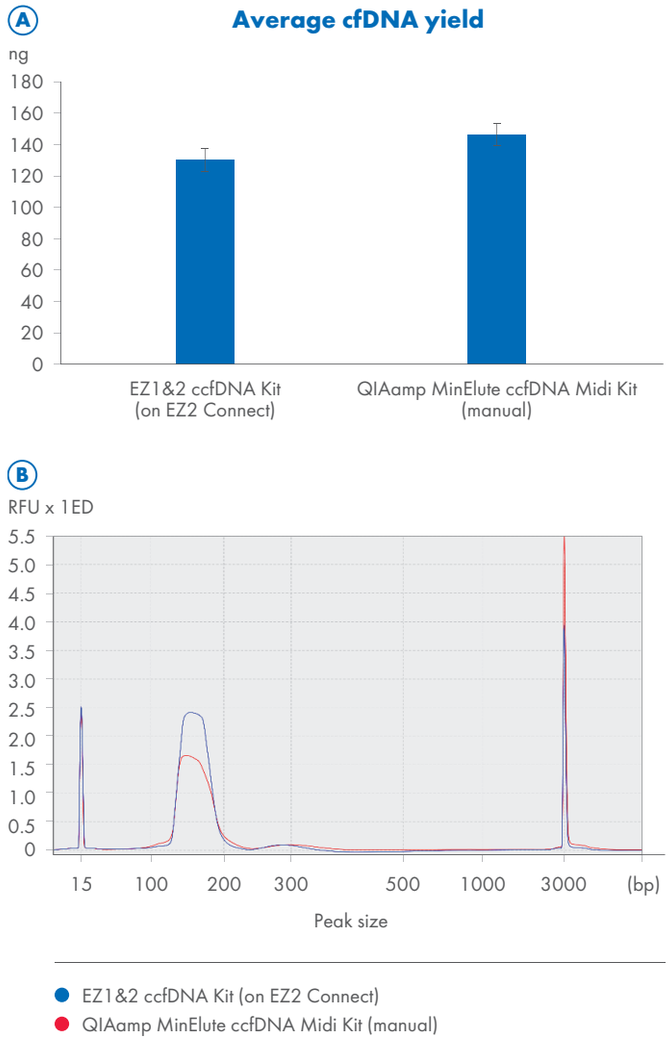

Figure 2. High quality and yield of cfDNA. A Both methods yielded more than 130 ng of cfDNA. The averages of three replicates are shown with standard deviations. **B** The extracted cfDNA shows the expected size distribution of the mononucleosomal (~160 bp) and dinucleosomal peaks (~300 bp) for both extraction methods on the QIAxcel Connect (DNA High-Sensitivity Cartridge).

The described procedure provides scalable input volumes of up to 10 mL, ensuring the detection of low-frequency actionable variants. Both methods allow low elution volumes for high concentrations of nucleic acids – a requirement for successful library preparation. ▷

Additionally, the EZ1&2 ccfDNA Kit enables fully automated processing of large sample volumes without manual pre-enrichment or preparation of plates or tubes.

Highly efficient and sensitive NGS library preparation workflow with integrated UMIs and optimized high-fidelity chemistry

QIAseq Targeted cfDNA Ultra Lung Cancer Panel, which targets relevant genes for hotpots and actionable variants, also includes genes targeting full coding regions to enable CNV analysis. Results show target specificity greater than 90% and uniformity of coverage greater than 99% across libraries prepared using samples from automated EZ1&2 ccfDNA Kit and manual QIAamp MinElute ccfDNA Midi Kit workflows (Figure 3). The consistent on-target rates across different library preparations (manual or automated) illustrate the efficient target capture of QIAseq target enrichment.

Accurate detection of cfDNA variants by error-correction bioinformatics pipeline

The secondary analysis pipeline for QIAseq Targeted cfDNA Ultra Panels accurately detected all SNVs and indels from SeraCare cfDNA reference material extracted using the manual and automated workflows (Table 1).

Two special considerations for variant calling of cfDNA variants for QIAseq Targeted cfDNA Ultra Panels were taken:

- The UMI grouping method needed to accommodate higher error rates in UMI sequences of the same group as cfDNA samples typically involves low DNA input and many PCR cycles, resulting in many reads per UMI. A more relaxed UMI grouping was introduced to allow more than one mismatch in the UMI sequence.
- The calling of low-frequency variants was optimized by calculating the quality score of consensus based on the frequency of errors instead of the Q-scores in the original reads (7).

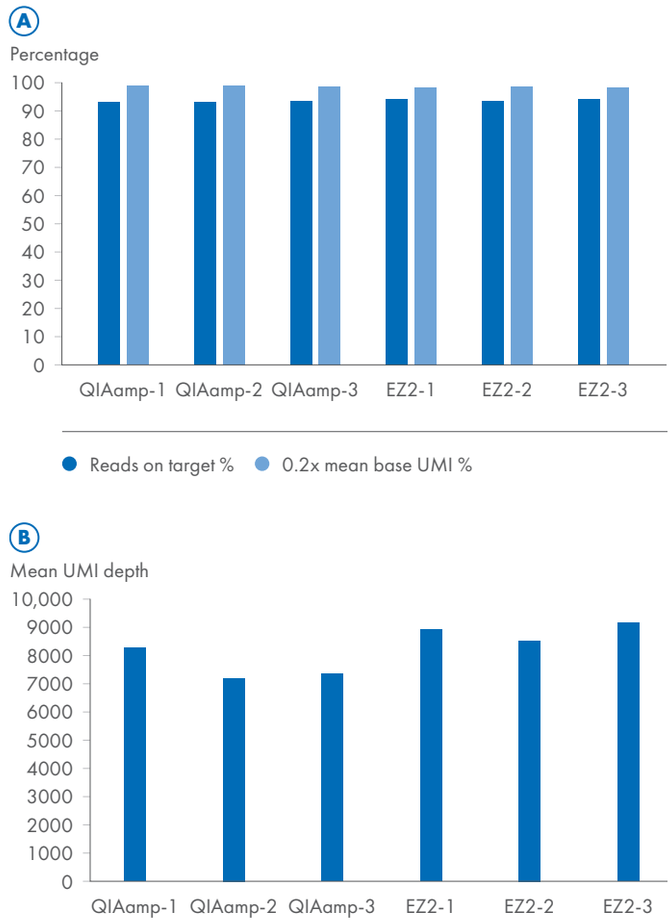

Figure 3. Highly efficient and sensitive NGS library preparation. **A** Target specificity and uniformity of coverage **B** Mean UMI depth per target base. No notable differences were seen in the libraries obtained using cfDNA prepared with the EZ1&2 ccfDNA Kit (automated) and QIAamp MinElute ccfDNA Midi Kit (manual) methods. Three replicates per method are shown.

Variant classification using QCI Interpret for QIAseq

The QCI Interpret report was created by uploading the vcf files from the secondary analysis to QCI Interpret, an online software tool that allows you to upload and analyze variant calls from genomic data from QIAseq panels for variant interpretation and reporting (Figure 4).

QCI Interpret uses the QIAGEN Knowledge Base, a database that contains over 5 million variant findings as well as data from third-party databases, and rules-based approaches to automatically compute pathogenicity classifications (pathogenic to benign) and actionability classifications (tier 1 to 4) for each variant in the vcf.

Table 1. SNV and indel allele frequencies of SeraCare reference material with 0.125 % mutation frequency with QIAamp MinElute ccfDNA Midi Kit (manual) and the EZ1&2 ccfDNA Kit (automated) workflows

	QIAamp MinElute ccfDNA Midi Kit		EZ1&2 ccfDNA Kit	
	Average	Standard deviation	Average	Standard deviation
BRAF V600E	0.14	0.11	0.10	0.07
EGFR L858R	0.14	0.02	0.19	0.01
EGFR D770_N771insG	0.20	0.02	0.22	0.04
EGFR E746_A750del	0.14	0.00	0.13	0.02
EGFR T790M	0.16	0.01	0.14	0.05
ERBB2 Y772_A886dup	0.15	0.03	0.14	0.03
KRAS G12D	0.21	0.04	0.28	0.06
NRAS/CSDE1 Q61R	0.18	0.03	0.16	0.02
PIK3CA E545K	0.16	0.04	0.05	0.07
PIK3CA H1047R	0.12	0.02	0.13	0.04
RET M918T	0.13	0.03	0.14	0.05
TP53 R175H	0.19	0.05	0.22	0.03
TP53 R273H	0.20	0.03	0.13	0.02
TP53 R248Q	0.20	0.02	0.24	0.04
*TP53 C242Afs*5*	0.14	0.02	0.14	0.03
*TP53 S90Pfs*33*	0.13	0.05	0.15	0.02

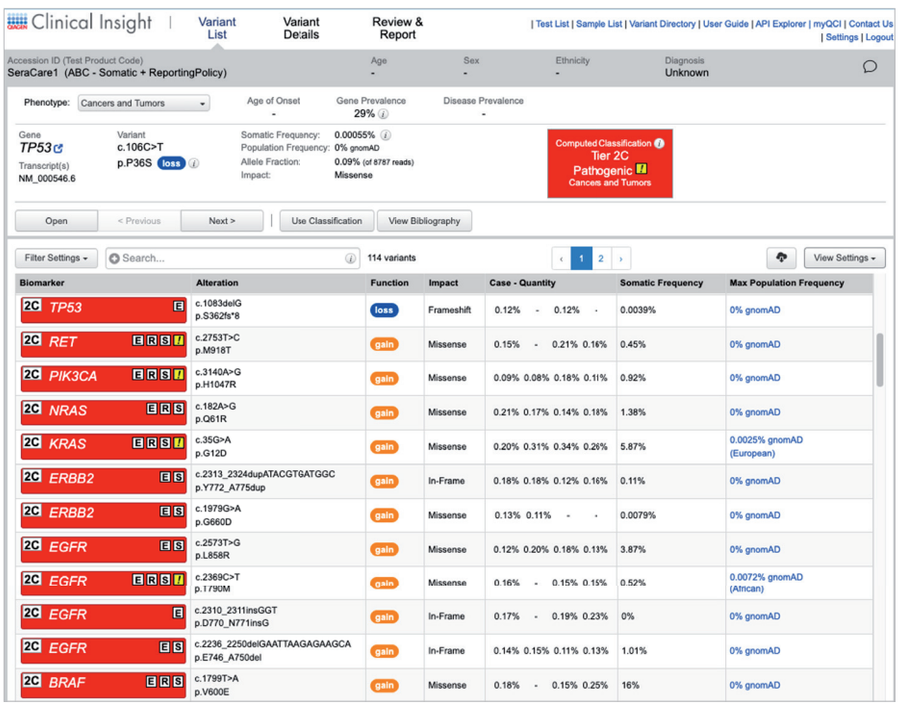

Figure 4. Variant pathogenicity classifications in QCI Interpret for QIAseq. The software ranks detected variants based on the evidence from the QIAGEN Knowledge Base. The software also identifies the alteration, position, molecular function, impact, somatic frequency and population frequency.

Conclusion

NGS-based detection of ultra-rare variants in cfDNA is technically challenging due to the low amount and quality of cfDNA, the high level of noise and errors in sequencing and the requirement for high accuracy and precision. Here, we describe a fast and efficient Sample to Insight workflow to confidently detect ultra-rare cfDNA variants – down to 0.1% VAF – with high sensitivity.

Both manual QIAamp MinElute ccfDNA and automated EZ1&2 ccfDNA workflows enable cfDNA extraction with high yield and purity from large plasma volumes of up to 10 mL. Additionally, the fully automatable workflow on the EZ2 Connect eliminates any manual pre-enrichment or preparation of plates, yielding sample to sequencing-ready libraries in just 8 hours and actionable insights in fewer than 4 days.

The QIAseq Targeted cfDNA Ultra Panels incorporate high-fidelity polymerase to lower the false positive rate, as well as UMIs to reduce PCR and sequencing errors. This leads to improved detection of genetic alterations, allowing for the detection of variants at ultra-low allele frequencies down to 0.1% with >99% specificity (from internal data not shown).

Empowered with the ability to detect ultra-rare variants from cfDNA, researchers can further push boundaries in various research areas, including stratification of individuals for prospective studies, biomarker identification and analysis, minimal residual disease (MRD) and therapy monitoring.

References

1. Nikanjam M, Kato S, Kurzrock R. Liquid biopsy: Current technology and clinical applications. *J Hematol* Oncol. 2022;15(1):131.
2. Song P, et al. Limitations and opportunities of technologies for the analysis of cell-free DNA in cancer diagnostics. *Nat Biomed Eng.* 2022;6(3):232–245.
3. Voss T, Ullius A, Schönborn M, Oelmüller U. Sensitivity assessment of workflows detecting rare circulating cell-free DNA targets: A study design proposal. *PLoS One.* 2021;16(7):e0253401.
4. Chen M, Zhao H. Next-generation sequencing in liquid biopsy: Cancer screening and early detection. *Hum Genomics.* 2019;13(1):34.
5. Ossandon MR, et al. Circulating tumor DNA assays in clinical cancer research. *J Natl Cancer Inst.* 2018;110(9):929–934.
6. Xu C, et al. smCounter2: An accurate low-frequency variant caller for targeted sequencing data with unique molecular identifiers. *Bioinformatics.* 2019;35(8): 1299–1309.
7. Shugay M, et al. MAGERI: Computational pipeline for molecular-barcoded targeted resequencing. *PLoS Comput Biol.* 2017;13(5):e1005480.

For up-to-date licensing information and product-specific disclaimers, see the PreAnalytiX website (**www.preanalytix.com/trademarks-disclaimers**) and respective PreAnalytiX product webpages. PreAnalytiX instructions for use and handbooks are available at **www.preanalytix.com**, **eifu.bd.com**, **www.qiagen.com**, or can be requested from BD or QIAGEN Technical Services or your local distributor.

QIAGEN products shown here are intended for molecular biology applications. These products are not intended for the diagnosis, prevention or treatment of a disease.

For up-to-date licensing information and product-specific disclaimers, see the respective QIAGEN kit instructions for use or user operator manual. QIAGEN instructions for use and user manuals are available at **www.qiagen.com** or can be requested from QIAGEN Technical Services (or your local distributor).

Scan the QR code or use the link to learn more: **www.qiagen.com/cfDNA**