Application Note
High-performance somatic variant analysis

KAPA HyperPETE: A novel, end-to-end target enrichment workflow for high-performance somatic variant analysis

Authors:

Nitya Furtado
Senior Scientist – Reagent and Assay Development

Mingyang Cai*
Principal Bioinformatics Scientist II

Jingchuan Li
Principle Scientist I – Reagent and Assay Development

Florence Crary-Dooley
Senior Scientist – Reagent and Assay Development

Liu Xi
Principle Bioinformatics Scientist I

Brian Godwin
Director – Reagent and Assay Development

Jimmy Tu
International Product Manager

Roche Molecular Labs,
Pleasanton, CA, USA

Mariana Fitarelli-Kiehl
Senior Applications Scientist – Support & Applications

Spencer Debenport*
Applications Manager

Roche Diagnostics Corp. US,
Wilmington, MA, USA

Date of first publication: 2022
Date of this publication: 2024

*No longer employed by Roche..

Targeted NGS with small panels is a well-established and sensitive tool for somatic oncology research. Current workflows continue to fall short with respect to sample preparation robustness and reproducibility, content, and integrated bioinformatics tools. KAPA HyperPETE Workflows comprise a new class of end-to-end solutions for flexible, high performance analysis of all major mutation classes from plasma cfDNA, as well as tissue and cell line DNA and RNA.

Introduction

Cancers are genetically complex and heterogeneous. Whilst germline (inherited) variants play a major role in 5 – 10% of all cancers,[1] it is the somatic (acquired) variants that drive disease progression and response to therapeutic intervention. Every individual cancer has a unique combination of genetic alterations, ranging from single nucleotide variants (SNVs) and short insertions/deletions (indels), to microsatellite instability (MSI) and large structural rearrangements, such as gene fusions and gene deletions/duplications causing copy number variation (CNV). Targeted next-generation sequencing (NGS) is an established and powerful tool for studying the genetic basis of cancer, offering the ability to survey large numbers of targets (potential variants) in parallel, at high sensitivity (variant allele frequencies of 5% or less).[2] NGS pipelines for the analysis of somatic variants require three critical components: (i) efficient methods for the preparation of enriched libraries from challenging sample types including cell free/circulating tumor DNA (cf/ctDNA), formalin-fixed paraffin-embedded (FFPE) tissue samples, and total RNA; (ii) relevant, optimized enrichment panels (content), and (iii) advanced bioinformatic tools.

The KAPA HyperPETE Workflow offers a streamlined, versatile, end-to-end solution for high-performance somatic variant analysis using small (≤300 kb catalog and ≤250 kb custom) panels. Based on the Roche novel Primer Extension Target Enrichment (PETE) technology (Figure 1), the Workflow enables analysis of DNA and/or RNA extracted from liquid biopsies, FFPE tissues, and cell lines—with catalog and custom panel content designed to interrogate all major mutation classes. The single-day library preparation workflow incorporates unique molecular identifiers (UMIs) and eliminates tedious processing steps required for hybridization capture, whilst offering improved performance over amplicon-based protocols.[3]

In this Application Note we demonstrate the performance and versatility of the KAPA HyperPETE technology for somatic variant analysis, achieving true positive rates of 97.9 – 100% for all major variant classes, across >450 libraries prepared from cfDNA, FFPE DNA, FFPE RNA and reference cell line DNA and RNA samples.

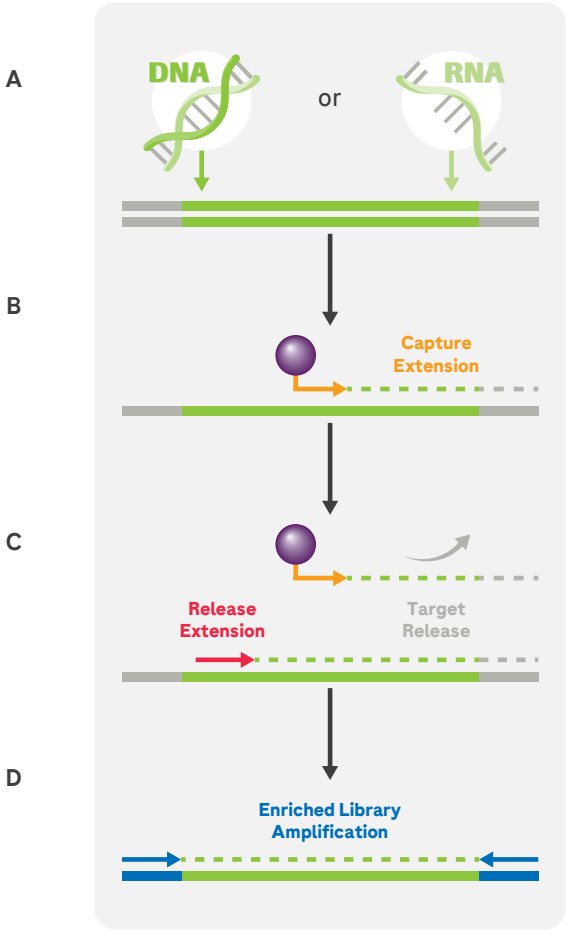

Figure 1. Overview of the KAPA HyperPETE (Primer Extension Target Enrichment) Technology. (A) Uniquely dual-indexed pre-capture libraries are generated from genomic DNA, FFPE DNA, cfDNA, or RNA using the appropriate (KAPA HyperPlus, KAPA HyperPrep, or KAPA RNA HyperPrep) library preparation kit and KAPA Universal UMI adapters. **(B)** Following a heated denaturation step, biotinylated target-specific capture primers (orange) are annealed and extended using a DNA polymerase. Paramagnetic streptavidin beads (purple) are used to capture the hybrid molecules. Off-target, uncaptured library fragments are washed away (not shown). **(C)** Target-specific release primers (red) are hybridized and extended by a DNA polymerase, specifically releasing target molecules from the beads into the supernatant (an intermediate wash step to remove non-hybridized primers is performed prior to the release extension). **(D)** Released products are amplified with universal primers (blue). Final libraries are purified, quantified, pooled, and prepared for sequencing. The workflow may be completed in 8 – 10 hours, depending on sample type and the number of samples that are processed.

Materials and methods

Experimental design

This study was designed to demonstrate the versatility and performance of the KAPA HyperPETE Workflow for somatic variant analysis. To this end, libraries were prepared from (i) cell line and plasma derived cfDNA, (ii) cell-line, tissue and xenograft-derived FFPE DNA, and (iii) cell-line and tissue derived FFPE RNA. Non-FFPE cell line controls were included in the Somatic Tissue DNA and RNA workflows. Fragment libraries were prepared with the KAPA HyperPrep Kit (Plasma cfDNA Workflow), KAPA HyperPlus Kit (Tissue DNA Workflow), or KAPA RNA HyperPrep Kit (Tissue RNA Workflow). Primer Extension Target Enrichment (PETE) was performed with catalog or custom KAPA HyperPETE Panels (capture sizes ranging from 7.5 – 301 kb) according to the standard workflow for each sample type and panel. Sequencing was performed on the Illumina® platform using standard protocols. Data analysis was performed with an internal Roche secondary analysis pipeline. Appropriate sequencing and variant calling metrics are reported for each workflow. A summary of the experimental design is given in Figure 2.

Samples and DNA Extraction

Plasma cfDNA Workflow: Four reference materials (purified cell line ctDNA or ctDNA mixtures; Table 1, lines 1 – 4) were obtained from a commercial supplier.

Sixteen plasma samples from healthy donors (Table 1, line 5) were obtained from a commercial supplier. For the KAPA HyperPETE Plasma cfDNA Workflow, cfDNA may be extracted with any appropriate method from blood collected in EDTA-containing collection tubes and handled according to standard procedures for plasma samples. Extracted cfDNA was quantified using a Qubit® Fluorometer and Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific). To confirm that cfDNA extracts were free of high-molecular weight genomic DNA, samples were subjected to electrophoretic analysis using a 4200 TapeStation system and Cell-free DNA ScreenTape Assay, or 2100 Bioanalyzer system and High Sensitivity DNA Kit (Agilent® Technologies).

Tissue DNA Workflow: Sixteen reference samples from characterized cell lines were obtained from commercial suppliers or internal sources (Table 2, lines 1 – 16). These included purified genomic DNA (Table 2, line 6 and lines 11 – 16), mixes of purified DNA (Table 2, lines 5, 7 – 10) from unprocessed cells (non-FFPE samples), and FFPE curls prepared from cell cultures (Table 2, lines 1 – 4). FFPE xenograft samples (curls; Table 2, lines 17 – 18) were obtained from a commercial supplier, and sixteen surgical tissue samples (FFPE curls, Table 2, line 19) from internal sources.

DNA was extracted from FFPE curls (Table 2, lines 1 – 4 and lines 17 – 19; 1 – 2 sections of up to 20 µm per extraction) with the KAPA NGS DNA Extraction Kit (Roche PN: 09189823001 or 09190023001) as described in the KAPA HyperPETE Somatic Tissue DNA Workflow Instructions for use (Version 1.0 or later).[4]

All DNA samples (FFPE and non-FFPE) were quantified using a Qubit® Fluorometer and Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific). The quality of all formalin-compromised DNA samples (Table 2, lines 1 – 4 and lines 17 – 19) was assessed with the qPCR-based KAPA NGS FFPE DNA QC Kit (Roche PN: 09217193001 or 09306889001) as described in the KAPA HyperPETE Workflow, Somatic Tissue DNA Preparation Instructions for use (Version 1.0 or later).4 DNA extracts were divided into high, mid or low quality based on the normalized Q (quality) score calculated for each sample.

Tissue RNA Workflow: Five reference samples from characterized cell lines were obtained from commercial suppliers (Table 3, lines 1 – 5). These included purified, total RNA from unprocessed cells (Table 3, lines 4 – 5) and FFPE curls prepared from cell cultures (Table 3, lines

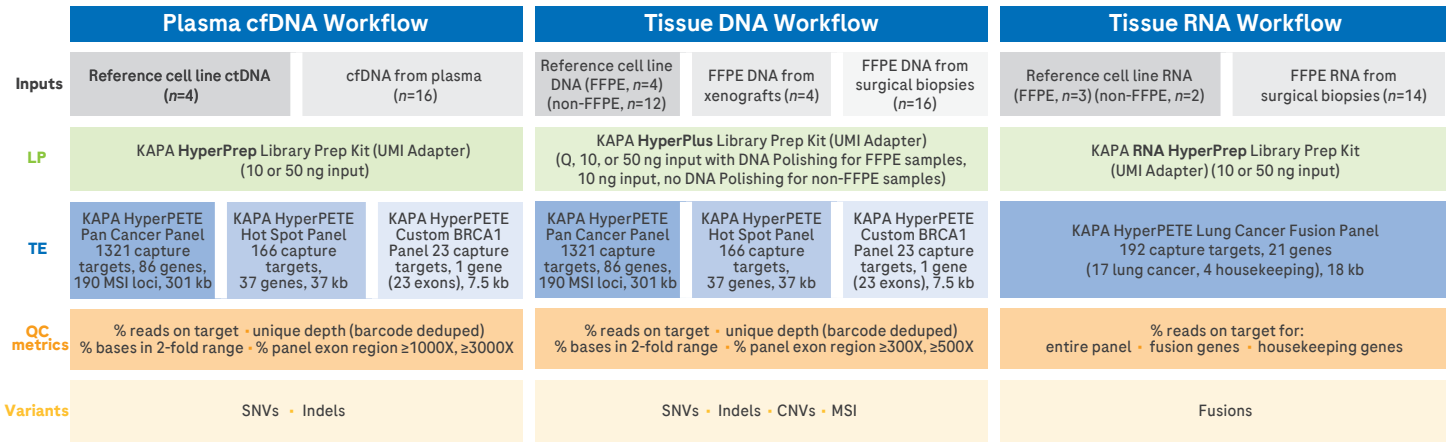

Figure 2. Summary of the experimental design for this study. Full details of reference (cell line), plasma, xenograft, and biopsy samples (inputs) for each of the three workflows are provided in Materials and Methods. Uniquely dual-indexed libraries were prepared as described. For FFPE DNA samples Q=input DNA mass calculated from normalized Q score. Quality control assays (quantification, library fragment size assessment) performed on pre- and post-enrichment libraries are not shown in the diagram, but were performed as described in Materials and Methods. Single (1-plex) captures were performed with all KAPA HyperPETE panels. Data analysis and variant calling was performed with an internal Roche secondary analysis pipeline. Full definitions of QC metrics are given in Results and Discussion. LP: Library Preparation, TE: Target Enrichment, UMI: Unique Molecular Identifier.

1 – 3). Fourteen surgical tissue samples (FFPE curls; Table 3, line 6) were obtained internally.

Total RNA was extracted from FFPE curls (1 – 2 sections per extraction) with the High Pure FFPET RNA Isolation Kit (Roche PN: 06650775001).[5] All RNA samples were quantified using a Qubit® Fluorometer and Qubit RNA HS Assay Kit (Thermo Fisher Scientific). RNA quality was assessed with a 2100 Bioanalyzer system and RNA 6000 Pico Kit (Agilent Technologies). Extracts were divided into high, mid, or low quality based on the DV_{200} value (the percentage of RNA fragments >200 nt) calculated for each sample.[6]

Library Preparation and Target Enrichment

Plasma cfDNA Workflow: A total of 144 cfDNA libraries were prepared from 10 ng or 50 ng inputs for target enrichment with three different KAPA HyperPETE Panels, as outlined in Table 4. Reference materials (cell line ctDNA) were processed in duplicate, whereas only one library per input was prepared from each of the plasma samples (which represent biological replicates).

All cfDNA libraries were prepared with the KAPA HyperPrep Kit (Roche PN: 07962347001 or 07962363001), KAPA Universal UMI Adapter (Roche PN: 09329862001 or 09329889001), KAPA UDI Primer Mixes 1 – 96 (Roche PN: 09134336001), and KAPA HyperPure Beads (Roche PN: 08963835001, 08963843001, 08963851001, 08963878001, or 08963860001) as described in the KAPA HyperPETE Somatic Plasma cfDNA Workflow Instructions for use (Version 1.0 or later).[7]

Tissue DNA Workflow: A total of 240 FFPE DNA libraries and 48 libraries from non-FFPE DNA controls were prepared for target enrichment with three different KAPA HyperPETE Panels, as outlined in Table 4. Depending on the panel, FFPE libraries were prepared from 10 ng, 50 ng, and/or Q ng of input DNA, where Q is the input mass calculated from the normalized Q score determined with the qPCR-based KAPA NGS FFPE DNA QC Kit. The value of Q ranged from 20 – 26 ng for high-quality FFPE DNA extracts, from 39 – 54 ng for mid-quality extracts, and from 60 – 248 ng for low-quality FFPE DNA preparations. Non-FFPE DNA samples were all regarded as high quality, and libraries were prepared from 10 ng inputs only. Reference materials (FFPE

Table 1. Samples used in the Plasma cfDNA Workflow

	Sample name/ part number	Sample description	Geno-type	Sample type	Sample format	Source/reference
1	0710-0143	Seraseq® ctDNA Mutation Mix v2 AF0.125%	Mut	Reference (cell line)	Purified ctDNA mix	https://www.seracare.com/Seraseq-ctDNA-Mutation-Mix-v2-AF0125-0710-0143/
2	0710-0141	Seraseq® ctDNA Mutation Mix v2 AF0.5%	Mut	Reference (cell line)	Purified cfDNA mix	https://www.seracare.com/Seraseq-ctDNA-Mutation-Mix-v2-AF05-0710-0141/
3	0710-0140	Seraseq® ctDNA Mutation Mix v2 AF1%	Mut	Reference (cell line)	Purified cfDNA mix	https://www.seracare.com/Seraseq-ctDNA-Mutation-Mix-v2-AF1-0710-0140/
4	0710-0144	Seraseq® ctDNA Mutation Mix v2 WT	WT	Reference (cell line)	Purified cfDNA	https://www.seracare.com/Seraseq-ctDNA-Mutation-Mix-v2-WT-0710-0144/
5	Various (n=16)	Samples from healthy donors	WT	Plasma	Plasma samples, cfDNA extracted in this study	BioCollections Worldwide, Inc.

Data on file with Roche.
For Research Use Only. Not for use in diagnostic procedures.

Table 2. Samples used in the Tissue FFPE DNA Workflow

	Sample name/ part number	Sample description	Geno-type	Sample type	Sample format	Source/reference
1	GM24143	Genome in a Bottle Ashkenazim PGP Mother Reference Standard	WT	Reference (cell line)	FFPE curls (10 – 15 µm), DNA extracted in this study	https://horizondiscovery.com/en/reference-standards/products/genome-in-a-bottle-ashkenazim-pgp-mother-reference-standard
2	GM24149	Genome in a Bottle Ashkenazim PGP Father Reference Standard	WT	Reference (cell line)	FFPE curls (10 – 15 µm), DNA extracted in this study	https://horizondiscovery.com/en/reference-standards/products/genome-in-a-bottle-ashkenazim-pgp-father-reference-standard?catalognumber=GM24149
3	HD200	Quantitative Multiplex Reference Standard (FFPE)	Mut	Reference (cell line)	FFPE curls (10 – 15 µm), DNA extracted in this study	https://horizondiscovery.com/en/reference-standards/products/quantitative-multiplex-reference-standard-ffpe?catalognumber=HD200
4	HD789	Structural Multiplex Reference Standard (FFPE)	Mut	Reference (cell line)	FFPE curls (10 – 15 µm), DNA extracted in this study	https://horizondiscovery.com/en/reference-standards/products/structural-multiplex-reference-standard-ffpe
5	HD753	Structural Multiplex Reference Standard (gDNA)	Mut	Reference (cell line)	Purified genomic DNA mix (non-FFPE)	https://horizondiscovery.com/en/reference-standards/products/structural-multiplex-reference-standard-gdna
6	Custom part no.	GM24385 Genomic DNA (RSS, Sample 9)	WT	Reference (cell line)	Purified genomic DNA (non-FFPE)	LGC SeraCare
7	Custom part no.	Custom gDNA Mutation Mix (RSS, Sample 1)	Mut	Reference (cell line)	Custom purified DNA mix (non-FFPE)	LGC SeraCare
8	Custom part no.	Custom gDNA Mutation Mix, (RSS, Sample 3)	Mut	Reference (cell line)	Custom purified DNA mix (non-FFPE)	LGC SeraCare
9	Custom part no.	Custom gDNA Mutation Mix, (RSS, Sample 4)	Mut	Reference (cell line)	Custom purified DNA mix (non-FFPE)	LGC SeraCare
10	0710-0412	Seraseq® Breast CNV Mix, +6 copies	Mut	Reference (cell line)	Purified DNA mix (non-FFPE)	https://www.seracare.com/Seraseq-Breast-CNV-Mix--6-copies-0710-0412/
11	MOLT-4	MSI cell line DNA	Mut	Reference (cell line)	Sequence-verified, purified DNA (non-FFPE)	Internal
12	SW48	MSI cell line DNA	Mut	Reference (cell line)	Sequence-verified, purified DNA (non-FFPE)	Internal
13	DLD-1	MSI cell line DNA	Mut	Reference (cell line)	Sequence-verified, purified DNA (non-FFPE)	Internal
14	A-172	MSS cell line DNA	Mut	Reference (cell line)	Sequence-verified, purified DNA (non-FFPE)	Internal
15	A549	MSS cell line DNA	Mut	Reference (cell line)	Sequence-verified, purified DNA (non-FFPE)	Internal
16	AU565	MSS cell line DNA	Mut	Reference (cell line)	Sequence-verified, purified DNA (non-FFPE)	Internal
17	DU145 SW48	Xenograft, MSI	Mut	Xenograft	FFPE curls (10 – 15 µm), DNA extracted in this study	CrownBio
18	BT474 MDA-MB-453	Xenograft, MSS	Mut	Xenograft	FFPE curls (10 – 15 µm), DNA extracted in this study	CrownBio
19	Various (*n*=16)	Surgical biopsy	WT	Tissue	FFPE curls (10 µm), DNA extracted in this study	Internal

and non-FFPE, including xenograft samples analyzed for MSI/MSS) were processed in duplicate, whereas only one library per input was prepared from surgical samples (which represent biological replicates).

All libraries were prepared with the KAPA HyperPlus Kit (Roche PN: 07962401001 or 07962428001), KAPA Universal UMI Adapter (Roche PN: 09329862001 or 09329889001), KAPA UDI Primer Mixes 1 – 96 (Roche PN: 09134336001), and KAPA HyperPure Beads (Roche PN: 08963835001, 08963843001, 08963851001, 08963878001, or 08963860001) as described in the KAPA HyperPETE Somatic Tissue DNA Preparation Workflow Instructions for use (Version 1.0 or later).4 All formalin-compromised samples (Table 2, lines 1 – 4 and lines 17 – 19) were subjected to DNA Polishing with the KAPA NGS FFPE DNA Polishing Kit (Roche PN: 09217215001 or 09217223001) prior to enzymatic fragmentation.

Tissue RNA Workflow: A total of 40 FFPE RNA libraries and eight libraries from non-FFPE RNA controls were prepared for target

Table 3. Samples used in the Tissue RNA Workflow

	Sample name	Sample description	Geno-type	Sample type	Sample format	Source/reference
1	HD784	ALK RET ROS RNA fusion	Mut	Reference (cell line)	FFPE curls (10 – 15 µm), RNA extracted in this study	https://horizondiscovery.com/en/reference-standards/products/alk-ret-ros-rna-fusion
2	0710-0496	Seraseq® FFPE Tumor Fusion RNA v4 Reference Material	Mut	Reference (cell line)	FFPE curls (10 – 15 µm), RNA extracted in this study	https://www.seracare.com/Seraseq-FFPE-Fusion-RNA-RM-v4-0710-0496/
3	0710-1031	Seraseq® FFPE NTRK Fusion RNA Reference Material	Mut	Reference (cell line)	FFPE curls (10 – 15 µm), RNA extracted in this study	https://www.seracare.com/Seraseq-FFPE-NTRK-Fusion-RNA-Reference-Material-0710-1031/
4	AM7832	Leukemia (K-562) Total RNA	Mut	Reference (cell line)	Purified total RNA (non-FFPE)	https://www.thermofisher.com/order/catalog/product/AM7832#/AM7832
5	AM7846	Breast Adenocarcinoma (MCF-7) Total RNA	Mut	Reference (cell line)	Purified total RNA (non-FFPE)	https://www.thermofisher.com/order/catalog/product/AM7846#/AM7846
6	Various (*n*=14)	Surgical biopsy, normal	WT	Tissue	FFPE curls (10 µm), RNA extracted in this study	Internal

Table 4. Breakdown of libraries prepared using the Plasma cfDNA and Tissue DNA Workflows

Panel	Input	Number of libraries prepared					
		Plasma cfDNA Workflow (*n*=144)		Tissue DNA Workflow (FFPE *n*=240, non-FFPE *n*=48)			
		Plasma	Cell line	HQ FFPE	MQ FFPE	LQ FFPE	Non-FFPE
KAPA HyperPETE Pan Cancer Panel	10 ng	16	8	9	14	9	24
	50 ng	16	8	9	14	9	–
	Q ng	N/A	N/A	9	14	9	N/A
	TOTAL	48		120			
KAPA HyperPETE Hot Spot Panel	10 ng	16	8	9	6	9	24
	50 ng	16	8	9	6	9	–
	Q ng	N/A	N/A	9	6	9	N/A
	TOTAL	48		96			
KAPA HyperPETE BRCA1 Panel	10 ng	16	8	–	–	–	–
	50 ng	16	8	–	–	–	–
	Q ng	N/A	N/A	27	18	27	N/A
	TOTAL	48		72			

HQ: high quality, MQ: mid quality, LQ: low quality (based on normalized Q scores); N/A: Not applicable.
Q ng: input mass calculated from normalized Q score, for HQ: 20 – 26 ng; MQ: 39 – 54 ng; LQ: 60 – 248 ng.

Table 5. Breakdown of libraries prepared using the Tissue RNA Workflow

Panel	Input	Number of libraries prepared			
		HQ FFPE	MQ FFPE	LQ FFPE	Non-FFPE
KAPA HyperPETE Lung Cancer Fusion Panel	10 ng	9	6	5	4
	50 ng	9	6	5	4
	TOTAL	48			

HQ: high quality, MQ: mid quality, LQ: low quality (based on DV_{200} values).

**Not available for sale in the United States. Contact your local Roche representative for availability in your region.*

enrichment with the KAPA HyperPETE Lung Cancer Fusion Panel, as outlined in Table 5. Libraries were prepared from 10 ng or 50 ng inputs. Reference materials (Table 3, lines 1 – 5) were processed in duplicate, whereas only one library per input was prepared from surgical samples (which represent biological replicates).

All RNA libraries were prepared with the KAPA RNA HyperPrep Kit (Roche PN: 08098093702 or 08098107702), KAPA Universal UMI Adapter (Roche PN: 09329862001 or 09329889001), KAPA UDI Primer Mixes 1 – 96 (Roche PN: 09134336001), and KAPA Pure Beads (included in the KAPA RNA HyperPrep Kit) as described in the KAPA HyperPETE Tissue RNA Fusion Transcript Workflow Instructions for use (Version 1.0 or later).[8]

Pre-capture Library QC: The concentration and fragment size distribution of amplified, pre-capture libraries were determined with an Agilent® 4200 TapeStation system and DNA High Sensitivity D1000 ScreenTape Assay (Agilent Technologies). All libraries were diluted 1:40 for analysis.

Target enrichment: Primer Extension Target Enrichment (PETE) was performed with the KAPA HyperPETE Reagent Kit (Roche PN: 09211624001 or 09211683001) and KAPA HyperCapture Bead Kit (Roche PN: 09075780001 or 09075798001), according to the standard single-plex protocol for each of the Plasma cfDNA, Tissue DNA and Tissue RNA Workflows.[4,7,8] The number of captures performed with each panel is outlined in Tables 4 and 5. Details of the four the KAPA HyperPETE Panels used in this study are summarized in Table 6.

Post-capture (PETE) Library QC: Sequencing-ready, post-capture (enriched) libraries were quantified and analyzed in the same manner as pre-capture libraries, but without any dilution.

Sequencing and Data Analysis

Sequencing: Libraries were pooled for multiplexed, paired-end Sequencing (2 x151 bp) on an Illumina® NextSeq™ 500, NextSeq 550Dx (RUO mode), or MiSeq® system. Pools were configured to meet the paired-end read requirement for each panel, as outlined in the Instructions for use for each workflow.[4,7,8]

Data analysis: Secondary data analysis was performed with an internal Roche secondary analysis pipeline. This included demultiplexing, adapter trimming, and alignment; as well as barcode and position deduplication, and the detection of single nucleotide variants (SNVs), short insertions and deletions (indels), fusions, microsatellite instability (MSI) and copy number variation (CNV). The performance of the KAPA HyperPETE technology across different sample types, inputs, and panels was evaluated by assessing (i) key sequencing QC metrics for each workflow, and (ii) variant calling results for characterized reference samples. Graphs were generated with ggplot2 (version 3.3.5) in R (version 3.4.4).[9]

Results and discussion

Library QC metrics

All of the pre-capture libraries prepared with each of the three workflows met the criteria for target enrichment, and all post-capture (enriched) libraries met the criteria for sequencing (Table 7).

Sequencing QC metrics

The internal Roche secondary analysis pipeline generates an extensive list of sample and analysis QC metrics for every run. A subset of metrics are reported for each of the three workflows. These are listed and defined in Table 8.

Table 6. KAPA HyperPETE Panels used in this study

Workflow	Panel name	Panel type	Capture size	Details	Paired-end (PE) reads required per library
Plasma cfDNA and Tissue DNA	KAPA HyperPETE Pan Cancer Panel	Catalog	301 kb	• 1321 capture targets, covering 86 cancer related genes and 190 MSI loci which have been found relevant to somatic oncology research applications. • Panel is optimized to deliver high uniformity and specificity from low quality, low quantity cfDNA and FFPE DNA samples.	• 110 M for Plasma cfDNA • 55 M for Tissue DNA
	KAPA HyperPETE Hot Spot Panel	Catalog	37 kb	• 166 capture targets across 37 genes from the genomic 'hot spot' regions that are frequently mutated in human cancer genes. • Panel is optimized to deliver high uniformity and specificity from low quality, low quantity cfDNA and FFPE DNA samples.	• 15 M for Plasma cfDNA • 7.5 M for Tissue DNA
	KAPA HyperPETE BRCA1 Panel	Custom	7.5 kb	• 23 capture targets across 23 exons of the BRCA1 gene.	• 3 M for Plasma cfDNA[a] • 1.5 M for Tissue DNA
Tissue RNA	KAPA HyperPETE Lung Cancer Fusion Panel	Catalog	18 kb	• Designed for use with the novel PETE technology to detect known/unknown fusions and fusion partners, including 17 lung cancer fusion genes (*ALK*, *AXL*, *BRAF*, *EGFR*, *FGFR1*, *FGFR2*, *FGFR3*, *MET*, *NRG1*, *NTRK1*, *NTRK2*, *NTRK3*, *PDGFRA*, *PDGFRB*, *PPARG*, *RET*, *ROS1*) and 4 housekeeping genes (*RAB7A*, *VCP*, *CHMP2A*, *HPRT1*) as positive controls.	• 1.8 M for Tissue RNA

[a]For custom panels, number of PE reads per library is calculated using the formulas: (400 x estimated coverage bp) for plasma cfDNA and (200 x estimated coverage bp) for tissue DNA.

Data on file with Roche.
For Research Use Only. Not for use in diagnostic procedures.

Table 7. Summary of library QC results

Workflow	Pre-capture libraries	Post-capture (PETE) libraries
Plasma cfDNA	• Concentration ≥35 ng/µL in the region of 150 – 1000 bp • Primary peak between 300 and 400 bp, with several minor peaks >400 bp	• Region molarity of ≥4 nM in the region of 150 – 1000 bp • Primary peak between 300 and 400 bp, with several minor peaks >400 bp
Tissue DNA	• Concentration ≥35 ng/µL in the region of 150 – 1000 bp • Average fragment size distribution between 250 and 350 bp	• Region molarity of ≥4 nM in the region of 150 – 1000 bp • Average fragment size distribution between 250 and 400 bp, depending on DNA quality
Tissue RNA	• Concentration ≥35 ng/µL in the region of 150 – 1000 bp • Average fragment size distribution between 250 and 350 bp, depending on input RNA quality	• Region molarity of ≥4 nM in the region of 150 – 1000 bp • Mean fragment size between 280 and 400 bp, depending on input RNA quality

Table 8. Sequencing QC metrics reported in this study

Metric	Explanation/definition	Workflow details
% reads on target	• Also referred to as on-target rate. • Defined as the percentage reads aligning to the capture panel's BED coordinate marked areas	• Reported for Plasma cfDNA and Tissue DNA Workflows • For Tissue RNA Workflow, on-target rates for entire panel, fusion genes in design and housekeeping genes reported separately
% bases in 2-fold range	• Measure of coverage uniformity • Calculated as the percentage of depth values that are bound within a range of: value ≥(median / factor) and value ≤(factor x median); where median is the median of the values and factor=2.	• Reported for Plasma cfDNA and Tissue DNA Workflows
Unique depth	• Median on-target depth after UMI (barcode) deduplication	• Reported for Plasma cfDNA and Tissue DNA Workflows
% Panel exon region ≥*d*X	• % of panel positions with a coverage depth of at least *d*	• Plasma cfDNA Workflow: *d*=1000X or 3000X • Tissue DNA Workflow: *d*=300X or 500X

Plasma cfDNA workflow: Results obtained with the three KAPA HyperPETE panels for libraries prepared from different inputs (10 ng and 50 ng) and sample types (plasma and cell line DNA), are shown in Figure 3 on p. 8. On-target rates were high (>70%) and highly reproducible across inputs and sample types for the Pan Cancer (301 kb) and Hot Spot (37 kb) panels. As expected for panels with a capture size <30 kb, on-target rates were lower and more variable for the 7.5 kb BRCA1 panel.

Coverage uniformity (% bases in 2-fold range) was high and consistent across inputs and sample types, irrespective of panel size. This confirmed that the KAPA HyperPETE design algorithm and novel enrichment strategy support high coverage uniformity from challenging samples. Coverage uniformity was slightly lower for libraries prepared from higher (50 ng) inputs due to intrinsically lower duplication rates.

Coverage of exonic regions at a depth ≥1000X was also high and highly reproducible across panels, inputs, and sample types. At a depth ≥3000X, coverage of exonic regions dropped significantly for 10 ng inputs, in accordance with the limited number of genome equivalents (<3000) in these libraries. Fewer genome equivalents in 10 ng samples also impacted unique coverage depth. When input is sufficient (i.e., for 50 ng libraries), unique depth is impacted by on-target rate, and was therefore lower for the small BRCA1 panel.

Together, these results demonstrate that the KAPA HyperPETE Plasma cfDNA Workflow yields high and uniform sequencing coverage across designs for both cell line and plasma samples, thereby supporting high-confidence variant calling (see below) from the recommended amount of paired-end reads for each panel (ref. Table 6). Panels <30 kb are expected to have lower on-target rates and, concomitantly, lower unique coverage depth. Since the number of genome equivalents in a library may limit performance, the highest available input of cfDNA should always be used to ensure maximum coverage and variant calling sensitivity from the amount of sequencing data generated per library.

Tissue FFPE DNA Workflow: When working with FFPE samples, it is critical to understand that spectrophotometric or fluorometric DNA quantification methods do not provide a reliable indicator of the amount of utilizable material (or genome equivalents) in a DNA preparation. Libraries prepared from the same input amount of FFPE DNA preparations of variable quality are therefore not expected to perform

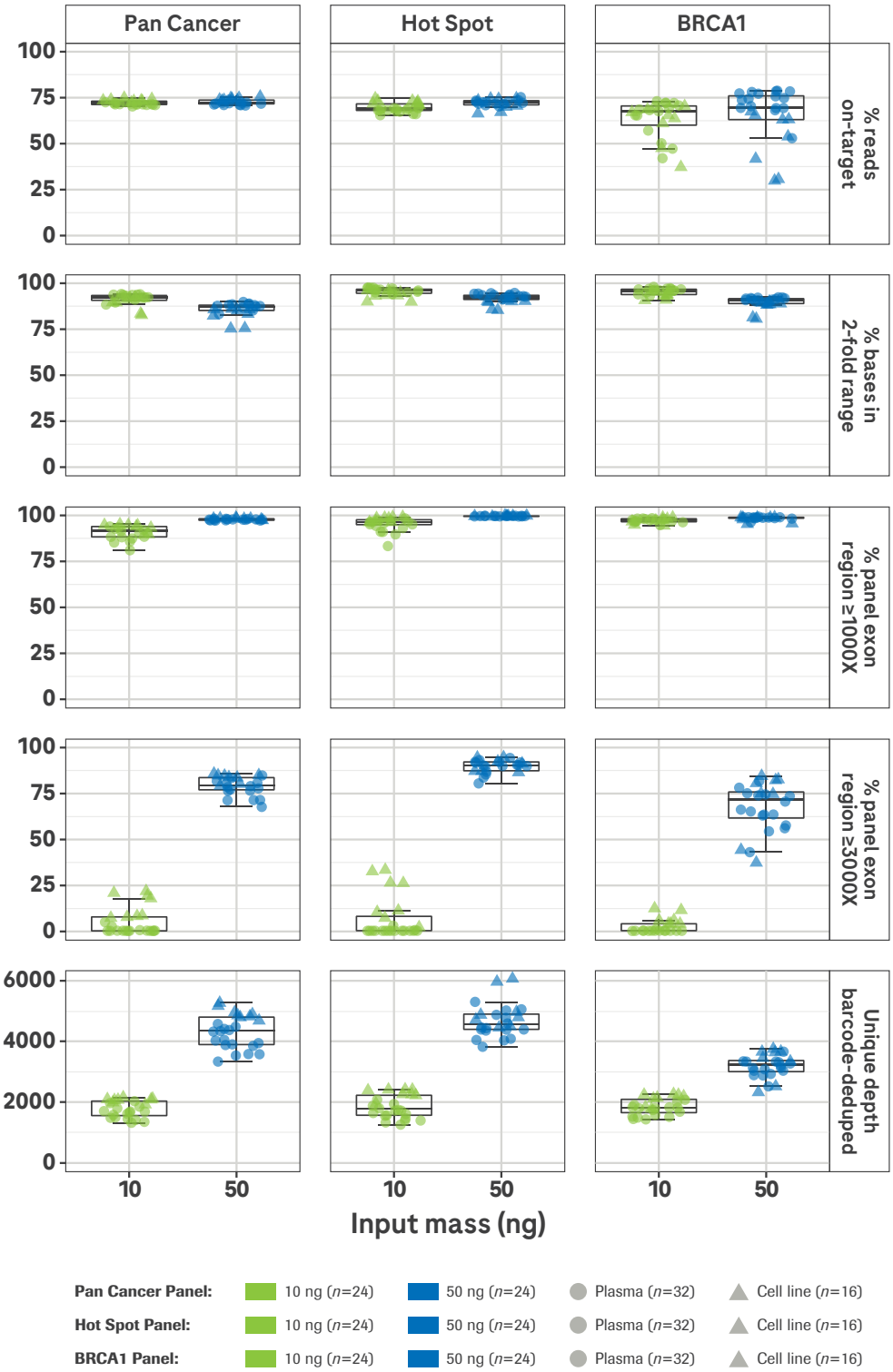

Figure 3. Key sequencing QC metrics for the Plasma cfDNA Workflow. DNA was extracted from plasma (dots) or cell line (triangles) samples and libraries were prepared, enriched with the three different panels, and sequenced as described in Materials and Methods. Data were analyzed and QC metrics were generated with an internal Roche secondary analysis pipeline. The number of libraries prepared and analyzed from each input/sample type with each panel is summarized in the legend. Refer to Table 4 for full details. Cell line samples were processed in duplicate. Plasma cfDNA samples represent biological replicates (one library per sample/input).

Figure 4. Key sequencing QC metrics for the Tissue DNA Workflow. FFPE DNA was obtained or extracted from tissue (dots) and cell line (triangles) samples and DNA quality was assessed with a qPCR-based method as described in Materials and Methods. Samples were divided into high, mid, and low quality based on normalized Q scores. Libraries were prepared, enriched with the different panels, and sequenced as described. Q is the amount of input DNA calculated from the normalized Q score, and ranged from 20 – 26 ng for high-quality DNA, from 39 – 54 ng for mid-quality DNA, and from 60 – 248 ng for low-quality DNA. Data were analyzed and QC metrics were generated with an internal Roche secondary analysis pipeline. The number of libraries prepared and analyzed from each input/sample type with each panel is summarized in the legend. Refer to Table 4 for full details. Reference materials (FFPE and non-FFPE cell line samples, and xenograft reference samples analyzed for MSI/MSS) were processed in duplicate. Biopsy samples represent biological replicates (one library per sample/input).

similarly. The qPCR-based KAPA NGS FFPE DNA QC Kit used in this study (recommended as part of the standard KAPA HyperPETE Tissue DNA Workflow) provides a means for adjusting library input according to DNA quality. FFPE DNA extracts were divided into three groups based on the normalized Q score determined with this kit: 0.614 – 1.000 (theoretically: ≥0.5) for high quality, 0.227 – 0.340 (theoretically: 0.22 ≤Q <0.5) for mid quality, and 0.042 – 0.200 (theoretically: 0.04 ≤Q <0.22) for low quality samples. Q scores were not experimentally determined for non-FFPE (cell-line) DNA extracts, but these were grouped with high quality samples (expected Q score ≥0.5) based on prior experience.

Similarly to the Plasma cfDNA Workflow, results obtained with the Tissue DNA Workflow (summarized in Figure 4) confirmed that the KAPA HyperPETE technology supports:

- high (>60% for all samples in this study; typically >65%) and reproducible on-target rates, irrespective DNA input and quality, across a wide (~8-fold) range of panel capture target sizes. Lower and more variable on-target rates are only expected for very small (<30 kb) panels.
- high (>70% bases in 2-fold range for all samples in this study; typically >85%) and consistent coverage uniformity across all inputs, sample types, and panel sizes/designs—even for low quality FFPE samples.
- extremely high (~100%) and reproducible coverage of exonic regions, provided that the number of genome equivalents in a library is not limiting. For HQ FFPE DNA samples, input amount (10 ng, 50 ng or Q ng=20 – 26 ng) had no impact on coverage ≥500X. However, as DNA quality decreased, coverage for limiting (10 ng) DNA inputs dropped significantly (≤25% for LQ samples) and became highly variable—reflecting the overall decrease in genome equivalents (and concomitant increase in sample-to-sample variability). Coverage remained very high for libraries prepared from a sufficient input (50 ng and quality-adjusted Q ng=39 – 248 ng) of mid- and low-quality FFPE DNA.

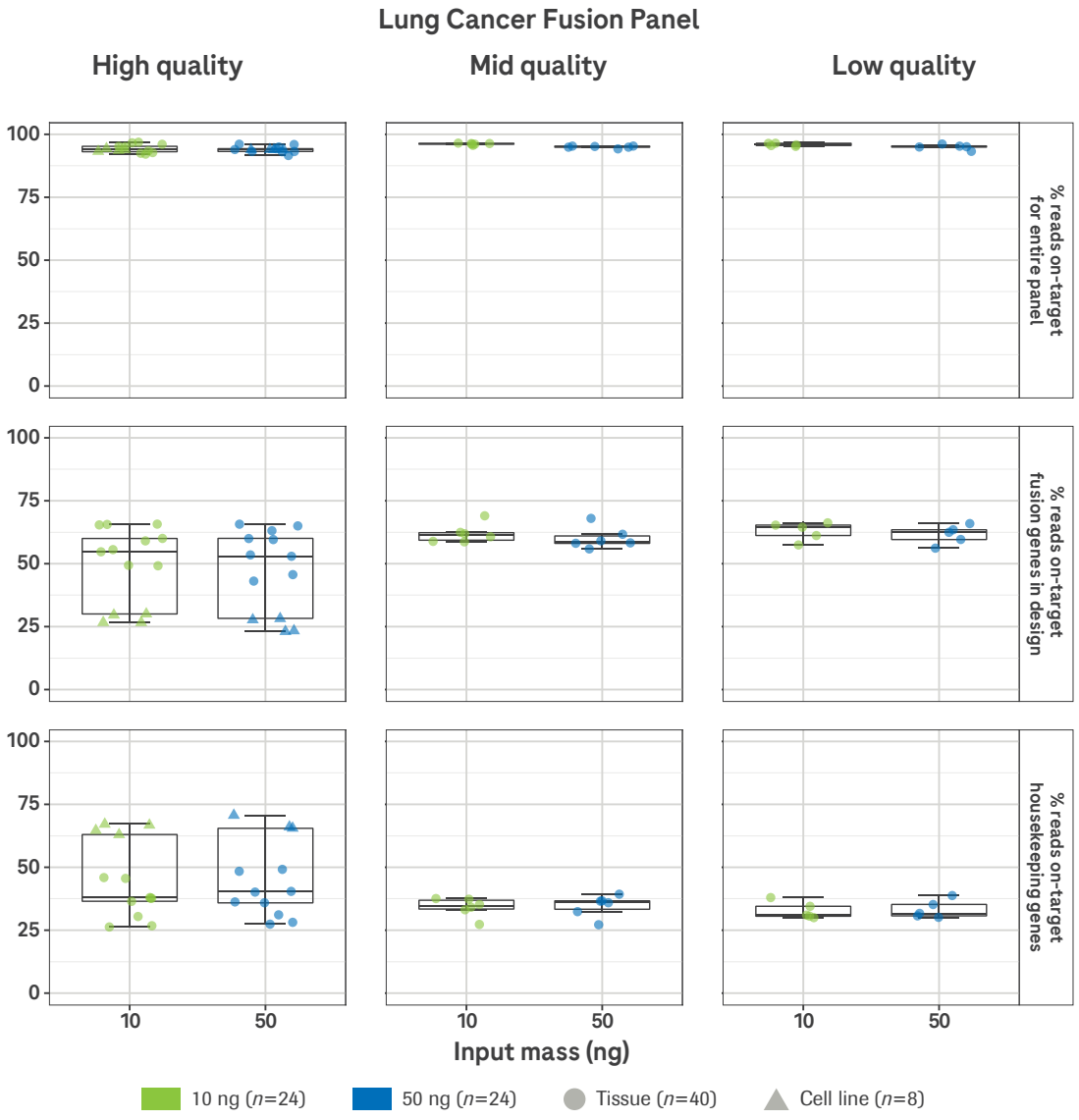

Figure 5. Key sequencing QC metrics for the Tissue RNA Workflow. RNA was extracted from tissue (dots) or cell line (triangles) samples and RNA quality was assessed with the Bioanalyzer DV_{200} method as described in Materials and Methods. FFPE samples were divided into high, mid, and low quality based on the percentage of fragments >200 nt. Libraries were prepared, enriched with the KAPA HyperPETE Lung Cancer Fusion Panel, and sequenced as described. Data were analyzed and QC metrics were generated with an internal Roche secondary analysis pipeline. The number of libraries prepared and analyzed from each input/sample type with each panel is summarized in the legend. Refer to Table 5 for full details. Reference materials (FFPE and non-FFPE cell line samples) were processed in duplicate. Normal/adjacent surgical biopsies represent biological replicates (one library per sample/input).

Table 9. Somatic variant calling performance summary

Somatic variant type	FFPE/cell line		cfDNA	
	True positive rate (%)	False positive rate (%)	True positive rate (%)	False positive rate (%)
SNV	99.5	<1 per 10 kb	99.1	<1 per 10 kb
Indel	100	0.2 per sample	97.9	0.6 per sample
CNV	100	0 per sample	N/A	
MSI	100	0 per sample		
RNA fusion	100	0.2 per sample		

As for the Plasma cfDNA Workflow, unique (barcode deduplicated) coverage depth correlated with input mass (genome equivalents). The nanogram value of the Q input varied depending on the quality of extracted FFPE DNA, and was <50 ng for all HQ extracts, close to 50 ng for MQ extracts, and >50 ng and highly variable (up to 248 ng) for LQ samples. This underlines the importance of using quality-adjusted DNA input (and the highest available inputs) to ensure successful library construction, enrichment, and sequencing.

Tissue FFPE RNA Workflow: FFPE RNA extracts were also divided into three quality groups, based on the degree of RNA fragmentation (DV_{200} score): high-quality extracts had a $DV_{200} \geq 70\%$; $50\% \leq DV_{200} < 70\%$ for mid-quality extracts, and $30\% \leq DV_{200} < 50\%$ for low-quality extracts. Twenty-four libraries were prepared from each input (10 or 50 ng), and all libraries were enriched with the relatively small (18 kb) KAPA HyperPETE Lung Cancer Fusion Panel. Reads mapping to ribosomal RNA (rRNA) ranged between 0.84% and 11.2% of total reads and were removed bioinformatically before calculating on-target rates. Results are shown in Figure 5.

On-target rates for the entire panel were very high (>90%) and comparable across inputs, irrespective of RNA quality. On-target rates for fusion genes and housekeeping genes (which are included in the Workflow and analysis as a positive enrichment control) are mathematically related: the on-target rate for housekeeping genes = (on-target rate for entire panel) minus (on-target rate for fusion genes). On-target rates for both sets of genes were highly consistent across inputs for mid- and low-quality RNA extracts. On-target rates for high-quality RNA extracts appeared to be highly variable. This was attributed to intrinsic differences between the expression levels of target genes between RNA extracted from FFPE tissue (surgical biopsies; circles) and non-FFPE cell line controls (triangles).

Variant calling results

Characterized reference materials with validated mutations and allele frequencies were included in all workflows in order to assess variant calling performance (true and false positive rates) for five different somatic variant classes:

- *single nucleotide variants (SNVs):* assessed for the Plasma cfDNA and Tissue DNA Workflows;
- *small insertions and deletions (indels):* assessed for the Plasma cfDNA and Tissue DNA Workflows;
- *copy number variants (CNVs):* assessed for the Tissue DNA Workflow;
- *microsatellite instability (MSI):* assessed for the Tissue DNA Workflow; and
- *RNA fusions:* assessed for the Tissue RNA Workflow.

Detailed results for each workflow are given in the Appendix. A summary of all results is shown in Table 9.

All variant classes were detected with a very high degree of accuracy and precision for all sample types and workflows. When reviewing the detailed results for individual workflows it is important to note the following:

- A reference sample with a verified allele frequency (AF) of 0.125% (for 40 clinically-relevant mutations across 28 genes) was included in the Plasma cfDNA Workflow, despite the fact that the standard protocol is not optimized for such low allele frequencies. This resulted in much lower true positive (TP) rates than for the same cfDNA mix with an AF of 0.5% or 1%. True positive rates (and sensitivity) may be increased by increasing the sequencing depth (amount of reads) for low AF samples. Results for the AF 0.125% sample were not taken into account in calculations for Table 9.
- For both the Plasma cfDNA and Tissue DNA Workflows, false positive (FP) SNV counts were higher for the KAPA HyperPETE Pan Cancer Panel than for the Hot Spot Panel. This is attributed to the ~8-fold larger capture size of the Pan Cancer Panel (301 kb vs. 37 kb for the Hot Spot Panel).
- For the Tissue DNA Workflows, adjacent normal tissues were used as negative control samples. None of the false positives detected with the internal Roche secondary analysis pipeline have been validated using an orthogonal or secondary assay (e.g., ddPCR, qPCR, or Sanger sequencing). Putative false positives may therefore

represent real biological signals arising from potential tumor content contamination.

- In the Tissue RNA Workflow, the *EGFR-SEPT14* variant in the Seraseq® FFPE Tumor Fusion RNA v4 Reference Material was manually curated as the fusion caller in the internal Roche secondary analysis pipeline identified a different *EGFR* partner that has a homologous sequence to *SEPT14*.

Conclusions

The novel KAPA HyperPETE technology from Roche enables a new class of end-to-end targeted sequencing solutions for high-performance somatic variant analysis using small (≤300 kb catalog and ≤250 kb custom) panels. KAPA HyperPETE Workflows combine deep content expertise and innovative panel design with industry-leading library preparation and target enrichment reagents to achieve the high on-target rates, coverage depth, and uniformity needed to support economical, sensitive, accurate, and reproducible analysis of all major variant classes (including MSI and RNA fusions) from the sample types commonly used in oncology research.

Key features of KAPA HyperPETE Workflows for Plasma cfDNA, and Tissue DNA/RNA highlighted in this study include:

- Comparable performance to gold-standard hybrid capture workflows with a convenient, single-day protocol.
- Flexibility to process a wide range of sample types and inputs with high efficiency, using optimized catalog or easy-to-order custom panels.
- Sample preparation enhancements, including DNA Polishing, quality adjustment of input amounts, unique molecular identifiers, and unique dual indexing, to achieve the best possible results from the most challenging and precious samples.

In addition, KAPA HyperPETE Workflows:

- employ the existing user-friendly HyperDesign Tool for custom panel design. The tool (which was originally developed to support custom probe design for the hybridization-based KAPA HyperCap Workflows) has been updated to include a new design/selection algorithm for KAPA HyperPETE primer panels.
- are automation-friendly, and compatible with all liquid handlers commonly used in NGS sample preparation pipelines.
- are supported from end-to-end with fully integrated service and technical support from a single vendor.

Appendix

Detailed variant calling results by Workflow

Table A1. SNV calling performance (true positive rate) for the Plasma cfDNA Workflow

Panel	Sample	Expected allele frequency (AF)	Input (ng)	No. of samples	No. of expected variants	True positives	False negatives	True positive rate (%)
KAPA HyperPETE Pan Cancer Panel	Seraseq® ctDNA Mutation Mix v2	1%	10	2	46	46	0	99.4
			50	2	46	46	0	
		0.5%	10	2	44	44	0	
			50	2	44	43	1[b]	
		0.125%[a]	10	2	46	25	21	62.0
			50	2	46	32	14	
KAPA HyperPETE Hot Spot Panel	Seraseq ctDNA Mutation Mix v2	1%	10	2	40	40	0	98.7
			50	2	40	40	0	
		0.5%	10	2	38	36	2[c]	
			50	2	38	38	0	
		0.125%[a]	10	2	40	17	23	55.0
			50	2	40	27	13	

[a]Reference sample with an expected AF of 0.125% was included in the experiment for demonstration purposes only. The standard Plasma cfDNA Workflow is not optimized for this AF level. Low true positive rates are expected to improve with higher sequencing depth (more reads per sample).

The following missing variants (false negatives) were the result of low support based on the heuristic rules in the SNV caller:
[b]*GNAQ*, Q209P, AF=0.11%, *ALT*=4, LOW_SUPPORT
[c]*AKT1*, E17K, *ALTDP*=5, *ALTDP2*=0, AF=0.23%, LOW_SUPPORT; *PIK3CA*, E545K, *ALTDP*=3, *ALTDP2*=0;AF=0.2%, LOW_SUPPORT

Data on file with Roche.
For Research Use Only. Not for use in diagnostic procedures.

Table A2. SNV calling performance (false positive rate) for the Plasma cfDNA Workflow

Panel	Input (ng)	No. of samples	False positives	Total SNV false positives*	False positives per base
KAPA HyperPETE Pan Cancer Panel	10	16	33	82	<1 per 10 kb (0.00002)
	50	16	49		
KAPA HyperPETE Hot Spot Panel	10	16	10	42	
	50	16	32		

*Because no AF cutoff was applied in plasma SNV calling, most of the false positives in plasma were with AF <0.5% and while not presented in databases, these could be Clonal hematopoiesis of indeterminate potential (CHIP) variants.

Table A3. Indel calling performance (true positive rate) for the Plasma cfDNA Workflow

Panel	Sample	Expected allele frequency (AF)	Input (ng)	No. of samples	No. of expected variants	True positives	False negatives	True positive rate (%)
KAPA HyperPETE Pan Cancer Panel	Seraseq® ctDNA Mutation Mix v2	1%	10	2	24	24	0	97.7
			50	2	24	24	0	
		0.5%	10	2	20	18	2[b]	
			50	2	20	20	0	
		0.125%[a]	10	2	26	6	20	38.5
			50	2	26	14	12	
KAPA HyperPETE Hot Spot Panel	Seraseq ctDNA Mutation Mix v2	1%	10	2	16	16	0	98.2
			50	2	16	16	0	
		0.5%	10	2	12	11	1[c]	
			50	2	12	12	0	
		0.125%[a]	10	2	16	6	10	43.8
			50	2	16	8	8	

[a]Reference sample with an expected AF of 0.125% was included in the experiment for demonstration purposes only. The standard Plasma cfDNA Workflow is not optimized for this AF level. Low true positive rates are expected to improve with higher sequencing depth (more reads per sample).

The following missing variants (false negatives) were the result of low support based on the heuristic rules in the Indel caller:
[b]LOW_SUPPORT: *ATM*, C353fs*5, *ALT*=5, AF=0.32%; *ERBB2*, A775_G776insYVMA, *ALT*=2, AF=0.06%
[c]*ERBB2*, A775_G776insYVMA, *ALTDP*=3;AF=0.09%, LOW_SUPPORT

Table A4. Indel calling performance (false positive rate) for the Plasma cfDNA Workflow

Panel	Input (ng)	No. of samples	False positives	False positives per sample	False positives per sample (for panel)	False positives per sample (overall)
KAPA HyperPETE Pan Cancer Panel	10	16	6	0.38	1.03	0.6
	50	16	27	1.69		
KAPA HyperPETE Hot Spot Panel	10	16	0	0	0.13	
	50	16	4	0.25		

Data on file with Roche.
For Research Use Only. Not for use in diagnostic procedures.

Table B1. SNV calling performance (true positive rate) for the Tissue DNA Workflow

Panel	Sample	Expected allele frequency (AF)	Input (ng)	No. of samples	No. of expected variants	True positives	False negatives	True positive rate (%)
KAPA HyperPETE Pan Cancer Panel	Custom gDNA Mutation Mix (RSS, Sample 1)	5%	10	2	38	38	0	100
	HD753	5%	10	2	4	4	0	
	HD789	5%	10	2	4	12	0	
			50	2	4	4	0	
			Q	2	4	4	0	
	HD200	5%	10	2	16	16	0	
			50	2	16	16	0	
			Q	2	16	16	0	
KAPA HyperPETE Hot Spot Panel	Custom gDNA Mutation Mix (RSS, Sample 1)	5%	10	2	34	33	1[a]	99.0
	HD753	5%	10	2	4	4	0	
	HD789	5%	10	2	4	4	0	
			50	2	4	4	0	
			Q	2	4	4	0	
	HD200	5%	10	2	16	16	0	
			50	2	16	16	0	
			Q	2	16	16	0	

The following missing variant (false negative) was the result of AF lower than the default setting of 2% or low support based on the heuristic rules in the SNV caller:
[a]*AKT1*, E17K, AF=1.69%, LOW_AF

Table B2. SNV calling performance (false positive rate)[a] for the Tissue DNA Workflow

Panel	Input (ng)	No. of samples	SNV false positives	Total SNV false positives	False positives per base
KAPA HyperPETE Pan Cancer Panel	10	13	6	16	<1 per 10 kb (0.000001)
	50	16	6		
	Q	16	4		
KAPA HyperPETE Hot Spot Panel	10	16	1	2	
	50	16	0		
	Q	16	1		

[a]Adjacent normal tissues were used as negative control samples. None of the false positives detected with the internal Roche secondary analysis pipeline have been validated using an orthogonal or secondary assay (e.g., ddPCR, qPCR, or Sanger sequencing). Putative false positives may therefore represent real biological signals arising from potential tumor content contamination.

Data on file with Roche.
For Research Use Only. Not for use in diagnostic procedures.

Table B3. Indel calling performance (true positive rate) for the Tissue DNA Workflow

Panel	Sample	Expected allele frequency (AF)	Input (ng)	No. of samples	No. of expected variants	True positives	False negatives	True positive rate (%)
KAPA HyperPETE Pan Cancer Panel	Custom gDNA Mutation Mix (RSS, Sample 1)	5%	10	2	6	6	0	100
	Custom gDNA Mutation Mix (RSS, Sample 3)	5%	10	2	16	16	0	
	Custom gDNA Mutation Mix (RSS, Sample 4)	5%	10	2	16	16	0	
	HD753	5%	10	2	4	4	0	
	HD789	5%	10	2	4	4	0	
			50	2	4	4	0	
			Q	2	4	4	0	
	HD200	5%	10	2	2	2	0	
			50	2	2	2	0	
			Q	2	2	2	0	
KAPA HyperPETE Hot Spot Panel	Custom gDNA Mutation Mix (RSS, Sample 1)	5%	10	2	4	4	0	100
	Custom gDNA Mutation Mix (RSS, Sample 3)	5%	10	2	8	8	0	
	Custom gDNA Mutation Mix (RSS, Sample 4)	5%	10	2	12	12	0	
	HD753	5%	10	2	4	4	0	
	HD789	5%	10	2	4	4	0	
			50	2	4	4	0	
			Q	2	4	4	0	
	HD200	5%	10	2	2	2	0	
			50	2	2	2	0	
			Q	2	2	2	0	

Table B4. Indel calling performance (false positive rate)[a] for the Tissue DNA Workflow

Panel	Input (ng)	No. of samples	False positives	False positives per sample	False positives per sample (for panel)	False positives per sample (overall)
KAPA HyperPETE Pan Cancer Panel	10	13	3	0.23	0.36	0.2
	50	16	5	0.31		
	Q	16	8	0.50		
KAPA HyperPETE Hot Spot Panel	10	16	2	0.13	0.13	
	50	16	2	0.13		
	Q	16	2	0.13		

[a]Adjacent normal tissues were used as negative control samples. None of the false positives detected with the internal Roche secondary analysis pipeline have been validated using an orthogonal or secondary assay (e.g., ddPCR, qPCR, or Sanger sequencing). Putative false positives may therefore represent real biological signals arising from potential tumor content contamination.

Table B5. CNV calling performance (true positive rate) for the Tissue DNA Workflow

Panel	Sample	Genes	Expected copies	Input (ng)	No. of samples	No. of expected variants	True positives	False negatives	True positive rate (%)
KAPA HyperPETE Pan Cancer Panel	0710-0412	*ERBB2, FGFR3, MYC*	~6	10	2	4	4	0	100
	0710-1286	*ERBB2*	4.7	10	2	2	2	0	
	HD753	*MET*	4.5	10	2	2	2	0	
	HD789	*MET*	4.5	10	2	2	2*	0	
				50	2	2	2	0	
				Q	2	2	2*	0	

*Low confidence CNV calls

Table B6. CNV calling performance (false positive rate) for Tissue DNA Workflow

Panel	Input (ng)	No. of samples	CNV False positives	Total CNV False positives	False positives per sample
KAPA HyperPETE Pan Cancer Panel	10	13	0	0	0
	50	16	0		
	Q	16	0		

Table B7. MSI calling performance (true positive rate) for the Tissue DNA Workflow

Tissue type	Expected status	Sample	No. of samples	Input (ng)	Reported status	True positive wrate (%)
FFPE	MSI	DU145	2	10	MSI	100
			2	50	MSI	
			2	Q	MSI	
	MSI	SW48	2	10	MSI	
			2	50	MSI	
			2	Q	MSI	
Non-FFPE	MSI	DLD-1	2	10	MSI	
	MSI	MOLT-4	2	10	MSI	
	MSI	SW48	2	10	MSI	

Table B8. MSI calling performance (false positive rate) for the Tissue DNA Workflow

Tissue type	Expected status	Sample	No. of samples	Input (ng)	Reported status	False positive rate (%)
FFPE	MSS	BT474	2	10	MSS	0
			2	50	MSS	
			2	Q	MSS	
	MSS	MDA-MB-453	2	10	MSS	
			2	50	MSS	
			2	Q	MSS	
Non-FFPE	MSS	A-172	2	10	MSS	
	MSS	A549	2	10	MSS	
	MSS	AU565	2	10	MSS	

Data on file with Roche.
For Research Use Only. Not for use in diagnostic procedures.

Table C1. Fusion calling performance (true positive rate) for the Tissue RNA Workflow

Panel	Sample	Input (ng)	No. of samples	No. of expected fusions	True positives	False negatives	True positive rate (%)
All reads (7 M – 34 M total reads)							
KAPA HyperPETE Lung Cancer Fusion Panel	Seraseq® FFPE Tumor Fusion RNA v4 Reference Material*	10	2	30	30	0	100
		50	2	30	30	0	
	Seraseq FFPE NTRK Fusion RNA Reference Material	10	2	24	24	0	
		50	2	24	24	0	
	HD784	10	2	6	6	0	
		50	2	6	6	0	
100 reads per base (1.8 M total reads)							
KAPA HyperPETE Lung Cancer Fusion Panel	Seraseq FFPE Tumor Fusion RNA v4 Reference Material*	10	2	30	30	0	100
		50	2	30	30	0	
	Seraseq FFPE NTRK Fusion RNA Reference Material	10	2	24	24	0	
		50	2	24	24	0	
	HD784	10	2	6	6	0	
		50	2	6	6	0	

*The *EGFR-SEPT14* variant was manually curated as the fusion caller in the internal Roche secondary analysis pipeline identified a different *EGFR* partner that has a homologous sequence to *SEPT14*.

Table C2. Fusion calling performance (false positive rate) for the Tissue RNA Workflow

Panel	Input (ng)	No. of samples	False positives	False positives per sample (per input)	False positives per sample (overall)
All reads (7 M – 34 M total reads)					
KAPA HyperPETE Lung Cancer Fusion Panel	10	14	2	0.14	0.2
	50	14	4	0.29	
100 reads per base (1.8 M total reads)					
KAPA HyperPETE Lung Cancer Fusion Panel	10	14	2	0.14	0.2
	50	14	4	0.29	

References

1. National Cancer Institute. The Genetics of Cancer. https://www.cancer.gov/about-cancer/causes-prevention/genetics. Accessed July 2024.
2. Bewicke-Copley F, Arjun Kumar E, Palladino G, et al. Applications and analysis of targeted genomic sequencing in cancer studies. Comput. Struct. Biotechnol. J. 2019;17(1348–1359). doi: 10.1016/j.csbj.2019.10.004.
3. Roche Sequencing Solutions. KAPA HyperPETE Brochure.
4. Roche Sequencing Solutions. KAPA HyperPETE Somatic Tissue DNA Workflow; Instructions for use, Version 1.0 or later.
5. Roche Diagnostics GmbH. High Pure FFPET RNA Isolation Kit Instructions for Use, Version 06 (Content version: October 2020).
6. Agilent Technologies. Simplified DV200 Evaluation with the Agilent 2100 Bioanalyzer System. https://www.agilent.com/cs/library/technicaloverviews/public/5991-8287EN.pdf. Accessed July 2024.
7. Roche Sequencing Solutions. KAPA HyperPETE Somatic Plasma cfDNA Workflow; Instructions for use, Version 1.0 or later.
8. Roche Sequencing Solutions. KAPA HyperPETE Tissue RNA Fusion Transcript Workflow; Instructions for use, Version 1.0 or later.
9. https://github.com/tidyverse/ggplot2. Accessed July 2024.

Acknowledgments

The authors wish to thank Liesl Joubert, Marius Snyman, Etienne Slabbert, and Jo-Anne Penkler (Roche Molecular Labs, Cape Town) and Seng Saelee (Roche Molecular Labs, Pleasanton), for their contributions to important precursor studies which enabled the work presented in this project.

Published by:

Roche Sequencing and Life Science
9115 Hague Road
Indianapolis, IN 46256

sequencing.roche.com

MC-US-15954 A841 11/2024