White Paper

Longitudinal detection of non-Hodgkin lymphoma ctDNA

Authors

Clara Bermejo
Principal Scientist II – Assay Development

Parul Agarwal
Scientist II – Assay Development

Richard Chien
Manager – Assay Bioinformatics

Bowdoin Su
Medical Affairs Leader – Medical Affairs

Alexandra Markus
Medical Affairs Leader – Medical Affairs

Vlada Melnikova
Clinical Development Lead – Clinical Development

Sunandini Chopra
Senior International Product Manager

Roche Molecular Labs
Pleasanton, CA, USA

The KAPA HyperCap Design Share NHL Panel enables highly sensitive, longitudinal detection of non-Hodgkin lymphoma circulating tumor DNA

The KAPA HyperCap Design Share NHL Panel is a research solution that covers SNVs in coding and/or untranslated regions of 383 genes, plus additional intergenic regions for a total capture size of 341 Kb. These genomic regions are enriched in genomic alterations associated with NHL. Used in combination with the KAPA HyperCap Workflow and open-source KAPA bioinformatics analysis for longitudinal detection of ctDNA, it offers a robust, streamlined, and fully integrated solution for highly sensitive detection and longitudinal study of NHL-associated SNVs in blood-derived samples.

Introduction

Non-Hodgkin lymphoma (NHL) is one of the most common hematological malignancies, estimated to have caused 544,000 new cases and 260,000 deaths worldwide in 2020.[1,2] Tumors are genetically diverse, with associated disease ranging from indolent to aggressive, curable to refractory. Global research efforts are not only focused on the introduction of new therapies (including stem cell transplants, chemotherapy, immuno-therapy, targeted therapies, and vaccines), but also on the development of genetic testing methods to improve disease detection and management.

Current diagnostic methods for lymphoid neoplasms include histopathology, flow cytometry, cytogenetics, immunohistochemistry, and molecular techniques. Next-generation sequencing (NGS)-based genomic profiling and gene expression analysis are playing an increasingly important role in accurate tumor classification, as this drives therapy selection.[3] Molecular monitoring during and after treatment is also critical, as the radiological and nuclear imaging methods typically used to assess treatment response are unable to provide information on clonal evolution and minimal residual disease (MRD)—both of which impact final outcomes.[4] Cell-free circulating tumor DNA (cfDNA/ctDNA), typically assessed using targeted deep sequencing, has emerged as an important non-invasive and highly sensitive biomarker in the monitoring of patient status.[5,6]

The Roche Design Share platform offers NGS target enrichment panels designed and developed by Roche in collaboration with leading researchers from around the world. In 2023, Roche released a non-Hodgkin lymphoma-focused Design Share panel suitable for the detection of both somatic and germline variants in blood and tissue samples. This white paper demonstrates the use of the KAPA HyperCap Design Share NHL Panel, KAPA HyperCap Workflow, and open source bioinformatic tools* for the detection of single nucleotide variants (SNVs) at various variant allele frequencies (AF, 0 – 5%) in commercially obtained cfDNA and genomic DNA (gDNA) samples. The use of this integrated solution for the longitudinal detection of NHL-associated variants is also described.

*Bioinformatic tools are described in detail in an accompanying white paper entitled *KAPA HyperCap Design Share bioinformatics analysis: Longitudinal detection of non-Hodgkin lymphoma circulating tumor DNA*.[7]

Materials and methods

Experimental design

The KAPA HyperCap Design Share NHL Panel (IRN: 1000028225) covers single nucleotide variants (SNVs) located in the coding and/or untranslated regions of 383 genes (listed in Table A.1 in the Appendix) previously identified in NHL patients, particularly those diagnosed with diffuse large B cell lymphoma (DLBCL).[5] The panel, which also contains genes associated with other B cell lymphomas, enables longitudinal detection of variants associated with NHL, as illustrated in Figure 1.

In this study, commercial reference samples (mixes of purified cfDNA and gDNA) were used to prepare "contrived" samples—with known variants at AF ranging from 0% (wild type, WT) to 5%—that mimic NHL samples. NGS libraries were prepared using the KAPA HyperPrep Kit and KAPA HyperCap cfDNA Workflow v1.1 ("plasma cfDNA workflow") or the KAPA HyperPlus Kit (with enzymatic fragmentation) and KAPA Hypercap Workflow v3.4 ("germline workflow"). Libraries were enriched by hybridization to the KAPA HyperCap Design Share NHL Panel. Sequencing was performed on an Illumina® NextSeq™ 500/550 instrument using standard protocols. Data analysis was performed using open source bioinformatic tools.[7] Performance of the library construction/enrichment workflows for cfDNA and gDNA samples was assessed via ten key sequencing metrics and variant calling results for known SNVs. Data generated from contrived samples were subsequently processed using the three-stage KAPA bioinformatics analysis for longitudinal detection of ctDNA to demonstrate the use of the KAPA HyperCap Design Share NHL Panel for longitudinal analysis of NHL-associated variants in circulating tumor DNA.

Samples

Reference materials (purified cell line cfDNA or cfDNA mixes; Table 1, lines 1 – 5) and plasma samples from healthy donors (Table 1, line 8) for the plasma cfDNA workflow were obtained from commercial suppliers. For the germline workflow, gDNA from two characterized B-lymphocyte cell lines (NA24631 and NA24149; Table 1, lines 6 – 7) were purchased from the Coriell Institute for Medical Research. DNA preparations were mixed in a ratio of 98:2 to generate a contrived sample with known SNVs with an AF of 1% (see Table A.4 in the Appendix for an expected variant list).

DNA Extraction and QC

Plasma cfDNA workflow: For the KAPA HyperCap cfDNA Workflow, any appropriate method may be used to extract cfDNA from blood collected in EDTA-containing collection tubes and handled according to standard procedures for plasma samples. For this study, the cobas® cfDNA Sample Preparation Kit (Roche PN:07247737190) was used starting from 10 mL of plasma and minor IFU modifications, as described in Table A.2. Extracted cfDNA was quantified using a fluorescence microplate reader and the Quant-iT™ dsDNA High-Sensitivity Assay Kit (Thermo Fisher Scientific), although a Qubit® Fluorometer and Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific) or any other equivalent method may be used.

Assessment of cfDNA quality with a qPCR-based method is highly recommended. In this study this was performed using primers used in Saelee SL, et al., 2022[8] (forward primers 66F: 5'-TTGCGGAAGTCAGTGTGG-3' and 330F: 5'-CAAACAACCCCATCAAAAAGTG-3' in combination with a single reverse primer, 5'-GATGGCTGGGTCAAATGGTA-3'), and with reagents from KAPA NGS FFPE DNA QC Kit (Roche PN: 09217193001 or 09217207001), as described in Table A.3.

Germline workflow: The KAPA NGS DNA Extraction Kit (Roche PN: 09189823001 or 09190023001 is recommended for the extraction of gDNA from buffy coat or plasma-depleted blood. DNA may be quantified using any of the methods and instruments listed above. Note that for this study, the germline background was generated using commercially available gDNA reference samples.

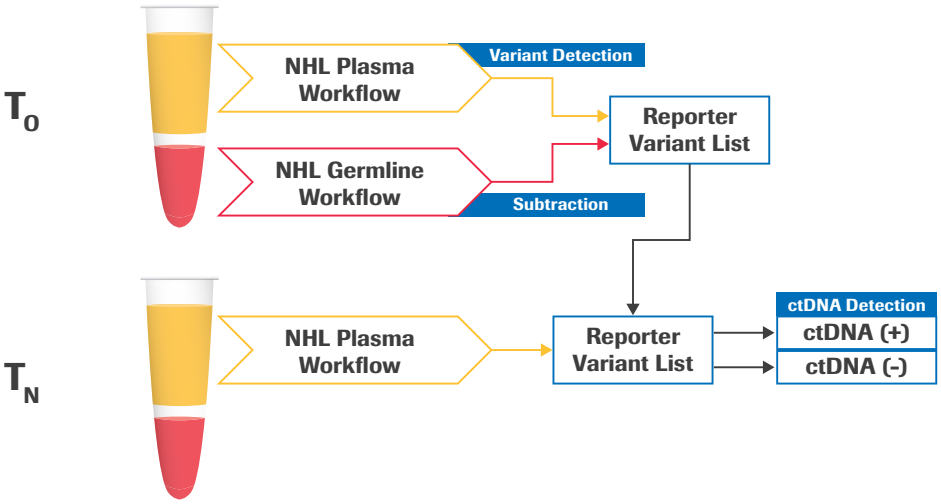

Figure 1. Longitudinal detection of non-Hodgkin lymphoma-associated variants using the KAPA HyperCap Design Share NHL Panel. Blood samples collected at an initial time point (T_0) are centrifuged to obtain plasma and plasma-depleted blood (PDB, composed of red cells and buffy coat). Cell-free circulating tumor DNA (cfDNA/ctDNA) is extracted from plasma, processed using the plasma cfDNA workflow, and sequenced to identify somatic variants. Genomic DNA (gDNA) is extracted from PDB, processed using the germline workflow and sequenced to identify germline variants. Subtraction of germline variants from a candidate list of somatic reporter variants results in a baseline reporter variant list that can be used for monitoring. At any subsequent time point (T_N), only the plasma cfDNA workflow is executed to assess the presence/absence of reporter variants. T_N samples are scored for longitudinal mutation positivity (akin to ctDNA detection or MRD analysis). Note that reference materials were used in this study to mimic cfDNA and gDNA that would normally be extracted from blood samples. Diluted mixtures were created to mimic longitudinal samples.

Data on file.

Table 1. Samples used in this study

	Sample name/ part number	Sample description	ctDNA/ Germline	Genotype	Sample type	Sample format
1	104542, 105900	Twist cfDNA Pan-cancer reference standard VAF 0% (WT)	ctDNA	WT	Reference (cell line)	Purified ctDNA
2	104548, 105906	Twist cfDNA Pan-cancer reference standard VAF 5%	ctDNA	Mut	Reference (cell line)	Purified ctDNA mix
3	0710-0533	Seraseq® ctDNA Complete™ Mutation Mix WT (0%)	ctDNA	Mut	Reference (cell line)	Purified ctDNA
4	0710-0528	Seraseq® ctDNA Complete™ Mutation Mix AF5%	ctDNA	Mut	Reference (cell line)	Purified ctDNA mix
5	0710-0531	Seraseq® ctDNA Complete™ Mutation Mix AF 0.5%	ctDNA	Mut	Reference (cell line)	Purified ctDNA mix
6	NA24149	GIAB NA24149	Germline	WT	Reference (cell line)	Purified gDNA
7	NA24631	GIAB NA24631	Germline	WT	Reference (cell line)	Purified gDNA
8	Various (*n*=23)	Samples from healthy donors	ctDNA	WT	Plasma	Plasma samples

Library Preparation and Target Enrichment

Plasma cfDNA Workflow: A total of 44 cfDNA libraries were prepared from 11 different reference DNA materials (30 ng inputs), as outlined in Table 2 (lines 1 – 11). Note that a subset of these (Table 1, lines 4, 5, 8, 9, 10, and 11) were contrived samples, prepared by combining the WT (AF 0%) and AF 5% from each commercial supplier in specific ratios to achieve allele frequencies of known mutations in the range of 0.01 – 0.05%. The number of replicate libraries prepared from each unique sample ranged from 2 – 8 per sample, with more replicates for libraries targeting known variants at lower allele frequencies. In addition, cfDNA libraries were generated from 30 ng inputs of 23 healthy donor samples (Table 2, line 13).

All cfDNA libraries were prepared with the KAPA HyperPrep Kit (Roche PN: 07962312001, 07962347001, or 07962363001), KAPA Universal UMI Adapter (Roche PN: 09329862001 or 09329889001), KAPA UDI Primer Mixes 1 – 384 (Roche PN: 09134336001, 09329838001, 09329846001, and 09329854001), and KAPA HyperPure Beads (Roche PN: 08963835001, 08963843001, 08963851001, 08963878001, or 08963860001) as described in the KAPA HyperCap cfDNA Workflow v1.1, Instructions for use.[9]

Germline workflow: A total of eight gDNA libraries were prepared from a mixture of NA24631 (98%) and NA24149 (2%) DNA for the germline workflow. Replicate libraries were prepared from 100 ng inputs using the KAPA HyperPlus Kit (Roche PN: 07962380001, 07962401001, or 07962428001), KAPA Universal UMI Adapter (Roche PN: 09329862001 or 09329889001), KAPA UDI Primer Mixes 1 – 384 (Roche PN: 09134336001, 09329838001, 09329846001, and 09329854001), and KAPA HyperPure Beads (Roche PN: 08963835001, 08963843001, 08963851001, 08963878001, or 08963860001) as described in the KAPA HyperCap Workflow v3.4, Instructions for Use.[10]

Pre-capture Library QC: Amplified pre-capture libraries were diluted 1/10 for analysis of fragment size distribution, performed according to the manufacturer's recommendations using an Agilent Bioanalyzer 2100 and High Sensitivity DNA Chips and Reagents (Agilent Technologies). The same diluted material was used for library quantification using a fluorescence microplate reader and the Quant-iT™ dsDNA High-Sensitivity Assay Kit (Thermo Fisher Scientific).

Target enrichment: Probe-based hybridization capture was performed with the KAPA HyperCap Design Share Panel ordered as KAPA HyperChoice MAX 3Mb T1 (Roche PN: 09052631001, IRN: 1000028225), KAPA HyperCapture Reagent Kit (Roche PN: 09075810001, 09075828001, or 09075917001) and KAPA HyperCapture Bead Kit (Roche PN: 09075780001, 09075798001, or 09075909001). Singleplex captures were performed as described in the standard protocols for the plasma cfDNA[9] and germline workflow,[10] respectively.

Post-capture Library QC: The concentration and fragment size distribution of sequencing-ready, post-capture (enriched) libraries from singleplex captures were determined using an Agilent 4200 TapeStation system and DNA High Sensitivity D1000 ScreenTape Assay (Agilent Technologies) according to the manufacturer's recommendations.

Sequencing and Data Analysis

Sequencing: Libraries were pooled for multiplexed, paired-end sequencing (2 x151 bp) on an Illumina® NextSeq™ 500/550 system. Pools of eight samples were configured to obtain approximately 90 M raw reads per library using standard sequencing protocols.

Data analysis: Data analysis was performed using *KAPA bioinformatics analysis for longitudinal detection of circulating tumor DNA*,[7] which comprises three main stages: (i) single-sample pre-processing, (ii) longitudinal mutation blocklist generation, and

Table 2. Breakdown of libraries prepared for this study

	Sample name	Purpose	Used for variant detection	Use in longitudinal analysis	No. of replicates
1	Seraseq® ctDNA Complete™ Mutation Mix WT (0%)	Test MRD analysis pipeline	No	T_0 (germline subtraction)	2
2	Seraseq® ctDNA Complete™ Mutation Mix AF5%	Test MRD analysis pipeline and plasma cfDNA workflow	Yes (cfDNA)	T_0 (plasma cfDNA workflow)	2
3	Seraseq® ctDNA Complete™ Mutation Mix AF0.5%	Test MRD analysis pipeline and plasma cfDNA workflow	Yes (cfDNA)	Not used	2
4	Seraseq® ctDNA Complete™ Mutation Mix AF0.1%*	Test MRD analysis pipeline	No	T_0 (plasma cfDNA workflow)	4
5	Seraseq® ctDNA Complete™ Mutation Mix AF0.05%*	Test MRD analysis pipeline	No	T_N (plasma cfDNA workflow)	6
6	Twist cfDNA Pan-cancer Reference Standard 0%	Test MRD analysis pipeline	No	T_0 (germline subtraction)	4
7	Twist cfDNA Pan-cancer Reference Standard 5%	Test MRD analysis pipeline and plasma cfDNA workflow	Yes (cfDNA)	T_0 (plasma cfDNA workflow)	4
8	Twist cfDNA Pan-cancer Reference Standard 0.5%*	Test MRD analysis pipeline	Yes (cfDNA)	T_N (plasma cfDNA workflow)	4
9	Twist cfDNA Pan-cancer Reference Standard 0.1%*	Test MRD analysis pipeline	No	T_N (plasma cfDNA workflow)	4
10	Twist cfDNA Pan-cancer Reference Standard 0.05%*	Test MRD analysis pipeline	No	T_N (plasma cfDNA workflow)	6
11	Twist cfDNA Pan-cancer Reference Standard 0.01%*	Test MRD analysis pipeline	No	T_N (plasma cfDNA workflow)	6
12	NA24149 (2%) / NA24631 (98%) mixture†	Test MRD analysis pipeline	Y (gDNA)	Not used (for germline workflow only)	8
13	Healthy donor samples (*n*=23)	Create blocklist	No	Use for blocklist	1 each

*Contrived samples, prepared by combination of commercially available WT (AF 0%) and AF 5% cfDNA samples to target specific allele frequencies of known mutations.
†Contrived sample (AF 1%) for the germline workflow, prepared by mixing the two GIAB samples.

(iii) longitudinal mutation analysis to detect previously identified reporter variants. (The blocklist is used to identify base-specific loci that are prone to high error rates; see below.)

The same pre-processing workflow is used to process sequencing data for (i) cfDNA and germline libraries, prepared from baseline (T0) samples, and (ii) cfDNA libraries derived from longitudinal samples. FASTQ files were generated from raw sequencing data using bcl-convert 3.10.5. UMI consensus identification was performed using tools from fgbio 1.3, GATK 4.2.0, bwa 0.7.17, and Samtools 1.13 to obtain UMI-deduplicated BAM files.

To generate the longitudinal mutation blocklist, the panel of normals (PON; data for the 23 cfDNA libraries prepared from healthy donor samples; Table 1, line 8 and Table 2, line 13) was processed using the single-sample pre-processing workflow. UMI deduplicated BAM files were used to generate the blocklist using the ctDNAtools package.[11] The blocklist is used to identify base-specific loci that are prone to high error rates. Since these loci have a high probably of generating false positive variants, it is critical to exclude them from the list of candidate reporter variants used for longitudinal mutation analysis. This stage of the analysis pipeline is only executed once to obtain a KAPA NHL Panel-specific blocklist file prior to performing longitudinal variant analysis for the first time.

For longitudinal mutation analysis, variant calling was first performed on the baseline (T_0 cfDNA) sample using VarDictJava 1.8.3 to identify reporter variant candidates. Parameters were set to retain SNVs that met the following criteria:* FILTER=PASS, AF >0.65% and AF <35%, DP >1000, VD >15, MQ >55, QUAL >45. Reporters were removed from the candidate list if their presence was detected in the germline sample or if they were included in the longitudinal mutation blocklist. Finally, the remaining reporter candidates were used to assess longitudinal mutations in T_N samples. Mutation positivity was determined using the Monte Carlo sampling empirical p-value approach,[12] based on the reference and alt (alternative) allele read counts and the background error rate. Calculations were performed using the ctDNAtools package.[11] The p-value cutoff for mutation positivity in T_N samples was set at 0.003, as this was the lowest p-value observed in the wild type (AF 0%) samples.

***AF**: Allele frequency; **DP**: Depth – total coverage; **VD**: AltDepth – variant coverage; **MQ**: mapping quality; **QUAL**: average base quality at a variant position.

Results and discussion

Assessment of workflow performance

Commercially available DNA preparations mimicking biological samples (pre-fragmented cfDNA and high molecular weight genomic DNA) were used to assess the performance of the plasma cfDNA and germline workflows outlined in Figure 1.

Library QC metrics: Pre-capture libraries prepared with both workflows met the yield and size distribution criteria for target enrichment. All post-capture (enriched) libraries and library pools met the yield and size distribution criteria for sequencing (data not shown; refer to KAPA HyperCap cfDNA Workflow v1.1 and KAPA HyperCap Workflow v3.4 Instructions for Use for details).

Sequencing performance metrics: KAPA bioinformatics analysis for longitudinal detection of ctDNA generates a list of sequencing QC metrics for every sample. A subset of these metrics is listed and defined in Table 3.

A subset of sequencing metrics for libraries prepared with the KAPA NHL Panel and plasma cfDNA workflow from 30 ng inputs of commercial cfDNA mixes are shown in Figure 2. (A: after UMI (Unique Molecular Identifier) deduplication, B: raw results). A median of 88 M raw reads were obtained across all libraries. After UMI deduplication, the median number reads returned for Complete Mutation Mix (Seraseq ctDNA) libraries was 44 M, compared to the median of 33 M reads for Pan-cancer Reference Standard (Twist cfDNA) libraries. This translated to a median coverage depth of 6100X and 5000X, respectively.

Other results from Figure 2A worth noting are:

- The uniformity metric *fold-80 base penalty*, which indicates the amount of additional sequencing required to ensure that the mean coverage is achieved for 80% of target bases. Penalty values of approximately 1.6 indicated good coverage uniformity across all libraries.
- The percentage of reads on the primary target (% selected bases) was high (median of 74% for all libraries) and highly reproducible across DNA samples from different suppliers and libraries prepared in different batches.
- The average percentage of target bases covered at ≥1000X was >93% for all libraries, whereas the average percentage covered at ≥2500X was >86%.
- The median library insert size was 146 bp and 165 bp for Complete Mutation Mix and Pan-cancer Reference Standard libraries, respectively. Intrinsic differences between the DNA preparations from different suppliers may explain the lower median coverage after duplicate removal in the libraries with longer inserts.

Corresponding raw sequencing metrics are given in Figure 2B. The percentage of duplicate reads prior to UMI deduplication was approximately 57% and 73% for Complete Mutation Mix and Pan-cancer Reference Standard libraries, respectively (not shown). The average error rate of 2.4×10^{-4} mismatches/read depth was consistent across DNA types and library replicates.

Sequencing metrics for the eight replicate libraries prepared with the germline workflow from 100 ng inputs of the NA24149/NA24631 gDNA mixture are given in Figure 3 (A: after UMI deduplication, B: raw results). A median of 89 M raw reads were obtained. Removal of UMI duplicates reduced this to a median of 48 M reads, yielding a median coverage depth of approximately 9100. Fold-80 base penalty values were low (<1.46) across all replicates. The median on-target bases (% selected bases) was >80% across all replicates, and the average percentage of target bases covered at ≥1000X and ≥2500X exceeded 97% and 94%, respectively. The median insert size for the gDNA libraries was 189 bp, and the average error rate was 3.1×10^{-4} mismatches/read depth. The mean percentage of duplicate reads across all replicates was approximately 40% prior to UMI deduplication.

Table 3. Key sequencing QC metrics reported by the KAPA bioinformatics analysis for longitudinal detection of ctDNA

Metric	Description
Total reads	Total number of reads including all PF and non-PF reads. When CATEGORY equals PAIR this value will be 2x the number of clusters.
Percent passing filter reads	Fraction of reads that are PF (PF_READS / TOTAL_READS)
Mean target coverage	Mean coverage of a target region
Median target coverage	Median coverage of a target region
Fold-80 base penalty	Fold over-coverage necessary to raise 80% of bases in "non-zero-cvg" targets to the mean coverage level in those targets
Percent selected bases	Fraction of PF_BASES_ALIGNED located on or near a baited region, calculated as follows: (ON_BAIT_BASES + NEAR_BAIT_BASES) / PF_BASES_ALIGNED
Percent target bases >1000X	Fraction of all target bases achieving 1000X or greater coverage
Percent target bases >2500X	Fraction of all target bases achieving 2500X or greater coverage
Median insert size	MEDIAN insert size of all paired end reads where both ends mapped to the same chromosome
Error rate	Mismatch rate calculation from ctDNAtools package get_background_rate.R. Defined as (the sum of mismatches) / (sum of read depths) for all bases in the targets.

Terms in CAPITALS are Picard metrics. See https://broadinstitute.github.io/picard/picard-metric-definitions.html for a complete list. The error rate calculation was performed with the ctDNAtools package.[11]

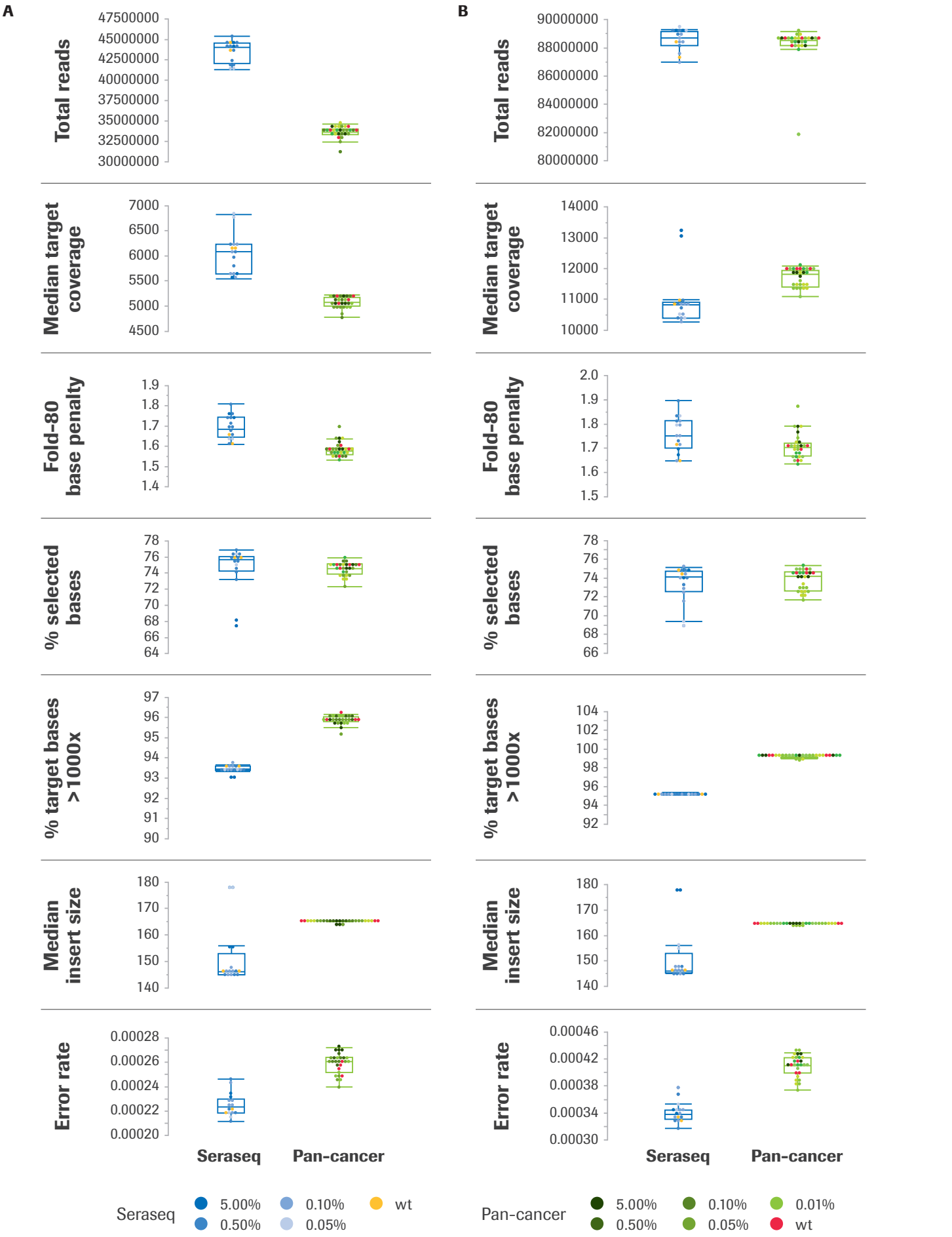

Figure 2. Key sequencing performance metrics for the KAPA NHL Panel in the plasma cfDNA workflow. (A) Results after UMI deduplication and **(B)** Results from raw data. Libraries were prepared from 30 ng inputs of commercial cfDNA mixes, enriched, and sequenced as outlined in *Materials and methods*. Data were analyzed with the KAPA bioinformatics analysis for longitudinal detection of ctDNA as described. The number of replicates for each cfDNA sample is summarized in Table 2.

Data on file.

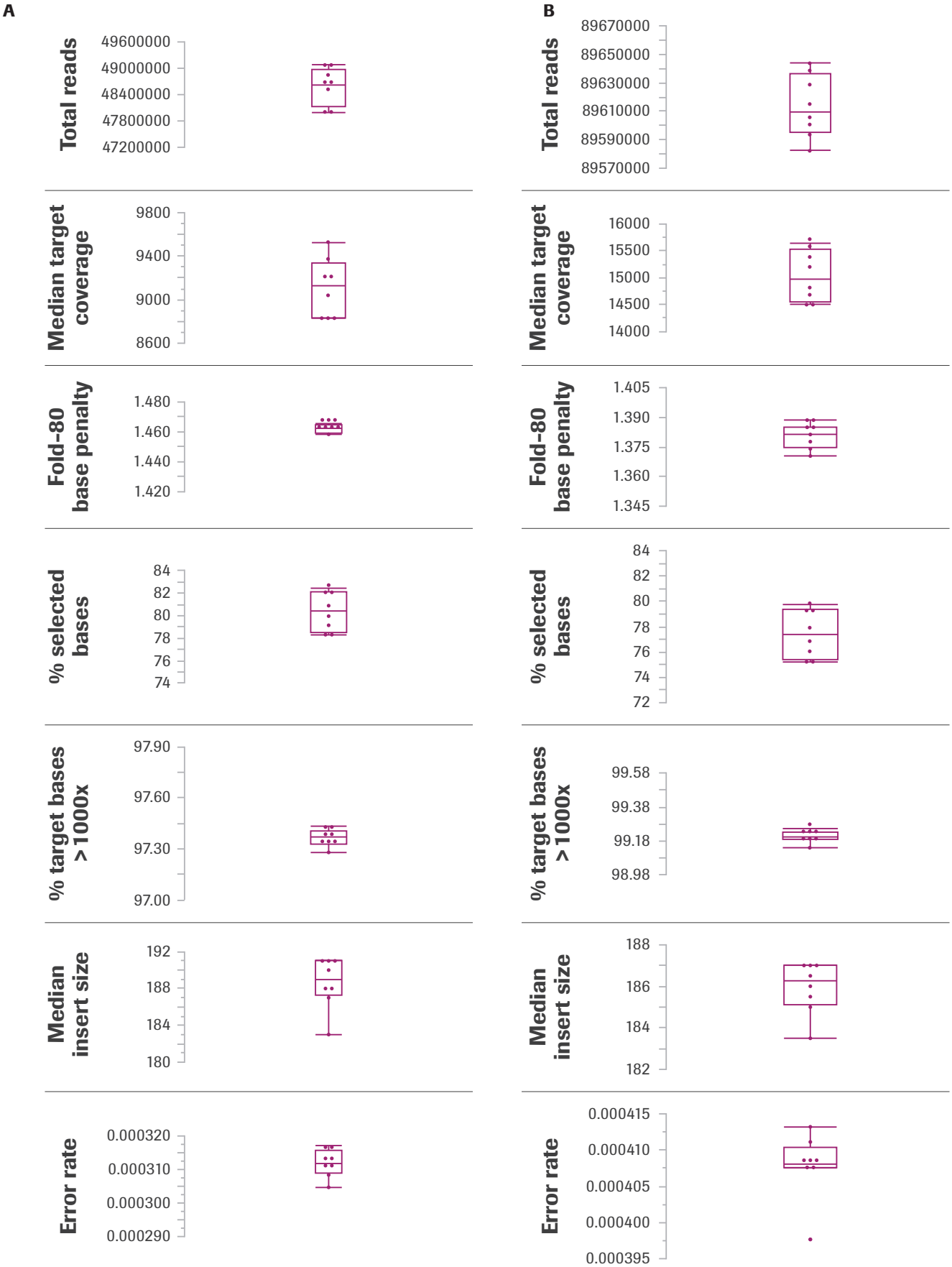

Figure 3. Key sequencing performance metrics for the KAPA NHL Panel in the germline workflow. (A) Results after UMI deduplication and **(B)** Results from raw data. Eight replicate libraries were prepared from 100 ng inputs of a 98:2 mixture of NA24631 and NA24149 gDNA, enriched, and sequenced as outlined in *Materials and methods*. Data were analyzed with the KAPA bioinformatics analysis for longitudinal detection of ctDNA as described.

Variant calling results: Reference materials with characterized mutations were included in this study to assess variant calling performance for somatic SNVs. Samples 2 and 3 in Table 2 contain three characterized SNVs covered by the KAPA NHL Panel (described in more detail in Table A.5 in the Appendix), whereas samples 7 and 8 contain twelve characterized SNVs (see Table A.6). All expected SNVs were found in all replicates of the relevant samples.

Germline variant calling performance was also assessed, using sample 12 in Table 2. Concordance between observed and expected variants in the NA24149/NA24631 mixture was very high. Ten out of ten true positives (TP) and 53 out of 53 true negatives (TN) were observed in all eight replicates, translating to very high sensitivity and specificity. In addition, high SNV calling specificity was demonstrated by 0.06 errors every 10 Kb of the panel (as calculated by a set of 23 healthy donor samples). Refer to Table A.4 in the Appendix for lists of TP and TN variants.

Longitudinal mutation analysis

Longitudinal mutation analysis was performed with the KAPA bioinformatics analysis for longitudinal detection of ctDNA, which utilizes three samples:

- the baseline (T_0) cfDNA sample, which is used to generate a candidate list of variants to be used as reporters for longitudinal mutation analysis;
- the T_0 germline sample, which is used to filter out candidate reporter variants found in the normal germline, to obtain a final list of reporter variants; and
- the T_N cfDNA sample, which is analyzed for the presence/absence of reporters to determine whether the sample is positive or negative for longitudinal mutations.

In order to generate a set of results that can be reproduced experimentally, longitudinal mutation analysis was not performed using real-world samples. Instead, commercially available paired wild type gDNA and pre-fragmented reference cfDNA samples from two different suppliers were used to create two sets of contrived (T_0 and T_N) samples, targeting known variants at specific allele frequencies: AF 0.5%, 0.1%, 0.05%, and 0.01% for Pan-cancer Reference Standards from Twist Biosciences, and AF 0.5%, 0.1%, and 0.05% for Complete™ Mutation Mixes from SeraCare/LGC Clinical Diagnostics.

Reporter variant candidates for the contrived Pan-cancer Reference Standard samples were obtained from the baseline AF 5% sample. Twelve vendor-verified SNVs were expected to be covered by the KAPA HyperCap Design Share NHL Panel after germline filtering (see Table A.6 in the Appendix). Of those, three appeared in the longitudinal mutation blocklist (and two of the three were also detected in the germline sample). The three blocklist variants were confirmed to have elevated background error rates in the PON samples (thereby increasing the chance of false positive calls) and were removed from the candidate list. The remaining nine reporter variants were used in longitudinal mutation analysis.

Reporter variants were successfully detected in all contrived T_N Pan-Cancer samples. The Monte Carlo p-value threshold for ctDNA positivity in simulated longitudinal samples was set at 0.003 since this was the lowest value observed in the wild type sample. Observed vs. expected allele frequencies for the nine reporter variants are shown in Figure 4, and results are summarized in Table 4. The number of reporter variants with non-zero supporting reads, as well as the total number of supporting alt reads, drops as the expected AF % decreases from 0.5% to 0.01%. Mutation positivity was accurately called in all replicates of the AF 0.5%, AF 0.1%, and AF 0.05% samples (Monte Carlo p-values <0.005). For the AF 0.01% sample, mutation positivity was accurately called in five out of six replicates. All replicates of the wild type sample were called negative.

In similar fashion, reporter variant candidates for the contrived Complete Mutation Mix samples were obtained from the corresponding baseline AF 5% sample. All three of the vendor-verified SNVs (listed in Table A.5 in the Appendix) were confirmed to be absent from the germline sample and blocklist and were used in longitudinal mutation analysis.

Reporter variants were successfully detected in all contrived T_N Complete Mutation Mix samples. Observed vs. expected allele frequencies are shown in Figure 5. Longitudinal mutation positivity was accurately called in all replicates of the AF 0.5%, AF 0.1%, and AF 0.05% samples, with corresponding Monte Carlo p-values <0.003 (Table 5). All replicates of the wild type sample were called negative.

Conclusion

The KAPA HyperCap Workflow with KAPA HyperCap Design Share NHL Panel offers a robust and streamlined method for preparing NGS libraries enriched for NHL-associated variants from a wide variety of sample types, including cell-free circulating tumor (cf/ctDNA). Preparing high-quality libraries from difficult samples is, however, only half of the challenge. Bioinformatic pipelines are needed to unlock the information captured in sequencing libraries—to elucidate tumor biology, classify neoplasms, and understand tumor behavior in response to treatment. The KAPA bioinformatics analysis for longitudinal detection of ctDNA used in this study (and described in detail elsewhere[7]) provides a bioinformatics solution for NHL research, composed from open-source tools.

Confirming that sequencing libraries are of a high quality (i.e., are able to support conclusions regarding biological phenomena) is an important aspect of data analysis. Leading institutions like the Broad Institute have published sequencing performance metrics that are used to assess library quality.[13] A subset of these metrics have been incorporated in the KAPA bioinformatics analysis for longitudinal detection of ctDNA, and ten of those were used to assess the quality of cfDNA libraries (and germline controls) in this study. In addition, reference samples with known NHL-associated SNVs were included to verify that the complete workflow—from sample to analysis—produces reliable and reproducible results.

Once library quality was confirmed, sequencing data from cfDNA samples and germline controls were used to perform longitudinal NHL mutation analysis. Results from contrived samples prepared from reference materials confirmed that reporter variants can be detected with high reproducibility at allele frequencies as low as 0.05%, and with good reproducibility at AF of 0.01%.

Overall, the fully integrated KAPA HyperCap workflow with the KAPA NHL panel can support the analysis of longitudinal dynamics of circulating tumor DNA and the detection of minimal residual disease to further advance research on molecular response and MRD detection in lymphoma.

References

1. https://seer.cancer.gov/statfacts/html/all.html. Accessed 22 June 2023.
2. Lou J, Craver A, Bahl K, et al. Etiology of non-Hodgkin lymphoma: A review from epidemiologic studies. *JNCC* 2022;2(4),226 – 234. doi: 10.1016/j.jncc.2022.08.003.
3. Campo E, Jaffe ES, Cook JR, et al. The International Consensus Classification of Mature Lymphoid Neoplasms: a report from the Clinical Advisory Committee. *Blood* 2022;140(11),1229 – 1253: doi: 10.1182/blood.2022015851.
4. Roschewski M, Rossi D, Kurtz D et al. Circulating Tumor DNA in Lymphoma: Principles and Future Directions. *Blood Cancer Discov* 2022; 3(1)5 – 15. doi: 10.1158/2643-3230.BCD-21-0029.
5. Scherer F, Kurtz, DM, Newman AM, et al. Distinct biological subtypes and patterns of genome evolution in lymphoma revealed by circulating tumor DNA. *Sci Transl Med* 2016;8(364),364ra155. doi: 10.1126/scitranslmed.aai8545.
6. Fernández-Miranda I, Pedrosa L, Llanos M, et al. Monitoring of Circulating Tumor DNA Predicts Response to Treatment and Early Progression in Follicular Lymphoma: Results of a Prospective Pilot Study. *Clin Cancer Res* 2023;29(209 – 220). doi: 10.1158/1078-0432.CCR-22-1654.
7. Chien, R. KAPA bioinformatics analysis for longitudinal detection of circulating tumor DNA. Roche white paper. MC--12095.
8. Saelee SL, Lovejoy AF, Hinzmann B, et al. Quantitative PCR–Based Method to Assess Cell-Free DNA Quality, Adjust Input Mass, and Improve Next-Generation Sequencing Assay Performance. *J Mol Diagn* 2022;24(6), 566 – 575: doi: 10.1016/j.jmoldx.2022.02.005.
9. Roche Sequencing Solutions. KAPA HyperCap cfDNA Workflow v1.1; Instructions for use. https://elabdoc-prod.roche.com/eLD/api/downloads/cbc03242-4122-ec11-0b91-005056a772fd?countryIsoCode=pi. Accessed 22 June 2023.
10. Roche Sequencing Solutions. KAPA HyperCap Workflow v3.4; Instructions for use. https://elabdoc-prod.roche.com/eLD/api/downloads/6a0292c9-f2de-ed11-1d91-005056a71a5d?countryIsoCode=pi. Accessed 22 June 2023.
11. Alkodsi A, Meriranta L, Pasenen A, Sirpa Leppä. ctDNAtools: An R package to work with sequencing data of circulating tumor DNA. *bioRxiv* 2020.01.27.912790. doi: 10.1101/2020.01.27.912790.
12. Newman AM, Lovejoy AF, Klass DM, et al. Integrated digital error suppression for improved detection of circulating tumor DNA. *Nature Biotechnol* 2016;34(547 – 555). doi: 10.1038/nbt.3520.
13. Picard Toolkit. Broad Institute, GitHub Repository. https://broadinstitute.github.io/picard/. Accessed 22 June 2023.

Acknowledgments

The authors would like to thank the following Roche colleagues: Ehsan Tabari and Todd Richmond for their contributions to the development of the bioinformatics workflow and KAPA HyperCap Design Share NHL panel design; Andrea Ramos and Renee Stokowski for assisting with the development of the wet lab workflow and panel validation; and Markos Mihalatos and Suzanne Cheng for their significant contributions to the conceptualization and execution of the project.

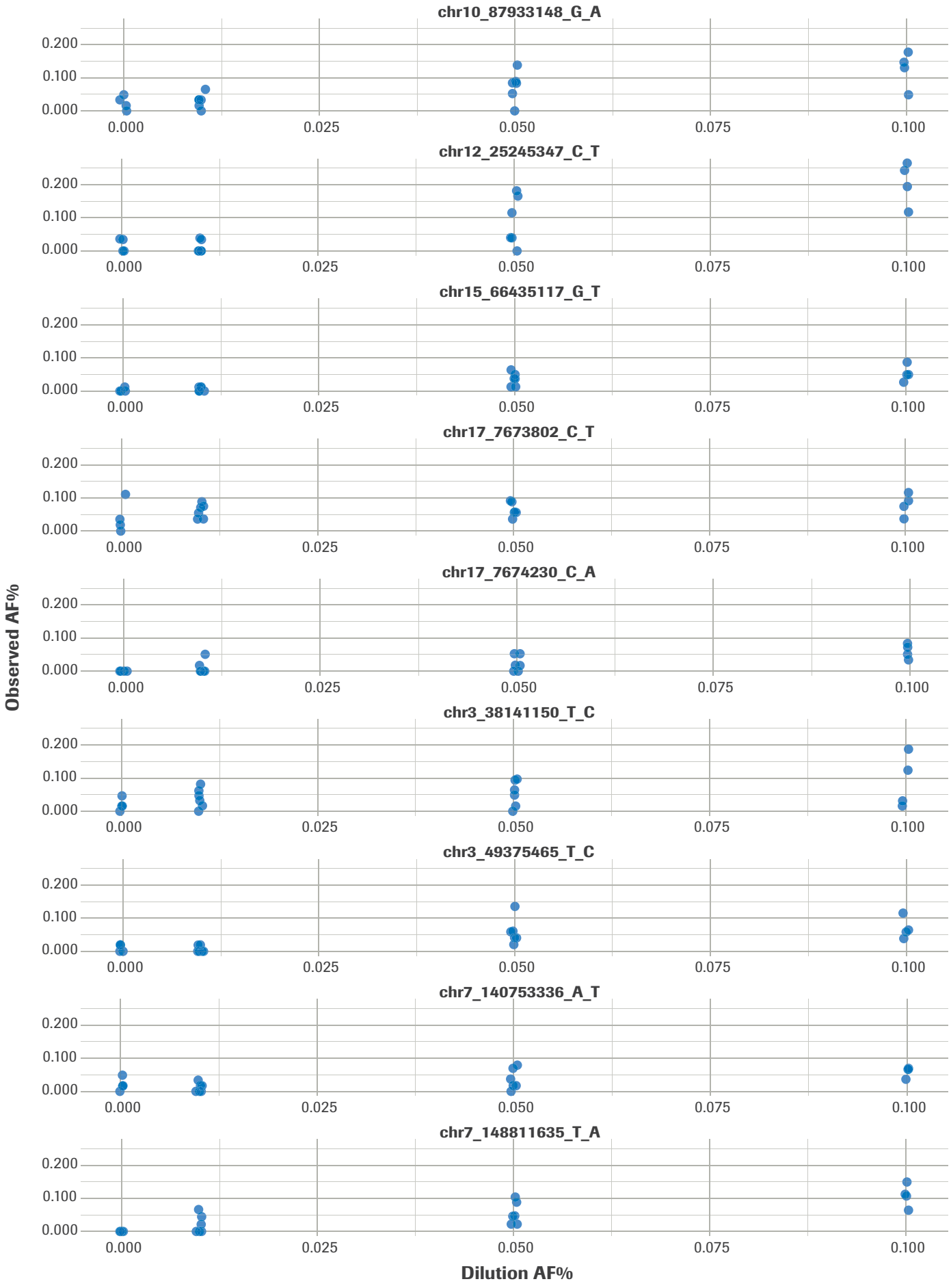

Figure 4. Observed vs. expected allele frequencies for nine reporter SNVs used in longitudinal mutation analysis of contrived Pan-cancer Reference Standard samples. Libraries were prepared and analyzed as described in *Materials and methods*. Of twelve vendor-confirmed SNVs, three were eliminated due to their presence in the longitudinal mutation blocklist or T_0 germline sample. Results are summarized in Table 4.

Data on file.

Table 4. Summary of longitudinal mutation analysis results for contrived Pan-cancer Reference Standard samples

Sample	Rep #	n_mutations	n_nonzero_alt	total_alt_reads	informative_reads	p-value	decision	total_af*	expected_af
AF 0.5%	1	9	9	205	50312	0.0001	positive	0.407%	0.50%
	2	9	9	196	49707	0.0001	positive	0.394%	0.50%
	3	9	9	197	51206	0.0001	positive	0.385%	0.50%
	4	9	9	198	51148	0.0001	positive	0.387%	0.50%
AF 0.1%	1	9	9	38	50383	0.0001	positive	0.075%	0.10%
	2	9	9	48	46862	0.0001	positive	0.102%	0.10%
	3	9	9	40	50633	0.0001	positive	0.079%	0.10%
	4	9	9	44	50145	0.0001	positive	0.088%	0.10%
AF 0.05%	1	9	7	17	48153	0.0001	positive	0.035%	0.05%
	2	9	8	27	48575	0.0001	positive	0.056%	0.05%
	3	9	8	25	48949	0.0001	positive	0.051%	0.05%
	4	9	9	20	48507	0.0001	positive	0.041%	0.05%
	5	9	8	29	49869	0.0001	positive	0.058%	0.05%
	6	9	8	31	47337	0.0001	positive	0.065%	0.05%
AF 0.01%	1	9	5	9	48945	0.0018	positive	0.018%	0.01%
	2	9	4	12	49792	0.0005	positive	0.024%	0.01%
	3	9	6	11	50665	0.0003	positive	0.022%	0.01%
	4	9	2	5	49152	0.2858	**negative**	0.010%	0.01%
	5	9	8	16	50211	0.0001	positive	0.032%	0.01%
	6	9	6	14	50613	0.0001	positive	0.028%	0.01%
WT (AF 0%)	1	9	5	7	51250	0.0297	**negative**	0.014%	0.00%
	2	9	4	10	50116	0.0031	**negative**	0.020%	0.00%
	3	9	3	7	51083	0.1217	**negative**	0.014%	0.00%
	4	9	5	6	51116	0.0522	**negative**	0.012%	0.00%

*Observed AF% (total_af) = (total_alt_reads / total_informative_reads) x 100%

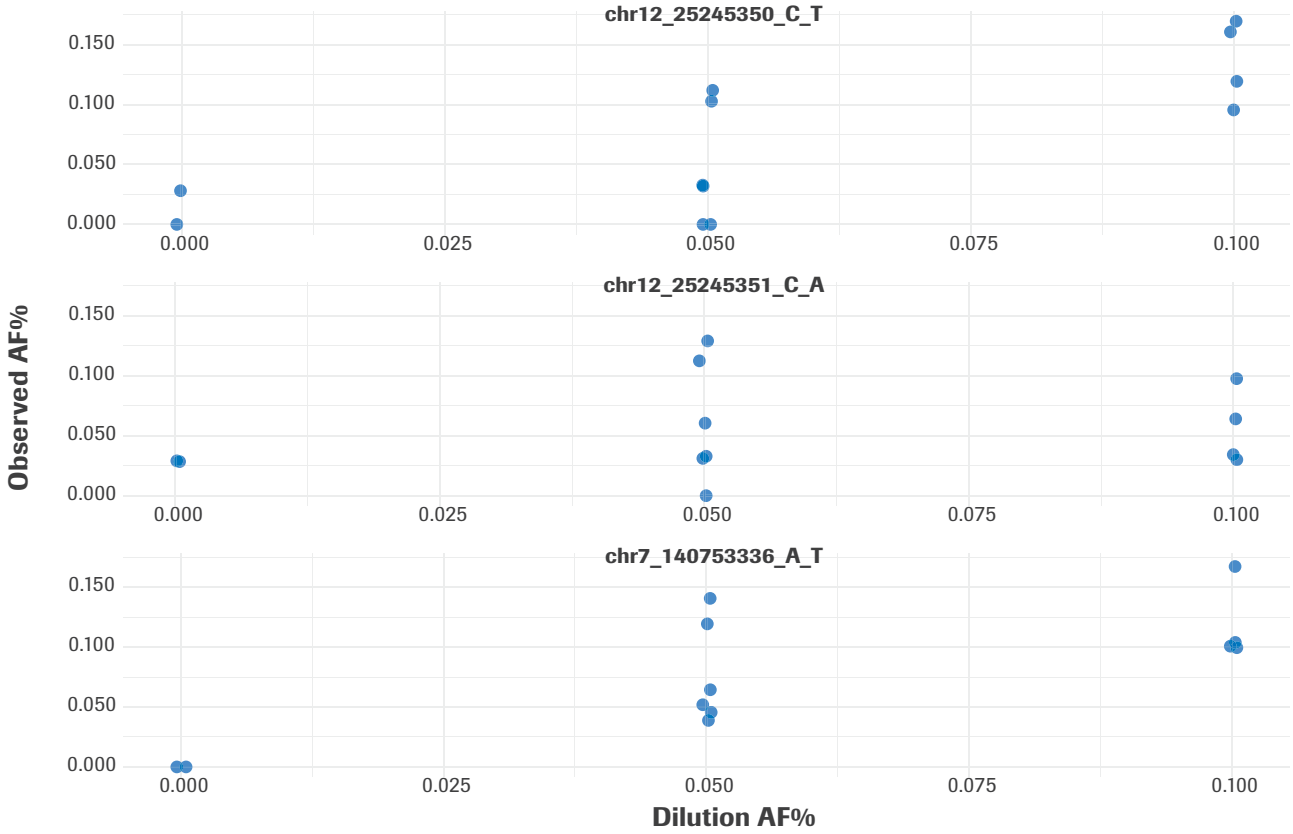

Figure 5. Observed vs. expected allele frequencies for three reporter SNVs used in longitudinal mutation analysis of contrived Complete™ Mutation Mix samples. Libraries were prepared and analyzed as described in *Materials and methods*. None of the three vendor-confirmed SNVs were used in the analysis. Results are summarized in Table 5.

Table 5. Summary of longitudinal mutation analysis results for contrived Complete Mutation Mix samples

Sample	Rep #	n_mutations	n_nonzero_alt	total_alt_reads	informative_reads	p-value	decision	total_af*	expected_af
AF 0.5%	1	3	3	96	12370	0.0001	positive	0.776%	0.50%
	2	3	3	97	12570	0.0001	positive	0.772%	0.50%
AF 0.1%	1	3	3	13	11100	0.0001	positive	0.117%	0.10%
	2	3	3	12	10223	0.0001	positive	0.117%	0.10%
	3	3	3	15	10100	0.0001	positive	0.149%	0.10%
	4	3	3	17	9556	0.0001	positive	0.178%	0.10%
AF 0.05%	1	3	3	11	11344	0.0001	positive	0.097%	0.05%
	2	3	2	6	11081	0.0022	positive	0.054%	0.05%
	3	3	2	11	10391	0.0001	positive	0.106%	0.05%
	4	3	3	10	9792	0.0001	positive	0.102%	0.05%
	5	3	3	8	9748	0.0001	positive	0.082%	0.05%
	6	3	2	7	9160	0.0002	positive	0.076%	0.05%
WT (AF 0%)	1	3	1	1	10786	0.2613	**negative**	0.009%	0.00%
	2	3	2	2	11159	0.0535	**negative**	0.018%	0.00%

*Observed AF % (total_af) = (total_alt_reads / total_informative_reads) x 100%

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

Appendix

Supplemental information

Table A.1. KAPA HyperCap Design Share NHL Panel gene list

ABCB11	CAPZA3	DSEL	H1-5	IGHV1-58	IGKV2-24	IGLV3-25	LIPH	NTNG1	RIMS2	WDPCP
ACAD8	CARD11	DTX1	H2AC11	IGHV1-69	IGKV2-30	IGLV3-27	LIPM	P2RY8	RNF213	WRAP73
ACSS1	CCND3	DUSP2	H2AC16	IGHV2-26	IGKV2D-26	IGLV3-32	LIPN	PABPC1	RNLS	XBP1
ACTA2	CD274	DUSP3	H2AC17	IGHV2-5	IGKV2D-29	IGLV3-9	LRP10	PANK1	S1PR2	XPO1
ACTB	CD36	EBF3	H2AC6	IGHV3-11	IGKV2D-30	IGLV4-3	LRP1B	PAPSS2	SCML4	YTHDF2
ACTG1	CD53	EDIL3	H2AC8	IGHV3-13	IGKV3-20	IGLV4-60	LRP2	PCDH17	SERPINE3	ZEB2
ACVR2A	CD58	EFEMP1	H2BC11	IGHV3-15	IGKV3D-11	IGLV4-69	LRRN3	PCDHA6	SGK1	ZFP36L1
ADAMTS1	CD70	EGR1	H2BC12	IGHV3-16	IGKV3D-20	IGLV5-37	LRRTM4	PCLO	SI	ZFP42
ADAMTS16	CD79B	EHBP1	H2BC14	IGHV3-20	IGKV4-1	IGLV5-45	MAGEB16	PCSK5	SLC22A16	ZMYM6
ADAMTS9	CD83	ELAC1	H2BC17	IGHV3-21	IGKV5-2	IGLV5-48	MAGEC2	PCSK7	SLC9A4	ZNF577
AFF1	CDH12	ELAVL1	H2BC4	IGHV3-23	IGLJ2	IGLV5-52	MAP2K1	PDCD1	SLITRK1	ZNF608
AHCYL1	CDH19	ENSG00000281179	H2BC5	IGHV3-30	IGLJ3	IGLV6-57	MAP3K13	PDCD1LG2	SMAD4	ZNF649
ANKRD22	CDKN2A	ENSG00000282988	H2BC8	IGHV3-33	IGLJ4	IGLV7-43	MC5R	PIK3C2G	SOCS1	ZNF678
APC	CDKN2B	ENSG00000285938	H2BC9	IGHV3-35	IGLJ5	IGLV7-46	MDH1	PIK3CD	SORCS2	
ARHGAP5	CEP104	ENSG00000285947	H3C1	IGHV3-38	IGLJ7	IGLV8-61	MED12	PIK3CG	SRRM2	
ARID1A	CFAP276	EP300	H3C10	IGHV3-43	IGLL5	IGLV9-49	MEF2B	PIK3R1	STAT3	
ATP1B4	CFL1	EPHA7	H3C11	IGHV3-48	IGLV10-54	IRAG2	MEX3C	PIM1	STAT6	
ATP8B1	CLSTN2	EPS8	H3C2	IGHV3-49	IGLV11-55	IRF1	MFHAS1	PLCG2	STT3A	
B2M	CNTNAP2	ERICH1	H3C3	IGHV3-64	IGLV1-36	IRF4	MPDZ	PLCL1	TAF1	
BCHE	COL22A1	EZH2	H3C4	IGHV3-7	IGLV1-40	IRF8	MPEG1	POM121L2	TAS2R16	
BCL10	COL24A1	F2RL2	H3C7	IGHV3-72	IGLV1-44	ITPKB	MRO	POU2F2	TBC1D22A	
BCL11A	CREBBP	FAS	H4C12	IGHV3-73	IGLV1-47	IZUMO3	MSH6	POU2F3	TBL1XR1	
BCL2	CRISPLD1	FAT1	HAS2	IGHV3-74	IGLV1-50	JUNB	MTAP	PPP2R1B	TCL1A	
BCL6	CRYAB	FBXO11	HDAC7	IGHV4-28	IGLV1-51	KCND2	MYC	PPP4C	TET2	
BCR	CSMD1	FBXW7	HMCN1	IGHV4-31	IGLV2-11	KHDRBS3	MYD88	PRDM1	TGFBI	
BNC2	CSMD3	FGFR4	ID3	IGHV4-34	IGLV2-14	KIF2B	MYO15A	PSD3	THBD	
BORCS8	CTNNA2	FOCAD	IGHA1	IGHV4-39	IGLV2-18	KLF10	NARS1	PSRC1	THYN1	
BORCS8-MEF2B	CTNND2	FOXO1	IGHA2	IGHV4-59	IGLV2-23	KLHL14	NCAPD3	PTEN	TLR2	
BRAF	CXCR4	FXYD6	IGHG2	IGHV5-51	IGLV2-33	KLHL25	NCSTN	PWWP3A	TMEM30A	
BRCA2	CXCR5	FXYD6-FXYD2	IGHG4	IGHV6-1	IGLV2-8	KLHL4	NEDD4L	PXDN	TMSB4X	
BRINP3	DFFB	GK2	IGHM	IGHV7-81	IGLV3-1	KLHL6	NEXMIF	RASSF9	TNFAIP3	
BTG1	DHX33	GNA13	IGHV1-18	IGKV1-5	IGLV3-10	KLK11	NFATC1	RBFA	TNFRSF14	
BTG2	DKC1	GRM7	IGHV1-2	IGKV1-6	IGLV3-12	KLK13	NFIA	REL	TNFRSF1A	
BTG3	DLAT	GSDMC	IGHV1-24	IGKV1-8	IGLV3-16	KMT2D	NFKBIA	RER1	TP53	
BTK	DNAH5	H1-2	IGHV1-3	IGKV1D-17	IGLV3-19	KRAS	NFRKB	RFTN1	TRPS1	
CACNA1E	DPAGT1	H1-3	IGHV1-45	IGKV1D-43	IGLV3-21	LAMA1	NOTCH1	RHOA	USP34	
CACNA1S	DRG2	H1-4	IGHV1-46	IGKV1D-8	IGLV3-22	LINGO2	NOTCH2	RHOH	VPS8	

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

Table A.2. Reagent volumes of cobas® cfDNA Sample Preparation Kit recommended for this white paper

Reagent	Volume*
PK	1 mL
DNA PBB	8 mL
Isopropanol	2 mL
WBI	500 µL
WBII	500 µL
Elution buffer	65 µL

*Reagent volumes for extraction of cfDNA for 6 mL (minimum) to 10 mL (highly recommended) of plasma (in this white paper using cobas® cfDNA Sample Preparation Kit)

Table A.3. Quantitative PCR reaction volumes to determine cfDNA quality and quantify gDNA contamination[8]

Amplicon 66 bp	Volume per reaction (µL)	Amplicon 330 bp	Volume per reaction (µL)
QC PCR Reaction Mix (2X)	5	QC PCR Reaction Mix (2X)	5
66F (8 µM)	0.5	330F (8 µM)	0.5
Reverse (8 µM)	0.5	Reverse (8 µM)	0.5
500X water-diluted cfDNA/ 500X water-diluted QC PCR standard (kit provided)/ NTC (PCR grade water)	4	500X water-diluted cfDNA/ 500X water-diluted QC PCR standard (kit provided)/ NTC (PCR grade water)	4
Total volume	10	Total volume	10

This assay is designed to amplify two amplicons of 66 bp and 330 bp in size. Each amplicon has its unique forward primer and shares the same reverse primer. A QC PCR DNA Standard is included in each qPCR run. Dilute the input DNA samples and QC PCR DNA Standard in water to a 500-fold final dilution. To increase accuracy and precision, it is recommended to split the 500-fold dilution into 2 steps: first, 100-fold dilution, and then 5-fold dilution. The qPCR is performed with an initial denaturation of 10 min followed by 40 cycles of denaturation (10 sec at 95°C), annealing (30 sec at 60°C) and extension (30 sec at 72°C), and one last cooling step. The quality score is determined by the following equation: Q-ratio = $2^{(averageCp66 - averageCp330)}$. A normalized Q score for each sample is obtained through the following equation: Normalized Q Score = sample Q score / QC PCR DNA standard Q score. The proportion of HMW DNA in a sample is derived from HMW = 1.106 x Q-ratio - 0.161. For a fixed total DNA concentration, the higher the HMW DNA proportion is, the lower the extracted cfDNA yield and ng cfDNA input, thus impacting the yield of amplified sample library.

Table A.4. SNVs in the 98:2 mixture of NA24631/NA24149 gDNA covered by the KAPA HyperCap Design Share NHL Panel

A. Variants expected at AF 1% (true positives)					
Chromosome	**Reference**	**Alternate**	**Position**	**Gene**	**Mutation**
chr2	T	C	140487635	LRP1B	c.9225A>G
chr3	C	T	16377802	RFTN1	c.742G>A
chr4	T	C	186636808	FAT1	c.3749A>G
chr5	C	T	11397076	CTNND2	c.567G>A
chr6	C	T	106105508	PRDM1	c.1348C>T
chr8	A	G	127743276	MYC-PVT1	n.127743276A>G
chr8	C	T	127939542	PVT1	n.406C>T
chr9	G	A	27950496	LINGO2	c.176C>T
chr13	G	T	57633577	PCDH17	c.1031G>T
chr22	C	T	23181882	BCR	c.922C>T

B. Loci with no expected variants (AF 0%, true negatives)							
Chromosome	**Position**	**Gene**	**Mutation**	**Chromosome**	**Position**	**Gene**	**Mutation**
chr1	85267691	BCL10	c.638G>A	chr14	105644801	IGHG2	c.-11C>T
chr2	140487592	LRP1B	c.9243+25A>G	chr14	105708754	IGHA1	c.-89C>T
chr3	183492477	KLHL6	c.1564+17T>C	chr14	105709265	IGHG1	n.105709265C>T
chr3	183555341	KLHL6	c.293+20G>C	chr14	105745668	IGHG1	c.437-2598C>T
chr3	186996999	ST6GAL1	c.-183+33073G>T	chr14	105772993	IGHG1	c.437-29923G>A
chr4	186636801	FAT1	c.3756C>T	chr14	105862082	IGHG1	c.436+1116T>C
chr6	26251920	HIST1H2BH	c.270C>T	chr14	105862166	IGHG1	c.436+1032C>T
chr6	41936045	CCND3	c.774C>T	chr14	106422367	IGHV4-39	c.-199G>C
chr6	106088378	PRDM1	c.220G>A	chr14	106422493	IGHV4-39	c.-325C>T
chr6	137881241	TNFAIP3	c.2295C>T	chr14	106715277	IGHV1-69	c.-157C>G
chr7	2945952	CARD11	c.225G>C	chr16	10880196	CIITA	c.52+2814G>A
chr8	121614628	HAS2	c.1140G>A	chr17	53823475	KIF2B	c.442C>T
chr8	127751897	MYC-PVT1	n.127751897C>T	chr18	63126261	BCL2	n.63126261G>A
chr8	130360831	ASAP1	c.60-2688C>T	chr18	63126316	BCL2	n.63126316G>A
chr8	136089081	RP11-149P24.1	n.205C>T	chr18	67513469	DSEL	c.1170A>G
chr9	24545772	IZUMO3	c.-123C>G	chr18	79448922	NFATC1	c.1527C>T
chr9	37025609	PAX5	c.47-4808A>G	chr18	79467473	NFATC1	c.1983G>A
chr9	37384705	RP11-397D12.4	n.-271C>G	chr19	19147847	MEF2B	c.259-15C>T
chr10	88820255	LIPM	c.1026C>T	chr22	22698529	IGLV2-23	c.*122C>T
chr11	134260976	ACAD8	c.706-68C>T	chr22	22698655	IGLV2-23	c.*248C>T
chr11	134260993	ACAD8	c.706-51G>A	chr22	22698869	IGLV2-23	c.*462G>C
chr11	134261186	ACAD8	c.841+7T>C	chr22	22889148	IGLL5	c.327+2666G>A
chr11	134263997	ACAD8	c.1196-911G>A	chr22	22889858	IGLL5	c.328-2125C>T
chr12	18648084	PIK3C2G	c.*56G>A	chr22	22890256	IGLL5	c.328-1727G>A
chr12	49044564	KMT2D	c.4964-42A>G	chr22	22892612	IGLL5	c.356+601G>T
chr12	113077467	DTX1	c.303G>T	chr22	22892799	IGLL5	c.356+788C>A
chr14	68791812	ZFP36L1	c.57+1070G>A				

Data on file.

Table A.5. SNVs in Seraseq® ctDNA Complete™ Mutation Mixes covered by the KAPA HyperCap Design Share NHL Panel[a]

Chromosome	Position	Reference	Alternate	Gene	Mutation	COSMIC ID[b]
chr7	140753336	A	T	BRAF	V600E	COSM476
chr12	25245351	C	A	KRAS	G12C	COSM516
chr12	25245350	C	T	KRAS	G12D	COSM521

[a]All mutations are expected at an AF of 0.05 in the AF5% Mix, and at an AF of 0.005 in the AF0.5% Mix.

[b]https://cancer.sanger.ac.uk/cosmic

Table A.6. SNVs in Twist ctDNA Pan-cancer Reference Standards covered by the KAPA HyperCap Design Share NHL Panel[a]

Chromosome	Position	Reference	Alternate	Gene	Mutation	COSMIC ID[b]
chr3	38141150	T	C	MYD88	L265P	COSM85940
chr3	49375465	T	C	RHOA	Y42C	COSM2849892
[c]chr4	152328233	G	A	FBXW7	R465C	COSM222932
chr7	140753336	A	T	BRAF	V600E	COSM476
chr7	148811635	T	A	EZH2	Y641F	COSM37028
chr10	87933148	G	A	PTEN	R130G	COSM5033
chr12	25245347	C	T	KRAS	G13D	COSM532
chr15	66435117	G	T	MAP2K1	K57N	COSM1235478
chr17	7673802	C	T	TP53	R273H	COSM10660
chr17	7674230	C	A	TP53	G245C	COSM11081
[c]chr17	7675088	C	T	TP53	R175H	COSM10648
[c]chr18	51065549	G	A	SMAD4	R361H	COSM14122

[a]All mutations are expected at an AF of 0.05 in the 5% Standard, and at an AF of 0.005 in the 0.5% Standard.

[b]https://cancer.sanger.ac.uk/cosmic

[c]Candidate reporter variants excluded from longitudinal mutation analysis due to their presence in the blocklist or germline.

Published by:

Roche Sequencing Solutions, Inc.
4300 Hacienda Drive
Pleasanton, CA 94588

sequencing.roche.com

MC--11981 06/2023