Roche KAPA HyperPETE Target Enrichment workflow based on DNBSEQ Sequencing Platform Facilitates Cancer Research

Performance Evaluation for KAPA HyperPETE Pan Cancer Panel and Lung Cancer Fusion Panel on DNBSEQ-G99 Genetic Sequencer

Roche has developed a variety of target enrichment panels based on its proprietary KAPA HyperPETE technology. This study has chosen two representative panels: KAPA HyperPETE Pan Cancer Panel and KAPA HyperPETE Lung Cancer Fusion Panel, to verify the compatibility of KAPA HyperPETE target enrichment workflow on MGI DNBSEQ sequencing platform. The comprehensive analysis showed that the performance of KAPA HyperPETE target enrichment workflow on DNBSEQ-G99 sequencer is outstanding, which is comparable to that of Vendor X platform.

Recommended application: Cancer Genomics

Recommended models: DNBSEQ-G99ARS, DNBSEQ-T7RS, DNBSEQ-G400RS

- ### Superior capture performance and variant calling capability

  Roche KAPA HyperPETE target enrichment workflow can achieve high quality target enrichment within 3 hours and is suitable for various samples, with a wide range of variant calling types.

- ### Perfect compatibility with DNBSEQ sequencing platform

  MGIEasy Universal Library Conversion kit enables KAPA HyperPETE target enrichment workflow perfectly work on DNBSEQ sequencing platform.

- ### Data output is efficient and high-quality

  DNBSEQ sequencing technology has significant features such as high accuracy, low repeat sequence rate, and low index hopping rate. The DNBSEQ-G99 sequencer has a fast speed and built-in computing module, integrating sequencing and bioinformatics. It can complete PE150 sequencing in 12 hours, with efficient and high-quality data output.

- ### Automatic operation compatible

  MGI provides automation solutions for experimental processes, which can greatly save labor costs and improve efficiency.

Background

Cancer is a disease with genetic complexity and heterogeneity. Germline variants play a dominant role in 5-10% of all tumors, while acquired somatic variants are the main factor driving disease progression and responding to therapeutic intervention[1,2]. Every individual cancer patient has unique genetic variation profiles, ranging from single nucleotide variants/polymorphisms (SNVs/SNPs), short insertions and deletions (InDels), microsatellite instability (MSI), copy number variations (CNVs), structural variations (SVs), and large-scale structural rearrangements. SVs leading to gene fusion are common events in solid tumors. Fusion gene refers to chimeric gene formed by the fusion of partial sequences of two genes, usually caused by chromosomal translocation, deletion or other reasons. This chimeric gene can generate aberrant transcripts or proteins in subsequent biological processes, which may cause or promote the occurrence of tumors[1]. Therefore, accurately detecting these variants is of great significance for the prevention, treatment, and comprehensive understanding of tumor diseases.

The rapid development of massively parallel sequencing (MPS) has overturned the public's understanding of cancer and clinical research[3]. This technology has been widely used in cancer research due to its high sensitivity, high throughput, and ability to detect different variations. Researchers can identify mutations in the genome through whole genome sequencing (WGS), whole exome sequencing (WES) and targeted sequencing (TS)[3]. Compared to WGS or WES, targeted sequencing can ensure high accuracy and sensitivity in identifying target regions while having the advantages of lower cost and less data volume[3]. The high sequencing depth of targeted sequencing makes it very suitable for the study of clinical samples, and it can detect mutation sites with variant allele frequencies as low as 0.1~0.2%[4].

Roche has developed a novel target capture technology——KAPA HyperPETE, based on the Primer Extension Target Enrichment (PETE) technology. PETE technology can use primer extension reactions to achieve specific capture and release targeted regions, and prepare targeted library through amplification. The PETE technology greatly shortens the workflow while retaining the capture performance of traditional hybridization and the NGS libraries can be prepared in 1 day. This technology can be used to detect all major somatic variants, including SNVs, short InDels, CNVs, MSI, and fusion genes (containing unknown fusion variants), in DNA or RNA samples extracted from liquid biopsies, FFPE tissues, and cell lines. It has been proven to be very suitable for research applications in small panel tumor studies[5].

Based on the PETE technology, Roche has developed a variety of panels covering hereditary oncology, oncology hotspots, pan-cancer variants (with an MSI module), and lung cancer fusion variants for researchers to choose from. In addition, researchers can also design customized panels based on Roche online design tool——HyperDesign[5].

The MGI's sequencing platform based on DNBSEQ technology has the advantages of high accuracy and sensitivity, ultra-low duplication rate, low index hopping rate, etc[6]. In this study, we have selected the KAPA HyperPETE Pan Cancer Panel and KAPA HyperPETE Lung Cancer Fusion Panel to test the comprehensive performance of KAPA HyperPETE target enrichment workflow on DNBSEQ sequencing platform.

Study description

To evaluate the performance of KAPA HyperPETE target enrichment workflow on MGI's DNBSEQ platform, this study has selected the KAPA HyperPETE Pan Cancer Panel and KAPA HyperPETE Lung Cancer Fusion Panel to test their performance on DNBSEQ-G99. The results showed that KAPA HyperPETE target enrichment workflow can be perfectly compatible with the DNBSEQ sequencing platform. Both panels have high capture efficiency on DNBSEQ platform, and are comparable to Vendor X.

Materials and Methods

Sample preparation

This study adopted commercial standard samples for performance testing and evaluation.

The performance evaluation of KAPA HyperPETE Pan Cancer Panel was carried out using the Seraseq® ctDNA MRD Panel Mix standard (PN: 0710-2146). Three standards with tumor fractions (TF) of 0% (Seraseq -WT), 0.5% (Seraseq -AF05), and 0.05% (Seraseq -AF005) were selected for subsequent experiments.

As to the evaluation of KAPA HyperPETE Lung Cancer Fusion Panel, Seraseq® Compromised FFPE WT (DNA/RNA) RM（PN：0710-1710）was used as a control, total RNA of which was extracted with RNeasy FFPE Kit (50) (QIAGEN PN：73504). Seraseq® FFPE Tumor Fusion RNA v4 Reference Material（PN：0710-0496）was used as experimental group, total RNA of which was extracted with High Pure FFPET RNA Isolation Kit (Roche PN：06650775001). In this study, "RNA-Seraseq WT" means "Seraseq® Compromised FFPE WT (DNA/RNA) RM", "RNA-Seraseq FC" means "Seraseq® FFPE Tumor Fusion RNA v4 Reference Material".

Library preparation and sequencing

The KAPA HyperPETE Pan Cancer Panel used in this study is a 302 kb capture target panel, covering 1321 target regions, 86 cancer-related genes, and 190 MSI loci related to somatic oncology research applications. The KAPA HyperPETE Lung Cancer Fusion Panel is a 18 kb capture target panel, including 17 lung cancer fusion genes and 4 housekeeping genes (as internal controls).

Performance evaluation for KAPA HyperPETE Pan Cancer Panel: All cfDNA libraries were prepared with KAPA HyperPrep Kit, the sample input is 50ng, Pre-PCR was performed for 6 cycles. Then the KAPA HyperPETE Pan Cancer Panel was used for target enrichment (3 plex), Vendor X universal primers were used for Post-PCR step (15 cycles). The detailed operation procedures can be checked in KAPA HyperPETE Somatic Plasma cfDNA Workflow v1.0.

Performance evaluation for KAPA HyperPETE Lung Cancer Fusion Panel: All RNA libraries were prepared with KAPA RNA HyperPrep Kit, the sample input is 10ng, Pre-PCR was performed for 18 cycles. Then, KAPA HyperPETE Lung Cancer Fusion Panel was utilized for target enrichment (2 plex), Vendor X universal primers were used in Post-PCR step (17 cycles). The detailed operation procedures can be checked in KAPA HyperPETE Tissue RNA Fusion Transcript Workflow V1.0.

When library preparation was completed, sequencing was carried out using PE150 strategy on the MGI DNBSEQ-G99 genetic sequencer and Vendor X platform. Among them, the library for sequencing on DNBSEQ-G99 should be converted using MGIEasy Universal Library Conversion Kit (App-A) before sequencing.

Bioinformatic analysis

The bioinformatic analysis process includes: sequencing read quality assessment, adapter trimming, read filtering, mapping against the reference genome, duplicate removal, coverage statistic assessment, variant calling, and variant filtering. The specific workflow can be referred to Roche official website(https://www.n-genetics.com/products/1104/1024/19049.pdf). KAPA HyperPETE Pan Cancer Panel performance evaluation refers to P1-19, VarDict is used for variant calling; KAPA HyperPETE Lung Cancer Fusion Panel performance evaluation refers to P29-50, STAR-Fusion is used for variant calling.

Sample collection	Library preparation and sequencing	Bioinformatics analysis	Result analysis
Standards purchased from Seracare	 KAPA HyperPrep Kit  KAPA HyperPETE Panel  MGIEasy Universal Library Conversion Kit (App-A)  DNBSEQ-G99	https://www.n-genetics. com/products/1104/1024/19049.pdf	Systematic evaluation of KAPA HyperPETE target enrichment workflow

Results

Performance evaluation of KAPA HyperPETE Pan Cancer Panel

Following the sequencing of three libraries (Seraseq-WT, Seraseq-AF05, and Seraseq-AF005) on DNBSEQ-G99 and Vendor X, the data were down-sampled to 110M reads (PE150) for further analysis. The analysis results show that no matter what platform or sample, reads on target rate are all above 87%, suggesting the panel provides high on-target rate and enrichment specificity for the target genes (Figure 1A), and Fold-80 Base Penalty are less than 1.5, indicating superior coverage uniformity of this panel (Figure 1B). The base rate of sequencing depths ≥10×, 50×, and 100× all exceed 99%, indicating this panel has a relatively broad target coverage (Figure 1C). All of the above QC metrics suggest KAPA HyperPETE Pan Cancer Panel is perfectly compatible with DNBSEQ-G99, and the performance of DNBSEQ-G99 is comparable to Vendor X.

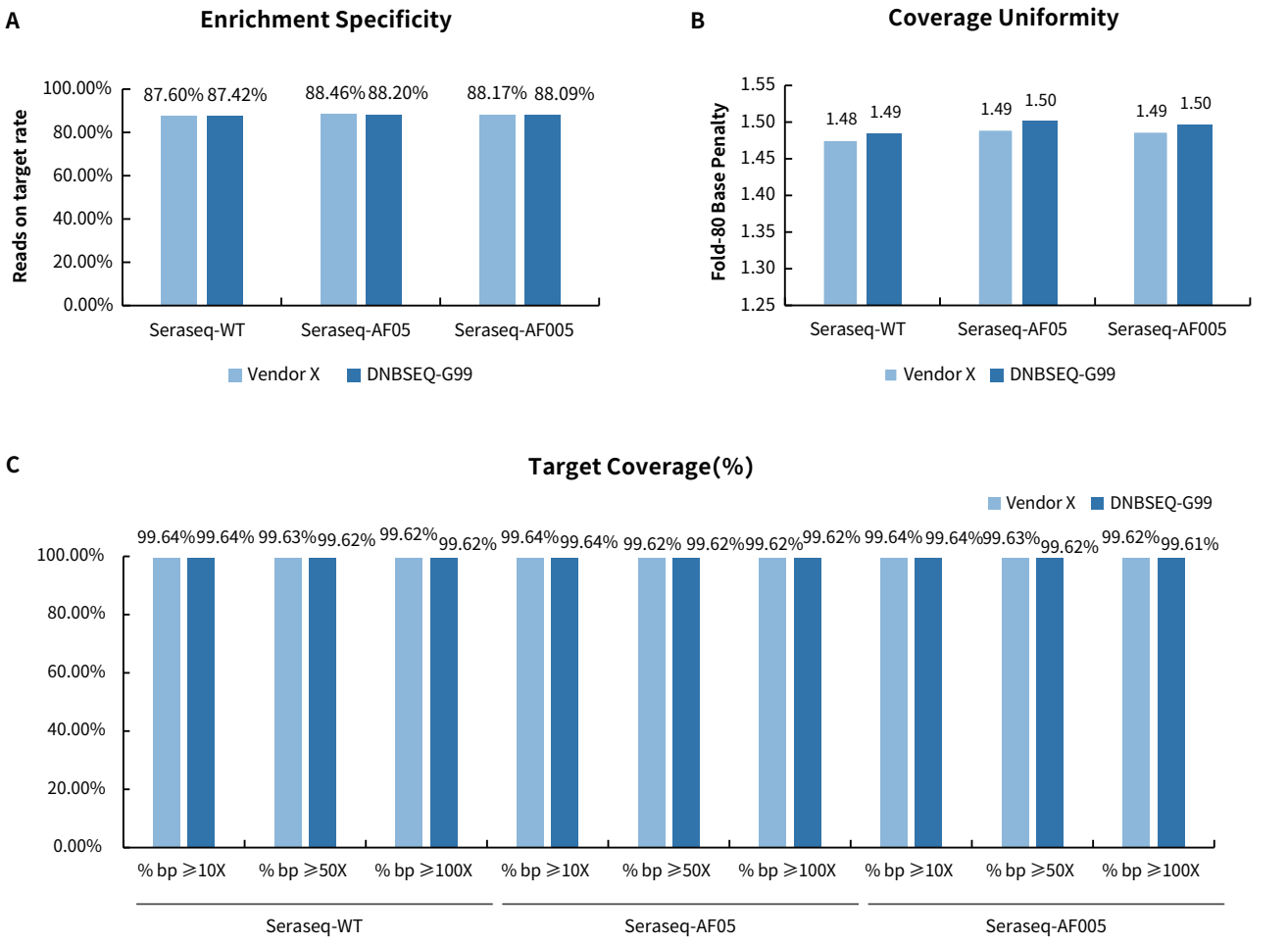

Figure 1. The comparison of QC metrics for KAPA HyperPETE Pan Cancer Panel compatible with DNBSEQ-G99 or Vendor X. (A) Reads on target rate: The specific enrichment rate of target genes; (B) Fold-80 Base Penalty: Coverage uniformity of the panel (closer to 1 is better); (C) Target Coverage: The coverage rate of the target.

For SNVs detection, both platforms were able to detect all variants except *PIK3CA* SNV in Seraseq-AF05. In Seraseq-AF005, Vendor X detected 7 SNVs, while DNBSEQ-G99 detected 6 SNVs. For InDels detection, both DNBSEQ-G99 and Vendor X could detect all of the variants in Seraseq-AF05. However, for Seraseq-AF005, Vendor X detected 3 InDels, while DNBSEQ-G99 detected 4 InDels, more than Vendor X platform. These variant calling results suggest that this panel compatible with both platforms can detect almost all SNVs and InDels in Seraseq-AF005, and the detection performance of both platforms is equivalent (Table 1).

Mutation type	Gene	Mutation information					Vendor X			DNBSEQ-G99		
							Seraseq WT	Seraseq-AF05	Seraseq-AF005	Seraseq WT	Seraseq-AF05	Seraseq-AF005
SNV	*AKT1*	c.49G>A	p.E17K	chr14	105246551	105246551		0.07%			0.26%	0.04%
	ALK	c.3604G>A	p.G1202R	chr2	29443613	29443613		0.18%	0.03%		0.21%	
	ALK	c.3522C>A	p.F1174L	chr2	29443695	29443695		0.20%			0.12%	
	BRAF	c.1799T>A	p.V600E	chr7	140453136	140453136	0.06%	0.56%	0.06%	0.06%	0.67%	0.06%
	EGFR	c.2369C>T	p.T790M	chr7	55249071	55249071	0.02%	0.21%	0.05%	0.02%	0.19%	0.05%
	EGFR	c.2573T>G	p.L858R	chr7	55259515	55259515		0.42%	0.03%		0.40%	
	KIT	c.2447A>T	p.D816V	chr4	55599321	55599321		0.40%	0.05%		0.45%	0.13%
	KRAS	c.183A>C	p.Q61H	chr12	25398285	25398285	0.06%	0.24%	0.04%		0.12%	
	KRAS	c.35G>A	p.G12D	chr12	25398284	25398284		0.34%	0.04%	0.03%	0.33%	0.03%
	KRAS	c.34G>T	p.G12C	chr12	25380275	25380275	0.02%	0.10%			0.33%	
	NRAS	c.182A>G	p.Q61R	chr1	115256529	115256529		0.36%			0.31%	
	PIK3CA	c.3140A>G	p.H1047R	chr3	178936082	178936082						0.04%
InDel	*BRCA1*	c.1961delA	p.K654fs*47	chr17	41245587	41245587		0.44%	0.20%		0.39%	0.17%
	BRCA2	c.7934del	p.R2645Nfs*3	chr13	32936788	32936788		0.20%	0.04%		0.12%	0.02%
	EGFR	c.2235_2249del	p.E746_A750del	chr7	55242465	55242479		0.18%			0.23%	
	EGFR	c.2240_2257del	p.L747_P753delinsS	chr7	55242470	55242487		0.24%	0.03%		0.17%	0.05%
	EGFR	c.2254_2277del	p.S752_I759del	chr7	55242484	55242507		0.25%			0.21%	
	PIK3CA	c.3204_3205insA	p.*1069Mext*3, p.*1069fs	chr3	178952149	178952149		0.15%			0.20%	0.04%
	ERBB2	c.2313_2324dup	p.Y772_A775dup,p.A775_G776insYVMA	chr17	37880981	37880982		0.27%			0.27%	

Table 1. The comparison of SNV and InDel variant calling rates for KAPA HyperPETE Pan Cancer Panel based on DNBSEQ-G99 or Vendor X.

Performance evaluation of KAPA HyperPETE Lung Cancer Fusion Panel

Two samples (RNA-Seraseq WT and RNA-Seraseq FC) were tested on DNBSEQ-G99 and Vendor X, and the data were down-sampled to 4M reads (PE150).Results showed that rRNA reads rates of the two samples on DNBSEQ-G99 is 1.91% and 1.38%, respectively, lower than those on Vendor X platform (Figure 2A), and the rate of reads on target for entire panel and reads on target for fusion genes in design of RNA-Seraseq FC in DNBSEQ-G99 were 94.91% and 57.98%, respectively, which are marginally superior than Vendor X (Figures 2B, C). The two platforms performed comparably for reads on target for housekeeping genes and uniquely mapped reads after rRNA removal (Figures 2D, E). The above results indicated that this panel has a higher on-target rate and the results obtained from DNBSEQ-G99 are marginally superior than Vendor X.

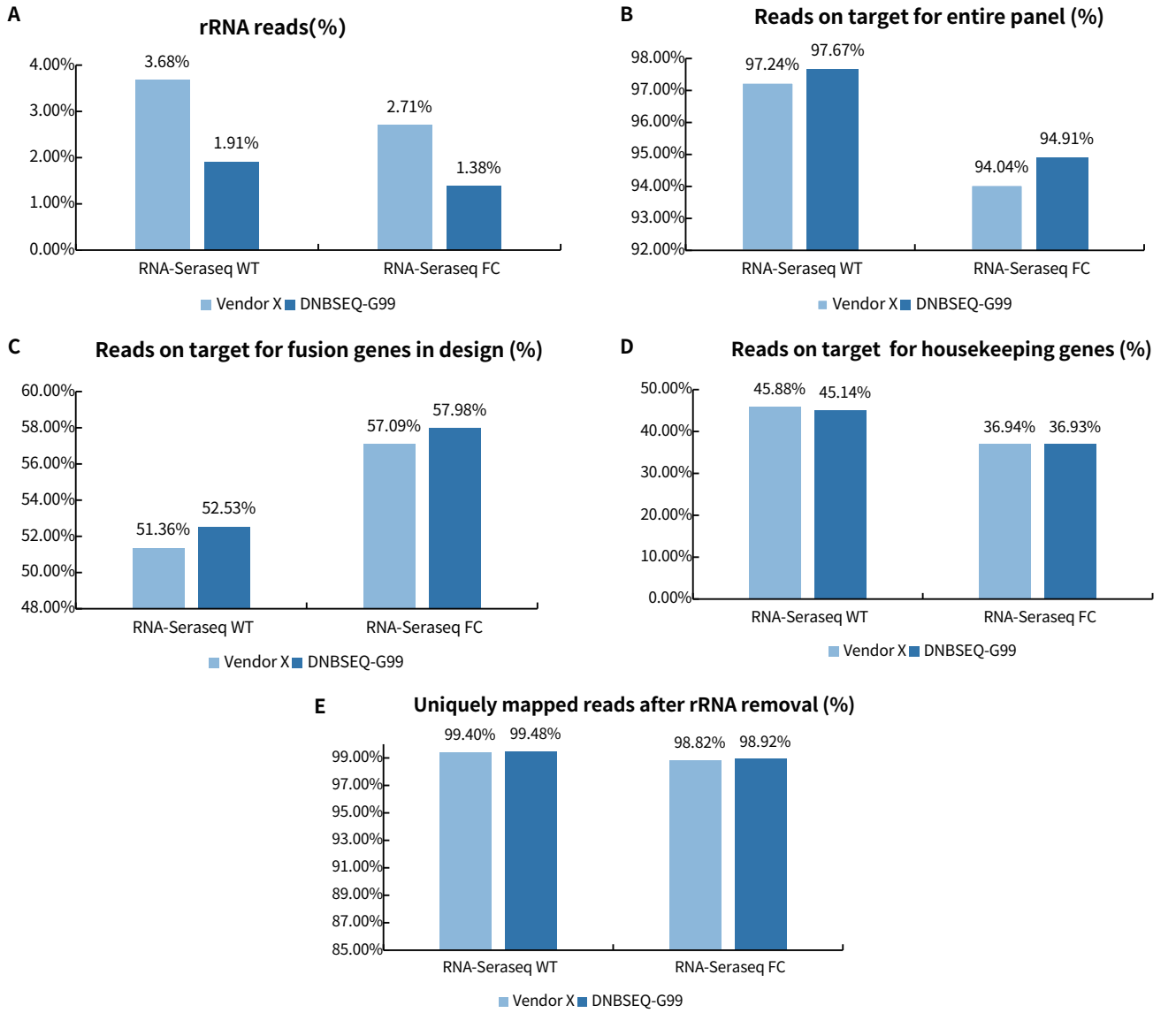

Figure 2. The comparison of QC metrics for KAPA HyperPETE Lung Cancer Fusion Panel based on DNBSEQ-G99 or Vendor X. (A) Ratio of rRNA reads. Before calculating reads on-target rate, the reads mapping to rRNA are removed. (B,C,D) On-target rate for housekeeping genes (positive enrichment control) = (on-target rate for entire panel)-(on-target rate for fusion genes). (E) Percentage of uniquely mapped reads after rRNA removal.

Variant calling analysis of RNA-Seraseq FC showed that there are 18 fusion variants in RNA-Seraseq FC, and the main genes of 17 fusion variants are covered by this panel. These 17 variants are detected in both DNBSEQ-G99 and Vendor X (Table 2).

RNA Fusion	5' Partner	3' Partner	HGVS Name	Vendor X		DNBSEQ-G99	
				RNA-SeraSeq WT	RNA-Seraseq FC	RNA-Seraseq WT	RNA-Seraseq FC
EML4--***ALK***	*EML4* ex 13	*ALK* ex 20	EML4{NM_019063.4}:r.1_1763_ALK{NM_004304.4}:r.4125_6265	N	Y	N	Y
CD74--***ROS1***	*CD74* ex 6	*ROS1* ex 34	CD74{NM_001025159.2}:r.1_812_ROS1{NM_002944.2}r.5757_7368	N	Y	N	Y
SLC34A2--***ROS1***	*SLC34A2* ex 4	*ROS1* ex 34	SLC34A2{NM_006424.2}:r.1_460_ROS1{NM_002944.2}:r.5757_7368	N	Y	N	Y
CCDC6--***RET***	*CCDC6* ex 1	*RET* ex 12	CCDC6{NM_005436.5}:r.1_435_RET{NM_020975.6}:r.2327_5617	N	Y	N	Y
KIF5B--***RET***	*KIF5B* ex 24	*RET* ex 11	KIF5B{NM_004521.2}:r.1_3231_RET{NM_020975.6}:r.2070_5617	N	Y	N	Y
NCOA4--***RET***	*NCOA4* ex 8	*RET* ex 12	NCOA4{NM_001145260.1}:r.1_1014_RET{NM_020975.6}:r.2327_5617	N	Y	N	Y
EGFR Variant III	*EGFR* ex 1	*EGFR* ex 8	EGFR{NM_005228.5}:r.350_1150del	N	Y	N	Y
EGFR--*SEPT14*	*EGFR* ex 24	*SEPT14* ex 10	EGFR{NM_005228.5}:r.1_3207_SEPT14{NM_207366.3}:r.1200_3752	N	Y	N	Y
LMNA--***NTRK1***	*LMNA* ex 2	*NTRK1* ex 10	LMNA{NM_170707.3}:r.1_762_NTRK1{NM_001012331.1}:r.1290_2647	N	Y	N	Y
TFG--***NTRK1***	*TFG* ex 5	*NTRK1* ex 9	TFG{NM_006070.5}:r.1_851_NTRK1{NM_001012331.1}:r.1234_2647	N	Y	N	Y
TPM3-***NTRK1***	*TPM3* ex 7	*NTRK1* ex 9	TPM3{NM_153649.3}:r.1_794_NTRK1{NM_001012331.1}:r.1234_2647	N	Y	N	Y
ETV6--***NTRK3***	*ETV6* ex 5	*NTRK3* ex 15	ETV6{NM_001987.4}:r.1_1283_NTRK3{NM_001012338.2}:r.1892_3004	N	Y	N	Y
FGFR3--*BAIAP2L1*	*FGFR3* ex 17	*BAIAP2L1* ex 2	FGFR3{NM_000142.4}:r.1_2530_BAIAP2L1{NM_018842.4}:r.315_3682	N	Y	N	Y
FGFR3--*TACC3*	*FGFR3* ex 17	*TACC3* ex 11	FGFR3{NM_000142.4}:r.1_2530_TACC3{NM_006342.3}:r.2066_2799	N	Y	N	Y
MET ex 14 Skipping	*MET* ex 13	*MET* ex 15	MET{NM_001127500.3}:r.3338_3478del	N	Y	N	Y
PAX8--***PPARG1***	*PAX8* ex 9	*PPARG1* ex 3	PAX8{NM_003466.4}:r.1_1253_PPARG{NM_138712.3}:r.246_1892	N	Y	N	Y
SLC45A3--***BRAF***	*SLC45A3* ex 1	*BRAF* ex 8	SLC45A3{NM_033102.3}:r.1_109_BRAF{NM_004333.5}:r.1206_4560	N	Y	N	Y
*TMPRSS2-ERG**	*TMPRSS2* ex 1 (5' UTR)	*ERG* ex 2	TMPRSS2{NM_005656.3}:r.1_78_ERG{NM_004449.4}:r.124_5042	N	N	N	N

Table 2. The comparison of fusion variants detection using KAPA HyperPETE Pan Cancer Panel based on DNBSEQ-G99 or Vendor X. The bolded words are the genes covered by the panel, while the non-bolded words are not covered. * indicates that this panel doesn't cover the relevant genes of the fusion variants.

Summary

The performance evaluation results of KAPA HyperPETE Pan Cancer Panel and KAPA HyperPETE Lung Cancer Fusion Panel on DNBSEQ and Vendor X platforms show that Roche KAPA HyperPETE target enrichment workflow is compatible with the DNBSEQ-G99 platform --ideal results can be obtained from different panels and sample types. KAPA HyperPETE panels designed based on Roche PETE technology are compatible with DNBSEQ-G99 platform, has excellent ability to detect variation and is comparable to Vendor X platform.

DNBSEQ-G99, based on DNBSEQ™ patented technology, can stably obtain high-quality sequencing data. As one of the fastest models among global sequencing instruments with equivalent throughput, it is particularly suitable for targeted gene sequencing. The throughput of a single DNBSEQ-G99 flow cell is 80M reads, and up to two flow cells can be run simultaneously. The full PE150 sequencing process can be completed twice in one day, maximizing sequencing efficiency. The built-in computing module integrates sequencing and bioinformatics, resulting in efficient and high-quality data output. The sequencing operation is simple and easy to use.

In KAPA HyperPETE rapid target enrichment workflow, DNA process only takes 6 hours and RNA process only takes 8 hours. Combined with DNBSEQ-G99 rapid sequencing, wet lab of NGS can be completed in a single day, helping to speed up NGS in an overall way.

DNBSEQ-G99ARS Genetic Sequencer

References

1. Schram, A. M., Chang, M. T., Jonsson, P. & Drilon, A. Fusions in solid tumours: diagnostic strategies, targeted therapy, and acquired resistance. *Nat Rev Clin Oncol* **14**, 735-748, doi:10.1038/nrclinonc.2017.127 (2017).
2. National Cancer Institute. The Genetics of Cancer. *https://www.cancer.gov/about-cancer/causes-prevention/genetics. Accessed 22 October 2021.*
3. Bewicke-Copley, F., Arjun Kumar, E., Palladino, G., Korfi, K. & Wang, J. Applications and analysis of targeted genomic sequencing in cancer studies. *Comput Struct Biotechnol J* **17**, 1348-1359, doi:10.1016/j.csbj.2019.10.004 (2019).
4. Shin, H. T. *et al.* Prevalence and detection of low-allele-fraction variants in clinical cancer samples. *Nat Commun* **8**, 1377, doi:10.1038/s41467-017-01470-y (2017).
5. https://sequencing.roche.com/us/en/products/product-category/primer-extension-based-target-enrichment.html#overview.
6. Porreca, G. J. Genome sequencing on nanoballs. *Nat Biotechnol* **28**, 43-44, doi:10.1038/nbt0110-43 (2010).

Recommended Ordering Information

Category	Product	Cat. NO.
Instruments	Genetic Sequencer DNBSEQ-G99ARS	900-000609-00
	MGISP-100RS Automated Sample Preparation System	900-000206-00
	MGISP-960RS Automated Sample Preparation System	900-000146-00
Library Prep	MGIEasy Universal Library Conversion kit (App-A)	1000004155
	KAPA HyperPrep Kit	07962347001*
	KAPA HyperPure Beads	08963835001*
	KAPA Universal UMI Adapter	09329862001*
	KAPA UDI Primer Mixes, 1 – 96	09134336001*
	KAPA RNA HyperPrep Kit	08098093702*
	KAPA HyperPETE Pan Cancer Panel	09329161001*
	KAPA HyperPETE LC Fusion Panel	09329471001*
	KAPA HyperPETE Reagent Kit	09211624001*
	KAPA HyperCapture Bead Kit	09075780001*
Sequencing Reagents	High-throughput Sequencing Set (G99 SM App-C FCL PE150)	940-000413-00

*The relevant products are available and can be ordered on the Roche official website

MGI Tech Co.,Ltd

Building 11, Beishan Industrial Zone, Yantian District, Shenzhen, CHINA, 518083

Version: August 2023

+86-4000-688-114
en.mgi-tech.com
MGI-service@mgi-tech.com
Author: Hanfei Zhang
Editor-in-Charge: Qiwei Wang
Reviewer: Yao Jiang

1. For StandardMPS and CoolMPS: Unless otherwise informed, StandardMPS and CoolMPS sequencing reagents, and sequencers for use with such reagents are not available in Germany, Spain, UK, Sweden, Belgium, Italy, Finland, Czech Republic, Switzerland, Portugal, Austria and Romania. Unless otherwise informed, StandardMPS sequencing reagents, and sequencers for use with such reagents are not available in Hong Kong. No purchase orders for StandardMPS products will be accepted in the USA until after January 1, 2023.
2. For HotMPS sequencers: This sequencer is only available in selected countries, and its software has been specially configured to be used in conjunction with MGI's HotMPS sequencing reagents exclusively.
3. For HotMPS reagents: This sequencing reagent is only available in selected countries.