Singlera MethylTitan High-Throughput Targeted Methylation Library Preparation Solution and MGI DNBSEQ Platform Enabled Cancer Detection

Comparison of the bisulfite-targeted sequencing performance between MGI's DNBSEQ-G400 and Illumina's NovaSeq 6000 sequencing platform

Singlera compared the performance of MGI's DNBSEQ-G400 sequencer and Illumina's NovaSeq 6000 platform in bisulfite-targeted sequencing. The results showed that the DNBSEQ-G400 sequencing platform was comparable to the NovaSeq 6000 platform in terms of sequencing quality, consistency of methylation level detection, sensitivity of cancer signal detection, and clinical accuracy, and could be widely used in clinical methylation research.

The related research results were published in an article entitled "Cross-platform comparisons for targeted bisulfite sequencing of DNBSEQ-G400 and NovaSeq 6000" on the *Clinical Epigenetics* in 2023[1].

Recommended Application: DNA Targeted Methylation Sequencing
Recommended Models: DNBSEQ-G400RS, DNBSEQ-T7RS

- **Singlera mTitan high-throughput targeted methylation sequencing technology can specifically detect multiple methylated regions**

  Singlera's proprietary specific methylation marker detection method and liquid biopsy technology can detect 10^2-10^3 DNA methylation regions in depth

- **Singlera mTitan library protocol is perfectly compatible with DNBSEQ platform**

  MGI's universal library conversion solution enables mTitan targeted methylation sequencing solution to be perfectly adapted to the DNBSEQ platform

- **Efficient and accurate sequencing quality**

  The unique DNBSEQ™ technology is characterized by high accuracy, low duplication rate and low index hopping rate.

- **Compatible with automated operations**

  MGI can provide automated solutions for the experimental process, which can greatly save labor costs and improve efficiency.

Background

DNA methylation is an important regulatory factor of gene expression and plays a key role in regulating various physiological and pathological regulatory processes[2]. The role of DNA methylation in the development and progression of cancer and in the assessment of drug resistance to targeted therapies has been widely studied[3]. Cell free DNA (cfDNA) are DNA fragments with a length of about 170 bp that is released into the blood after apoptosis and disintegration of tissue cells. DNA fragments released by tumor cells in the blood are known as circulating tumor DNA (ctDNA)[2]. An increasing number of studies have shown that aberrant ctDNA methylation may be associated with cancer development and progression, and the elevated level of methylation of tumor suppressor genes is an early event in many tumors, making ctDNA methylation status a feasible biomarker for clinical noninvasive cancer detection[1,4]. Compared with detection of ctDNA mutations, methylation status analysis has the following advantages: higher detection sensitivity and detection range, as evidenced by the fact that there are more methylation target intervals in patients and more variable CpG methylation sites can be present within the same target interval, etc.[5]. Bisulfite sequencing is currently the golden standard for methylation detection, which can detect DNA methylation status at single-base resolution. However, the application of this method in clinical noninvasive diagnosis faces some limitations. The bisulfite reaction can lead to DNA degradation and library sequence imbalance, resulting in low sequencing data output, poor quality and high error rate[1]. It is therefore necessary to evaluate the quality and performance of the sequencing platform used in bisulfite sequencing before formal clinical test.

In recent years, MGI has launched a series of genetic sequencers based on DNA Nanoballs (DNB) rolling circle amplification and combined primer anchor synthesis (cPAS) technology, including DNBSEQ-G400RS, DNBSEQ-T7RS, DNBSEQ-G99RS, etc.[1,6] . With its high-quality data output performance, MGI has become a tier 1 supplier of high-throughput sequencers. DNB rolling circle amplification has the following advantages over bridge PCR amplification: DNB rolling circle amplification adopts the linear amplification mode, which always uses the same original DNA as the template, avoids the accumulation of errors caused by exponential amplification, and reduces the probability of index hopping and coverage preference. A recent comparative study on different platforms has shown that MGI's DNBSEQ sequencing platform is comparable to Illumina's in terms of targeted sequencing, whole genome sequencing (WGS), whole exome sequencing (WES), RNA-seq, single cell RNA-seq (scRNA-seq), and metagenomic sequencing[1].

In this study, the performance of the DNBSEQ platform in DNA methylation sequencing was evaluated in all aspects, using the Illumina NovaSeq 6000 sequencer as control. During the study, libraries were constructed from simulated cfDNA standards and clinical pancreatic ductal adenocarcinoma (PDAC) cfDNA samples and sequenced on DNBSEQ-G400 to comprehensively evaluate its sequencing quality, detection of methylation levels, cancer signal detection ability, and clinical diagnostic accuracy.

Materials and Methods

DNA extraction and sample preparation

The PDAC FFPE tissue samples and 24 clinical plasma samples (12 preoperative PDAC plasma samples and 12 healthy control samples) used in this study were purchased from ProteoGenex. The universal methylated DNA standard (meDNA) and NA12878 standard were purchased from Zymo and Coriell, respectively. QIAamp Circulating Nucleic Acid Kit (Qiagen) was used to extract cfDNA from clinical plasma samples according to the procedures described in the relevant instructions. Reliaprep FFPE gDNA Miniprep System (Promega) was used to extract gDNA from FFPE samples.

The gDNA, meDNA and NA12878 gDNA of PDAC samples were fragmented using Bioruptor NGS. Simulated cfDNA samples were prepared by mixing fragmented PDAC gDNA with NA12878 gDNA at a certain ratio, and diluted to samples with a tumor fraction (TF) of 0%, 0.1%, 0.5%, 1%, 5% and 10%, respectively.

Library preparation and sequencing

The library of simulated cfDNA samples was constructed according to the MethylTitan protocol of Singlera, and two libraries were prepared for each sample based on the MGI associated kit (phosphorylated i5 sequencing primers) and Illumina associated kit (i5 sequencing primers) and were named as "mLib" and "iLib", respectively. The mLib libraries were used for PE150 sequencing on both MGI DNBSEQ-G400 and Illumina NovaSeq 6000 sequencing platforms, while the iLib libraries were used for PE150 sequencing on Illumina NovaSeq 6000 only. Circularized WGS libraries and PhiX libraries were used as balance libraries for library sequencing on DNBSEQ-G400 platform and NovaSeq 6000 platform, respectively. "mLib-MGISEQ" is the dataset obtained by sequencing the mLib library on the DNBSEQ-G400 sequencing platform, which also applies to other names.

Data processing

The output data from DNBSEQ-G400 sequencer were pre-processed for base identification using Zebre call, a base identification software independently developed by MGI, and then converted into a format consistent with that of the Illumina data using FastQC software (v0.11.7). The PE150 reads were spliced into single-ended reads using pear software (v0.9.6), with the analysis parameters set to "-j 4 -v 20 -t 30 -n30". Splices and low-quality bases were removed using the default parameters of trim_galore (v0.4.0). UMI information in the reads was extracted and sequencing data were aligned to the hg19 reference genome using bismark software (v0.17.0), with analysis parameters set to "-bowtie2 -l 32 -n 1 -non_directional". Only on-targeted reads were retained for subsequent analysis. PCR duplicate amplified sequences were filtered using the umi_tools (version 1.1.2) based on UMI information. The sequencing data with PCR repeats removed were used to calculate the following quantitative indexes: average methylation fractions (AMF), methylation haplotype fractions (MHF), and methylation haplotype loads (MHL) of the target region. Please refer to the text of Reference 1 for the quantification formula.

Sample collection	Library preparation and sequencing	Bioinformatics analysis	Analysis of results
Clinical plasma cfDNA samples, simulated cfDNA samples	Singlera MethylTitan high-throughput targeted methylation library preparation protocol  MGIEasy Universal Library Conversion Kit (App-A)  DNBSEQ-G400 Genetic Sequencer	AMF, MHF, MHL quantitative analysis, LOD analysis, PDACatch	Systematic evaluation of bisulfite-targeted sequencing performance of the DNBSEQ-G400 platform

Results

DNBSEQ-G400 sequencer can provide high-quality targeted methylation sequencing data

In this study, fully methylated gDNA (meDNA) was diluted into NA12878 gDNA at ratios of 0, 0.002, 0.01, 0.02, and 0.05 to prepare libraries. Then, WGS balance libraries were mixed at different ratios (50%, 30%, 10% and 0%) on the DNBSEQ-G400 sequencing platform for PE150 sequencing (Figure 1A).

Except the library(0% WGS) yielded slightly lower sequencing data (375 M), the outputs of other libraries met the expected standard, and the proportion of unbarcoded data decreased with the increase in the proportion of balance libraries (Figure 3B). As sequencing quality, high-quality reads (Phred score ≤ 30) ratio in the 0% WGS library was significantly lower than other libraries (Figure 3C). The sequencing error rate of 30% WGS library was slightly lower than others (Figure 3D).

In addition, this study also analyzed the consistency of the average methylation fraction (AMF) between the methylated libraries and the target regions in libraries with different ratios of WGS. The study found that the correlation coefficient of those two libraries was 0.999, indicating very high consistency (Figure 1E). Principal component analysis (PCA) of the AMF revealed that the PC1 fluctuated with the proportion of meDNA, while the variations of PC2 and PC3 mainly reflected the deviation of library preparation (Figure 1F). This result suggested that the variation in sequencing was small than the variation in library preparation. In addition, in terms of quantitative accuracy of methylation sequencing, both the estimated meDNA ratios and the expected meDNA ratios had good consistency (R^2 = 0.95) (Figure 3G).

The above results indicate that for the DNBSEQ-G400 platform, the mixing ratio of balance libraries mainly affects the data output, sequencing quality, and sequencing error rate, and has almost no effect on the quantification accuracy and consistency of the final methylation level.

Table1 The summary QC of the data of MGISEQ-2000

WGS library ratio (%)	Total data (M)	BS Data (M)	High-quality reads ratio (%)	Mapping ratio (%)	On-target ratio (%)	Uniformity ratio (%)
50	405.96	160.38	78–84	50–61	72–84	55–59
30	429.89	272.31	80–85	50–62	72–84	55–59
10	396.09	315.93	79–85	50–61	72–84	55–59
0	359.59	312.14	74–81	50–61	72–84	55–58

Table 1. Sequencing data quality of DNBSEQ-G400.

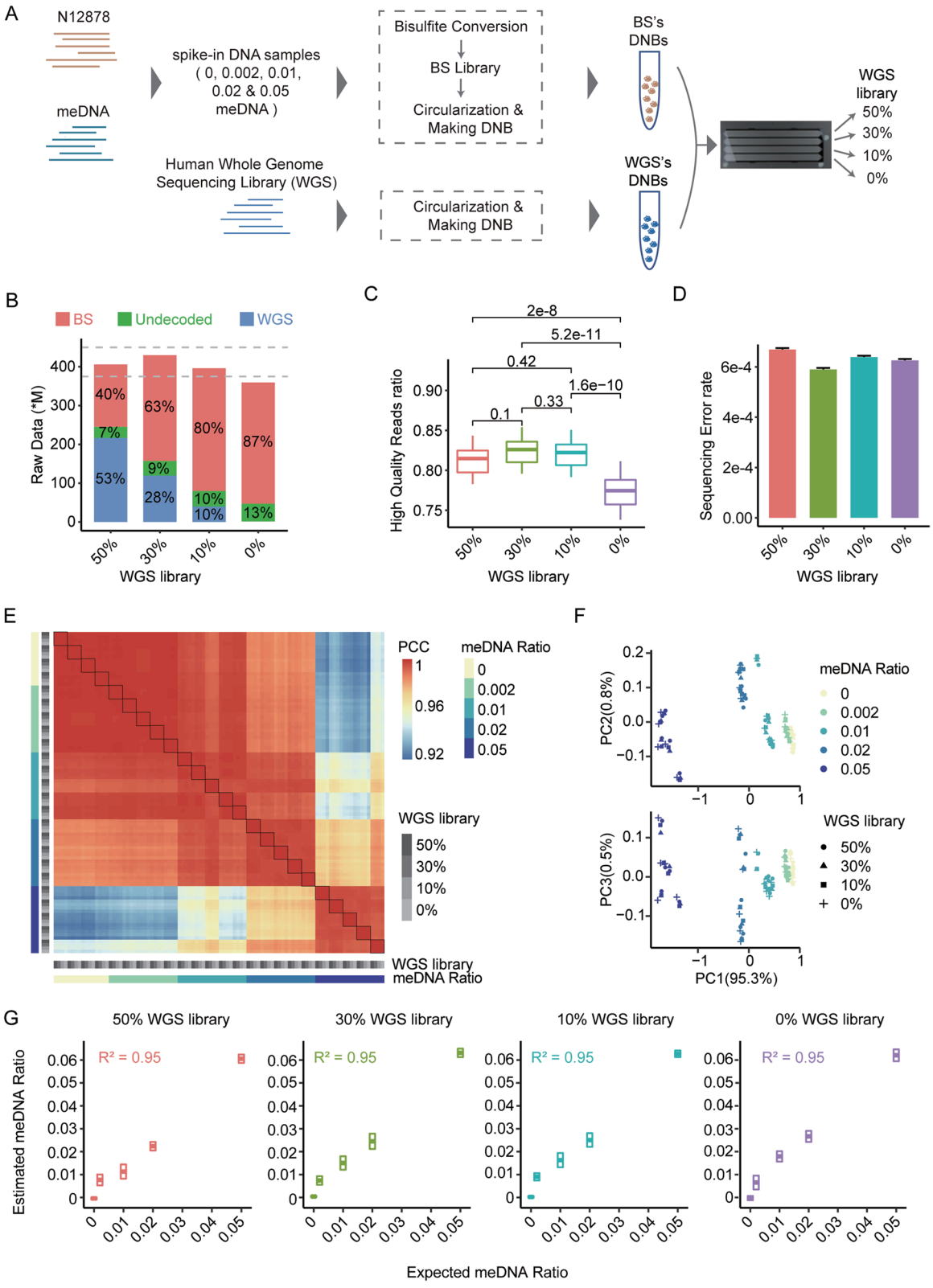


Figure 1. DNBSEQ-G400 sequencer exhibited qualified performance in targeted methylation sequencing.

The sequencing data quality of DNBSEQ-G400 was comparable to NovaSeq 6000

In this study, quality analysis was performed to three datasets, mLib-MGISEQ, iLib-NovaSeq, and mLib-NovaSeq, to compare the data quality of methylation sequencing between DNBSEQ-G400 and NovaSeq 6000 platforms. The results showed that the three datasets were comparable in terms of Q30, mapping ratio and on-target ratio (Table 2). The sequencing error rate of mLib-MGISEQ was significantly lower than that of mLib-NovaSeq, but was comparable to that of iLib-NovaSeq, indicating that better effects could be obtained when using the same manufacturer's library preparation kits and sequencing platforms (Figure 2B). In addition, the size distribution of inserted fragments varied greatly between mLib-MGISEQ and iLib-NovaSeq sequencers (Figure 2C), which was consistent with the previous report. The above results indicate that the data quality of the DNBSEQ-G400 sequencing platform is comparable to that of the NovaSeq 6000 sequencing platform.

The sequencing data quality of DNBSEQ-G400 was comparable to NovaSeq 6000

In this study, quality analysis was performed to three datasets, mLib-MGISEQ, iLib-NovaSeq, and mLib-NovaSeq, to compare the data quality of methylation sequencing between DNBSEQ-G400 and NovaSeq 6000 platforms. The results showed that the three datasets were comparable in terms of Q30, mapping ratio and on-target ratio (Table 2). The sequencing error rate of mLib-MGISEQ was significantly lower than that of mLib-NovaSeq, but was comparable to that of iLib-NovaSeq, indicating that better effects could be obtained when using the same manufacturer's library preparation kits and sequencing platforms (Figure 2B). In addition, the size distribution of inserted fragments varied greatly between mLib-MGISEQ and iLib-NovaSeq sequencers (Figure 2C), which was consistent with the previous report. The above results indicate that the data quality of the DNBSEQ-G400 sequencing platform is comparable to that of the NovaSeq 6000 sequencing platform.

Table2 The summary QC of the data of MGISEQ-2000 and NovaSeq6000

Data type	Library type	Sequencer	High-quality reads ratio (%)	Mapping ratio (%)	On-target ratio (%)	Uniformity ratio (%)
mLib-MGISEQ	mLib	MGISEQ-2000	86–88	46–49	75–79	56–60
iLib-NovaSeq	iLib	NovaSeq6000	86–89	49–52	70–75	59–64
mLib-NovaSeq	mLib	NovaSeq6000	86–88	48–52	72–75	59–64

"Library Type" represented the kit used to prepare libraries. "Sequencer" represented the sequencer to generate data. High-quality reads ratio" demonstrated the ratio of high-quality reads (phred > 30). "Mapping Ratio" represented the ratio of reads that aligned to human genome. 'On-target Ratio' represented the ratio of mapping reads which were amplified by panel primers and located in targeted genome regions. 'Uniformity Ratio' demonstrated the uniformity of panel targeted priming, which was calculated using the ratio of CpGs whose coverages were larger than 25% median coverage. The values before "–" represented the minimum values, while those after "–" represented the maximum values

Table 2. Comparison of data quality between DNBSEQ-G400 and NovaSeq 6000 sequencers.

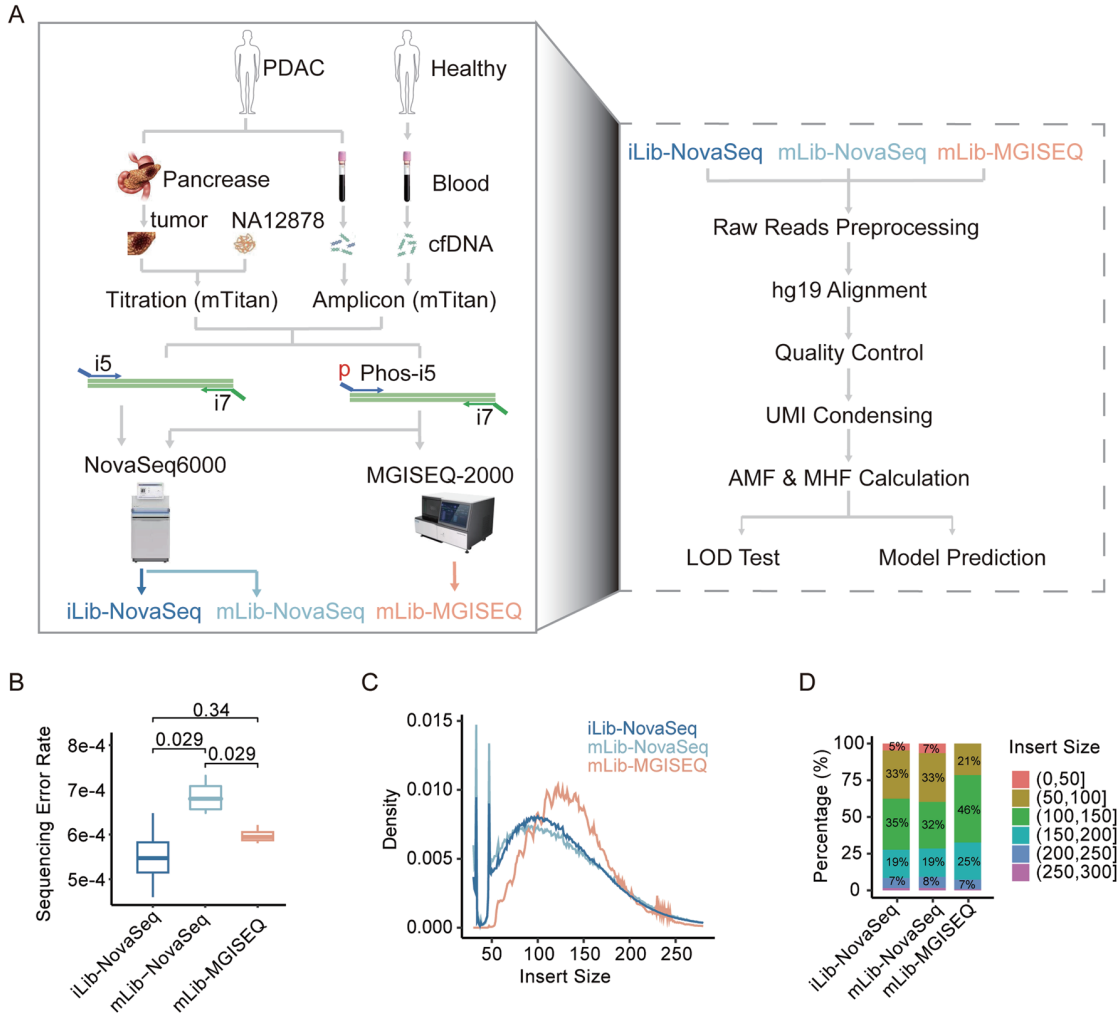


Figure 2. The data quality of DNBSEQ-G400 was comparable to that of NovaSeq 6000.

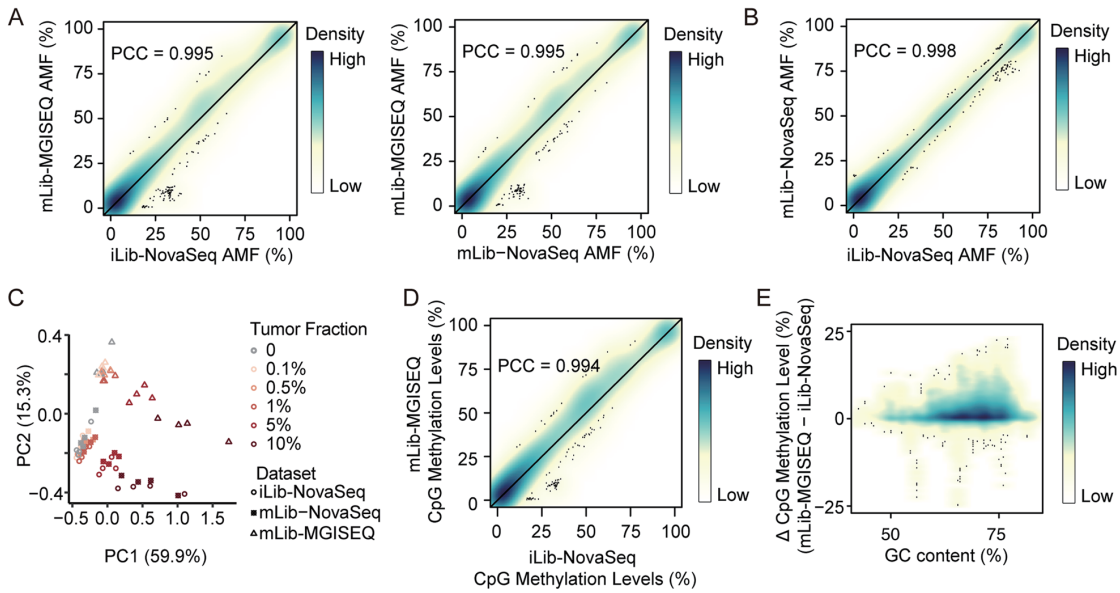


Figure 3. The ability of DNBSEQ-G400 to detect methylation levels is highly consistent with NovaSeq 6000.

The sensitivity of DNBSEQ-G400 to detect cancer signal was comparable to NovaSeq 6000

In order to test the detection sensitivity of different platforms, the limit of detection (LOD) of sequencing data was analyzed in this study. It was found that for simulated cfDNA samples with a tumor fraction of 0.1%, DNBSEQ-G400 was also able to detect cancer signals as significantly as NovaSeq 6000 (Figure 4A).

The markers were detected for the dataset of simulated cfDNA samples with a tumor fraction of 10%.The results showed that 74% of the detected markers in the three datasets of iLib-NovaSeq, mLib-NovaSeq, and mLib-MGISEQ were consistent, and only 5% of the detected markers had a batch effect (Figure 4B).

The above results indicate that DNBSEQ-G400 is comparable to NovaSeq 6000 in detecting cancer signals, and has high potential for clinical application.

DNBSEQ-G400 demonstrated comparable performance to NovaSeq 6000 in clinical applications

To evaluate the performance of DNBSEQ-G400 in clinical applications, this study constructed "iLib" and "mLib" methylation libraries and performed sequencing of 12 cfDNA samples from healthy people and 12 cfDNA samples from PDAC patients, respectively. The results showed that the iLib and mLib libraries had similar size distributions, with lengths of about 200-600 bp and main peaks close to 320 bp. Different from the sequencing results of simulated cfDNA samples, the fragment sizes of mLibs obtained from clinical cfDNA were comparable to those of iLibs (Figures 5A,B).

Subsequently, the overall systematic deviations of the methylation levels of AMF and individual CpG sites in the target region for mLib-MGISEQ and iLib-NovaSeq data were analyzed and

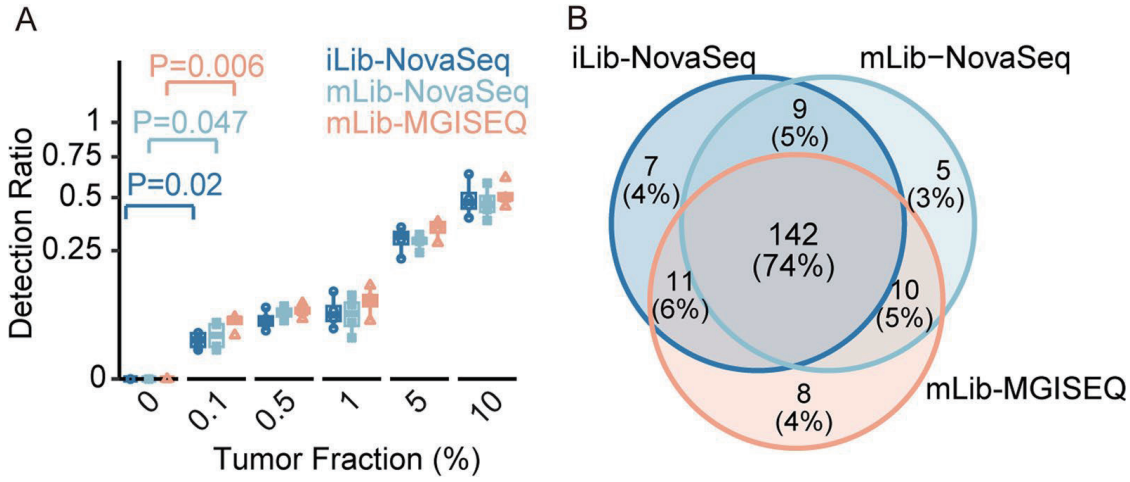
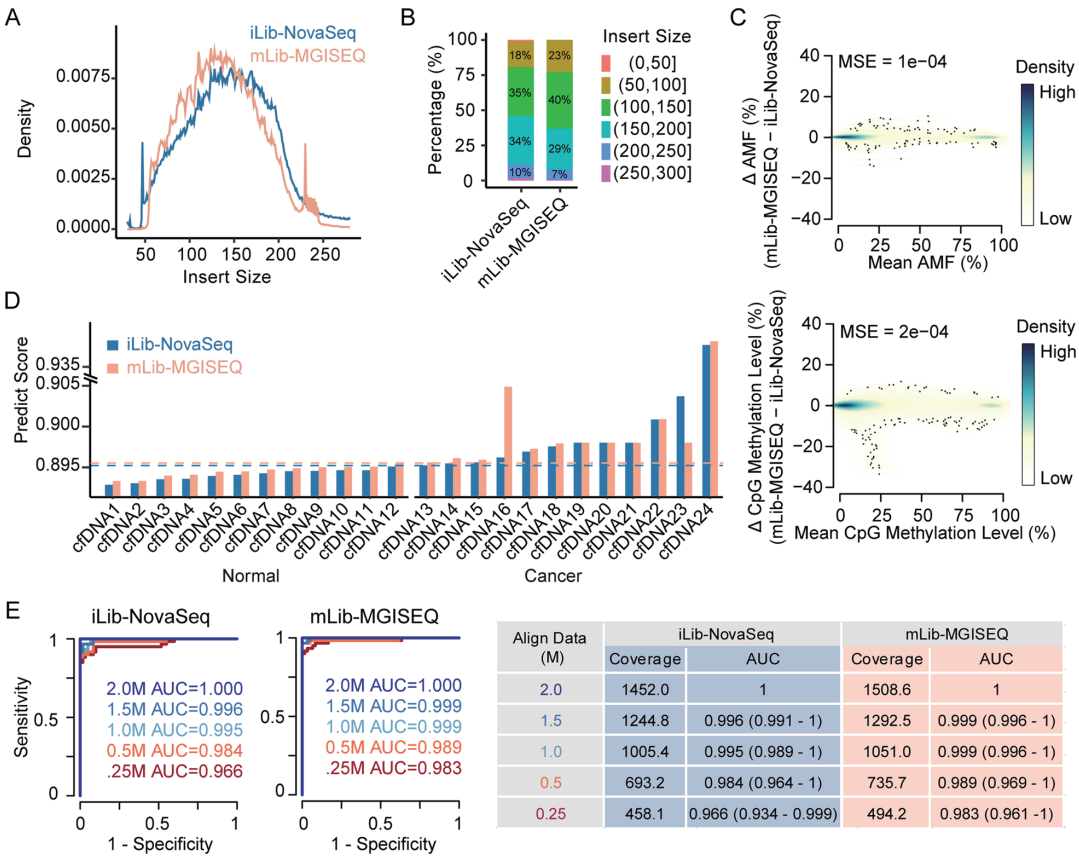


Figure 4. DNBSEQ-G400 displayed comparable cancer signal detection ability to NovaSeq 6000 at a tumor fraction of 0.1%.

compared. The result showed that the Pearson correlation coefficients for AMF and CpG methylation levels were 0.999 and 0.998, respectively. The mean squared error (MSE) of AMF measured by two platforms was as low as 1e-04, and the MSE of CpG methylation level was as low as 2e-04 (Figure 5C).

PDACatch was used to predict the predict scores of individual samples. The results showed that out of the 24 cfDNA samples, only two samples (cfDNA16 and cfDNA23) showed a significant difference in predict scores between mLib-MGISEQ and iLib-NovaSeq, while the predict scores of the other 22 samples were very close to each other (Figure 5D). In addition, both platforms were able to distinguish cfDNA of PDAC patients from healthy people with 100% accuracy at two very close thresholds (0.8952 for iLib-Novaseq and 0.8956 for mLib-MGISEQ) (Figure 5E). 2 M, 1.5 M, 1.0 M, 0.5 M and 0.25 M of data were intercepted from the aligned data of the sequenced data of each platform to evaluate the effectiveness of the PDCA classifier in the test samples. The results showed that the PDACatch classifier performed more efficiently for the analysis of mLib-MGISEQ data than iLib-NovaSeq data.

The above results indicate that the DNBSEQ-G400 platform is highly accurate and performs well in clinical cfDNA-targeted methylation sequencing, with very small systematic difference from NovaSeq 6000.



Align Data (M)	iLib-NovaSeq		mLib-MGISEQ	
	Coverage	AUC	Coverage	AUC
2.0	1452.0	1	1508.6	1
1.5	1244.8	0.996 (0.991 - 1)	1292.5	0.999 (0.996 - 1)
1.0	1005.4	0.995 (0.989 - 1)	1051.0	0.999 (0.996 - 1)
0.5	693.2	0.984 (0.964 - 1)	735.7	0.989 (0.969 - 1)
0.25	458.1	0.966 (0.934 - 0.999)	494.2	0.983 (0.961 -1)

Figure 5. The performance of DNBSEQ-G400 in clinical applications was comparable to NovaSeq 6000.

Conclusion

In order to study the performance of MGI DNBSEQ sequencing platform in targeted methylation sequencing, this study systematically compared the targeted methylation sequencing data of DNBSEQ-G400 and NovaSeq 6000 sequencing platforms. The results showed that the performance of DNBSEQ-G400 was comparable to NovaSeq 6000 in terms of sequencing quality, consistency of methylation levels, cancer signal detection rate, and accuracy of clinical diagnosis results. This indicates that the DNBSEQ-G400 platform can be used for the detection of clinical DNA methylation status, especially the detection of cfDNA methylation level, and has a broad application prospect in the clinical fields of early tumor screening, auxiliary diagnosis, and condition monitoring.

As an all-round desktop sequencer, DNBSEQ-G400 has flexible throughput, supports 1-2 slides per run, independent operation of different sizes of slides (FCS small slides, FCL large slides), and a variety of sequencing read lengths, can comprehensively satisfy a wide range of sequencing needs, and is one of the preferred models for large and medium-sized sequencing laboratories.

Singlera MethylTitan high-throughput targeted methylation library construction process, together with the MGI DNBSEQ platform, can comprehensively facilitate DNA methylation detection, and accelerate scientific research and clinical diagnosis process of cancers.

Genetic Sequencer DNBSEQ-G400RS

References

1. Sun, J. *et al.*, Cross-platform comparisons for targeted bisulfite sequencing of MGISEQ-2000 and NovaSeq 6000. *CLIN EPIGENETICS* 15 130 (2023).
2. Stroun, M. *et al.*, The origin and mechanism of circulating DNA. *ANN NY ACAD SCI* 906 161 (2000).
3. Galardi, F. *et al.*, Cell-Free DNA-Methylation-Based Methods and Applications in Oncology. *BIOMOLECULES* 10 (2020).
4. Xu, R. H. *et al.*, Circulating tumour DNA methylation markers for diagnosis and prognosis of hepatocellular carcinoma. *NAT MATER* 16 1155 (2017).
5. Pishvaian, M. J. *et al.*, A pilot study evaluating concordance between blood-based and patient-matched tumor molecular testing within pancreatic cancer patients participating in the Know Your Tumor (KYT) initiative. *Oncotarget* 8 83446 (2017).
6. Tobias Fehlmann, S. R. C. G., cPAS-based sequencing on the BGISEQ-500 to explore small non-coding RNAs. *CLIN EPIGENETICS* (2016).

Recommended Ordering Information

Category	Product	Cat. NO.
Instruments	Genetic Sequencer DNBSEQ-G400RS	900-000170-00
Bioinformatics	MegaBOLT Bioinformatics analysis accelerator	900-000555-00
Library prep	MGIEasy Whole Genome Methylation Sequencing Library Prep Kit (16 RXN)	940-001530-00
	MGIEasy universal library conversion kit (App-A) (16 RXN)	1000004155
Sequencing reagents	DNBSEQ-G400RS High-throughput Sequencing Reagent Set (G400 FCL PE150)	940-001356-00

MGI Tech Co.,Ltd

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

Version: June 2024

+86-4000-688-114

en.mgi-tech.com

MGI-service@mgi-tech.com

Authors: Hanfei Zhang

Editor-in-Charge: Qiwei Wang

Reviewer: Yao Jiang