Performance and compatibility evaluation of Celemics' NGS panels with MGI DNBSEQ-G400 Platform

ABSTRACT

The development of Next Generation Sequencing (NGS) allows significant advancement to researchers and clinicians in oncology field for finding and screening tumor-specific biomarkers or other novel variants. This study aims to evaluate the performance of two Celemics NGS panels, WES and OncoRisk, and library preparation on MGI DNBSEQ-G400. Celemics whole exome sequencing (WES) panel is the most comprehensive whole exome panel available in the market with the target size of 37 Mb. As for OncoRisk, it is a germline specific panel designed to target 31 well-known oncogenes, which allows cost-effective sequencing compared to whole genome sequencing (WGS). To address the compatibility and performance of Celemics panels and library preparation kit on MGI platform, the equivalent sequencing of corresponding panels was also conducted on Illumina platform for detailed comparison. The results showed 100% sensitivity for detecting the known variants in reference standard sample, and the two platforms showed 100% match rate for detecting all relevant variants. This concludes that MGI platform can generate high-quality sequencing results comparable to Illumina platforms and that Celemics NGS panels and kits are fully compatible with not only the MGI's DNBSEQ-G400 but also other various sequencing instruments available in the market for robust sequencing results.

INTRODUCTION

The use of next-generation sequencing (NGS) technologies, combined with decreasing sequencing cost, has become a more cost-effective method for screening in clinical settings. With NGS, researchers and clinicians are able to make significant advances in oncology by performing whole-genome sequencing (WGS), whole-exome sequencing (WES), targeted gene profiling, targeted panel sequencing, and more. NGS also offers the sensitivity to detect rare somatic variants, tumor subclones, and circulating DNA fragments.

This application note evaluates the performance of two Celemics library preparation kits and target enrichment panels on MGI DNBSEQ-G400. (1) For whole exome sequencing, Celemics provides WES kit, which is the most comprehensive whole exome panel with a target size of 37 Mb that covers the regions of major WES panels in the market. (2) As for OncoRisk Panel, the panel was designed to target 31 well-known oncogenes, leading to a reduction of total sequencing cost compared to WGS or WES. Along with the evaluation of Celemics' kits and panels, this application note also aims to include sequencing instrument comparison data between MGI's DNBSEQ-G400 and Illumina NextSeq500 sequencing platform, to address detection sensitivity and compatibility of two of Celemics' kits in two different sequencing platforms.

MATERIALS AND METHODS

Sample Library Preparation and target enrichment

For the assessments of panels and different sequencing platforms for comparison, 200 ng of standard reference material samples, NA12878, NA12891, and NA12892 were used for each panel and corresponding libraries were prepared by Celemics Inc. (Korea). Celemics designed and produced capture probes for corresponding target regions of 96 Kb for OncoRisk, and 37 Mb for WES, respectively. Using the standard reference samples as noted above, the initial input amount for all samples were set to 200 ng for library preparation process. The process included following steps; fragmentation, end repair, dA-tailing, adapter ligation, and pre-PCR for indexed NGS library. Each sample library went under QC with TapeStation (Agilent Technologies, USA) D1000 and High Sensitivity D1000 screen tape to check for appropriate library construction. Each sample library was prepared in duplicates and distinguished as L01 and L02, for library 1 and library 2, respectively, to confirm reproducibility. NGS-prepared DNA library and capture probes were hybridized in buffer to capture target regions of interest through the use of a Celemics target enrichment kit. After the hybridization and washing process were finished, the captured libraries were amplified by the post-PCR process. The post-PCR products were then sequenced by MGI and Illumina platform.

Sequencing

The captured libraries were subsequently processed to make DNA nanoball (DNB) prior to be sequenced on the DNBSEQ-G400 and also processed accordingly for Illumina NextSeq500 sequencing platform as well. Samples and sequencing details are provided in **Table 1**.

The sequencing on DNBSEQ-G400 was conducted by MGI, while the sequencing on NextSeq500 platform was performed by Celemics.

	Sample and Sequencing Details
Sample type	NA12878 NA12891 NA12892
Sequencing panel	Whole Exome Sequencing Panel OncoRisk Panel
Sequencing kit	DNBSEQ-G400RS High-Throughput Sequencing Rapid Sequencing Kit NextSeq 500/550 High Output Kit v2.5
Sequencing device	MGI DNBSEQ-G400RS Illumina NextSeq 500
Sequencing strategy	PE150 for both platforms

Table 1. Sample, kits, and sequencing details used in this evaluation.

Bioinformatics

The data analyses were performed using MegaBOLT, a bioinformatics suite performing various analyses, such as alignment (invoking BWA), variant calling (invoking GATK), and quality control, for NGS sequencing data with high efficiency and accuracy. The variant validation analysis is conducted using RTG Tools and benchmark NA12878 SNP and INDEL sets. Human genome reference hg19 was used for the analysis.

RESULTS

Sequencing statistics

Sequencing metrics are compared over various categories of samples (**Table 2**; only NA12878 sample is shown out of three samples, NA12878, NA12891 and NA12892). The data volumes of the libraries vary among the samples. Q30 of raw and clean reads are comparable for any library of any sample (~95%), whereas clean rates for OncoRisk are relatively lower than those of WES.

	OncoRisk (L01)	OncoRisk (L02)	WES (L01)	WES (L02)
#Raw reads	16,112,328	17,951,376	67,818,040	73,274,128
Raw data (Gb)	2.22	2.48	9.65	10.42
Rate of clean reads	91.57%	91.68%	94.46%	94.42%
Q30 (raw)	95.76%	95.96%	94.68%	94.97%
Q30 (clean)	95.69%	95.90%	94.64%	94.93%

Table 2. Sequencing metrics of samples NA12878 on MGI's DNBSEQ-G400

Alignment statistics

Table 3A and **3B** present alignment metrics for both OncoRisk and WES panels using NA12878 samples. The comparison between MGI and Illumina sequencing platforms reveals the respective values for each category. Mapping rates, representing the percentage of mapped reads, exceeded 99% for all libraries in both panels. The mapping rates, coverage and on-target ratios for both OncoRisk and WES panels on the MGI platform were slightly higher compared to Illumina platform. However, overall metrics of both platforms were relatively similar, indicating the comparable sequencing performance for both instruments.

"Fold 80 base penalty" is the fold over-coverage needed to raise 80% of bases in covered targets to the mean coverage. It is a measure of sequencing uniformity. The lower the value, the better the uniformity. For OncoRisk panel, the Fold-80 base penalty was calculated to be 1.49 for MGI, which was lower than the value of 1.84 obtained from Illumina platform. Conversely, for WES panel, the Fold-80 base penalty on Illumina's platform was lower (2.11) compared to MGI's DNBSEQ (2.24). However, both platforms demonstrated relatively similar and small values, highlighting the outstanding sequencing competencies for both instruments. Coverage plots again demonstrate the uniformity of the sequencing (**Figure 1**).

	OncoRisk (MGI)	OncoRisk (Illumina)
Mapping rate	99.93%	99.79%
Duplication rate	4.46%	3.71%
Fold-80 base penalty	1.49	1.84
On-target ratio	72.59%	69.19%
Coverage at least 100X	99.50%	99.39%

Table 3A. Alignment metrics of NA12878 samples for OncoRisk

	WES (MGI)	WES (Illumina)
Mapping rate	99.88%	99.83%
Duplication rate	1.56%	1.68%
Fold-80 base penalty	2.24	2.11
On-target ratio	71.59%	68.65%
Coverage at least 20X	91.66%	90.50%

Table 3B. Alignment metrics of NA12878 samples for WES

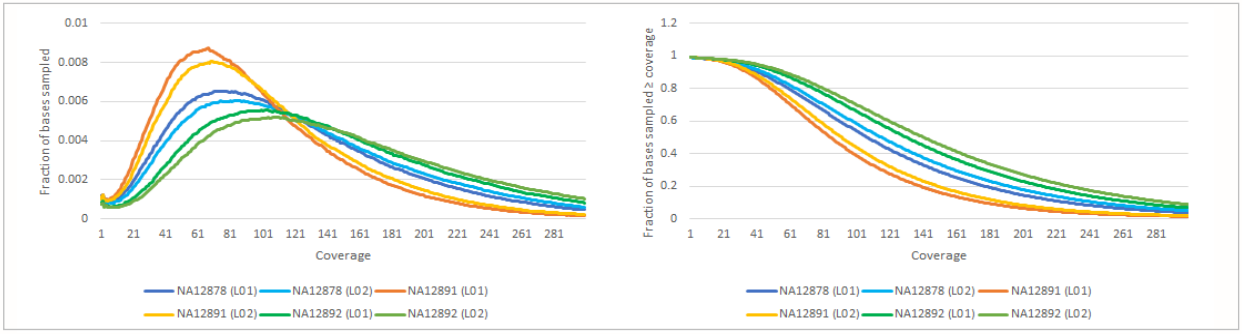

Figure 1 : Coverage plots for WES samples on MGI platform. Replicate libraries yielded similar coverage profiles, with a sharp peak indicating coverage uniformity. The mean coverage for each sample was ~100x with a high proportion of bases covered >30x.

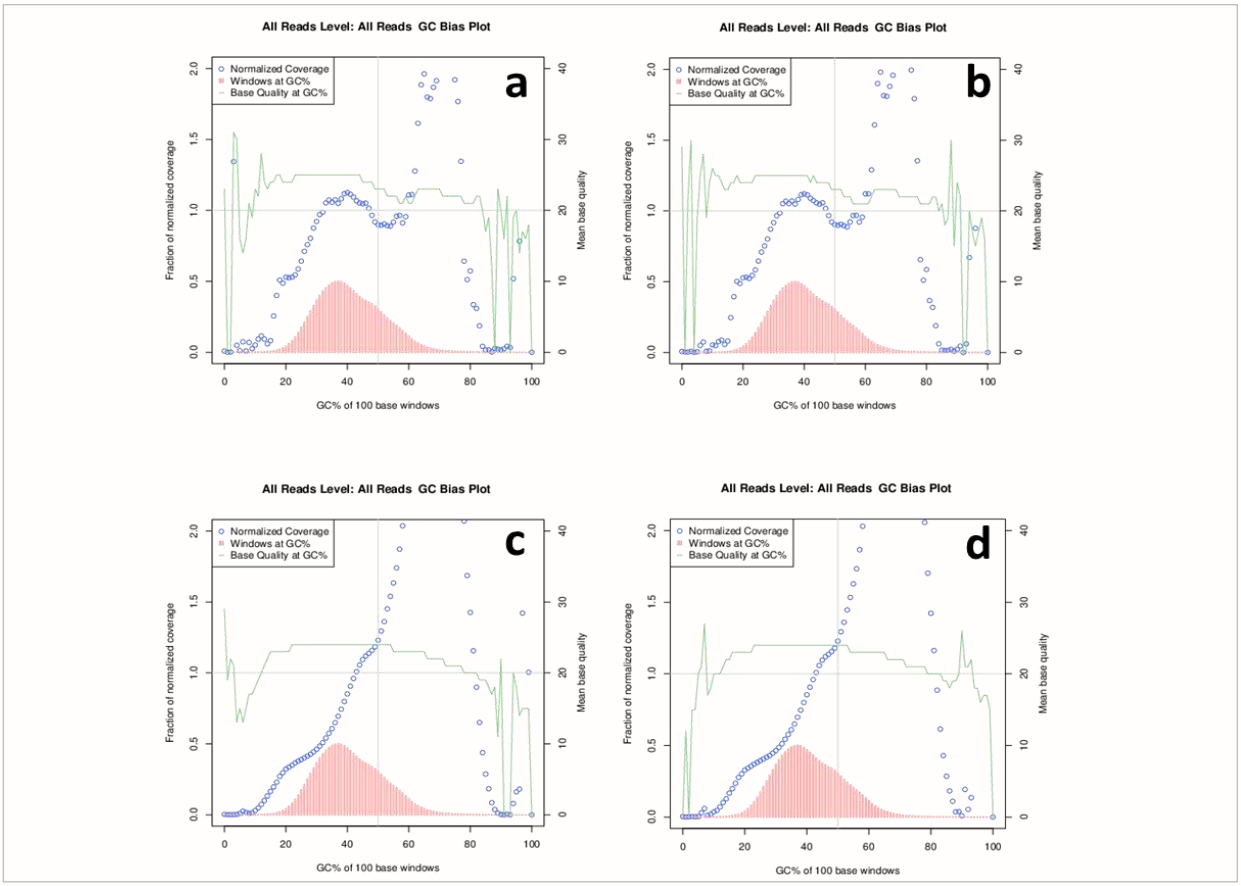

Figure 2. GC-bias plots for NA12878 samples: (a) OncoRisk L01 (b) OncoRisk L02 (c) WES L01 (d) WES L02 on MGI. Red histograms represent the distribution of genomic GC content, calculated for the reference sequence in 100 bp bins. GC-bias was assessed by plotting normalized coverage for each bin (blue circles). If all sample-to-data processes were completely unbiased, all bins would be equally represented, i.e., the plot for each workflow would be a horizontal distribution centered on a normalized coverage of 1.

GC bias plots show the performance of sequencing in different GC enrichment regions. As mentioned previously, duplicated libraries are labelled L01 and L02 to distinguish for each panel. The plots for all four NA12878 samples illustrate that balanced GC regions (30-50%) are well represented (~1), higher GC regions (60-80%) are over-represented (~2) and extreme GC regions are under-represented (~0) (**Figure 2)**. For NA12891 and NA12892 samples, the pattern observed in OncoRisk and WES libraries closely resemble those of NA12878 samples.

Panel Sensitivity Assessment

The sensitivity of variant calling were conducted by comparing detected variants with the list of known NA12878 standard variants for both platforms and analyzed using software "RTG Tools" and the corresponding data for OncoRisk panel is provided in **Table 4**. Of the generated data set, the germline variants were selectively identified as this particular panel was designed to investigate germline mutations.

OncoRisk	Sensitivity
OncoRisk (L01)	100%
OncoRisk (L02)	100%

Table 4. The sensitivity of variant calling of sample NA12878 for OncoRisk for both Illumina and MGI platforms

Also, the number of sensitivity values for OncoRisk panel variants generated with MGI's DNBSEQ were compared with those of the variants detected and verified on Illumina platform. The number of variants for NA12878 for OncoRisk panel detected in both MGI and Illumina sequencing platforms are listed in **Table 5**.
Of the germline variants successfully detected, we have observed the 100% concordance in result regardless of the sequencing platforms. Moreover, the duplicated libraries sequenced for both MGI and Illumina platforms also showed absolute concordance, yielding 100% match in both groups, leading to 100% sensitivity for OncoRisk illustrating good performance of the panel and the instruments.

		No. of Germline variants	Sensitivity
MGI	OncoRisk (L01)	52	100 %
	OncoRisk (L02)	52	100 %
Illumina	OncoRisk (L01)	52	100 %
	OncoRisk (L02)	52	100 %

Table 5. The numbers of germline variants and the sensitivity values for NA12878 from MGI and Illumina platforms for OncoRisk panel

Moreover, the total number of SNVs detected for sample library 1 and 2 for WES panel were 29,097 and 30,450, respectively. Based on the number of SNVs detected, the sensitivity values were indicated as 98.27% and 98.42% for WES L01 and L02, respectively (**Table 6**), demonstrating robust panel and sequencing performance.

	SNVs
	Sensitivity
WES (L01)	98.27%
WES (L02)	98.42%

Table 6. The sensitivity of variants called for NA12878 samples for WES panel on MGI

Reporting

MegaBOLT generates report for each sample analyzed. The report includes sequencing, alignment and variant calling metrics and graphs. Some part of an example report is demonstrated in **Figure 3**.

Figure 3. Part of an example of MGI's MegaBOLT report.

Conclusion

Celemics' OncoRisk and WES panels were sequenced on MGI's DNBSEQ and Illumina's NextSeq to evaluate panel performance and compatibility across different sequencing technologies. The constructed NGS data sets were analyzed using MGI's MegaBOLT pipeline and were also cross-referenced with Celemics' pipeline for confirmation.

Both MGI and Illumina instruments produced comparable sequencing data, yielding similar performance metrics for both OncoRisk and WES panels. These metrics demonstrated strong reproducibility in all duplicated libraries. Furthermore, the variants detected in both OncoRisk and WES panels across two platforms were found to be either identical or comparable, indicating a significant level of concordance.

The objective of this study was to explore the compatibility of Celemics' library preparation kit and NGS panels with MGI's DNBSEQ, while also comparing the sequencing efficiency and performance between MGI and Illumina sequencing platforms. The results and performance data lead to the conclusion that MGI's DNBSEQ platform can generate sequencing data of equivalent, if not superior, quality compared to Illumina's platform. Additionally, it highlights that Celemics' libraries and NGS panels exhibit high compatibility with various sequencing platforms, enabling the generation of robust sequencing results.

About

CELEMICS

Celemics has developed and manufactured over a thousand different panels to our customers, including hospitals, clinical labs, research institutes, and biopharma companies. The outstanding performance of the panel is enabled by Celemics' proprietary probe design technology and its RNA-based biotinylated probes, providing an even greater binding capacity than DNA-based probes. Celemics has been providing services to the top CROs and clinical labs in Korea and other countries since the company foundation, also collaborating with major hospitals in Korea including Asan Medical Center, Samsung Medical Center, and Seoul National University Hospital. Celemics is certified to ISO 13485, GMP, ISO 9001, and CE-IVD.

About

MGI

MGI Tech Co., Ltd. (referred to as MGI) is committed to building core tools and technology to lead life science through intelligent innovation. With a focus on R&D, production and sales of DNA sequencing instruments, reagents, and related products, MGI provides real-time, panoramic, and life course equipment and systems for precision medicine, precision agriculture, precision healthcare and other relevant industries. MGI is a leading producer of clinical high-throughput gene sequencers, and its multi-omics platforms include genetic sequencing, medical imaging, and laboratory automation.