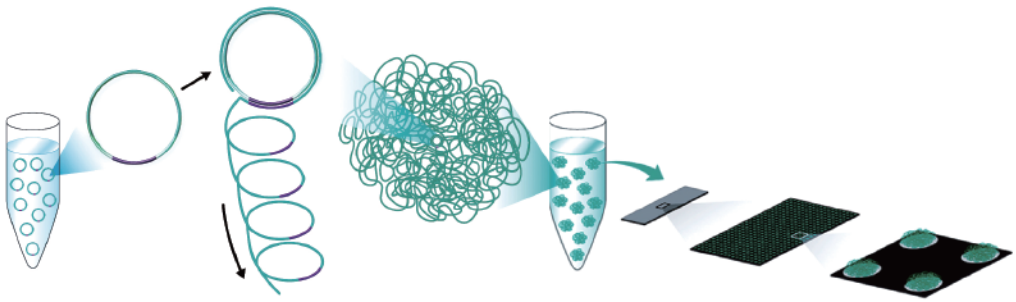
WE SEQUENCE, YOU DISCOVER

Towards the Perfect Genome

How DNBSEQ™ NGS Technology Addresses Sequencing Errors



We Sequence, You Discover

Towards the Perfect Genome

Innovation for a Cause

Next Generation Sequencing revolutionized genomics research by vastly improving the speed and cost of RNA and DNA sequencing. Commonly used NGS technology is limited in utility, especially in healthcare-related applications because of technology-related sequencing errors and cost factors.

The innovation of BGI's NGS technology[1] was motivated by the need for more accurate sequencing data. This was achieved by systematically eliminating common sequencing errors, while improving throughput and further reducing cost. The result is a new NGS technology - DNBSEQ™, which produces high quality genomic data at a lower cost. DNBSEQ™ technology enables us to approach the "Perfect Genome", at a cost that will allow for much broader application in research, drug development and personalized medicine. DNBSEQ™ sequencing technology has already been widely adopted by scientists around the world.

DNBSEQ™ NGS

More accurate, more affordable data to benefit research, drug development and the advance of Personalized Medicine.

Sources of Sequencing Errors

A major source of sequencing errors in common NGS technologies are the PCR amplification steps in the library preparation and in the sequencing process itself. Conventional NGS methods rely on cluster generation (bridge amplification) which leads to an exponential amplification of the single molecule to achieve a strong and robust sequencing signal, at the price of significantly amplified errors, poor coverage uniformity, and significant GC gaps.

In addition, signal-to-noise issues in the optical signal detection process cause misclassification of bases, which is a common error in conventional NGS platforms. Low signal-to-noise ratio will affect the base call accuracy, and sometimes brings in optical duplication. A widely used strategy to increase system throughput and reduce cost, is to pool multiple samples on a single flow cell and use barcoding techniques to assign reads to the original samples. In the last few years, it has been found that this strategy, when used with common NGS technology, introduces significant errors by mis-classification of reads[2].

So, strategies used to improve throughput and lower cost in conventional NGS methods have been offset with compromised data quality.

How DNBSEQ™ Technology Reduces Errors

DNBSEQ™ Sequencing technology systematically addresses the common causes of sequencing errors:

- **Circularized DNA Strands**

Genomic DNA or cDNA is fragmented either mechanically or enzymatically into DNA fragments between 200 bp and ~800 bp. The selected DNA fragments are blunted and 5′ phosphorylated, while the 3′ ends are A-tailed and ligated to the double-stranded adapters[3] with the T-overhang at the 3' ends. The adapter ligated library constructs are then denatured and the resulting single-stranded DNA with adapters at both ends is ligated with a splint oligo that is complimentary to adapter sequences, to form a specific intramolecular circle.

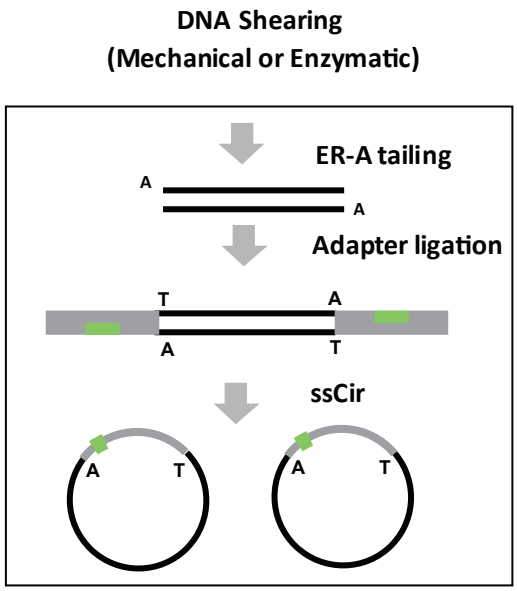
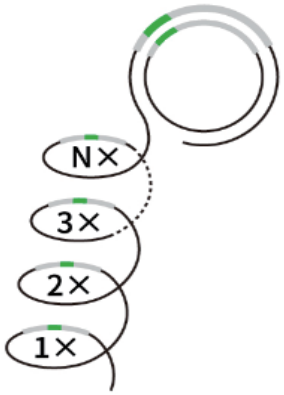


Figure 1. Overview of DNBSEQ™ library prep. Size normalization, end reparing, A-tailing and adapter ligation steps are illustrated before circularization of the constructed library for DNB generation.

• Rolling Circle Replication (RCR) and DNA Nanoball (DNB™) formation

The circularized DNA is converted into DNA nanoballs (DNBs) by a linear amplification process called Rolling Circle Replication (RCR), with specific adapters and a proprietary DNA polymerase. The number of replications is tightly controlled to 200 – 500 copies. The adaptor sequences and primers work to systematically fold the RCR product into nanoballs[4]. The DNBs are concatemers that are well-organized structures of around 200 nanometers in size.

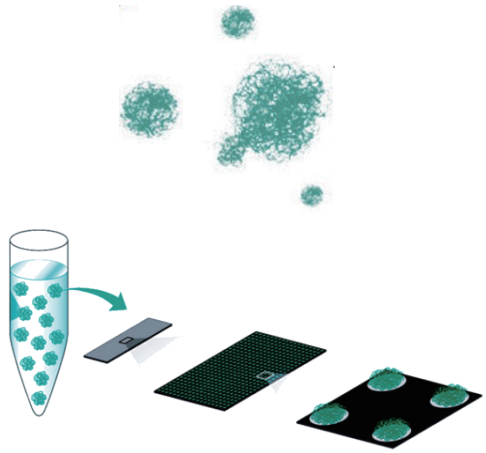


As RCR always copies the original DNA circle, there is no accumulation of clonal errors. Even if one of the three hundred original DNA copies has an error, it will not present an accumulated error because we measure the sum of all three hundred copies, resulting in almost zero clonal errors.

• DNA Nanoball (DNB™) loading and High-Density nano array formation

The adaptor sequences and primers work to collapse the RCR product into DNA Nanoballs[4].

The DNBs are then loaded onto a very tightly patterned nano array within a microfluidic flow cell with 200nm activated spots at 715nm pitch[5]. Only one DNB can occupy each spot because of their size and because once a DNB fills the spot, it presents a negative charge to repels other DNBs, allowing 95% occupancy rates for efficient use of flow cell.



The resulting high-density array has distinct spots of very concentrated DNA that consequently produce very bright fluorescence signal at well-defined spots and without crosstalk.

Since the DNB loading processing has a wide dynamic range, the DNB particles are loaded efficiently at low numbers as well as at high numbers without overloading. The three-dimensional structure of the DNB generates more intense signal which results in 7 times brighter spots on the DNB array compared to cluster arrays. The resulting higher signal to noise ratio enables easier and more accurate base calling. This ensures the accuracy of SNP and indel calls as well as negligible possibility of index hopping.

Furthermore, an extra bonus of the DNB array is that, unlike the cluster arrays, there is no need for a single molecule of library to find its complementary primer on the array surface, which is usually at the one nanometer precision and increases the cost and error rate. The DNB particles have the same size as the active spot on the array surface and the loading process is simple yet very efficient and error-proof, without the need for library concentration titration to avoid overloading as with other NGS platforms.

• Sequencing Process

The sequencing of DNBs is accomplished with combinatorial Probe Anchor Synthesis (cPAS) chemistry[4]. During each sequencing cycle, a fluorescent probe is incorporated into a DNA anchor on the DNB, followed by high-resolution digital image acquisition, digitization of the optical signals and finally conversion to the nucleotide sequence of the DNB.

Superior Data Clarity

Evaluation studies have demonstrated the superior performance of DNBSEQ™ True PCR-free sequencing technology in several areas.

• Significantly Reduced Duplication Rate

An immediate result of True PCR-free DNB sequencing is the significantly lower duplication rate, as a result the PCR-free sequencing process. Thus, nearly all the qualified sequencing data is usable, unlike the conventional sequencing data from which a good amount of duplicated reads will have to be filtered out or merged.

	DNBSEQ™ PCR-free	DNBSEQ™ PCR	NovaSeq PCR-free	NovaSeq PCR
Raw reads (M)	622	622	758	758
Raw bases (Gb)	93.4	93.4	114.1	114.5
Clean data rate (%)	99.85	99.87	98.86	98.95
GC content (%)	41.38	41.39	41.61	42.02
Mapping rate (%)	99.89	99.80	99.94	99.05
Unique mapping rate (%)	96.4	94.8	90.44	85.6
Duplicate rate (%)	0.66	2.2	6.37	11.47
Mismatch rate (%)	0.61	0.66	0.55	0.42
Average seq depth (X)	30.65	30.09	34.37	32.42
Coverage (%)	99.15	99.15	99.18	99.20
Coverage at least 20X (%)	93.48	93.36	97.65	94.60

Figure 2. DNBSEQ™ platforms demonstrate a much lower duplication rate. Human reference genome NA12878 was sequenced on two sequencing platforms (DNBSEQ™ and Illumina NovaSeq 6000) using two different library construction methods (PCR-free and PCR enabled) respectively, at the coverage of 30X. Four sets of sequencing data were analyzed and the overall performance shown above indicates that the DNBSEQ™ platform is superior to the NovaSeq platform in terms of clean data rate, unique mapping rate, mismatch rate, and especially the duplication rate (highlighted), even though the NovaSeq runs have deeper sequencing coverage at the beginning.

• Improved SNP and Indel Calls

For the reference sample NA12878, the concordance of SNP and InDel calling in the high confidence regions between the two sequencing platforms is high at 99.7% and 97.5%, respectively, confirming the high consistency of variant calls between DNBSEQ™ True PCR-free WGS and NovaSeq PCR-Free WGS.

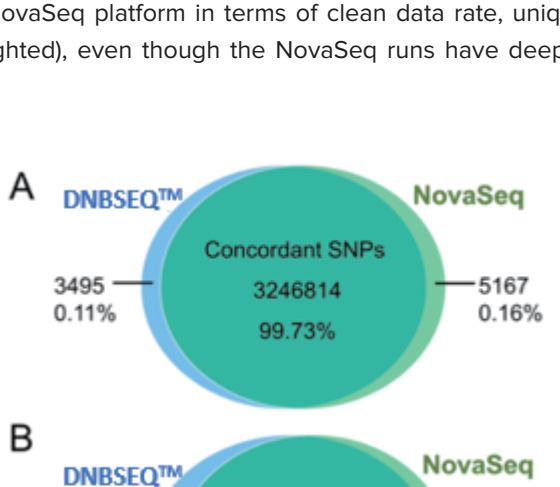


Figure 3. DNBSEQ™ provides concordance of SNP and InDel calling in the high confidence regions compared with the Illumina NovaSeq 6000. A high concordance of 99.73% was observed in SNP detection (A) and 97.50% in InDel detection (B) between DNBSEQ™ platform and the NovaSeq, respectively, using human reference genome NA12878.

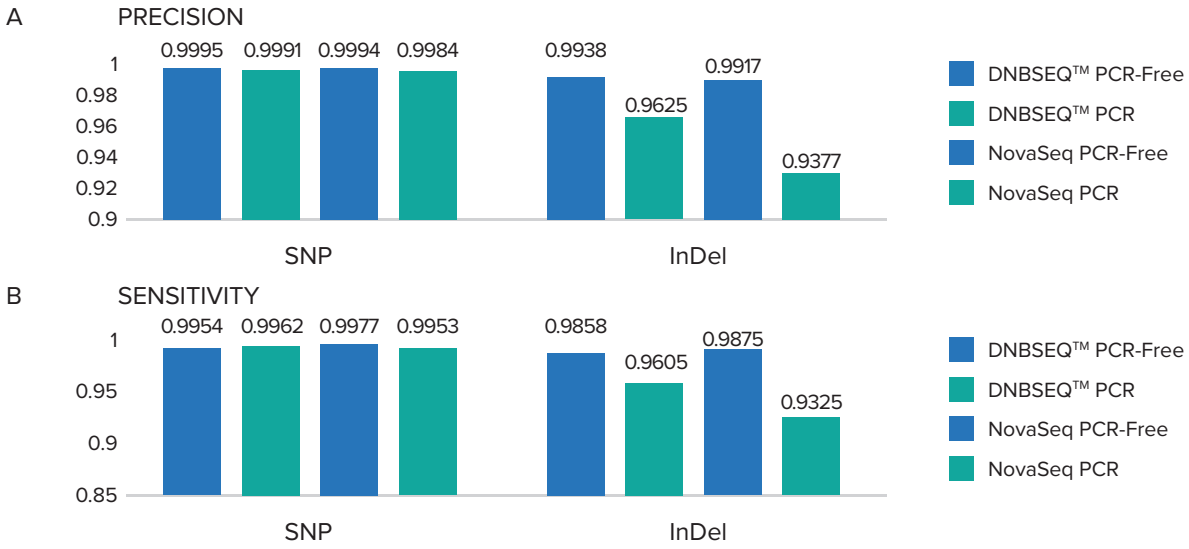


Figure 4. DNBSEQ™ performs at par with the Illumina NovaSeq 6000 in SNP detection but better in InDel detection. Precision (A) and sensitivity (B) were compared between the two different sequencing platforms and two library prep workflows, respectively, when analyzing variant calls based on the human reference genome NA12878.

• Higher Coverage in GC-rich Regions

The complete lack of PCR amplification in the DNBSEQ™ PCR-free workflow reduces library bias and improves coverage for traditionally challenging genomic content including GC-rich regions. Increased coverage of DNBSEQ™ PCR-free libraries resulted in fewer coverage gaps in the GC-rich regions, eventually leading to more covered genomic areas and much less "unknown' regions yet to be discovered by alternative genomic tools.

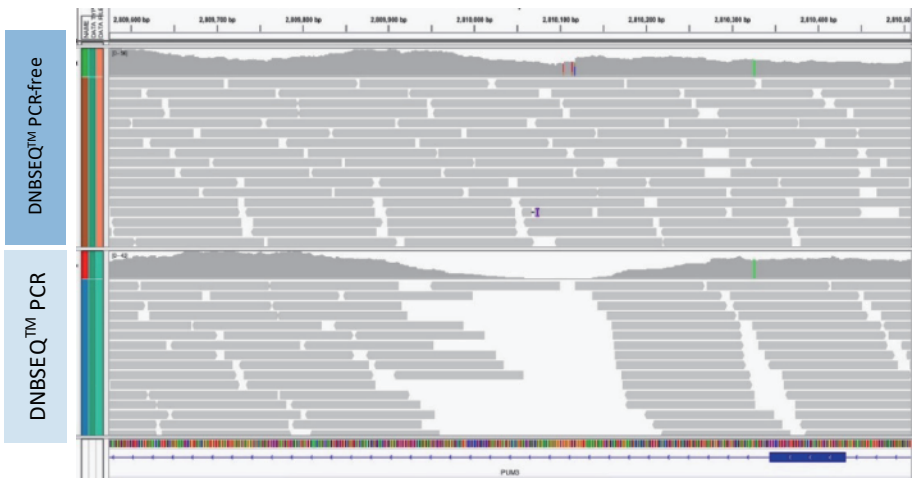


Figure 5. DNBSEQ™ reduces coverage gaps in the GC-rich regions. WGS data analysis results demonstrated that the GC-rich gaps located in the promoter region of the PUM3 gene can be sequenced by the DNBSEQ™ PCR-free workflow (upper panel) which can be difficult for regular WGS (data not shown) and the PCR-enabled DNBSEQ™ workflow (lower panel).

• Virtually No Index Mis-assignment

Due to the complete elimination of PCR throughout the entire DNBSEQ™ sequencing process, there is virtually no index mis-assignment when samples are pooled into one sequencing run, preserving data integrity while optimizing throughput and cost [4].

Conclusion: True benefits from DNBSEQ™ NGS Technology

The unique DNBSEQ™ True PCR-free Next Generation Sequencing technology, developed by BGI's Complete Genomics subsidiary in Silicon Valley, offers significant benefits in data quality and cost, as a result of specific innovations in technology and chemistry.

These benefits offer more accurate genomic data for research and further democratization of sequencing technology. Furthermore, it provides easier access for scientists involved in basic and translational research, for developers of pharmaceutical drugs or personalized medicine, and for investigators in environmental research and agriculture applications.

DNBSEQ™ Technology Benefits	
True PCR-free sequencing	• Reduced coverage bias • Better coverage of GC regions • Virtually no index miss-assignment with pooled samples, for maximum throughput and lower cost without loss of data integrity
DNA Nanoballs on high-density nano arrays produce a bright, concentrated signal	• Benefit deep sequencing applications like Transcriptome Sequencing • Improved signal to noise for accurate base calling

True PCR-free Sequencing Technology	Frequency of Mis-assignment
ExAmp sequencing with non-combinatorial dual indexes	0.08%
DNB sequencing with single index	0.0001%
DNB sequencing with empty controls	0.0000028%

References

1. Reliable Multiplex Sequencing with Rare Index Mis-Assignment on DNB-Based NGS Platform. BMC Genomics volume 20, Article number: 215 (2019)
2. Effects of Index Misassignment on Multiplexing and Downstream Analysis (Illumina white paper, 2017)
3. Vesicular Linker and Uses Thereof in Nucleic Acid Library Construction and Sequencing. US Patent No. US20170356039A1
4. Comparative Analysis of Novel MGISEQ-2000 Sequencing Platform vs Illumina HiSeq 2500 for Whole-Genome Sequencing. BioRxiv 552588 (2019)
5. Method and System for Accurate Alignment and Registration of Array for DNA sequencing. US Patent No. US9359641B2

Experience DNBSEQ™ Data For Yourself

DNBSEQ™ technology is available in the US through the sequencing services of BGI Americas.

Find out why over 75% of our customers have already benefitted from quality data at a lower cost from DNBSEQ™ NGS technology, by visiting www.bgi.com or by contacting your local BGI Sequencing expert.

BGI Americas
One Broadway, 14th Floor
Cambridge, MA 02142,
USA
Tel:+1 617 500-2741

BGI Europe
Ole Maaløes Vej 3,
DK-2200 Copenhagen N,
Denmark
Tel:+45 7026 0806

BGI Asia-Pacific
16 Dai Fu Street,
Tai Po Industrial Estate,
New Territories, Hong Kong
Tel:+852 36103510

Published January 2020.

BGI
We Sequence, You Discover