MGI High Throughput Sequencing Platform

Innovate Life Technology
MGI Tech Co., Ltd

MGI 华大智造

*DNBSEQ-G400

About MGI

MGI Tech Co., Ltd. (MGI) a subsidiary of BGI Group, is committed to enabling effective and affordable healthcare solutions for all. Based on its proprietary technology, MGI produces sequencing devices, equipment, consumables and reagents to support life science research, medicine and healthcare. MGI's multi-omics platforms include genetic sequencing, mass spectrometry and medical imaging. Providing real-time, comprehensive, life-long solutions, its mission is to develop and promote advanced life science tools for future healthcare.

MGI 华大智造

Outline

MGI Sequencing platform

MGI's DNA sequencing instruments utilize the state of art core technology called DNBSEQTM. DNBSEQTM includes all technology related to DNA nanoballs (DNB), such as DNA single strand circularization and DNB preparation technology, Patterned Arrays, DNB loading, cPAS (combinatorial Probe Anchor Synthesis), Pair-End Sequencing technology on DNB's, fluidics and detection systems, base calling algorithms, etc. CoolMPS is an advanced technology developed from cPAS. cPAS technology has been widely used on various sequencing platforms including DNBSEQ-G50, DNBSEQ-G400, DNBSEQ-T7, etc. In addition, MGI has developed a series of automated sample preparation systems and libraries preparation kits for total solutions of various applications.

*DNBSEQ-G50

MGI sequencing technology

DNB

DNB is the unique technology that allows DNA linear amplification in a single-tube solution. The workflow includes DNA fragmentation, adapter ligation and single-stranded circulation to produce DNB.

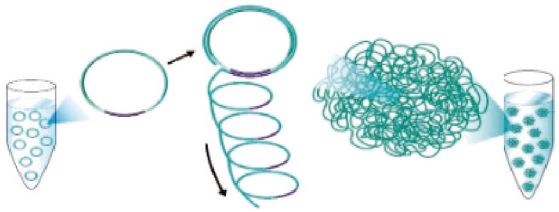
▶ High intensity of fluorescent signals and high sequencing accuracy

RCR increases the quantity of "DNA fragments (to be processed)" and thus enhance the intensity of fluorescent signals and accuracy.

▶ Zero-error accumulation using RCR

RCR enables each copy to be amplified from the original template and generate no accumulated replication errors, delivering high sequencing accuracy.

▶ Maximized efficiency of DNA loading on patterned array



Patterned array

Patterned array is the key in DNB loading technology. The coating surface of the semiconductor chip makes each loading spot positively charged. Therefore, negatively charged DNB can easily attach to the positively charged chip surface through electrostatic adhesion. Patterned array ensures each spot is attached to a single DNB and thus prevents the signal interference. Overall, patterned array enables high sequencing accuracy and high chip utilization.

▶ Increasing sequencing signals using nanoscale patterned chip

The Patterned Array of spots on Flow Cells prevents cross interference between fluorescent signals and enable high intensity of signals.

▶ Precision machined semiconductor chip enhances the attachment of DNB to chip

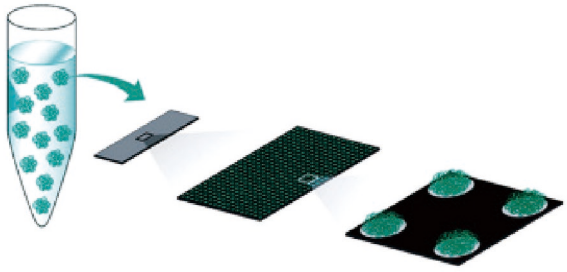
The size of DNB matches the area of effective spot on the chip.
Thus, each spot is only attached to a single DNB, avoiding the signal interference.

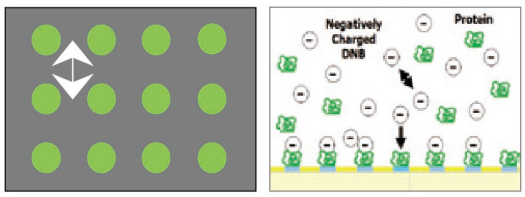
▶ Precise pattern enables efficient imaging system and DNB loading

Patterned array and DNB technology maximize the utilization of imaging system and chip surface area, and sequencing accuracy.

▶ Low Duplication , No Index Hopping

DNB is amplified in solution and loaded without PCR reagents. Hence the duplicate rate is extremely low using DNB technology and patterned array.





cPAS

cPAS is the innovative technology developed by BGI. By improving combinatorial Probe-Anchor Ligation(cPAL), cPAS increases the sequencing accuracy with higher speed and read length of sequencing.

▶ MGI proprietary enzymes react completely with 60 seconds

Our biochemistry team has studied a large quantity of reaction conditions and screened tens of thousands of sequencing enzymes to successfully complete biochemical reactions within just 60 seconds.

▶ Real-time sub-pixel registration, image processing and base calling

Advanced Real-time Image Processing software, Sub-pixel Registration and Multi-thread Parallel Compression Algorithms result in accurate real-time imaging and base calling and high industry-leading data-processing speed. Furthermore, GPU empowers DNBSEQ-G400 with high speed to analyze high empowers DNBSEQ-G400 with high speed to analyze high

▶ CoolMPS

CoolMPSFluorescent labeling of dNTPs and capture of nucleotide signal are key steps in high-throughput sequencing. CoolMPS sequencing chemistry is a novel antibody-based sequencing product. The dNTPs of CoolMPS are without fluorescent labeled (called “cold dNTPs”) which incorporated into the sequencing strand by DNA polymerase, and base calling is achieved by specific binding of fluorescently labeled antibodies. During this process, the incorporated bases are unmodified, ultimately resulting in clearer base calling.

MGI Genetic Sequencer

DNBSEQ-G400

Versatile and flexible large-scale sequencer

The DNBSEQ-G400 provides cost-effective solutions for exome, transcriptome and whole-genome sequencing. Its dual flowcell (FC) system supports two FC and several different read length to meet a wide range of sequencing needs.

Features

- Competitive pricing and low running cost
- Various read lengths, short turnaround time, a wide range of applications
- Multiple throughput modes support different sequencing project needs

Performance parameters

FCS chip					
Lane/chip	2 lanes	Read length	PE150, PE100, SE100*	Run time	≤26 hours
Number of reads	Single chip : ~550 M Dual chip : ~1100 M	Output per run	Single chip: ~165 Gb Dual chip: ~330 Gb		
FCL chip					
Lane/chip	4 lanes	Read length	SE50, SE100, SE400, PE100 PE150, PE200	Run time**	≤48 hours
Number of reads	Single chip:~1800 M Dual chip:~3600 M	Output per run	Single chip: 75 ~ 720 Gb Dual chip: 150 ~ 1440 Gb		
Data quality	Q30≥85% ***				

* SE100 is in development.
** Run time for PE100 with dual chip at full capacity.
*** The percentage of bases > Q30 is averaged over the entire run. This result is performed under BGI-E8 quality control library.

Application

Research

WGS, WES, targeted sequencing, whole transcriptome sequencing, RNA-Seq, non-coding RNA sequencing, ChIP-Seq, WGBS, ddRAD sequencing, single-cell sequencing (DNA, RNA)

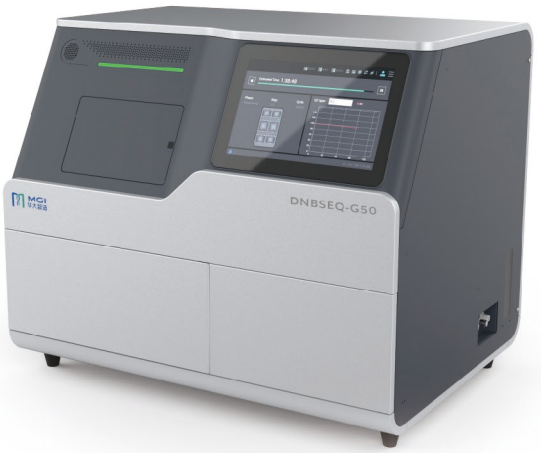
Clinical Application

NIPT, PGS/PGD, Chromosome CNV detection, complex disease detection, monogenic disorder detection, tumor gene mutation detection and pathogenic microorganism detection

DNBSEQ-G50

Dedicated high-efficiency desktop sequencer

DNBSEQ-G50 is a dedicated benchtop sequencer which meets different requirements of read length, time and output. Its key applications are targeted sequencing, small WGS and pathogen screening.



Features

- Various read length options and short turnaround time
- Build-in automated system saves the costs of reagents
- Integrated LIMS (Laboratory Information Management System) enables the fully-automated sample tracking

Performance parameters

FCS					
Lane/flow cell	1 lanes	Read length	SE50, SE100,PE50, PE100, PE150*	Run time**	≤48 hours
Number of reads	~300 M	Output	15 ~ 60 Gb		
Data quality	Q30≥75%-80% ***				

* Number of reads will be upgrade to 100-500M togeter with PE150 in 2020 Q2.
** Run time for PE100 at full capacity
*** The percentage of bases > Q30 is averaged over the entire run. This result is performed under BGI-E8 quality control library

Application

Clinical Application

NIPT, PGS/PGD, Chromosome CNV detection, pathogenic microorganism detection, tumor gene mutation detection, complex disease detection and monogenic disorder detection.

Research

small WGS, targeted sequencing, WES, transcriptome sequencing, RNA-seq, small RNA sequencing, ChIP-Seq and single-cell sequencing (DNA/RNA)

Automated sample preparation system

DNBSEQ-T7

Turbocharge your sequencing

DNBSEQ-T7 can generate 1-6Tb of high quality data per day, for a wide range of applications.
Powered by DNBSEQTM Technology, DNBSEQ-T7 makes sequencing more efficient and productive with advances in biochemical, fluidics, and optical systems.

Features

- High-speed: 24 HOURS for PE150 sequencing
- High-flexibility: 4 FLOWCELLS, PE150 and PE100 at the same time
- Ultra-high Throughput: up to 6 Tb/DAY, High quality data 24/7

Performance parameters

Reads Lengths	Max. read number/chip	Data Output	Data QualityQ30*	Run time**
PE100	5000M	1~4Tb	>85%	20h
PE150	5000M	1.5~6Tb	>80%	~24h

* The percentage of base above Q30 is the average of an internal standard library over the entire run.
The actual performance is affected by factors such as sample type, library quality, and insert fragment length.
** Run time was calculated based on Dual-Flow Cell mode, and includes sample loading, sequencing, base calling and data processing.

Application

Research

Whole Genome Sequencing, Deep Exome Sequencing, Epigenome Sequencing, Transcriptome Sequencing, and targeted panel projects.

Solutions for a wide range of applications

Sequencer	DNBSEQ-G50	DNBSEQ-G400		DNBSEQ-T7
Highlights	Rapid and efficient solution for targeted sequencing and small genome sequencing	Cost-effective solution for genomic research		high-speed, high flexibility and ultra-high throughput
Key application	Targeted sequencing, microbial genome sequencing	Exome, transcriptome, genome sequencing etc.		Whole genome sequencing, Exome sequencing
Flow cell type	FCS	FCS	FCL	FCL
Lane/chip	1 lane	2 lane	4 lane	1
Throughput	Mid throughput	Mid throughput	High throughput	Ultra high throughput
Chip/run	1	2	2	4
Data output	~ 60 Gb	~ 330 Gb	~ 1440 Gb	~6Tb
Run time*	≤48 hours	≤26 hours	≤48 hours	≤24 hours
Number of reads/ chip	~300 M	~550 M	1500M~1800 M	~5000M
Maximum read length	PE100	PE150	PE200	PE150

MGISP-100

Automated sample preparation system MGISP-100 is a dedicated library preparation platform for NGS. Its automated workflow enables batch operation and eliminates complicated manual operation, delivering affordable reproducible results on NGS library preparation.

Function

- Automated nucleic acid extraction, PCR, library preparation etc.
- Supports a wide range of NGS applications, such as NIPT, PGS, human whole genome sequencing, pathogen identification *, single bacteria identification*

Features

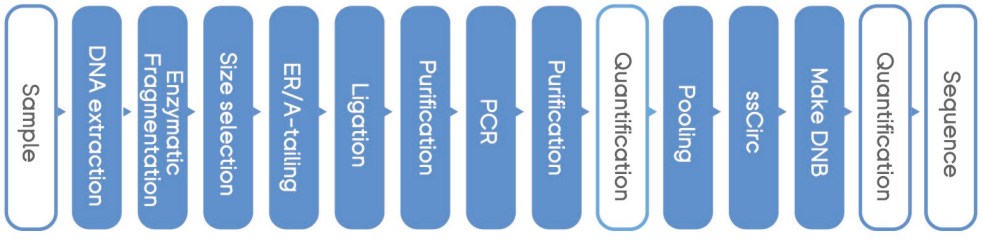
- One stop solution integrating PCR, temperature control and many functional modules
- Simplified library preparation on a single instrument. Minimal labor requirement and human errors
- Cost-effective, short turnaround time. Batch operation is ideal for small and medium laboratory
- Simple-to-follow interface and instruction
- Real-time sample tracking
- LIMS (Laboratory Information Management System) supports automated sample tracking throughout the sequencing workflow
- Full contamination control. Build-in ISO5 standard cleanroom fan filter unit
- Open platform and customizable system setting

Performance parameters

Sample/run	16 samples		
Turnaround time*	Manual operation 30min + run time 6hours		
Sample type	plasma, saliva, FFPE, gDNA, WGA product etc.		
Volume of liquid handler	2 - 200 μl		
Accuracy of liquid handler accuracy	liquid	2 μl	200 μl
	CV	< 5%	< 1%
	accuracy	< ±10%	< ±1%

* Turnaround time of library preparation for human whole genome sequencing. Turnaround time might be different between library preparation for different applications.

The automated library preparation workflow

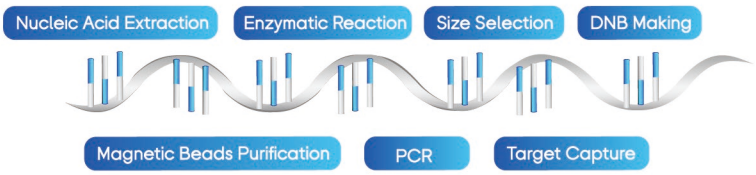


Note: white shows manual operation and blue shows automated steps.

MGISP-960

MGISP-960 High-throughput Automated Sample Preparation System is a flexible, fully automated workstation with a built-in NGS sample-prep process and more tailored workflows to satisfy each lab's needs and budget. Handling 8 to 96 sample during a run, it has a 24-position board supporting multiple combinations of functional modules.

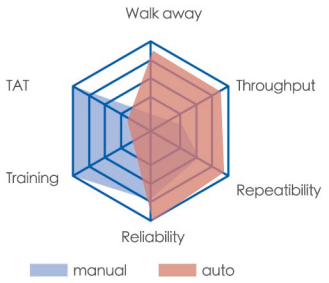
Function



- Applications are available for genomics, cell biology, drug discovery, protein science, analytical chemistry, etc.

Features

- One stop solution integrating PCR, temperature control, shaker, magnetic rack and many functional modules
- Simplified library preparation on a single instrument. Minimal labor requirement and human errors
- Cost-effective, short turnaround time
- Simple-to-follow interface and instruction
- LIMS (Laboratory Information Management System) supports automated sample tracking throughout the sequencing workflow
- Full contamination control. Build-in ISO5 standard cleanroom fan filter unit
- Open platform and customizable system setting

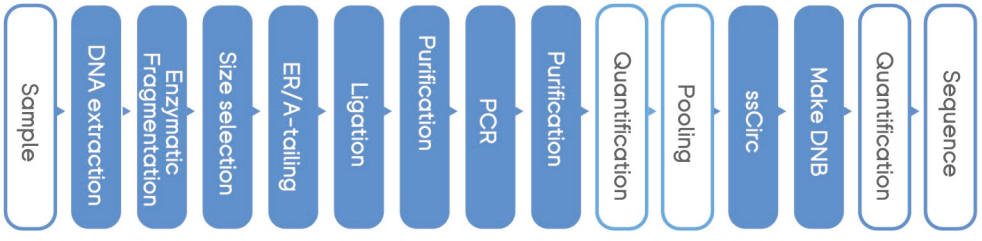


Performance parameters

Sample/run	8-96 samples		
Turnaround time*	Manual operation 2hours + run time 6hours		
Sample type	plasma, saliva, FFPE, gDNA, WGA product etc.		
Volume of liquid handler	2 - 200 μl		
Accuracy of liquid handler accuracy	liquid	2 μl	200 μl
	CV	< 5%	< 1%
	accuracy	< ±10%	< ±1%

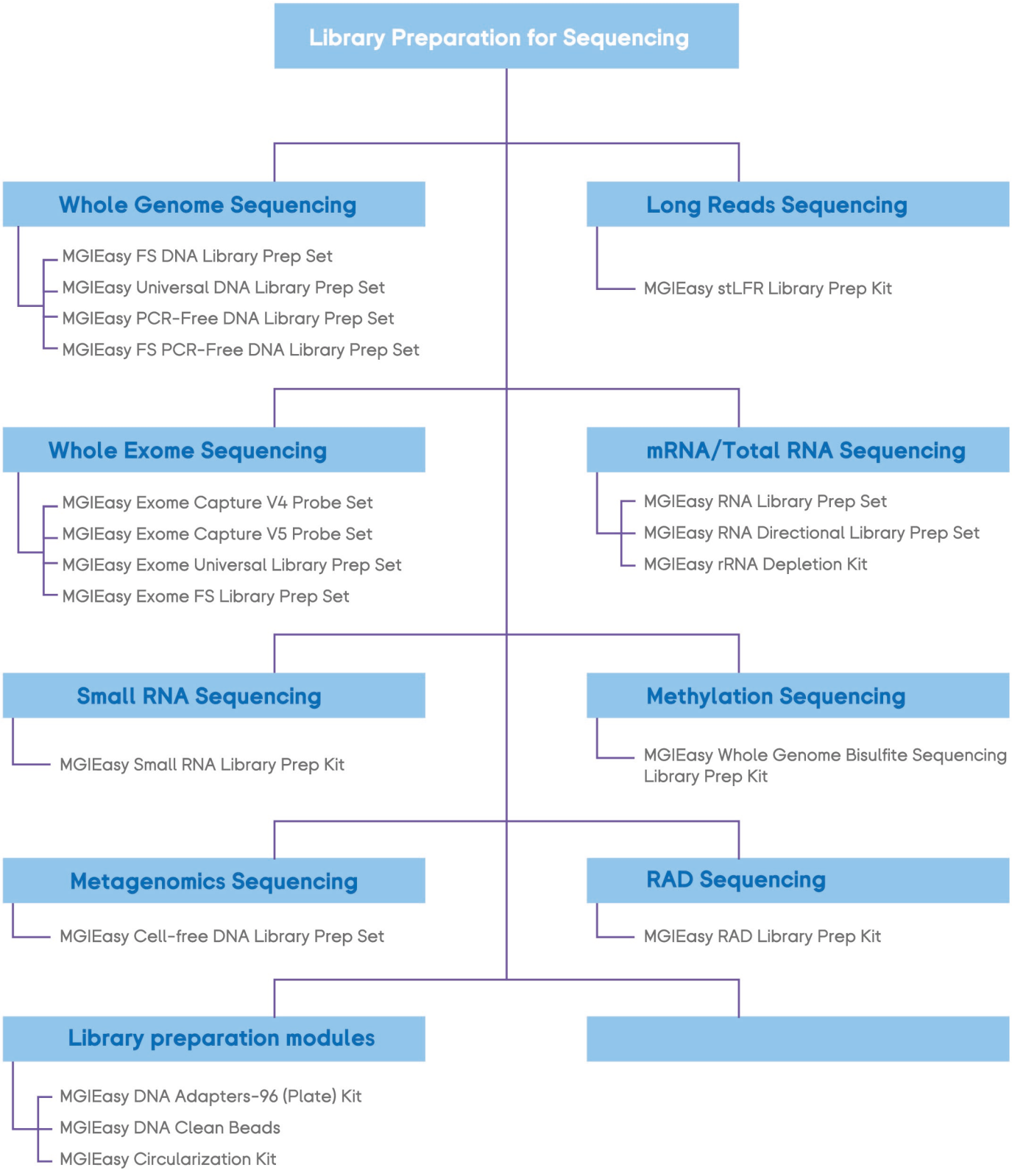
* Turnaround time of library preparation for human whole genome sequencing. Turnaround time might be different between library preparation for different applications.

The automated library preparation workflow



Note: white shows manual operation and blue shows automated steps.

Library Preparation for Sequencing



Whole Genome Sequencing

Whole-Genome Sequencing (WGS) is the process by which an organism's entire genome, both nuclear and cytoplasmic, is unraveled and read. It is useful in elucidating the underlying genomic details in studies at either an individual or population scale. WGS has applications in selective breeding, population evolution, disease research, clinical diagnosis and drug developments.
The DNBSEQ-based WGS solutions can accurately and rapidly identify the distribution of SNPs across the genome and annotate the variants based on the genome database.

Recommended MGI Process

Library preparation	Sequencing	Bioinformatics Analysis
MGIEasy FS DNA Library Prep Set MGIEasy Universal DNA Library Prep Set MGIEasy FS PCR-Free DNA Library Prep Set MGIEasy PCR-Free DNA Library Prep Set	DNBSEQ-G400 DNBSEQ-G50 DNBSEQ-T7	MegaBOLT Bioformics analysis accelerator (WGS Analysis System)
Automated sample preparation system MGISP-100 and MGISP-960		

WGS library preparation

MGI has developed a series of DNA library preparation kits for WGS to meet the needs of various sample types and applications.

▶▶ MGIEasy FS DNA Library Prep Set

The MGIEasy FS DNA Library Prep Set can quickly prepare 5-400 ng genomic DNA into a library using enzyme fragmentation method.

▶▶ MGIEasy Universal DNA Library Prep Set

The MGIEasy Universal DNA Library Prep Set provides an efficient and universal library construction solution for various DNA sample types such as fragmented gDNA, FFPE DNA, Meta DNA , cfDNA, ChIP DNA and capture DNA.

▶▶ MGIEasy PCR-Free DNA Library Prep Set

The MGIEasy PCR-Free DNA Library Prep Set is used for broad application of WGS library without PCR amplification. For removing PCR amplification during library preparation, there is no PCR error or PCR bias. Combined with DNBSEQ™ technology, is the true PCR-free sequencing workflow without PCR amplification. It has the advantages of no error accumulation, high coverage uniformity and excellent performance of variation detection.

▶▶ MGIEasy FS PCR-Free DNA Library Prep Set

The MGIEasy FS PCR-Free DNA Library Prep Set is used for broad application of WGS library without PCR amplification. The kit include high-quality and low-bias fragmentation enzyme, and could fragment 50-1000 ng gDNA.

Product specification

Products	MGIEasy FS DNA Library Prep Set	MGIEasy Universal DNA Library Prep Set	MGIEasy PCR-Free DNA Library Prep Set	MGIEasy FS PCR-Free DNA Library Prep Set
Assay Time	<5.5 h	<4.5 h	~3.5 h	~3.5 h
Input Quantity	5- 400 ng gDNA	0.5-50 ng fragmented DNA (200-500 bp)	80-200 ng fragmented DNA	50-1000 ng gDNA
Sample types	gDNA, FFPE DNA, Meta DNA, cfDNA	gDNA, FFPE DNA, Meta DNA, Amplicons, cfDNA, ChIP DNA, capture DNA, etc.	gDNA, meta DNA	gDNA, meta DNA
Species Compatibility	Human, animals, plants, fungi and bacteria, such as mouse, rice, Arabidopsis, yeast and E.coli.		Human, animals, plants, fungi, bacteria, metagenomics	
Fragmentation method	Enzyme	Physical	Physical	Enzyme
Automation option	Automation-friendly: MGISP-100,MGISP-960			
Applications	Human genome sequencing, Plant and Animal genome sequencing, Microbial genome sequencing, Metagenomics sequencing			
Platform Compatibility	DNBSEQ-G400, DNBSEQ-G50,DNBSEQ-T7			
Recommended Read Length	SE50, PE100, PE150 and etc.			

MGIEasy FS DNA library Prep Set

Features

- Low DNA input
- Compatible with intact and degraded specimens, including FFPE, plasma, and prokaryotic samples with different GC content from human, animal, plant and fungi samples
- Easy to use and suitable for automatic library preparation instrument
- Outstanding quality sequencing data with excellent coverage of genome and variation detection capacity enabling excellent usability of data

Performance

▶ Meeting the needs of different inserts

Using NA12878 standard as a template, by controlling the enzyme digestion time of the fragmentation step and the magnetic bead selecting step, PCR products with different sizes can be consistently obtained, satisfying the needs of various inserts and sequencing read lengths.

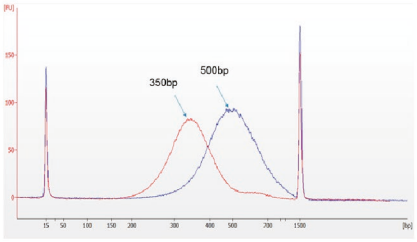


Fig.1 Library quality control graph of different inserts.

▶ Compatible with microbial library preparation of different GC content

Microbial samples with different GC content were used as templates to achieve a stable range of 600 ng within the recommended amount of gDNA. The yield of the PCR library above meets the requirements for subsequent circularization and sequencing. The coverage plots of high GC bacteria and low GC bacteria are similar to middle GC bacteria, and close to the expected normalized coverage of 1.0. This indicates the FS DNA Library Prep Set has uniform GC coverage over a broad range of GC content.

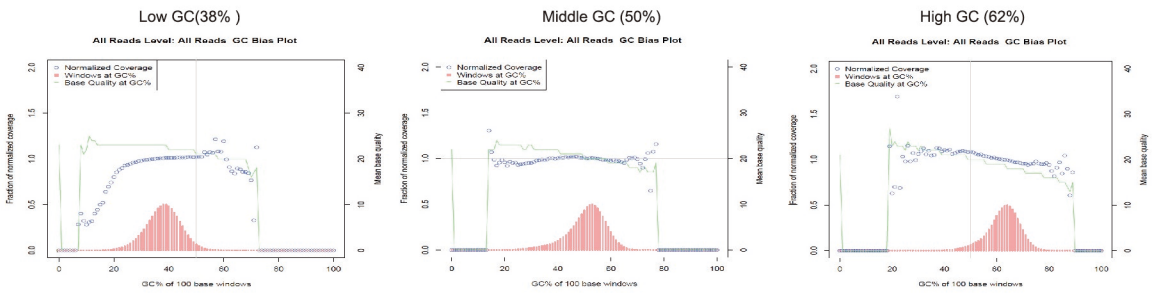


Fig.2b GC bias plot of bacteria with different genome GC content (Bacillus megaterium, 38% GC; E.Coli. , 50% GC and Olsenella, 62% GC).

MGIEasy PCR-Free DNA library Prep Set

▶ Feature

- By eliminating PCR amplification steps, WGS PCR-free prepared and sequenced by MGI DNBSEQ™ platform has no amplification error accumulation, resulting in better genome fidelity.
- Compatible with human, animals, plants, bacteria, fungi etc., e.g. human (blood, saliva, fresh tissue), mice, rice, E.coli and metagenomics
- Compared with traditional WGS (PCR amplification), WGS PCR-free reduces GC bias and improves coverage uniformity across the genome, such as GC-rich region, promoter and repetitive region.
- Compared with traditional WGS (PCR amplification), WGS PCR-free shows higher sensitivity and accuracy of variant detection, especially indels.

▶ Performance

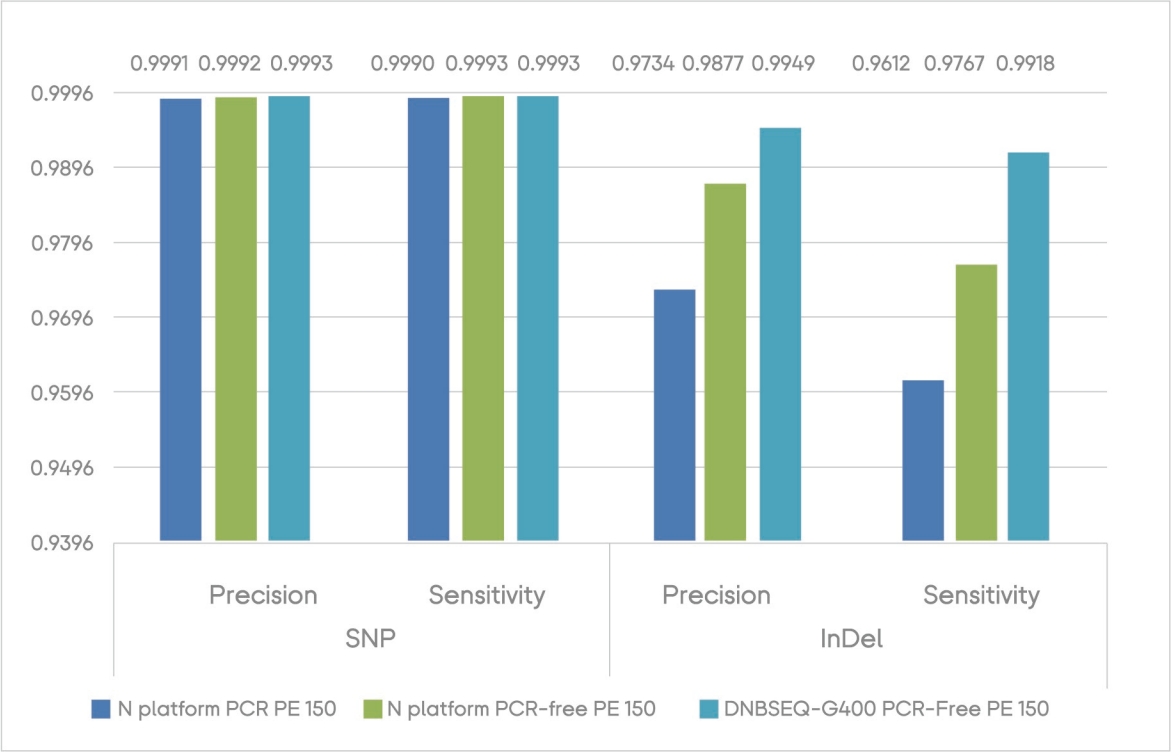


Figure 1 Comparison of variant detection performance of PCR-free WGS and PCR WGS on two sequencing platform. Libraries were prepared using NA12878, prepared with MGIEasy PCR-Free DNA Library Prep Set, sequenced on DNBSEQ-G400 (PE150), and analyzed for variant detection performance. The results were compared with competing products (T PCR-Free kit and T PCR kit) on N sequencing platform, using data downloaded from their official website.

MGIEasy FS PCR-Free DNA library Prep Set

▶ Feature

- Use a high-quality，low bias fragmentase for different species and input amounts shearing with the same incubation condition can get the consistent and concentrated fragment range.
- Library construction can be completed within 3.5 hours.
- Compatible with automatic sample preparation systems to provide an automated solution.
- The variant detection performance of FS PCR-free is similar to Covaris PCR-free, and better than traditional PCR, especially InDels.

▶ Performance

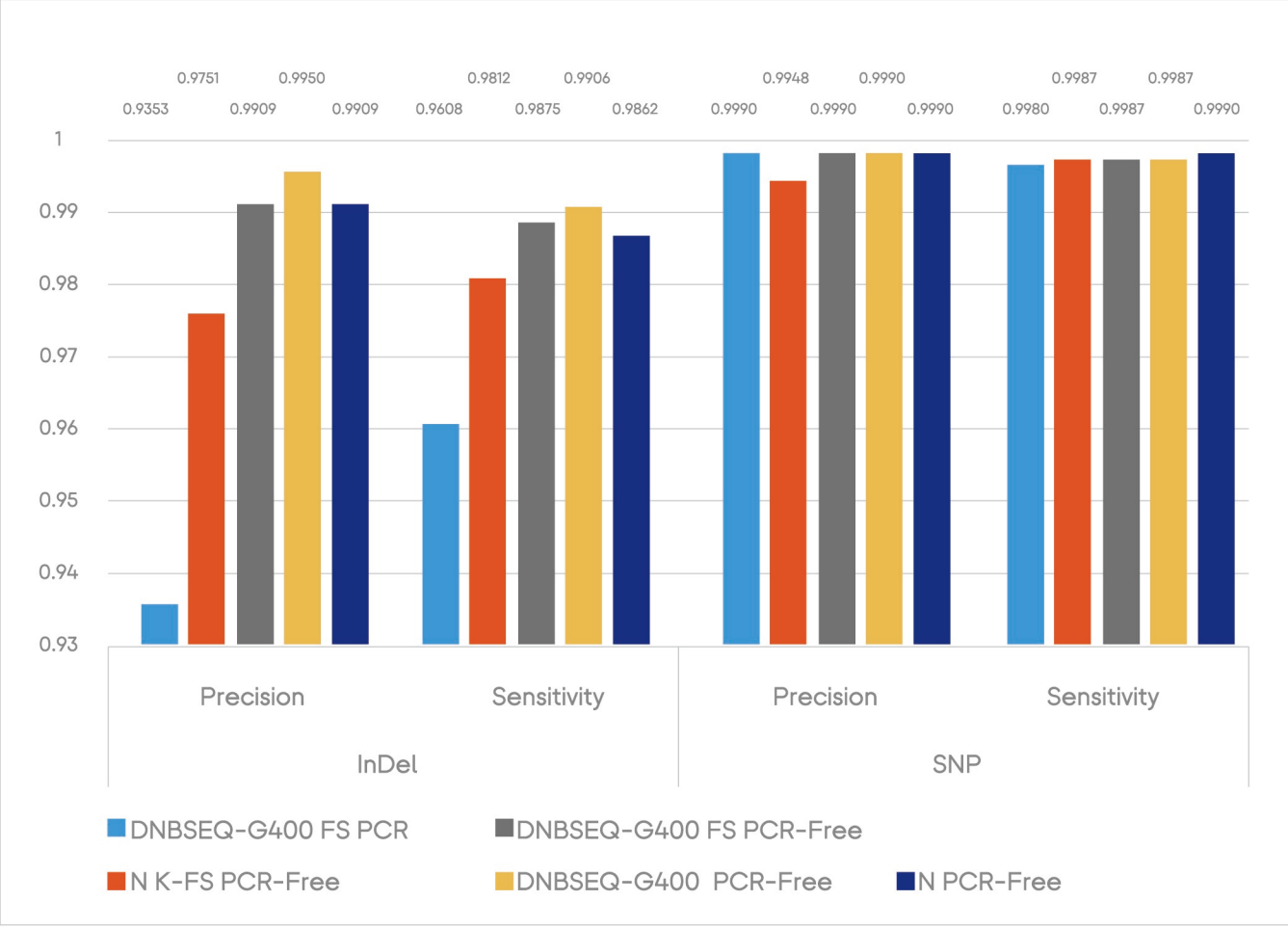


Figure 2 Comparison of the variant detection performance of FS PCR-free with different library prep kits and sequencing platform. The samples are NA12878 gDNA. Libraries were prepared with MGIEasy FS DNA Library Prep Set (DNBSEQ-G400 FS PCR)， vendor K FS Library Prep Set (with PCR-Free) （N K-FS PCR-free）， MGIEasy FS PCR-Free Library Prep Set (DNBSEQ-G400 FS PCR-free), MGIEasy PCR-Free Library Prep Set (DNBSEQ-G400 PCR-free) and vendor i PCR-Free Library Prep Kit (N PCR-free); And those libraries were sequenced on different sequencing platform, DNBSEQ-G400 (PE150) and N sequencing platform of vendor i （PE150）. The sequencing data were analyzed for variant detection performance with 30X average depth.

Long Reads sequencing

Since the development of Next Generation sequencing, the limitations of short read lengths in providing diploid information and in the detection of some genome variants have increasingly been recognized. In order to address these limitations, we introduce single tube Long Fragment Read (stLFR) (Wang et al. 2018), a technology based on DNA co-barcoding (Peters et al. 2014), that is adding the same barcode sequence to sub-fragments of the original long DNA molecule . With MGI's DNBseq™, the world's most accurate sequencing technology, stLFR enables high quality small variants calling, phasing of over 99% of the human genome, detection of structure variations, de novo assembly, and other long read applications.

Recommended MGI Process

Library preparation	Sequencing	Bioinformatics Analysis
MGIEasy stLFR Library Prep Kit	DNBSEQ-G400 DNBSEQ-T7	MegaBOLT Bioformics analysis accelerator (stLFR Analysis System)

MGIEasy stLFR Library Prep Kit is the world's first partition-less long fragment DNA co-barcoding library prep kit. Based on the patented DNA co-barcoding technology, that is adding the same barcode sequence to sub-fragments of the original long DNA molecule, stLFR could read the long range genetic information very accurately. With MGI's DNBSEQ™ the world's most accurate sequencing technology, stLFR enables high quality small variants calling, phasing diploid genomes, detection of structure variations and other long read applications.

Features

- Long range information with accurate short-read sequencing.
- No pre-amplification, high quality WGS libraries from only 1 ng DNA.
- Over 10 Mb of Haplotype Contig N50 and powerful detection of structure variations, such as deletions, inversions, translocations and insertions.
- One tube reaction, no need for nanoliter liquid handling or microfluidic systems
- Magnetic beads based, easily automatable solution.

Performance

▶ Read long fragment information

The MGIEasy stLFR can analyze long DNA fragments with an average length of 50-70 kb (maximum length up 300 kb. Benefiting from over 30 million molecule barcodes, more than 85% of long DNA fragments can be co-barcoded by single unique barcode. This makes stLFR co-barcoded reads analogous to direct single molecule sequencing, but without the high error rates and low throughput.

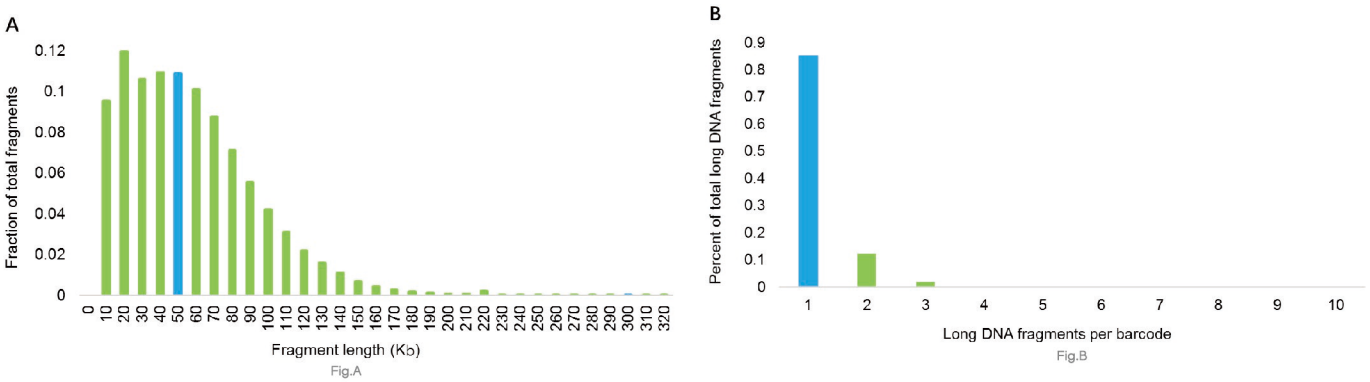


Figure Characteristics of stLFR. (A) Fragment length distribution. The typical fragment length of the stLFR kit is approximately 50 kb with molecules up to 300 kb. (B) Long fragment per barcode distribution. When starting from 1 ng of high molecular weight DNA, over 85% of DNA can be co-barcoded by a single unique barcode.

SNP & InDel calling

At 30X coverage, stLFR demonstrates high quality variant calling performance equivalent to that of standard short-read WGS libraries. Positive predictive values (PPV) and sensitivities of SNP detection above 0.99 are possible. In addition, F-measures of InDel detection above 0.95 are achievable (Figure 4).

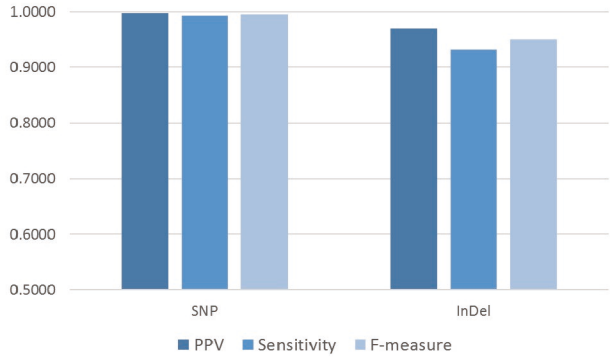


Figure 2 Variant calling performance with 30X coverage. The PPV and sensitivity was calculated after comparing variant calls to the high confidence truth dataset from Genome in a Bottle (GIAB).

Diploid genome phasing

stLFR co-barcoded reads can accurately assign heterozygous SNPs into phasing blocks with N50 sizes in excess of 10 Mb (Figure 3). This enables resolution of the combination of variants in regulatory and coding regions inherited from each parent for most genes in the human genome.

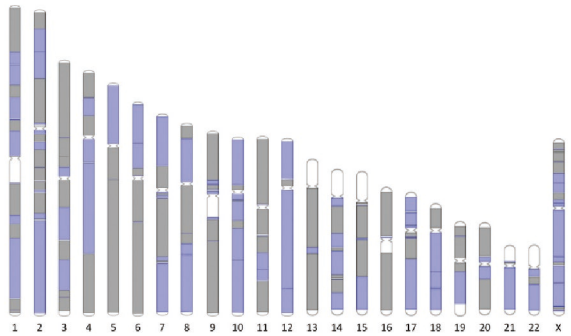


Figure 3 Ideogram of stLFR phase blocks.
With 40X coverage, the phasing block N50 of stLFR library was 34 Mb and 99.7% of heterozygous SNPs were phased.

Structural variation detection

stLFR barcode information can be used to detect multiple types of structural variations. Figure 4 demonstrates the detection of a balanced translocation between chromosomes 5 and 12 in a patient sample. Figure 6B shows the identification of an inversion within chromosome 2 in the GM20759 cell line.

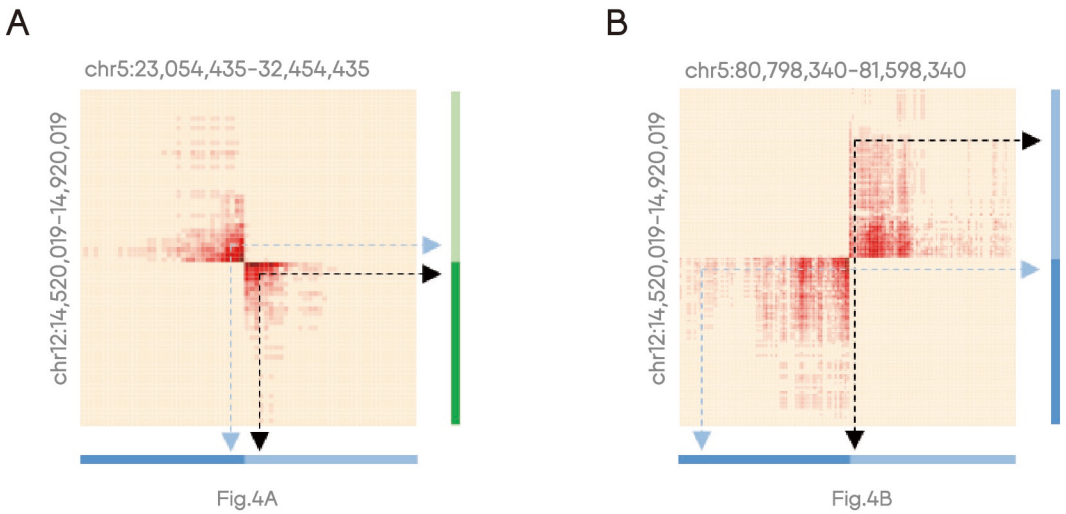


Figure 4 Structure variation detection using stLFR.
(A) Heat maps of barcode overlap between chromosomes 5 and 12 for a patient sample with a translocation and (B) GM20759, a cell line with a known transversion in chromosome 2.

Whole Exome Sequencing

Whole Exome Sequencing (WES) is a targeted sequencing technology that focuses on only 2% of an entire region's genome, but covers most of the gene coding regions and >99% of variations in the ClinGen database. Data from WES can be used to detect SNVs, indels, gene arrangements and copy number variations (CNVs). The decreasing cost of DNA sequencing has made WES more accessible, resulting in widespread adoption of the technology for genetic disease scanning, tumor molecular diagnosis and immunotherapy. The DNBSEQ-based WES solutions deliver full coverage of coding regions and help identify the exomes with the most significant phenotypic impact.

Recommended MGI Process

Library preparation	Sequencing	Bioinformatics Analysis
MGIEasy Exome Capture V4 Probe Set MGIEasy Exome Capture V5 Probe Set MGIEasy Exome FS Library Prep Set MGIEasy Exome Universal Library Prep Set	DNBSEQ-G400 DNBSEQ-G50 DNBSEQ-T7	MegaBOLT Bioformics analysis accelerator (WES Analysis System)
Automated sample preparation system MGISP-100 and MGISP-960		

WES library preparation

▶▶ MGIEasy Exome Capture V4 Probe Set

MGIEasy Exome Capture V4 Probe Set is the first exome capture product designed and synthesized by MGI. CCDS、GENCODE、RefSeq、miRBase database are used for the probe design. As a universal exome capture kit, it is compatible with DNBSEQ and other high-throughput sequencing platforms.

▶▶ MGIEasy Exome Capture V5 Probe Set

MGIEasy Exome Capture V5 Probe Kit is designed to cover areas covered by traditional exon probes, in addition to optimized probes targeting reproductive health, neonatal, cardiovascular, cerebrovascular, diseases, and hereditary tumors, as well as genes for monogenic diseases, pharmacogenetics, personalized genomics, hereditary deafness, immunodeficiency, and mitochondrial defects.

▶▶ MGIEasy Exome FS Library Prep Set

The MGIEasy Exome FS Library Prep Set can construct a WES library from 200-300 ng genomic DNA using an enzyme fragmentation method in conjunction with MGIEasy Exome Capture V4 Probe Set.

▶▶ MGIEasy Exome Universal Library Prep Set

The MGIEasy Exome Universal Library Prep Set can prepare 200-500ng genomic DNA into a WES library using a non-enzymatic fragmentation method combined with MGIEasy Exome Capture V4 Probe Set.

Product specification

Products	MGIEasy Exome FS Library Prep Set	MGIEasy Exome Universal Library Prep Set
Assay Time	<6 h	
Input Quantity	200-300 ng gDNA	200-500ng gDNA
Sample type	tissues, blood, tumor fresh tissue and FFPE sample	tissues, blood, tumor fresh tissue and FFPE sample
Fragmentation method	enzyme	physical
Compatible with major target-enrichment platforms	MGI Exome Capture, Agilent SureSelect, Roch Nimblegen SeqCap EZ, and IDT xGen Lockdown Probes	
Platform Compatibility	DNBSEQ-G400, DNBSEQ-G50	
Recommended Read Length	PE100, PE150 and etc	

Performance

▶ Excellent sequencing data quality

With low duplicate and high genome coverage rates, the MGI library sequencing data has a high effective utilization ratio.

Sample	Mapping rate	Duplicate rate	Capture rate	Average Depth	Coverage (≥1X)	Coverage (≥4X)	Coverage (≥10X)	Coverage (≥ 20X)
NA12878	99.71%	5.17%	58.14%	100.36%	99.61%	99.02%	97.36%	92.96%

Library was prepared from 1ug NA12878 gDNA using the Exome Capture V4 Universal Kit and the MGIEasy Exome Library Preparation Kit, and sequenced on DNBSEQ-G400 using PE100.

▶ The accuracy and sensitivity of SNP and Indel detection

MGI sequencing platforms demonstrated high accuracy and sensitivity on SNP and Indel detection.

Variation	True-pos	False-pos	False-neg	Precision	Sensitivity
SNPs	40,153	371	748	99.08%	98.17%
Indel	2,802	485	289	85.24%	90.65%

Libraries was prepared from 1ug NA12878 gDNA using the Exome Capture V4 Universal Kit and the MGIEasy Exome Library Preparation Kit, and sequenced on DNBSEQ-G400 using PE100

mRNA/ Total RNA Sequencing

RNA sequencing is a powerful method for comprehensive and rapid analysis of gene expression changes, examination of rare and novel transcripts, and discovery of alternate splicing events, gene fusions, SNPs and allele-specific expression in tissues or cells. It has been widely applied to many research fields, including biological, disease, drug development, agricul ture and environmental studies.

Recommended MGI Process

Library preparation ▶ **Sequencing**

MGIEasy RNA Library Prep Set
MGIEasy RNA Directional Library Prep Set
MGIEasy rRNA Depletion Kit

Automated sample preparation system MGISP-100 and MGISP-960

DNBSEQ-G400
DNBSEQ-G50
DNBSEQ-T7

mRNA/ Total RNA library preparation

MGI has developed the following library preparation products to support mRNA/ Total RNA sequencing.

▶▶ MGIEasy RNA Library Prep Set

The MGIEasy RNA Library Prep Set provides a solution for RNA quantification and transcriptome studies with all kinds of eukaryotic and prokaryotic species.

▶▶ MGIEasy RNA Directional Library Prep Set

The MGIEasy RNA Directional Library Prep Set enables strand specific studies to more accurately detect and quantify transcripts, reveal gene structure, and probe sense and antisense transcripts.

Product specification

Products	MGIEasy RNA Library Prep Set	MGIEasy RNA Directional Library Prep Set
Assay Time	<7 h	
Input Quantity	10 ng- 1 μg total RNA	
Objects	mRNA enriched by oligo (dT) beads, RNA enriched by rRNA depleted with rRNA depletion kit	
Sample types	tissues and FFPE sample	
Species Compatibility	Human, animals, plants, fungi and bacteria, such as mouse, rice, Arabidopsis, yeast and E.coli.	
Applications	RNA-Seq, Transcriptome Sequencing, total RNA sequencing, lncRNA Sequencing	
Platform Compatibility	DNBSEQ-G400, DNBSEQ-G50,DNBSEQ-T7	
Recommended Read Length	SE50/PE100/PE150	
Recommended sequencing data per sample	25 M raw reads（SE50）/8 Gb raw data（PE100/PE150）	
Detection of strand orientation	Not available	Available

Features

- As low as 10 ng total RNA per sample
- Compatible with human, animal, plant, fungi and bacteria samples with high integrity or degraded total RNA such as FFPE and plasma samples
- Easy to use with simple protocol and short operation time in 7 hours
- Provide two options of size selection of 150-300 bp or 200-400 bp
- Accurate and comprehensive mapping of transcripts
- High uniformity enables high uniform from transcripts 5 'to 3' ends and superior 3'end coverage
- Accurate identification transcript strand orientation to enhance transcript annotation and detect antisense transcripts using MGIEasy RNA Directional Library Prep Set

Performance

▶ High concordance of gene expression

The concordance of gene expression was measured using different input quantities of Universal Human Reference RNA (UHRR). Pearson and Spearman correlation tests demonstrated, stable and accurate results, yielding r values >0.995 for all data analyzed (Fig. 1).

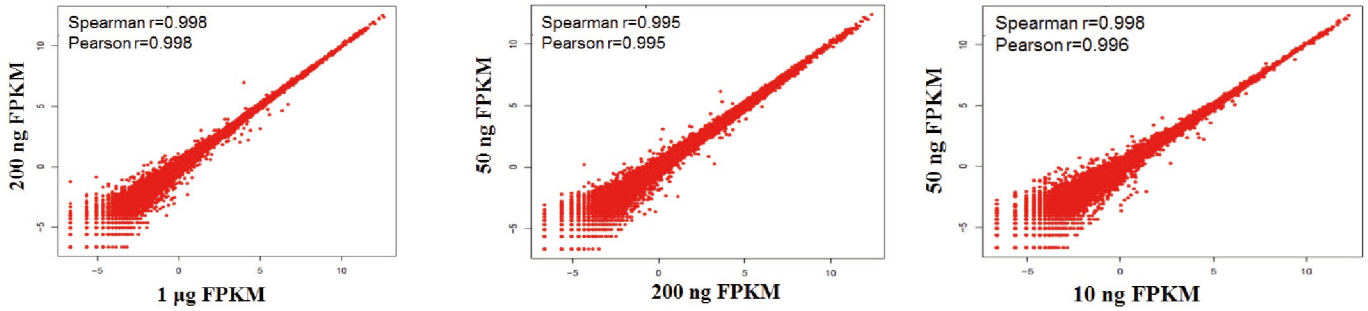


Fig Gene expression level reproducibility and concordance in different input amounts of total RNA

The libraries were prepared from an input of UHRR ranging from 10 ng -1 μg using the MGIEasy RNA Library Prep Set and sequenced on DNBSEQ-G400 at PE100 read-length. After data filtering, approximately 8 Gb of data was collected per library for analysis.

▶ High uniformity

The results in Fig demonstrate that the libraries constructed using MGIEasy RNA Library Prep Set have high uniformity throughout the entire length of a transcript with supreme 3'end coverage.

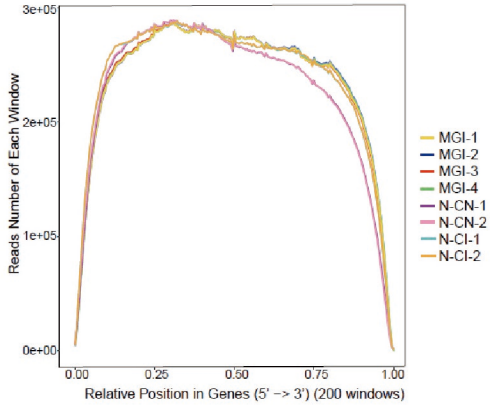


Fig Reads randomness of different platforms and kits

The libraries MGI -1/2/3/4 were generated from UHRR using the MGIEasy RNA Library Prep Set and sequenced on DNBSEQ-G400 at PE150 read-length. The libraries N-CN-1/2 and N-CI-1/2 were prepared using Company-N and Company-I kits respectively and sequenced on "N" platform at PE150 read-length. After data filtering, about 10 Gb were collected per library for analysis. The analysis of sequencing data was based on the same instruction.

▶ High-Quality Stranded Information

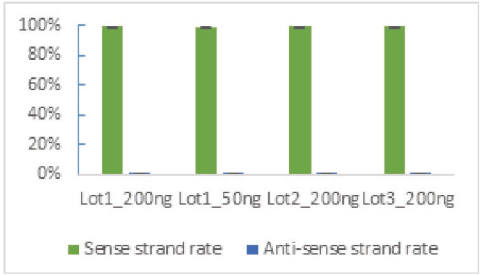


Fig.3a Sense strand rate of MGIEasy RNA Directional Library Prep Set

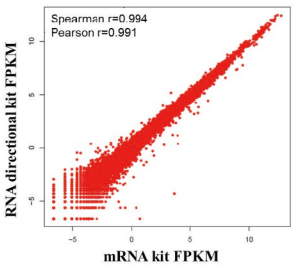


Fig.3b Concordance of MGIEasy RNA Directional Library Prep Set and MGIEasy RNA Library Prep Set RNA

The percentage of unique mapped reads that present accurate strand origin information from libraries of enriched UHRR poly(A) mRNA measured above 99% (Fig. 3a), confirming accurate preservation of transcript directionality. RNA abundance is reflected in the high consistency between libraries constructed using MGIEasy RNA Directional Library Prep Set and MGIEasy RNA Library Prep Set (Fig. 3b).

Fig Libraries were prepared with from an enriched poly(A) mRNA of UHRR using the MGIEasy RNA Directional Library Prep Set and sequenced on DNBSEQ-G400 at PE100 read-length. After data filtering, approximately 8 Gb were collected per library for analysis.

MGIEasy rRNA Depletion Kit

The MGIEasy rRNA Depletion Kit can efficiently deplete rRNA from 10 ng-1 μg of total RNA from human, mouse and rat samples, including HMW material and even degraded samples (such as FFPE), to enrich messenger and non-coding RNA, increasing the proportion of useful data.

Product specification

Assay Time	~2 hours
Input Quantity	10 ng - 1 μg of total RNA
Sample types	tissues and FFPE sample
Species Compatibility	Human, mouse, rat
Applications	total RNA sequencing, lncRNA Sequencing

Performance

▶ High efficiency on 99% rRNA depletion

Compared with other depletion kits, MGIEasy rRNA Depletion Kit has greater efficiency.
Less than 1% of the initial quantity of rRNA remains after a single depletion.

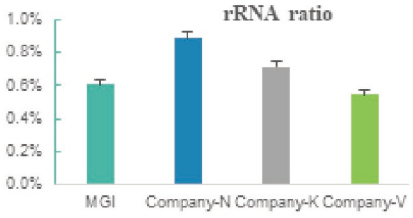


Fig.4a Depletion efficiency of different depletion kits

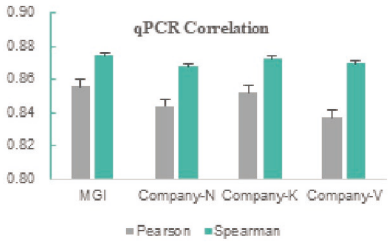


Fig.4b qPCR correlation of different depletion kits

Fig.4 Ribosomal RNA in 200 ng UHRR was depleted with different rRNA depletion kits. The resulting rRNA-depleted RNA were prepared to construct libraries using the MGIEasy RNA Directional Library Prep Set and sequenced on DNBSEQ-G400 at PE100 read-length. After data filtering, approximately 8 Gb of data was used per library for analysis.

Small RNA Sequencing

Small RNAs are an endogenous non-coding single-stranded RNA molecules formed of 20-24 nucleotides and include miRNAs, siRNA and piRNA, snRNA and snoRNA. Small RNAs are involved in gene silencing and post-transcriptional regulation of gene expression through RNA-RNA interactions. They are crucial regulators in cell growth and differentiation in development processes. Small RNA Sequencing is a technique used to isolate and sequence small RNA for biological, disease, drug development, plant and animal research.

The DNBSEQ-based Small RNA sequencing solution uses supported kits to prepare single-stranded circularized cDNA libraries from total RNA, sequence those libraries, and generates base calling sequencing data of hundreds of small RNA sequences.

Recommended MGI Process

Library preparation	Sequencing
MGIEasy Small RNA Library Prep Kit	DNBSEQ-G400
MGISP-100	

▶▶ MGIEasy Small RNA Library Prep Kit

The MGIEasy Small RNA Library Prep Kit provides an efficient solution for generating libraries suitable for MGI high-throughput sequencing platforms from 10 ng - 1 μg of total RNA

Product specification

Assay Time	~13 hours
Input Quantity	10 ng - 1 μg of total RNA
Sample types	Tissues
Species Compatibility	Human, animals and plants, such as mouse, rice and Arabidopsis
Applications	Small RNA Sequencing
Platform Compatibility	DNBSEQ-G400
Recommended Read Length	SE50
Recommended sequencing data per sample	25 M raw reads

Features

- As low as 10 ng total RNA per sample
- Compatible with various species (human, animal and plant) samples
- Provides two options of size selection based on either bead or gel purification to meet a wide range of library needs

Methylation Sequencing

DNA methylation is an important epigenetic phenomenon. It is essential to know every methylated cytosine in the genome to understand temporal and spatial gene expression and chromatin remodeling. The DNBSEQ-based Whole Genome Bisulfite Sequencing (WGBS) solution leverages bisulfite conversion during library preparation and high throughput sequencing technology. The solution enhances epigenetic research with high accuracy, producing a genome-wide DNA methylation map.

Recommended MGI Process

Library preparation	Sequencing
MGIEasy Whole Genome Bisulfite Sequencing Library Prep Kit	DNBSEQ-G400 DNBSEQ-T7

▶▶ MGIEasy Whole Genome Bisulfite Sequencing Library Prep Kit

The MGIEasy Whole Genome Bisulfite Sequencing Library Prep Kit is specifically designed for preparing WGBS libraries for MGI high-throughput sequencing platform from 10-100 ng DNA fragments. This kit is compatible with a variety of sample types, including challenging FFPE samples, and can be widely applied to many applications, such as cell development and differentiation research, human disease research, animal and plant breeding, species and population evolution studies.

Product specification

Assay Time	~8 hours
Input Quantity	10- 100 ng fragmented DNA (200-500 bp)
Sample types	tissues and FFPE sample
Species Compatibility	Human, animals, plants and fungi with reference genome sequence, such as Arabidopsis and yeast
Fragmentation method	Physical (Fragmentation is not needed for plasma free DNA)
Applications	Methylation sequencing
Platform Compatibility	DNBSEQ-G400,DNBSEQ-T7
Recommended Read Length	PE100, PE150
Recommended sequencing data per sample	30X clean base (e.g. 90G clean bases for human)

Features

- As low as 10 ng fragmented DNA per sample
- Compatible with human, animal, plant and fungi samples with high integrity or degraded total RNA such as FFPE and plasma samples
- Low duplicate rate and high genome mapping rate enabling high usability of sequencing data
- Excellent CpG coverage and accuracy of methylation detection

Performance

▶ High usability of sequencing data

MGI_NA library was prepared from 100 ng NA12878 gDNA using the MGIEasy Whole Genome Bisulfite Sequencing Library Prep Kit and sequenced at DNBSEQ-G400. X_NA library was constructed using the corresponding products for X NGS platform and sequenced on X platform. The result shows that MGI_NA has higher mapping rate and lower duplication rate, resulting in a greater sequencing depth at the same clean data amount, enabling high usability of sequencing data with the MGIEasy Whole Genome Bisulfite Sequencing Library Prep Kit (Table 1).

Table 1 Mapping and coverage rate

Sample ID	X_NA	MGI_NA	Sample ID	X_NA	MGI_NA
input DNA amounts (ng)	1000	100	Duplication rate (%)	28.49	16.29
Clean data amount (Gb)	111	111	Mismatch Rate (%)	0.71	0.72
Read length	PE150	PE100	BS Conversion Rate (%)	99.45	99.55
Clean rate (%)	99.47	96.26	Average sequencing depth	23X	28X
Mapping Rate (%)	87.04	93.52	Coverage at least 4X (%)	97.24	97.08
Unique mapping rate (%)	84.15	89.29	Coverage at least 10X (%)	93.29	94.25
			Coverage at least 20X (%)	61.73	78.15

▶ High CpG coverage

Data produced from NA12878 DNA using Infinium® MethylationEPIC BeadChip was used to compare the CpG coverage between MGI_NA and X_NA. The result shows that MGI_NA data has higher concordance of CpG coverage with microarray compared with X_NA and detects more CpG sites with sequencing depth at 10X and 20X.

Table 2 Analysis data on Concordance of CpG coverage

Sample	Sequencing depth	Genome CG number	Concordance CG rate* (%)	Unique CG rate**(%)
MGI_NA	4X	56,434,896	83	8
X_NA	4X			9
MGI_NA	10X		48	35
X_NA	10X			18
MGI_NA	20X		15	70
X_NA	20X			15

*The concordance CG rate of the two platforms in all CG sites detected

**The percentage of CG sites detected by only one of the NGS platforms in all CG sites detected

▶ Excellent accuracy of methylation detection

The genome methylation rate of CpG pattern of MGI_NA is close to that of X_NA. The methylation rate in the CH pattern (CHG and CHH) of NA12878 (b-lymphocyte) should be less than 1%. The chromosomal methylation rates of MGI_NA datasets are at the expected level, while those of X_NA dataset are considerably higher (mostly around 3%), indicating high accuracy regarding methylation levels (Fig.5).

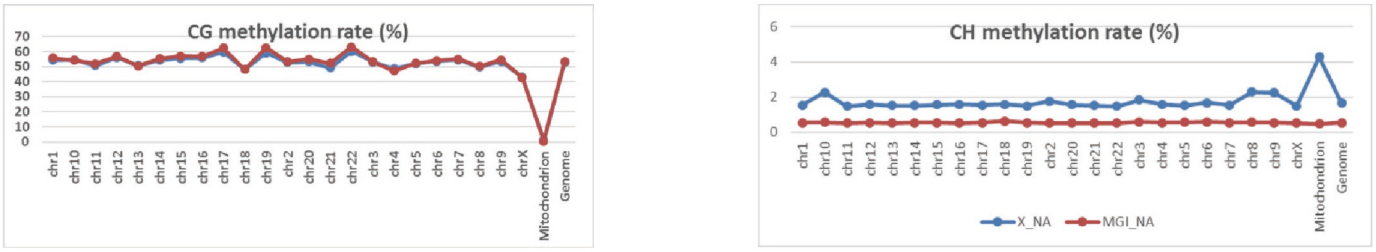


Fig.1 Methylation rate of CG pattern and CH pattern

Metagenomics Sequencing

Metagenomics sequencing is a method for directly sequencing separate microbial populations in pooled samples without isolating and culturing each individual population. It can be used to study the community structure, species classification, phylogenetic evolution, gene function and metabolic network of environmental microorganisms, etc. The method is an effective tool for environmental microbe studies and has been widely used in applications, such as the identification of pathogenic infections, where traditional methods are not useful. MGI high throughput sequencing delivers an accurate and rapid pathogen identification solution for diagnosing infectious disease using microbial genetic information.

Recommended MGI Process

Library preparation	Sequencing	Bioinformatics Analysis
MGIEasy FS DNA Library Prep Set MGIEasy Universal DNA Library Prep Set MGIEasy Cell-free DNA Library Prep Set	DNBSEQ-G400 DNBSEQ-G50 DNBSEQ-T7	Fast identification of pathogenic infections
Automated sample preparation system MGISP-100 and MGISP-960		

Metagenomics Sequencing library preparation

MGI has developed a series library preparation kits for Metagenomics Sequencing.

Product specification

Products	MGIEasy FS DNA Library Prep Set	MGIEasy Universal DNA Library Prep Set	MGIEasy Cell-free DNA Library Prep Set
Assay Time	<5.5 h	<4.5 h	<3 h
Input Quantity	5- 400 ng gDNA	0.5-50 ng fragmented DNA (200-500 bp)	2-6 ng cfDNA or DNA fragment (150-250 bp)
Sample types	Meta DNA, cfDNA	Meta DNA, cfDNA, Amplicons	cfDNA from plasma, nasopharyngeal swab, sputum, alveolar lavage, cerebrospinal fluid, stool, etc.
Species Compatibility	Human, animals, plants, fungi and bacteria, such as mouse, rice, Arabidopsis, yeast and E.coli.		Meta sample from Human, animals and environment sample
Fragmentation method	Enzyme	Physical	No need
Automation option	Automation-friendly: MGISP-100,MGISP-960		
Applications	Human genome sequencing, Plant and Animal genome sequencing, Microbial genome sequencing, Metagenomics sequencing,Pathogen identification		Pathogen identification
Platform Compatibility	DNBSEQ-G400, DNBSEQ-G50,DNBSEQ-T7		
Recommended Read Length	SE50, PE100, PE150 and etc.		SE50, PE100

RAD Sequencing

▶▶ MGIEasy RAD Library Prep Kit

The MGIEasy RAD Library Prep Kit supports library preparation for reduced-representation sequencing on the MGI high throughput sequencing platform. The kit provides reagents to prepare a library from 1 μg animal or plant gDNA. This kit offers restriction enzyme digestion, adaptor ligation, samples pooling, size selection using magnetic beads, PCR, purification and single-strand circularization.

Product specification

Assay Time	~7 hours
Input Quantity	1 μg gDNA
Sample types	Tissues
Species Compatibility	Animals and plants, such as pig and rich
Fragmentation method	Enzyme
Applications	Agriculture research
Platform Compatibility	DNBSEQ-G400
Recommended Read Length	PE100

Features

- Compatible with animal and plant samples. Not limited to reference genome. Effectively reduces the genomic complexity of different species.
- Simple and fast operation without ER/A tailing.
- High throughput library preparation solution enables the pooling of 16-64 samples.

Library preparation modules

In addition to high quality Library Preparation Kit, MGI offers flexible library preparation modules and reagents. Library preparation modules are used in a single step or multiple steps. They can be used with other kits to meet the requirements of the workflow. A variety of reagents are available to meet the full range of project needs.

Library preparation module

Type	Products	Configuration	Version	Catalog No.
Adapter	MGIEasy DNA Adapters-96 (96-well plate) Kit	96 × 10 μL	V1.0	1000005282
Bead Purification	MGIEasy DNA Purification Magnetic Bead Kit	50 mL	V1.0	1000005279
Circularization	MGIEasy Circularization Kit	16 RXN	V2.0	1000005259

Ordering information

Sequencing platform

Classifications	Product	Catalog No.
Genetic Sequencer	Genetic Sequencer DNBSEQ-G50RS	900-0001822-00
	Genetic Sequencer DNBSEQ-G400RS	900-000170-00
	Genetic Sequencer DNBSEQ-T7RS	900-000128-00
DNB Loader	Portable DNB Loader MGIDL-200H	900-000115-00
	MGIDL-T7RS	900-000261-00
Automated sample preparation system	High-throughut Automated Sample Preparation Systm MGISP-960RS	900-000152-00
	DNA Sequencing Library Preparation System MGISP-100RS	900-000206-00

Sequencing reagents

Sequencing platform	Product	Configuration	Catalog No.
DNBSEQ-G400RS	DNBSEQ-G400RS High-throughput Sequencing Set	SE50	1000016941
	DNBSEQ-G400RS High-throughput Sequencing Set	PE100	1000016950
	DNBSEQ-G400RS High-throughput Sequencing Set	PE150	1000016952
	DNBSEQ-G400RS High-throughput Sequencing Set	PE200	1000016955
	CoolMPS High-throughput Sequencing Reagent Set	SE50	1000017992
	CoolMPS High-throughput Sequencing Reagent Set	SE100	1000016933
	CoolMPS High-throughput Sequencing Reagent Set	PE100	1000016935
DNBSEQ-G50RS	DNBSEQ-G50RS High-throughput Sequencing Set	SE50	1000016959
	DNBSEQ-G50RS High-throughput Sequencing Set	PE50	1000016963
	DNBSEQ-G50RS High-throughput Sequencing Set	PE100	1000016965
DNBSEQ-T7RS	DNBSEQ-T7RS High-throughput Sequencing Set	PE100	1000016105
	DNBSEQ-T7RS High-throughput Sequencing Set	PE150	1000016106

Libraries preparation kits

Applications	Product	Configuration	Catalog No.
Whole Genome Sequencing	MGIEasy Universal DNA Library Prep Set	16 RXN（with 16rxn circularization）	1000006985
		96 RXN（with 16rxn circularization）	1000006986
		96 RXN（with 96rxn circularization）	1000017571
	MGIEasy FS DNA Library Prep Set	16 RXN（with 16rxn circularization）	1000006987
		96 RXN（with 16rxn circularization）	1000006988
		96 RXN（with 96rxn circularization）	1000017572
	MGIEasy PCR-Free DNA Library Prep Set	16 RXN	1000013452
		96 RXN	1000013453
	MGIEasy FS PCR-Free DNA Library Prep Set	16 RXN	1000013454
		96 RXN	1000013455
Whole Exome Sequencing	MGIEasy Exome Capture V4 Probe Set	16 RXN	1000007745
	MGIEasy Exome Capture V5 Probe Set	16 RXN	1000007746
	MGIEasy Exome Capture Accessory Kit	16 RXN	1000007743
	MGIEasy Exome Universal Library Prep Set	16 RXN	1000009657
	MGIEasy Exome FS Library Prep Set	16 RXN	1000009658
mRNA/Total RNA Sequencing	MGIEasy RNA Library Prep Set	16 RXN	1000006383
		96 RXN	1000006384
	MGIEasy RNA Directional Library Prep Set	16 RXN	1000006385
		96 RXN	1000006386
	MGIEasy rRNA Depletion Kit	32 RXN	1000005953
Small RNA Sequencing	MGIEasy Small RNA Library Prep Kit	24 RXN	1000005269
WGBS	MGIEasy Whole Genome Bisulfite Sequencing Library Prep Kit	16 RXN	1000005251
cfDNA	MGIEasy Cell-free DNA Library Prep Set	48 RXN	1000007037
RAD-seq	MGIEasy RAD Library Prep Kit	64 RXN	1000005242

For Research Use Only. Not for use in diagnostic procedures.

Innovate Life Technology
MGI Tech Co., Ltd

High Throughput Sequencing Platform

Innovate Life Technology

MGI Tech Co., Ltd

▶ Contact Us

Address: Building 11, Beishan Industrial Zone,
Yantian District, Shenzhen, China, 518083

Website: en.mgitech.cn
Email: MGI-service@genomics.cn
Telephone: (+86)4000-966-988

▶ Copyright Disclaimer

Version: December 2019

Official Wechat

Official Website