MGIEasy single tube Long Fragment Read (stLFR) solution

Abstract

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 (Figure 1). With the world leading MGI' s DNBseq™ 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.

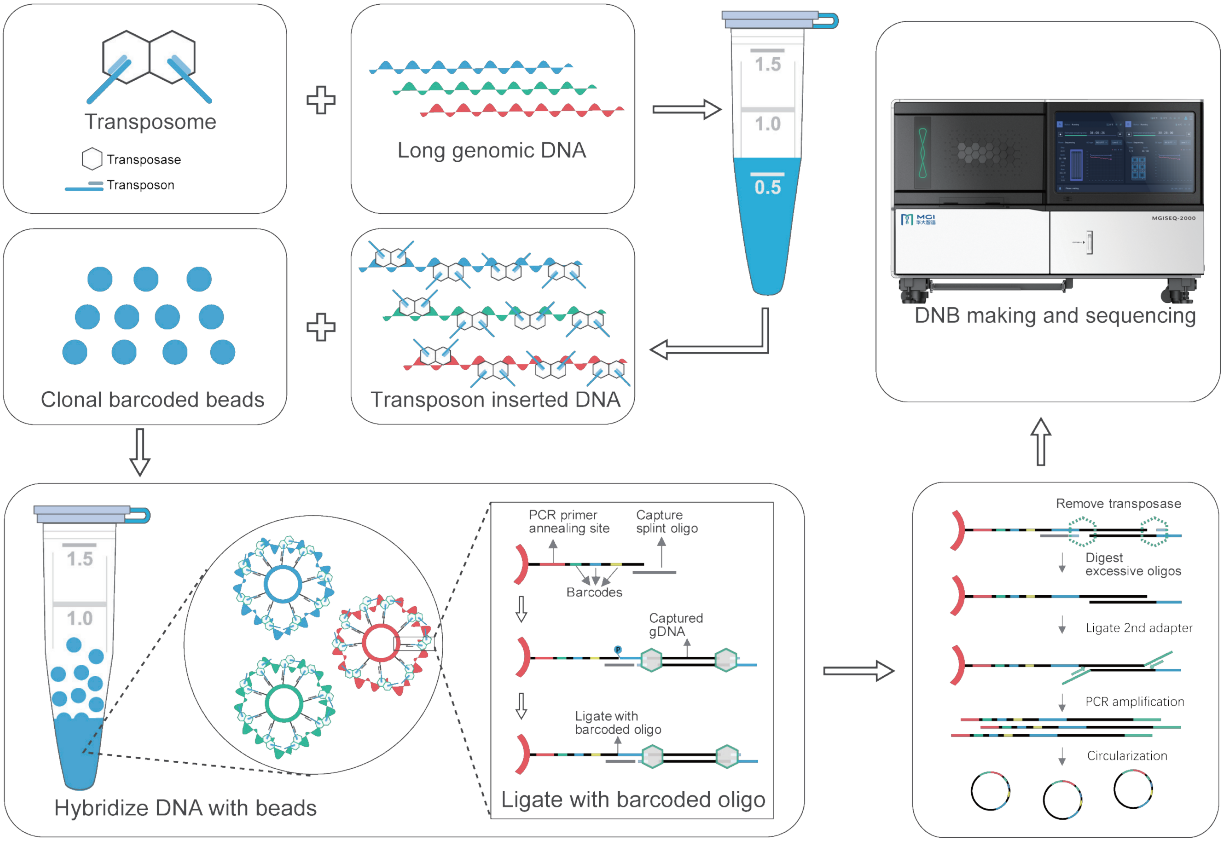

Figure1. stLFR library process overview. (A)stLFR starts with inserting transposons into long genomic DNA followed by hybridization of the transposon integrated DNA onto clonally barcoded beads. After a few additional library processing steps the co-barcoded sub-fragments are ready for sequencing.

Highlights

- 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.

MGIEasy stLFR Library Prep Kit

MGIEasy stLFR Library Prep Kit is the world' s first partition-less long fragment DNA co-barcoding library prep kit. The kit enables the entire process to be performed in a single microcentrifuge tube without any physical compartments or emulsion generation. In addition, the MGIEasy stLFR Library Prep Kit is capable of generating 30 million unique barcodes (Wang et al. 2018) per sample without requiring any special equipment to perform co-barcoding.

Technical 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 (Figure 2A))[1]. Benefiting from over 30 million molecule barcodes, more than 85% of long DNA fragments can be co-barcoded by single unique barcode (Figure 2B) [2]. This makes stLFR co-barcoded reads analogous to direct single molecule sequencing, but without the high error rates and low throughput.

1: The DNA fragment length measurement highly depends on the length of original isolated DNA.
2: The long DNA fragment per barcode measurement highly relies on the length of original isolated DNA and the input amount of DNA.

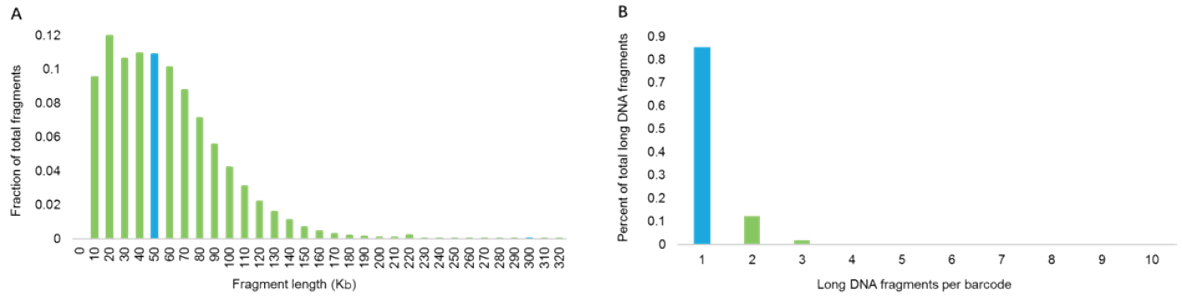

Figure 2. Characteristics of stLFR. (A) Fragment length distribution. The typical fragment length of the stLFR kit is approximately 50 kb with molecules up to 300 kb. (B) Long fragment per barcode distribution. When starting from 1 ng of high molecular weight DNA, over 85% of DNA can be co-barcoded by a single unique barcode.

Coverage uniformity

Starting with 1 ng of high molecular weight DNA, stLFR library can obtain comparable sequencing coverage uniformity with that of standard short-read whole genome sequencing (WGS) libraries starting from 100 fold more DNA (Figure 3).

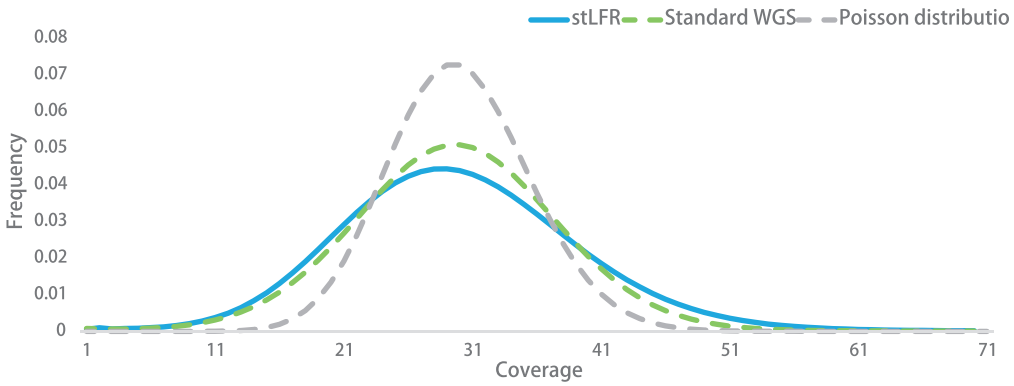

Figure 3. Sequencing uniformity of an stLFR library (blue line), a standard whole genome sequencing (WGS) library (green dash line), and the Poisson distribution (grey dash line) all at 30X genome coverage are plotted.

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).

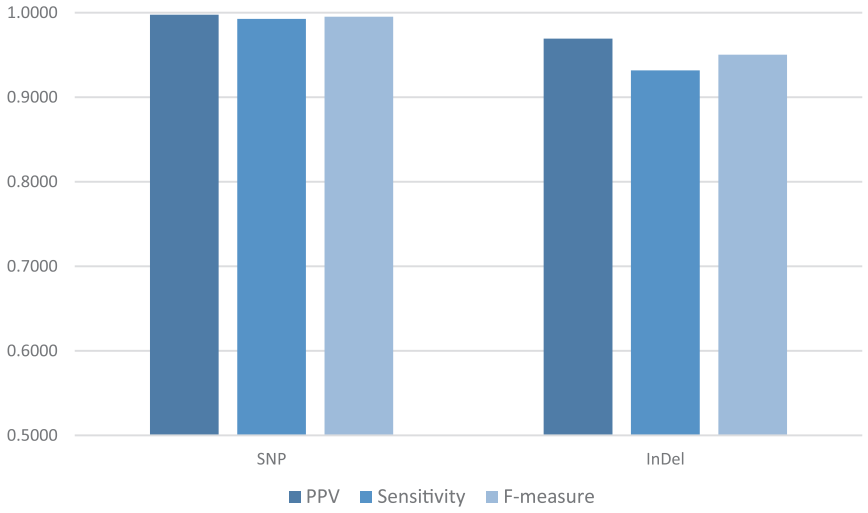

Figure 4. Variant calling performance with 30X coverage. The PPV and sensitivity was calculated after comparing variant calls to the high confidence truth dataset from Genome in a Bottle (GIAB).

Diploid genome phasing

stLFR co-barcoded reads can accurately assign heterozygous SNPs into phasing blocks with N50 sizes in excess of 10 Mb (Figure 5)[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.

3: The phasing block N50 measurement highly relies on the original fragment size of isolated DNA and total sequencing coverage.

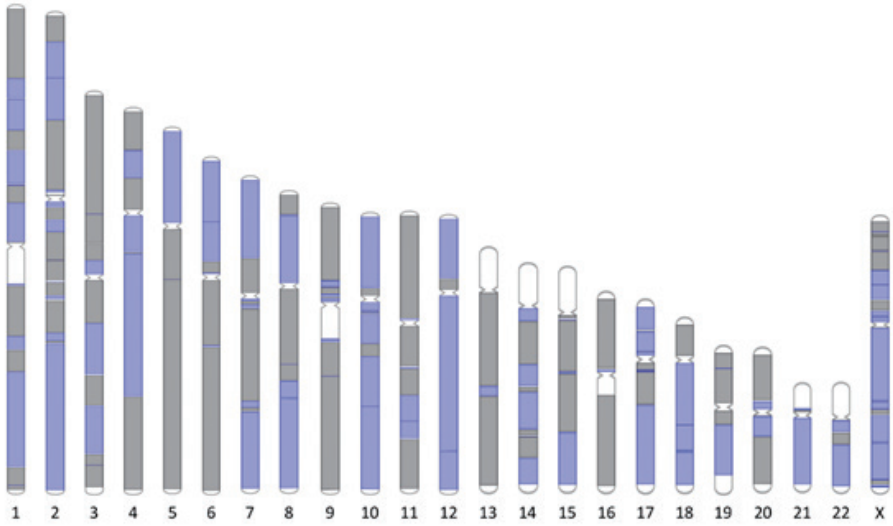

Figure 5. Ideogram of stLFR phase blocks. With 40X coverage, the phasing block N50 of stLFR library was 34 Mb and 99.7% of heterozygous SNPs were phased.

Structural variation detection

stLFR barcode information can be used to detect multiple types of structural variations. Figure 6A 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.

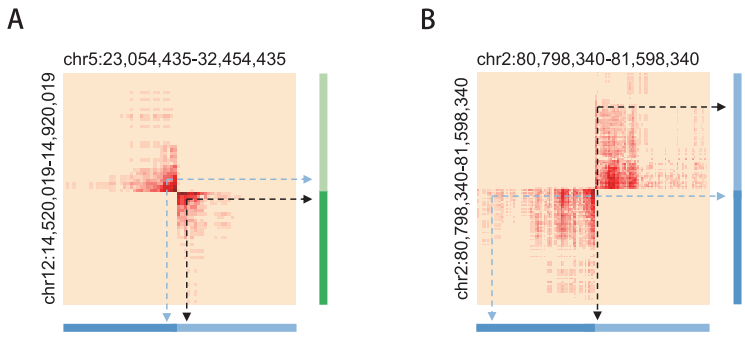

Figure 6. Structure variation detection using stLFR. (A) Heat maps of barcode overlap between chromosomes 5 and 12 for a patient sample with a translocation and (B) GM20759, a cell line with a known transversion in chromosome 2.

Summary of variant calling ability of stLFR library

Variant type	Call and Phased
SNVs	Yes
InDels	Yes
Structural variations (SVs including translocations, inversions, deletions and insertions)greater than 20 Kb	Yes
Deletions and duplications greater than 5 Kb (CNVs)	Yes
Deletions greater than 50 bp(CNVs)	Coming soon

References

Peters BA, Liu J, Drmanac R. 2014. Co-barcoded sequence reads from long DNA fragments: a cost-effective solution for "perfect genome" sequencing. *Frontiers in genetics* 5: 466.

Wang O, Chin R, Cheng X, Wu M, Mao Q, Tang J, Sun Y, Lam H, Chen D, Zhou Y et al. 2018. Single tube bead-based DNA co-barcoding for cost effective and accurate sequencing, haplotyping, and assembly. *bioRxiv.*

Contact Us

MGI Tech Co., Ltd
Add.: Building11, Beishan Industrial Zone,
Yantian District, Shenzhen, CHINA 518083
Email: MGI-service@genomics.cn
Website: www.mgitech.cn
Tel: 4000-966-988
Revision: 2018.09

https://www.linkedin.com/company/mgi-bgi

@MGI_BGI