Multi-throughput Iso-Seq

Take Full Advantage of Long Read Sequencing with the Same Volume of Data - But More Transcripts

Introduction

"Isoform Sequencing" (Iso-seq) developed by Pacific Biosciences (PacBio), is based on long-read sequencing technology. The unique long-read sequencing feature allows this method to identify new isoforms with extraordinary precision. The Iso-Seq application generates full-length cDNA sequences — from the 5' end of transcripts to the poly-A tail — eliminating the need for transcriptome reconstruction using isoform-inference algorithms. The Iso-Seq method generates accurate information about alternatively spliced exons and transcriptional start sites.

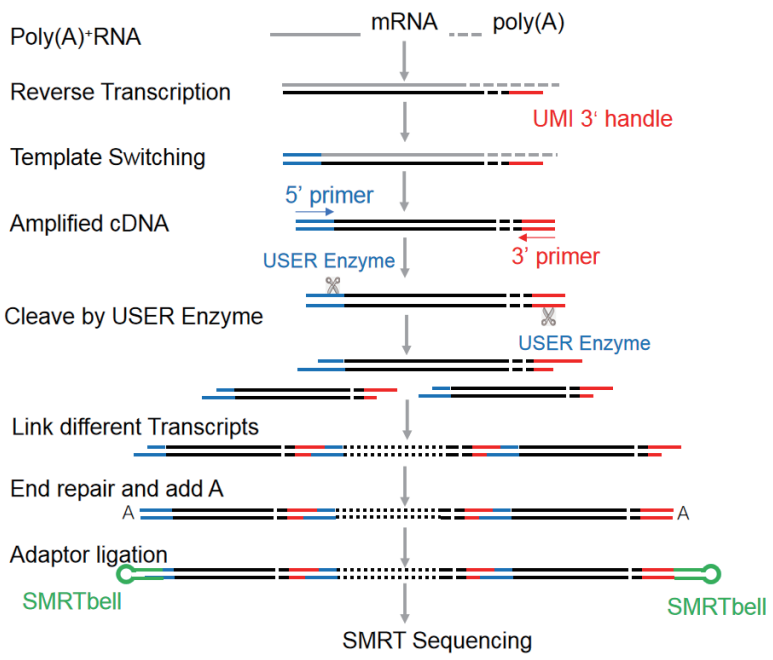
As long-read sequencing technology evolves, the sequencing reads have become longer and longer. However, most of the transcripts (inserts) are shorter than 2kb, and the read length of the PacBio platform is much longer than the insert size. So, each insert will be sequenced multiple times by repeating cycles until the termination of sequencing reaction*. More sequencing data with longer reads will not detect more transcripts, in fact possibly even fewer.

* PB Iso-Seq is sequenced with the circular consensus sequencing (CCS) mode.

To take full advantage of long read sequencing, BGI has developed a unique protocol called 'Multi-throughput Iso-Seq'.

Multi-throughput Iso-Seq Protocol

The total RNA is first converted into full length cDNA by reverse transcription. Then, the cDNA is cleaved by USER enzyme to enable the ligation of different transcripts. As a result, a circular consensus sequence (CCS) will contain more than one inserts, each of which links to a different unique molecular identifier (UMI). Therefore, the same volume of sequencing data will detect more transcripts.



Workflow of Multi-throughput Iso-Seq

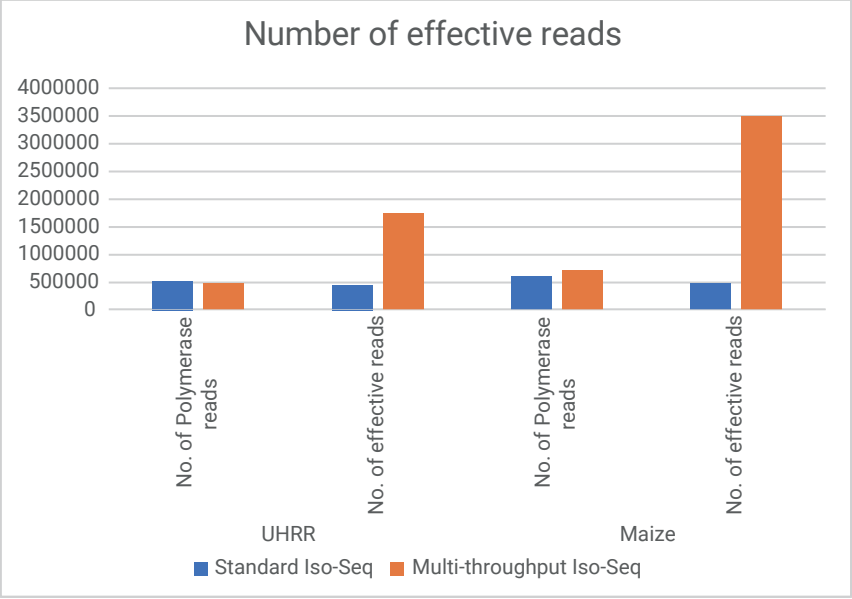
Key Benefits

· Gene/Transcript quantification: UMI technology enables highly accurate quantification of genes/transcripts.
· Higher Transcripts Detection rate: The same volume of sequencing data will generate up to 5x more effective reads, and can even double the transcript detection rate.

Example Data

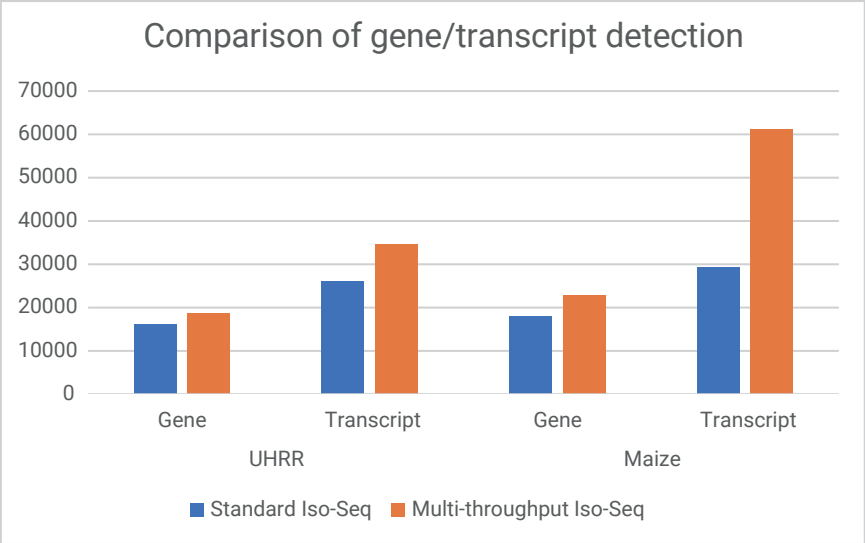
Greater Transcript Detection

Compared with Standard Iso-Seq, Multi-throughput Iso-Seq can obtain 3-5 times more effective reads and can even detect double the amount of transcripts with the same volume of sequencing data.



More effective reads detected by Multi-throughput Iso-Seq. In the tested human UHRR samples, the total number of polymerase-reads from Multi-throughput Iso-Seq was slightly less than the total reads from the standard Iso-seq (93%), but the Multi-throughput Iso-Seq was able to generate 3.8 times as many effective reads as the standard Iso-Seq. We also tested maize samples. The Multi-throughput Iso-Seq was able to yield 6.2 times more effective reads than the standard Iso-Seq with almost the same amount of total polymerase-reads (1.2 times total polymerase reads of standard Iso-Seq). An average of 3.5 and 4.9 effective reads in one polymerase-read were achieved in human and maize samples respectively.

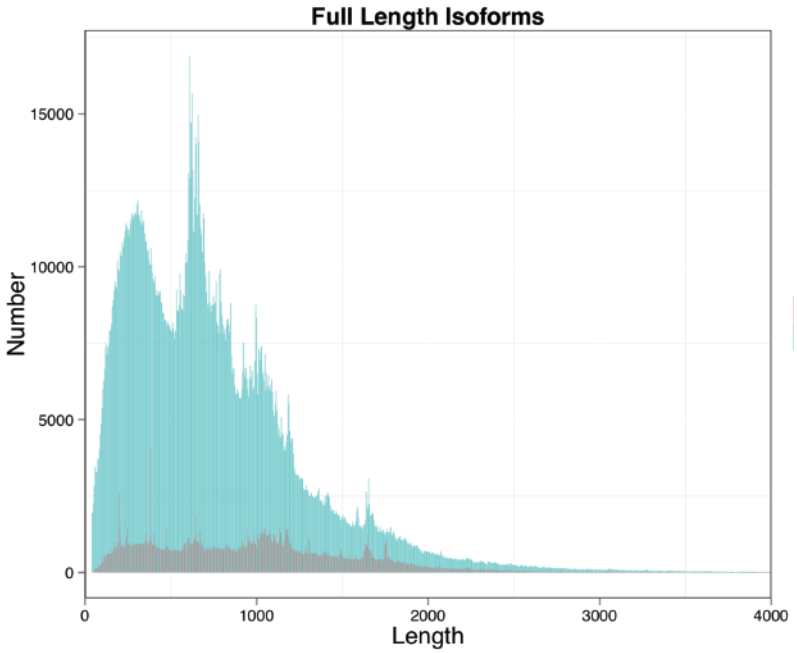
Notes:
Total sequencing data amount as follows:
UHRR-9.47 Gb (Standard Iso-Seq); 9.57 Gb (Multi-throughput Iso-Seq);
Maize-11.04 Gb (Standard Iso-Seq); 19.3 Gb (Multi-throughput Iso-Seq).
UHRR: UHRR means universal human reference RNA, which is composed of total RNA from 10 human cell lines and designed by Agilent for gene-profiling experiments.
Polymerase-reads : The number of polymerases generated high quality reads. Polymerase reads will be then trimmed to preserve only the high-quality region, which includes bases from adaptors and single or multiple passes around a circular template.
Effective reads: Each cDNA template molecule is considered as an “insert” and each pass through the insert is called a effective read. A polymerase read made by Multi-throughput Iso-Seq can contain more than one unique inserts.



More genes and transcripts detected by Multi-throughput Iso-Seq. The amount of genes detected by Multi-throughput Iso-Seq was 15% and 27% more than that of Standard Iso-seq in UHRR and maize respectively. And the amount of transcripts detected by Multi-throughput Iso-Seq was 33% and 109% more than that of Standard Iso-seq in UHRR and maize respectively.

No Transcript Length Bias

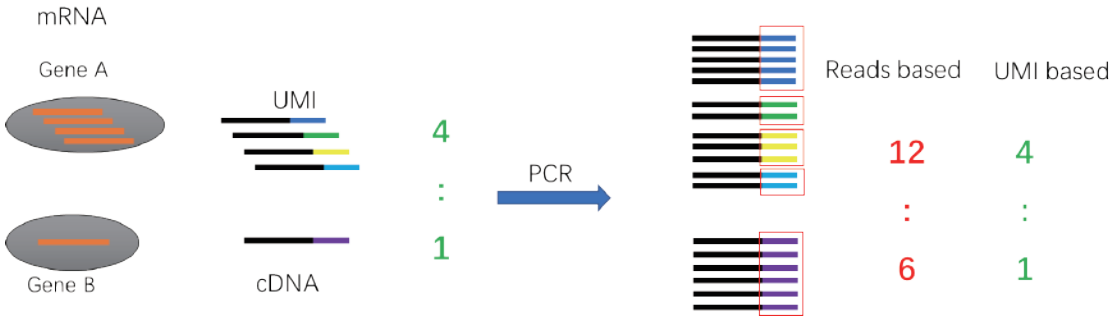
Different length transcripts are linked randomly in a Multi-throughput Iso-Seq library. Compared to Standard Iso-Seq, Multi-throughput Iso-Seq can detect the transcripts at various length ranges, without length bias.



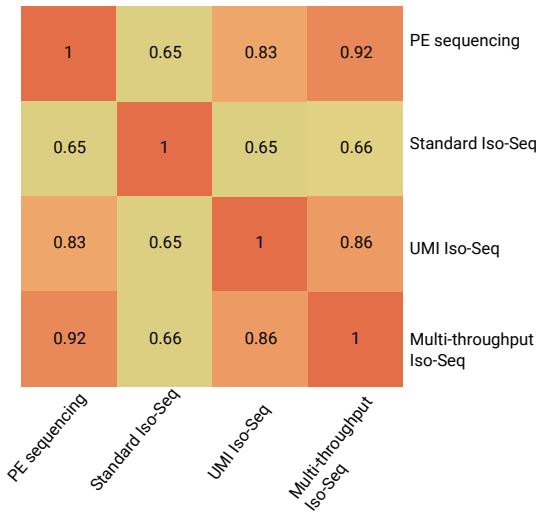
The length distribution of full-length isoforms. The grey and green areas indicate the data from Standard Iso-Seq and Multi-throughput Iso-Seq respectively.

Absolute Quantification with UMI

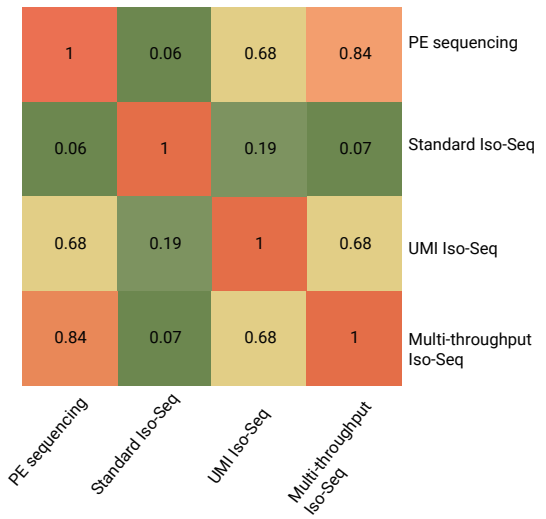
Each original transcript is linked to a unique molecular identifier (UMI), which includes 6-8 random bases. Counting the copy number of transcripts utilizing UMI technology enables accurate isoform quantification without the interference of sequencing duplication.



UMI enables more accurate gene quantification. There will be deviation from the actual Isoform expression based on sequencing reads because of the PCR amplification bias induced during library construction. Multi-throughput Iso-Seq makes able to absolutely quantify genes with UMI technology.



Spearman's correlation coefficient



Pearson 's correlation coefficient

The correlation coefficient of gene quantification between Multi-throughput Iso-Seq and short reads RNA-Seq (PE sequencing) is the highest.

Conclusion

Multi-throughput Iso-Seq was developed by BGI to obtain full-length transcripts. This method takes full advantage of the long sequencing reads generated by the PacBio platform. Compared to the Standard Iso-seq approach, Multi-throughput Iso-Seq can detect more genes/transcripts from the same amount of sequencing data and can quantify genes more accurately.

Request for Information or Quotation

Contact your BGI account representative for the most affordable rates in the industry and to discuss how we can meet your specific project requirements or for expert advice on experiment design, from sample to bioinformatics.

info@bgi.com
www.bgi.com

BGI Offices

BGI Americas

One Broadway, 3rd Floor
Cambridge, MA 02142,
USA

BGI Europe

Ole Maaløes Vej 3,
DK-2200 Copenhagen N,
Denmark

BGI Asia

Building NO.7, BGI Park,
No.21 Hongan 3rd Street,
Yantian District, Shenzhen

BGI Australia

L6, CBCRC, 300 Herston
Road, Herston, Brisbane,
Queensland 4006, Australia

BGI Genomics BGI_Genomics

We Sequence, You Discover