Title
Development of High Throughput Processes for Constructing Illumina Libraries

Authors
Tang, Eric
Hack, Christopher
Ewing, Aren
et al.

Publication Date
2010-03-25

DEVELOPMENT OF HIGH THROUGHPUT PROCESSES FOR CONSTRUCTING ILLUMINA LIBRARIES

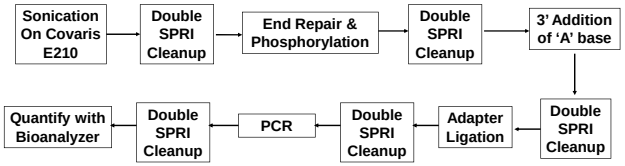
Eric Tang[1], Christopher Hack[1], Aren Ewing[1], Jeff Froula[1], Susan Lucas[2], Shweta Deshpande[1], and Jan-Fang Cheng[1]
[1]Lawrence Berkeley National Laboratory, [2]Lawrence Livermore National Laboratory

Abstract

As the demand of constructing Illumina libraries increases, we have started to modify the library construction protocol to adapt the use of multichannel pipette and 96-well plates. With the few simple modification steps, we have doubled the library production efficiency. These modifications include the shearing of DNA with Covaris E210, and the cleaning of enzymatic reactions and fragment size selection with SPRI beads and a magnetic plate holder. We have also designed a set of molecular barcodes to enable the sequencing of many libraries in parallel. The requirements of these barcodes include 4 bases, balanced GC, and at least 2 bases difference between barcodes. The barcode is attached to the adaptor so it does not require third sequencing primer and the barcoded library can be run on the same flowcell/run with other non-barcoded libraries. We have begun to assess the ability to assign reads and the potential bias towards certain barcodes after pooling different number of libraries.

We have recently programmed the Biomek FX robot to carry out the library construction process. Although this process still require manual transfer of plates from robot to other work stations, the processing of 96 Illumina libraries takes approximately 6 - 8 hours. This semi-automated process represents a significant increase of library capacity comparing to the manual process. We will present the progress and the challenges of these scale-up processes.

Scaling Up Illumina Library Construction

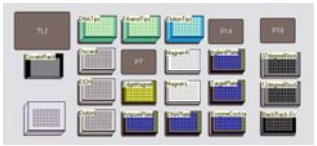


The Illumina library construction process above was constructed to be performed either robotically on the BioMek FX or manually. When constructing a barcoded library a barcoded adapter was used at the adapter ligation step instead of the standard Illumina adapter. Barcoded libraries can be created using the manual or robotic approach as well.

Programming the BioMek Robot

Currently, a single operator is able to construct 24 Illumina libraries in parallel within 3 days. These 3 days take into account the numerous pipetting steps and database work, while providing an allowance of time to adhere to ergonomic limitations. A method to construct 96 Illumina libraries in parallel using the Beckman-Coulter BioMek FX robot is currently in development. This process automates the repetitive pipetting involved in the library construction process and will enable a single operator to construct 96 Illumina libraries in 3 days with minimal ergonomic risk.

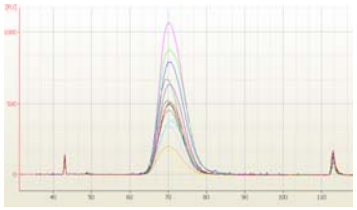
Deck Layout of the BioMek Robot



We have successfully developed a method that has been tested and been used to create libraries. Currently we are working to optimize the efficiency, quality, and reproducibility of the libraries created on the BioMek.

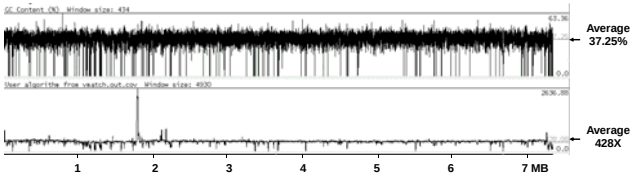
Quality of the Illumina Libraries Created by BioMek

(A) Library Quantification



We assessed the quantity of libraries constructed using the Agilent Bioanalyzer High Sensitivity DNA Chip. We require a sample to have a concentration of at least 10 nM. The graph on the left shows the Bioanalyzer traces of 13 libraries that were constructed by the semi-automated process on the BioMek FX robot.

(B) Genome Coverage



We assessed the sequence coverage of one of the BioMek created Illumina libraries using the Artemis software (Sanger). The above graph shows the GC content plot (top panel) and the sequence depth plot (bottom panel) of the 7.3 Mb cyanobacteria *Pleurocapsa sp. PCC 7319* genome. We were able to generate an average of 428-fold coverage of this BioMek generated library. The high sequence depth region represents a 15kb plasmid whose sequence matches with plasmids of other cyanobacteria. Aside from this peak, the sequence depth shows even coverage of the genome. The low coverage regions correspond to the gaps depicted in the GC content plot.

Pooling of Illumina Libraries with Molecular Barcodes

Rationale: The samples selected for barcoding were derived from different recombinant inbred lines of the monkey flower *Mimulus guttatus*. Illumina sequencing produces millions of reads per run, and barcoding libraries allows pooling of multiple samples to be run in the same lane. This approach has progressively become important as the total number of reads and raw bases per lane increases.

BARCODE SEQUENCE DESIGN REQUIREMENTS
- Oligo length: 4 nucleotides
- 50% GC content
- Adapter sequences differ by at least 2 bases

We have designed 43 unique molecular barcode sequences for Illumina library construction based on the above requirements. We selected 16 of these molecular barcodes to create 16 individually barcoded Illumina libraries from *Mimulus guttatus* samples. A 10nM aliquot of each sample was created according to the Bioanalyzer readings. We then pooled an equal molar ratio of each of the 16 samples into 2 pools (GUBZ and GUCA) containing 8 uniquely barcoded libraries in each.

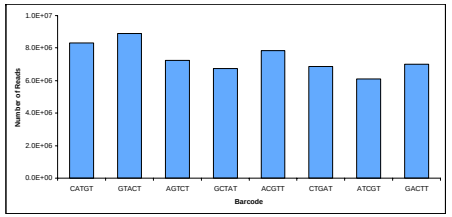
We have analyzed the sequence data using the "number of allowed errors" parameter to asses the ability of assigning these barcodes :

GUBZ - Allowing no error: 96.62% of the reads can be assigned to a barcode, the missing 3.38% could be due to error rate and contamination.

GUCA - Allowing no error: 96.44% of the reads can be assigned to a barcode, the missing 3.56% could be due to error rate and contamination.

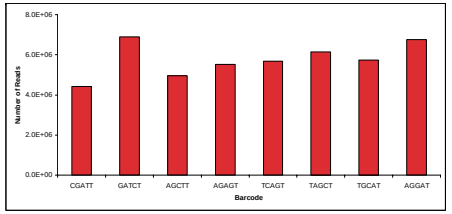
Read Distribution of Barcoded Libraries

Each pool of 8 barcoded libraries was sequenced in a single lane of an Illumina run. For each pooled libraries created, equal quantities 8 barcoded libraries were used. Ideally we would expect about 12% of the total reads to correspond to each barcoded library. The data below shows that each barcode corresponds to about 10%-15% of the total reads for the pooled library.



8 Barcoded and Pooled Libraries
GUBZ

This pool yielded 58,970,748 total reads that corresponded to existing barcodes. If divided equally, we expect each barcode to correspond to about 7,371,344 reads (12%). The actual reads sorted into each barcode ranged from a low of 6,109,125 (10.4%) to a high of 8,874,898 (15%).



8 Barcoded and Pooled Libraries
GUCA

This pool yielded 46,105,087 total reads that corresponded to existing barcodes. If divided equally, we expect each barcode to correspond to about 5,763,136 reads (12%). The actual reads sorted into each barcode ranged from a low of 4,418,407 (10%) to a high of 6,891,596 (15%).

Conclusions

1. We have created a protocol to create 96 Illumina libraries on the Beckman-Coulter BioMek FX robot. The library quantity and genome coverage show no irregularities. We are currently working to optimize the efficiency, quality, and reproducibility of the libraries created on this platform.
2. We have successfully designed 43 unique molecular barcodes that can be used in Illumina library creation. These unique barcodes have been used to create quality Illumina libraries.
3. We have demonstrated success in using our set of molecular barcodes to create and pool libraries for Illumina sequencing. This approach has allowed us to sequence multiple samples in the same run lane. The data presented here indicates that our method produces a relatively even number of reads for each unique barcode used.

I would like to acknowledge the JGI Production Sequencing Group for sequencing the libraries used in this poster.

LLNL-POST-427519

This work was performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Berkeley National Laboratory under contract No. DE-AC02-05CH11231, Lawrence Livermore National Laboratory under Contract No. DE-AC52-07NA27344, and Los Alamos National Laboratory under contract No. DE-AC02-06NA25396.