Meeting the Demand for Long Range Sequencing:

A Scalable Automated Workflow for the 10X Chromium Whole Genome Platform.

MICHELLE CIPICCHIO, COLE WALSH, JOHN WALSH, JUSTIN ABREU, DANIELLE PERRIN, TIM DESMET, SHEILA DODGE, STACEY GABRIEL

BROAD INSTITUTE
GENOMICS

Introduction

Background

Long range sequencing is essential to understand the full variability of a whole genome, enabling phasing and revealing large structural variations.

The Chromium Genome Solution from 10X Genomics enables users to recover long range sequence information from short read sequencing data.

Inherent challenges for this type of library preparation currently limit sample processing to low throughput:

- small batch size
- sensitivity of microfluidics
- precise requirements for DNA input and length.

In collaboration with 10X Genomics, the Genomics Platform at the Broad Institute has addressed these challenges by implementing an automated workflow to prepare Chromium libraries from high quality DNA.

Enabling high throughput

Here we present an automated workflow enabling the preparation of libraries for whole genome phasing. These protocols enable

- Precise automated normalization and quantification, with specific liquid handling to preserve the fragment length of high molecular weight DNA
- Increased overall throughput capacity by combining emulsions from up 12 chips onto a single PCR plate for library construction.
- Sample and process tracking through integration with our LIMS, decreasing the risk of errors

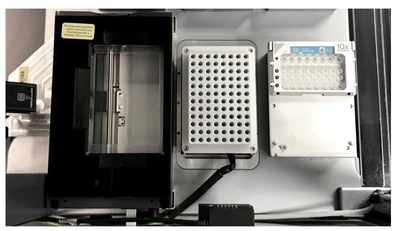


Figure 1. The layout for automated chip loading, from sample 2D barcode tube scan, to a barcoded chilled PCR plate, to 10X chip.

End-to-End Workflow

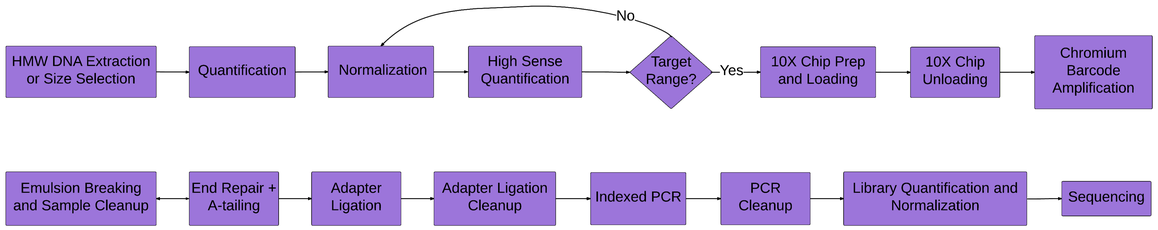


Figure 2. Fully automated end-to-end workflow, from sample preparation, through library construction, to sequencing.

Challenges for Throughput Capacity

Increasing Input DNA Length

Data quality is highly sensitive to the length of the input gDNA molecules.

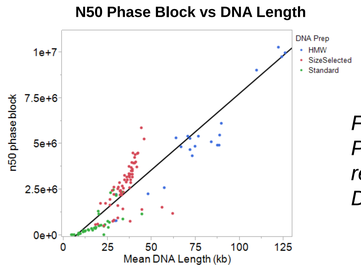


Figure 3. Phasing ability relies on input DNA length

Challenge:
High molecular weight DNA is too fragile to withstand freeze/thaws, vortexing, storage for long periods of time, and pipette-mixing.

Solution:
Implemented HMW extraction, size selection, and new protocols with gentle liquid handling or wide bore tips.

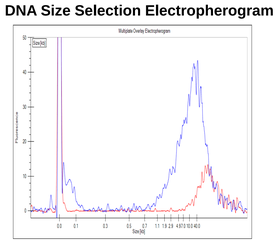


Figure 4. Input gDNA length can be increased using an electrophoresis-based size selection.

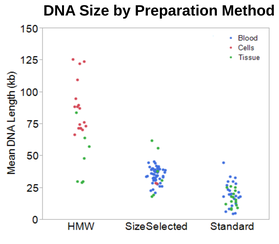


Figure 5. The source and extraction of gDNA significantly affects average length.

Low Input

Challenge:
Target input of 1.2 ng gDNA is below the optimized range of our standard quantification process.

Solution:
Modified quantification assay to increase precision in the low range.

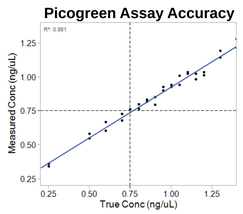


Figure 6. Validation of the accuracy of the target range for the quantification assay.

Chip Capacity

Challenge:
The microfluidic chip has a capacity for 8 samples.

Solution:
Through a collaboration with 10X, implemented

- Custom chip adapters and liquid handler settings to enable automated chip loading and unloading.
- LIMS tracking enabling combination of multiple chips into larger batches without risk of sample swaps.

Microfluidics Sensitivity

Challenge:
The chip is very sensitive to particulates or bubbles from the environment or from plasticware.

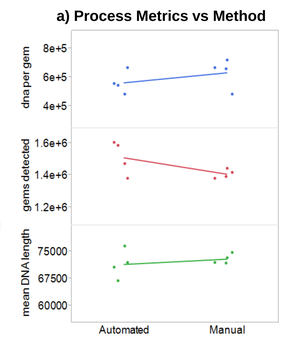
Solution:

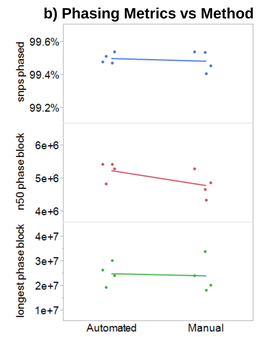
- Evaluated labware for particulates
- Moved process to positive pressure clean room
- Optimized liquid handling for difficult or viscous liquid classes

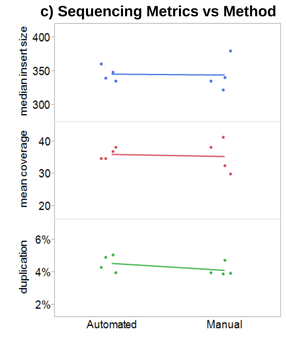
Assay Equivalency Metrics

Validation

Comparison of automation workflow to manually prepared samples using NA12878 in quadruplicate for each condition. Samples were sequenced to ~30X mean coverage on a HiSeqX and analyzed in Long Ranger.







Figures 7a-c. Quality metrics comparison for validation samples. We compared (a)Process metrics for chip loading, (b)Phasing metrics and (c)Sequencing metrics and found no significant difference due to preparation methods.

Conclusion

Altogether, this workflow is a robust, scalable solution to address the increasing demand for long range sequencing, enabling larger cohort studies with an efficient turnaround time.

We have processed over 200 Chromium whole genome samples. This workflow can scale up to 96 samples per week for each Chromium instrument in use. The next challenge will be integrating the data analysis pipeline into the workflow.

Acknowledgments

Data used in this poster was generated at the Broad Institute, for more information please visit: http://genomics.broadinstitute.org/

Data was produced in collaboration with 10X Genomics.