Rapid, low input library preparation using sequential tagmentation

Michelle Cipicchio[1]*, Amirali Kia [2]*, Trina Osothprarop[2], Megan Shand[1], Yossi Farjoun[1], Justin Abreu[1], Maura Costello[1], Danielle Perrin[1] and Molly He[2]
*These authors contributed equally to this work [1]Genomics Platform, The Broad Institute of MIT and Harvard, Cambridge, MA
[2] Protein Engineering Group, Illumina Inc., San Diego, CA

Introduction

Traditional next generation sequencing library preparation is time-consuming, and requires specialized shearing instrumentation as well as several enzymatic reactions and cleanup steps.

Transposase-based methods present well-known advantages:

- Decreased sample input
- Decreased turnaround time
- Increased throughput capacity.

However, transposase-based library construction methods have been known to increase duplication as well as bias in difficult genome regions, which has limited this approach for whole genome libraries.

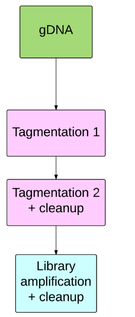
Here we present a novel method for library preparation from human gDNA using sequential tagmentation by two transposases. Initial data shows improved quality over current transposase-based methods:

- More balanced representation across the GC spectrum
- Increased library diversity and decreased duplication

Workflow Detail

Simple overall workflow enables:

- rapid overall processing time
- no requirement for shearing instrumentation
- high throughput using automation for every step



2.5 - 3.5 hours total

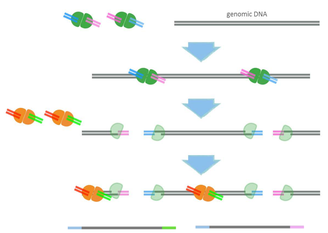


Figure 1. Sequential tagmentation inserts P5 and P7 primers in both steps, maximizing the diversity of fragments for amplification.

Library Construction QC Metrics

Whole genome libraries can be prepared from as little as 25 ng of input DNA

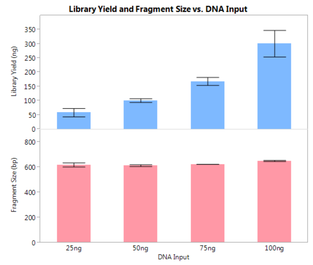


Figure 2. Library yield is dependent on gDNA input, but input amount does not significantly affect library fragment size

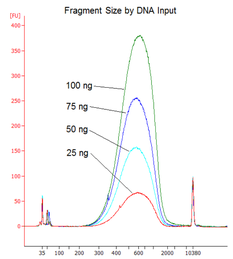


Figure 3. Post-PCR fragment size as measured by Bionanalyzer is independent of DNA input amount

Sequencing QC Metrics

Analysis of 3 sequential tagmentation workflows compared here to traditional Nextera tagmentation as a control method. The 3 workflows differ in concentration and order of transposases added.

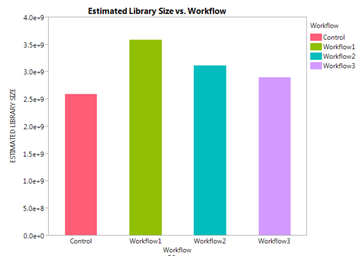


Figure 4. Sequential tagmentation libraries have higher diversity (estimated library size) than control Nextera libraries.

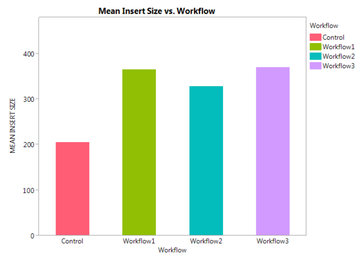


Figure 5. Sequential tagmentation libraries have larger mean insert size than control Nextera libraries.

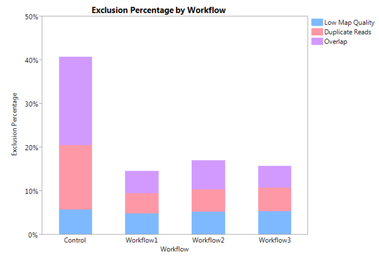
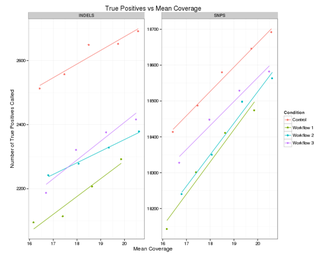


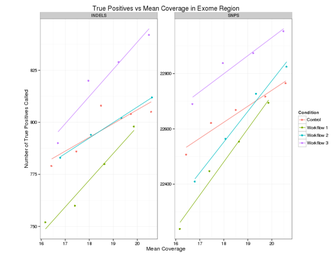
Figure 6. Sequential tagmentation libraries have significantly lower exclusion due to insert overlap and duplication rates

- Low map quality – bases filtered due to mapping quality <20
- Duplicate reads – bases filtered out due to PCR duplication or optical duplication
- Overlap - bases filtered as a second observation from an insert with overlapping reads.

Sensitivity and Accuracy

Precision and heterozygous concordance are similar between these workflows and traditional Nextera, with >99% for SNPs and >98% for Indels on Chromosome 20.





Figures 8 and 9. Sensitivity for SNPs and Indels on Chromosome 20 is higher for the control workflow at ~20x coverage, however in the exome region one of the sequential workflows appears to be more sensitive than the control. Down-sampling was used to examine sensitivity at comparable coverage between all methods.

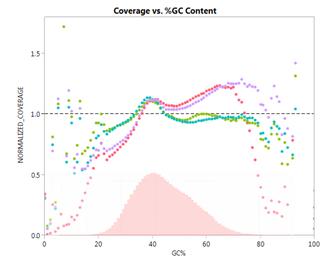


Figure 10. GC bias varies between workflows. Two of the workfows show improved balance of across the GC spectrum as compared to Nextera.

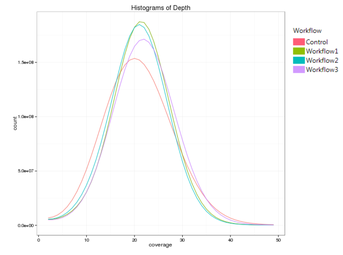


Figure 11. All three workflows show improved evenness of coverage as compared to Nextera, enabling greater sequencing depth.

Conclusion

- The initial data review suggests that sequential tagmentation using two novel tranposases can be used to generate high-quality libraries from human gDNA, enabling benefits such as reduced input and sample preparation time.
- Whole genome analysis of these libraries indicates low duplication and high diversity, with improved evenness of coverage as compared to Nextera.
- While the sensitivity is superior in the exome regions, further development will aim to improve sensitivity overall.