Transposase-mediated sample preparation improvements enable high-throughput variant detection using human whole exome sequencing

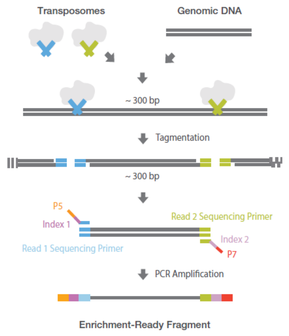
Michelle Cipicchio, Andrew Cheney, Cole Walsh, John Walsh, Jon Thompson, Maura Costello, Justin Abreu, Chris Friedrich, Niall Lennon, Danielle Perrin, Sheila Dodge, and Stacey Gabriel

Broad Institute Genomics Platform, 320 Charles St., Cambridge MA

Introduction

- With over 70,000 samples to prepare for whole exome sequencing in 2015, the Genomics Platform at the Broad Institute has scaled up our process for increased capacity and faster turnaround times by implementing a modified workflow using transposase-mediated library construction.
- Through a series of process optimizations, we have incorporated Illumina's Nextera Rapid Capture Exome into our fully automated workflow for efficient sample preparation that produces diverse libraries, enabling superior exome coverage.

Illumina's Nextera Rapid Capture Exome



How it works:
- Transposase fragments and adds adapter sequences to genomic DNA in a single reaction
- PCR using indexed primers generates fragments ready for selection

Benefits to this method:
- Lower DNA input requirements
- No need for specialized shearing instrumentation
- Reduced process time for library construction

Process Improvements

Quadrupling the throughput achievable by a single user. Enable preparation of four times the number of libraries in less than half the time.

- **Capacity**: 1600 samples/week with ability to scale up to 3200 samples/week
- **Input Normalization:** Increased automated process efficiency
- **Library preparation:** Four process steps replaced with single tagmentation reaction, fully automated and LIMS tracked
- **Cleanup:** Improvements using automated liquid handling for 192 samples simultaneously
- **Indexing**: 96x96 unique barcodes for increased multiplexing scale
- **In process QCs:** Ensures library quality and even representation

Previous Workflow (96 samples)		New Workflow (384 samples)	
Timing	Step	Step	Timing
2 hours	Normalize genomic DNA	Normalize genomic DNA	2.5 hours
5.5 hours	Shearing + cleanup	Tagmentation + cleanup	1 hour
1.5 hours	End repair + cleanup	Library amplification + cleanup	2 hours
1.5 hours	A-tailing + cleanup		
1.5 hours	Adapter ligation + cleanup		
2 hours	Library amplification + cleanup		
Total: 96 samples in 12 hours		**Total: 384 samples in 5.5 hours**	

Quality Improvements

Metric	Manual Nextera	Automated Nextera with Process Improvements
% Selected Bases	66.48%	80.11%
HS Library Size	36,287,668	65,172,128
Penalty 20x	9.09	5.54
Gb Needed for %80 at 20x (Per Samples)	10.14	6.21

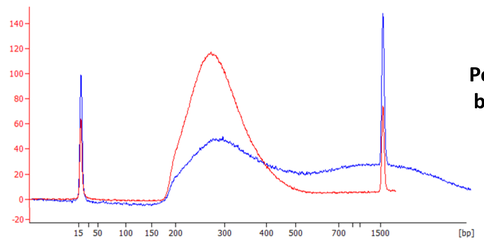
* Comparison of averages for 6 samples per condition

Improved Process Adaptions for Better Sequencing Quality:
- Fully automated process
- Development of 96x96 unique barcoded adapters
- Reaction conditions optimized for PCR and Tagmentation

PCR amplification:

Reaction adjusted to increase efficiency and library quality; indexing allows pooling of 96 samples.

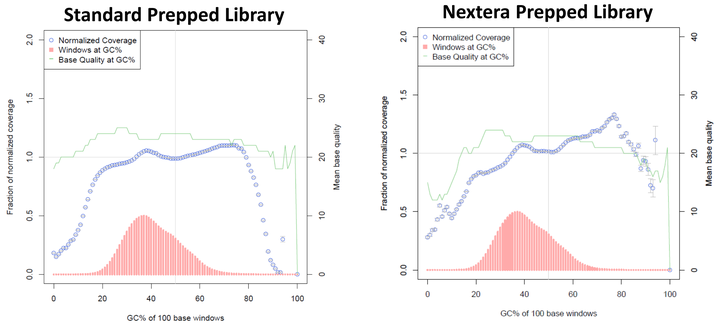
Optimization of PCR allows for a better constructed library utilizing more of the sample input. This yields better sample quantification with increased %Selected Bases and Library Size after Selection.



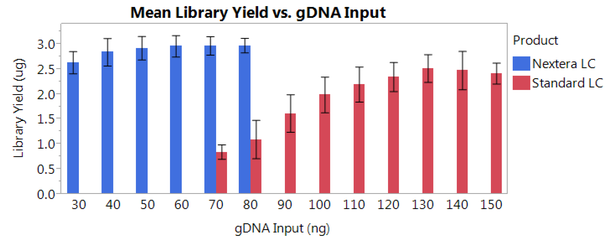
Pond enrichment before and after optimization

Nextera Produces Improved GC Coverage of Samples:

A library created with higher GC regions allows for a more successful coverage of targets after sample selection.



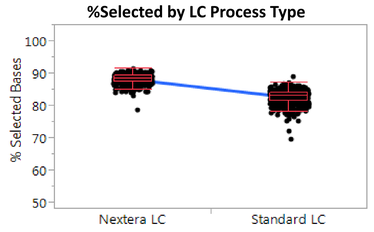
High throughput production



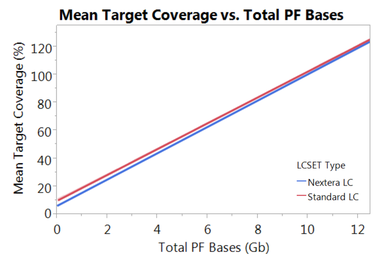
Superior Exome Coverage from 50% Lower Genomic DNA Input

Nextera LC performance highest with 50-70 ng gDNA input, compared to over 100 ng required for Standard LC.

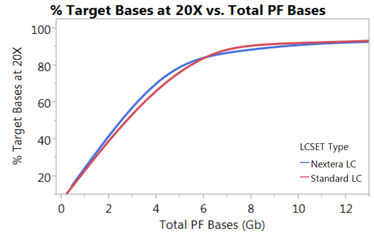
- Typical % selected is high, over 87%
- 5% higher than standard LC
- Little variability between samples



- With less than half the DNA input, exome coverage goals can be met with the same amount of sequencing



- This data compiled from ~1500 Nextera Exomes and ~1500 Standard LC Exomes prepared in full throughput production



Acknowledgements

National Human Genome Research Institute

illumina®