BROAD INSTITUTE

GENOMICS

Development of a fully automated high throughput process for long read sequencing library preparation

Evan McDaid[1], Michelle Cipicchio[1], Ally Day, Corey Nolet, Roberto Luis-Fuentes, John Walsh, Scott Anderson,
Brendan Blumenstiel, Conrad Lavoie, Baiyu Zhou, Cole Walsh, Catie Ramnarine, Erin LaRoche, Michael DaSilva, Katie Larkin,
Jon Thompson, Mariela Mihaleva, Atanas Mihalev, Tom Howd, Niall Lennon, Stacy Gabriel
Broad Institute of MIT and Harvard, 320 Charles St Cambridge MA 02141

Introduction

Background

Long read sequencing enables a deeper understanding of the most complex and challenging regions in the human genome, however due to the complexity and cost for these technologies they have not thus far been implemented at high scale.

The All of Us Research Program aims to build a diverse database of health and genetic data to further our understanding of the genetic basis of common disease. In order to enable a more comprehensive database that includes SV's, it is critical to enable long read sequencing at scale. Figure 1 shows that increased read length yields better structural variant detection.

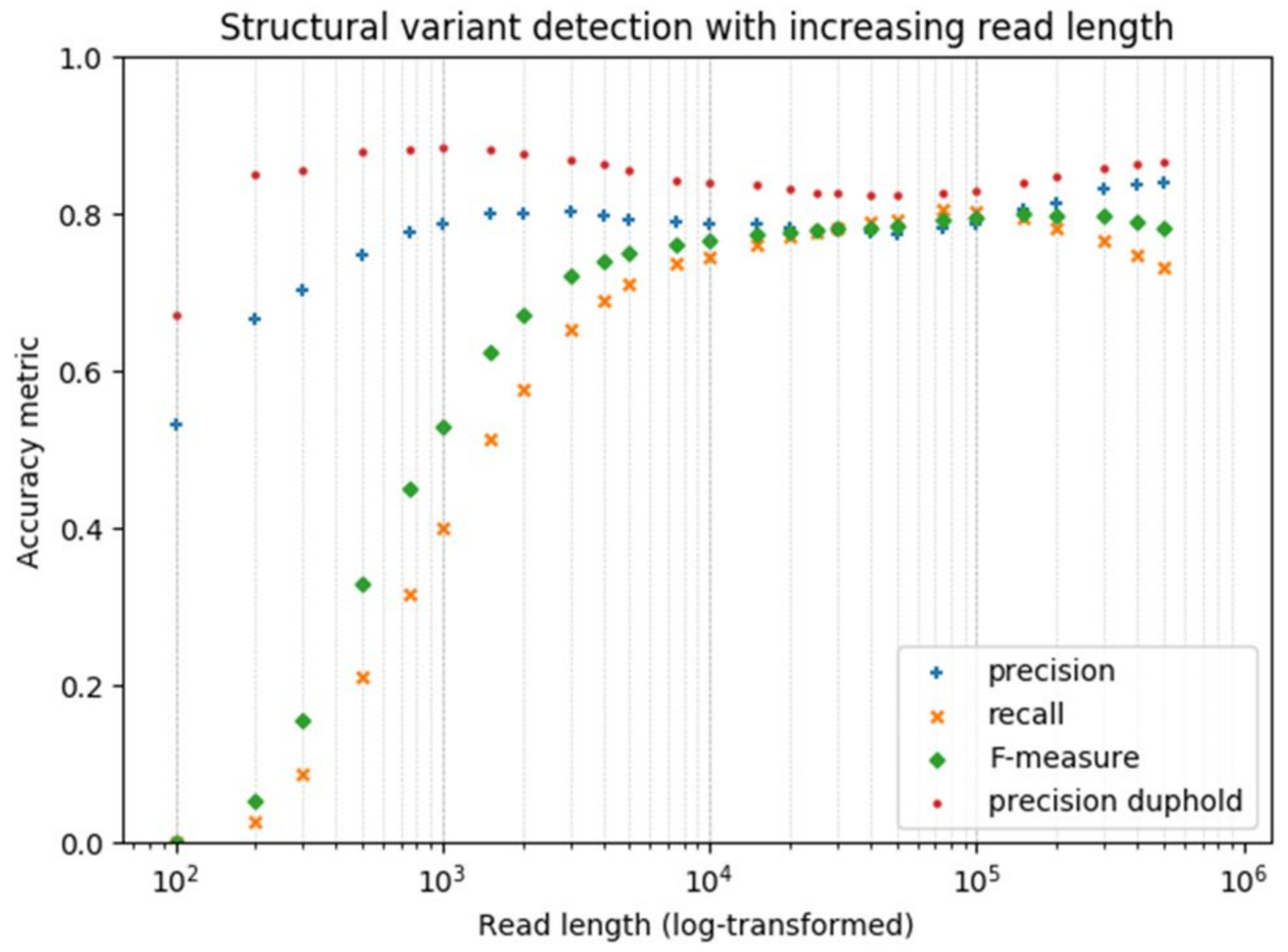


Figure 1. The *F*-measure indicates that optimal performance is reached approximately from reads of 15 kb and longer. Figure credit: De Coster, et al. *NAR Genom Bioinform*, Volume 2, Issue 1, March 2020

High Throughput Automation

Until recently these library preparation methods have not been suitable for high throughput scale.

- Long fragments of DNA are highly susceptible to manual shearing.
- Complex sample preparation methods were not readily automatable.

A new version of HiFi long reads sequencing from Pacific Biosciences has provided a solution to the challenges described here.

HiFi sample preparation is amenable to automated liquid handling, and yields highly accurate reads >10kb for SV calling.

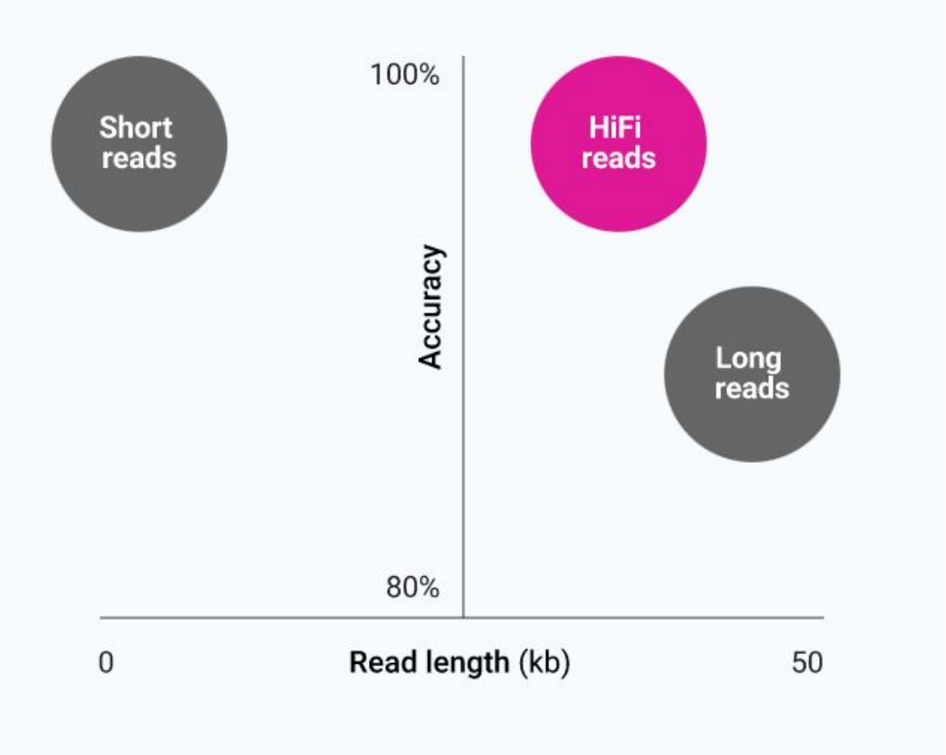
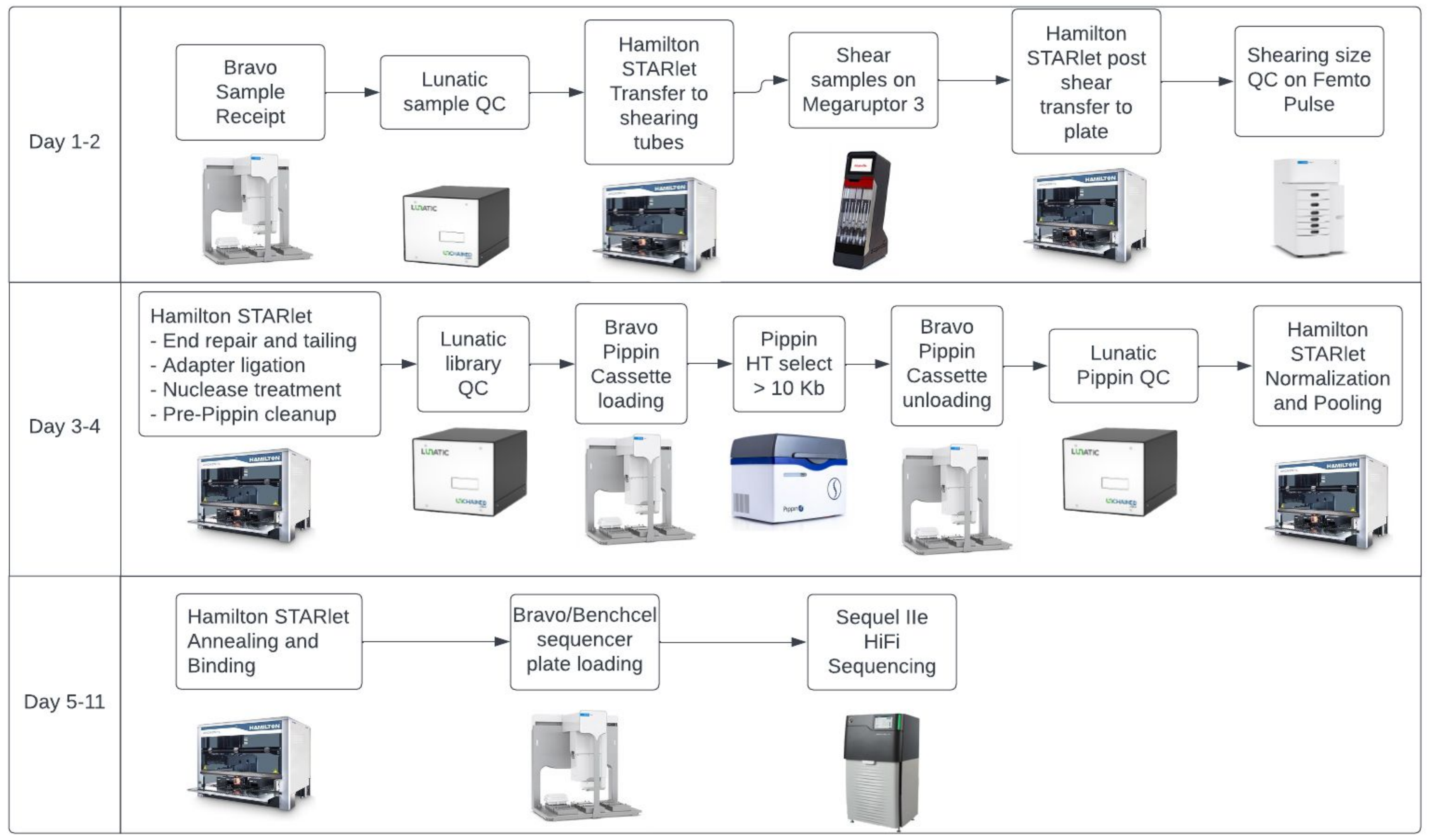


Figure 2. PacBio HiFi sequencing bridges the gap between long reads and accuracy

Automated Workflow Design


Day 1-2
Bravo Sample Receipt
Lunatic sample QC
Hamilton STARlet Transfer to shearing tubes
Shear samples on Megaruptor 3
Hamilton STARlet post shear transfer to plate
Shearing size QC on Femto Pulse
Day 3-4
Hamilton STARlet
- End repair and tailing
- Adapter ligation
- Nuclease treatment
- Pre-Pippin cleanup
Lunatic library QC
Bravo Pippin Cassette loading
Pippin HT select > 10 Kb
Bravo Pippin Cassette unloading
Lunatic Pippin QC
Hamilton STARlet Normalization and Pooling
Day 5-11
Hamilton STARlet Annealing and Binding
Bravo/Benchcel sequencer plate loading
Sequel IIe HiFi Sequencing

Data Quality: Manual vs Automated

We validated the automated workflow on a pilot set of samples and compared the quality metrics to confirm equivalent sample performance.

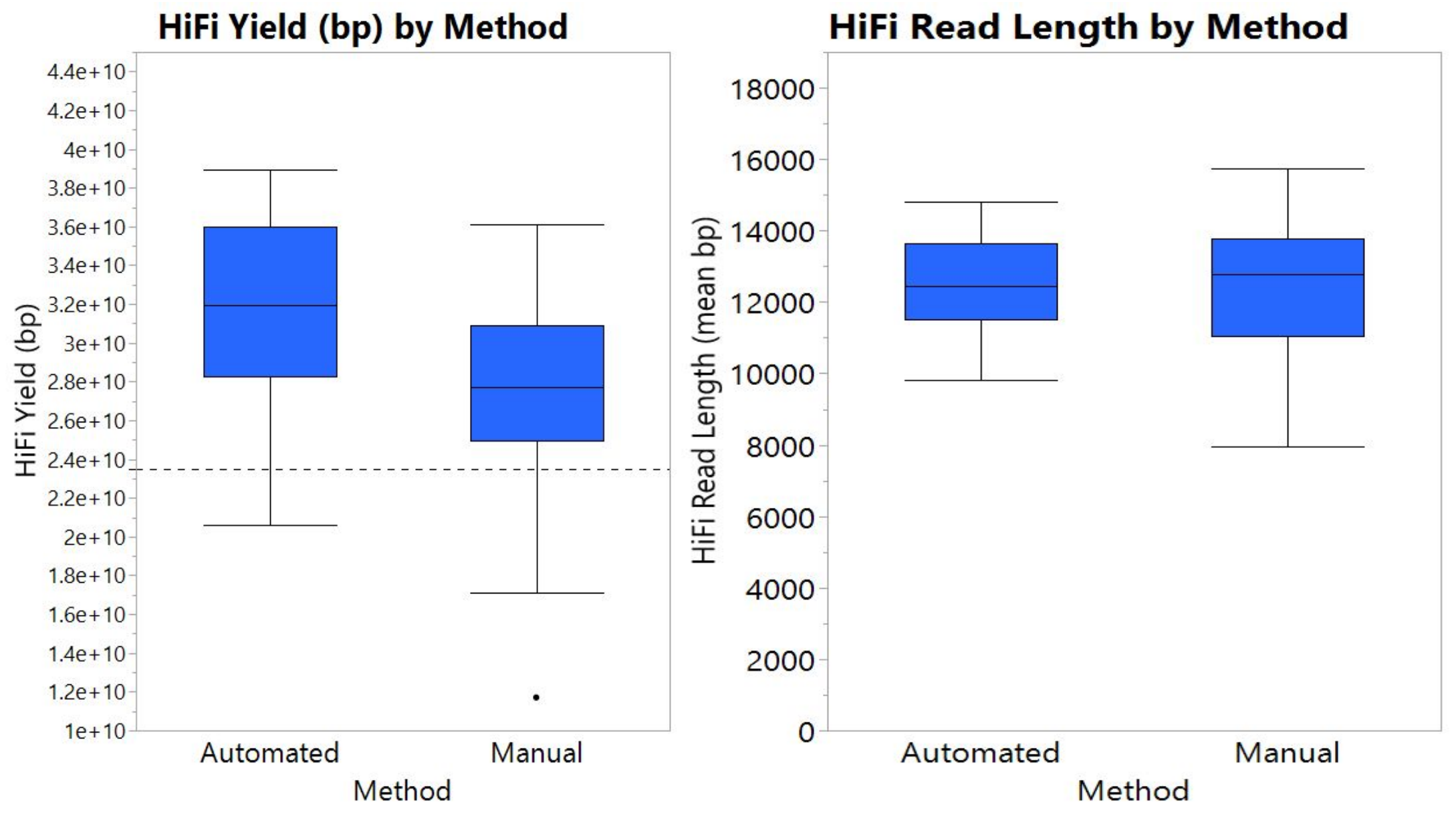


Figure 3. HiFi Yield and Read Length for samples prepared using the automated workflow performed similarly or better than samples prepared manually on the same sample preparation method.

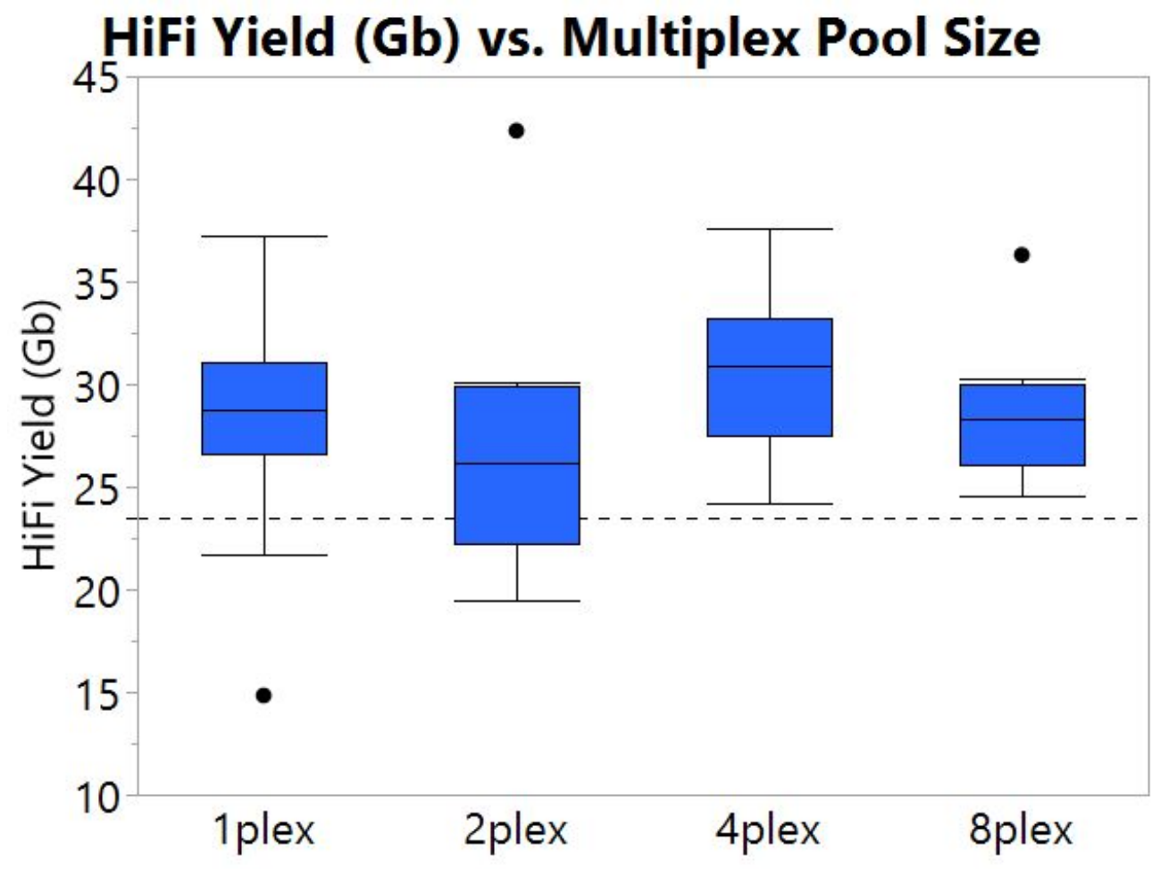


Figure 4. Variability in yield is reduced by pooling libraries and sequencing across multiple SMRT cells. Multiplexing reduces the number of samples that do not meet deliverable due to run-specific issues.

Size Selection: Bead-based vs Pippin HT

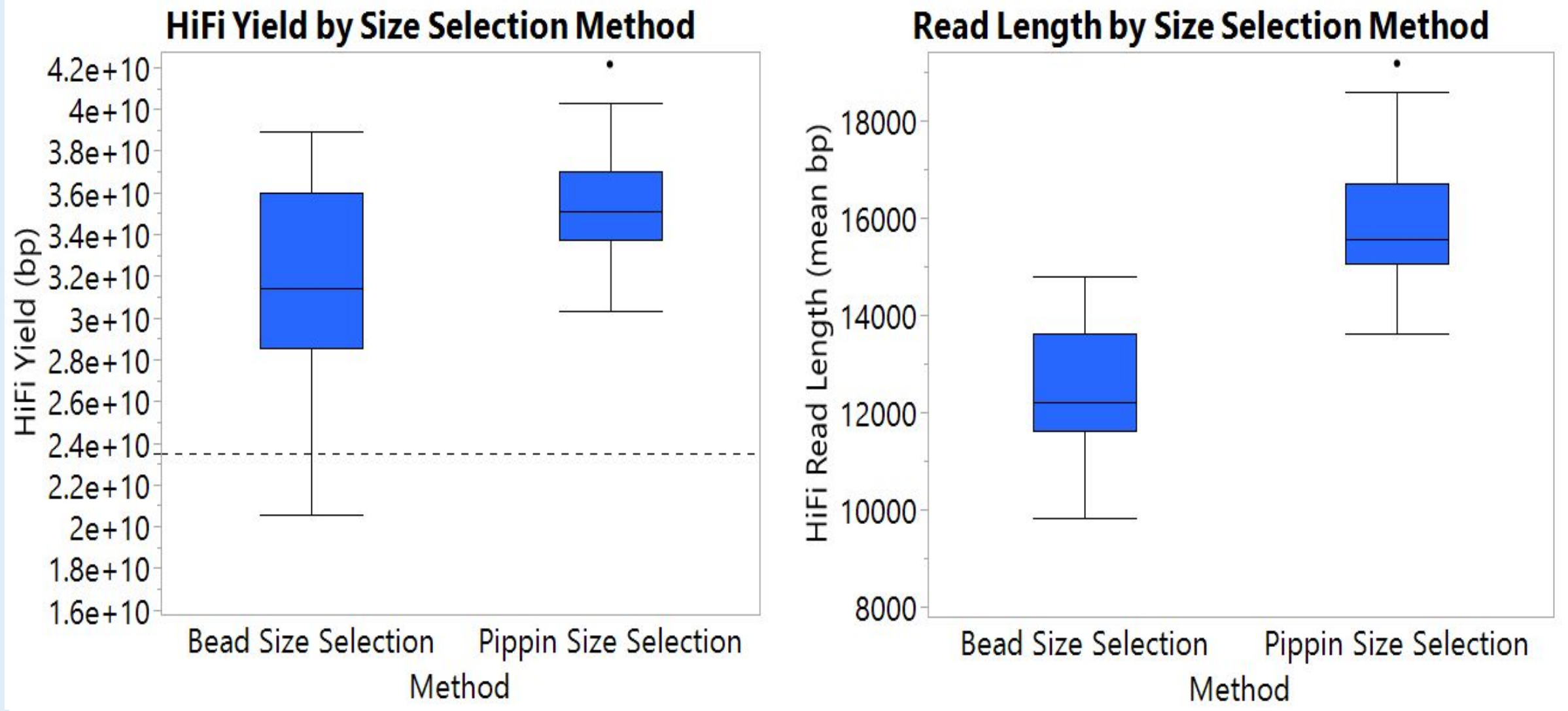


Figure 5. HiFi Yield and Read Length for libraries after size selection using the PippinHT is significantly higher than after size selected using beads.

Bead-based Size selection

- Ideal for high throughput applications
- Can be performed on deck
- Does not require additional instrumentation
- Samples remain in SBS format
- Inaccurate/Unreliable size selection

Pippin HT size selection

- Not optimized for high throughput automation
- Additional equipment\consumable purchase
- More time consuming
- Precise size selection
- Consistent size selection results

Equipment and Scale

	Set 1	Set 2	Set 3	Set 4
Week 1	Sample QC Shearing Shearing QC Library Prep Multiplex Pooling	Sample QC Shearing		
Week 2	Seq Prep Sequencing	Shearing QC Library Prep Multiplex Pooling Seq Prep Sequencing	Sample QC Shearing Shearing QC	
Week 3			Library Prep Multiplex Pooling Seq Prep Sequencing	Sample QC Shearing Shearing QC Library Prep Multiplex Pooling
Week 4				Seq Prep Sequencing

Figure 6. Monthly laboratory schedule. Sample batches are processed concurrently to reduce instrument downtime and meet sequencer capacity

- Current automation equipment enables processing 500 samples a month.
- Instrument redundancy removes any gaps or processing downtime for instrument repairs, maintenance ect.
- Automation equipment:
  - 2 Hamilton STARlet liquid handlers
  - 4 Diagenode Megaruptor 3
  - 2 Agilent Bravo's
  - 1 Agilent BenchCel
  - 2 Pippin HT
  - 1 Agilent Femto Pulse
  - 1 Unchained Labs Lunatic.

Conclusions

The inclusion of a fully automated process has allowed us to increase our scale 3.5 fold, from 39 cells a week to 140, while ensuring high quality results. We show that switching to an automated process does not hurt metrics in terms of turnaround time, reproducibility, and data quality. Additionally, we gained the ability to easily track and process multiplexed samples at scale reducing sequencing costs.

A fully automated workflow will allow researchers across the globe access to population scale long read studies, such as the All of Us Research Program. Continuous chain of custody, reagent tracking, and advances in the analytical pipeline have enabled a pathway for large scale clinical long read genomes and the potential to unlock previously unknown causes of human disease.

Multiplexing enables the flexibility to increase or decrease pool size to adjust for the amount of data we want to generate.

Acknowledgments

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