A re-designed and unified workflow for exome and custom targeted panels to support a variety of projects at scale

MICHELLE CIPICCHIO[1], GINA VICENTE[1], MATTHEW DEFELICE[1], SOPHIE LOW[1], MARK FLEHARTY[1], JUSTIN ABREU[1], BRENDAN BLUMENSTIEL[1], MAURA COSTELLO[1], CARRIE CIBULSKIS[1],TIM DESMET[1], STACEY GABRIEL[1]

1 Broad Genomics, Broad Institute of MIT and Harvard, 320 Charles St. Cambridge, MA 02142

Introduction

The Broad Institute Genomics Platform has developed a complete end-to-end redesign of our targeted sequencing products, enabling

- lower DNA input
- faster turnaround time
- lower cost
- improved sequencing quality

Our goal was to provide the flexibility to support a wide variety of projects, while streamlining the workflow to enable reduced cost and turnaround times.

Flexible and Customizable

- Input DNA – this workflow accommodates a variety of sample types, including cell-free, FFPE, as well as low and high quality genomic DNA.
- Unified selection – our fully automated protocol utilizes IDT's xGen Hybridization and Wash Kit, with a standardized protocol independent of target size or library input.
- Custom targets – we can design and order baits for projects of any size. Our selection process is validated to work equally well with panel sizes ranging from 50 kb to 35 Mb
- Analysis - libraries can be UMI-enabled for optional analysis of low allele fraction variants using deep sequencing

Features of Standard Germline Exome

- High throughput scale with rapid turnaround time. We currently process 2000 samples/week.
- Large insert library to maximize coverage while minimizing sequencing costs (385 bp shear)
- Twist Human Core Exome design plus ~2 Mb of additional custom content

Features of Somatic Exome and Custom Targeted Panels

- Short insert to maximize yield from low quality sample types such as FFPE DNA (150 bp shear)
- Optimized for ultra low input (down to 5 ng) to maximize yield from limited samples such as cell free DNA

A Universal Streamlined Workflow to Accommodate a Variety of Custom Projects

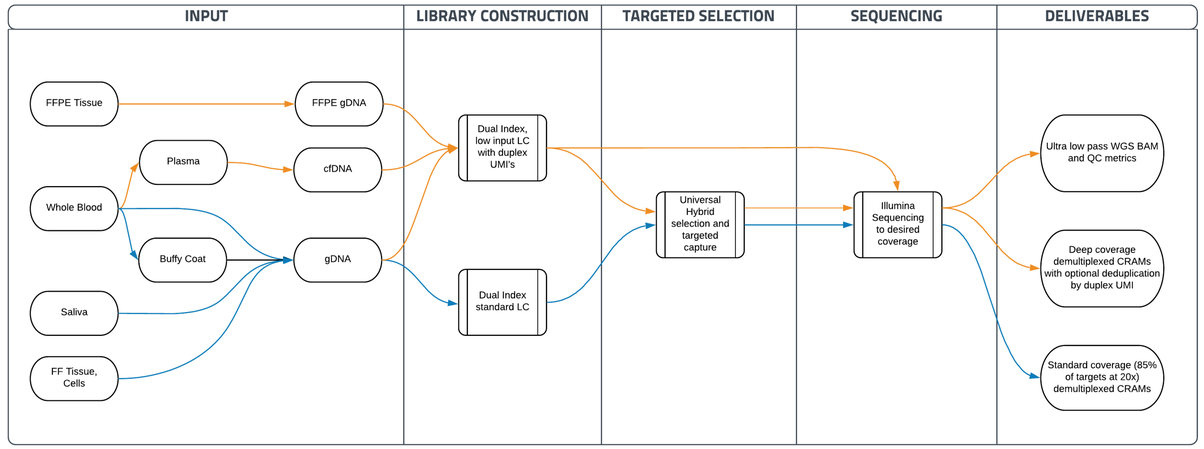

Figure 1. *Our innovative ultra high throughput automated selection workflow far exceeds our current library construction capacity for any given product, enabling us to combine libraries from different products into one universal selection protocol. (1900/week germline exomes, 480/week somatic exomes or targeted products, 480/week cell free DNA exomes or targeted products.) The same library with UMIs can be selected multiple times, if after exome sequencing there is interest in deeper coverage of a small panel of genes.*

Library Construction

Our workflow is agnostic to library construction method and enables batching of different library types together for streamlined processing.

A single library preparation enables:

- ultra low pass whole genome sequencing for tumor purity QC
- followed by somatic exome sequencing
- and/or deeper targeted sequencing for genes of interest.

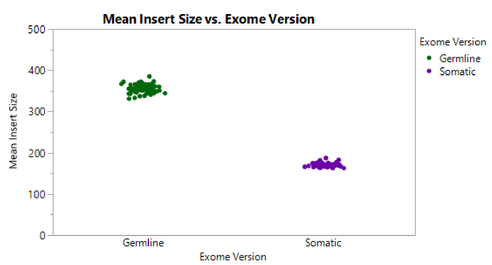

Figure 2*. The standard germline exome version uses a 385 bp shear fto maximize coverage while minimizing sequencing costs . The somatic exome uses a 150 bp shear to optimize for lower quality input DNA, and incorporates UMIs to empower deeper sequencing.*

Content and Coverage

Our new exome product features more complete exome coverage, with content curated from users for both population and somatic studies.

Design details:

- Twist Human Core Exome design (~35 Mb) plus ~2 Mb of additional custom content
- Population genetics:
  - Full list of ACMG59 genes
  - Additional RefSeq and OMIM putative gene sequences
- Cancer genetics:
  - COSMIC (Catalogue of Somatic Mutations in Cancer) variants
  - Broad ONCO panel genes
  - key promoters and other motifs that have been identified as potential cancer hot spots
- Complete mitochondrial genome

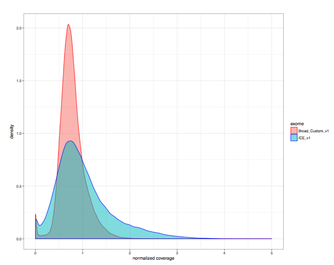

Figure 3*. The normalized coverage of exome content in our new panel (red) is significantly more even than our previous exome product (blue).*

Quality Metrics

Our new exome surpasses our prior version in virtually every measurement of quality. Higher % Selected, lower Penalty 20X, and improved evenness of representation across the targets enable samples to reach higher coverage with fewer bases sequenced. Our deliverable for a standard exome (85% of bases at 20X coverage) can be reached with 65% less sequencing than previously required, allowing for dramatic cost savings to our users.

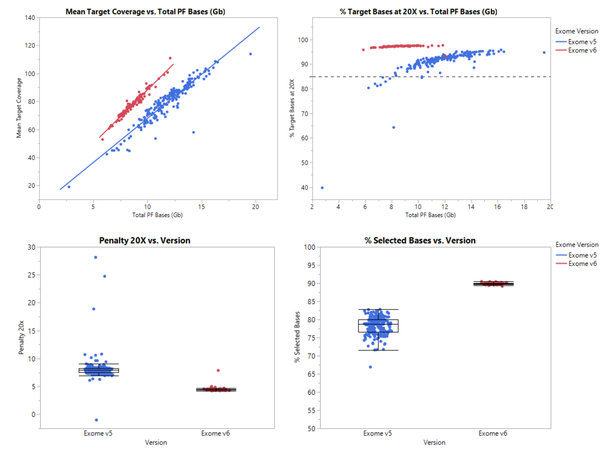

Figure 4*. Our new v6 exome (red) surpasses our prior v5 exome (blue) in virtually every measurement of quality.*

Conclusion

Beginning in January 2019, we implemented a new exome version as part of a complete redesign of our targeted sequencing workflow. This new streamlined protocol has enabled us to provide higher quality exome data to users at a reduced cost, while simplifying our internal workflow for high throughput scale.

The next steps for optimization will focus on increasing the throughput capacity for our library construction processes to match the scale of our targeted selection workflow.

Acknowledgments

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