Development of Exome 2.0, the best-in-class exome and improved target enrichment system

Elaine Maggi, Steve Oh, Michael Bocek, Leonardo Arbiza, Rebecca Nugent, Venkatesh Moktali, Adriana Arneson, Derek Murphy, Esteban Toro, and Siyuan Chen

1. Abstract

Target enrichment offers distinct advantages over whole genome sequencing. Specifically, the complexity and costs of sample analysis are reduced through focused sequencing of select regions of interest. Leveraging target enrichment, whole exome sequencing has become a powerful and well established tool used to interrogate the coding regions of the genome. There are several target enrichment systems and exome capture panels available on the market, including options from Twist Bioscience. Here we present data from our most up-to-date whole exome capture solution. Featuring an improved hybridization system and optimized capture panel design, Exome 2.0 enables best-in-class performance with options for same day or overnight target enrichment workflows.

Our newly designed panel has been expanded to include an exhaustive list of regions of clinical significance. This additional content includes ClinVar coverage of over 99% representing an appreciable improvement over previous exomes. The probe design has also been rebalanced and specifically tailored to produce increased uniformity of coverage on the Illumina Novaseq sequencing platform. In addition to these improvements, the updated overnight hybridization system reduces undesirable off-bait by up to 50% compared to previous offerings. This increased uniformity of coverage combined with an improved on-target rate results in lower overall sequencing needs and costs per sample.

Further improvements include the addition of a bead-based alternative to the speedvac drydown step for both the overnight and single-day hybridization workflows. The configuration of the hybrid-capture system also allows for customization of targeted regions and the ability to multiplex up to 16 samples in a single capture reaction.

We demonstrate that Exome 2.0 has improved performance in several key areas including uniformity of coverage and percent on bait. These results combined with our flexible workflow makes our system an ideal option for whole exome sequencing.

2. Panel Content

Desired content was selected based on several different criteria. Final panel design blended desired content with performance by removing some poor preforming probes in regions without significant clinical value.

Coding Content

- 36.5 Mb total targets designed against hg38
- Coding sequences in all major gene databases
- These also fully cover UCSC and Refgene

Major Gene Databases

Annotations Database	Version	Target territory (Mb)
Ensembl	v101, time stamp 7/11/20, 8:02:00 PM	35.67
RefSeq	v109, time stamp 8/19/20, 9:10:00 AM	35.77 Mb
Gencode	v35, time stamp 08/4/20 9:45 AM	35.79 Mb
CCDS	v22, 2018-06-14 12:08	33.13Mb
Refgene	Nov 2019	35.85 Mb
Gencode Basic	Subset of gencode above	32.59 Mb
UCSC	Has fully switched to gencode & used gencode basic	--

Additional Content

- Cohesive retention of targets in important genes
- All coding targets for:
  - ✓ Genes with 500+ PubMed citations
  - ✓ COSMIC pathogenic variants
  - ✓ ACMG73 genes
  - ✓ Key genes identified by customers
  - ✓ Genes with 50+ tests in the Genetic Testing Registry
- Recovery of the vast majority of filtered regions for:
  - ✓ Genes with 150 or more citations
  - ✓ Genes with 7 or more tests in the Genetic Testing Registry

Noncoding and Other Content

- ClinVar:
  - ✓ Pathogenic variants
  - ✓ Likely pathogenic
  - ✓ Other variants near genes of interest/passing filters that cover UTRs, Introns, and other areas of interest
- Targets of interest from the field
- Pharmacogenomics variants from PharmGKB, PharmVar, CPIC
- Tracking SNPs within targets including two popular sets:
  - ✓ Pengelly *et al.* SNPs
  - ✓ EuroGenetest SNPs
- The TERT promoter and enhancer regions (Figure 1)

Figure 1: IGV Genome Browser view of the TERT promoter region showing target regions in the Twist Exome 2.0 and competitor exome panels.

Comprehensive Spike-in Panel available to include all regions covered in previous Twist exome

3. Protocol Changes

Standard Hybridization v2			Fast Hybridization	
DNA	50 ng	Library Prep	DNA	50 ng
DNA	500 ng		DNA	500 ng
DNA Drydown	Speedvac or Bead Based		DNA Drydown	Speedvac or Bead Based
Hybridization	70°C, 16 hr ST Hyb Buffer	Hybridization	Hybridization	60°C, 2-4 hr FH Hyb Buffer
Streptavidin	68°C, 5 min		Streptavidin	RT, 30 min
Wash 1 (x1)	68°C, 5 min Std Wash 1 Buffer		Wash 1 (x2)	70°C, 5 min Fast Wash Buffer 1
Wash 2 (x3)	48°C, 5 min Wash Buffer 2	Capture and Wash	Wash 2 (x3)	48°C, 5 min Wash Buffer 2
PCR & Clean up			PCR & Clean up	

Protocol changes relative to current published methods are noted in green text.

Financial Disclosures: All authors are current or previous employees and shareholders of Twist Bioscience

4. Bake-off Performance of Exome 2.0 with Standard Hyb v2

Bake-off Study Design

- Two third party laboratories were hired to conduct an exome comparison between Twist Exome 2.0, Competitor I, Competitor K, and Competitor A, to present an unbiased dataset
- Vendor-specific protocols and library preparation were all performed in duplicate

Assay Parameters

- Sample used was NA12878
- 50ng input to library prep, except vendor K which requires 100ng
- Twist Exome 2.0 was run with Enzymatic fragmentation Kit 2.0
- NovaSeq S2 Flow Cell, PE100, 6Gb/sample

Uniformity (Fold 80 Base Penalty)

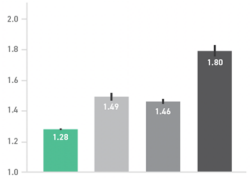


Figure 2: Fold 80 base penalty is a measure of the uniformity of the capture across all targets. Lower fold 80 indicates less sequencing is needed to achieve the required coverage depth for all target regions.

Targets

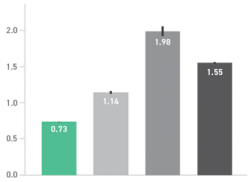


Figure 3: Percent Zero coverage targets highlights the number of targets with no reads assigned. Lower values indicate fewer target dropouts.

Pct On Target

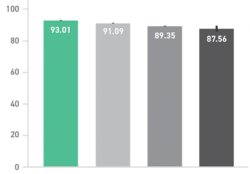


Figure 4: Percent on target is defined as 1-PCT_OFF_BAIT. Higher on target means less wasted sequencing.

Pct Target Bases with 30X Coverage

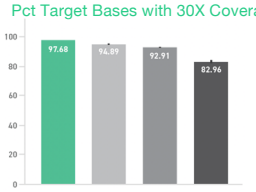
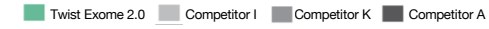


Figure 5: Percent target bases with a minimum of 30X coverage is influenced by both panel uniformity and on target. Higher values mean more of the target regions have at sufficient coverage for downstream analysis.



Normalized Coverage by GC Content

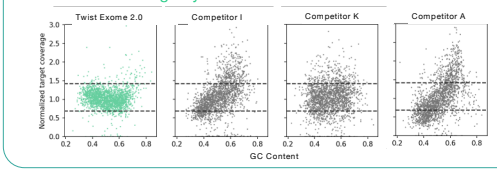


Figure 6: To assess the effects of coverage bias by target GC content, normalized target cover was plotted against the percent GC in each of the target regions. A flatter and narrower distribution signifies less GC bias in capture efficiency and more uniform coverage between target regions.

5. Comparison of Results with Different Configurations

Study Design

Captures were done in house using different configurations of hybridization method, drydown method, and sequencer. These provide flexibility for different customer needs. Data for comprehensive spike-in is also shown.

Uniformity (Fold 80 Base Penalty)

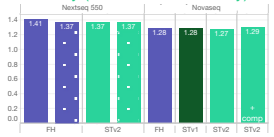


Figure 7: Fold 80 is slightly higher on the Nextseq 550 instrument due to intentional optimization for the Novaseq system. There is no difference between hyb methods.

Pct Zero Coverage Targets

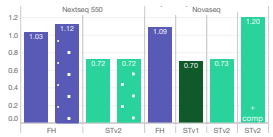


Figure 8: Fast hyb has higher zero coverage targets than Standard hyb system, however, it has fewer dropouts than competitor systems.

Pct On Target

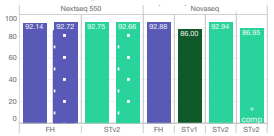


Figure 9: STv2 improves on target relative to the initial Standard hyb system. Fast hyb and STv2 have similar percent on target bases.

Pct Target Bases with 30X Coverage

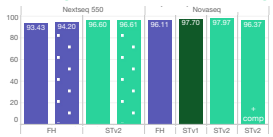


Figure 10: Coverage is very similar for all hyb types. There is slightly lower coverage with fast hyb sequenced on the Nextseq 550.

6. Conclusions

- Exome 2.0 has been designed to be balanced for the Novaseq 600 system and to contain the most clinically relevant content.
- The updated standard hyb system provides decreased off bait without affecting dropouts or uniformity.
- There are no differences between beads based drydown and speedvac methods.