Optimized Workflows for Custom Targeted Enrichment, From Oligo Probe Design to NGS Data Analysis

Susanne Angelow[1], Jonathan Nery[1], Yoon Hye Shin[1], Shiyu Chen[1], Yonmee Han[1], Nidhi Shah[1], Priyanka Prashar[1], Tayler Mazurski[1], Deresa Lee[1], Seung Hyeon Lee[1], Joel Geoghegan[1], Scott Kuersten[1], Beth Frey[1]

[1]Illumina, Inc. 5200 Illumina Way, San Diego, CA 92122

Abstract

Targeted enrichment allows sequencing of a subset of genes or genomic regions of interest by hybridization to sequence-specific biotinylated DNA probes and isolation with streptavidin linked magnetic beads. This method allows researchers the flexibility to sequence a large range of targets, from whole exomes to just several genes or variant hotspots. Successful target enrichment relies on a high-quality probe pool and library prep workflow to provide high target coverage uniformity and to minimize off-target reads and dropouts. Here we describe the end-to-end workflow from panel design in DesignStudio, to secondary analysis in BaseSpace Sequencing Hub for custom DNA and RNA targeted enrichment panels. Multiple DNA Illumina Custom Enrichment Panels (ICEP) v2 were designed using Illumina DesignStudio, an online tool for custom assay design, to a variety of genomic content resulting in a range of panel sizes. DesignStudio uses proprietary design algorithms to design and select 120bp, double-stranded probes that provide high on-target enrichment and uniformity. DNA panels were designed as either standalone panels or as a spike-in to the Twist Bioscience for Illumina Exome 2.5 Panel. Spike-in panels can be used to enhance performance of existing regions or supplement additional content into the panel. RNA panels were designed by Illumina Concierge, which used an exon informed design strategy for transcripts that allows for detection of novel isoforms or fusion transcripts. In addition, we demonstrate the compatibility of ICEPv2 across multiple sample types from circulating tumor (ct) and formalin fixed paraffin embedded (FFPE) DNA with Illumina cfDNA DNA Prep with Enrichment and RNA samples including viral pathogens with Illumina RNA Prep with Enrichment.

End-to-End Solution from Design Through Analysis

A.

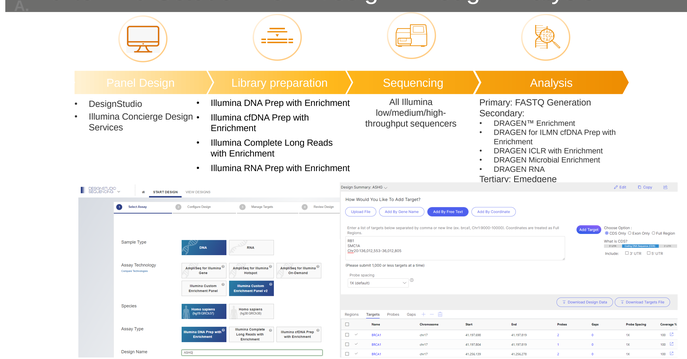


Figure 1. End to end workflow from probe design through data analysis. Researchers can submit targets of interest in DesignStudio or work with Illumina Concierge to create ICEPv2 panels compatible across Illumina's targeted enrichment NGS workflows, sequencing platforms and analysis apps in BaseSpace Sequencing Hub.

Custom Supplemental Panel to Boost Coverage in WES

Figure 2.ICEPv2 can be used as standalone or to supplement custom or exome panels to deepen coverage of existing target regions or add new content. Shown here are representative IGV data for CSF1R and BCORL1 generated from libraries enriched with the Twist Bioscience for Illumina Exome 2.5 Panel with and without a supplemental custom panel targeting genes associated with Clonal hematopoiesis of indeterminate potential (CHIP). Mean target coverage (MTC) of the covered regions improved by ~2-fold with the addition of the supplemental panel. Samples: Coriell NA24143, NA24149, and NA24385. Workflow: Illumina DNA Prep with Exome 2.5 Enrichment. Sequencing NovaSeq 6000 downsampled to 50M reads. Analysis: DRAGEN Enrichment.

DNA Applications

Table 1. Custom Sequencing Panel specifications used in Illumina DNA Prep with Enrichment workflow

	Panel A*	Panel B	Panel C	Panel D
No. of target genes	66	242	4,811	6,704
No. of unique probes	2,081	7,877	99,432	137,503
Cumulative target region size	191kb	732kb	11,925kb	16,420kb

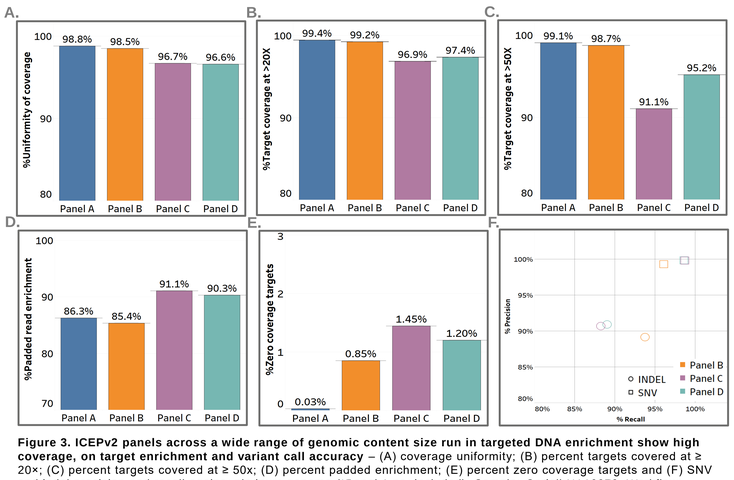


Figure 3. ICEPv2 panels across a wide range of genomic content size run in targeted DNA enrichment show high coverage, on target enrichment and variant call accuracy – (A) coverage uniformity; (B) percent targets covered at ≥ 20×; (C) percent targets covered at ≥ 50x; (D) percent padded enrichment; (E) percent zero coverage targets and (F) SNV and indel precision and recall against platinum genome (*Panel A not included). Sample: Coriell NA12878. Workflow: Illumina DNA Prep with Enrichment. Sequencing NovaSeq 6000 downsampled to 300-400x MTC. Analysis: DRAGEN Enrichment.

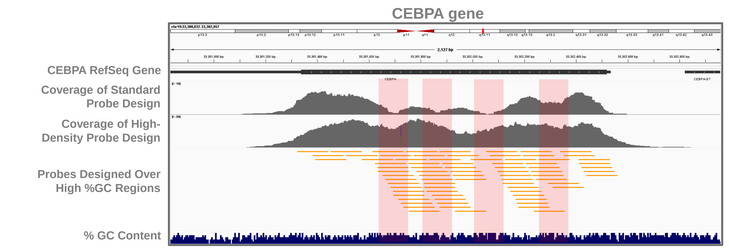


Figure 4. Increasing probe density improves coverage across the GC-rich regions of CEBPA, exon 1 – IGV screenshot showing CEBPA exon 1 region coverage using standard and high probe density panels. The high probe density design results in improved coverage of the GC-rich regions of exon 1 (highlighted in red), allowing more complete capture of this critical region. Enhanced coverage is essential for accurate detection of CEBPA mutations linked to acute myeloid leukemia (AML). Sample: Coriell NA12878. Workflow: Illumina DNA Prep with Enrichment. Sequencing: NovaSeq 6000. Analysis: DRAGEN Enrichment.

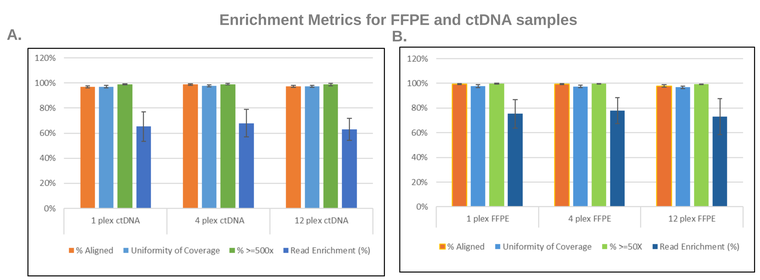
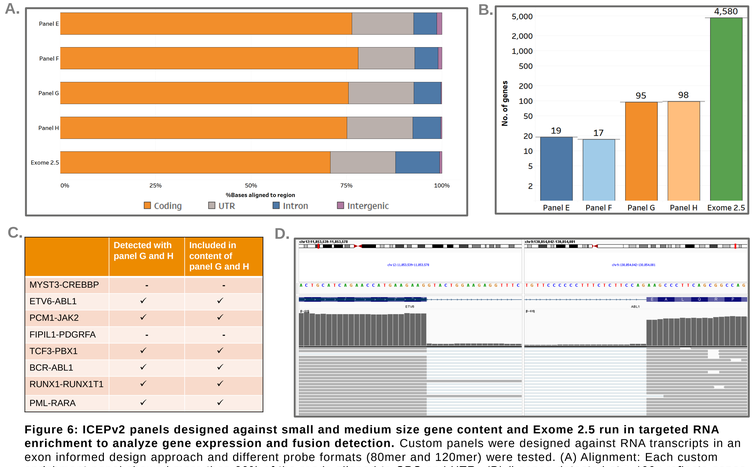


Figure 5. Use of ICEPv2 panel with circulating tumor (ct) DNA and formalin-fixed paraffin embedded (FFPE) samples in targeted DNA enrichment shows high coverage, uniformity and enrichment. A custom panel designed against 79 genes associated with solid tumors was run with (A) ctDNA and (B) FFPE samples at different sample plexities in the hyb reaction (1, 4, and 12 plex). Consistent enrichment and coverage metrics were seen across the different sample types and plexities that were tested. Samples: ctDNA and Covaris-sheared DNA extracted from FFPE. Workflow: Illumina cell-free DNA Prep with Enrichment. Sequencing: NextSeq 2000, MTC ~2500x (ctDNA) and ~900x (FFPE). Analysis: DRAGEN for ILMN Cell-Free DNA Prep with Enrichment.

RNA and Infectious Disease Applications

Table 2. Custom Sequencing Panel specifications used in Illumina RNA Prep with Enrichment workflow

	Panel E	Panel F	Panel G	Panel H	Exome 2.5
No. of target genes	20	20	109	109	~4,800
No. of probes	978	792	4,305	3,794	-
Probe length	80mer	120mer	80mer	120mer	120mer



C.

	Detected with panel G and H	Included in content of panel G and H
MYST3-CREBBP	-	-
ETV6-ABL1	✓	✓
PCM1-JAK2	✓	✓
FIPIL1-PDGRFA	-	-
TCF3-PBX1	✓	✓
BCR-ABL1	✓	✓
RUNX1-RUNX1T1	✓	✓
PML-RARA	✓	✓

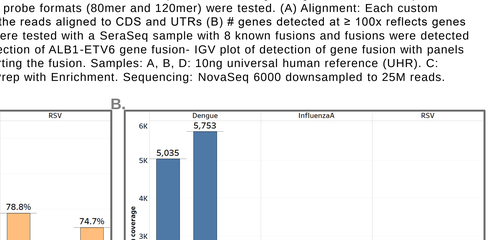


Figure 6: ICEPv2 panels designed against small and medium size gene content and Exome 2.5 run in targeted RNA enrichment to analyze gene expression and fusion detection. Custom panels were designed against RNA transcripts in an exon informed design approach and different probe formats (80mer and 120mer) were tested. (A) Alignment: Each custom enrichment panel showed more than 90% of the reads aligned to CDS and UTRs (B) # genes detected at ≥ 100x reflects genes targeted in each panel (C) Panels G and H were tested with a SeraSeq sample with 8 known fusions and fusions were detected specific to panel content. (D) Successful detection of ALB1-ETV6 gene fusion- IGV plot of detection of gene fusion with panels G and H. The track shows only reads supporting the fusion. Samples: A, B, D: 10ng universal human reference (UHR). C: SeraCare SeraSeq. Workflow: Ilumina RNA Prep with Enrichment. Sequencing: NovaSeq 6000 downsampled to 25M reads. Analysis: DRAGEN RNA

Figure 7: **Application of ICEPv2 panel designed against representative sampling of viruses in detection and quantification of viral RNA** in serum (sample1-3) and nasopharyngeal swab (4-9). Data shown represent three (RSV, FluA, Dengue) out of multiple viruses included in the panel. (A) % Genome detected at >10x depth. (B) Median Coverage. Workflow: Illumina RNA Prep with Enrichment. Sequencing: NextSeq 2000. Analysis: DRAGEN Metagenomics

Conclusion

- Comprehensive end-to-end solution, from probe design to data analysis, for multiple targeted sequencing applications.
- Custom probe panels designed to ensure high uniformity and coverage.
- Enhanced coverage of GC-rich and difficult to sequence areas in Exome and custom panels.
- Wide range of applications and sample types, including germline variant calling, FFPE and ctDNA sequencing, gene expression and fusion detection and viral RNA sequencing.

CAP#- M-GL-03226