LOW-COVERAGE, LONG-READ WHOLE GENOME SEQUENCING FOR STRUCTURAL VARIATION
BEST PRACTICES

With the Sequel® System, Sequel Chemistry 3.0, and SMRT® Link 6.0, you can affordably and sensitively characterize structural variants (SVs) and large indels of all types ranging in size from tens to thousands of base pairs. Low-coverage, long-read whole genome sequencing data provides rapid discovery of common SVs for population genetics studies and resolves rare SVs unique to an individual or disease cohort, with a very low false discovery rate.

FROM gDNA TO MOST COMPREHENSIVE SV CALL SET

SAMPLE PREPARATION RECOMMENDATIONS

- Use recommended high-quality unamplified genomic DNA input (>3 μg) from blood, saliva or cells
- Prepare >15 kb library using SMRTbell® Express Template Preparation Kit
- Enrich for longest inserts with size selection
- Automated library-prep solutions available
- Sequence to desired coverage based on study needs:
  - 5 to 10-fold: population genetics studies - sensitivity limited per individual, but high for variants shared in the population using joint calling
  - 10-fold: rare undiagnosed disease studies - sensitivity high per individual allowing discovery of pathogenic SVs
  - 10 to >20-fold: genetic disease studies - identify a variant or gene that causes disease in a cohort of individuals with a shared phenotype; higher coverage required for de novo SV detection in trios
- ~10 Gb per Sequel SMRT Cell 1M

SV DISCOVERY POWER AT VARIOUS COVERAGE LEVELS

A diploid human (HG00733) was sequenced to 70-fold coverage on the Sequel System. The reads were randomly sampled to various coverage levels, and the SV calls at each coverage were evaluated against the calls at full 70-fold coverage.

Sensitivity increases sharply with coverage until about 10-fold, where it begins to level off. At 10-fold coverage, 10,854 insertions and 7,692 deletions are called (83% and 90.5% sensitivity, respectively).
**DATA ANALYSIS SOLUTIONS WITH SMRT ANALYSIS AND IGV**

- Detect the broadest range of SV types including insertions, deletions, inversions and translocations.
- Discover more SVs with the highest sensitivity – up to 20,000 SVs and 22,000 large indels per genome.
- Identify common SVs across multiple samples with joint calling.
- Resolve breakpoints to sequence level.
- Limit costly validation efforts with a low false-discovery rate of 5-10%.
- Output data in standard file formats – BAM and VCF – for seamless integration with downstream analysis tools.
- Confirm SV calls visually with IGV/GenomeRibbon.

**LONG-READ SMRT SEQUENCING PROVIDES HIGHER SENSITIVITY FOR SV DISCOVERY**

**VISUALLY VALIDATE SV CALLS BY EXPLORING READ EVIDENCE IN IGV**

A comparison of SV sensitivity across short-read sequencing and PacBio long-read sequencing for SV discovery in published human genomes. PacBio long-read sequencing uncovers many SVs missed by short-read methods, providing up to five-fold higher sensitivity for true SV discovery.

Visualization of long-read sequencing evidence in support of an insertion and deletion structural variant in IGV. Insertions are indicated with a purple box with the width of the box proportional to the size of the insertion, and the base pair size is written on the box. Deletions are indicated by a black line. The base pair size of the deletion is written on a white box at the center of the line.

**KEY REFERENCES**