

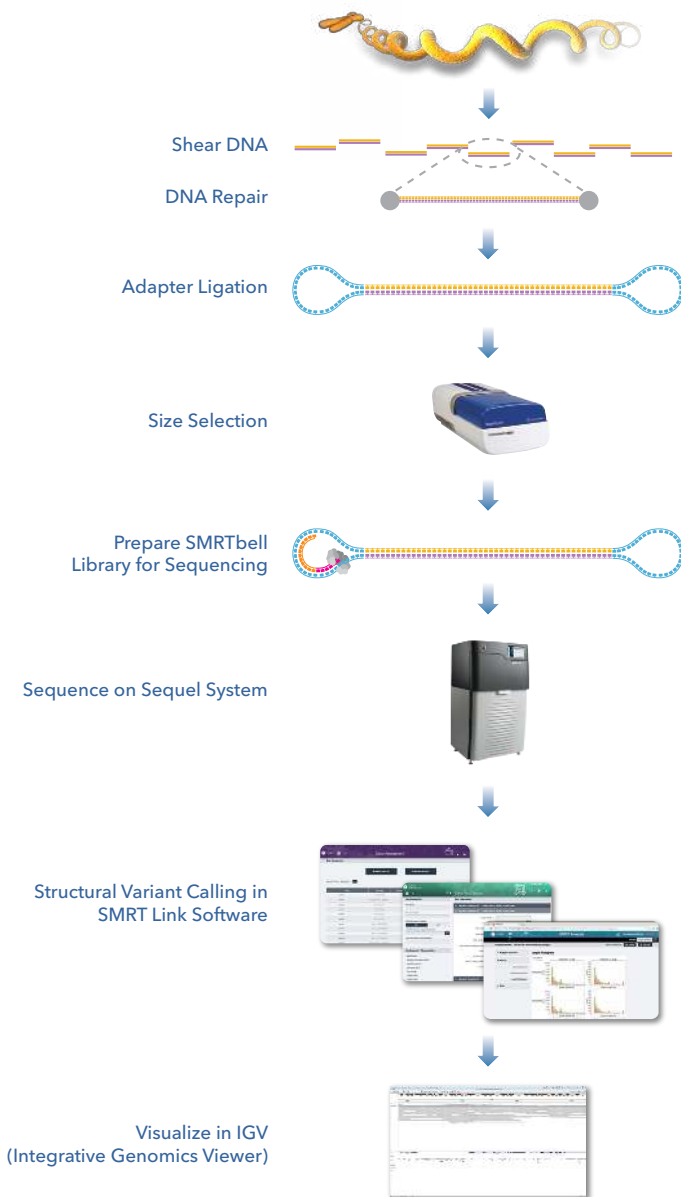


HUMAN BIOMEDICAL RESEARCH

# 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

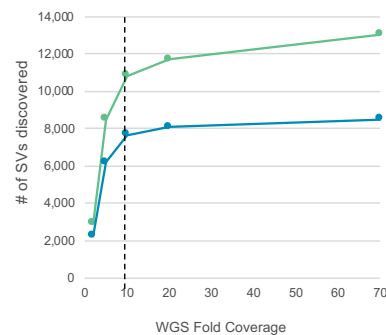


## SAMPLE PREPARATION RECOMMENDATIONS

- 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<sup>1</sup>
- Enrich for longest inserts with size selection<sup>1</sup>
- Automated library-prep solutions available<sup>2</sup>
- Sequence to desired coverage based on study needs<sup>3</sup>:
  - 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

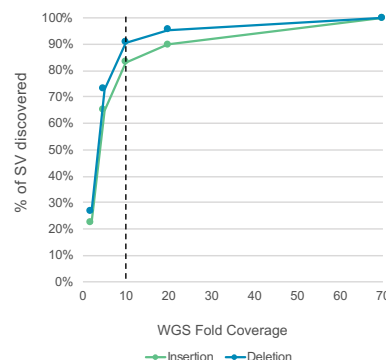
## SV DISCOVERY POWER AT VARIOUS COVERAGE LEVELS

Number of SVs Discovered vs. Coverage



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 for SV Discovery vs. 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)<sup>4</sup>.

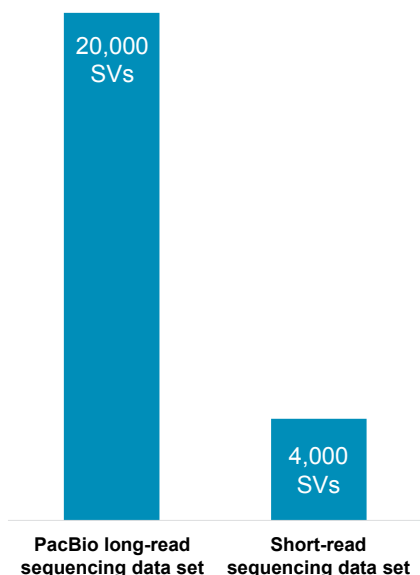
## DATA ANALYSIS SOLUTIONS WITH SMRT ANALYSIS AND IGV

- Detect the broadest range of SV types including insertions, deletions, inversions and translocations<sup>5</sup>
- Discover more SVs with the highest sensitivity - up to 20,000 SVs and 22,000 large indels per genome<sup>4,6</sup>
- Identify common SVs across multiple samples with joint calling
- Resolve breakpoints to sequence level<sup>7</sup>
- Limit costly validation efforts with a low false-discovery rate of 5-10%<sup>8</sup>
- Output data in standard file formats - BAM and VCF - for seamless integration with downstream analysis tools
- Confirm SV calls visually with IGV<sup>7,9</sup> and GenomeRibbon<sup>10</sup>



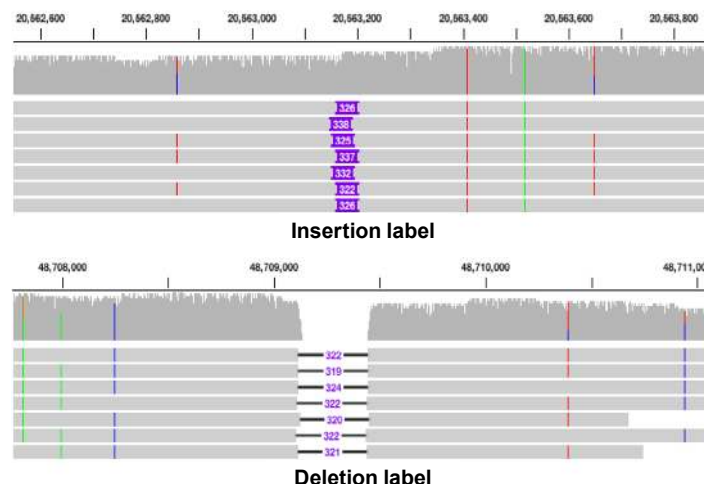
SMRT Link and IGV provide a complete workflow to call and visualize structural variants in PacBio® Sequel System Reads

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



A comparison of SV sensitivity across short-read sequencing and PacBio long-read sequencing for SV discovery in published human genomes<sup>8,11,12</sup>. PacBio long-read sequencing uncovers many SVs missed by short-read methods, providing up to five-fold higher sensitivity for true SV discovery<sup>4,13</sup>.

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



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

1. Procedure and Checklist - Preparing >15 kb Libraries Using SMRTbell® Express Template Preparation Kit.
2. "SMRT Compatible Workflow Products, Template-Prep Automation." *PacBio Products and Services*. PacBio, 10 May 2017.
3. Wenger, A., et al. (2018, February 12) PacBio Structural Variation Project Calculator. <https://www.pacb.com/calculator-structural-variation/>
4. "The Human Genome Structural Variation Consortium". The 1000 Genomes Project. IGS: The International Genome Sample Resource. PacBio Single molecule real time (SMRT) Sequel System high-coverage dataset for HG00733. March 2017.
5. Toepfer, A. and Wenger, A. (2018) pbsv - Calling All Variants: Fast, Accurate, Population-Scale Structural Variant Analysis from Single-Molecule PacBio Reads.
6. Hoischen, A. and Wenger, A. (2018) Sequencing Structural Variants for Disease Gene Discovery and Population Genetics. Nature.com Webcasts.
7. Robinson, J. et al. (2018) The Integrative Genomics Viewer (IGV).
8. Huddleston, J. et al. (2016) Discovery and genotyping of structural variation from long-read haploid genome sequence data. *Genome Research*. (5), 677-685.
9. Robinson JT, et al. (2011) Integrative genomics viewer. *Nature Biotechnology*. 29(1), 24-26.
10. Nattestad M. et al. (2016) Ribbon: Visualizing complex genome alignments and structural variation. *bioRxiv*. <https://doi.org/10.1101/082123>.
11. Seo, J. et al. (2016) De novo assembly and phasing of a Korean human genome. *Nature*. 538, 243-247.
12. Shi, L. et al. (2016) Long-read sequencing and de novo assembly of a Chinese genome. *Nature Communications*. 7, 12065.
13. Sudmant, P. et al. (2015) An integrated map of structural variation in 2,504 human genomes. *Nature*. 526, 75-81.

For Research Use Only. Not for use in diagnostic procedures. © Copyright 2018, Pacific Biosciences of California, Inc. All rights reserved. Information in this document is subject to change without notice. Pacific Biosciences assumes no responsibility for any errors or omissions in this document. Certain notices, terms, conditions and/or use restrictions may pertain to your use of Pacific Biosciences products and/or third party products. Please refer to the applicable Pacific Biosciences Terms and Conditions of Sale and to the applicable license terms at <http://www.pacb.com/legal-and-trademarks/terms-and-conditions-of-sale/>.

Pacific Biosciences, the Pacific Biosciences logo, PacBio, SMRT, SMRTbell, Iso-Seq, and Sequel are trademarks of Pacific Biosciences. BluePippin and SageELF are trademarks of Sage Science. NGS-go and NGSengine are trademarks of GenDx. FEMTO Pulse and Fragment Analyzer are trademarks of Advanced Analytical Technologies. All other trademarks are the sole property of their respective owners.

