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

BEST PRACTICES

With the Sequel® System, you can affordably and sensitively characterize structural variation (SV) of all types ranging in size from tens to thousands of base pairs. Low-coverage, long-read whole genome sequencing (WGS) data provides rapid discovery of common SVs for population genetics studies and resolves rare SVs unique to an individual, 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 ug)
- 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.
  - ~7.5 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 deletions, insertions, duplications, inversions, tandem repeats and translocations.
- Discover more SVs with the highest sensitivity - up to 20,000 SVs per genome.
- Identify common structural variants 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 and GenomeRibbon.

SMRT Link and IGV provide a complete workflow to call and visualize structural variants in PacBio 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. PacBio long-read sequencing uncovers many structural variants missed by short-read methods, providing up to five-fold higher sensitivity for true SV discovery.

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 basepair size is written on the box. Deletions are indicated by a black line. The basepair size of the deletion is written on a white box at the center of the line.

KEY REFERENCES


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/.

PacBio, 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.