

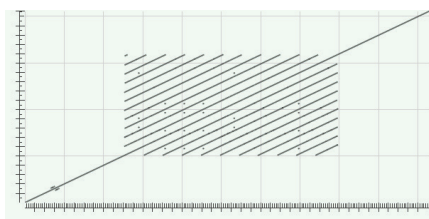
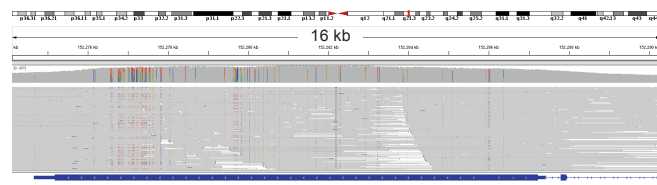
TRACK THE FULL EXTENT OF STRUCTURAL VARIATION IN A GENOME



Uncovering a connection between genotype and phenotype requires examination of all the variant types in a genome. Single Molecule, Real-Time (SMRT®) Sequencing provides contiguous long reads, high consensus accuracy, and uniform coverage opening up access to a broad range of structural variants (SV) across a wide range of sizes, types, and GC content.

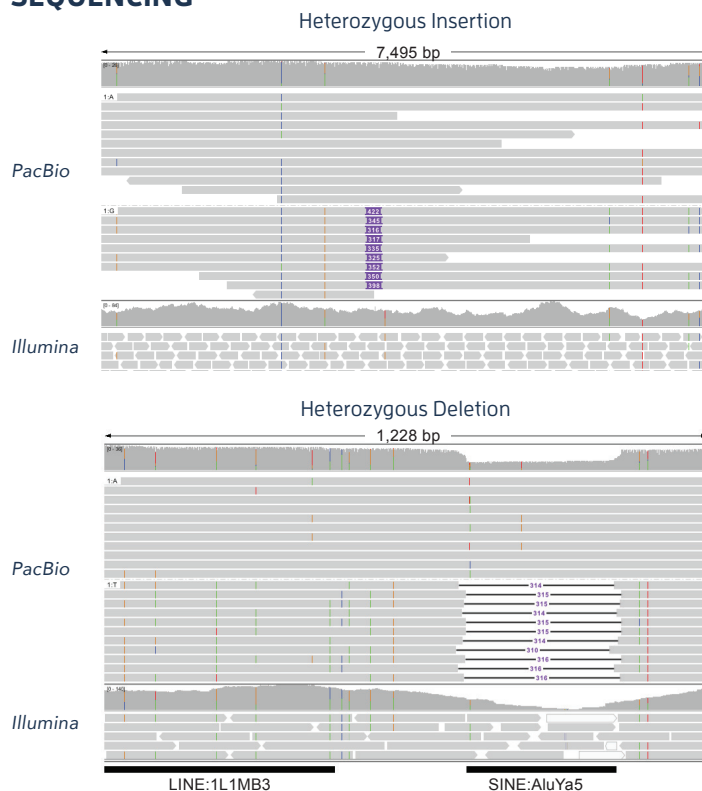
- Discover both common and rare SVs unique to an individual with low-coverage, long-read whole genome sequencing (WGS)
- Target complex regions to resolve allelic indels, copy number variants, novel repeat expansions, and pseudogenes in order to finely map causal variants and understand functional disease mechanisms
- Identify novel SVs with *de novo* long-read WGS assembly
- Phase SVs and single nucleotide variants into allele-specific haplotype blocks
- Increase sensitivity and specificity for SV calls in previously generated sequencing data

TARGET AND RESOLVE COMPLEX STRUCTURAL VARIANTS



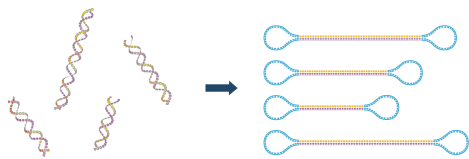
The NimbleGen Neurology gene panel (exon only) was used to capture the Filaggrin gene region in a sample. SMRT Sequencing of the enriched sample shows uniform coverage across the gene (top) and cleanly indicates this individual is homozygous for 12 repeats in the complex exon 1 region, demonstrated by the dot plot (bottom) between the two phased alleles¹.

IMPROVE DETECTION OF STRUCTURAL VARIATION WITH LOW-COVERAGE, LONG-READ WHOLE GENOME SEQUENCING



An NA12878 SV call set was generated with 10-fold coverage of Sequel data. 84% of known SVs were recovered and 1000s of potentially novel SVs were discovered, consistent with previous studies^{2,3,4}.

FROM DNA TO STRUCTURAL VARIANT CALLS



Library Preparation

- Prepare long-insert SMRTbell™ templates for WGS
 - Prepare >7 kb libraries
 - Use recommended gDNA library input (~5 – 10 ug of purified gDNA)
 - Enrich for longest inserts with size-selection options
- Utilize standard target enrichment methods
 - Amplify a few genes, using standard or long-range PCR (<1 kb to >10 kb)
 - Capture ~7 kb fragments with probe enrichment methods for multiple genes or large genomic regions (>50 kb)
 - Multiplex 384 barcoded targeted regions
- Library automation supported

SMRT Sequencing with PacBio Systems

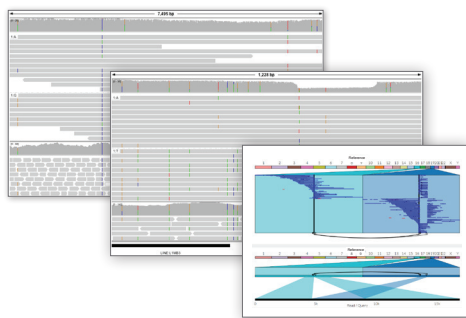
- Take advantage of the Sequel™ System for high-throughput projects to reduce costs and generate ~7-fold more reads compared with the PacBio RS II
- Achieve ~10 kb average read lengths, with some reads as long as 60 kb
- Scale throughput based on project needs and to optimize SV sensitivity
 - 5-fold WGS coverage: Survey and discover common SV across a population
 - 10-fold WGS coverage: Discover and genotype common and rare SV within individuals
 - >30-fold WGS coverage: *De novo* assembly of an individual genome
 - >60-fold WGS coverage: *De novo* diploid-aware assembly and reconstruction of haplotigs representing individual chromosomes
 - >30 to 100-fold targeted coverage: Finely resolve SV within targeted region



PacBio RS II System



Sequel System



Data analysis and visualization with SMRT Analysis or PacBio DevNet

- Utilize community-developed SV calling methods such as PBHoney and Parliament⁴, included in the DNAnexus cloud-based analysis toolset
- Visualize long reads mapped against the human reference genome with the Integrative Genomics Viewer (IGV)⁵ to observe evidence for insertions and deletions
- Directly phase SV by allele using concordant SNV and SV calls and the “Quick Phase” Visualization mode in IGV
- Explore complex structural variants, including inversions and translocations, using Genome Ribbon⁶

KEY REFERENCES

1. Gu, J. et al. (Sep, 2016) Multiplex target enrichment using barcoded multi-kilobase fragments and probe-based capture technologies. *Poster presented at Advances in Genome Biology Technology - Precision Health Medicine*. Scottsdale, AZ.
2. Wenger, A. et al. (Oct, 2016) Effect of coverage depth and haplotype phasing on structural variant detection with PacBio long reads. *Poster #3206 presented at 65th Annual Meeting of The American Society of Human Genetics*. Vancouver, BC.
3. Chaisson, M. J. P. et al. (2014) Resolving the complexity of the human genome using single molecule sequencing. *Nature*. 517(7536), 608-611.
4. English, A. et al. (2015) Assessing structural variation in a personal genome – towards a human reference diploid genome. *BMC Genomics*. 16, 286.
5. Robinson JT, et. al., (2011) Integrative Genomics Viewer. *Nature Biotechnology*. 29, 24-26
6. Nattestad M, (2016) Analysis and visualization tools for long reads, assemblies and complex variation. <https://www.youtube.com/watch?v=1h4Wf2U10-4> Accessed on Sept 30, 2016.
7. Gordon, D. et al. (2016) Long-read sequence assembly of the gorilla genome. *Science*. 352(6281), aae0344.
8. McFarland, K. N. et al. (2015) SMRT Sequencing of long tandem nucleotide repeats in SCA10 reveals unique insight of repeat expansion structure. *PLoS One*. 10(8), e0135906.
9. Qiao W, et. al, (2016) Long-read Single Molecule Real-Time full gene sequencing of cytochrome P450-2D6. *Human Mutation*. 37(3), 315-323.
10. Rafati, N. (2016) Large deletions at the *SHOX* locus in the pseudoautosomal region are associated with skeletal atavism in Shetland ponies. *G3*. 6(7), 2213-2223.
11. Wang, M. et al. (2015) PacBio-LITS: a large-insert targeted sequencing method for characterization of human disease-associated chromosomal structural variations. *BMC Genomics*. 16, 214.

For Research Use Only. Not for use in diagnostic procedures. © Copyright 2016, 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. Fragment Analyzer is a trademark of Advanced Analytical Technologies. All other trademarks are the sole property of their respective owners.

