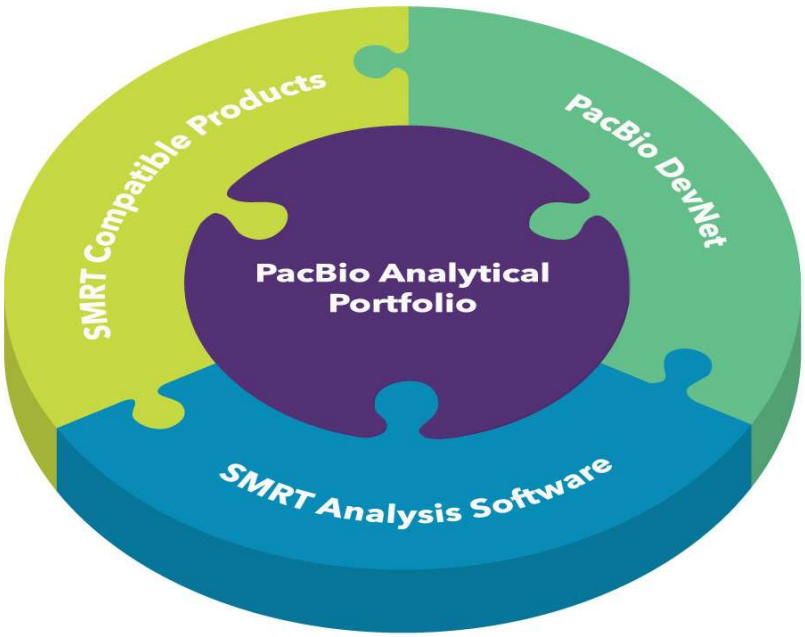
GAIN A DEEPER UNDERSTANDING
OF YOUR SEQUENCING DATA

PACBIO®

Unlock the potential of PacBio® Single Molecule, Real-Time (SMRT®) Sequencing using our portfolio of tools to analyze, visualize and annotate your sequencing data. This analytical portfolio can be utilized across our applications and includes SMRT Link with SMRT Analysis Software, the PacBio DevNet of community-developed analytical tools and SMRT compatible products with software solutions by trusted partners.

- **SMRT Analysis Software** - Open source software platform for analysis and interpretation of SMRT Sequencing data
  - Intuitive GUI (Graphical User Interface) enables analysis setup and monitoring, review, and exploration of results
  - Command line interface provides extended analysis flexibility
  - Extensive APIs (Application Programing Interface) facilitate easy integration with analytical workflows
- **PacBio DevNet** – Advanced bioinformatics methods and novel applications for PacBio data, developed through our continuous collaboration with the open-source community. Applications include *de novo* assembly, structural variants detection, RNA analysis, epigenetic modifications analysis.
- **SMRT Compatible Products** - Products from qualified partners that complement our analytical offerings and enable end-to-end solutions. Products include High Performance Compute and cloud infrastructure, Laboratory Information Management Systems, analytical services, analysis and annotation applications.



Join the SMRT Community to download analysis tools for the SMRT Sequencing data:

www.pacb.com/software

SMRT ANALYSIS SOFTWARE

SMRT Analysis is web-based software for PacBio data. It includes a comprehensive set of applications for *de novo* assembly, variant identification, RNA analysis, and detection of epigenetic modifications. SMRT Analysis provides analytical support for barcoding and sequencing multiple samples on the same SMRT Cell.

- **Hierarchical Genome Assembly Process (HGAP)** - Generates high-quality *de novo* assemblies with long sequence continuity and high accuracy (>QV50)
- **Structural Variant Calling** - Identifies variants ≥20 base pairs in a sample or set of samples
- **Iso-Seq® Analysis** - Characterizes transcripts and splice variants (*de novo* or reference-based)
- **Circular Consensus Sequences (CCS)** - Creates High-Fidelity (Q20+) reads by identifying consensus sequences for single DNA molecules
- **Long Amplicon Analysis (LAA)** - Identifies phased consensus sequences from a heterogeneous pool of amplicons – no reference needed
- **Minor Variant Analysis –** Detects, quantifies and phases minor variants in comparison to a reference sequence
- **Base Modification Detection** - Finds specific modified sites in unamplified genomes
- **Base Modification and Motif Analysis** - Determines common bacterial epigenetic modifications and analyzes the methyltransferase recognition motifs
- **Demultiplex Barcodes** - Identifies barcodes for multiplexed samples

COMPUTATIONAL AND STORAGE REQUIREMENTS

Table below includes High Performance Compute (HPC) requirements for efficient and reliable analysis of PacBio SMRT Sequencing data using SMRT Analysis.

HPC Configurations / HPC Components	Targeted Applications HPC[1]	Targeted Applications PLUS HPC[2]	Whole Genome Applications HPC[3]
Head Node			
64 GB RAM 32 Cores			
Compute Nodes			
Cores	18 physical or 36 hyper-threaded	96 physical or 192 hyper-threaded	192 physical or 384 hyper-threaded
Min RAM per Slot (1 slot = 1 core)	4 GB	4 GB	8 GB
Local Disk per Node ($SMRT_ROOT/userdata/tmpdir)	100 GB	1 TB	1 TB
HDD Operation Speed	N/A	**15 TB**[4] serving 1800 IOPS	**30 TB**[4] serving 1800 IOPS
Long Term Data Storage ($SMRT_ROOT/userdata)			
	10 TB	38 TB	70 - 100 TB
Network			
10 GBE recommended (1 GBE required)			

1 Targeted sequencing applications (CCS, LLA, resequencing), microbial genome assembly, and targeted Iso-Seq analysis. Long-term storage is calculated based on moderate usage of Sequel® System per year.

2 Targeted sequencing applications noted above plus moderate throughput of large genome de novo assemblies and whole transcriptome Iso-Seq analysis. Long-term storage is calculated based on moderate usage of Sequel System per year.

3 For human-scale genomes with 50-fold coverage and target assembly time approximately 72 hours. Long-term storage is calculated for one Sequel System assuming 50 human-scale genomes per year at 50-fold coverage.

4 Non-redundant storage dedicated to this compute environment - choose from NFS, Open Source or Commercial DFS, or CIFS.

PN: SW100-100918