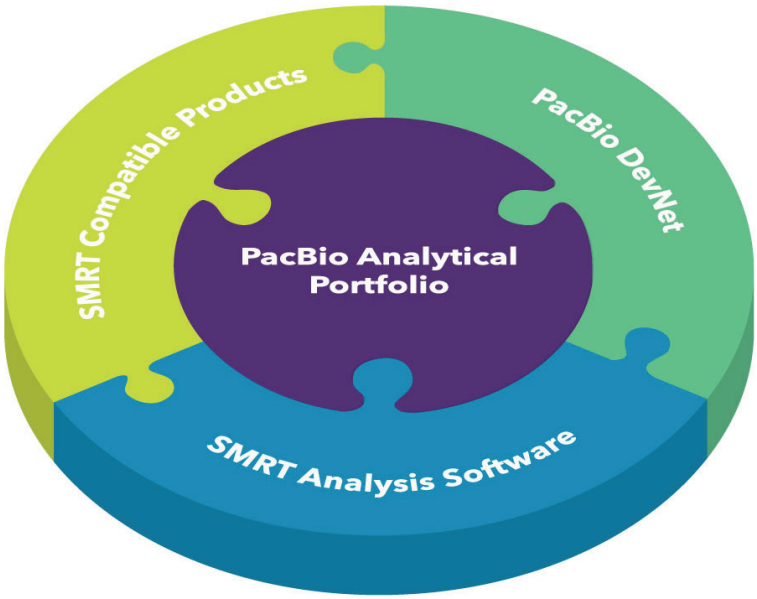
GAIN A DEEPER UNDERSTANDING
OF YOUR SEQUENCING DATA

PACBIO®

The PacBio® Platform includes an extensive software portfolio that employs key advantages of SMRT® (Single Molecule, Real-Time) Sequencing technology: extraordinarily long reads, highest consensus accuracy, uniform coverage and simultaneous epigenetic characterization. Core elements of our analytical portfolio include:

- **SMRT Analysis Software** - Our 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
- **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, etc.
- **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, etc.



Join the SMRT Community to download open source software tools and data from our developers' network:
www.pacb.com/devnet

SMRT ANALYSIS SOFTWARE

SMRT Analysis is web-based software for PacBio data. It includes a comprehensive set of applications for *de novo* assembly, sequence alignment and consensus calling, variant identification, RNA analysis, and methylation detection. 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)
- **BridgeMapper** - Provides quality control for *de novo* assemblies, identification of circular DNA genetic elements, and detection of gene fusion events for resequencing applications
- **BLASR** - Aligns reads to a reference sequence
- **Quiver** - Generates highly accurate consensus sequence, calls SNPs and indels
- **Arrow -** Generates highly accurate consensus sequence based on statistical methods
- **Long Amplicon Analysis (LAA)** - Identifies phased consensus sequences from a heterogeneous pool of amplicons
- **Circular Consensus Sequence (CCS)** - Identifies consensus sequences for single DNA molecules
- **Minor Variant** - Detects minor variants in comparison to a reference sequence
- **Iso-Seq™ Analysis** - Characterizes transcripts and splice variants (*de novo* or reference-based)
- **Modification Detection** - Finds specific modified sites in unamplified genomes
- **Modification and Motif Analysis** - Determines common bacterial base modifications and analyzes the methyltransferase recognition motifs
- **Barcoding Analysis** - Identifies barcodes for multiplexed samples

COMPUTATIONAL AND STORAGE REQUIREMENTS

HPC Configurations / HPC Components	Targeted Applications HPC*	Targeted Applications PLUS HPC**	Large-genome *de novo* HPC***
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
RAM per Node (GB)	256	256	256
Local Disk per Node	100 GB	1 TB	1 TB
Intermediate Pipeline Storage	N/A	**15 TB**** serving 1800 IOPS	**30 TB**** serving 1800 IOPS
Long Term Data Storage			
	10 TB	38 TB	70 - 100 TB
Network			
10 GBE recommended (1 GBE required)			

* Targeted Sequencing applications (CCS, LAA, resequencing), assembly of bacterial genomes, Iso-Seq application. Long-term storage is calculated based on moderate usage of Sequel™ System per year.

** Targeted Sequencing applications as noted above plus occasional large-genome *de novo* assemblies. Long-term storage is calculated based on moderate usage of Sequel System per year.

*** 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 52 human genomes per year at 50-fold coverage.

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

PN: SW100-011617