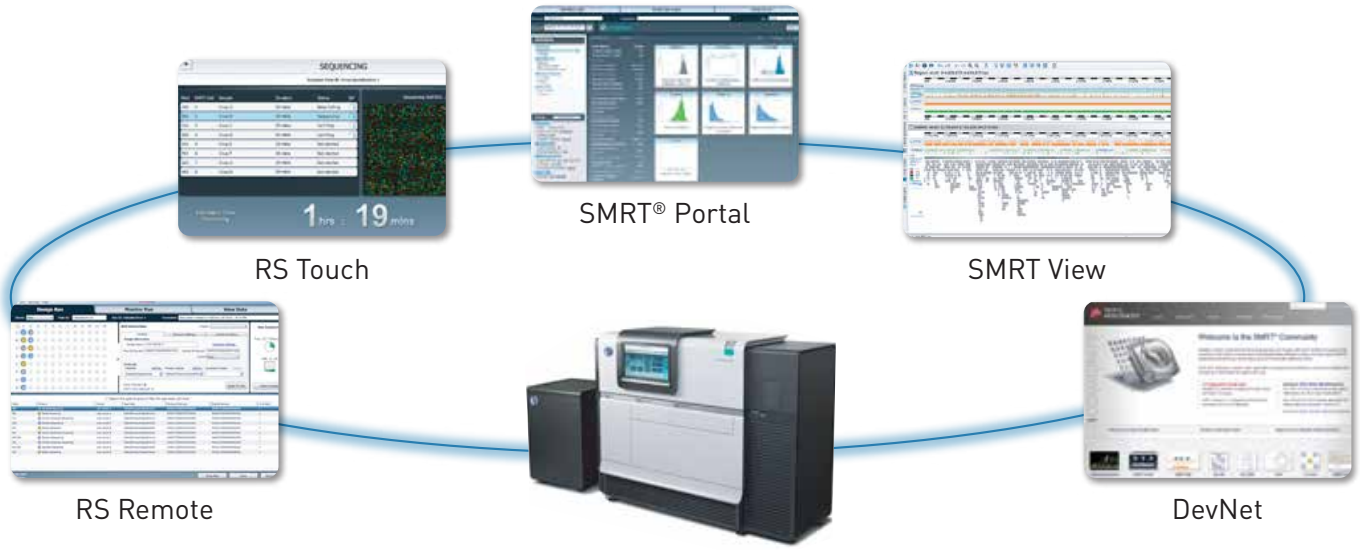
PACIFIC BIOSCIENCES®

# PacBio Software and Analysis Overview

The PacBio® Sequencing System includes an integrated software solution from beginning to end to support the key benefits of SMRT® Sequencing data: extraordinarily long reads, extremely high accuracy, exquisite sensitivity to variants, and unique epigenetic and methylation detection.

- Fully automated analysis from run setup, with the option to manually run later
- Efficient integration with LIMS and third-party analysis tools
- User-friendly UI design for advanced informatics researchers as well as biologists and clinicians
- Industry-standard output formats: FASTA, FASTQ, SAM/BAM, VCF, etc.
- Open source through our developers' network




## PacBio Instrument Software

- **RS Touch** uses a touchscreen interface to start and monitor runs from the instrument
- **RS Remote** is used to design runs directly from your desktop
- **RS Dashboard** is a web-based tool used to remotely monitor and evaluate runs

# SMRT® Analysis Software

## SMRT Pipe and SMRT Portal

SMRT Pipe provides command-line access to a range of tools for PacBio data. SMRT Portal lets you create and monitor SMRT Pipe jobs and download the results from a rich web interface.

### Targeted Sequencing

| | |
|---|---|
| BLASR | Map reads against a reference with BLASR |
| Quiver | Call haploid SNPs and indels with 99.999% accuracy |
| GATK | Identify haploid and diploid SNPs using the Broad's Unified Genotyper |
| GMAP | Align full-length cDNA transcripts against genomic DNA to discover splicing |
| SMRT View | Browse coverage, variants and annotations |

### *De Novo* Genome Assembly

| | |
|---|---|
| ALLORA | Assemble pure PacBio long reads, then polish with Quiver |
| AHA | Fill gaps and join existing scaffolds with PacBio long reads |
| Celera® Assembler | Combine PacBio long reads with short reads from other technologies |
| HGAP | Generate high quality assemblies from PacBio long reads alone |
| SMRT View | QC assemblies |

### DNA Base Modification Analysis

| | |
|---|---|
| Modification detection | Find specific modified sites in unamplified genomes |
| Bacterial methylomes | Discover recognition motifs for adenine and cytosine methylation |
| SMRT View | Visualize modified sites and sequence contexts |

## Join the SMRT Community

Find more open source tools from our developers' network, PacBio DevNet. View and download open source software, data, documentation and tips for SMRT Sequencing.

## IT Requirements

Network: 1GbE connection required; 2GbE recommended. Four (4) static IP addresses. Network address, subnet mask, default gateway. DNS servers, DNS default search domain, NTP servers, outbound NAT address or range, outbound internet, ssh/scp access, port 3306 open

Compute cluster: 5 nodes x 8 cores/node. 2GB RAM per core. 250GB/node

Storage: Minimum 10TB recommended

## File Formats and Sizes

| | | |
|---|---|---|
| Primary analysis | Base calls and quality values | HDF5*, FASTQ, FASTA |
| Primary analysis | Kinetics | HDF5* |
| Resequencing | Aligned reads and consensus calls | HDF5, SAM/BAM |
| Resequencing | Variants | GFF, VCF |
| *De novo* assembly | Assembled reads | HDF5, ACE, SAM/BAM |

*Base calls, quality values and kinetics are stored in the same file



PN: BR104-041113