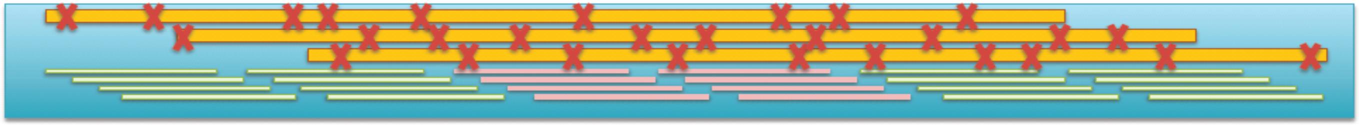
天津生物芯片技术有限责任公司
www.tjbiochip.com

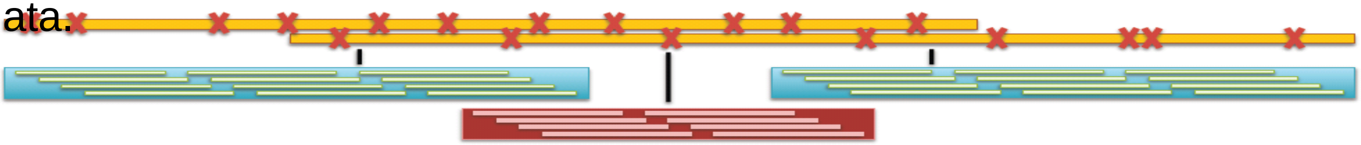
PacBio Assembly Softwares

End to End

SMRT Hybrid: Refers to the hybrid *de novo* assembly of error-corrected PacBio Continuous Long Read (“CLR”) data (lower accuracy) with a second data type with higher accuracy - either PacBio Circular Consensus Sequence reads (“CCS”) data or 2nd generation short-read data.

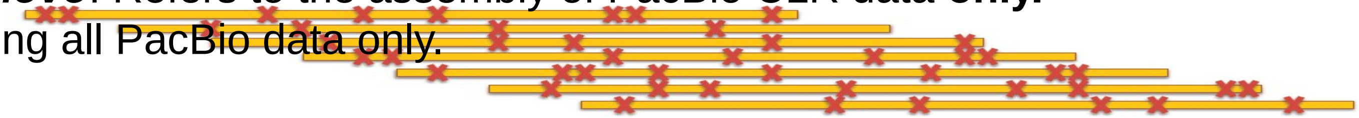


SMRT Scaffolding: Refers to using PacBio CLR to scaffold existing contigs generated from short-read data.

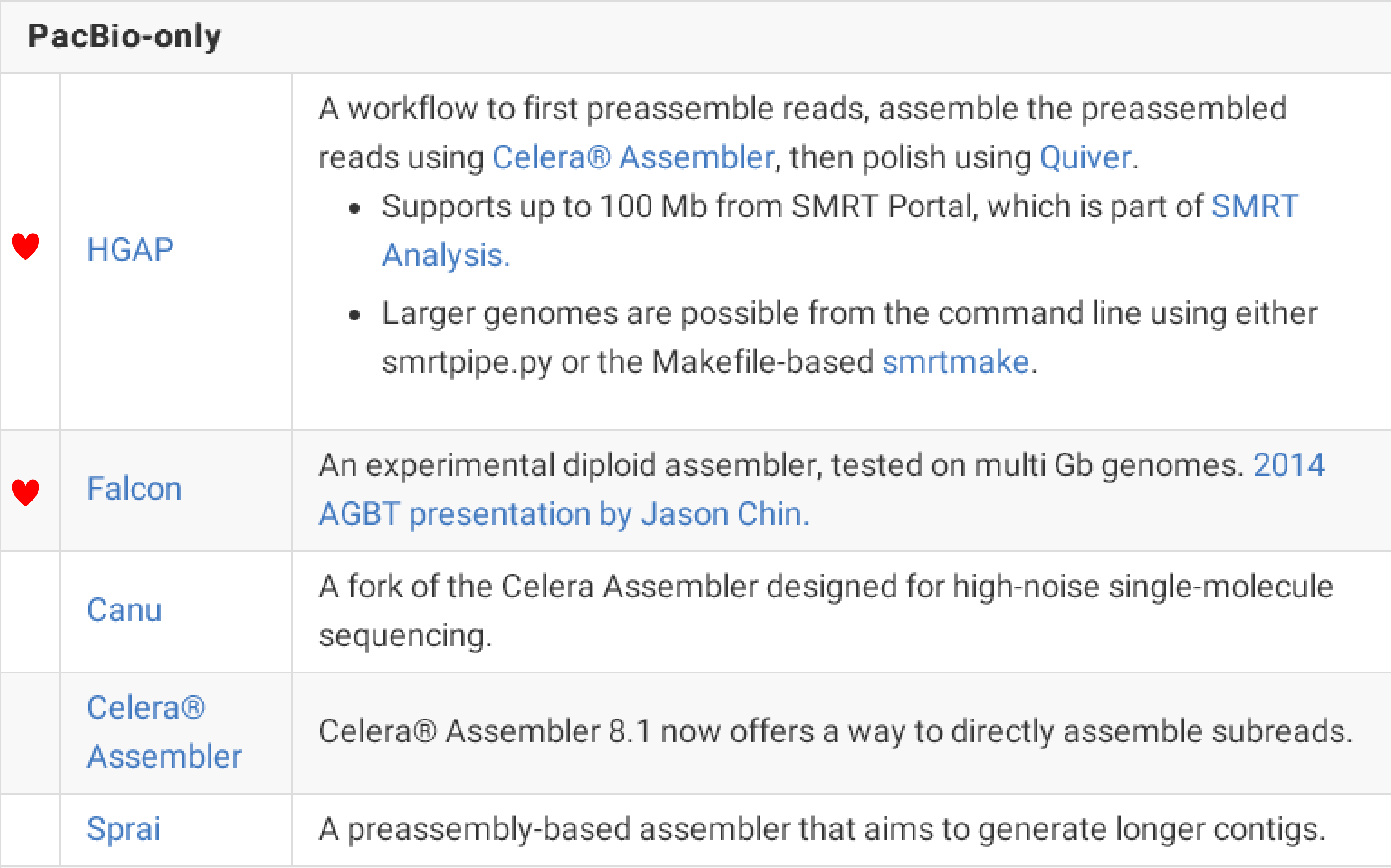


SMRT *de novo*: Refers to the assembly of PacBio CLR data **only.**

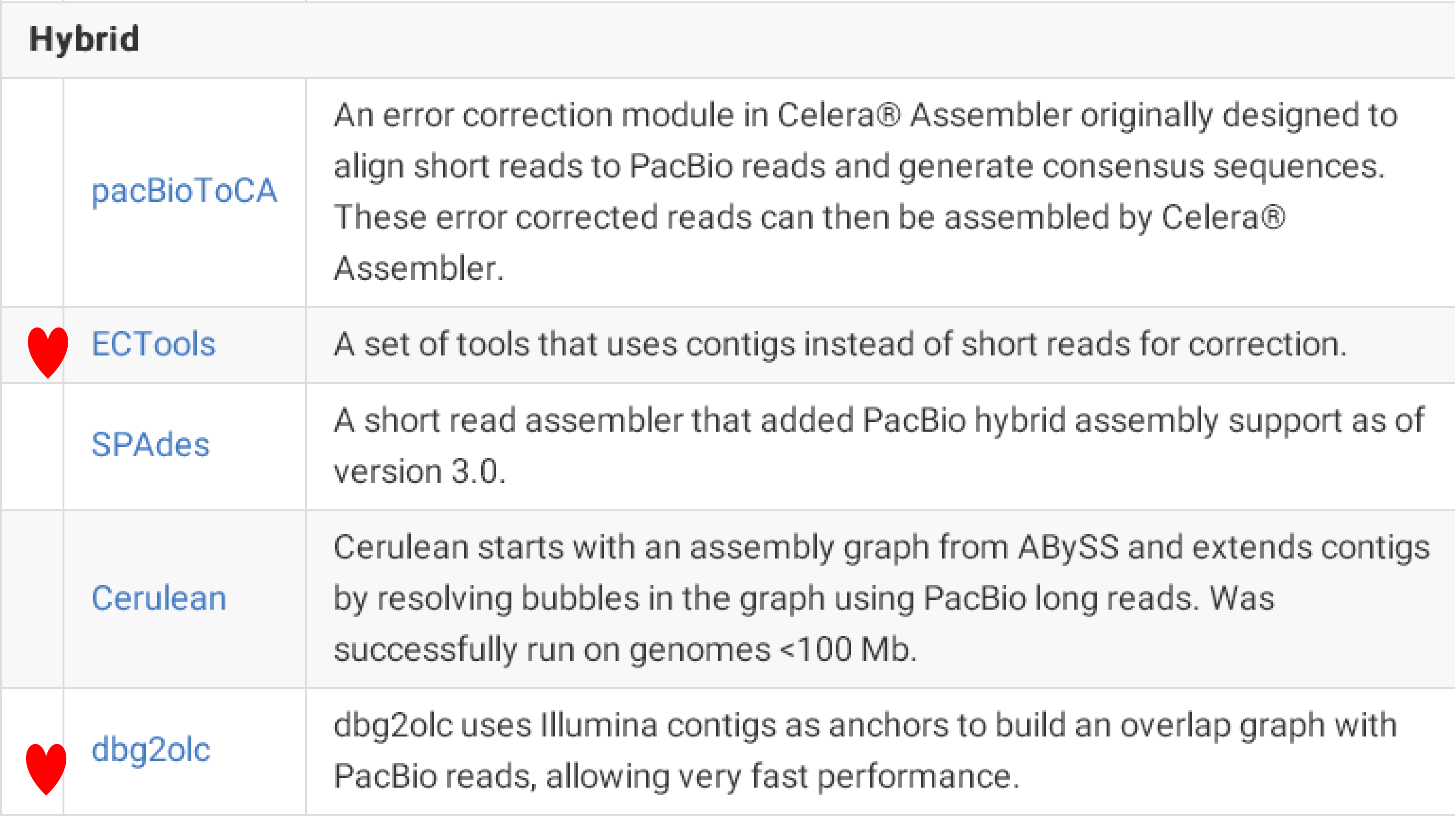
HGAP: Using all PacBio data only.



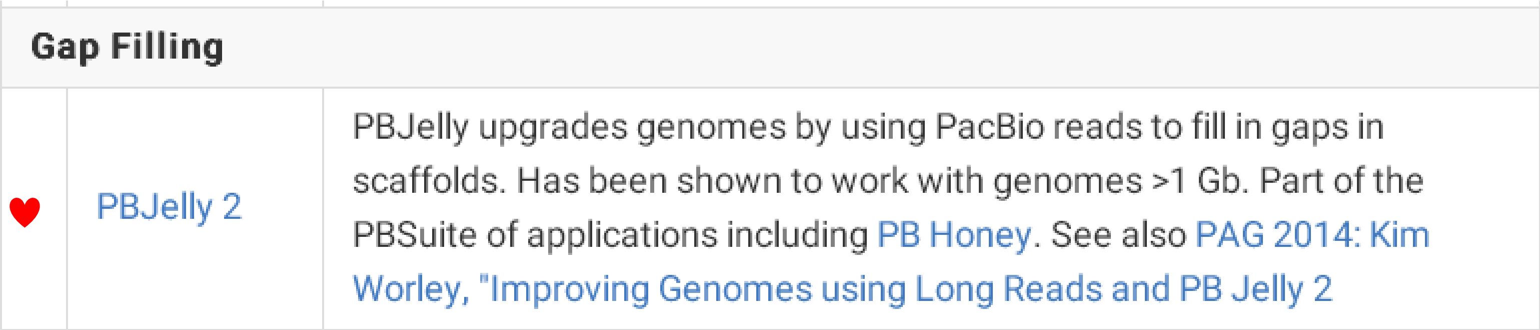
SMRT Gap Filling: Refers to using PacBio CLR to fill gaps in existing mate pair-based scaffolds



PacBio-only		
♥	HGAP	A workflow to first preassemble reads, assemble the preassembled reads using Celera® Assembler, then polish using Quiver. • Supports up to 100 Mb from SMRT Portal, which is part of SMRT Analysis. • Larger genomes are possible from the command line using either smrtpipe.py or the Makefile-based smrtmake.
♥	Falcon	An experimental diploid assembler, tested on multi Gb genomes. 2014 AGBT presentation by Jason Chin.
	Canu	A fork of the Celera Assembler designed for high-noise single-molecule sequencing.
	Celera® Assembler	Celera® Assembler 8.1 now offers a way to directly assemble subreads.
	Sprai	A preassembly-based assembler that aims to generate longer contigs.

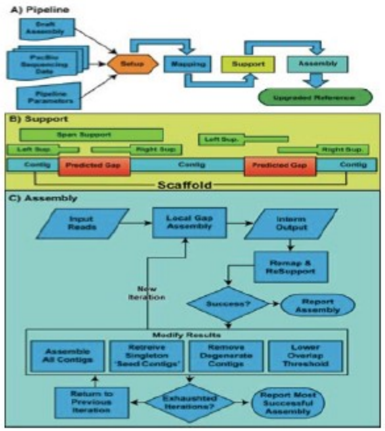


Hybrid		
	pacBioToCA	An error correction module in Celera® Assembler originally designed to align short reads to PacBio reads and generate consensus sequences. These error corrected reads can then be assembled by Celera® Assembler.
♥	ECTools	A set of tools that uses contigs instead of short reads for correction.
	SPAdes	A short read assembler that added PacBio hybrid assembly support as of version 3.0.
	Cerulean	Cerulean starts with an assembly graph from ABySS and extends contigs by resolving bubbles in the graph using PacBio long reads. Was successfully run on genomes <100 Mb.
♥	dbg2olc	dbg2olc uses Illumina contigs as anchors to build an overlap graph with PacBio reads, allowing very fast performance.



Gap Filling		
♥	PBJelly 2	PBJelly upgrades genomes by using PacBio reads to fill in gaps in scaffolds. Has been shown to work with genomes >1 Gb. Part of the PBSuite of applications including PB Honey. See also PAG 2014: Kim Worley, "Improving Genomes using Long Reads and PB Jelly 2

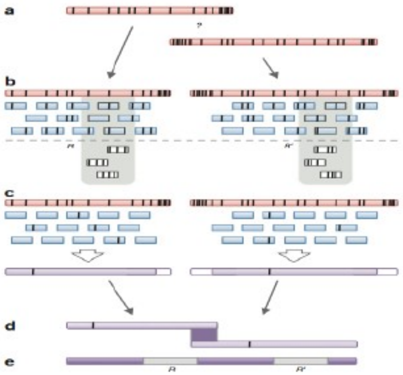
PBJelly



Gap Filling and Assembly Upgrade

English *et al* (2012)
PLOS One. 7(11): e47768

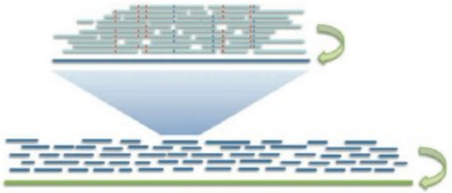
PacBioToCA & ECTools



Hybrid/PB-only Error Correction

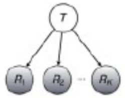
Koren, Schatz, *et al* (2012)
Nature Biotechnology. 30:693–700

HGAP & Quiver



$\Pr(\mathbf{R} \mid T)$

$$\Pr(\mathbf{R} \mid T) = \prod_k \Pr(R_k \mid T)$$



Quiver Performance Results
Comparison to Reference Genome
(*M. ruber* ; 3.1 MB ; SMRT® Cells)

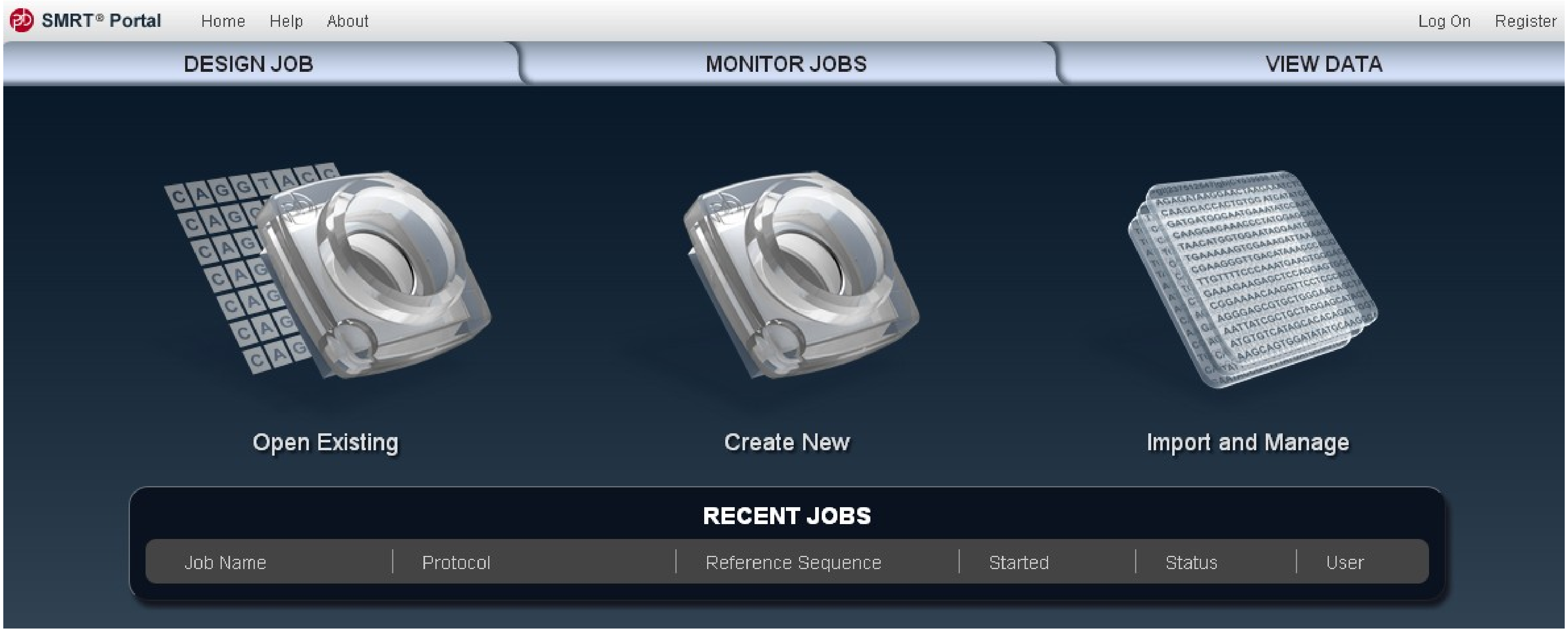
	Initial Assembly	Quiver Consensus
QV	43.4	54.5
Accuracy	99.99540%	99.99964%
Differences	141	11

PB-only Correction & Polishing

Chin *et al* (2013)
Nature Methods. 10:563–569

< 5x PacBio Coverage > 50x

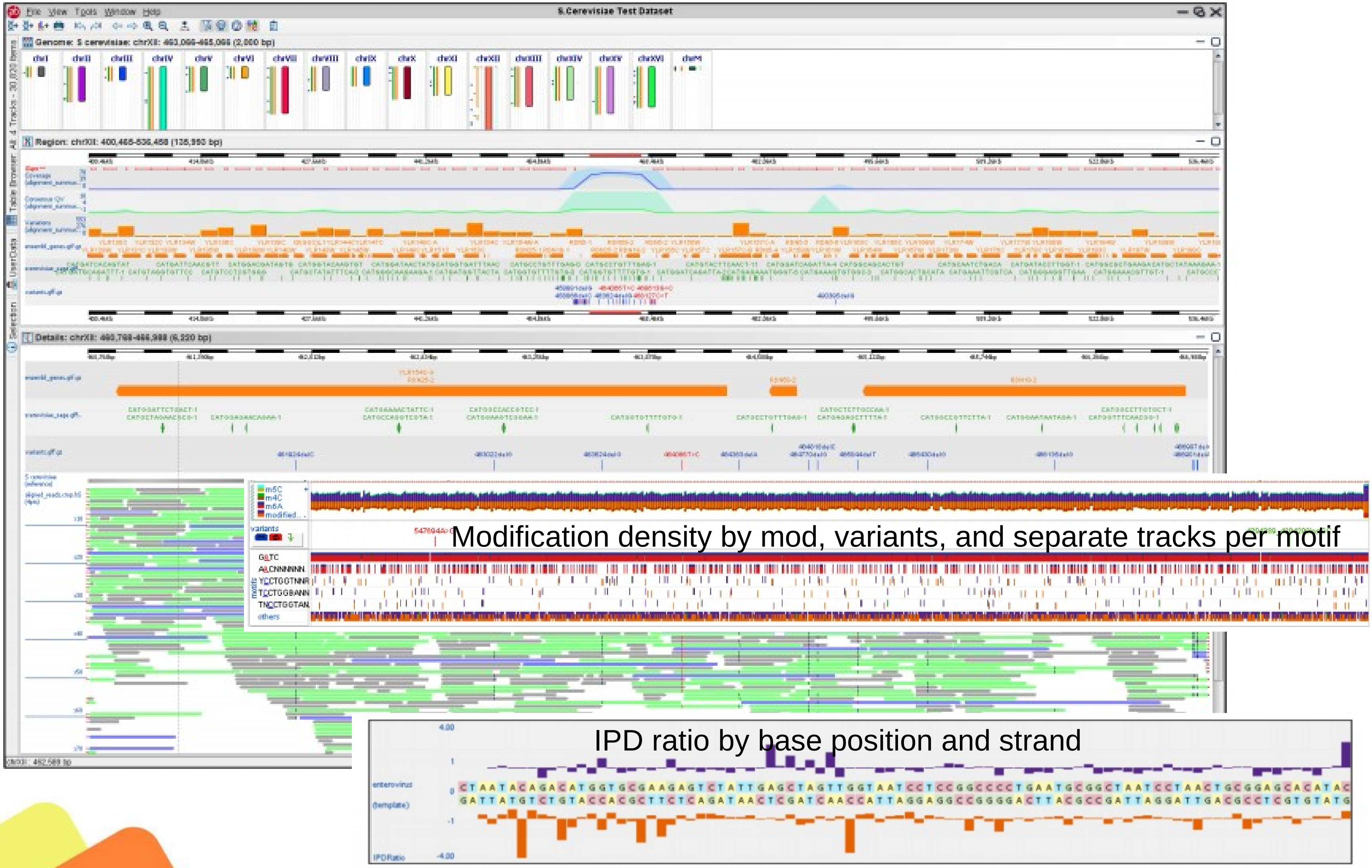
SMRT Portal Overview



RECENT JOBS

Job Name	Protocol	Reference Sequence	Started	Status	User

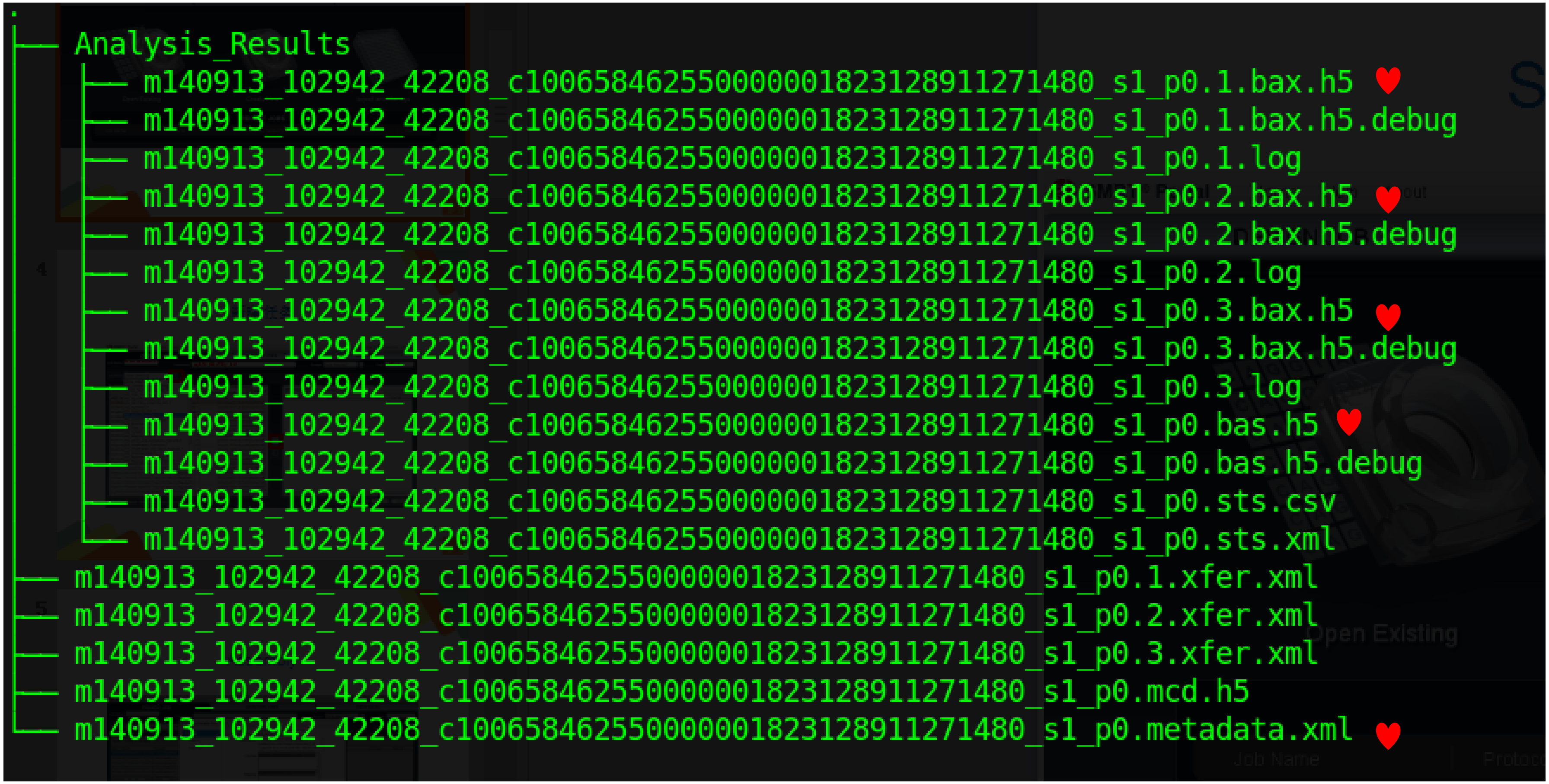
SMRT VIEW


File View Tools Window Help
S.Cerevisiae Test Dataset
Genome: S cerevisiae: chrXII: 463,066-465,066 (2,000 bp)
chrI chrII chrIII chrIV chrV chrVI chrVII chrVIII chrIX chrX chrXI chrXII chrXIII chrXIV chrXV chrXVI chrM
Region: chrXII: 400,465-536,458 (135,993 bp)
Coverage
Consensus QV
Variants
ensembl_genes.gff.gz
variants.gff.gz
Details: chrXII: 460,768-466,988 (6,220 bp)
m5C
m4C
m6A
modified...
variants
GATC
AACNNNNNN
YCCTGGTNNR
TCCTGGBANN
TNCCTGGTAN
others
4.00
1
0
-1
-4.00
enterovirus
(template)
IPDRatio
CTAATACAGACATGGTGCGAAGAGTCTATTGAGCTAGTTGGTAATCCTCCGGCCCCTGAATGCGGCTAATCCTAACTGCGGAGCACATAC
GATTATGTCTGTACCACGCTTCTCAGATAACTCGATCAACCATTAGGAGGCCGGGGACTTACGCCGATTAGGATTGACGCCTCGTGTATG

Modification density by mod, variants, and separate tracks per motif

IPD ratio by base position and strand

下机数据文件夹内容



```
.
├── Analysis_Results
│   ├── m140913_102942_42208_c100658462550000001823128911271480_s1_p0.1.bax.h5 ♥
│   ├── m140913_102942_42208_c100658462550000001823128911271480_s1_p0.1.bax.h5.debug
│   ├── m140913_102942_42208_c100658462550000001823128911271480_s1_p0.1.log
│   ├── m140913_102942_42208_c100658462550000001823128911271480_s1_p0.2.bax.h5 ♥
│   ├── m140913_102942_42208_c100658462550000001823128911271480_s1_p0.2.bax.h5.debug
│   ├── m140913_102942_42208_c100658462550000001823128911271480_s1_p0.2.log
│   ├── m140913_102942_42208_c100658462550000001823128911271480_s1_p0.3.bax.h5 ♥
│   ├── m140913_102942_42208_c100658462550000001823128911271480_s1_p0.3.bax.h5.debug
│   ├── m140913_102942_42208_c100658462550000001823128911271480_s1_p0.3.log
│   ├── m140913_102942_42208_c100658462550000001823128911271480_s1_p0.bas.h5 ♥
│   ├── m140913_102942_42208_c100658462550000001823128911271480_s1_p0.bas.h5.debug
│   ├── m140913_102942_42208_c100658462550000001823128911271480_s1_p0.sts.csv
│   └── m140913_102942_42208_c100658462550000001823128911271480_s1_p0.sts.xml
├── m140913_102942_42208_c100658462550000001823128911271480_s1_p0.1.xfer.xml
├── m140913_102942_42208_c100658462550000001823128911271480_s1_p0.2.xfer.xml
├── m140913_102942_42208_c100658462550000001823128911271480_s1_p0.3.xfer.xml
├── m140913_102942_42208_c100658462550000001823128911271480_s1_p0.mcd.h5
└── m140913_102942_42208_c100658462550000001823128911271480_s1_p0.metadata.xml ♥
```

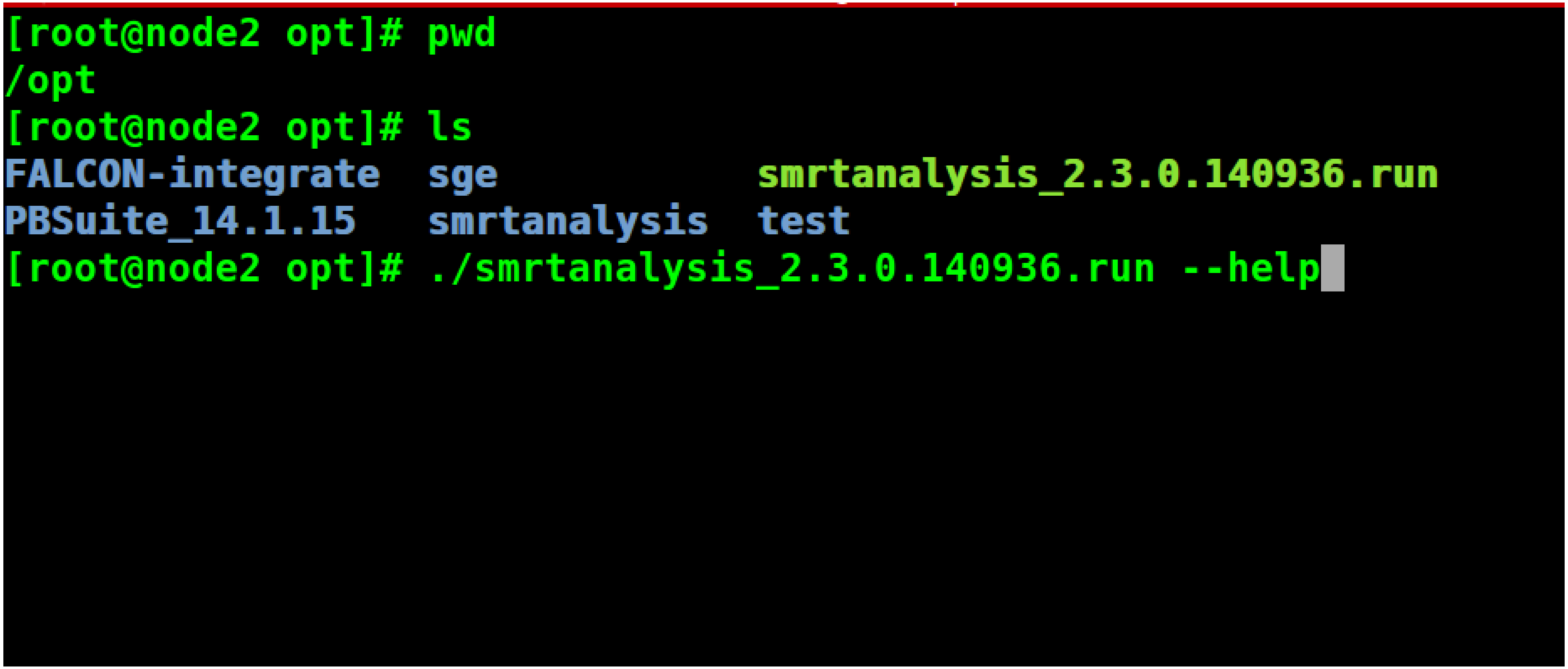
SGE

集群，简单的说即设置一台电脑用作主控主机（ qmaster ），收集集群信息、分配任务、调节负载均衡等；设置多台电脑用作执行主机

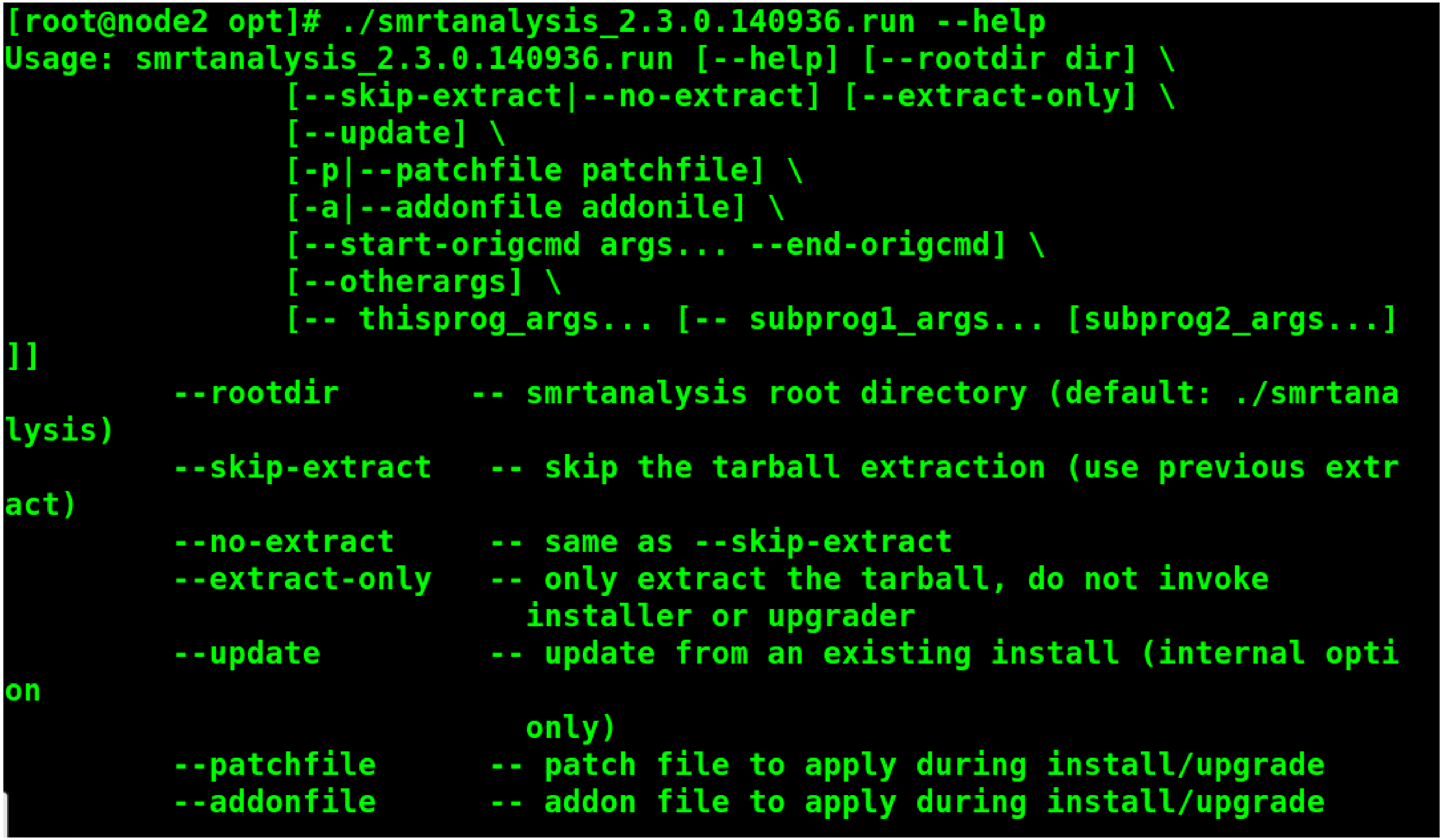
qsub – 此命令是将作业提交到 Sun Grid Engine 系统的用户界面。
qhost – 此命令显示 Sun Grid Engine 执行主机的状态信息。
qdel
qstat
...

SMRT Analysis 的安装

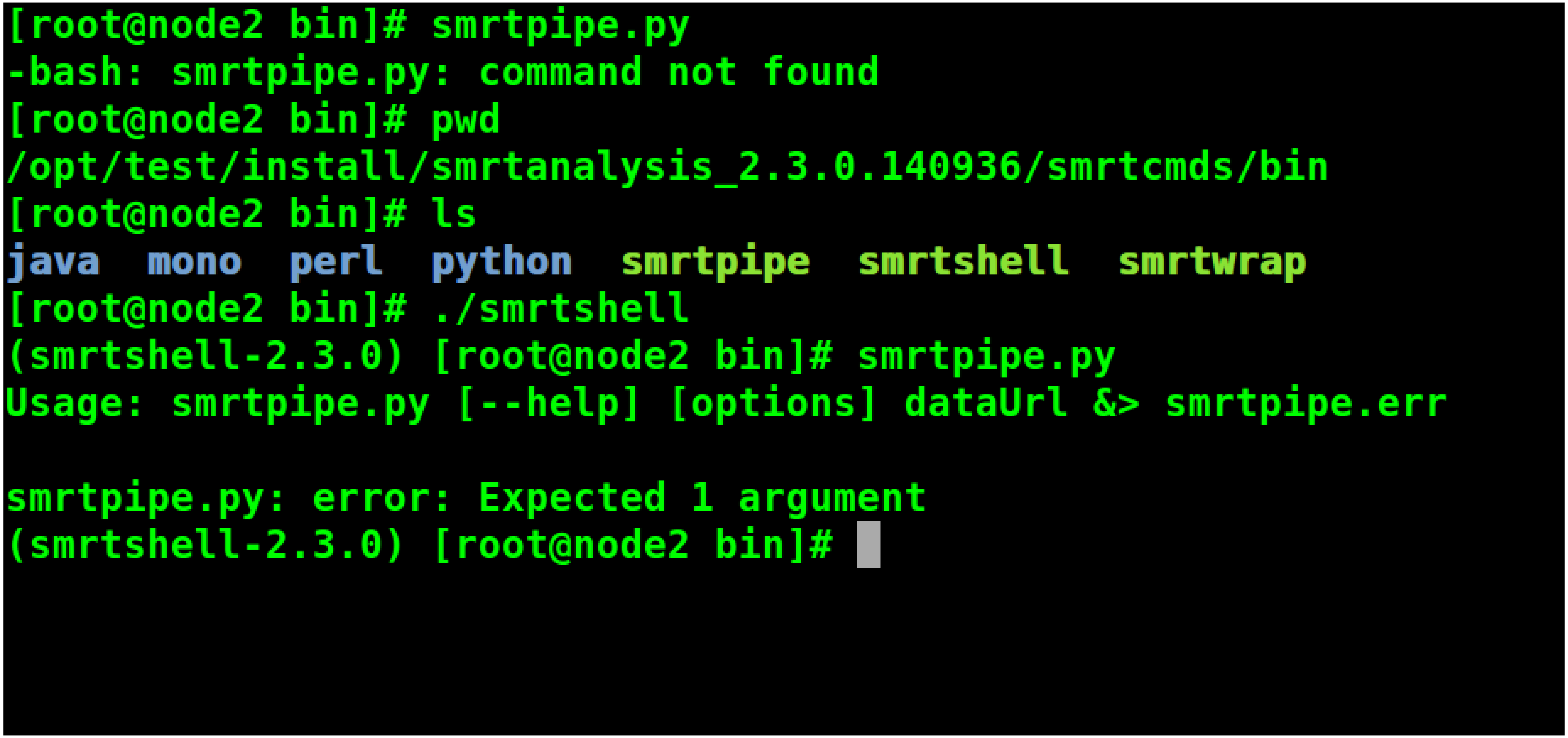
```
Usage: smrtanalysis_2.3.0.140936.run [--help] [--rootdir dir] \
              [--skip-extract|--no-extract] [--extract-only] \
              [--update] \
              [-p|--patchfile patchfile] \
              [-a|--addonfile addonile] \
              [--start-origcmd args... --end-origcmd] \
              [--otherargs] \
              [-- thisprog_args... [-- subprog1_args... [subprog2_args...]]]
        --rootdir        -- smrtanalysis root directory (default: ./smrtanalysis)
        --skip-extract   -- skip the tarball extraction (use previous extract)
        --no-extract     -- same as --skip-extract
        --extract-only   -- only extract the tarball, do not invoke
                          installer or upgrader
        --update         -- update from an existing install (internal option
                          only)
        --patchfile      -- patch file to apply during install/upgrade
        --addonfile      -- addon file to apply during install/upgrade
        --start-origcmd  -- original args to parent program, until
                          until --end-origcmd arg
        --otherargs      -- unrecognized args passed to subprogs until
        -- args...       -- force args to be handled by this program
        -- -- args...    -- force args to be handled by this subprog1
                          (first level subprog)
        -- -- -- args... -- force args to be handled by this subprog2
                          (second level subprog)
                          recognized
        --help           -- print this usage
        -- --help        -- print this usage
        --helpall        -- print this usage and usage of subprogs
```



```
[root@node2 opt]# pwd
/opt
[root@node2 opt]# ls
FALCON-integrate  sge             smrtanalysis_2.3.0.140936.run
PBSuite_14.1.15   smrtanalysis   test
[root@node2 opt]# ./smrtanalysis_2.3.0.140936.run --help
```



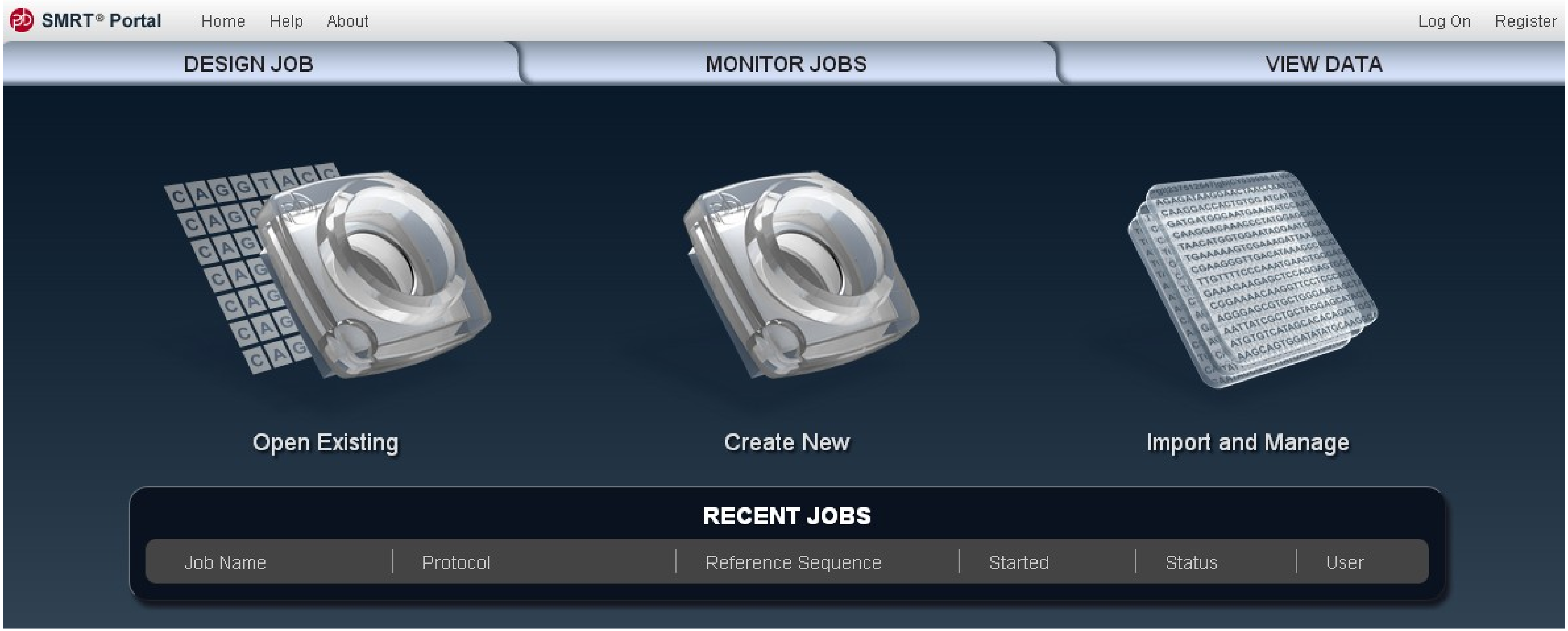
```
[root@node2 opt]# ./smrtanalysis_2.3.0.140936.run --help
Usage: smrtanalysis_2.3.0.140936.run [--help] [--rootdir dir] \
               [--skip-extract|--no-extract] [--extract-only] \
               [--update] \
               [-p|--patchfile patchfile] \
               [-a|--addonfile addonile] \
               [--start-origcmd args... --end-origcmd] \
               [--otherargs] \
               [-- thisprog_args... [-- subprog1_args... [subprog2_args...]
]]
         --rootdir       -- smrtanalysis root directory (default: ./smrtana
lysis)
         --skip-extract   -- skip the tarball extraction (use previous extr
act)
         --no-extract     -- same as --skip-extract
         --extract-only   -- only extract the tarball, do not invoke
                            installer or upgrader
         --update         -- update from an existing install (internal opti
on
                            only)
         --patchfile      -- patch file to apply during install/upgrade
         --addonfile      -- addon file to apply during install/upgrade
```



```
[root@node2 bin]# smrtpipe.py
-bash: smrtpipe.py: command not found
[root@node2 bin]# pwd
/opt/test/install/smrtanalysis_2.3.0.140936/smrtcmds/bin
[root@node2 bin]# ls
java  mono  perl  python  smrtpipe  smrtshell  smrtwrap
[root@node2 bin]# ./smrtshell
(smrtshell-2.3.0) [root@node2 bin]# smrtpipe.py
Usage: smrtpipe.py [--help] [options] dataUrl &> smrtpipe.err

smrtpipe.py: error: Expected 1 argument
(smrtshell-2.3.0) [root@node2 bin]# 
```

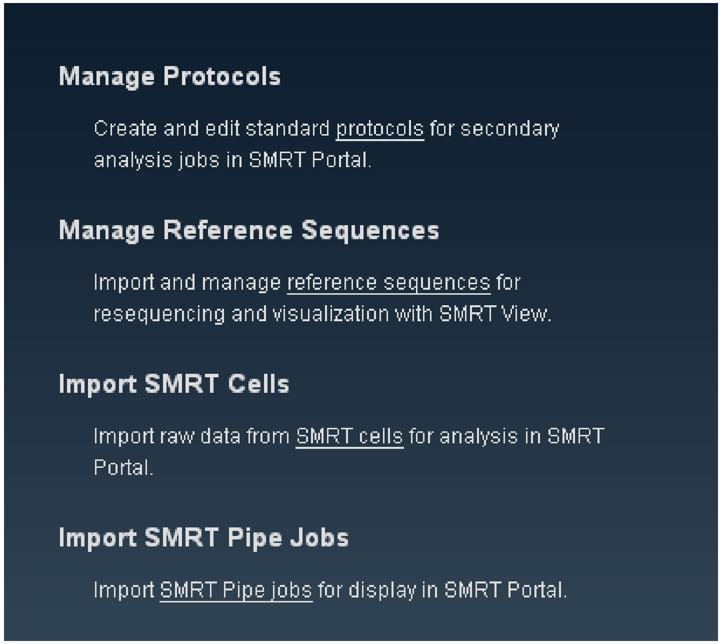
SMRT Portal Overview



RECENT JOBS

Job Name	Protocol	Reference Sequence	Started	Status	User

数据导入



Manage Protocols

Create and edit standard protocols for secondary analysis jobs in SMRT Portal.

Manage Reference Sequences

Import and manage reference sequences for resequencing and visualization with SMRT View.

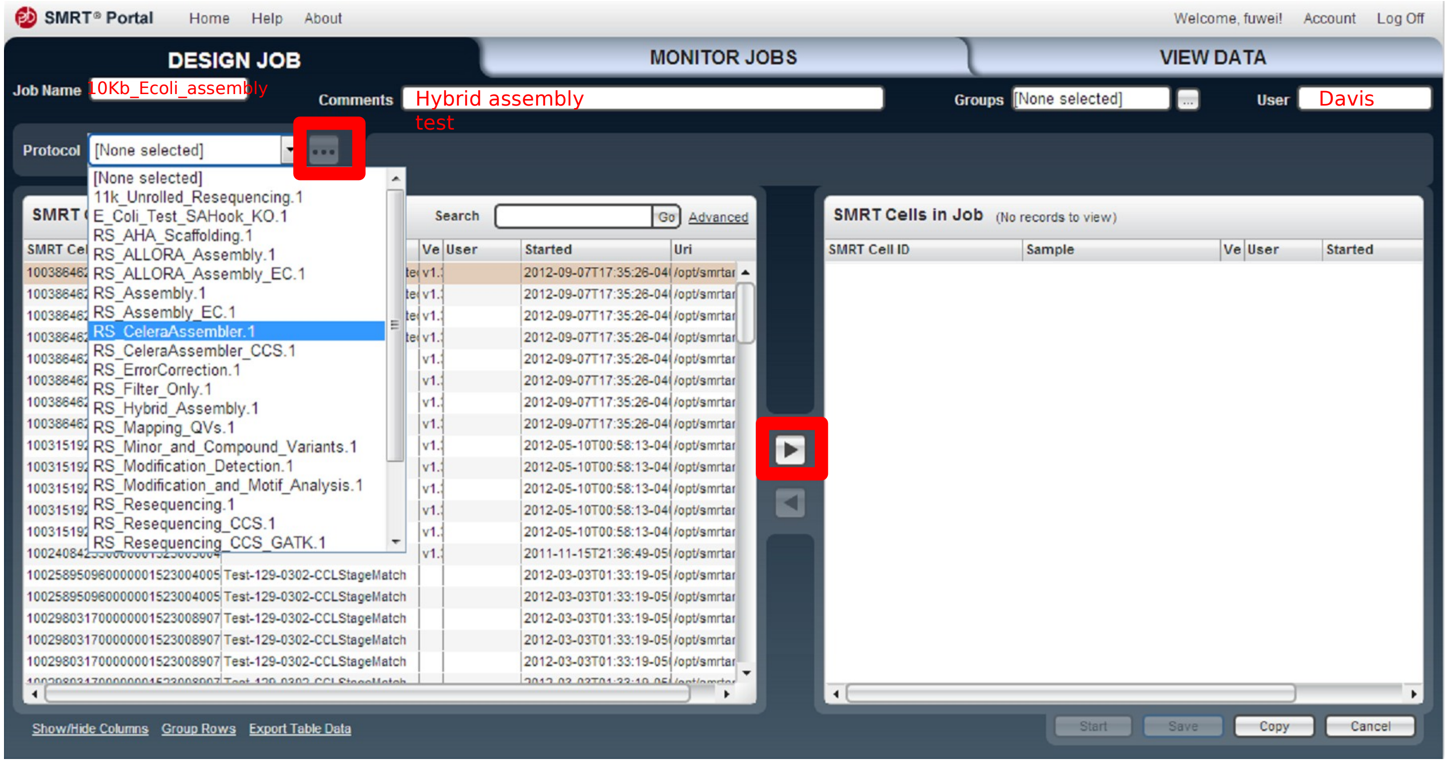
Import SMRT Cells

Import raw data from SMRT cells for analysis in SMRT Portal.

Import SMRT Pipe Jobs

Import SMRT Pipe jobs for display in SMRT Portal.

新建任务



Job Name: 10Kb_Ecoli_assembly

Comments: Hybrid assembly test

Groups: [None selected]

User: Davis

Protocol: [None selected]

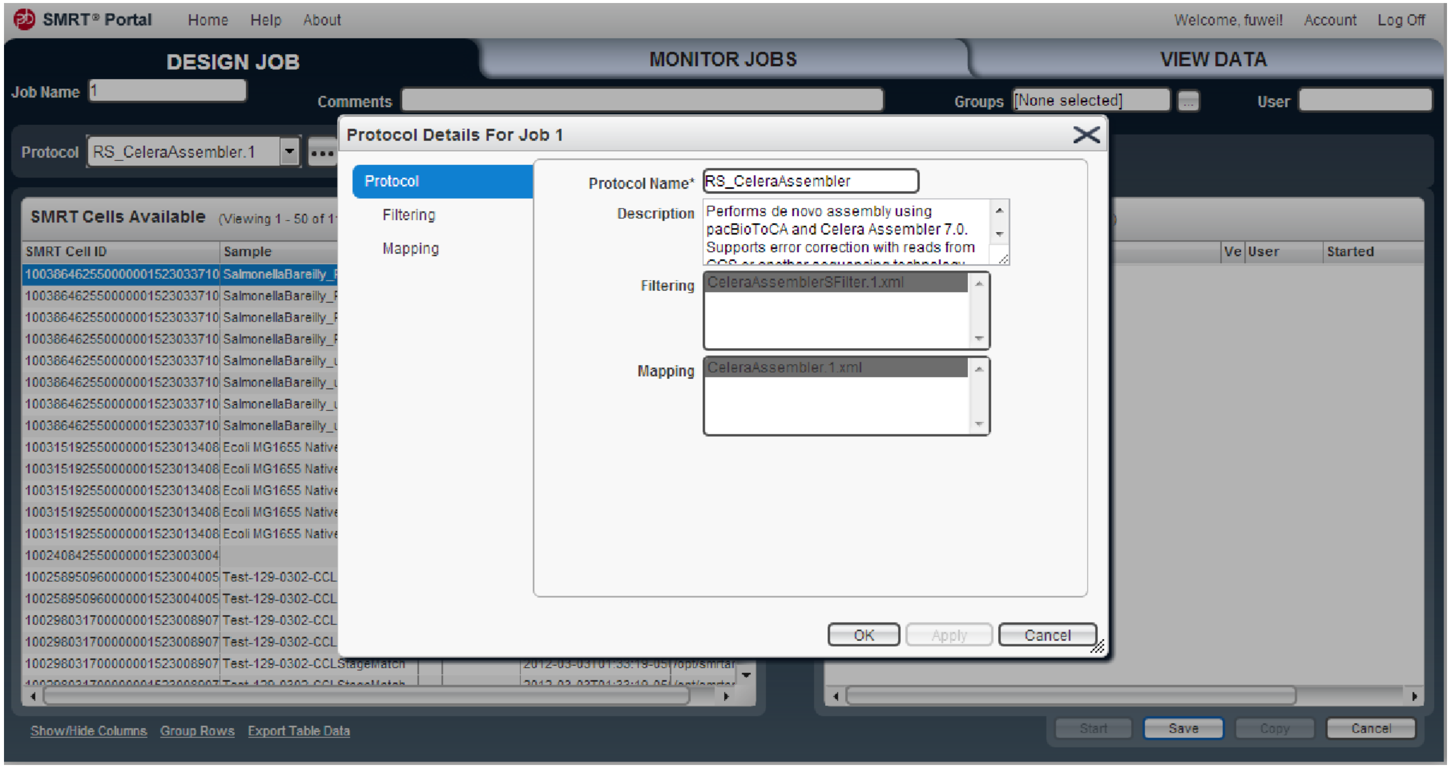
SMRT Cells

SMRT Cel		Ve	User	Started	Uri
100386462	te	v1.:		2012-09-07T17:35:26-04(	/opt/smrtar
100386462	te	v1.:		2012-09-07T17:35:26-04(	/opt/smrtar
100386462	te	v1.:		2012-09-07T17:35:26-04(	/opt/smrtar
100386462	te	v1.:		2012-09-07T17:35:26-04(	/opt/smrtar
100386462		v1.:		2012-09-07T17:35:26-04(	/opt/smrtar
100386462		v1.:		2012-09-07T17:35:26-04(	/opt/smrtar
100386462		v1.:		2012-09-07T17:35:26-04(	/opt/smrtar
100386462		v1.:		2012-09-07T17:35:26-04(	/opt/smrtar
100315192		v1.:		2012-05-10T00:58:13-04(	/opt/smrtar
100315192		v1.:		2012-05-10T00:58:13-04(	/opt/smrtar
100315192		v1.:		2012-05-10T00:58:13-04(	/opt/smrtar
100315192		v1.:		2012-05-10T00:58:13-04(	/opt/smrtar
100315192		v1.:		2012-05-10T00:58:13-04(	/opt/smrtar
10024084		v1.:		2011-11-15T21:36:49-05(	/opt/smrtar
1002589509600000001523004005	Test-129-0302-CCLStageMatch			2012-03-03T01:33:19-05(	/opt/smrtar
1002589509600000001523004005	Test-129-0302-CCLStageMatch			2012-03-03T01:33:19-05(	/opt/smrtar
1002980317000000001523008907	Test-129-0302-CCLStageMatch			2012-03-03T01:33:19-05(	/opt/smrtar
1002980317000000001523008907	Test-129-0302-CCLStageMatch			2012-03-03T01:33:19-05(	/opt/smrtar
1002980317000000001523008907	Test-129-0302-CCLStageMatch			2012-03-03T01:33:19-05(	/opt/smrtar

SMRT Cells in Job (No records to view)

SMRT Cell ID	Sample	Ve	User	Started

Assembly



Job Name 1

Comments

Groups [None selected]

User

Protocol RS_CeleraAssembler.1

SMRT Cells Available (Viewing 1 - 50 of 1

SMRT Cell ID	Sample			
100386462550000001523033710	SalmonellaBareilly_F			
100386462550000001523033710	SalmonellaBareilly_F			
100386462550000001523033710	SalmonellaBareilly_F			
100386462550000001523033710	SalmonellaBareilly_F			
100386462550000001523033710	SalmonellaBareilly_u			
100386462550000001523033710	SalmonellaBareilly_u			
100386462550000001523033710	SalmonellaBareilly_u			
100386462550000001523033710	SalmonellaBareilly_u			
100315192550000001523013408	Ecoli MG1655 Native			
100315192550000001523013408	Ecoli MG1655 Native			
100315192550000001523013408	Ecoli MG1655 Native			
100315192550000001523013408	Ecoli MG1655 Native			
100315192550000001523013408	Ecoli MG1655 Native			
100240842550000001523003004				
100258950960000001523004005	Test-129-0302-CCL			
100258950960000001523004005	Test-129-0302-CCL			
100298031700000001523008907	Test-129-0302-CCL			
100298031700000001523008907	Test-129-0302-CCL			
100298031700000001523008907	Test-129-0302-CCLStageMatch		2012-03-03T01:33:19-05	/opt/smrtar

Protocol Details For Job 1

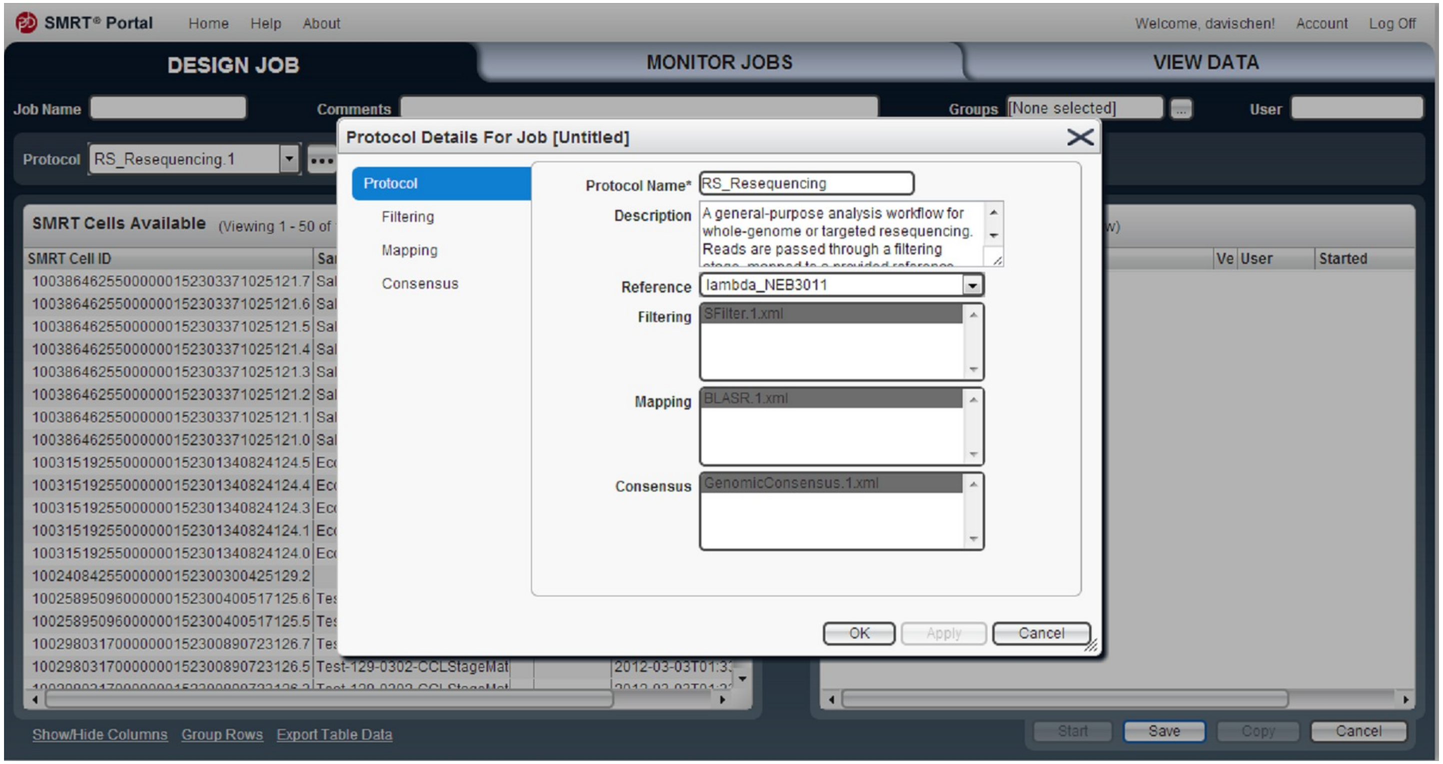
Protocol Name* RS_CeleraAssembler

Description: Performs de novo assembly using pacBioToCA and Celera Assembler 7.0. Supports error correction with reads from

Filtering: CeleraAssemblerSFilter.1.xml

Mapping: CeleraAssembler.1.xml

Resequencing



Job Name
Comments
Groups [None selected]
User

Protocol: RS_Resequencing.1

SMRT Cells Available (Viewing 1 - 50 of

SMRT Cell ID	Sa
10038646255000000152303371025121.7	Sal
10038646255000000152303371025121.6	Sal
10038646255000000152303371025121.5	Sal
10038646255000000152303371025121.4	Sal
10038646255000000152303371025121.3	Sal
10038646255000000152303371025121.2	Sal
10038646255000000152303371025121.1	Sal
10038646255000000152303371025121.0	Sal
10031519255000000152301340824124.5	Ec
10031519255000000152301340824124.4	Ec
10031519255000000152301340824124.3	Ec
10031519255000000152301340824124.1	Ec
10031519255000000152301340824124.0	Ec
10024084255000000152300300425129.2	
10025895096000000152300400517125.6	Tes
10025895096000000152300400517125.5	Tes
10029803170000000152300890723126.7	Tes
10029803170000000152300890723126.5	Test-129-0302-CCLStageMat

2012-03-03T01:3

Ve User Started

Protocol Details For Job [Untitled]

Protocol Name*: RS_Resequencing

Description: A general-purpose analysis workflow for whole-genome or targeted resequencing. Reads are passed through a filtering

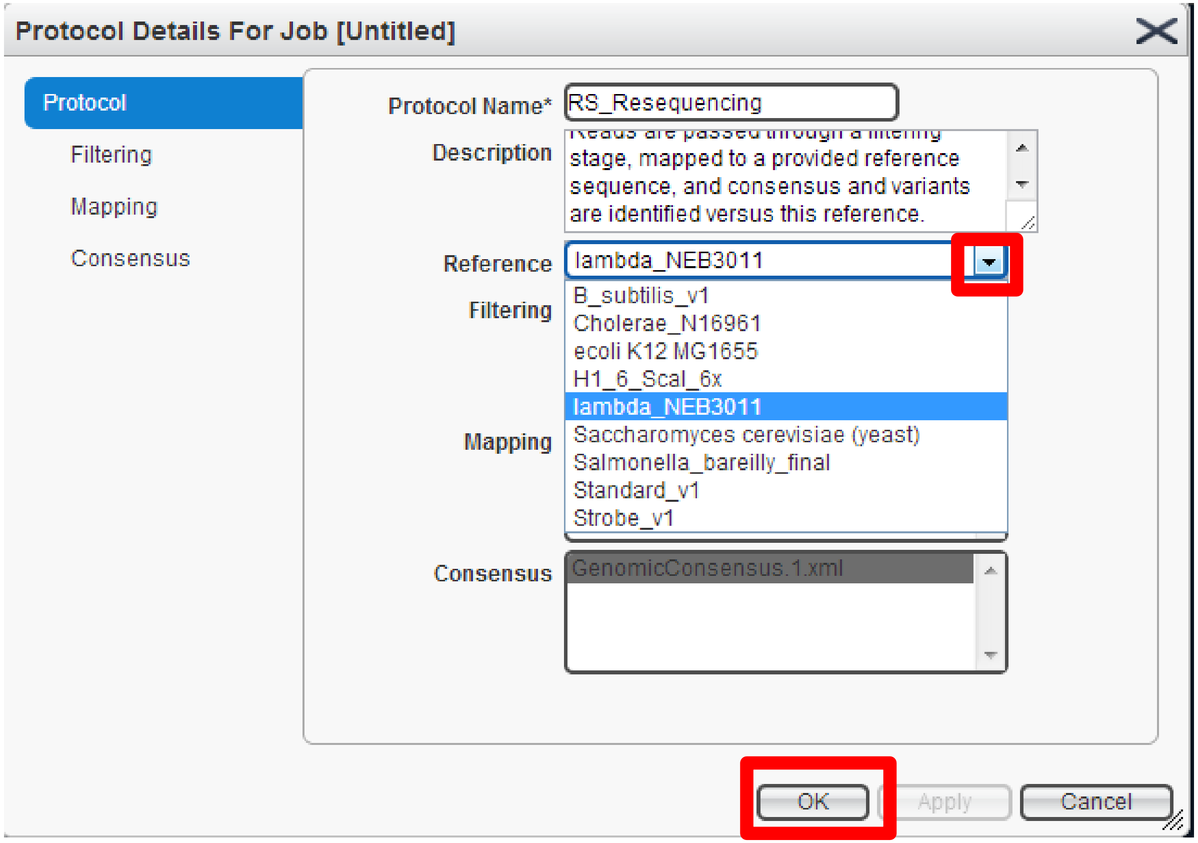
Reference: lambda_NEB3011

Filtering: SFilter.1.xml

Mapping: BLASR.1.xml

Consensus: GenomicConsensus.1.xml

Ref selection



Protocol Details For Job [Untitled]

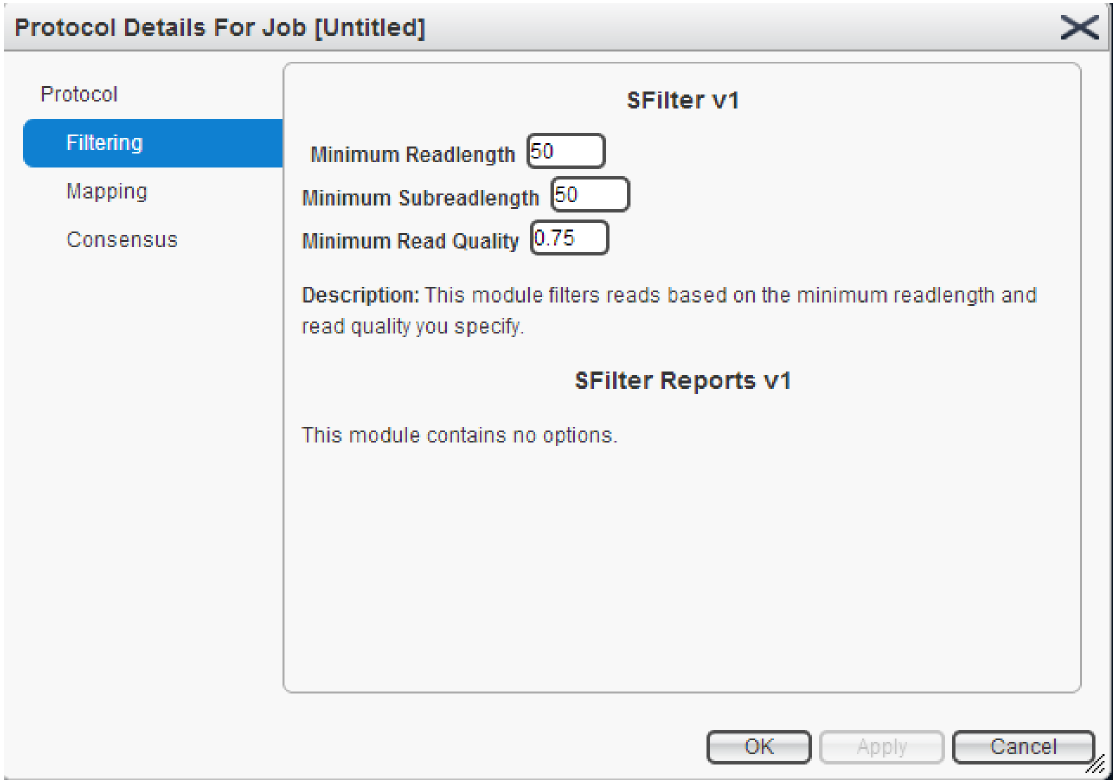
Protocol Name* RS_Resequencing

Description Reads are passed through a filtering stage, mapped to a provided reference sequence, and consensus and variants are identified versus this reference.

Reference lambda_NEB3011

Consensus GenomicConsensus.1.xml

Parameter



SFilter v1

Minimum Readlength: 50

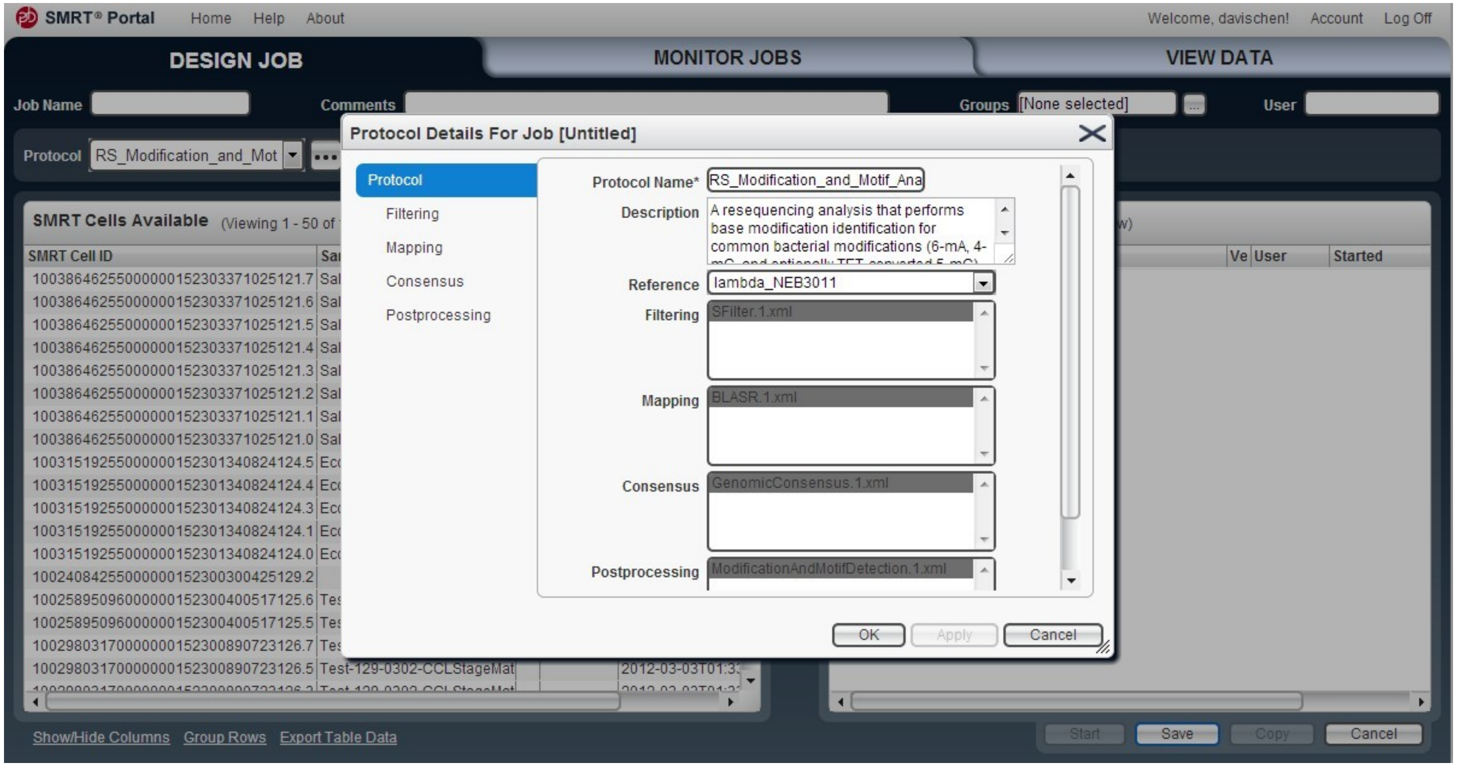
Minimum Subreadlength: 50

Minimum Read Quality: 0.75

Description: This module filters reads based on the minimum readlength and read quality you specify.

SFilter Reports v1

This module contains no options.



Job Name

Comments

Groups [None selected]

User

Protocol RS_Modification_and_Mot

Protocol Details For Job [Untitled]

Protocol Name* RS_Modification_and_Motif_Ana

Description A resequencing analysis that performs base modification identification for common bacterial modifications (6-mA, 4-mC, and optionally TET-converted 5-mC)

Reference lambda_NEB3011

Filtering SFilter.1.xml

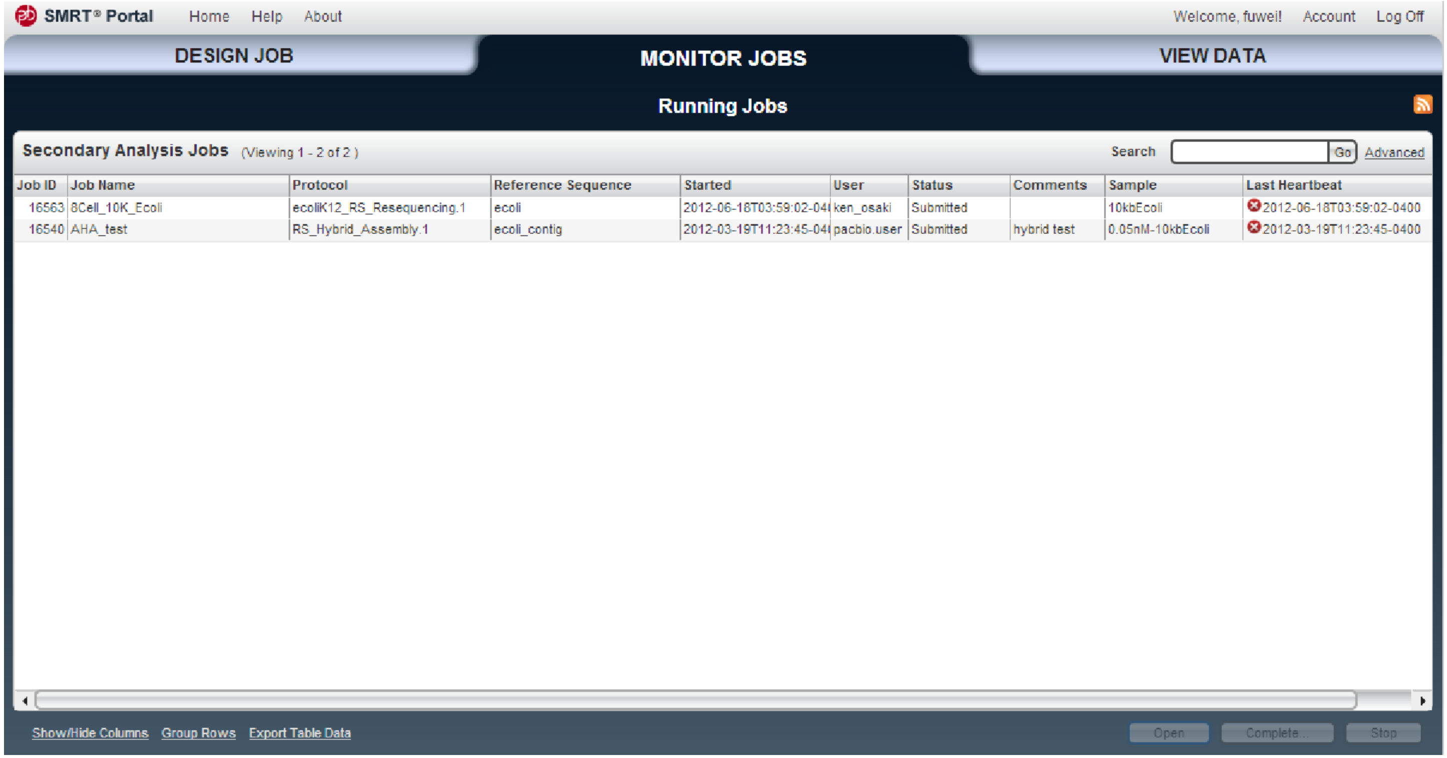
Mapping BLASR.1.xml

Consensus GenomicConsensus.1.xml

Postprocessing ModificationAndMotifDetection.1.xml

SMRT Cells Available (Viewing 1 - 50 of

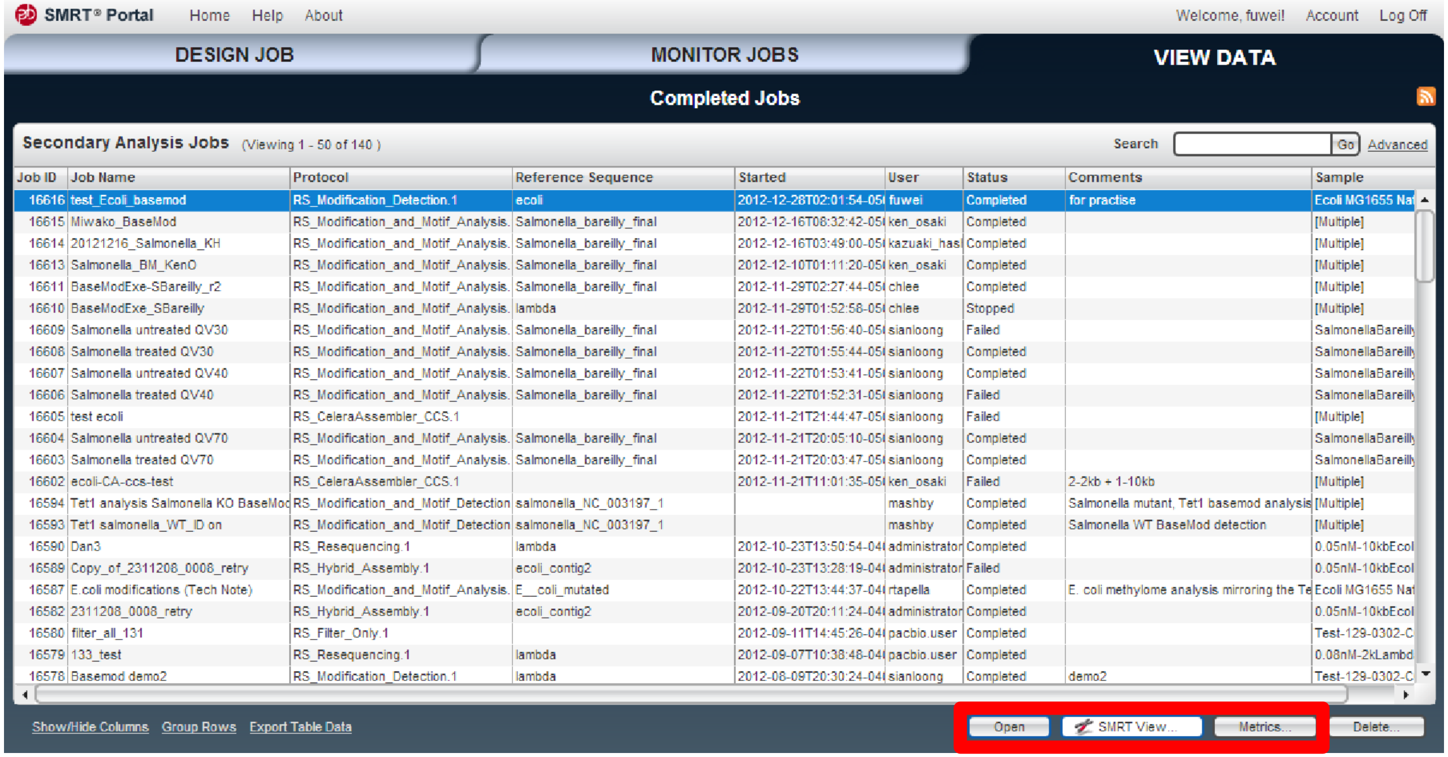
SMRT Cell ID	Sa	
100386462550000001523033710251 21.7	Sal	
10038646255000000152303371025121.6	Sal	
10038646255000000152303371025121.5	Sal	
10038646255000000152303371025121.4	Sal	
10038646255000000152303371025121.3	Sal	
10038646255000000152303371025121.2	Sal	
10038646255000000152303371025121.1	Sal	
10038646255000000152303371025121.0	Sal	
10031519255000000152301340824124.5	Ec	
10031519255000000152301340824124.4	Ec	
10031519255000000152301340824124.3	Ec	
10031519255000000152301340824124.1	Ec	
10031519255000000152301340824124.0	Ec	
10024084255000000152300300425129.2		
10025895096000000152300400517125.6	Tes	
10025895096000000152300400517125.5	Tes	
10029803170000000152300890723126.7	Tes	
10029803170000000152300890723126.5	Test-129-0302-CCLStageMat	2012-03-03T01:3



Running Jobs

Secondary Analysis Jobs (Viewing 1 - 2 of 2)

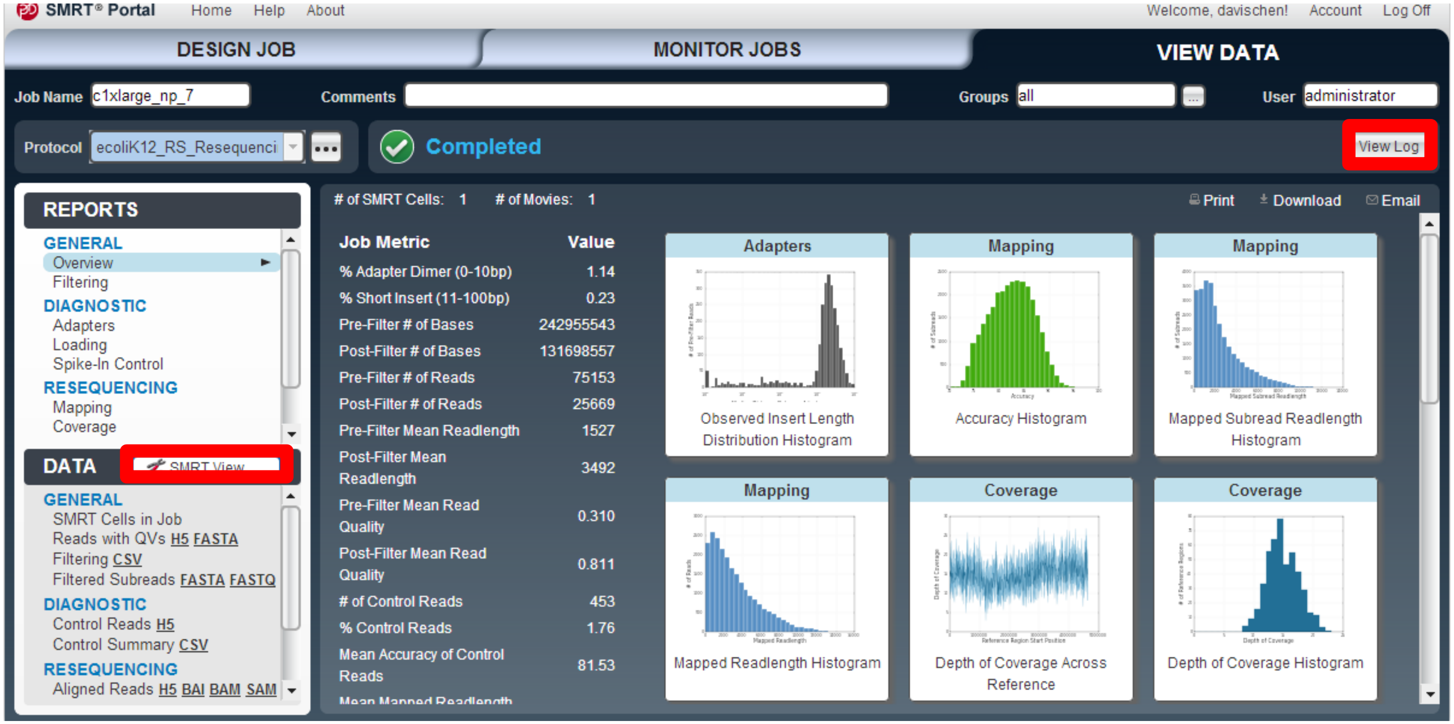
Job ID	Job Name	Protocol	Reference Sequence	Started	User	Status	Comments	Sample	Last Heartbeat
16563	8Cell_10K_Ecoli	ecoliK12_RS_Resequencing.1	ecoli	2012-06-18T03:59:02-04	ken_osaki	Submitted		10kbEcoli	⊗2012-06-18T03:59:02-0400
16540	AHA_test	RS_Hybrid_Assembly.1	ecoli_contig	2012-03-19T11:23:45-04	pacbio.user	Submitted	hybrid test	0.05nM-10kbEcoli	⊗2012-03-19T11:23:45-0400



Completed Jobs

Secondary Analysis Jobs (Viewing 1 - 50 of 140)

Job ID	Job Name	Protocol	Reference Sequence	Started	User	Status	Comments	Sample
16616	test_Ecoli_basemod	RS_Modification_Detection.1	ecoli	2012-12-28T02:01:54-05(	fuwei	Completed	for practise	Ecoli MG1655 Na(
16615	Miwako_BaseMod	RS_Modification_and_Motif_Analysis.	Salmonella_bareilly_final	2012-12-16T08:32:42-05(	ken_osaki	Completed		[Multiple]
16614	20121216_Salmonella_KH	RS_Modification_and_Motif_Analysis.	Salmonella_bareilly_final	2012-12-16T03:49:00-05(	kazuaki_has|	Completed		[Multiple]
16613	Salmonella_BM_KenO	RS_Modification_and_Motif_Analysis.	Salmonella_bareilly_final	2012-12-10T01:11:20-05(	ken_osaki	Completed		[Multiple]
16611	BaseModExe-SBareilly_r2	RS_Modification_and_Motif_Analysis.	Salmonella_bareilly_final	2012-11-29T02:27:44-05(	chlee	Completed		[Multiple]
16610	BaseModExe_SBareilly	RS_Modification_and_Motif_Analysis.	lambda	2012-11-29T01:52:58-05(	chlee	Stopped		[Multiple]
16609	Salmonella untreated QV30	RS_Modification_and_Motif_Analysis.	Salmonella_bareilly_final	2012-11-22T01:56:40-05(	sianloong	Failed		SalmonellaBareilly
16608	Salmonella treated QV30	RS_Modification_and_Motif_Analysis.	Salmonella_bareilly_final	2012-11-22T01:55:44-05(	sianloong	Completed		SalmonellaBareilly
16607	Salmonella untreated QV40	RS_Modification_and_Motif_Analysis.	Salmonella_bareilly_final	2012-11-22T01:53:41-05(	sianloong	Completed		SalmonellaBareilly
16606	Salmonella treated QV40	RS_Modification_and_Motif_Analysis.	Salmonella_bareilly_final	2012-11-22T01:52:31-05(	sianloong	Failed		SalmonellaBareilly
16605	test ecoli	RS_CeleraAssembler_CCS.1		2012-11-21T21:44:47-05(	sianloong	Failed		[Multiple]
16604	Salmonella untreated QV70	RS_Modification_and_Motif_Analysis.	Salmonella_bareilly_final	2012-11-21T20:05:10-05(	sianloong	Completed		SalmonellaBareilly
16603	Salmonella treated QV70	RS_Modification_and_Motif_Analysis.	Salmonella_bareilly_final	2012-11-21T20:03:47-05(	sianloong	Completed		SalmonellaBareilly
16602	ecoli-CA-ccs-test	RS_CeleraAssembler_CCS.1		2012-11-21T11:01:35-05(	ken_osaki	Failed	2-2kb + 1-10kb	[Multiple]
16594	Tet1 analysis Salmonella KO BaseMod	RS_Modification_and_Motif_Detection	salmonella_NC_003197_1		mashby	Completed	Salmonella mutant, Tet1 basemod analysis	[Multiple]
16593	Tet1 salmonella_WT_ID on	RS_Modification_and_Motif_Detection	salmonella_NC_003197_1		mashby	Completed	Salmonella WT BaseMod detection	[Multiple]
16590	Dan3	RS_Resequencing.1	lambda	2012-10-23T13:50:54-04(	administrator	Completed		0.05nM-10kbEcol
16589	Copy_of_2311208_0008_retry	RS_Hybrid_Assembly.1	ecoli_contig2	2012-10-23T13:28:19-04(	administrator	Failed		0.05nM-10kbEcol
16587	E.coli modifications (Tech Note)	RS_Modification_and_Motif_Analysis.	E__coli_mutated	2012-10-22T13:44:37-04(	rtapella	Completed	E. coli methylome analysis mirroring the Te	Ecoli MG1655 Na(
16582	2311208_0008_retry	RS_Hybrid_Assembly.1	ecoli_contig2	2012-09-20T20:11:24-04(	administrator	Completed		0.05nM-10kbEcol
16580	filter_all_131	RS_Filter_Only.1		2012-09-11T14:45:26-04(	pacbio.user	Completed		Test-129-0302-C
16579	133_test	RS_Resequencing.1	lambda	2012-09-07T10:38:48-04(	pacbio.user	Completed		0.08nM-2kLambd
16578	Basemod demo2	RS_Modification_Detection.1	lambda	2012-08-09T20:30:24-04(	sianloong	Completed	demo2	Test-129-0302-C



Job Name c1xlarge_np_7

Comments

Groups all

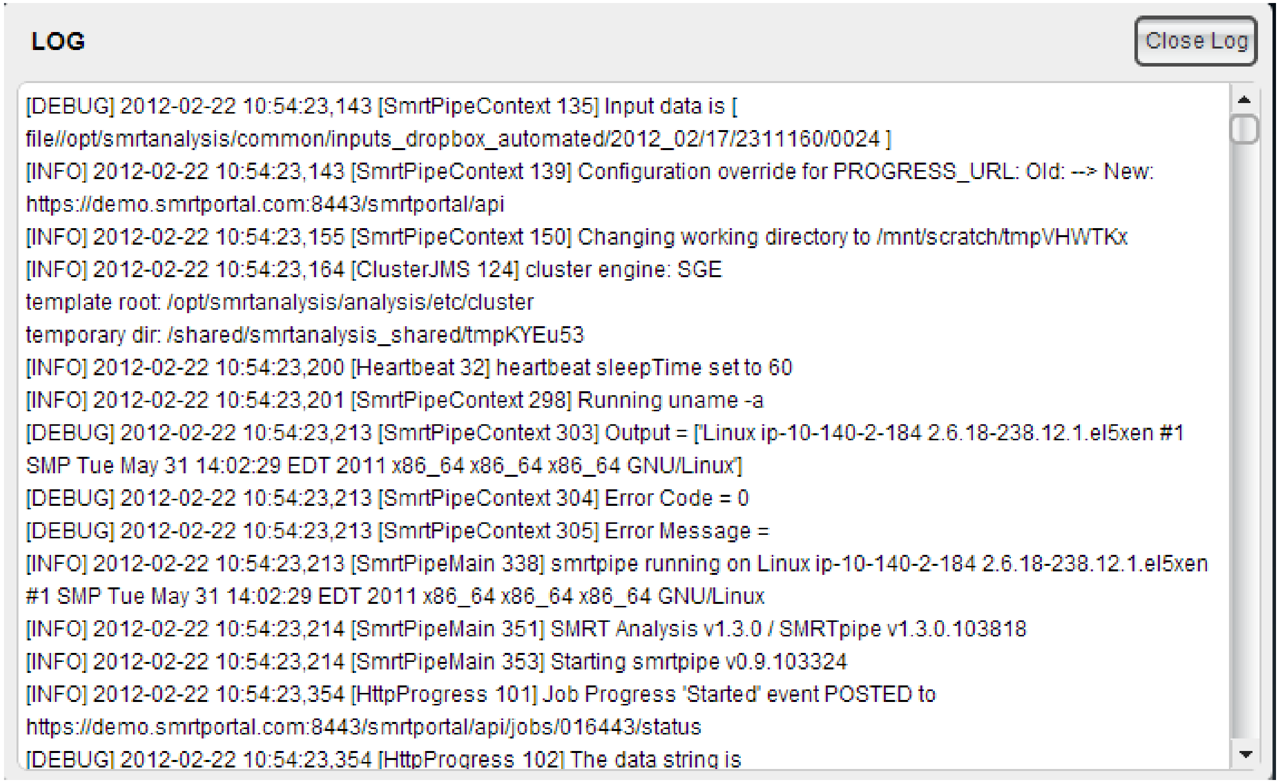
User administrator

Protocol ecoliK12_RS_Resequenci

Completed

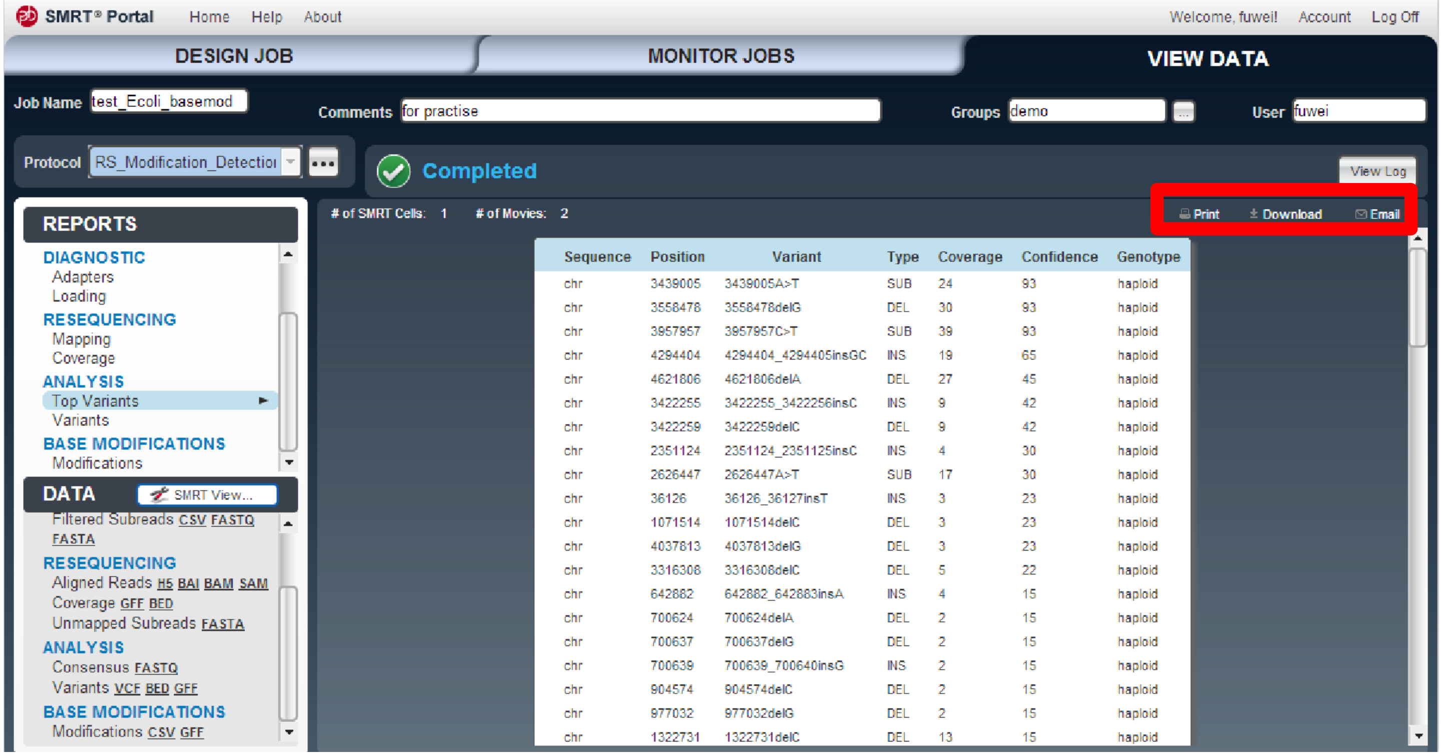
# of SMRT Cells: 1 # of Movies: 1

Job Metric	Value
% Adapter Dimer (0-10bp)	1.14
% Short Insert (11-100bp)	0.23
Pre-Filter # of Bases	242955543
Post-Filter # of Bases	131698557
Pre-Filter # of Reads	75153
Post-Filter # of Reads	25669
Pre-Filter Mean Readlength	1527
Post-Filter Mean Readlength	3492
Pre-Filter Mean Read Quality	0.310
Post-Filter Mean Read Quality	0.811
# of Control Reads	453
% Control Reads	1.76
Mean Accuracy of Control Reads	81.53
Mean Mapped Readlength	



```
[DEBUG] 2012-02-22 10:54:23,143 [SmrtPipeContext 135] Input data is [
file//opt/smrtanalysis/common/inputs_dropbox_automated/2012_02/17/2311160/0024 ]
[INFO] 2012-02-22 10:54:23,143 [SmrtPipeContext 139] Configuration override for PROGRESS_URL: Old: --> New:
https://demo.smrtportal.com:8443/smrtportal/api
[INFO] 2012-02-22 10:54:23,155 [SmrtPipeContext 150] Changing working directory to /mnt/scratch/tmpVHWTKx
[INFO] 2012-02-22 10:54:23,164 [ClusterJMS 124] cluster engine: SGE
template root: /opt/smrtanalysis/analysis/etc/cluster
temporary dir: /shared/smrtanalysis_shared/tmpKYEu53
[INFO] 2012-02-22 10:54:23,200 [Heartbeat 32] heartbeat sleepTime set to 60
[INFO] 2012-02-22 10:54:23,201 [SmrtPipeContext 298] Running uname -a
[DEBUG] 2012-02-22 10:54:23,213 [SmrtPipeContext 303] Output = ['Linux ip-10-140-2-184 2.6.18-238.12.1.el5xen #1
SMP Tue May 31 14:02:29 EDT 2011 x86_64 x86_64 x86_64 GNU/Linux']
[DEBUG] 2012-02-22 10:54:23,213 [SmrtPipeContext 304] Error Code = 0
[DEBUG] 2012-02-22 10:54:23,213 [SmrtPipeContext 305] Error Message =
[INFO] 2012-02-22 10:54:23,213 [SmrtPipeMain 338] smrtpipe running on Linux ip-10-140-2-184 2.6.18-238.12.1.el5xen
#1 SMP Tue May 31 14:02:29 EDT 2011 x86_64 x86_64 x86_64 GNU/Linux
[INFO] 2012-02-22 10:54:23,214 [SmrtPipeMain 351] SMRT Analysis v1.3.0 / SMRTpipe v1.3.0.103818
[INFO] 2012-02-22 10:54:23,214 [SmrtPipeMain 353] Starting smrtpipe v0.9.103324
[INFO] 2012-02-22 10:54:23,354 [HttpProgress 101] Job Progress 'Started' event POSTED to
https://demo.smrtportal.com:8443/smrtportal/api/jobs/016443/status
[DEBUG] 2012-02-22 10:54:23,354 [HttpProgress 102] The data string is
```

SMRT Portal



Job Name test_Ecoli_basemod

Comments for practise

Groups demo

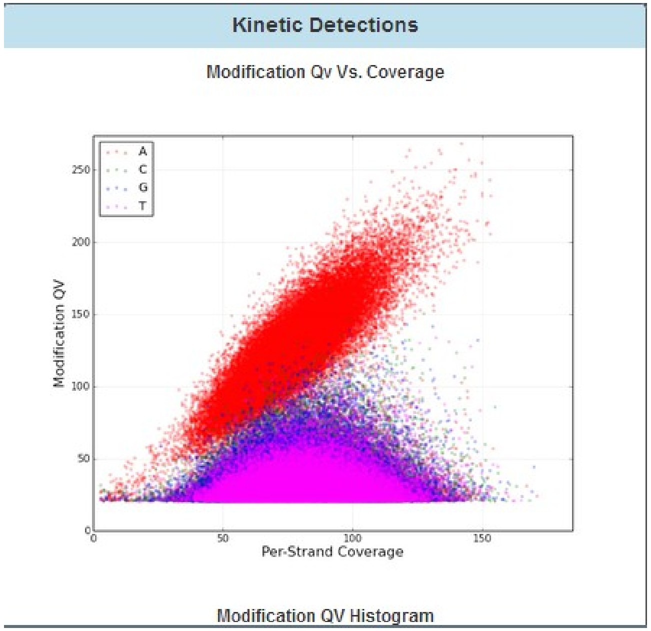
User fuwei

Protocol RS_Modification_Detectioı

Completed

# of SMRT Cells: 1 # of Movies: 2

Sequence	Position	Variant	Type	Coverage	Confidence	Genotype
chr	3439005	3439005A>T	SUB	24	93	haploid
chr	3558478	3558478delG	DEL	30	93	haploid
chr	3957957	3957957C>T	SUB	39	93	haploid
chr	4294404	4294404_4294405insGC	INS	19	65	haploid
chr	4621806	4621806delA	DEL	27	45	haploid
chr	3422255	3422255_3422256insC	INS	9	42	haploid
chr	3422259	3422259delC	DEL	9	42	haploid
chr	2351124	2351124_2351125insC	INS	4	30	haploid
chr	2626447	2626447A>T	SUB	17	30	haploid
chr	36126	36126_36127insT	INS	3	23	haploid
chr	1071514	1071514delC	DEL	3	23	haploid
chr	4037813	4037813delG	DEL	3	23	haploid
chr	3316308	3316308delC	DEL	5	22	haploid
chr	642882	642882_642883insA	INS	4	15	haploid
chr	700624	700624delA	DEL	2	15	haploid
chr	700637	700637delG	DEL	2	15	haploid
chr	700639	700639_700640insG	INS	2	15	haploid
chr	904574	904574delC	DEL	2	15	haploid
chr	977032	977032delG	DEL	2	15	haploid
chr	1322731	1322731delC	DEL	13	15	haploid

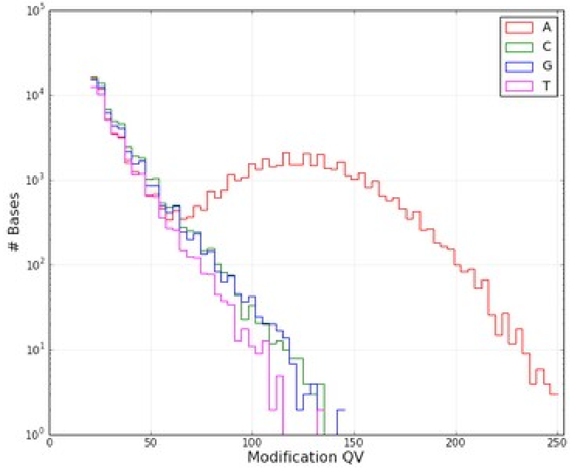


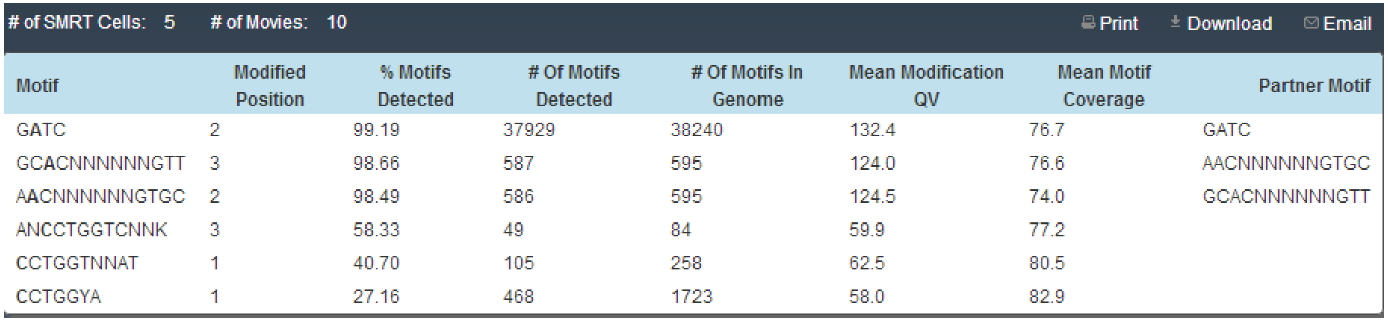
Kinetic Detections

Modification Qv Vs. Coverage

Modification QV Histogram

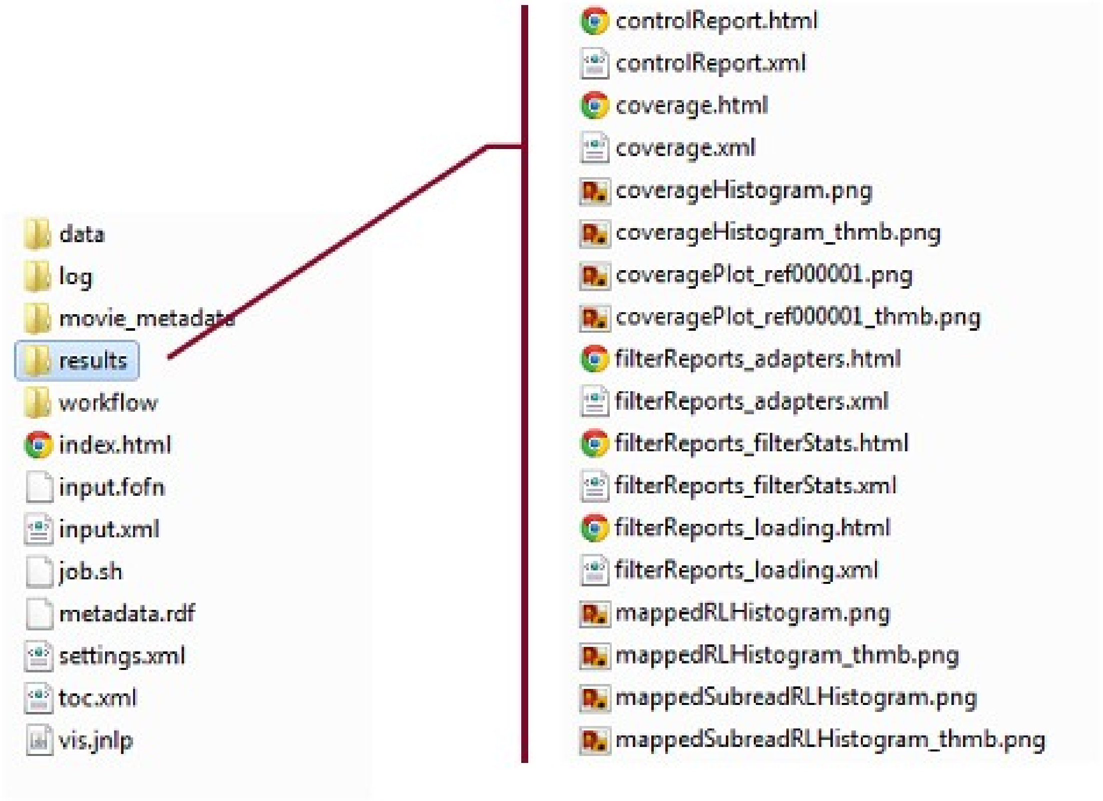
Modification QV Histogram





Motif	Modified Position	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
GATC	2	99.19	37929	38240	132.4	76.7	GATC
GCACNNNNNNGTT	3	98.66	587	595	124.0	76.6	AACNNNNNNGTGC
AACNNNNNNGTGC	2	98.49	586	595	124.5	74.0	GCACNNNNNNGTT
ANCCTGGTCNNK	3	58.33	49	84	59.9	77.2	
CCTGGTNNAT	1	40.70	105	258	62.5	80.5	
CCTGGYA	1	27.16	468	1723	58.0	82.9	

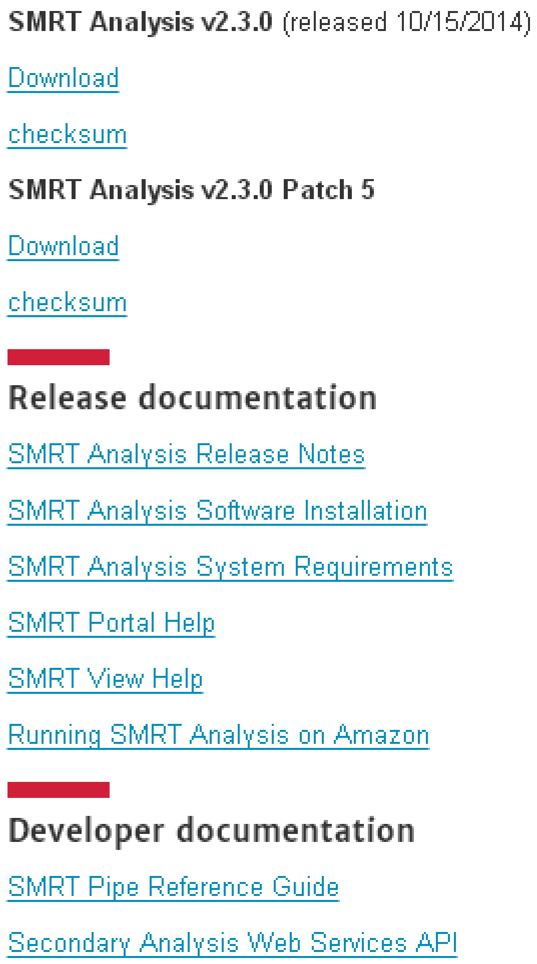
Results Directory



- data
- log
- movie_metadata
- results
- workflow
- index.html
- input.fofn
- input.xml
- job.sh
- metadata.rdf
- settings.xml
- toc.xml
- vis.jnlp

- controlReport.html
- controlReport.xml
- coverage.html
- coverage.xml
- coverageHistogram.png
- coverageHistogram_thmb.png
- coveragePlot_ref000001.png
- coveragePlot_ref000001_thmb.png
- filterReports_adapters.html
- filterReports_adapters.xml
- filterReports_filterStats.html
- filterReports_filterStats.xml
- filterReports_loading.html
- filterReports_loading.xml
- mappedRLHistogram.png
- mappedRLHistogram_thmb.png
- mappedSubreadRLHistogram.png
- mappedSubreadRLHistogram_thmb.png

http://www.pacb.com/support/software-downloads/



SMRT Analysis v2.3.0 (released 10/15/2014)

Download

checksum

SMRT Analysis v2.3.0 Patch 5

Download

checksum

Release documentation

SMRT Analysis Release Notes

SMRT Analysis Software Installation

SMRT Analysis System Requirements

SMRT Portal Help

SMRT View Help

Running SMRT Analysis on Amazon

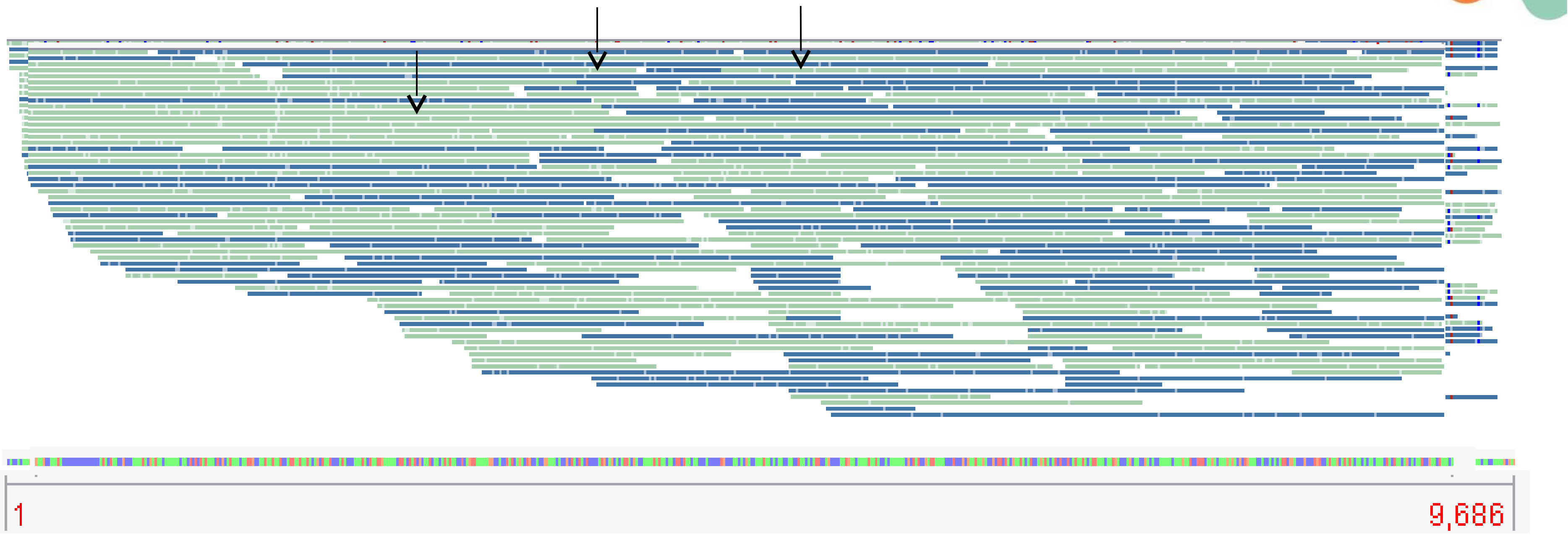
Developer documentation

SMRT Pipe Reference Guide

Secondary Analysis Web Services API

smrtanalysis 实例

Hierarchical Genome Assembly Process (HGAp)



1． 选定长度作为骨架
2． 比对其他 reads
3． 矫正错误
4． 构建准确的 consensus

- Utilizes every bit of data:
  - Long reads for continuity
  - Shorter reads for improving accuracy
- Accuracy: 85.7% 🖧 99.3%, 9089 bp
- Chimera / low quality regions can be filtered out early
- Accurate long consensus reads easier to assemble

SMRTPIPE

Falcon 实例

DBG2OLC 实例

从源代码编译，使用下面的命令：

```
g++ -O3 -o SparseAssebmler *.cpp
g++ -O3 -o DBG2OLC *.cpp
g++ -O3 -o Sparc *.cpp
```

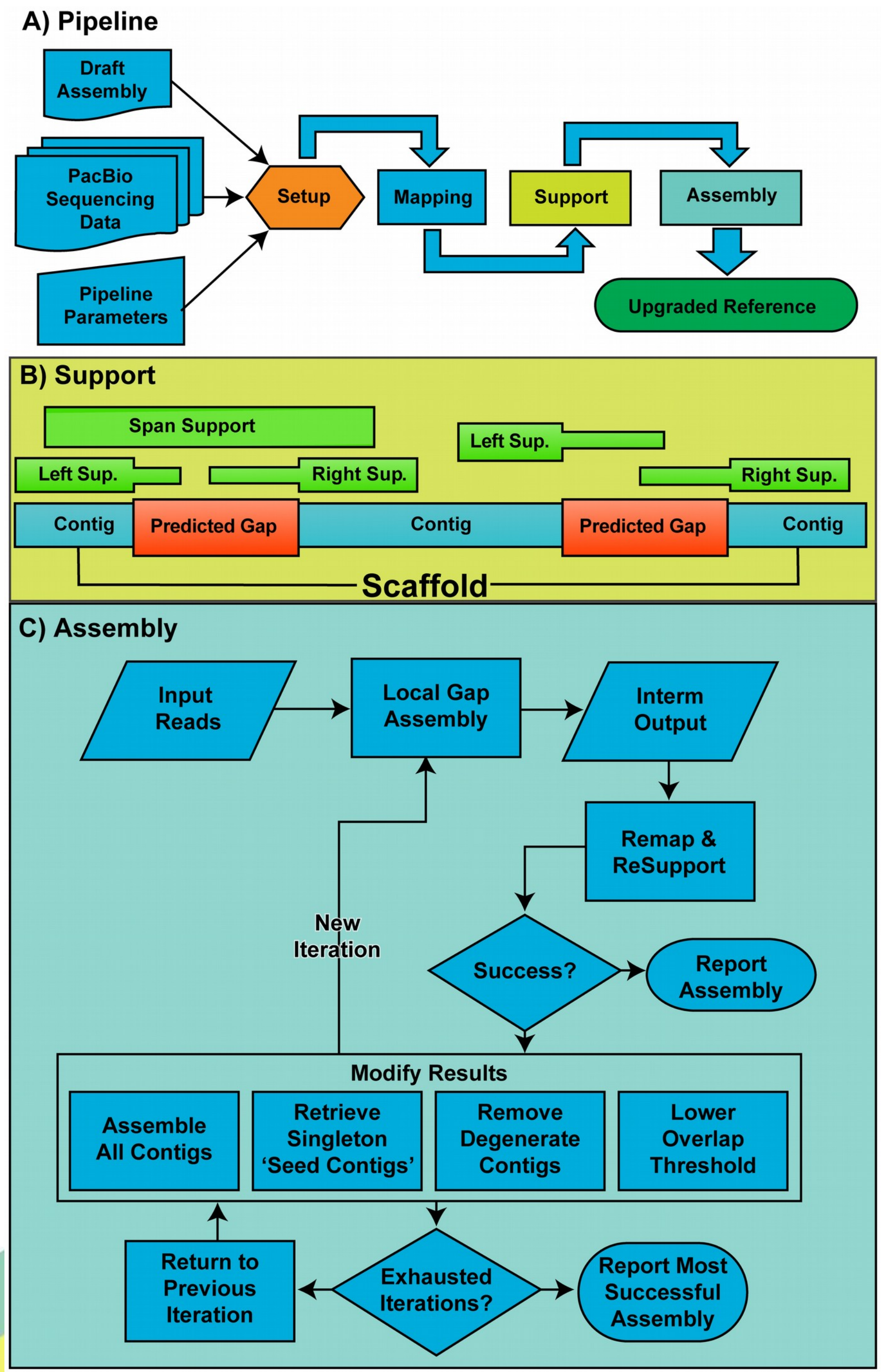
选择部分数据

```
./SelectLongestReads sum 600000000 longest 0 o Illumina_50x.fastq f Illumina_500bp_2x300_R1.fastq
./SelectLongestReads sum 260000000 longest 0 o Pacbio_20x.fasta f Pacbio.fasta
```

smrtdenovo 实例

```
1.    awk 'NR%4==1||NR%4==2' selfSampleData/pacbio_filtered.fastq | sed 's/^@/>/g' > reads.fa
2.    smartdenovo/smartdenovo.pl reads.fa > wtasm.mak
3.    make -f wtasm.mak
```

PBJelly 实例



对于 de novo 组装，长度超过读长的重复序列会产生缺口，导致片段化的组装。很难检测重复区域的变异，而这些对了解某些疾病可能很重要。

PBjelly 能够将 PacBio 长读长序列与组装的草图比对，进行 gapfilling。研究人员将这种方法应用在四个基因组上，解决了 63%-99% 的 gap

数据： /home/tbc/jellyExample
程序： source /opt/PBSuite_14.1.15/setup.sh

1) Create your Protocol.xml

2) Run each stage

```
Jelly.py <stage> yourProtocol.xml
```

The stages, in order, and their descriptions are

1. setup Tag sequence names, find gaps, and index the reference
2. mapping Use blasr to map the sequences to the reference
3. support Identify which reads support which gaps
4. extraction For each gap, consolidate all reads supporting it into a local-assembly folder.
5. assembly Build the consensus gap-filling sequence
6. output Stitch the reference sequences and gap-fillling sequences together.

3) Passing Parameters through Jelly.py

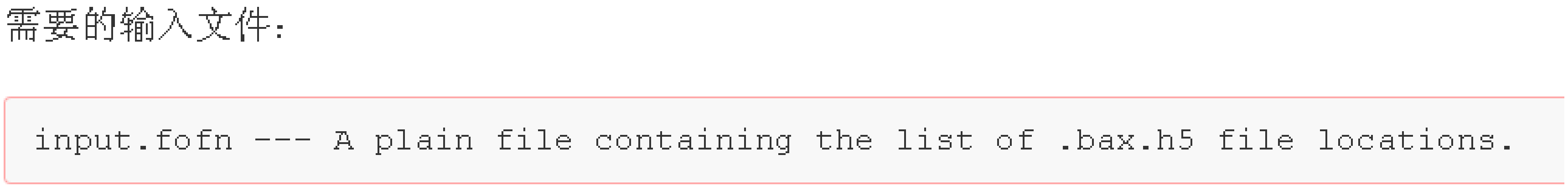
If you would like to pass a parameter to the stage you are running, use "-x". For example, when running the support stage, if you only wanted Jelly to attempt to fill captured-gaps (i.e. no inter-scaffold gaps), and you wanted to require that a read must have a minimum mapping QV of >= 250 to support a gap, you'd use the command:

```
> Jelly.py support Protocol.xml -x "--capturedOnly --minMapqv=250"
```

All parameters you add need to be enclosed in double quotes after the -x

Isoseq 分析

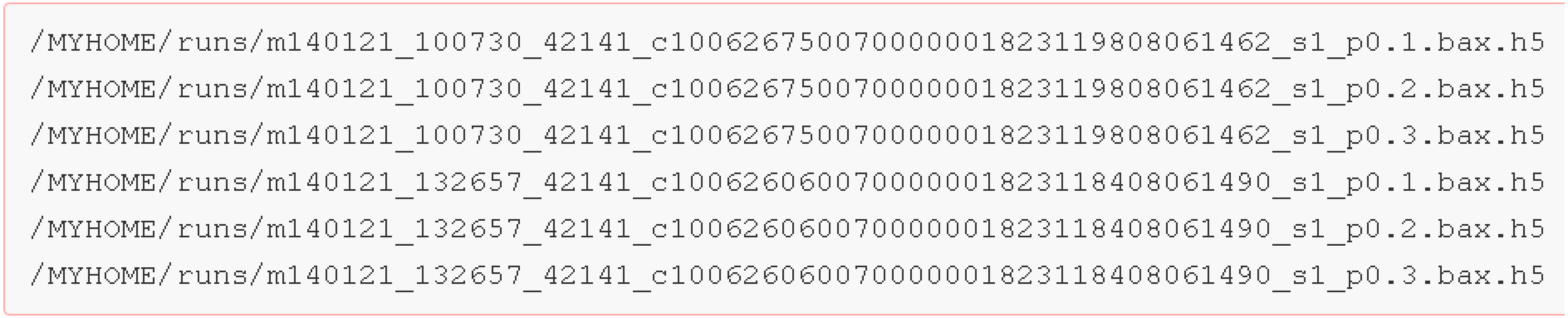
命令行运行 ReadsOfInser 的命令是 ConsensusTools . 可以到 smrt 安装目录 <smrtanalysis_directory>/doc/bioinformatics-tools/ConsensusTools/doc/index.html 下面查看相关文档。



需要的输入文件:

```
input.fofn --- A plain file containing the list of .bax.h5 file locations.
```

下面是一个 input.fofn 文件的例子 . 一共有两个 movie ，每个包含三个 .bax.h5 文件



```
/MYHOME/runs/m140121_100730_42141_c100626750070000001823119808061462_s1_p0.1.bax.h5
/MYHOME/runs/m140121_100730_42141_c100626750070000001823119808061462_s1_p0.2.bax.h5
/MYHOME/runs/m140121_100730_42141_c100626750070000001823119808061462_s1_p0.3.bax.h5
/MYHOME/runs/m140121_132657_42141_c100626060070000001823118408061490_s1_p0.1.bax.h5
/MYHOME/runs/m140121_132657_42141_c100626060070000001823118408061490_s1_p0.2.bax.h5
/MYHOME/runs/m140121_132657_42141_c100626060070000001823118408061490_s1_p0.3.bax.h5
```

示例命令：

```
ConsensusTools.sh CircularConsensus  --minFullPasses 0  --minPredictedAccuracy 75 \
   --parameters <smrtanalysis_directory>/analysis/etc/algorithm_parameters/2014-03 \
   --numThreads 24 --fofn /MYHOME/test_dir/input.fofn \
   -o /MYHOME/test_dir/data
```