PACBIO®

SMRT Link and Analysis Tools for PacBio Data

SMRT Informatics Developers Conference - January 17, 2018

AGENDA

- 5.1.0 release goals
- New features
  - Support for key analysis applications
    - Multiplexed microbial assembly
    - Structural variant calling
    - De novo assembly
  - Barcoding workflow redesign
  - Sample Setup redesign
  - Data Management and SMRT Analysis usability improvements
  - Release of Iso-Seq 2 [Beta]
  - Minor Variants support for custom target configs

Release Goals

HIGH-LEVEL GOALS

- Increased system throughput
  - Support for longer movies
  - Chemistry improvements
- End to end support for high-priority applications
- Focus on usability aspects of SMRT Link
- Preliminary R&D work on
  - Building out No-Amp targeted pulldown (Cas9) solution
  - Diploid consensus

SMRT TOOLS - NEW FEATURES

Applications

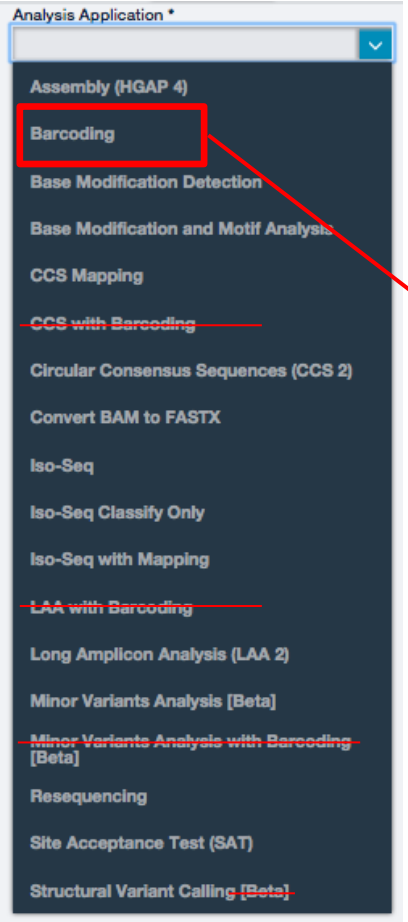
- Multiplexed Microbial Assembly
  - Completely redesigned barcode workflow
  - New barcode calling software
- Structural Variation
  - Support for multi-sample (joint) calling
  - Moved out of Beta status
- De Novo Assembly
  - Support for running unzip after HGAP.4
  - Support for generating GFA output
  - Binary release now available
- Iso-Seq
  - Address scalability issues
- Minor Variants
  - Support for adding custom target configurations (gene annotations)

SMRT LINK – NEW FEATURES

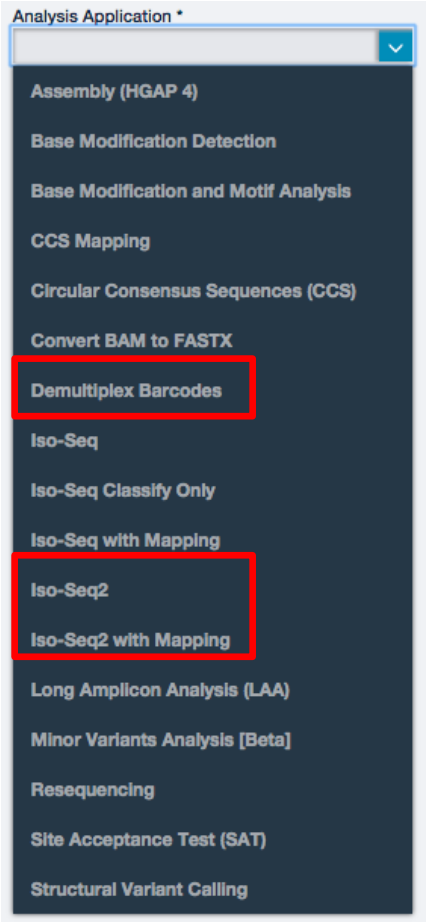
SMRT Link

- Barcoding workflow redesign
- Redesign of Sample Setup
- Usability improvements
  - Data Management
  - SMRT Analysis
- Analysis import/export

5.0.0



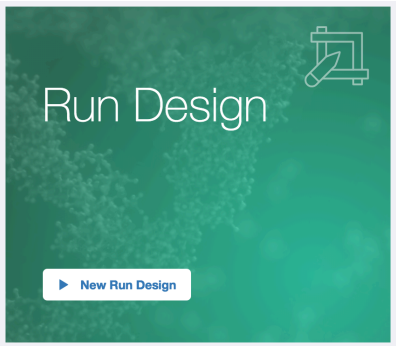
5.1.0



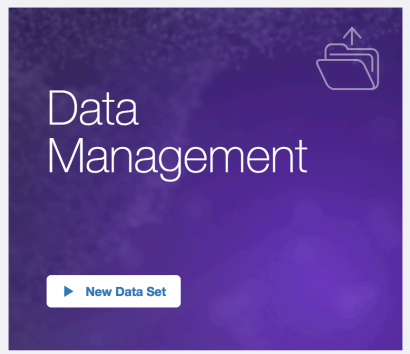
PACBIO®

Multiplexed Microbial Assembly

REDESIGN OF BARCODING WORKFLOW



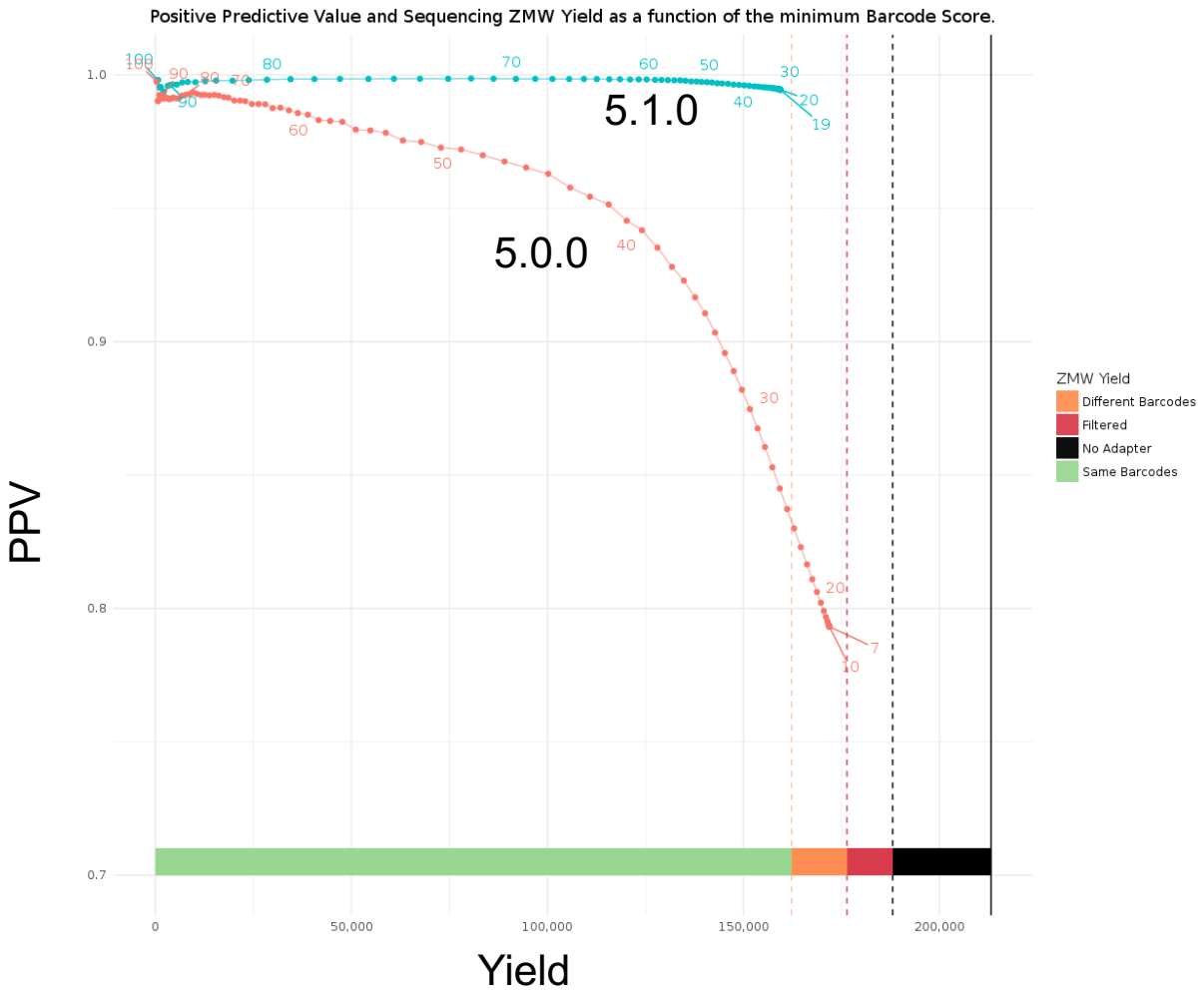
- Specify sample names of individual barcodes



- Better assignment of reads to barcodes
- New, more efficient demultiplexing algorithm
- Extensive new QC metrics

- Flexibility to launch analyses on a single barcode or as a batch with custom parameters

DE-MULTIPLEXING PERFORMANCE – 5.0.0 VS 5.1.0



RUN DESIGN: NEW SECTION FOR BARCODES

SMRT Link

Not Secure | https://smrtlink-bihourly.nanofluidics.com:8243/sl/#/run-design/new

PACBIO

Run Design

mbudagyan

Admin

New Run Design

DELETE | CANCEL | VIEW SUMMARY | SAVE

Run Information

Run Name: Run 11.10.2017 17:17

Run Comments

Experiment Name

Experiment Id

Estimated Run Duration (mins): 540

Run Reagents / Consumables

1 SMRT cell
1 sequencing reagent plate
1 OS enzyme tube
1 mineral oil tube
3 boxes of tips
1 mixing plate
1 sample plate

Sample Information

Binding Kit

Sequencing Kit

DNA Control Complex

Insert Size (bp): 10000

Movie Time per SMRT Cell (mins): 360

▸ Advanced Options

▾ Barcoded Sample Options

Sample is Barcoded: YES | NO

Barcode Set: Sequel_RSII_16_barcodes_v1

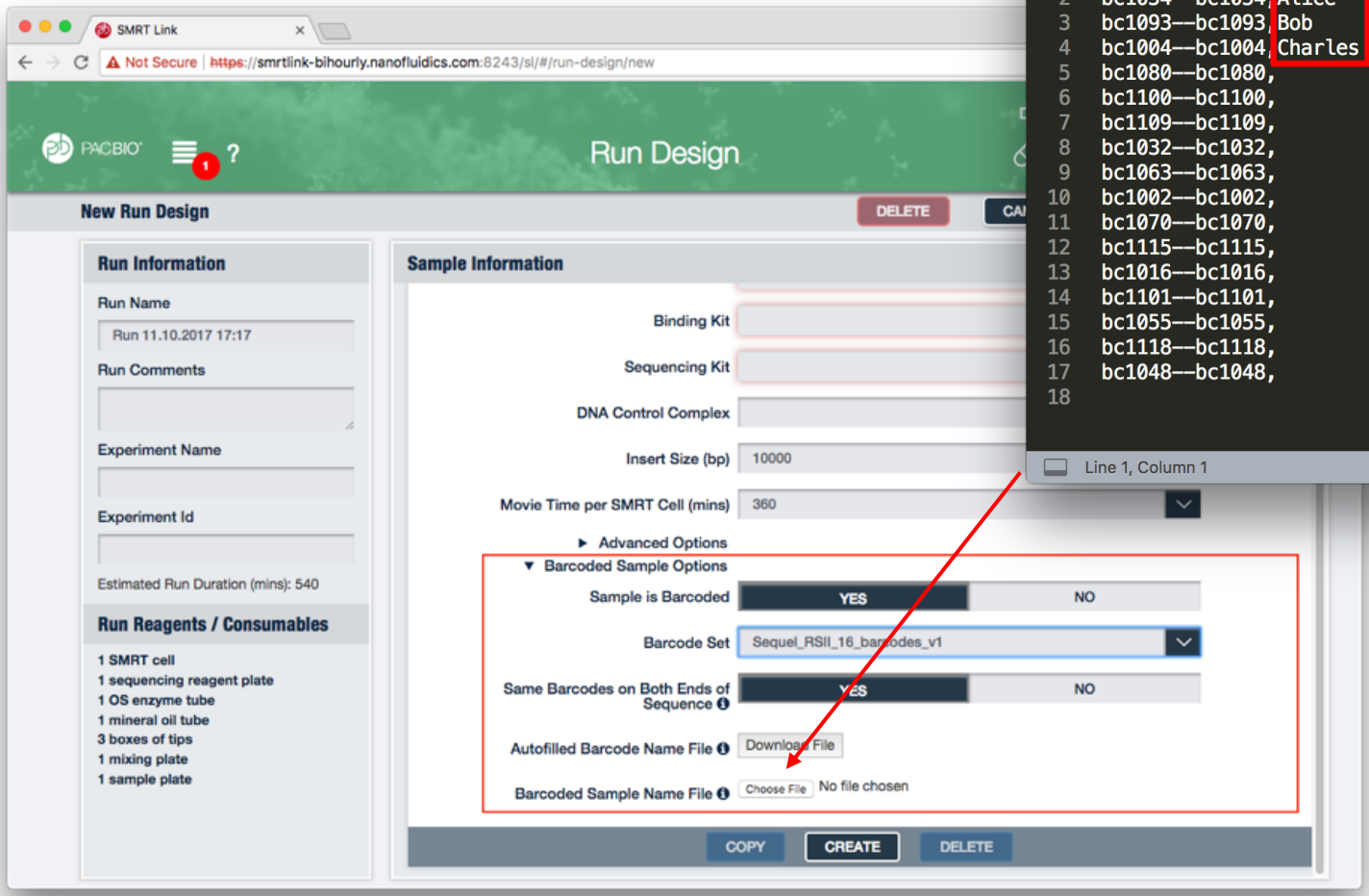
Same Barcodes on Both Ends of Sequence: YES | NO

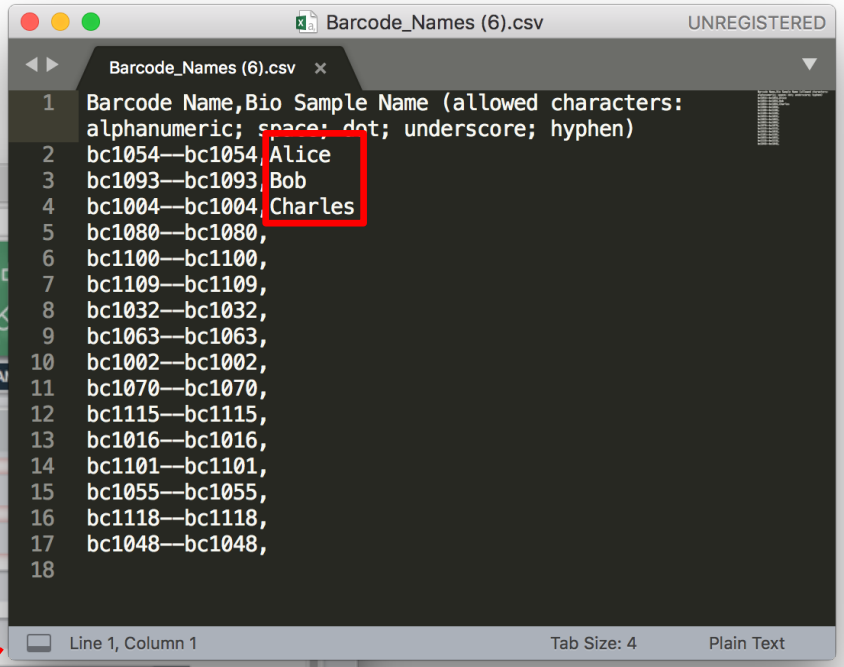
Autofilled Barcode Name File: Download File

Barcoded Sample Name File: Choose File No file chosen

COPY | CREATE | DELETE

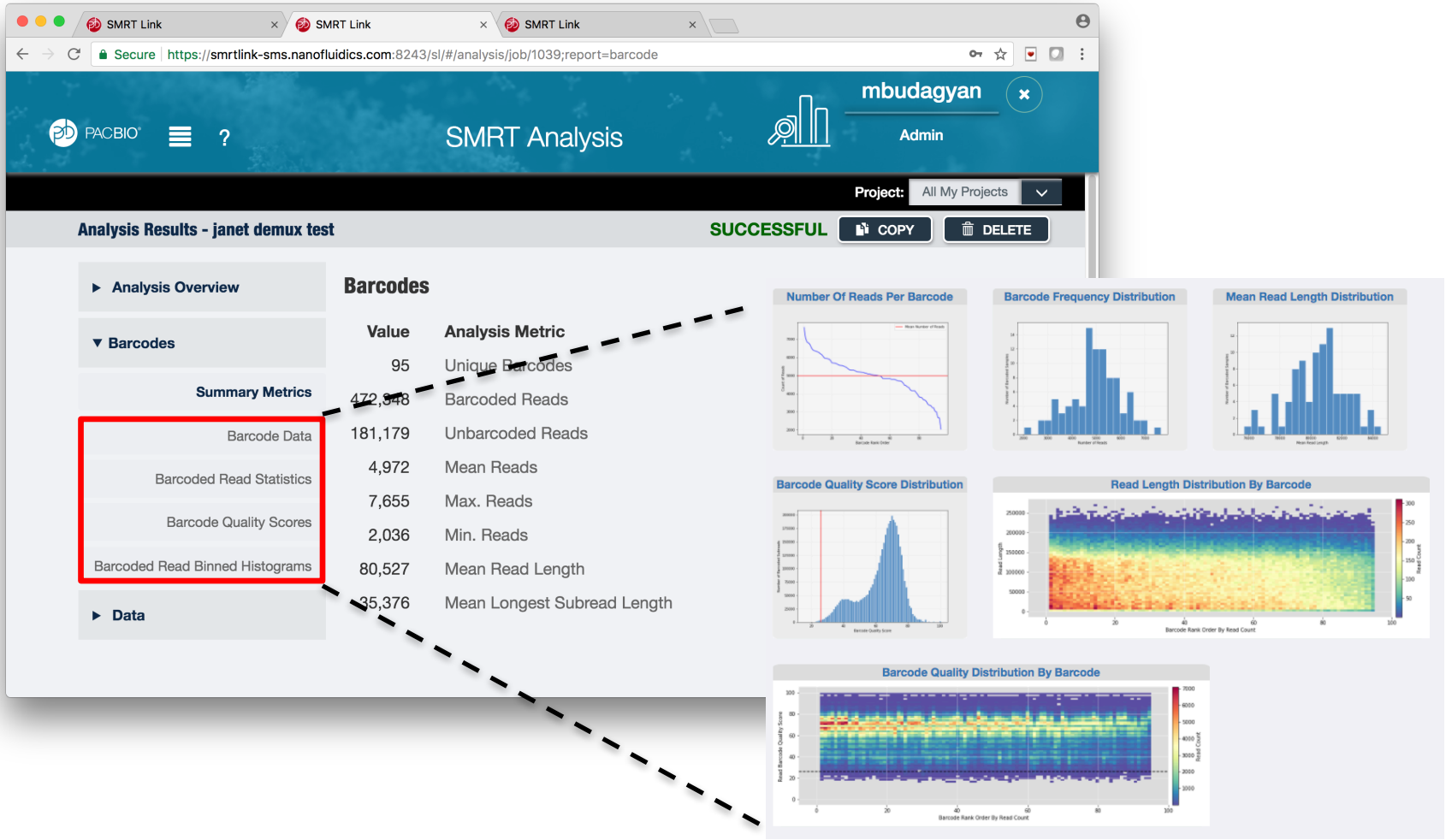
BARCODED SAMPLE NAME FILE



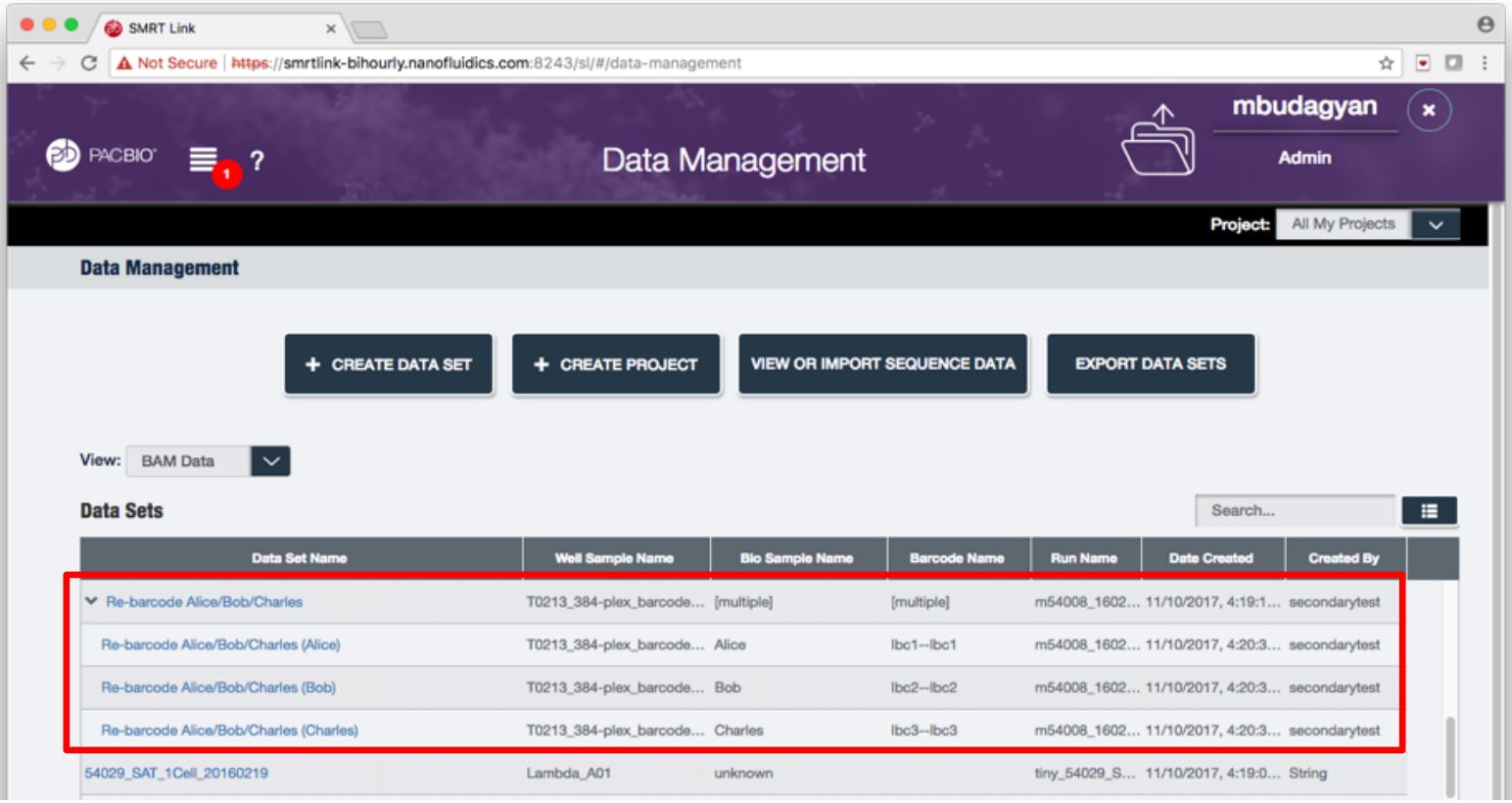


- Fill in the sample names for the barcodes used
- Upload the file

QC METRICS IN RESULTS OF DEMULTIPLEX BARCODES



ONE DATASET GENERATED PER BARCODE IN HIERARCHICAL DISPLAY



- Bio Sample Name assigned by Barcode Name if not provided
  - CSV is optional
- Ability to modify Bio Sample Name and demux parameters

ANALYZE EACH BARCODED DATASET WITH CUSTOM PARAMETERS

SMRT Link

Not Secure | https://smrtlink-bihourly.nanofluidics.com:8243/sl/#/analysis/settings

PACBIO

SMRT Analysis

mbudagyan

Admin

Project: All My Projects

Create New Analysis - Settings

CANCEL START

Analysis Application *

Assembly (HGAP 4)

Analysis of Multiple Datasets

Analysis Type *

One Analysis on All Data Sets

- One Analysis on All Data Sets
- One Analysis per Data Set - Identical Parameters
- One Analysis per Data Set - Custom Parameters

ADVANCED ANALYSIS PARAMETERS

Name *

View: BAM Data

Data Sets

Search...

	Data Set Name	Well Sample Name	Bio Sample Name
☐	Re-barcode Alice/Bob/Charles All None	T0213_384-plex_barcode...	[multiple]
☑	Re-barcode Alice/Bob/Charles (Alice)	T0213_384-plex_barcode...	Alice
☑	Re-barcode Alice/Bob/Charles (Bob)	T0213_384-plex_barcode...	Bob
☑	Re-barcode Alice/Bob/Charles (Charles)	T0213_384-plex_barcode...	Charles
☐	54029_SAT_1Cell_20160219	Lambda_A01	unknown
☐	Run_02.19.2016 00:27 384-plex barcode A8 All None	T0213_384-plex_barcode...	[multiple]
☐	Run_02.19.2016 00:27 384-plex barcode A8	T0213_384-plex_barcode...	unknown
☐	Run_20160304_SQ02_Run0753_PB1052	PB1052_A01	unknown
☐	pacbio_dataset_subreadset-170531_223853351	20170525_A12_IguanaSA...	unknown
☐	pacbio_dataset_subreadset-160216_202541225	unknown	unknown

EDIT PARAMETERS FOR EACH DATASET, THEN LAUNCH

SMRT Link

Not Secure | https://smrtlink-bihourly.nanofluidics.com:8243/sl/#/analysis/settings

PACBIO

SMRT Analysis

mbudagyan

Admin

Project: All My Projects

Create New Analysis - Settings

CANCEL

START AND CREATE NEXT

Analysis Application *

Assembly (HGAP 4)

Analysis of Multiple Datasets

Analysis Type *

One Analysis per Data Set - Custom Parameters

Assembly

Genome Length

5000000

ADVANCED ANALYSIS PARAMETERS

Name *

Mary's custom assembly

_Re-barcode Alice/Bob/Charles (Alice)

View: BAM Data

Data Sets

Search...

	Data Set Name	Well Sample Name	Bio Sample Name
	Re-barcode Alice/Bob/Charles All None	T0213_384-plex_barcode...	[multiple]
☑	Re-barcode Alice/Bob/Charles (Alice)	T0213_384-plex_barcode...	Alice
☑	Re-barcode Alice/Bob/Charles (Bob)	T0213_384-plex_barcode...	Bob
☑	Re-barcode Alice/Bob/Charles (Charles)	T0213_384-plex_barcode...	Charles
	54029_SAT_1Cell_20160219	Lambda_A01	unknown
	Run_02.19.2016 00:27 384-plex barcode A8 All None	T0213_384-plex_barcode...	[multiple]
	Run_02.19.2016 00:27 384-plex barcode A8	T0213_384-plex_barcode...	unknown
	Run_20160304_SQ02_Run0753_PB1052	PB1052_A01	unknown
	pacbio_dataset_subreadset-170531_223853351	20170525_A12_IguanaSA...	unknown

- SMRT Link highlights the current input dataset in yellow
- Autopopulates the dataset name to the (editable) analysis name

PACBIO®

Structural Variation

Joint calling

MULTI-SAMPLE SELECTION FOR STRUCTURAL VARIATION

SMRT Link

Secure | https://smrtlink-alpha.nanofluidics.com:8243/sl/#/analysis/settings

PACBIO

SMRT Analysis

mbudagyan

Admin

Project: All My Projects

Create New Analysis - Settings

CANCEL | START

Analysis Application *

Structural Variant Calling

Analysis Name *

Multi-sample SV Demo

Analysis of Multiple Datasets

View: BAM Data

Analysis Type *

One Analysis on All Data Sets

Data Sets

hg00

Associated Inputs

Reference *

Hg19

	Data Set Name	Well Sample Name	Bio Sample Name	Barcode Name
☑	HG00731-5fold J V&V	[multiple]	HG00731	
☑	HG00732-5fold J V&V	[multiple]	HG00732	
☑	HG00733-5fold J V&V	[multiple]	HG00733	
☐	AW Merge Test	[multiple]	Testing	
☐	HG00733_Jaguar_10hx3_20hx2	unknown	unknown	
☐	HG00731 5-fold JV&V	unknown	unknown	
☐	HG00733 10-fold JV&V	unknown	unknown	
☐	HG00732 5-fold JV&V	unknown	unknown	
☐	HG00733 5-fold JV&V	unknown	unknown	
☐	Jaguar_SV_VV_5cells_10hx3_20hx2	[multiple]	HG00733_Jaguar_10hx3...	

Structural Variants

Minimum Length of Structural Variant (bp)

50

Minimum Reads That Support Variant (Count)

2

Minimum Percentage of Reads That Support Variant (%)

20

MULTI-SAMPLE SUPPORT

SMRT Link | SMRT Link | SMRT Link

Secure | https://smrtlink-alpha.nanofluidics.com:8243/sl/#/analysis/job/27158;report=structural_variants;table=structural_variants.sample_table

PACBIO ? SMRT Analysis mbudagyan Admin

Project: All My Projects

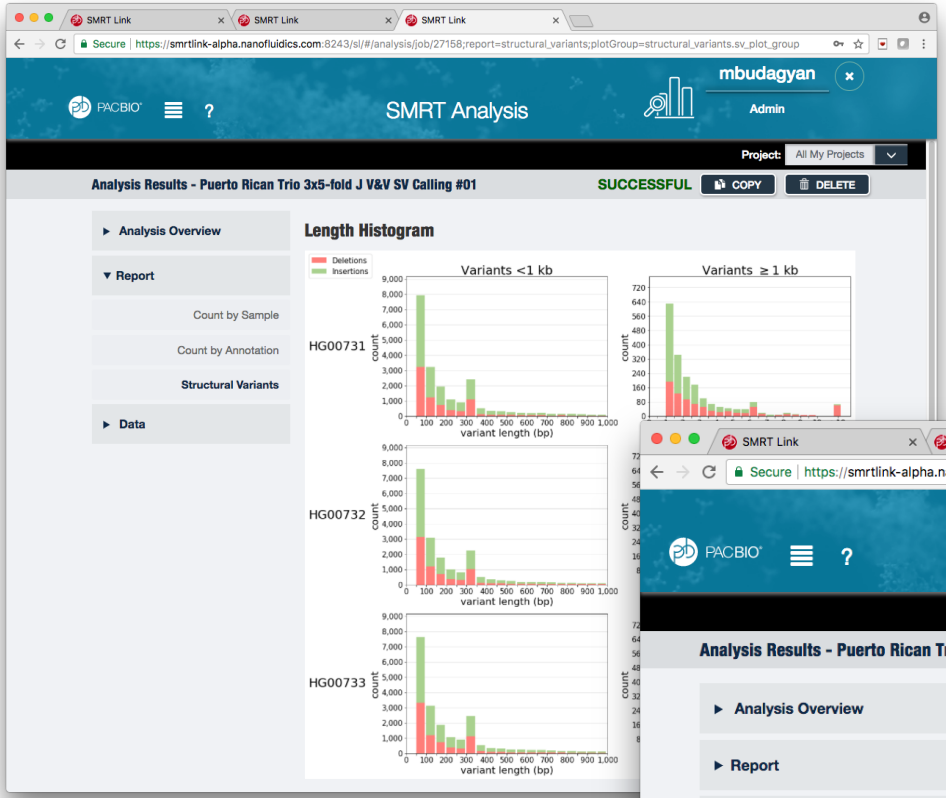
Analysis Results - Puerto Rican Trio 3x5-fold J V&V SV Calling #01 SUCCESSFUL COPY DELETE

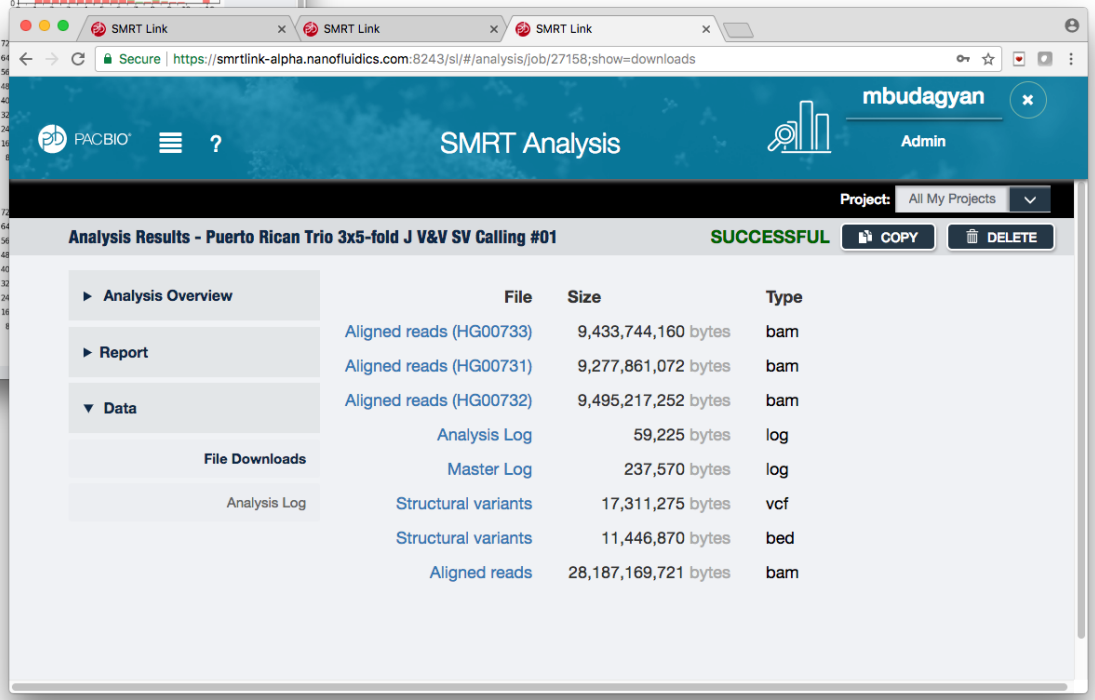
- Analysis Overview
- Report
  - Count by Sample
  - Count by Annotation
  - Structural Variants
- Data

Count by Sample

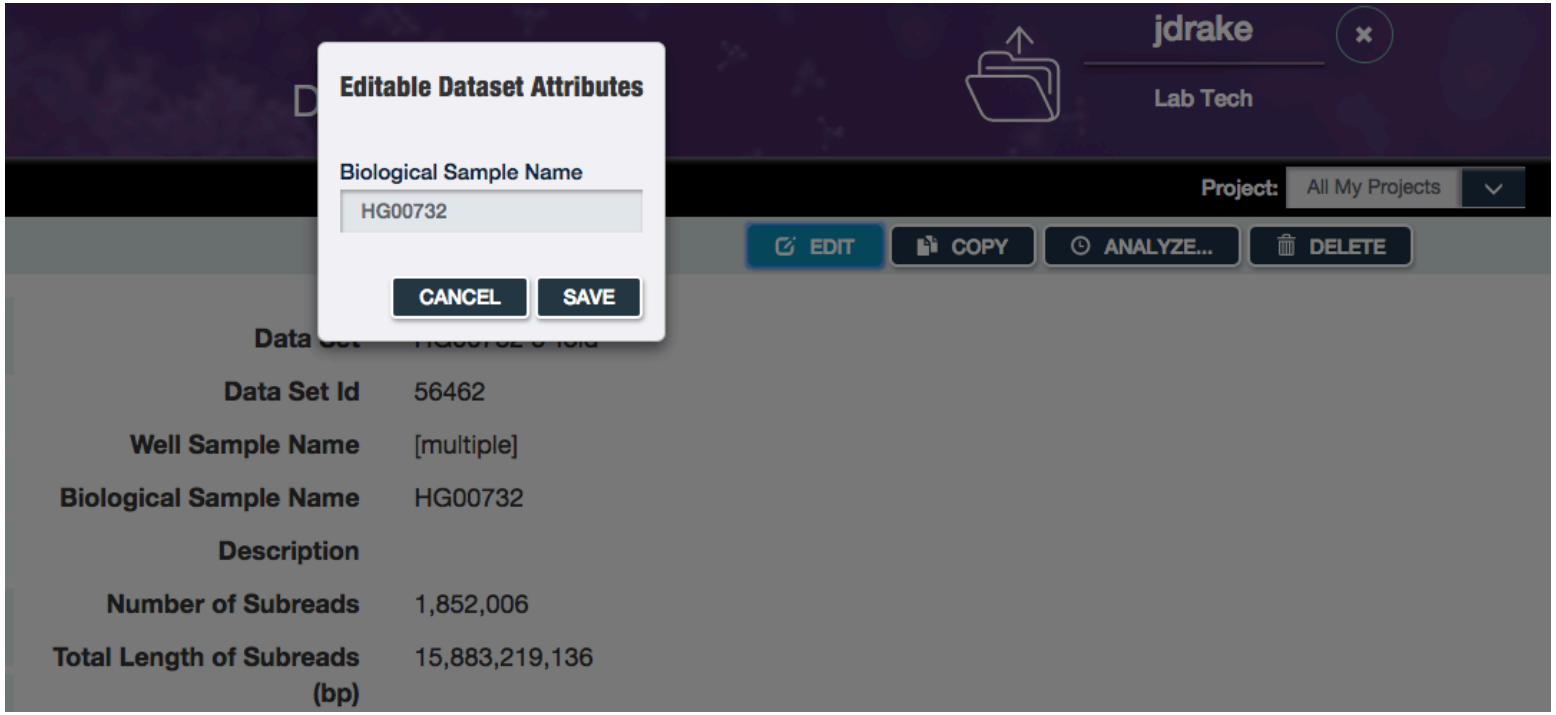
	Insertions (total bp)	Deletions (total bp)	Homozygous Variants	Heterozygous Variants	Total Variants (total bp)
HG00731	13,684 (5,389,546)	8,589 (5,113,599)	4,383	17,890	22,273 (10,503,145)
HG00732	12,859 (4,968,941)	8,124 (4,341,261)	4,259	16,724	20,983 (9,310,202)
HG00733	13,299 (5,595,503)	8,716 (5,194,418)	4,735	17,280	22,015 (10,789,921)

INFORMATION BROKEN UP BY SAMPLE



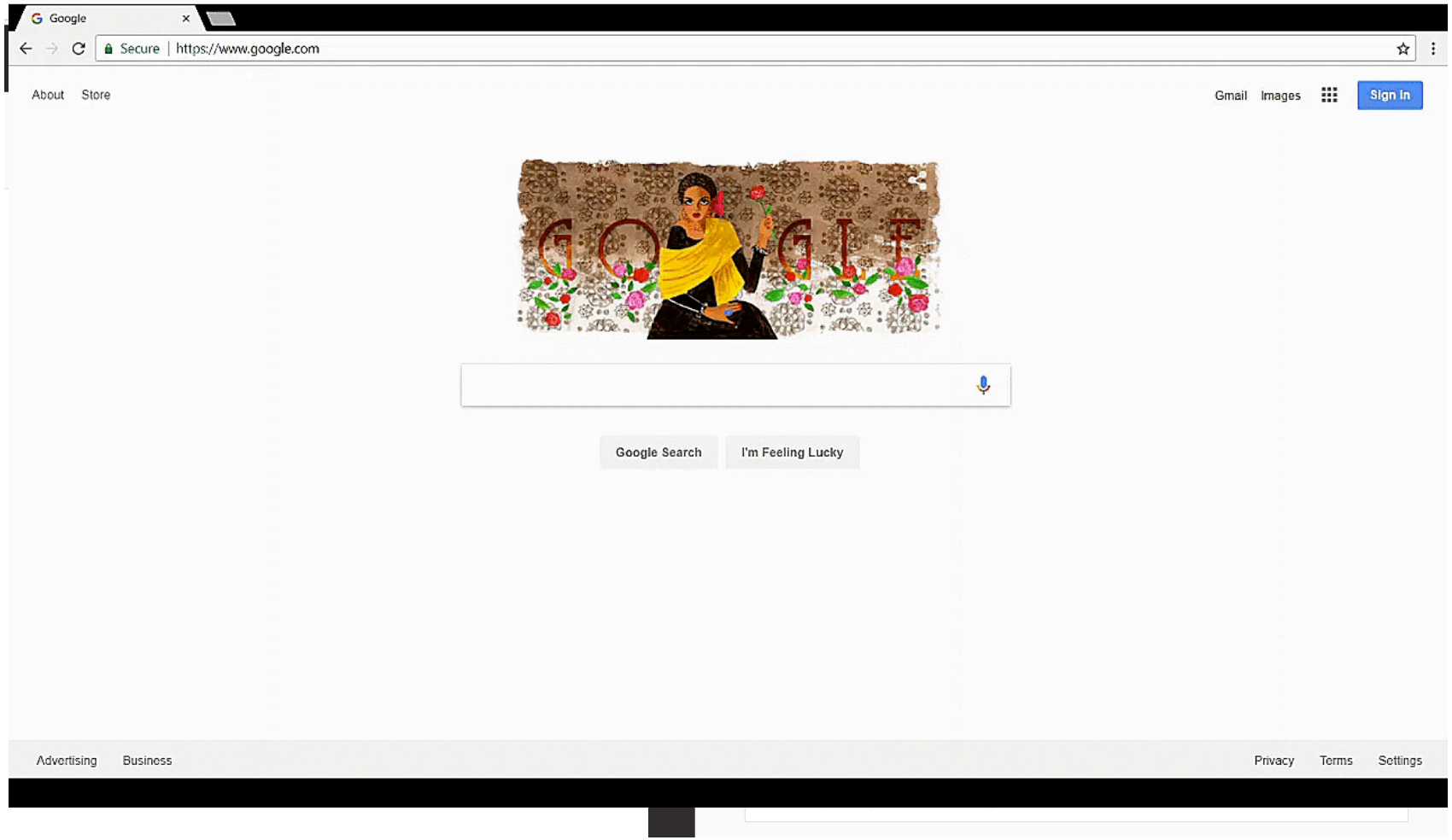


BIOSAMPLE NAME



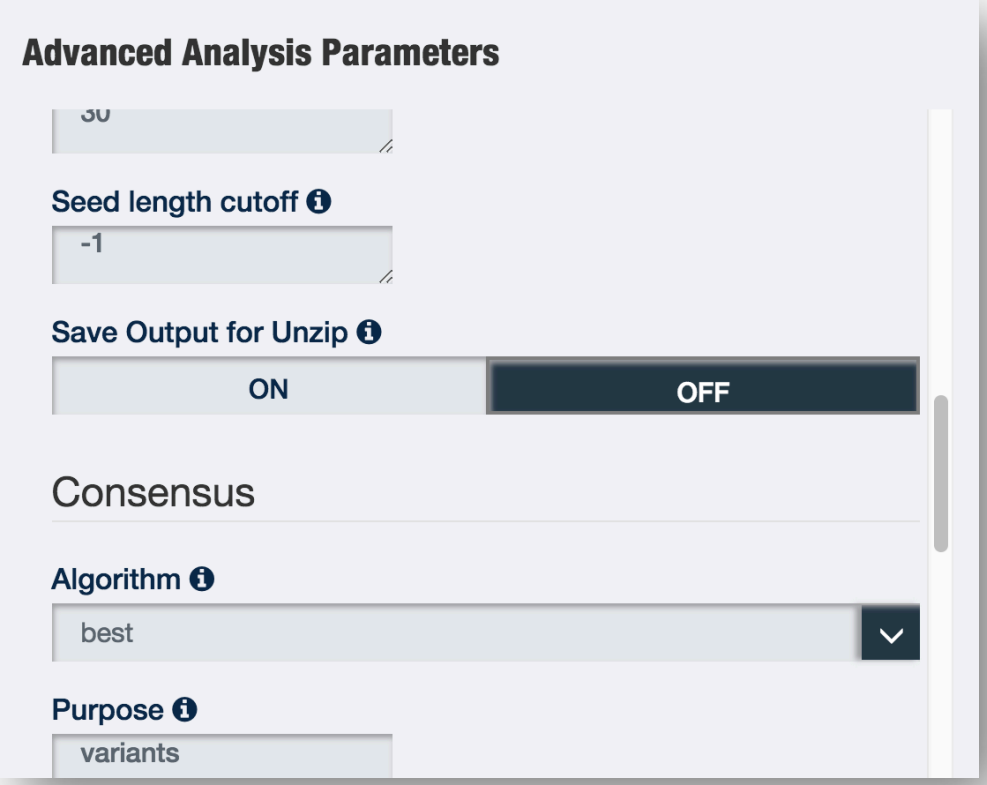
De Novo Assembly

BINARY RELEASE – FALCON/UNZIP

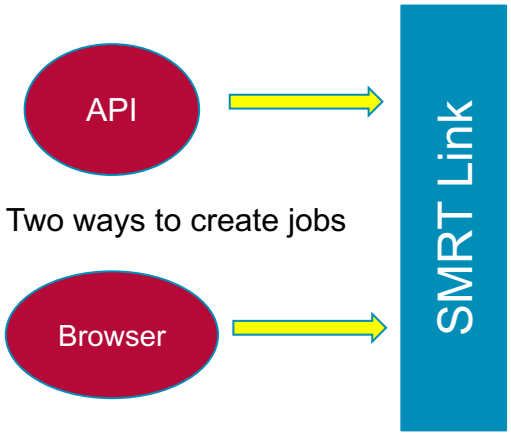


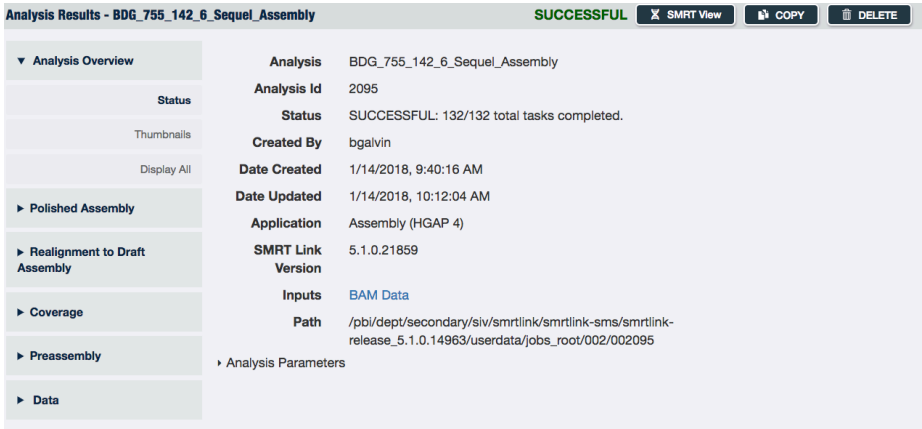
RUNNING UNZIP ON HGAP.4 OUTPUT

- HGAP 4 advanced parameter Save Output for Unzip, **off** by default
- Optionally retain the final set of `.las` files after overlapping raw reads, which can consume large amounts of disk space, especially for larger genomes
- Unnecessary when assembling haploid genomes, e.g., bacteria or when there is no intention to unzip



HGAP.4 TO UNZIP






```
# From the falcon code base
(venv) % python -m falcon_kit.mains.hgap4_adapt --help
usage: hgap4_adapt.py [-h] [--job-output-dir JOB_OUTPUT_DIR]

Given a full HGAP4 run, generate directories and symlinks to make it look like a pypeflow run.

optional arguments:
  -h, --help            show this help message and exit
  --job-output-dir JOB_OUTPUT_DIR
                        Directory of HGAP4 job_output. (A symlink or relative path is fine.) Task-dirs are under here in "tasks/" (de

Typically:
    mkdir mydir/
    cd mydir/
    python -m falcon_kit.mains.hgap4_adapt --job-output-dir=../job_output/
```

GFA OUTPUT – ASSEMBLY DEBUGGING

Script available in both SMRT Tools and GitHub

Bandage

https://github.com/rrwick

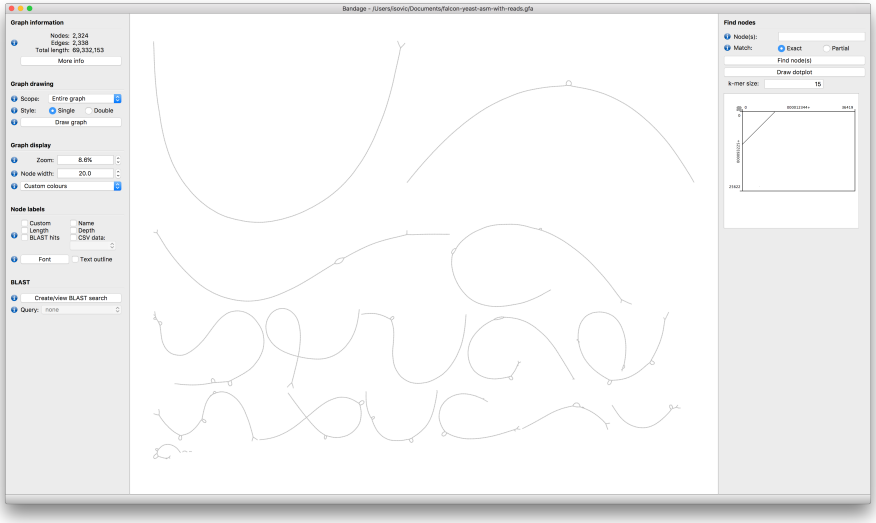
Falcon Assembly

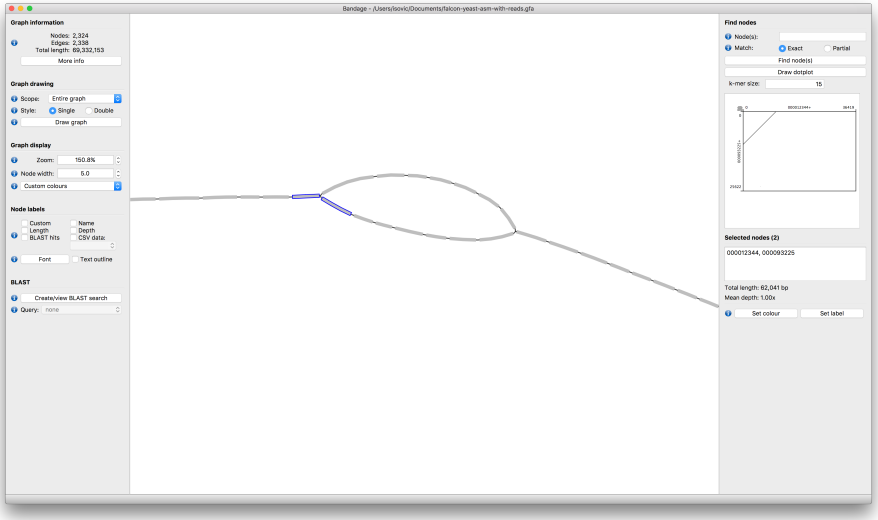
```
% python ./falcon_kit/mains/gen_gfa_v1.py -h

usage: gen_gfa_v1.py [-h] [--p-ctg-tiling-path P_CTG_TILING_PATH]
                     [--a-ctg-tiling-path A_CTG_TILING_PATH]
                     [--preads-fasta PREADS_FASTA] [--p-ctg-fasta P_CTG_FASTA]
                     [--a-ctg-fasta A_CTG_FASTA]
                     [--sg-edges-list SG_EDGES_LIST] [--utg-data UTG_DATA]
                     [--ctg-paths CTG_PATHS] [--add-string-graph]
                     [--write-reads] [--write-contigs] [--min-p-len MIN_P_LEN]
                     [--min-a-len MIN_A_LEN]

Generates GFA output (on stdout) from FALCON's assembly.
```

GFA File





PACBIO®

Usability Improvements

Data Management and SMRT Analysis

COPY AND RELAUNCH ANALYSIS

SMRT Link | SMRT Link

Secure | https://smrtlink-alpha.nanofluidics.com:8243/sl/#/analysis/job/28612

PACBIO ? SMRT Analysis mbudagyan Admin

Project: All My Projects

Analysis Results - Demo SV Send Log Files FAILED COPY DELETE

▼ Analysis Overview

Status

Display All

▶ Data

Analysis	Demo SV
Analysis Id	28612
Status	FAILED: Task pbsvtools.tasks.align-18 FAILED in 295.08 sec.
Created By	mbudagyan
Date Created	12/12/2017, 10:47:31 AM
Date Updated	12/12/2017, 10:55:59 AM
Application	Structural Variant Calling
SMRT Link Version	5.1.0.20106
Inputs	BAM Data Reference

SMRT Link | SMRT Link

Secure | https://smrtlink-alpha.nanofluidics.com:8243/sl/#/analysis/settings

PACBIO ? SMRT Analysis mbudagyan Admin

Create New Analysis - Settings CANCEL START

Analysis Application *: Structural Variant Calling

Analysis Name *: Copy of Demo SV

Associated Inputs

Reference *: Hg19

View: BAM Data

Data Sets Search...

Structural Variants

Minimum Length of Structural Variant (bp): 50

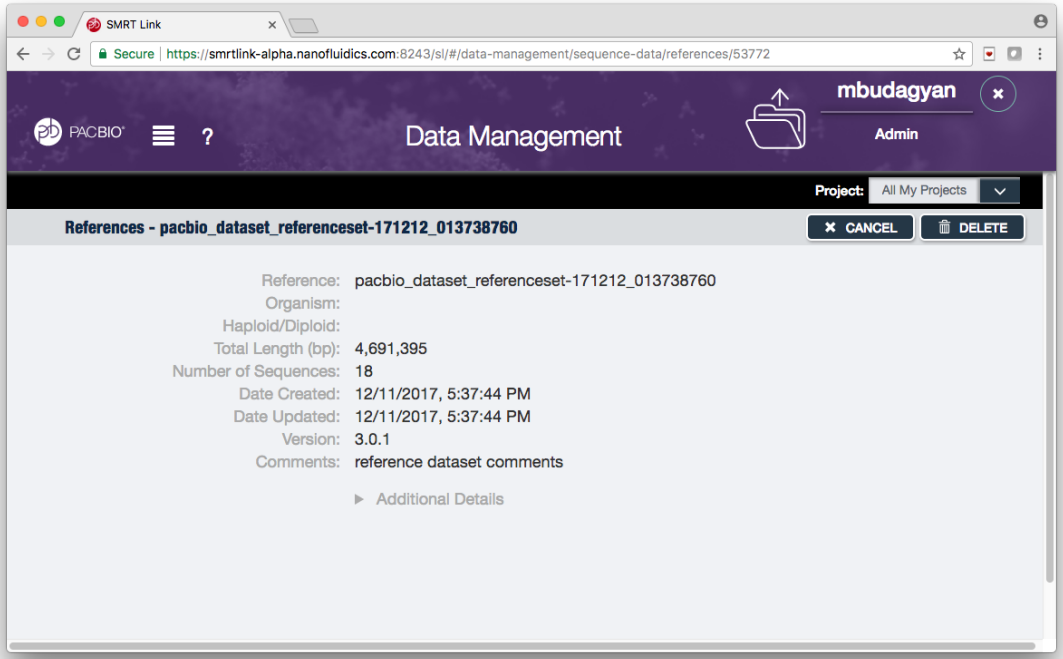
Minimum Reads That Support Variant (Count): 2

Minimum Percentage of Reads That Support Variant (%): 20

	Data Set Name	Well Sample Name
	BYU_Turneri-Cell5	BYU_Turneri
☑	HG00733 Sequel	unknown
	3_Cell-1_Workflow_Diffusion.py-Cell8	3_Cell-1_Workflow_Diffusi...
	15kb_Column Cleanup_5pM-Cell6	15kb_Column Cleanup_5pM
	2_Cell-1_Workflow_Diffusion.py-Cell7	2_Cell-1_Workflow_Diffusi...
	BYU_Turneri-Cell4	BYU_Turneri
	M1028-8pM-Cell4	M1028-8pM

DATA MANAGEMENT – REMOVING THE UNWANTED

- Deletes from the UI only
  - Collections
  - References
  - Barcode sets
- Helps reduce clutter
- UI responsiveness



CREATE ANALYSIS - NEW LAYOUT

SMRT Link

Not Secure | https://smrtlink-bihourly.nanofluidics.com:8243/sl/#/analysis/settings

PACBIO

SMRT Analysis

mbudagyan

Admin

Project: All My Projects

Create New Analysis - Settings

CANCEL | START

Analysis Application *

- Assembly (HGAP 4)
- Base Modification Detection
- Base Modification and Motif Analysis
- CCS Mapping
- Circular Consensus Sequences (CCS)
- Convert BAM to FASTX
- Demultiplex Barcodes
- Iso-Seq

Analysis Name *

View: BAM Data

Data Sets

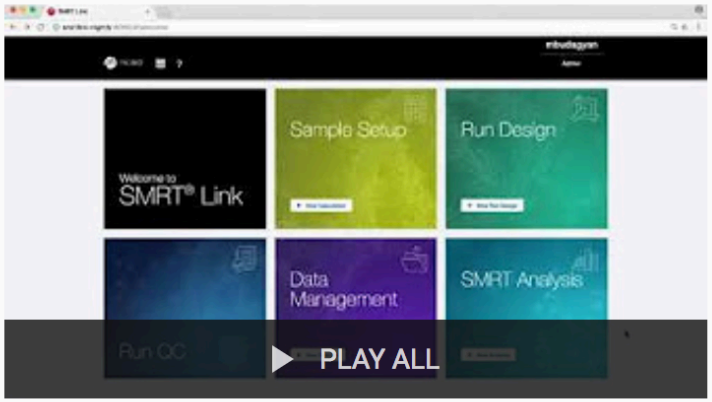
Search...

	Data Set Name	Well Sample Name	E
☐	hdfsubreads	unknown	un
☐	Output-dataset-for-protractor-barcoding-test All None	BCS23 diffusion 70pM P6...	un
☐	Auto-merged subreads @ 1513096418273	[multiple]	un
☐	pacbio_dataset_subreadset-171114_215558341	10-10-2017_14hr_Lambd...	un
☐	lambda/0007_tiny	Inst42267-040315-SAT-10...	un
☐	pacbio_dataset_subreadset-170313_050639539	unknown	un
☐	LVP2_D09-5253_BA008270-1st_4hrs_PkmidC525	"LVP2_D09-5253_BA0082...	un

SMRT LINK REFERENCE MATERIALS

pacb.com > Support > Software Downloads

- SMRT Link Documentation
  - Release Notes
  - Installation Instructions
  - SMRT Link User Guide
  - Barcoding Overview
- Developer Documentation
  - SMRT Tools Reference Guide
  - SMRT Link APIs
- Training materials
  - SMRT Link and SMRT Analysis Tutorials



PacBio SMRT Link Training Series

9 videos • 324 views • Last updated on Dec 4, 2017

PacBio

SUBSCRIBE

PACBIO®

www.pacb.com