Overview:

1. Assembly and methylation prediction with SMRTanalysis
2. QC of the SMRTanalysis results (also updating logs with the results of the SMRTanlysis run)
3. Structural Variation prediction/Assembly QC with PBHoney
4. Final assembly QC (update logs with PBhoney output)
5. Troubleshooting – what to do if things don't run correctly

1. **Assembly and methylation prediction with SMRTanalysis pipelines:**

Command to run the smrtanalysis:

```
cd $GROUPHOME/data/genomes
```

```
make –f $GROUPHOME/data/depot/assembly/excecute-smrtanalysis.mk  isolate.fasta EMAIL=youremail@blah.com
```

Example:

```
cd $GROUPHOME/data/genomes
```

```
make –f $GROUPHOME/data/depot/assembly/excecute-smrtanalysis.mk 1-0007.fasta EMAIL=sramirezbusby@sdsu.edu
```

Notes:

SMRTanalysis does the qsub for you.

This means you need to start a screen session (Screen Tutorial)

To start a screen session type screen

To detach from the screen (e.g. keep the screen active but close it) Ctl + a d (hold down Control key and press a then release the control key and press d

To reattach to a screen session type screen –r (if you have multiple screens, ask Sarah for help)

2. **QC of the SMRTanalysis results**

Command to run the "QC script", which will verify that outputs of each SMRTanalysis step have been created:

```
<cd $GROUPHOME/data/depot/assembly/>
```

```
< ./denovo_asmb_initial_qc.py -log  /path/to/repository/  -id [isolate id]>
```

Example:

```
< ./denovo_asmb_initial_qc.py -log ~/assembly_qc/denovo_initial_qc/clinical-isolates.pub/doc/  -id 1-0007 >
```

To run multiple isolates based on regular expressions replace the wildcard (*) in the example command with a pattern.

Example:

(using pattern"1-01*" to run all the isolates beginning with the prefix "1-01")

```
for f in 1-01* ;  do ./denovo_asmb_initial_qc.py -log ~/assembly_qc/denovo_initial_qc/clinical-isolates.pub/doc/ -id $f ; done
```

Notes:

~This script must be run from inside $GROUPHOME/data/depot/assembly/

~This script does not create log files, it only modifies existing ones. For this reason, the file path to the logs is required.

~The log files must still be added to the Lab's git repository (commit, push, merge) in order for the updates to the log files to be saved!

~Qsubs for you

This means you need to start a screen session (Screen Tutorial)

To start a screen session type screen

To detach from the screen (e.g. keep the screen active but close it) Ctl + a d (hold down Control key and press a then release the control key and press d

To reattach to a screen session type screen –r (if you have multiple screens, ask Sarah for help)

3. **Structural Variation prediction with PBHoney**

Command to run PBHoney (and an associated QC script):

```
< cd $GROUPHOME/data/depot/assembly/[isolate]/ >
```

```
< make –f ../assembly-qc.mk isolate=[isolate]>
```

Example:

```
< cd $GROUPHOME/data/depot/assembly/1-0007 >
```

```
< make –f ../assembly-qc.mk isolate=1-0007>
```

Notes:

~This script must be run from inside $GROUPHOME/data/depot/assembly/[isolate]/

~This script executes two programs: PBHoney, a structural prediction tool, and an in-house script to evaluate the PBHoney output. The output files are then not only the pbhoney output (isolate.hon.tails) but also a temp file that contains the status, either "pass" or "fail", of the PBhoney run as it relates to assembly QC. This temp file will be read and then removed by the QC script in step 4.

~Takes a long time to run so it is a good idea to use qsub or screen to run this step.

4. **Final assembly QC (repeat of step 2!)**

Re-run the QC script, which will detect that a temp file has been created by assembly-qc.mk script that was executed in step 3. The script will update logs to reflect PBhoney status.

Command to run the "QC script", which will verify that the outputs of each SMRTanalysis step have been created: (same as step 2!)

```
<cd $GROUPHOME/data/depot/assembly/>
```

```
<./denovo_asmb_initial_qc.py -log  /path/to/repository/  -id [isolate id]>
```

Example:

```
./denovo_asmb_initial_qc.py -log ~/assembly_qc/denovo_initial_qc/clinical-isolates.pub/doc/  -id 1-0007
```

5. **Troubleshooting:**

Potential Issues:

Issue 1: failed assembly

Issue 2: failed circularization

Issue 3: failed methylation

Issue 4: failed QC with PBhoney

Issue 1: In the event of a failed assembly:

Nothing...yet. Fixes coming soon!!!!

Issue 2: In the event of a completed assembly but failed circularization:

You may run an alternate circularization tool called CircLator. Commands to run:

```
< cd $GROUPHOME/data/depot/assembly/[isolate] >
< make –f ../circlator.mk circularization/circularized.fasta >
```

Example:

```
< cd $GROUPHOME/data/depot/assembly/1-0007>
< make –f ../circlator.mk circularization/circularized.fasta >
```

Re-run SMRTanalysis pipeline (Step 1), it will detect the circularized file and begin at the appropriate step in the pipeline automatically. Continue with the rest of the steps normally. **You do not need to run denovo_asmb_initial_qc.py after circlator.mk.**

Issue 3: In the event of a completed assembly and circularization but failed Methylation:

(this is due to a time-out error)

Rerun SMRTanalysis pipeline (Step 1). It will detect that methylation is incomplete and begin at the appropriate step in the pipeline automatically. Continue with the rest of the steps normally.

Issue 4: In the event of a completed assembly, circularization, and methylation but a failed PBHoney QC (I.e. the PBHoney reported significant structural variations):

Manually examine the output of the PBHoney run and open the output file found here:

$GROUPHOME/data/depot/assembly/[isolate-id]/assembly-qc/*hon.tails

The variation(s) that caused the PBhoney QC to fail are identified in the "Notes" section of the logfiles. Add additional info, in any, to log file (manually)