Check-In Procedure for GBS Data from KSU Genomics Facility

Prerequisites

An Illumina account is required to be able to access the GBS data produced by the KSU Genomics Facility. You can register for an account at https://login.illumina.com

The Illumina Basespace CLI must be installed (e.g. on Beocat) in order to access the Illumina data via the automated scripts that have been developed. Instructions for installing the CLI can be found at https://developer.basespace.illumina.com/docs/content/documentation/cli/cli-overview

The following shell scripts and programs that are stored in /homes/mlucas/scripts are required:

checkin_ksu_gbs_data.sh

rename_gbs_file

filter_FASTQ_byLength_outgz.pl

compute_gbs_file_metadata

generate_barcode_distribution

generate_blank_dna_quantification_report

Accept the Share Notification of New Data from Basespace

1. Notification of sharing of new data will be received via email from basespace-noreply@illumina.com - Subject will contain Project Share Request
2. Click on the share link in the email to open a browser with an Illumina login screen to accept the share of the new data.
3. Login to Illumina
4. Click on the link to share the data and logout.
5. Note the name of the Illumina project that was shared. The KSU Genomics Facility name this using the GBS Number and Flowcell e.g. 1370_HVY7JBGX

Execute the Script to Download and Check-in the GBS Data

6. Logon to Beocat
7. cd to /homes/mlucas/scripts and locate the script **checkin_ksu_gbs_data.sh**
8. Execute the script on the command line with the Illumina project name
   Example:

   ./checkin_ksu_gbs_data.sh 1370_HVY7JBGX

   The script will perform the following steps:

   a. Create a folder with the Illumina project name e.g. 1370_HVY7JBGX
   b. Download the GBS data from Basespace using the Basepace CLI (bs-cp)
   c. Verify the checksums of the 4 GBS files that were downloaded.
   d. Update the gbs database table for the associated gbs_id with the flowcell and lane values. (Lane is always set to 0 for GBS files produced by the KSU Genomics Facility).
   e. Create the target GBS file by concatenating the 4 GBS files that were downloaded into a new file with the correct naming format

      Example:

      GBS1370xStrawberryP01P02_HVY7JBGX7_s_0_fastq.txt.gz

   f. Filter out short reads of less than 75 bp in length. This will create a new filtered GBS file which will have an “F” after the GBS number part of the name.

      Example:

      GBS1370FxStrawberryP01P02_HVY7JBGX7_s_0_fastq.txt.gz

   g. Compute the MD5 checksum and line count for the filtered GBS file and update the gbs table ms5sum and num_lines columns for the gbs_id associated with the file.

   h. Generate read-barcode distribution report

      This report will allow the user to check % valid reads and % reads found in any blank well in the GBS file.

      The report will have the following naming format:

      GBSnnnn_sample_summary.txt

      Example:

      GBS1370_sample_summary.txt

i. Generate DNA quantification report

This report will allow the user to check the DNA quantification values for blank wells in the GBS library

The report will have the following naming format:

GBSnnnn_blank_dna_quant_report.csv

Example:

```
GBS1370_blank_dna_quant_report.csv
```

Review QC Reports and Cleanup

9. Review the GBSnnnn_sample_summary.txt report and verify that the following thresholds have not been exceeded:

   % Valid reads > 90%
   % Reads in any BLANK well < 0.01%

   If either threshold is violated, investigate potential causes:
   i. Incorrect blank well in DNA plate record
   ii. Poor sequencing run quality

10. Review the GBSnnnn_blank_dna_quant_report.csv to make sure that DNA quantification values in the blank wells are within tolerance.

    If the values reported are all NULL, this means that the dnaQuant table has not been updated yet for this GBS plate.

11. Change the group on the GBS file to ksu-plantpath-jpoland and remove write permissions from the file.

```
chgrp ksu-plantpath-jpoland GBS1370FxStrawberryP01P02_HVY7JBGX7_s_0_fastq.txt.gz

chmod a-w GBS1370FxStrawberryP01P02_HVY7JBGX7_s_0_fastq.txt.gz
```

12. Move the GBS file to /bulk/jpoland/sequence directory on Beocat.

```
mv GBS1370FxStrawberryP01P02_HVY7JBGX7_s_0_fastq.txt.gz /bulk/jpoland/sequence/.
```