

NEXT-GENERATION SEQUENCING SERVICE REPORT

CAGRIF19032

AGRF BIOINFORMATICS

Email: bioinformatics@agrif.org.au

AGRF NEXT-GENERATION SEQUENCING

Email: NextGen@agrif.org.au

CONTENTS

Bioinformatics Analyst details	1
Project Overview	2
Primary analysis	2
Data Yield	2
Data Delivery	3
AGRF Data Verification	4
Windows:	4
Unix:	4

Name	Matthew Stevens
Contact number	(03) 9345 2679
Email address	Matthew.Stevens@agrif.org.au

Project Overview

Project code: CAGRF19032
Type of service: Illumina Sequencing
Project start date: Tuesday 24th September, 2019

This document reports the data generation of the Illumina sequencing run and follow up analysis if applicable.

Primary analysis

Image analysis was performed in real time by the NovaSeq Control Software (NCS) v1.6.0 and Real Time Analysis (RTA) v3.4.4, running on the instrument computer. RTA performs real-time base calling on the NovaSeq instrument computer. Then the Illumina bcl2fastq 2.20.0.422 pipeline was used to generate the sequence data. The data generated here meet the AGRF quality standards.

Data Yield

The data yield is as follows:

Table 1:

150bp Paired End - Flowcell ID: HTH7NDSXX

Lane	Sample Name	Paired End	Data Yield(bp)
1	S11	193,895,491	58.56 Gb
	S13	183,180,085	55.32 Gb
	S14	254,146,829	76.75 Gb
	S17	220,677,187	66.64 Gb
	S20	227,655,689	68.75 Gb
	S22	227,607,718	68.74 Gb
	S23	190,636,099	57.57 Gb
	S27	204,027,795	61.62 Gb
	S3	252,646,110	76.30 Gb
	S31	260,033,513	78.53 Gb
	S35	202,081,413	61.03 Gb
	S5	192,299,753	58.07 Gb

Continued on next page

Table 1 – Continued from previous page

Lane	Sample Name	Paired End	Data Yield(bp)
	S8R	168,089,731	50.76 Gb
	S9R	199,187,487	60.15 Gb
2	S36	200,342,622	60.50 Gb
	T1	190,069,080	57.40 Gb
	T16	206,341,810	62.32 Gb
	T18	201,629,889	60.89 Gb
	T21	284,594,169	85.95 Gb
	T23	183,694,789	55.48 Gb
	T27	227,503,574	68.71 Gb
	T29	185,011,236	55.87 Gb
	T36	273,617,925	82.63 Gb
	T5	216,533,782	65.39 Gb
	T8	165,058,134	49.85 Gb
Total		5,310,561,910	1603.79Gb

Data Delivery

The data has been transferred and made available to you.

Below is a brief description of the files to expect.

<sample_name>_<flowcell_ID>_<index>_<lane>_<readNum>_fastq.gz - compressed FastQ formatted sequence file containing untrimmed 150 bp reads, where <readNum> specifies the first or second read of the pair.

These files contain the read **sequence output** with the corresponding Illumina **quality scores**. The quality is encoded in symbolic ASCII format in the following way: char (phred-like quality score + 33).

The files have been compressed using gzip and can be uncompressed using gunzip (*nix) or 7-Zip (Windows).

Kindly review the sequencing files after you have received them. Files will **remain on the server** until the **24th of October, 2019**, after that date they will be removed from the server. Files will be stored at the AGRF until the 2nd of January, 2020. After that date the FASTQ files will be able to be recreated, until the 20th of July, 2020. Please note that file recreation will be charged for.

If you have any questions or require further assistance please do not hesitate to contact us at NextGen@agrif.org.au.

AGRF Data Verification

As part of the general quality assurance AGRF tests the integrity and completeness of the data before it is sent to you.

Your packaged data has checksum information provided by AGRF should you wish to verify that your data is complete and error-free. Verification methods for both Windows and Unix systems are outlined below:

Windows:

The checksum information of the files included has been generated using Exactfile, which is a free third-party tool for Windows:
<http://www.exactfile.com/>.

Included in your directory are two files, “checksums.exf”, which contains the MD5 hashes of the original files on AGRF’s data server and “TestFiles.exe”, an applet created by the Exactfile program to check files against the hashes listed in “checksums.exf”.

To test the integrity of your files, you can simply start “TestFiles.exe”.

Please be patient as it may take up to several minutes before the file check report appears, because of the size of the fastq files.

Unix:

The checksum information of the files included has been generated using the md5sum program. The calculated checksums of your data files are included in the “checksums.md5” file.

To test the integrity of your files, type the following command while in the directory containing checksums.md5:

```
md5sum -c checksums.md5
```

*Please note: The checksum verification can only be performed on the original data as received from AGRF, before any files are moved or modified. AGRF accepts no responsibility regarding the performance of Exactfile program.