

# NEXT-GENERATION SEQUENCING SERVICE REPORT

---

**CAGRIF21303**

**AGRIF BIOINFORMATICS**

**Email: [bioinformatics@agrif.org.au](mailto:bioinformatics@agrif.org.au)**

**AGRIF NEXT-GENERATION SEQUENCING**

**Email: [NextGen@agrif.org.au](mailto:NextGen@agrif.org.au)**

## CONTENTS

Bioinformatics Analyst details . . . . .	1
<b>Project Overview</b>	<b>2</b>
<b>Primary analysis</b>	<b>2</b>
Data Yield . . . . .	2
Data Delivery . . . . .	6
<b>AGRF Data Verification</b>	<b>6</b>
Windows: . . . . .	6
Unix: . . . . .	7

Name	Christopher Nouné
Contact number	+61 3 9321 3745
Email address	Christopher.Nouné@agrif.org.au

## Project Overview

**Project code:** CAGRF21303  
**Type of service:** Illumina Sequencing  
**Project start date:** Monday 23<sup>rd</sup> 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 MiSeq Control Software (MCS) v2.6.2.1 and Real Time Analysis (RTA) v1.18.54, running on the instrument computer. RTA performs real-time base calling on the MiSeq 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:

### 300bp Paired End - Flowcell ID: 000000000-CPC4M

Lane	Sample Name	Paired End	Data Yield(bp)
1	E10-Exon2_AGRF	108,074	0.07 Gb
	E10-Exon2_Client	47,117	0.03 Gb
	E10-Exon3_AGRF	37,295	0.02 Gb
	E12-Exon2_AGRF	98,729	0.06 Gb
	E12-Exon2_Client	80,358	0.05 Gb
	E12-Exon3_AGRF	80,459	0.05 Gb
	E14-Exon2_AGRF	54,527	0.03 Gb
	E14-Exon2_Client	55,233	0.03 Gb
	E14-Exon3_AGRF	21,828	0.01 Gb
	E18-Exon2_Client	49,214	0.03 Gb
	E18-Exon3_AGRF	27,783	0.02 Gb
	E19-Exon2_Client	61,713	0.04 Gb

*Continued on next page*

Table 1 – *Continued from previous page*

Lane	Sample Name	Paired End	Data Yield(bp)
	E2-Exon2_AGRF	93,601	0.06 Gb
	E2-Exon2_Client	70,910	0.04 Gb
	E2-Exon3_AGRF	38,130	0.02 Gb
	E23-Exon2_Client	49,697	0.03 Gb
	E23-Exon3_AGRF	37,418	0.02 Gb
	E28-Exon2_Client	6,089	0.00 Gb
	E32-Exon2_AGRF	135,506	0.08 Gb
	E32-Exon2_Client	92,468	0.06 Gb
	E32-Exon3_AGRF	54,362	0.03 Gb
	E36-Exon2_AGRF	92,487	0.06 Gb
	E36-Exon2_Client	90,347	0.05 Gb
	E36-Exon3_AGRF	64,178	0.04 Gb
	E41-Exon2_AGRF	73,518	0.04 Gb
	E41-Exon2_Client	53,211	0.03 Gb
	E41-Exon3_AGRF	62,075	0.04 Gb
	E44-Exon2_AGRF	112,821	0.07 Gb
	E44-Exon2_Client	118,511	0.07 Gb
	E44-Exon3_AGRF	92,187	0.06 Gb
	E50-Exon2_AGRF	45,829	0.03 Gb
	E50-Exon2_Client	61,245	0.04 Gb
	E50-Exon3_AGRF	42,428	0.03 Gb
	E52-Exon2_Client	42,994	0.03 Gb
	E52-Exon3_AGRF	47,748	0.03 Gb
	E63-Exon2_Client	45,841	0.03 Gb
	E64-Exon2_Client	60,110	0.04 Gb
	E65-Exon2_Client	84,726	0.05 Gb
	E67-Exon2_AGRF	47,858	0.03 Gb
	E67-Exon2_Client	63,672	0.04 Gb
	E67-Exon3_AGRF	47,197	0.03 Gb
	E68-Exon2_AGRF	54,419	0.03 Gb
	E68-Exon2_Client	84,074	0.05 Gb
	E68-Exon3_AGRF	15,528	0.01 Gb
	E69-Exon2_Client	4,173	0.00 Gb
	E69-Exon3_AGRF	24,443	0.01 Gb
	E70-Exon2_Client	123,267	0.07 Gb
	E82-Exon2_AGRF	211,502	0.13 Gb
	E82-Exon2_Client	260,993	0.16 Gb
	E82-Exon3_AGRF	1,590	0.00 Gb

*Continued on next page*

Table 1 – *Continued from previous page*

Lane	Sample Name	Paired End	Data Yield(bp)
	E85-Exon2_Client	131,189	0.08 Gb
	E85-Exon3_AGRF	60,448	0.04 Gb
	E97-Exon2_AGRF	48,588	0.03 Gb
	E97-Exon2_Client	100,906	0.06 Gb
	E97-Exon3_AGRF	113,849	0.07 Gb
	S11-Exon2_AGRF	108,792	0.07 Gb
	S11-Exon2_Client	56,146	0.03 Gb
	S11-Exon3_AGRF	56,117	0.03 Gb
	S13-Exon2_AGRF	52,794	0.03 Gb
	S13-Exon2_Client	36,884	0.02 Gb
	S13-Exon3_AGRF	54,926	0.03 Gb
	S14-Exon2_AGRF	55,888	0.03 Gb
	S14-Exon2_Client	61,080	0.04 Gb
	S14-Exon3_AGRF	43,246	0.03 Gb
	S17-Exon2_AGRF	52,243	0.03 Gb
	S17-Exon2_Client	61,070	0.04 Gb
	S17-Exon3_AGRF	38,052	0.02 Gb
	S20-Exon2_Client	35,636	0.02 Gb
	S20-Exon3_AGRF	41,408	0.02 Gb
	S22-Exon2_Client	16,086	0.01 Gb
	S22-Exon3_AGRF	1,966	0.00 Gb
	S23-Exon2_Client	43,080	0.03 Gb
	S23-Exon3_AGRF	68,183	0.04 Gb
	S27-Exon2_AGRF	38,434	0.02 Gb
	S27-Exon2_Client	46,855	0.03 Gb
	S27-Exon3_AGRF	32,563	0.02 Gb
	S3-Exon2_AGRF	54,246	0.03 Gb
	S3-Exon2_Client	43,872	0.03 Gb
	S3-Exon3_AGRF	49,639	0.03 Gb
	S31-Exon2_AGRF	77,107	0.05 Gb
	S31-Exon2_Client	50,505	0.03 Gb
	S31-Exon3_AGRF	74,715	0.04 Gb
	S35-Exon2_Client	59,545	0.04 Gb
	S36-Exon2_Client	55,090	0.03 Gb
	S5-Exon2_Client	66,811	0.04 Gb
	S5-Exon3_AGRF	23,858	0.01 Gb
	S8-Exon2_AGRF	76,985	0.05 Gb
	S8-Exon2_Client	57,729	0.03 Gb

*Continued on next page*

Table 1 – Continued from previous page

Lane	Sample Name	Paired End	Data Yield(bp)
	S8-Exon3_AGRF	40,017	0.02 Gb
	S9-Exon2_AGRF	64,388	0.04 Gb
	S9-Exon2_Client	33,185	0.02 Gb
	S9-Exon3_AGRF	45,834	0.03 Gb
	T1-Exon2_AGRF	56,754	0.03 Gb
	T1-Exon2_Client	57,574	0.03 Gb
	T1-Exon3_AGRF	39,153	0.02 Gb
	T16-Exon2_AGRF	57,766	0.03 Gb
	T16-Exon2_Client	56,464	0.03 Gb
	T16-Exon3_AGRF	39,560	0.02 Gb
	T18-Exon2_AGRF	24,751	0.01 Gb
	T18-Exon2_Client	40,418	0.02 Gb
	T18-Exon3_AGRF	47,956	0.03 Gb
	T21-Exon2_AGRF	37,823	0.02 Gb
	T21-Exon2_Client	55,274	0.03 Gb
	T21-Exon3_AGRF	30,780	0.02 Gb
	T24-Exon2_AGRF	9,098	0.01 Gb
	T24-Exon2_Client	13,951	0.01 Gb
	T24-Exon3_AGRF	46,877	0.03 Gb
	T27-Exon2_AGRF	134,331	0.08 Gb
	T27-Exon2_Client	90,121	0.05 Gb
	T27-Exon3_AGRF	75,553	0.05 Gb
	T29-Exon2_AGRF	74,381	0.04 Gb
	T29-Exon2_Client	63,468	0.04 Gb
	T29-Exon3_AGRF	44,027	0.03 Gb
	T36-Exon2_AGRF	157,833	0.10 Gb
	T36-Exon2_Client	52,155	0.03 Gb
	T36-Exon3_AGRF	61,970	0.04 Gb
	T38-Exon2_Client	34,497	0.02 Gb
	T5-Exon2_AGRF	121,620	0.07 Gb
	T5-Exon2_Client	67,193	0.04 Gb
	T5-Exon3_AGRF	68,175	0.04 Gb
	T8-Exon2_AGRF	54,609	0.03 Gb
	T8-Exon2_Client	54,194	0.03 Gb
	T8-Exon3_AGRF	35,986	0.02 Gb
<b>Total</b>		<b>7,563,750</b>	<b>4.55Gb</b>

## 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 300 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 **23rd of October, 2019**, after that date they will be removed from the server. Files will be stored at the AGRF until the 1st of January, 2020. After that date the FASTQ files will be able to be recreated, until the 19th 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](mailto: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.