

## GENEWIZ PacBio Data Analysis Report

**Customer:** [REDACTED]

**Institute/Company:** United States Bureau of Reclamation

**Project ID:** [REDACTED]

**NGS Data:** BAM files from PacBio Sequel

**Bioinformatics Service:** Circular Consensus Sequences

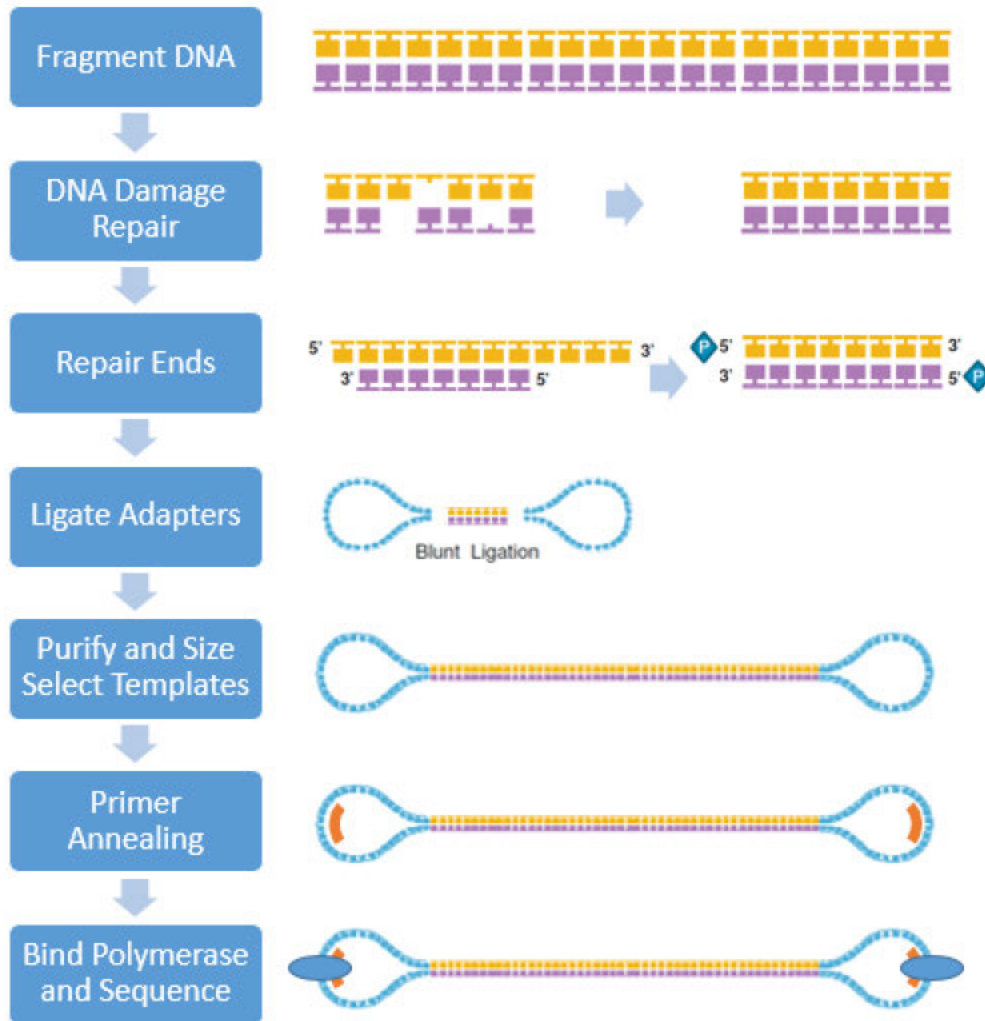
**Date:** January 2<sup>nd</sup>, 2019

**GENEWIZ Contact :** Phone: 908-222-0711

**Email:** [ngs@genewiz.com](mailto:ngs@genewiz.com)

# 1. Description of Workflow

## 1.1 PacBio Sequel Sequencing Workflow



## 1.2 PacBio Circular Consensus Sequence (CCS) Workflow

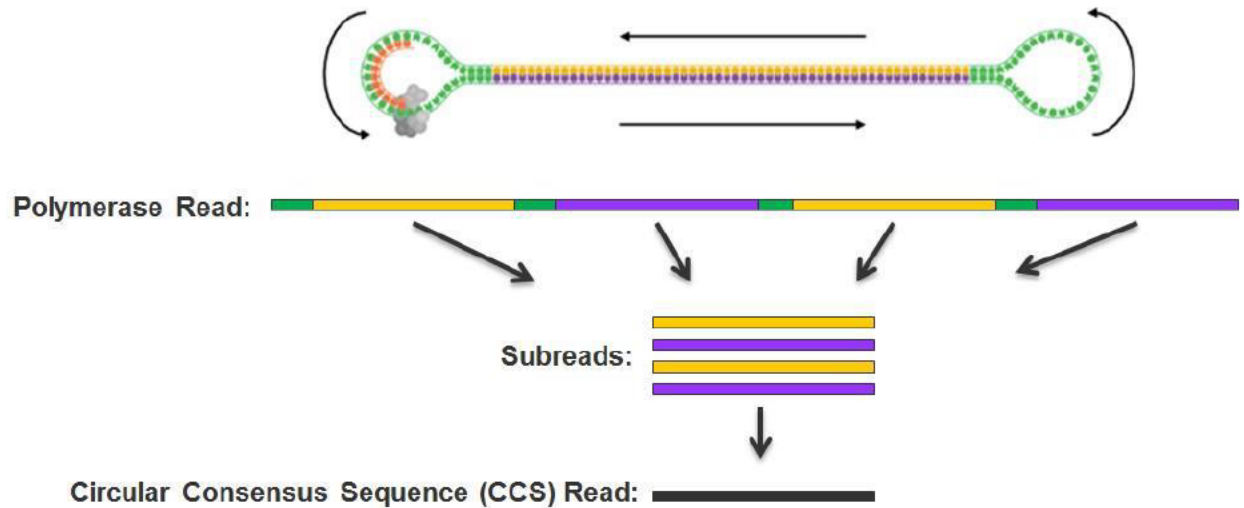


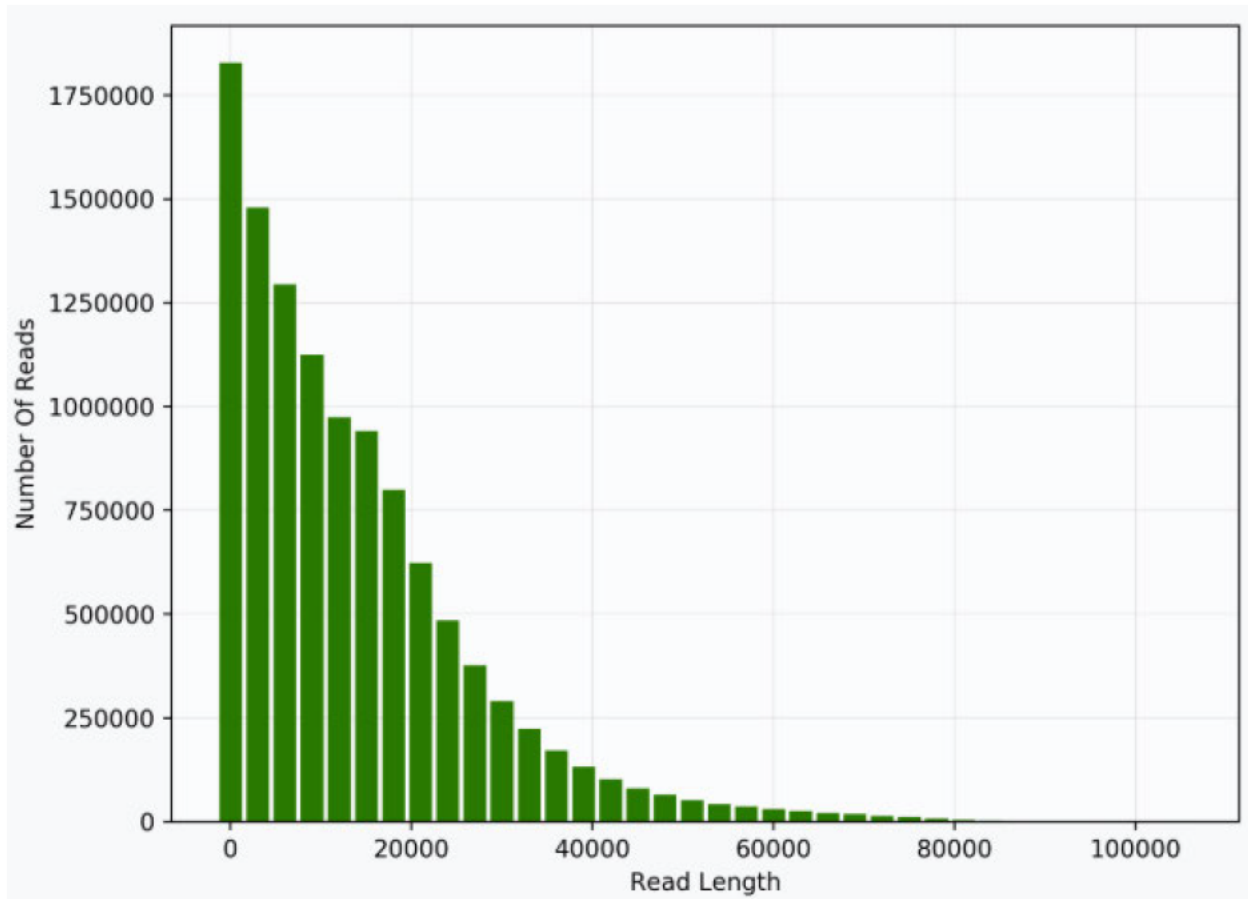
Figure 1.2 Bioinformatics analysis workflow.

## 2. Data Overview

### 2.1 Polymerase Read Statistics

Cell	Polymerase Read Bases	Polymerase Reads	Polymerase Read Length (mean)	Polymerase Read N50	Subread Length (mean)	Subread N50	Insert Length (mean)	Insert N50
1	8,014,503,036	484,137	16,554	24,840	14,007	20,582	14,795	21,479
2	8,421,852,259	497,891	16,915	25,425	14,226	20,862	15,038	21,784
3	8,202,465,181	503,275	16,298	24,419	13,849	20,358	14,608	21,195
4	5,703,531,282	338,691	16,840	24,766	14,114	20,693	15,099	21,544
5	7,520,847,101	516,333	14,566	22,402	12,886	19,588	13,462	20,207
6	7,839,158,418	514,199	15,245	22,903	13,379	19,994	14,052	20,650
7	8,868,031,605	574,605	15,433	23,825	13,209	19,887	13,851	20,675
8	8,341,878,158	577,831	14,437	22,654	12,590	19,354	13,137	20,043
9	7,877,789,917	544,887	14,458	21,979	12,757	19,314	13,351	19,899
10	8,082,488,498	552,858	14,619	22,484	12,900	19,660	13,502	20,313
11	10,311,998,305	683,128	15,095	23,945	12,762	19,580	13,381	20,412
12	8,825,030,551	661,464	13,342	21,174	11,812	18,556	12,291	19,096
13	5,684,563,289	416,797	13,639	20,747	12,115	18,653	12,716	19,106
14	5,159,222,836	380,335	13,565	20,702	11,951	18,608	12,640	19,073
15	4,278,688,589	314,595	13,601	20,744	12,071	18,773	12,743	19,197
16	5,197,387,638	395,600	13,138	20,241	11,497	18,211	12,244	18,647
17	5,461,902,736	393,454	13,882	20,982	12,183	18,722	12,879	19,198
18	8,683,415,262	579,120	14,994	23,024	13,043	19,714	13,646	20,384
19	8,392,786,196	585,826	14,326	22,163	12,552	19,151	13,084	19,745
20	8,533,113,852	574,970	14,841	22,389	13,173	19,756	13,819	20,385
21	8,458,659,422	596,286	14,186	22,394	12,528	19,389	13,016	19,998
22	7,985,443,493	512,943	15,568	23,875	13,326	20,144	14,054	20,901

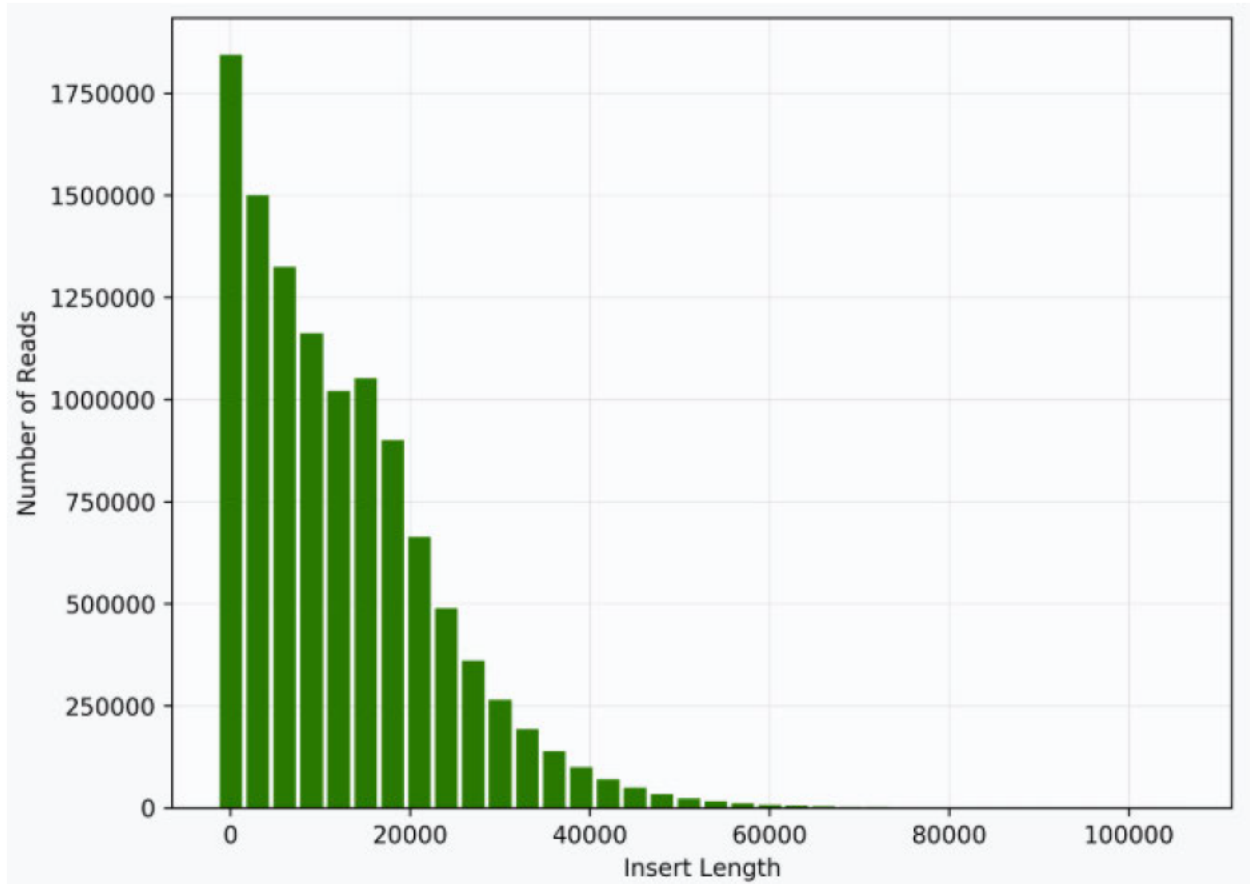
## 2.2 Polymerase Read Length Distribution



**Figure 2.2:** Distribution of Polymerase read lengths. X-axis: Read lengths in intervals of 20Kb. Y-axis: Number of reads in intervals of 5,000

Cell1 shown above and is representative of all cells in this project. Remaining cells included with deliverables.

## 2.3 Insert Length Distribution



**Figure 2.3:** Distribution of Insert sizes. X-axis: Insert lengths in intervals of 20Kb. Y-axis: Number of Reads in intervals of 50,000

### 3. Circular Consensus Sequences

PacBio SMRTLink v.6.0 was used to obtain CCS reads from the demultiplexed bam files using the CCS2 algorithm. Raw reads with a minimum number of passes greater than or equal to 3 were used to generate the CCS reads.

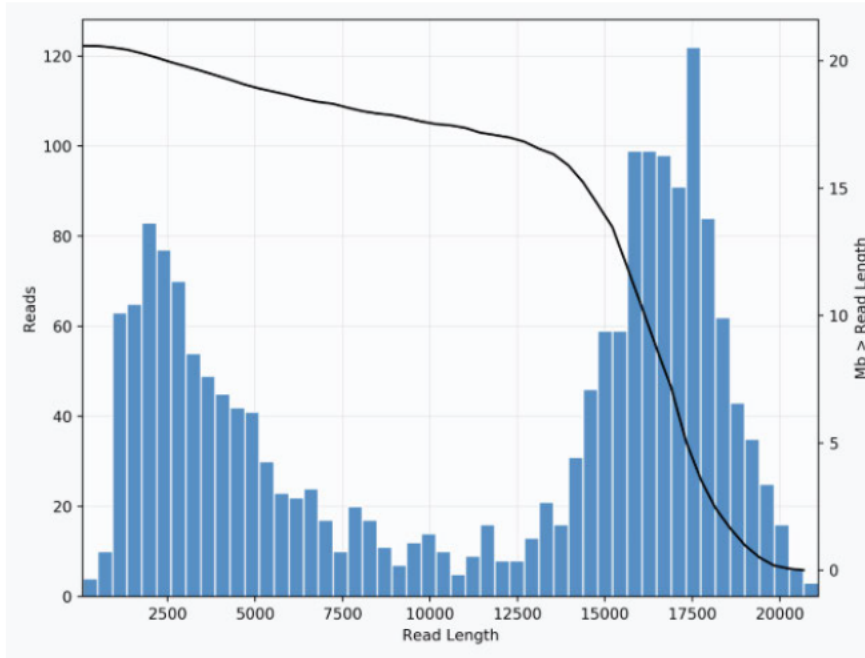
#### 3.1 CCS read overview:

##### 3.1.1 CCS Read Statistics:

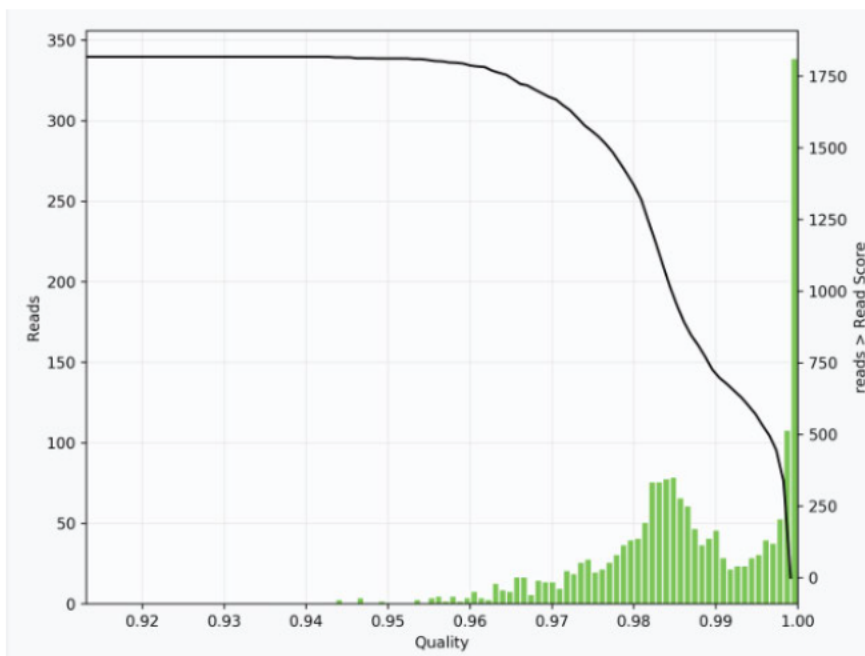
Cell	CCS reads	Number of CCS bases	CCS Read Length (Mean)	CCS Read Score (Mean)	Number of Passes (Mean)
1	1,894	20,995,168	11,085	0.986	8
2	1,978	21,667,170	10,954	0.987	8
3	1,843	19,833,923	10,761	0.986	8
4	1,622	15,391,344	9,489	0.988	10
5	1,215	10,775,988	8,869	0.988	8
6	1,383	13,234,931	9,569	0.988	9
7	1,971	20,579,071	10,440	0.987	8
8	1,494	15,258,280	10,213	0.987	8
9	1,426	13,302,623	9,328	0.987	9
10	1,328	11,770,422	8,863	0.988	9
11	2,456	28,136,581	11,456	0.985	7
12	1,621	14,398,551	8,882	0.987	9
13	1,268	10,124,968	7,984	0.99	11
14	1,199	7,248,936	6,045	0.992	14
15	962	6,381,178	6,633	0.992	14
16	1,307	7,628,526	5,836	0.991	15
17	1,237	8,509,161	6,878	0.992	14
18	1,534	13,651,301	8,899	0.988	9
19	1,508	13,136,594	8,711	0.988	9
20	1,051	8,081,706	7,689	0.989	10
21	1,228	11,605,057	9,450	0.986	8
22	1,825	16,705,071	9,153	0.989	10

\*Below are images for SMRT cell 1. Images for all cells are included as deliverables

### 3.1.2 CCS Read Length:

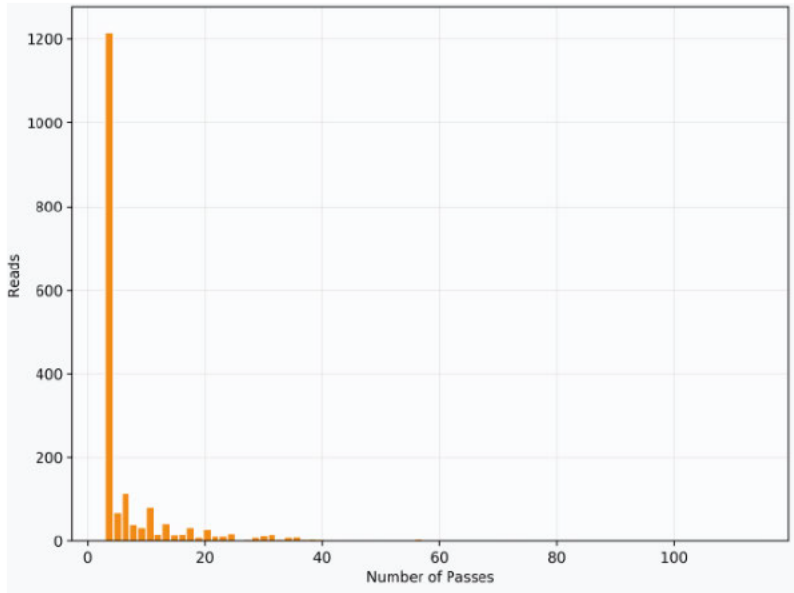


### 3.1.3 CCS Read Score:

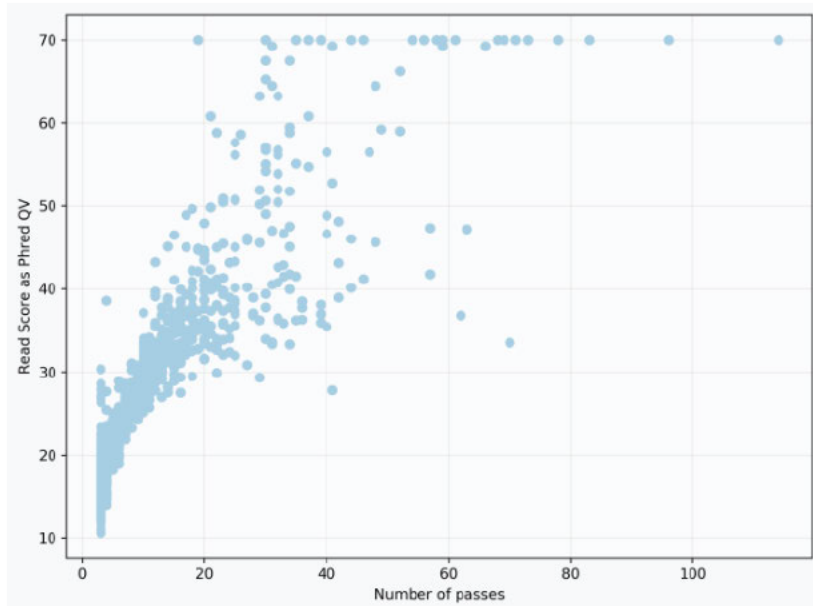




### 3.1.4 Number of Passes:



### 3.1.5 Number of Passes vs Read Score:



## 4. Summary

- Starting from an amplicon provided by the customer, PacBio SMRTbell libraries were prepared per the manufacturer's instructions.
- SMRTbell libraries were sequenced on the PacBio Sequel platform with v2.0 chemistry.
- The subreads were demultiplexed and CCS reads were obtained using CCS2 within the SMRTLink suite. The algorithm was run using the default parameters.

## 5. Deliverables

- Data Analysis Report.
- Raw BAM output and supporting files.
- Fasta and Fastq files with CCS reads for each SMRT cell.
- CCS statistics image files.