How to assess an Ion S5™ or Ion GeneStudio™ S5 sequencing run report

Pub. No. MAN0017983 **Rev.** A.0

This user bulletin describes an approach to evaluate an Ion S5™ or Ion GeneStudio™ S5 sequencing run report. Although each sequencing run should be evaluated as a whole, there are individual metrics that can help you assess a sequencing run. This document provides an overall view of expected run performance.

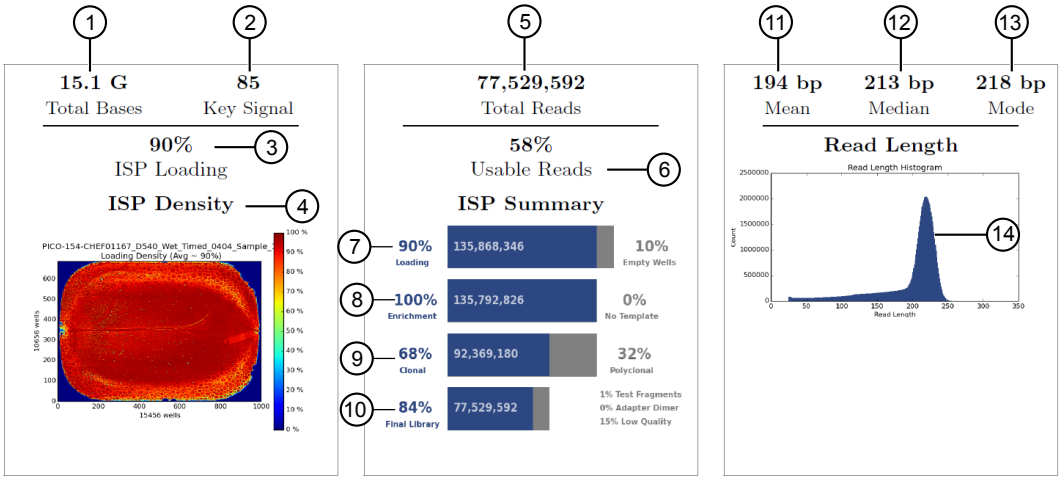
For cases that involve further help, contact Thermo Fisher Scientific Technical Support at **www.thermofisher.com/support**, or your contract support personnel.

An example run report is included and used in examples throughout this guide. The example run report is from a 200-base read run (500 flows) on an Ion 540™ Chip with a CEPH control library that is templated on an Ion Chef™ Instrument. For more information, see "Ion 540™ run report example" on page 9.

Torrent Suite™ Software 5.10 is used to view the report and report metrics throughout this guide.

Run report metrics before alignment

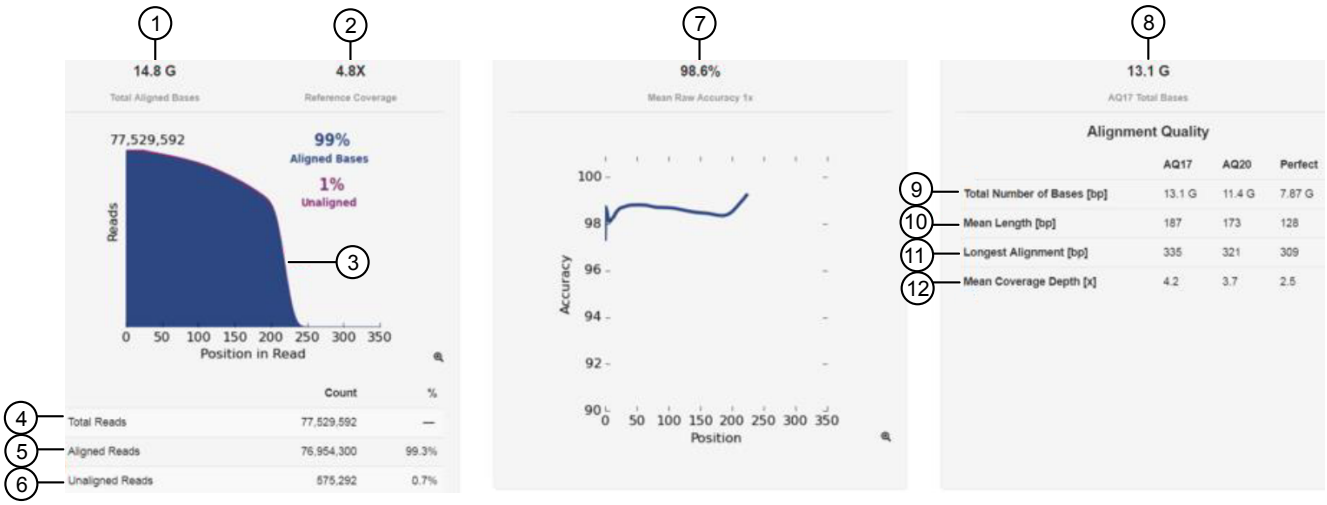
To assess the sequencing run, you can review the quality metrics for the unaligned reads. Primary pipeline processing, basecalling, and signal processing generate these metrics. The following prealignment metrics are provided in the run report. Access the run report in the Torrent Suite™ Software using the **Data** tab.



Number	Metric	Description
1	Total Bases	The number of filtered and trimmed base pairs reported in the output BAM file.
2	Key Signal	The average signal for all library ISPs with library key (TCAG).
3	ISP Loading	The percentage of chip wells that contain an Ion Sphere Particle (ISP). The percentage value considers only addressable wells.
4	ISP Loading Density	A visual representation of well loading distribution on the chip surface. Red color indicates areas of high loading and blue indicates areas of low loading.
5	Total Reads	The total number of filtered and trimmed reads independent of length reported in the output BAM file.
6	Usable Reads	The percentage of library ISPs that pass the polyclonal, low quality, and primer-dimer filters. This percentage is calculated by dividing final library ISPs by library ISPs.
7	ISP Summary-Loading	The percentage of chip wells that contain an ISP. The percentage value considers addressable wells.
8	ISP Summary-Enrichment	The predicted number of live ISPs that have a key signal identical to the library key signal or test fragment (TF) key signal. The Percent Enrichment value reported is the number of loaded wells with live ISPs that are Library ISPs or TF ISPs. This number is calculated by dividing wells with live ISPs by the number of wells loaded with ISPs.
9	ISP Summary-Clonality	The percentage of clonal ISPs (all library and TF ISPs that are clonal, not polyclonal). An ISP is clonal if all of its DNA fragments are cloned from a single original template. All the fragments on such an ISP are identical and they respond in unison as each nucleotide is flowed in turn across the chip. This percentage is calculated by dividing the number of ISPs with a single DNA template by the number of live wells.
10	ISP Summary-Final Library	The percentage of reads, which pass all filters, and which are recorded in the output BAM file. This value can be different from the Total Reads due to technicalities associated with read trimming beyond a minimal requirement that results in Total Reads being slightly less than Final Library.
11	Mean Read Length	The average length, in base pairs, of called reads.
12	Median Read Length	The median length, in base pairs, of called reads.
13	Mode Read Length	The mode length, in base pairs, of called reads.
14	Read Length Histogram	The read length histogram is a histogram of the trimmed lengths of all reads present in the output files.

Run report metrics after alignment

To assess the sequencing run when an accurate reference is available, you can review the quality metrics for aligned reads. Reads are aligned to a reference genome sequence. The following post-alignment metrics are provided in the run report.



Number	Metric	Description
1	Total Aligned Bases	The number of filtered and trimmed aligned base pairs reported in the output BAM file that are aligned to the reference sequence, excluding the library key, barcodes, and 3' adapter sequences.
2	Reference Coverage	The ratio of the total aligned bases divided by the number of bases in the reference sequence. Reference coverage does not account for enrichment done to selectively amplify a subset of the reference sequence.
3	Alignment plot	A plot of the number of aligned reads (blue) and unaligned (purple) by position in an aligned sequence.
4	Total Reads	The total number of reads after filtering.
5	Aligned Reads	The number of reads that align to the reference sequence expressed as a total count and percentage of the total aligned reads.
6	Unaligned Reads	The number of reads that do not align to the reference sequence expressed as a total count and percentage of the total reads.
7	Mean Raw Accuracy 1x	The mean raw accuracy across each individual base position in a read calculated as 1– (total errors in the sequenced base) / total bases sequenced.
8	AQ17 Total Bases	The total number of bases over all positions that align with an error rate of 2% or less.
9	Total Number of Bases (bp)	The total number of bases over all positions that align with a given error rate. (AQ17 ≤2% error rate, AQ20 ≤1% error rate, Perfect = no measurable error)
10	Mean Length (bp)	The average length, in base pairs, for aligned reads at a given error rate. (AQ17 ≤2% error rate, AQ20 ≤1% error rate, Perfect = no measurable error)
11	Longest Alignment (bp)	The maximum sequence read length for a given error rate. (AQ17 ≤2% error rate, AQ20 ≤1% error rate, Perfect = no measurable error)
12	Mean Coverage Depth (x)	The ratio of the total aligned bases at a given error rate to the size of the target region. (AQ17 ≤2% error rate, AQ20 ≤1% error rate, Perfect = no measurable error)

Throughput considerations

To assess a run, examine the throughput or total bases. Consider whether these numbers make sense for the application and chip type.

The table provides information regarding the number of wells per chip type and sequencing throughput specifications for sequencing runs from an Ion GeneStudio™ S5 System or Ion S5™ System.

Chip type	Number of addressable wells	Number of reads	Throughput	
			200 Base Read	400 Base Read
Ion 510™ Chip	~6 million	2–3 million	0.3–0.5 Gb	0.6–1 Gb
Ion 520™ Chip	~12 million	4–6 million	0.6–1 Gb	1.2–2 Gb
Ion 530™ Chip	~37 million	15–20 million	3–4 Gb	6–8 Gb
Ion 540™ Chip	~150 million	60–80 million	10–15 Gb	N/A
Ion 550™ Chip	~260 million	100–130 million	20–25 Gb	N/A

For more information, see the Ion S5™ System Specification Sheet and the Ion GeneStudio™ S5 System Next-Generation Sequencing Series Specifications available on **thermofisher.com/ngsresources**.

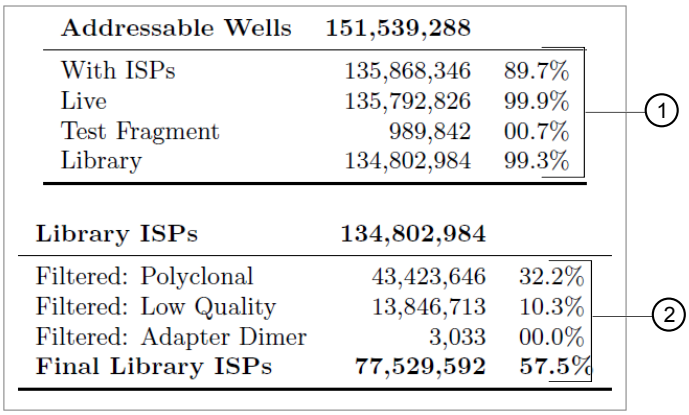
Read filtering and trimming

Empty and loaded wells are separated by the differences in buffering and signal over the chip during the nucleotide key flows (flows 1–8). Wells that are loaded with ISPs and associated polymerase have greater buffering capacity and higher signal than empty wells, and Torrent Suite™ Software uses these differences to identify and classify loaded versus empty wells.

After well classification, the software further processes the identified test fragments and library reads, including read filtering and trimming. This processing affects the total number of library reads and bases. You can see both well classification and library read filtering results that are displayed in the run report. Read trimming operations trim bases off the read, thereby making reads shorter, while read filtering operations completely remove them from the output BAM files. By default, reads that have a trimmed read length of less than 25 bases are being filtered. The different categories of filtered reads shown in the run report are in the table below.

Filter	Description
Polyclonal	Filters reads from ISPs with >1 unique library template population. Occasionally, low or unexpected signal ISPs can also get caught in this filter.
Low Quality	Filters reads with unrecognizable key signal, low signal quality, and reads trimmed to <25 bases.
Primer Dimer	Filters reads where no or only a very short sequencing insert is present. Reads that, after P1 adapter trimming, have a trimmed length of <25 bases are considered primer dimers.

You can see both well classification and library read filtering results that are displayed in the run report.

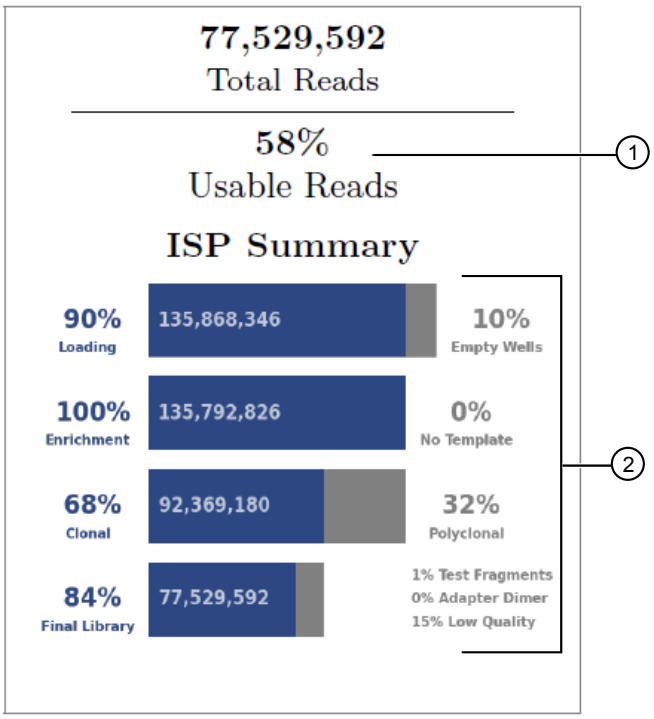


Addressable Wells	**151,539,288**	
With ISPs	135,868,346	89.7%
Live	135,792,826	99.9%
Test Fragment	989,842	00.7%
Library	134,802,984	99.3%
Library ISPs	**134,802,984**	
Filtered: Polyclonal	43,423,646	32.2%
Filtered: Low Quality	13,846,713	10.3%
Filtered: Adapter Dimer	3,033	00.0%
Final Library ISPs	**77,529,592**	**57.5%**

① Well classification results

② Library ISP filtering results

The ISP summary panel also presents the well classification and library ISP summary table using slightly different calculations.



① Usable reads percentage

② ISP summary percentages

The usable reads percentage is calculated by

$$\% \text{ usable reads} = \frac{\text{library reads passing filters (77,529,592) x 100}}{\text{number of library ISPs identified (134,802,984)}}$$

Each ISP summary percentage is calculated by dividing the current value by the previous values. For example,

$$\% \text{ final library} = \frac{\text{library reads passing filters (77,529,592) x 100}}{\text{number of clonal reads (92, 369,180)}}$$

ISP loading density

ISP loading density is the percentage of chip wells that contain an Ion Sphere™ Particle (ISP, templated and non-templated, or live and dud ISPs). This percentage value considers the addressable wells and is a result of the software well classification step.

What can affect ISP loading density?

ISP loading density can be affected by template preparation on the Ion Chef™, Ion OneTouch™ 2, or Ion OneTouch™ ES instruments. Bubbles and missing chip blocks or tiles can also affect ISP loading density.

Factors that can affect loading density.

Instrument	Factors
Ion Chef™ Instrument	Various template preparation problems on an Ion Chef™ Instrument can cause poor ISP loading density, including: • Low enriched ISP recovery that is caused by low library input. • High library input that can cause overloading, poor template and sequencing quality of poor enrichment efficiency. • Improper chip adapter placement or removal. • Air bubbles in reagent cartridge can affect various steps throughout the run including loading. **Note:** Sample-to-sample differences on an Ion Chef™ Instrument are normal, as long as the read counts meet throughput requirements as defined in "Throughput considerations" on page 4.
Ion OneTouch™ 2 System	A number of template preparation problems on an Ion OneTouch™ 2 System can cause poor ISP loading density on sequencing runs, including: • Poor ISP recovery from the Ion OneTouch™ ES Instrument. • Inadequate resuspension of ISPs before addition of primer. • Missed reagents such as sequencing primer or sequencing polymerase.

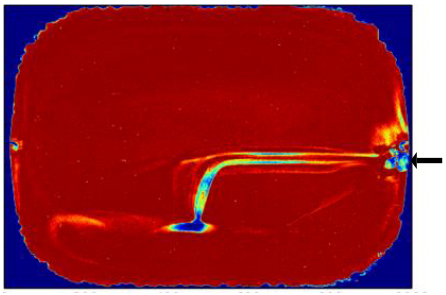
Variable images

Variability between ISP loading density images is expected and is not always indicative of a sequencing problem. If throughput meets threshold requirements, the sequencing results are usable, regardless of the appearance of the chip loading image. The following are examples of variation in chip loading images.

- Phenotypes of small bubbles in center or at edges of the chip
- Streaks across the middle of the chip
- Missing chip blocks or tiles

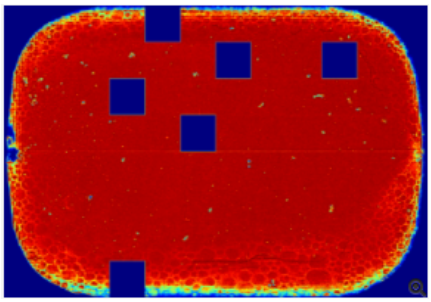
Bubbles

If you see bubbles in your ISP loading density image, ensure that reagent handling procedures are being followed to reduce bubbles in future runs. Bubbles can be introduced during chip loading or sequencing. During loading this can prevent ISPs from depositing into wells and during sequencing it can interfere with reagent flow across the chip. Small bubbles do not typically affect throughput or performance.



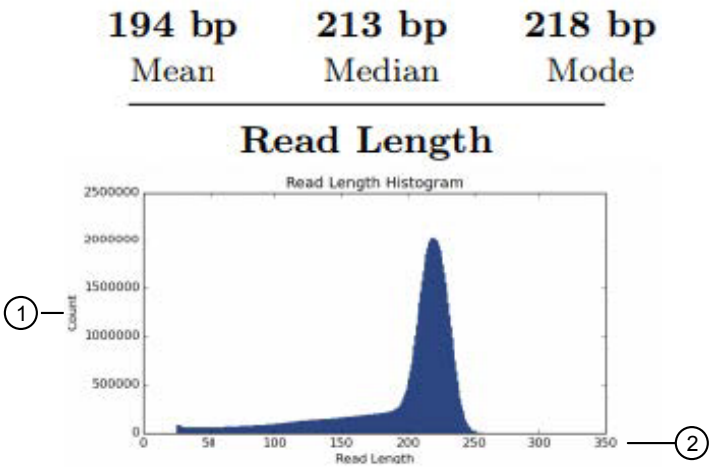
Missing chip blocks or tiles

If you see an ISP loading density image with missing blocks but good sequencing quality appearance, this could indicate that something might have gone wrong on the sequencer analysis, data transfer, or the analysis on the server. Usually, this does not impact the quality of the data. If the problem repeats, contact Technical Support or your support representative.



Read length

Read length value is the length of called reads measured in base pairs. The Read Length histogram presents all filtered and trimmed library reads reported in the output BAM file and the mean read length in base pairs. The shape of the histogram should closely resemble the library size distribution trace, without the adapter sequences.



① The y-axis provides the read count.

② The x-axis provides the read length in base pairs.

In addition to the loading density and read filtering and trimming, the average library read length also affects the total sequencing run throughput. Read length is considered in the total sequencing run throughput. For example, a sequencing run on an Ion 530™ chip produces about 15 million final library reads:

- If the average read length is 300 bp, then the approximate throughput is 15,000,000 reads * 300 bp = 4.5 Gbp.
- If the average read length is 400 bp, then the approximate throughput is 15,000,000 * 400 bp = 6 Gbp.

Read length histogram interpretation

The read length histogram reflects the library fragment lengths after read filtering and 3' quality and adapter trimming. In ideal cases, the read length histogram mimics the library size distribution minus the Ion A and P1 adapter sequences. For 200-base read libraries, the read length histogram should have a peak near 200 bp. Shown is the Ion Torrent Control library, fragmented DNA library with average size of 290 bp including adapter sequences.

A small degree of 3' quality trimming is normal for most samples. When there is a significant amount of 3' quality trimming, it usually indicates a problem with the template preparation or sequencing run quality. To help evaluate the root cause, start by checking the test fragment (TF) performance.

- If TF performance is as expected (key signal, %50AQ17), then investigate library preparation. To prevent problems, ensure that the library size is not larger than recommended for the template kit that is used, and ensure that the library is properly prepared (correct Ion adapter sequences). To confirm that the library input amount is appropriate the Ion Sphere™ Quality Control Kit, Cat. No. 4468656, can be used.
- If TF performance is below expectations/normal observations, then investigate the template preparation and sequencing. Ensure proper template reaction and sequencing reagent. For assistance download the sequencing run Customer Support Archive (CSA) file and contact Technical Support.

Note: To evaluate if the shorter-than-expected reads are a result of the library construction (true short fragments in the library) or a result of the 3′ quality trimming, the sequencing run can be reanalyzed with the 3′ quality trimming turned off.

Quality of library read trimming

The purpose of read trimming is to remove nontemplate bases at the 5' and 3' end of the read, as well as potential low-quality bases on the 3' end of the read.

- Trimmed nontemplate bases at the 5' end of reads include the key sequence, barcode, and barcode adapter sequence.
- Trimmed nontemplate bases at the 3' end of reads include the adapter that binds the library templates to the ISPs, the end barcode, and the end barcode adapter sequence.
- Low-quality base calls at the 3' end of reads are trimmed based on the base quality values.
- Reads that are filtered out entirely fall into the *low quality* filtering category.

Total reads

The total reads value is the total number of reads that are written to barcode or no-match output BAM files. Filtered reads are not included in this count.

Ion 540™ run report example

This run report example is from a 200-base read run (500 flows) on an Ion 540™ chip with a CEPH control library templated on an Ion Chef™ Instrument.

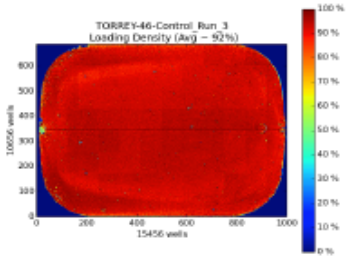
Run Report for Auto_user_TORREY-46-Control_Run_3_2147

Run Summary

17.3 G	83
Total Bases	Key Signal

92%
ISP Loading

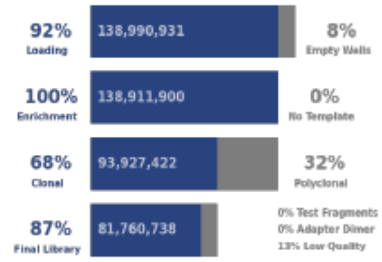
ISP Density



81,760,738
Total Reads

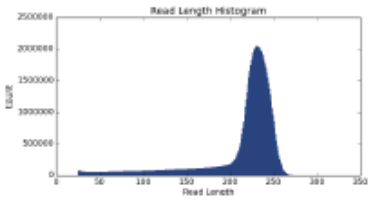
59%
Usable Reads

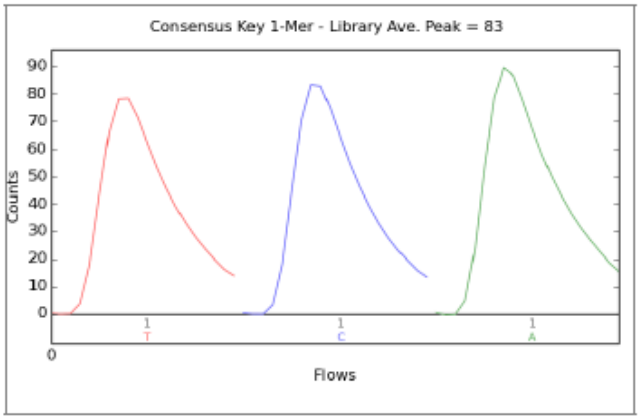
ISP Summary



211 bp	228 bp	230 bp
Mean	Median	Mode

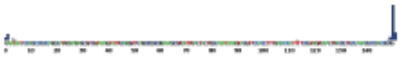
Read Length





Addressable Wells	151,539,288	
With ISPs	138,990,931	91.7%
Live	138,911,900	99.9%
Test Fragment	459,014	00.3%
Library	138,452,886	99.7%

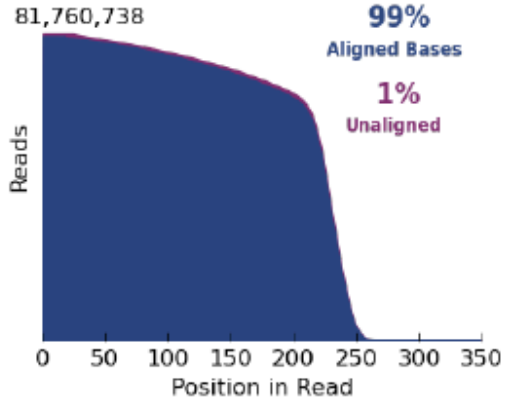
Library ISPs	138,452,886	
Filtered: Polyclonal	44,984,478	32.5%
Filtered: Low Quality	11,705,299	08.5%
Filtered: Adapter Dimer	2,371	00.0%
Final Library ISPs	**81,760,738**	**59.1%**

Test Fragment	Reads	Percent 50AQ17	Read Length Histogram
TF_1	**183,624**	**81**	

Alignment Summary *(aligned to hg19 from zip)*

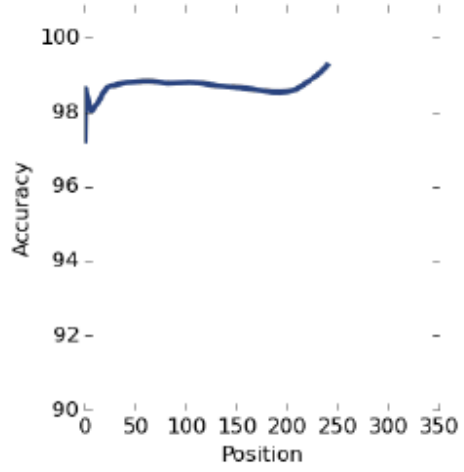
17.1 G
Total Alignment Bases

5.5X
Average Coverage Depth of Reference



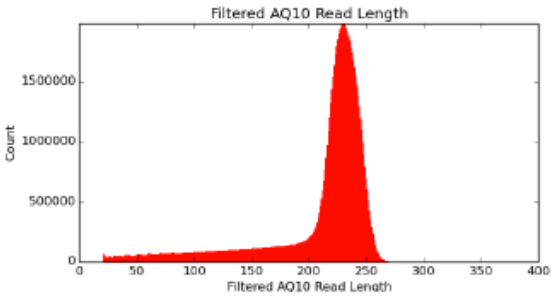
	Count	%
Total Reads	81,760,738	–
Aligned Reads	81,207,699	99.3%
Unaligned Reads	553,039	0.7%

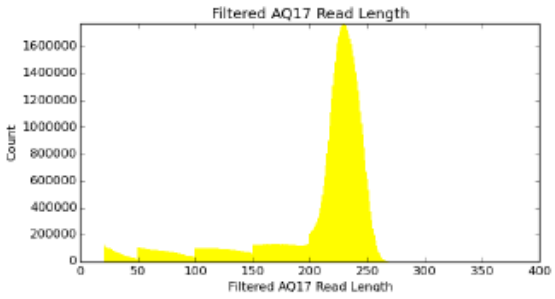
98.7%
Mean Raw Accuracy 1x

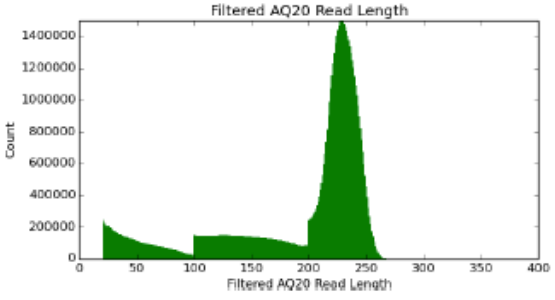


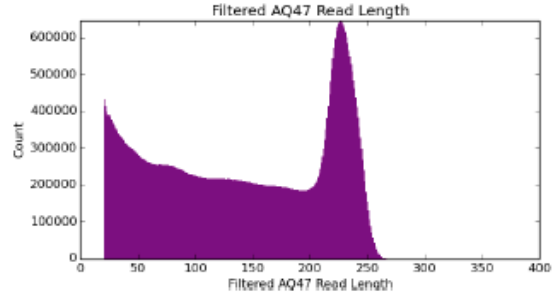
Alignment Quality

	AQ17	AQ20	Perfect
Total Number of Bases [Mbp]	15.2 G	13.4 G	9.17 G
Mean Length [bp]	205	192	142
Longest Alignment [bp]	355	350	334
Mean Coverage Depth	4.9	4.3	3.0









coverageAnalysis

Library type: Generic Sequencing
Read filters: Non-duplicate

Number of mapped reads	81,207,699
Average base coverage depth	5.483
Uniformity of base coverage	88.29%

Analysis Details

Run Name	R_2019_04_30_11_01_09_user_TORREY-46-Control_Run_3
Run Date	April 30, 2019, 11:03 a.m.
Run Flows	500
Projects	GLAS
Sample	Sample_1
Reference	
Instrument	TORREY
Operation Mode	Customer mode
Flow Order	TACGTACGTCTGAGCATCGATCGATGTACAGC
Library Key	TCAG
TF Key	ATCG
Chip Barcode	DAEJ00517
Chip Check	Passed
Chip Type	540
Chip Data	tiled
Chip Lot Number	QVS138
Chip Wafer	12
Barcode Set	
Analysis Name	Auto_user_TORREY-46-Control_Run_3_2147
Analysis Date	April 30, 2019, 10:03 p.m.
Analysis Flows	0
runID	PR4GR
BeadFind Args	justBeadFind –args-json /opt/ion/config/args_540_beadfind.json
Analysis Args	Analysis –args-json /opt/ion/config/args_540_analysis.json
Pre-BaseCaller	BaseCaller –barcode-filter-minreads 10 –phasing-residual-filter=2.0 –max-phasing-levels 2 –wells-normalization on
Calibration Args	Calibration
BaseCaller Args	BaseCaller –barcode-filter-minreads 10 –phasing-residual-filter=2.0 –max-phasing-levels 2 –num-unfiltered 1000 –barcode-filter-postpone 1 –wells-normalization on
Alignment Args	tmap mapall -q 50000 ... stage1 map4
IonStats Args	ionstats alignment
Analysis Parameters	default

Chef Summary

Chef Template Prep Information:

Chef Last Updated	April 30, 2019, 5:31 a.m.
Chef Instrument Name	242471368
Chef Operation Mode	Customer Mode
Sample Position	1
Tip Rack Barcode	48B020136
Chip Type 1	540v1
Chip Type 2	540v1
Chip Expiration 1	None
Chip Expiration 2	None
Templating Kit Type	Ion 540 Kit-Chef
Chef Flexible Workflow	
Reagent Expiration	200229
Reagent Lot Number	2065571
Reagent Part Number	A27758C
Reagent Cartridge Serial Number	None
Solution Lot Number	2065570
Solution Part Number	A27754C
Templating Protocol Planned	(use instrument default)
Solution Cartridge Serial Number	None
Solution Expiration	200331
Templating Protocol Executed	(use instrument default)
Chef Script Version	803
Chef Package Version	IC.5.10.0
Start Time	April 29, 2019, 3:13 p.m.
Completion Time	April 30, 2019, 5:19 a.m.

S5 Consumables Summary

Chip Type	540v1
Chip Barcode	DAEJ00517

Product Description	Part Number	Lot Number	Exp. Date	Remaining Uses
Ion S5 Cleaning Solution	100031096	2049959	2019/12/31	5
Ion S5 Sequencing Reagent	INS1012841B	2049947	2020/01/31	1
Ion S5 Wash Solution	100031091B	2044840	2019/12/31	1

Software Version

Torrent_Suite	5.10.1
host	sputnik04
ion-analysis	5.10.11-1
ion-gpu	5.10.0-1
ion-pipeline	5.10.10-1
ion-torrentpy	5.10.9-1
S5 Script	0.1.31
LiveView	2389
DataCollect	3606
OIA	51000
OS	35
Graphics	97
Ion_Chef	IC.5.10.0

Documentation and support

Customer and technical support

Visit **thermofisher.com/support** for the latest service and support information.

- Worldwide contact telephone numbers
- Product support information
  - Product FAQs
  - Software, patches, and updates
  - Training for many applications and instruments
- Order and web support
- Product documentation
  - User guides, manuals, and protocols
  - Certificates of Analysis
  - Safety Data Sheets (SDSs; also known as MSDSs)

    Note: For SDSs for reagents and chemicals from other manufacturers, contact the manufacturer.

Limited product warranty

Life Technologies Corporation and/or its affiliate(s) warrant their products as set forth in the Life Technologies' General Terms and Conditions of Sale at **www.thermofisher.com/us/en/home/global/terms-and-conditions.html**. If you have any questions, please contact Life Technologies at **www.thermofisher.com/support**.

Life Technologies Holdings Pte Ltd | Block 33 | Marsiling Industrial Estate Road 3 | #07-06, Singapore 739256
For descriptions of symbols on product labels or product documents, go to **thermofisher.com/symbols-definition**.

Revision history: Pub. No. MAN0017983

Revision	Date	Description
A.0	10 June 2019	New sequencing run assessment bulletin.