psomagen

Run Only

Report

2025.08

RAWDATA REPORT

Research use only

Table of Contents

Order Information

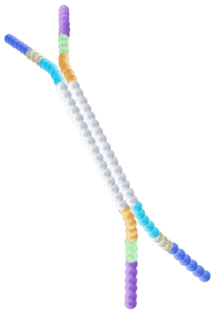
Client Name	Sam White
Client Organization	Univ. of Washington
Order Number	AN00025267
Application	Run Only
Type of Read	Paired-end
Read Length	151
Library Kit	Accel-NGS Methyl-Seq DNA Library Kit EZ DNA Methylation-Gold Kit (Zymo Research)
Library Protocol	Accel-NGS Methyl-Seq DNA Library Kit for Illumina Platforms
Type of Sequencer	Illumina platform

Experimental Workflow

The samples are prepared according to NGS library preparation workflow, and sequenced using Illumina platform.
The workflow illustrated below shows the common ligation based method of library preparation. The process may differ based on the library preparation protocol followed.

Sample Preparation

DNA/RNA is first extracted from the sample, and samples which meet quality control standards proceed to library construction.



Ligate Adapters

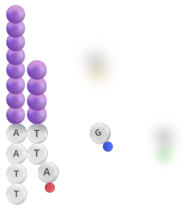
The sequencing library is prepared by random fragmentation of the DNA or cDNA sample, followed by 5' and 3' adapter ligation. Alternatively, "tagmentation" combines the fragmentation and ligation reactions into a single step which greatly increases the efficiency of the library preparation process.

Final library Construction

Adapter-ligated fragments are then PCR amplified with a PCR primer solution which anneals to the ends of each adapters.
The library templates undergo quality control and quantification process.

Cluster generation using bridge amplification

The library is loaded onto a flow cell where fragments are captured on a lawn of surface-bound oligos complementary to the library adapters.
Each fragment is then amplified into distinct clonal clusters through bridge amplification.
Once cluster generation is complete, the templates are ready for sequencing.



Sequencing by synthesis (SBS) technology

Illumina SBS technology utilizes a proprietary reversible terminator-based method that detects single bases as they are incorporated into DNA template strands. As all 4-reversible, terminator -bound dNTPs are present during each sequencing cycle, natural competition minimizes incorporation bias and greatly reduces raw error rates compared to other technologies.
The result is highly accurate base-by-base sequencing that virtually eliminates sequence-context-specific errors, even within repetitive sequence regions and homopolymers.

Generation of Raw data

The Illumina sequencer generates raw images utilizing sequencing control software for system control and base calling, through integrated primary analysis software called RTA (Real Time Analysis).
The BCL/cBCL (base call) binary files are converted into FASTQ files using bcl2fastq, which is an Illumina provided package. Adapters are not trimmed away from the reads.

Raw Data Statistics

- The total number of bases, reads, GC (%), Q20 (%), and Q30 (%) are calculated for the 32 samples. For example, in 01B sample, 109,565,892 reads are produced, and total read bases are 16.5 Gbp. The GC content (%) is 27.7% and Q30 is 92.5%.

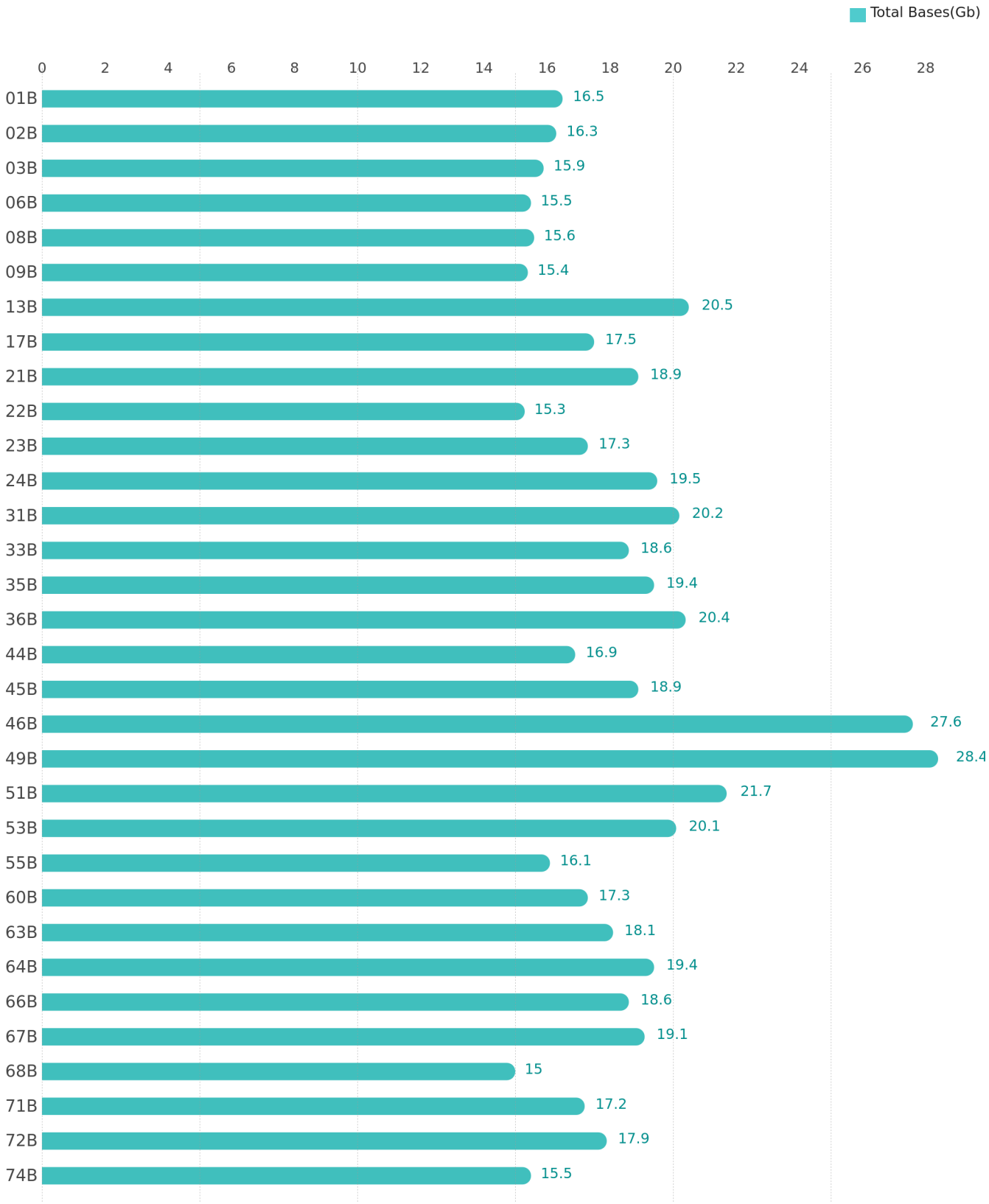
* Raw Data

Sample ID	Total bases(bp)	Total reads	GC(%)	AT(%)	Q20(%)	Q30(%)
01B	16,544,449,692	109,565,892	27.7	72.3	97.1	92.5
02B	16,332,423,948	108,161,748	27.7	72.3	97.1	92.5
03B	15,927,579,660	105,480,660	27.6	72.4	97.2	92.9
06B	15,500,589,108	102,652,908	27.6	72.4	97.2	93.0
08B	15,554,316,116	103,008,716	27.8	72.2	97.1	92.8
09B	15,363,618,820	101,745,820	27.5	72.5	97.1	92.9
13B	20,469,045,996	135,556,596	27.9	72.1	97.2	92.9
17B	17,526,984,042	116,072,742	27.7	72.3	97.1	92.5
21B	18,921,111,474	125,305,374	27.5	72.5	97.2	92.8
22B	15,255,094,214	101,027,114	27.7	72.3	97.2	92.7
23B	17,316,325,754	114,677,654	27.4	72.6	97.2	93.1
24B	19,478,749,340	128,998,340	27.5	72.5	97.4	93.5
31B	20,211,864,608	133,853,408	27.7	72.3	97.2	93.1
33B	18,571,341,114	122,989,014	27.5	72.5	97.4	93.4
35B	19,426,569,478	128,652,778	27.4	72.7	97.3	92.8
36B	20,395,834,250	135,071,750	27.6	72.4	97.3	93.2
44B	16,887,337,472	111,836,672	27.5	72.5	97.3	93.4
45B	18,939,209,428	125,425,228	27.9	72.2	97.2	92.6
46B	27,611,802,620	182,859,620	28.1	71.9	97.2	93.2
49B	28,357,033,524	187,794,924	27.6	72.5	97.2	93.1
51B	21,684,896,184	143,608,584	27.7	72.3	97.1	92.8
53B	20,067,387,204	132,896,604	27.6	72.4	97.3	93.3
55B	16,083,359,414	106,512,314	27.5	72.5	97.2	93.1
60B	17,344,634,630	114,865,130	27.8	72.2	97.2	92.7
63B	18,072,778,676	119,687,276	27.8	72.2	97.3	93.0
64B	19,371,601,854	128,288,754	27.8	72.2	97.2	93.0
66B	18,601,507,592	123,188,792	27.8	72.2	97.3	93.0
67B	19,061,075,186	126,232,286	27.7	72.3	97.2	92.6
68B	15,019,429,420	99,466,420	27.6	72.4	97.3	93.4
71B	17,191,880,614	113,853,514	27.6	72.4	97.3	93.4
72B	17,857,451,468	118,261,268	27.5	72.5	97.3	93.3
74B	15,506,326,202	102,690,902	27.6	72.4	97.2	92.9

- Sample ID : Sample name.
- Total bases(bp) : Total number of bases sequenced.
- Total reads : Total number of reads. For illumina paired-end sequencing, this value refers to the sum of read1 and read2.
- GC(%) : Ratio of GC content.
- AT(%) : Ratio of AT content.
- Q20(%) : Ratio of bases that have phred quality score of over 20.
- Q30(%) : Ratio of bases that have phred quality score of over 30.

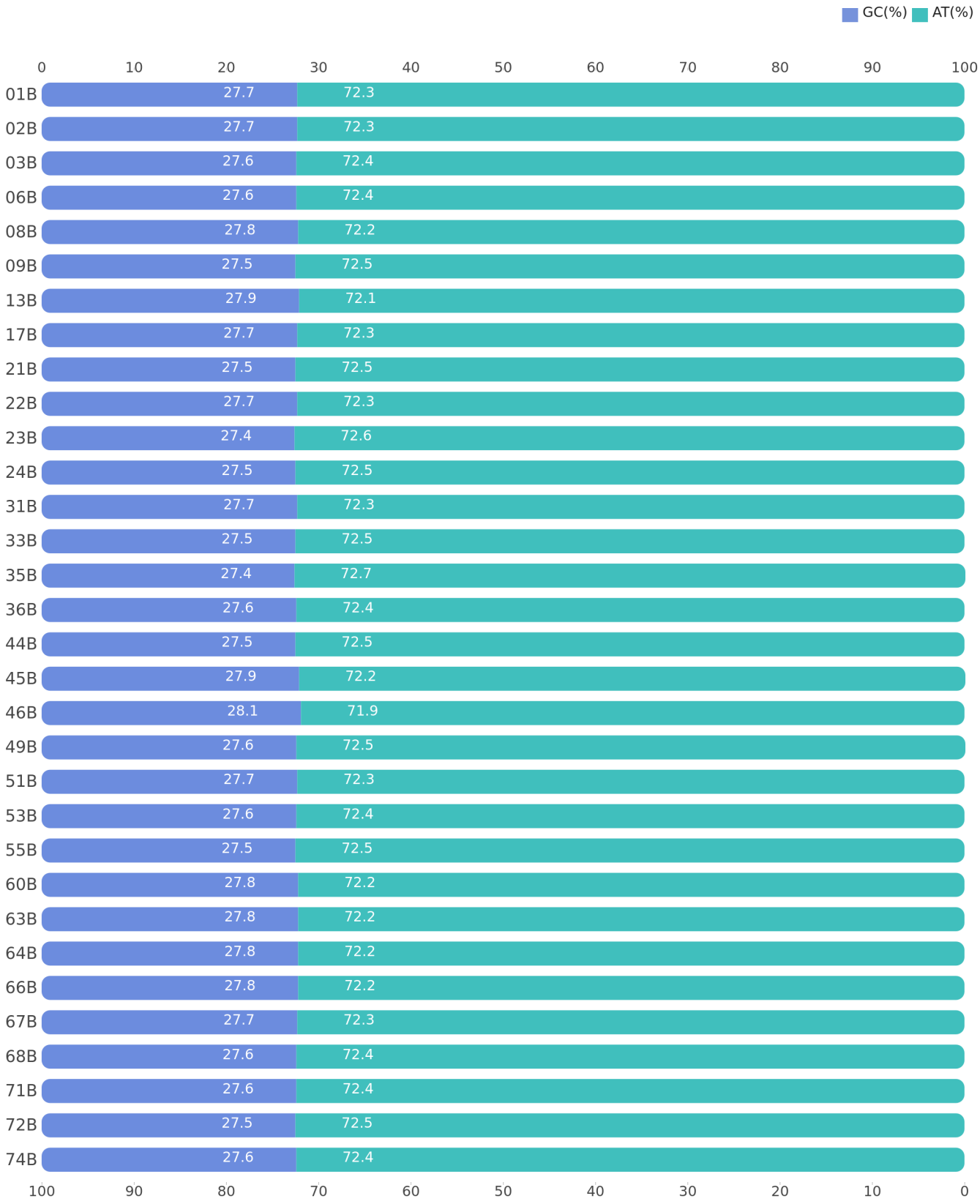
Total Bases

* Raw Data



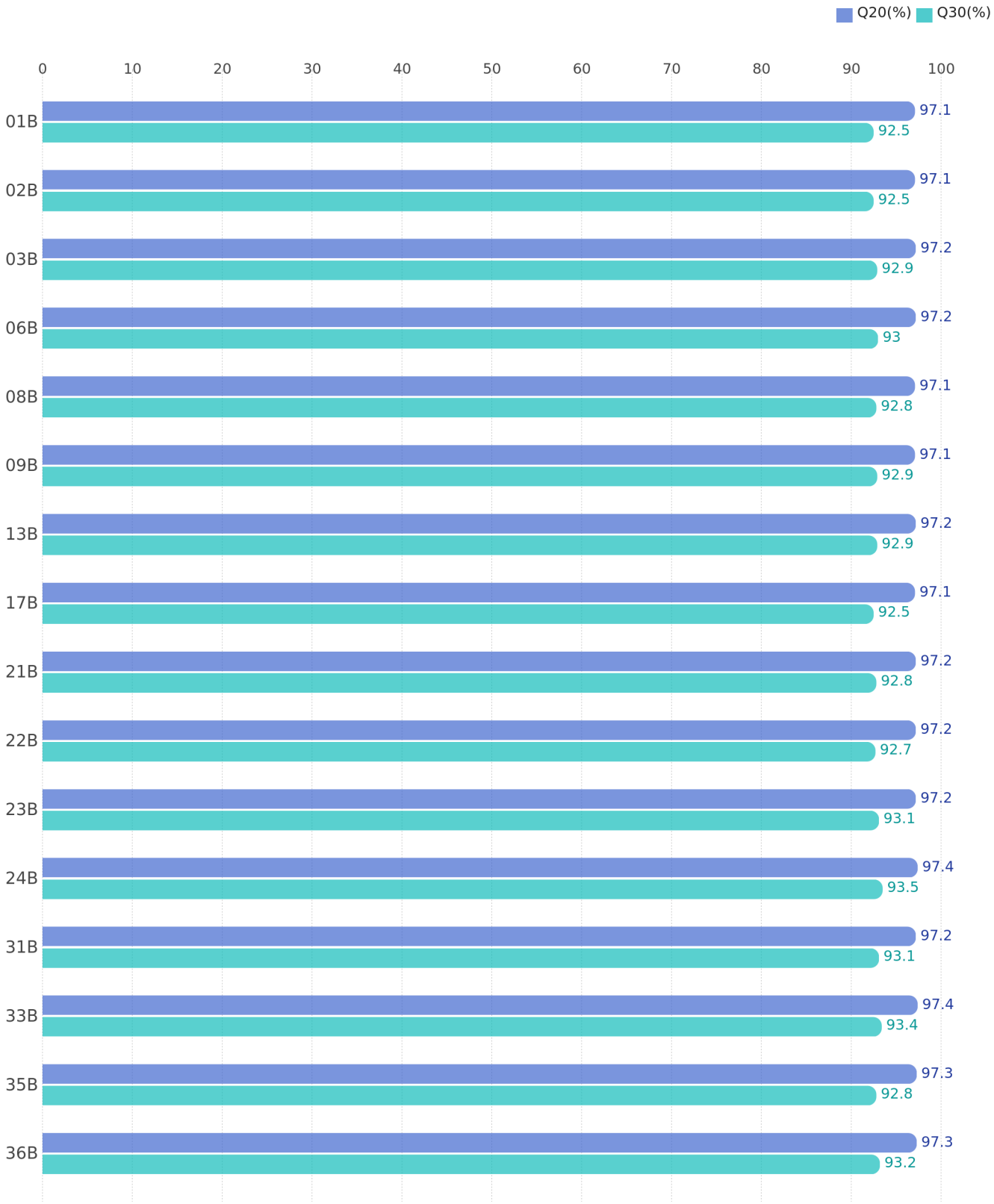
GC/AT Content

* Raw Data



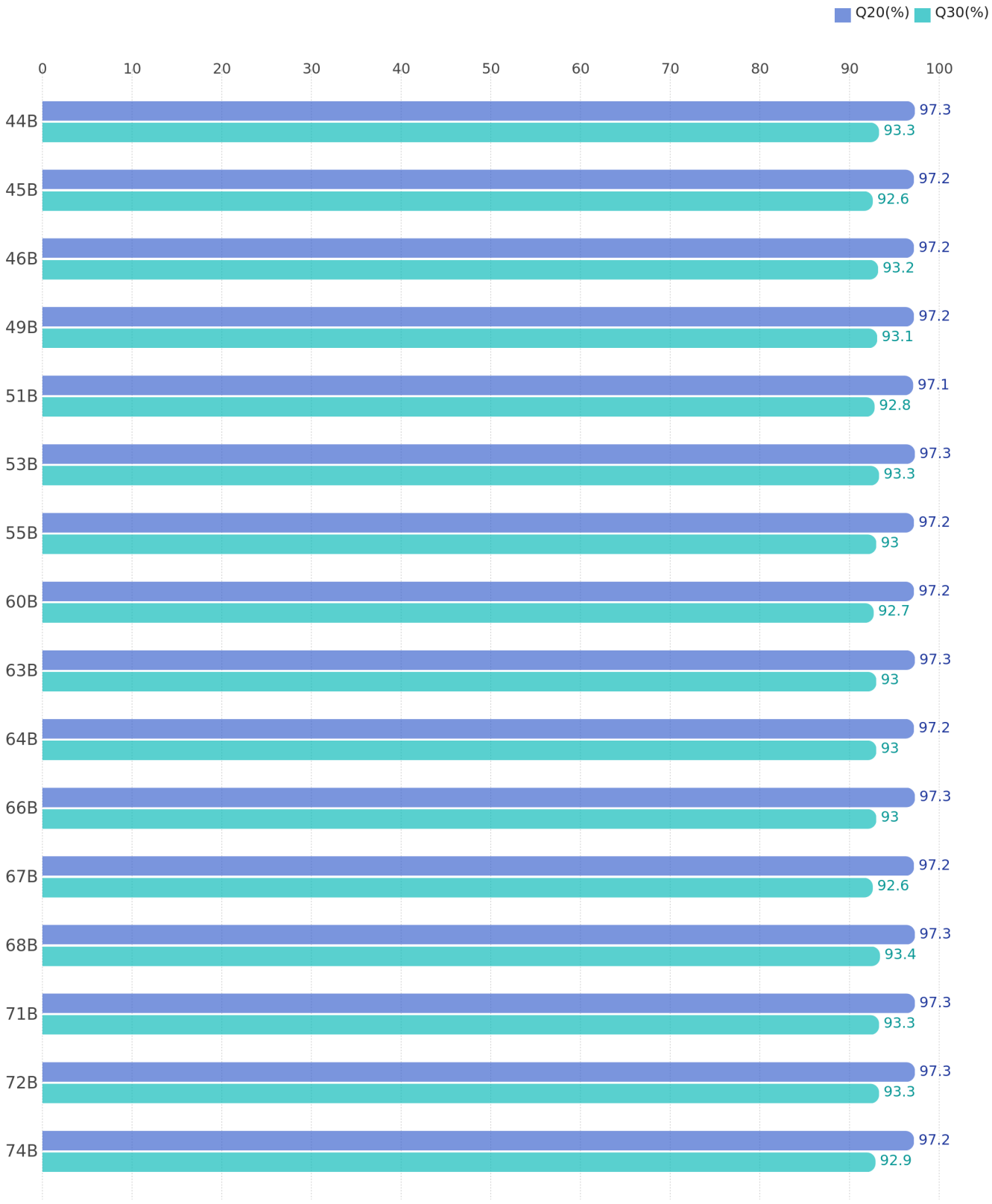
Q20/Q30 (%)

* Raw Data



Q20/Q30 (%)

* Raw Data



Download List

- The data has been uploaded to the cloud server.
- Once you receive/download the data, please make sure to check the integrity of the files.
  Please note that the sequencing files will be deleted from our server 3 months after the analysis report is released;
  please contact us within 3 months if you encounter a problem with the data.

* Raw Data Download

File Name	File Size(byte)	md5sum
01B_1.fastq.gz	3,670,497,946	993463c06e15b9933379fbc7d552edee
01B_2.fastq.gz	3,763,054,147	0c66b58fee27c38e6a83a03aa1cba519
02B_1.fastq.gz	3,627,813,914	cd150739102715a31c9e1fcccceb364e
02B_2.fastq.gz	3,697,994,184	6cd238aa193a47b0e1eace05a4f2d12b
03B_1.fastq.gz	3,520,092,334	60cc8f7ff8596ee9fc42108813f97585
03B_2.fastq.gz	3,597,700,405	9dda2b05b01c1b18d10ea35d2b4baf3a
06B_1.fastq.gz	3,365,710,684	69704b48ce9ba4ed3e71abfc3e531ad0
06B_2.fastq.gz	3,529,544,119	faa1eade13ec886939d3e499e4ce1038
08B_1.fastq.gz	3,398,035,743	ef313bda5895f5ca9e57dcccecb9dc2e
08B_2.fastq.gz	3,561,889,860	86f8ace83d3ed9e64b2b7fc50809891a
09B_1.fastq.gz	3,336,735,297	31333e2d0650fbf48108b6c14b4e7f24
09B_2.fastq.gz	3,497,770,966	3525c4d0b0d88d1f05f4d59068597c8b
13B_1.fastq.gz	4,498,302,521	55990b664205308a97262917be979617
13B_2.fastq.gz	4,593,012,950	b69de81c6be6a03bb6300c2ec1f39b04
17B_1.fastq.gz	3,881,342,477	3366cc36f0af2587f0e04943f8209c75
17B_2.fastq.gz	3,977,461,480	d2012ece4877f6f8a20dddf2642fa75a
21B_1.fastq.gz	4,149,282,396	5ccde43c2f58fb1b25dc9abbbf28bfb6
21B_2.fastq.gz	4,240,122,151	5202bdc33026edafa2b99d537b8e416e
22B_1.fastq.gz	3,387,168,764	133ec2e94ad48140b895c47dac9f78b1
22B_2.fastq.gz	3,476,000,687	301348768034b58253ab9f1f0cc3c103
23B_1.fastq.gz	3,743,473,899	70861caa5ec44d649fff9fdd195f7826
23B_2.fastq.gz	3,928,541,896	c1ec6b0eeedf5a175d7098e9428a1d08
24B_1.fastq.gz	4,179,213,252	1aeae4f2c43f5490430bf43a6473b6c0
24B_2.fastq.gz	4,391,412,940	3189046a72a02f85af4adfff273b83d9
31B_1.fastq.gz	4,370,182,478	6b99af0798d0184f2a40822cb45f9afe
31B_2.fastq.gz	4,576,968,647	f5695179452f3646a36223a48ce8d94d
33B_1.fastq.gz	4,022,081,237	57dd0c98c751455178deccab069515ad
33B_2.fastq.gz	4,200,058,698	1500177dd348c634f838fe9189afa863
35B_1.fastq.gz	4,293,424,588	4e592b26714a6ce38b21b9721adae175
35B_2.fastq.gz	4,390,945,141	504dd5cc6dc97e1e591646643d07c4de
36B_1.fastq.gz	4,416,534,103	36c33d389b90d7cc2ac54f92780bc10f
36B_2.fastq.gz	4,643,639,959	f2550b4b28d56094bb4501417ad75f0d

File Name	File Size(byte)	md5sum
44B_1.fastq.gz	3,631,381,548	20788bcc643da4839f72d57013f6ba8d
44B_2.fastq.gz	3,803,857,547	50dd2cce195b3088e421ca4d84462f82
45B_1.fastq.gz	4,202,447,338	7a2e24f4f0c927c700a9724f351f67d3
45B_2.fastq.gz	4,276,606,734	9e0f745e306ff699bb462d18e3d41bd8
46B_1.fastq.gz	5,980,629,723	07b7538d841a31e7a8e1b2e928f1f15f
46B_2.fastq.gz	6,277,320,363	d361f05efadd5eca29473eb898440c91
49B_1.fastq.gz	6,116,725,162	bcc58d38474f016eb400972cf2ecd8fc
49B_2.fastq.gz	6,412,923,503	48b1e572b8799f848c8761a337f01273
51B_1.fastq.gz	4,710,364,293	a953d7dcac3306f01ef2bc3eaf67133d
51B_2.fastq.gz	4,962,671,104	07ae52ecb62f5e47bc4e6e83905258cc
53B_1.fastq.gz	4,297,072,566	4dd9642eb96b7dc79c044c20408f8c97
53B_2.fastq.gz	4,520,504,189	ba8222679790bff760adf768110ce4cf
55B_1.fastq.gz	3,484,837,692	fa2c150724f7eaf8685c8f064738d075
55B_2.fastq.gz	3,654,052,221	0e47d5a4066651d20cb73ca793579340
60B_1.fastq.gz	3,844,887,834	e174d4ca00f515889fe135493e7243bc
60B_2.fastq.gz	3,938,443,822	e4dc28643ea4905f21d4fc3c2ae62a87
63B_1.fastq.gz	3,963,723,697	4ab2cccebe6122fb46b9921bdd52abab
63B_2.fastq.gz	4,054,137,741	47237b79a28c43220dea301a93475ebf
64B_1.fastq.gz	4,203,722,460	5f58257d480aafb08f4e510e8109f5e1
64B_2.fastq.gz	4,405,866,684	0bcf6ff1e07182fb7f042d97a46fae0b
66B_1.fastq.gz	4,080,294,600	26540ab6763138497525201e94a0b098
66B_2.fastq.gz	4,166,097,173	e57d7580585c0e899f652b8183fbc6e3
67B_1.fastq.gz	4,222,547,500	a1fee53bd35631da5a406ab9acc0af9d
67B_2.fastq.gz	4,326,391,432	4821f25d59755ca9fa5e2ff6b002408f
68B_1.fastq.gz	3,238,674,415	4523f71b01451ac18fbb2f07f7b617d7
68B_2.fastq.gz	3,382,691,722	728a838c9c84a1d4e68ce71a60f13c21
71B_1.fastq.gz	3,700,177,099	3d8ccba0a6c32b1716a0ddf7d25fa1ea
71B_2.fastq.gz	3,869,888,926	87d3116e5baf10ab53842d83b778ba7a
72B_1.fastq.gz	3,845,881,709	cee3bdf95f73341206dc7d53130f40aa
72B_2.fastq.gz	4,024,480,474	9bc8d204e20940443822de796898976f
74B_1.fastq.gz	3,366,650,306	22c2c2a6928194b1389374eed64af03d
74B_2.fastq.gz	3,546,997,199	dc227bb763e5d532c6d30cbe1a0a8a26

FAQ

Q **Why do I need to check the md5sum values, and how can I check it? (Windows system)**

A NGS data tend to have a large files size which makes them more likely to be corrupted during file transfer. So it's important that you check the md5sum of the files after receiving them to make sure what you received are what we gave.

Checking md5 hash in a Windows system

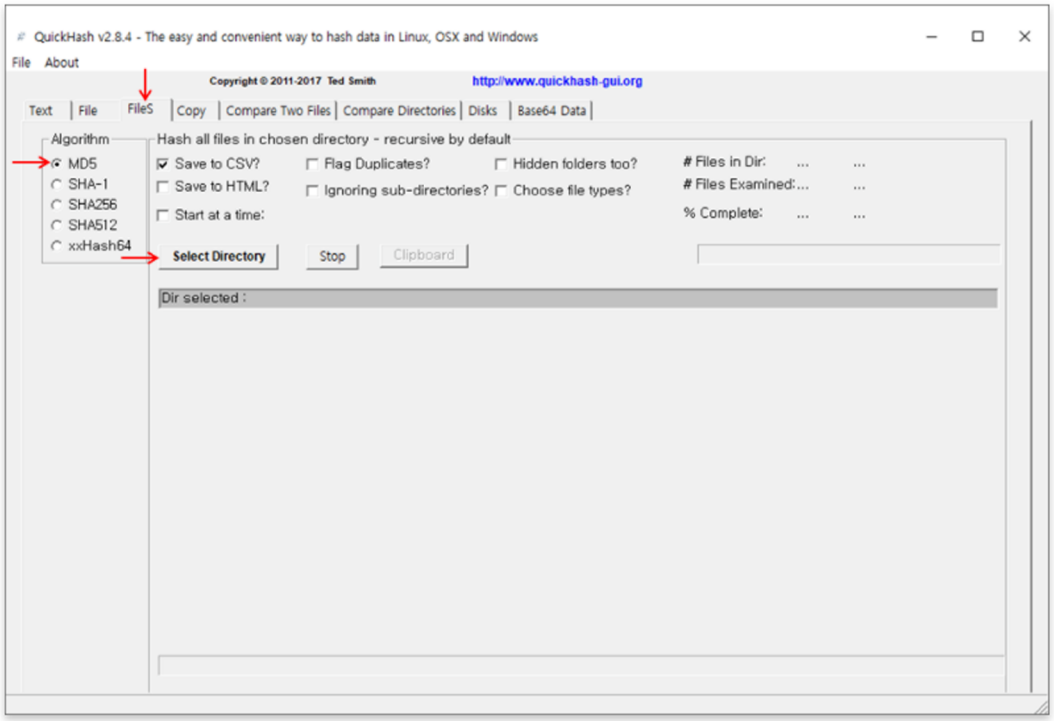
Windows does not provide a program for checking md5sum by default. An external program such as QuickHash-Windows can be used instead.

STEP 1 Download QuickHash-Windows from the website, and unzip the file.

STEP 2 Take a look at the UserManual.pdf file inside the zip file, and execute the .exe file.

Quickhash-GUI.exe	2,090,414	6,505,472
sqlite3-win32.dll	429,646	852,754
sqlite3-win64.dll	717,149	1,742,848
UserManual.pdf	512,697	576,987

STEP 3 Click on the "FileS" tab, and select [MD5] as the Algorithm.



STEP 4 Click "Select Directory" and choose the directory where the files to be checked are located in.
The output can be saved as a csv or txt file.
The process may take some time depending on the performance of the system being used.

STEP 5 Compare the newly calculated md5 value with the md5 value provided to you through the Analysis Report.

FAQ

Q Why do I need to check the md5sum values, and how can I check it? (Linux system)

A NGS data tend to have a large files size which makes them more likely to be corrupted during file transfer. So it's important that you check the md5sum of the files after receiving them to make sure what you received are what we gave.

Checking md5 hash in a Linux system

Linux systems have an internal md5sum program under /user/bin/md5sum.
md5sum has a "-c" option, which reads the md5 sums from the input file and checks them simultaneously.

Usage: $ md5sum -c [input file name]

STEP 1 Psomagen provides a text file containing the md5sum of deliverables you'll be receiving, which you can use to validate the integrity of the files. You can download this file by clicking on the "md5sum List" button in the "Download List" page. The text file will have the following name and format:

- Format : **<OrderNumber>_#samples_md5sum.txt**

- You can also find **"md5sum.txt"** located alongside your delivered files.

STEP 2 Use "md5sum -c" to validate the integrity of the file you've received. The input file for md5sum -c has to be delimited by two spaces with the md5sum column appearing before the file name, just like the sample image of "md5sum.txt" file shown above. As you can see, the two other files above are not formatted this way and need to altered to be used as input for md5sum -c. You can manually exclude the header and cut out "File" and "md5sum" column from the files, or simply run the following command:

$ awk '{print $3 " " $1}' <md5sum_file> | grep -v File

STEP 3 **"md5sum -c"** reads the input containing the md5 value of a file, and checks whether the md5 value of that file matches what's written inside the input file. This action outputs **"OK"** if the md5 value matches, and **"FAILED"** if otherwise. Check if the command outputs **"OK"** for all the files. (Refer to image below)

FAQ

Q I want to see the data produced by Psomagen. How can I open the files?

A NGS data tend to have large file sizes, and are not user-friendly to work with in a Windows environment. We recommend that you use Linux system for smoother operation.

Q Where can I find information for Illumina adapter sequences?

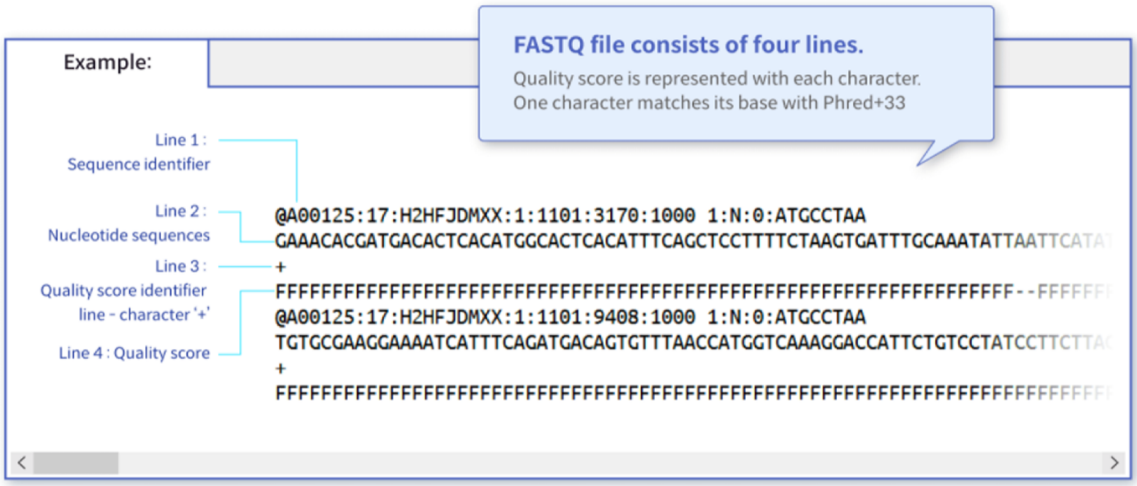
A Information on Illumina adapters can be found in this support document: Adapter Sequences Intro

Result File Description

Deliverables List

File Type	File Name	Description
FASTQ	[Sample name]_[read1].fastq.gz	Raw read1 sequence data
	[Sample name]_[read2].fastq.gz	Raw read2 sequence data
md5sum	[Order#]_[#samples]_md5sum[_DownloadLink].txt	You can download this file by clicking on the "md5sum List" button found on the "Download List" page. The file is slightly different in terms content, depending on how you're receiving your data. If you're receiving via download link, the file contains the following information : File name, File size, md5sum, FTP link. Otherwise, if your receiving your data via HDD the file only contains : File name, File size, and md5sum. MD5 is a string of 32 hexadecimal values, which represents a 'fingerprint' of a file. By comparing the supplied MD5 value to the actual value computed by the MD5sums utility, you can make sure that the file that you downloaded off of the internet has not been tampered with or modified from the original file stored in our server.

FASTQ Format



Phred Quality Score

Phred quality score numerically expresses the accuracy of each nucleotide. Higher Q number signifies higher accuracy.
For example, if Phred assigns a quality score of 30 to a base, the chances of having base call error are 1 in 1000.
Phred Quality Score Q is calculated with $-10\log_{10}(P)$, where p is probability of erroneous base call.

Quality of phred score	Probability of incorrect base call	Base call accuracy
10	1 in 10	90%
20	1 in 100	99%
30	1 in 1000	99.9%
40	1 in 10000	99.99%

Research use only

HEADQUARTER

Macrogen, Inc.

Laboratory, IT and Business Headquarter & Support Center

[08511] 1001, 10F, 254, Beotkkot-ro, Geumcheon-gu, Seoul, Republic of Korea (Gasan-dong, World Meridian 1)
Tel: +82-2-2180-7000
Email1: ngs@macrogen.com(Overseas)
Email2: ngskr@macrogen.com (Republic of Korea)
Web: www.macrogen.com
LIMS: dna.macrogen.com

SUBSIDIARY

Macrogen Europe

Laboratory, Business & Support Center

Meibergdreef 57, 1105 BA, Amsterdam, the Netherlands
Tel: +31-20-333-7563
Email: ngs@macrogen.eu

Macrogen Singapore

Laboratory, Business & Support Center

3 Biopolis Drive #05-18, Synapse, Singapore 138623
Tel: +65-6339-0927
Email: info-sg@macrogen.com

Psomagen (Macrogen USA)

Laboratory, Business & Support Center

1330 Piccard Drive, Suite 103, Rockville, MD 20850, United States
Tel: +1-301-251-1007
Email: inquiry@psomagen.com

Macrogen Japan

Laboratory, Business & Support Center

16F Time24 Building, 2-4-32 Aomi, Koto-ku, Tokyo 135-0064 JAPAN
Tel: +81-3-5962-1124
Email: ngs@macrogen-japan.co.jp

BRANCH

Macrogen Spain

Laboratory, Business & Support Center

Av. Sur del Aeropuerto de Barajas, 28. Office B-2, 28042 Madrid, Spain
Tel: +34-911-138-378
Email: info-spain@macrogen.com

psomagen