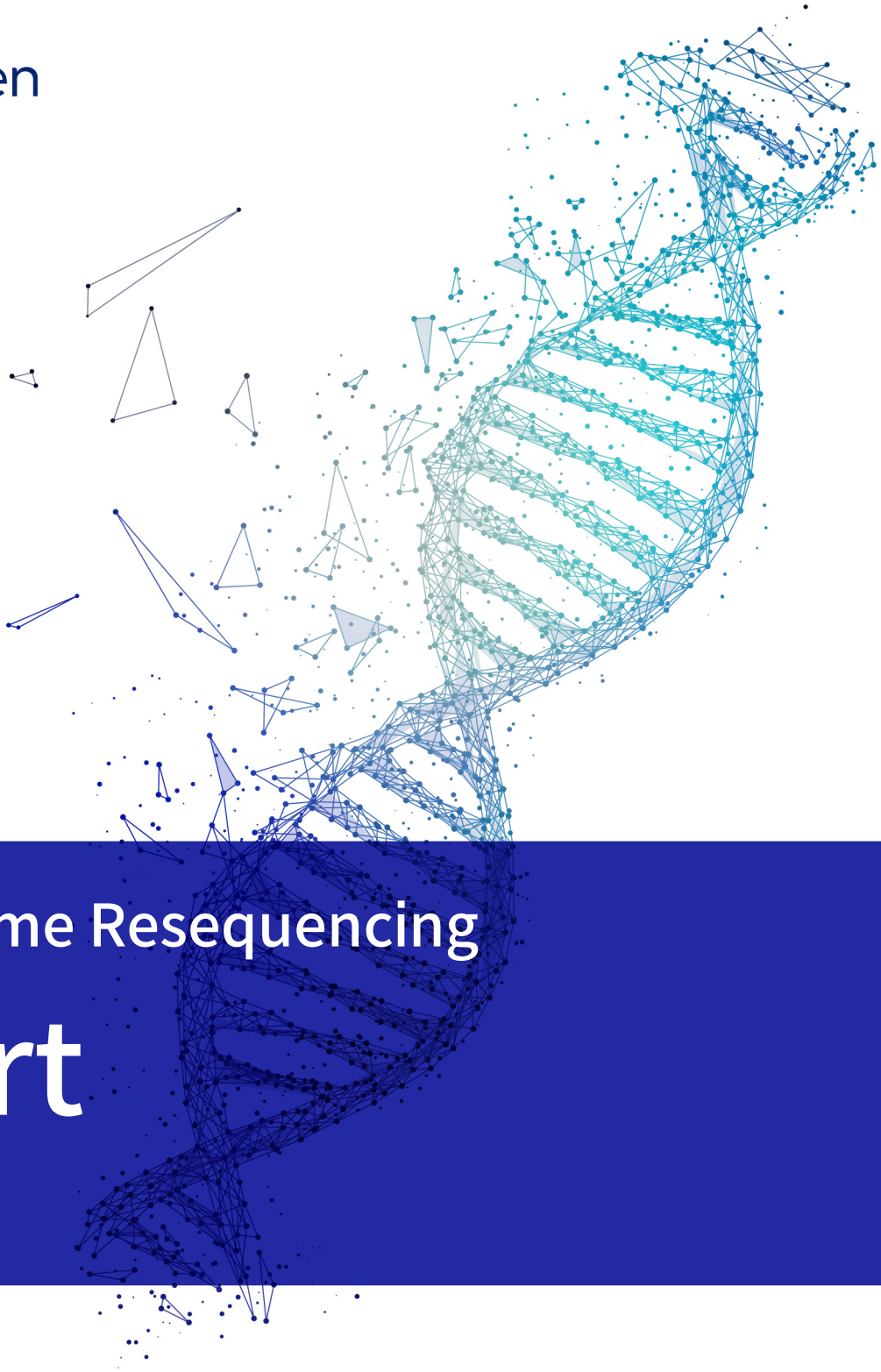
psomagen



# Transcriptome Resequencing

# Report

2025.08

RAWDATA REPORT

Research use only

# Table of Contents

# Order Information

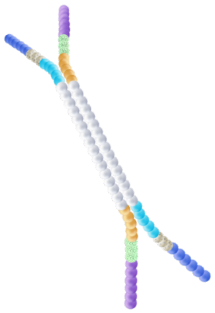
| | |
|---|---|
| Client Name | Sam White |
| Client Organization | Univ. of Washington |
| Order Number | AN00025268 |
| Application | Transcriptome Resequencing |
| Type of Read | Paired-end |
| Read Length | 151 |
| Type of Sequencer | Illumina platform |

# 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

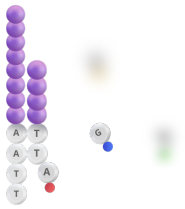
The sequencing library is prepared by random fragmentation of the DNA or cDNA sample, followed by 5' and 3' adapter ligation. Alternatively, "tagmentation" combines the fragmentation and ligation reactions into a single step which greatly increases the efficiency of the library preparation process.

## 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, 78,652,622 reads are produced, and total read bases are 11.9 Gbp. The GC content (%) is 52.8% and Q30 is 92.4%.

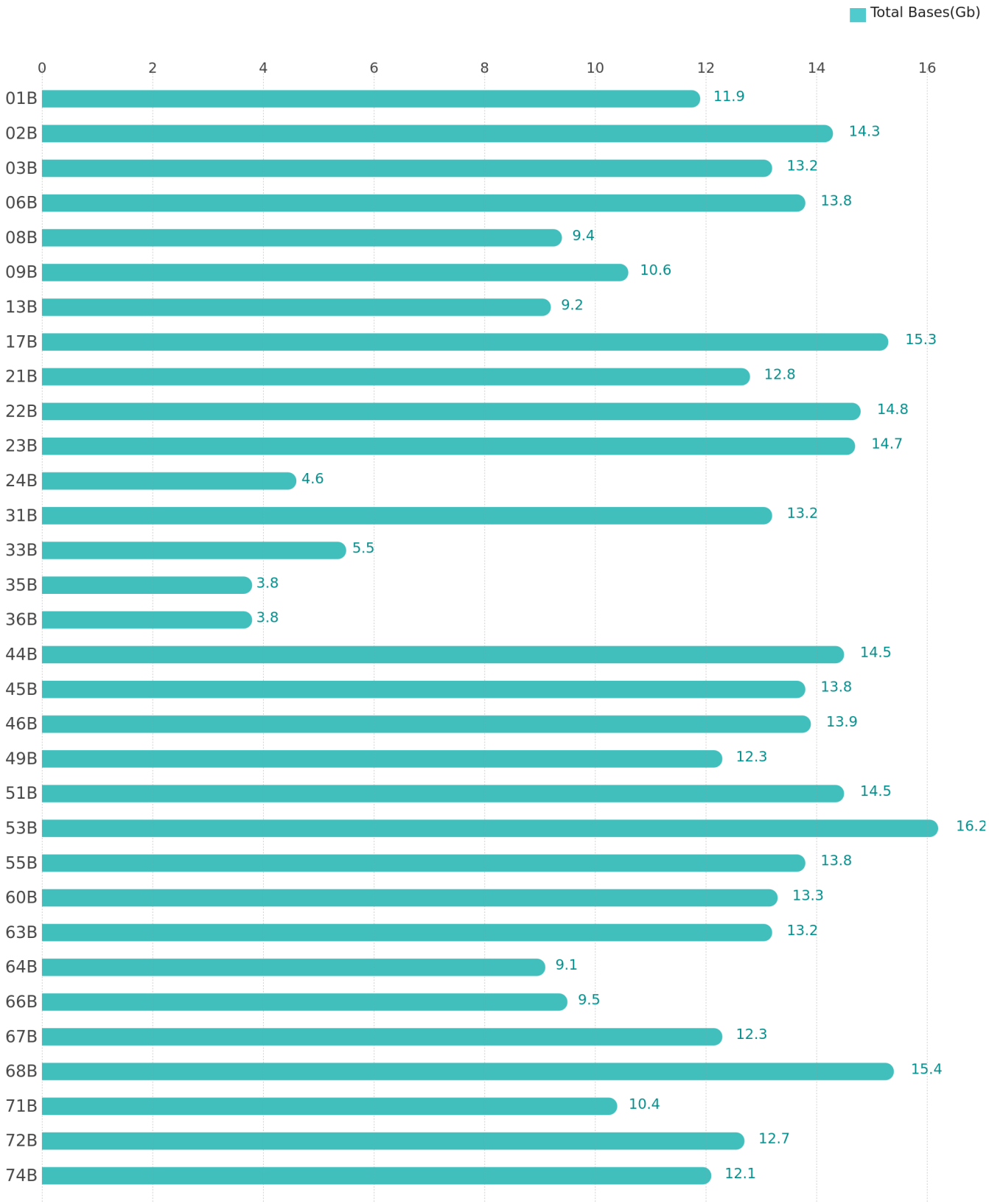
## * Raw Data

| Sample ID | Total bases(bp) | Total reads | GC(%) | AT(%) | Q20(%) | Q30(%) |
|---|---|---|---|---|---|---|
| 01B | 11,876,545,922 | 78,652,622 | 52.8 | 47.2 | 97.0 | 92.4 |
| 02B | 14,308,846,070 | 94,760,570 | 52.9 | 47.1 | 97.0 | 92.4 |
| 03B | 13,159,037,544 | 87,145,944 | 52.4 | 47.6 | 97.0 | 92.3 |
| 06B | 13,754,483,394 | 91,089,294 | 53.1 | 46.9 | 97.2 | 92.9 |
| 08B | 9,429,776,954 | 62,448,854 | 52.7 | 47.3 | 97.4 | 92.9 |
| 09B | 10,627,356,746 | 70,379,846 | 52.8 | 47.2 | 97.4 | 93.3 |
| 13B | 9,160,813,942 | 60,667,642 | 51.4 | 48.6 | 97.6 | 93.4 |
| 17B | 15,328,006,678 | 101,509,978 | 53.1 | 47.0 | 97.4 | 93.1 |
| 21B | 12,778,322,452 | 84,624,652 | 53.3 | 46.7 | 97.3 | 92.8 |
| 22B | 14,835,884,700 | 98,905,898 | 54.7 | 45.3 | 96.3 | 90.6 |
| 23B | 14,678,971,230 | 97,211,730 | 56.6 | 43.4 | 95.7 | 89.5 |
| 24B | 4,604,331,860 | 30,492,264 | 64.2 | 35.8 | 93.6 | 84.6 |
| 31B | 13,249,132,600 | 87,742,600 | 56.3 | 43.7 | 96.2 | 90.4 |
| 33B | 5,465,554,880 | 36,195,728 | 64.3 | 35.7 | 94.8 | 87.0 |
| 35B | 3,788,774,214 | 25,091,220 | 60.7 | 39.3 | 94.8 | 87.3 |
| 36B | 3,822,796,930 | 25,316,536 | 64.7 | 35.3 | 93.5 | 84.2 |
| 44B | 14,503,217,800 | 96,047,800 | 52.8 | 47.2 | 97.3 | 93.1 |
| 45B | 13,756,684,672 | 91,103,872 | 53.7 | 46.3 | 96.7 | 91.8 |
| 46B | 13,877,683,690 | 91,905,190 | 53.1 | 46.9 | 97.1 | 92.6 |
| 49B | 12,276,470,932 | 81,301,132 | 53.1 | 46.9 | 97.1 | 92.6 |
| 51B | 14,527,157,038 | 96,206,338 | 52.7 | 47.3 | 97.0 | 92.5 |
| 53B | 16,244,221,200 | 108,294,808 | 53.8 | 46.2 | 95.9 | 89.7 |
| 55B | 13,806,212,672 | 91,431,872 | 52.4 | 47.6 | 97.1 | 92.7 |
| 60B | 13,298,311,186 | 88,068,286 | 52.9 | 47.1 | 96.5 | 91.4 |
| 63B | 13,164,127,754 | 87,179,654 | 53.4 | 46.6 | 96.2 | 90.7 |
| 64B | 9,102,623,072 | 60,282,272 | 52.7 | 47.3 | 97.3 | 92.9 |
| 66B | 9,549,661,592 | 63,242,792 | 50.5 | 49.6 | 97.4 | 93.0 |
| 67B | 12,312,866,764 | 81,542,164 | 53.4 | 46.6 | 97.2 | 92.7 |
| 68B | 15,359,481,118 | 101,718,418 | 51.5 | 48.5 | 97.1 | 92.6 |
| 71B | 10,392,800,896 | 68,826,496 | 51.1 | 48.9 | 97.4 | 93.2 |
| 72B | 12,682,773,880 | 83,991,880 | 50.5 | 49.6 | 97.4 | 93.2 |
| 74B | 12,148,607,152 | 80,454,352 | 52.5 | 47.5 | 96.9 | 92.1 |

- 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.

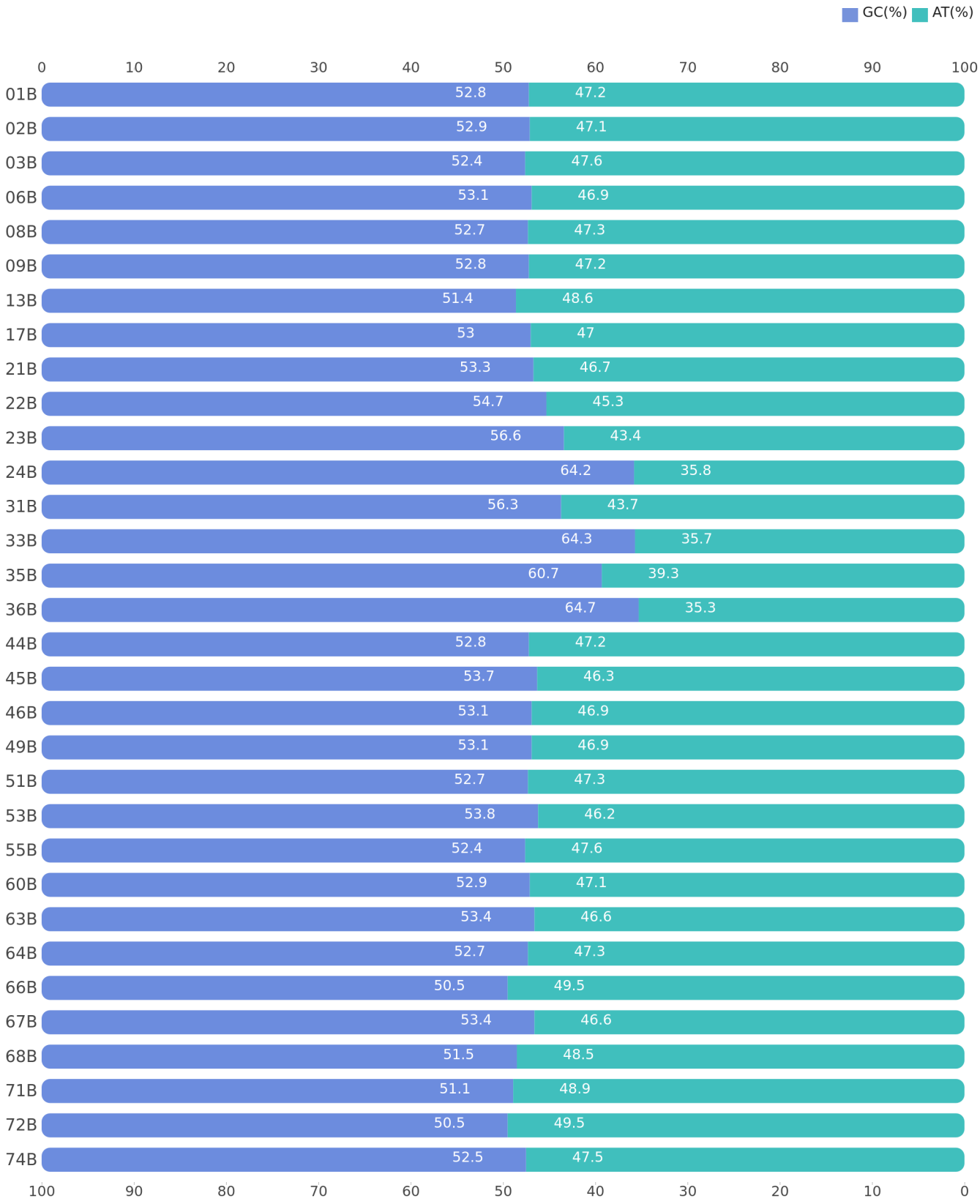
# Total Bases

## * Raw Data



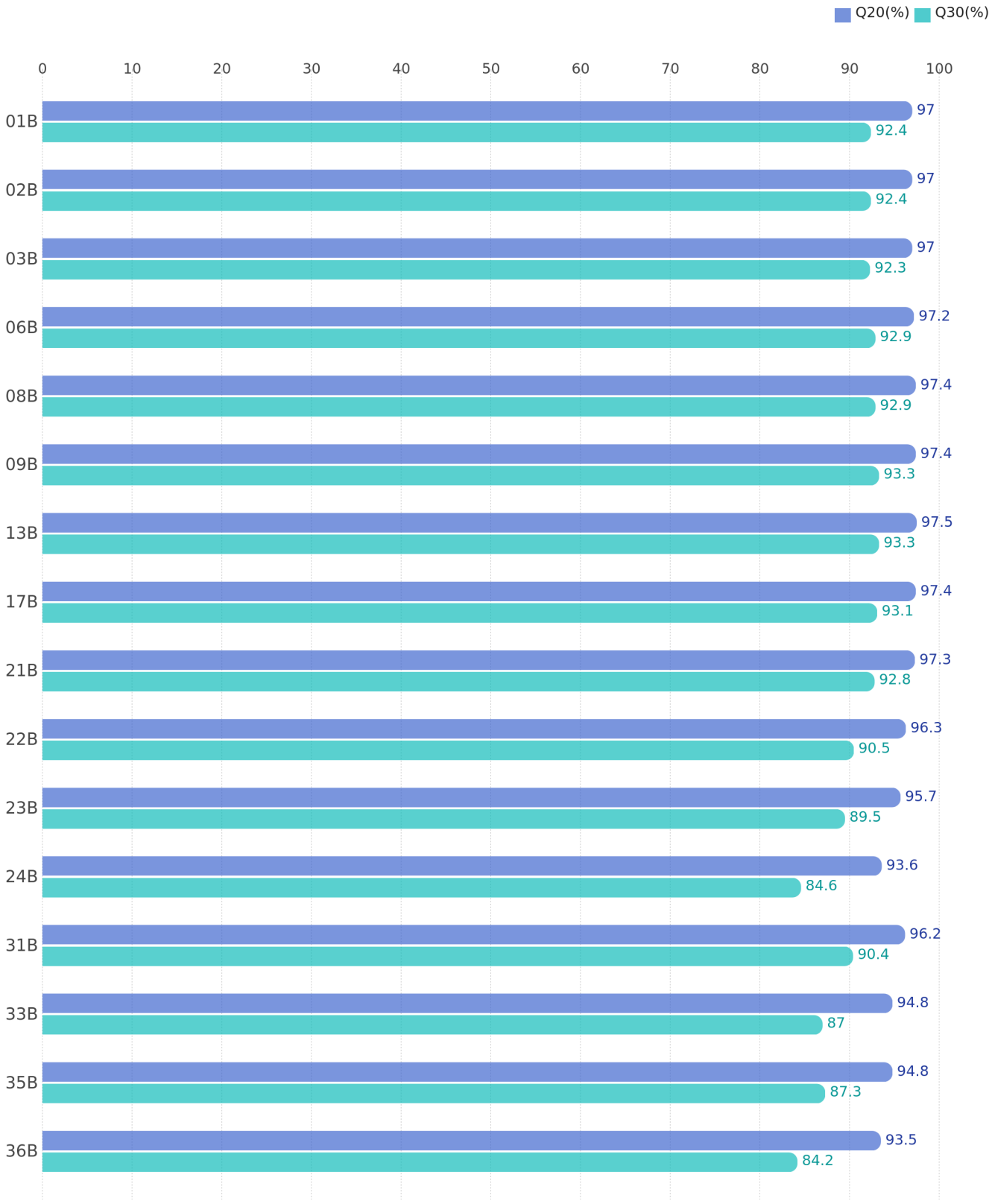
# GC/AT Content

## * Raw Data



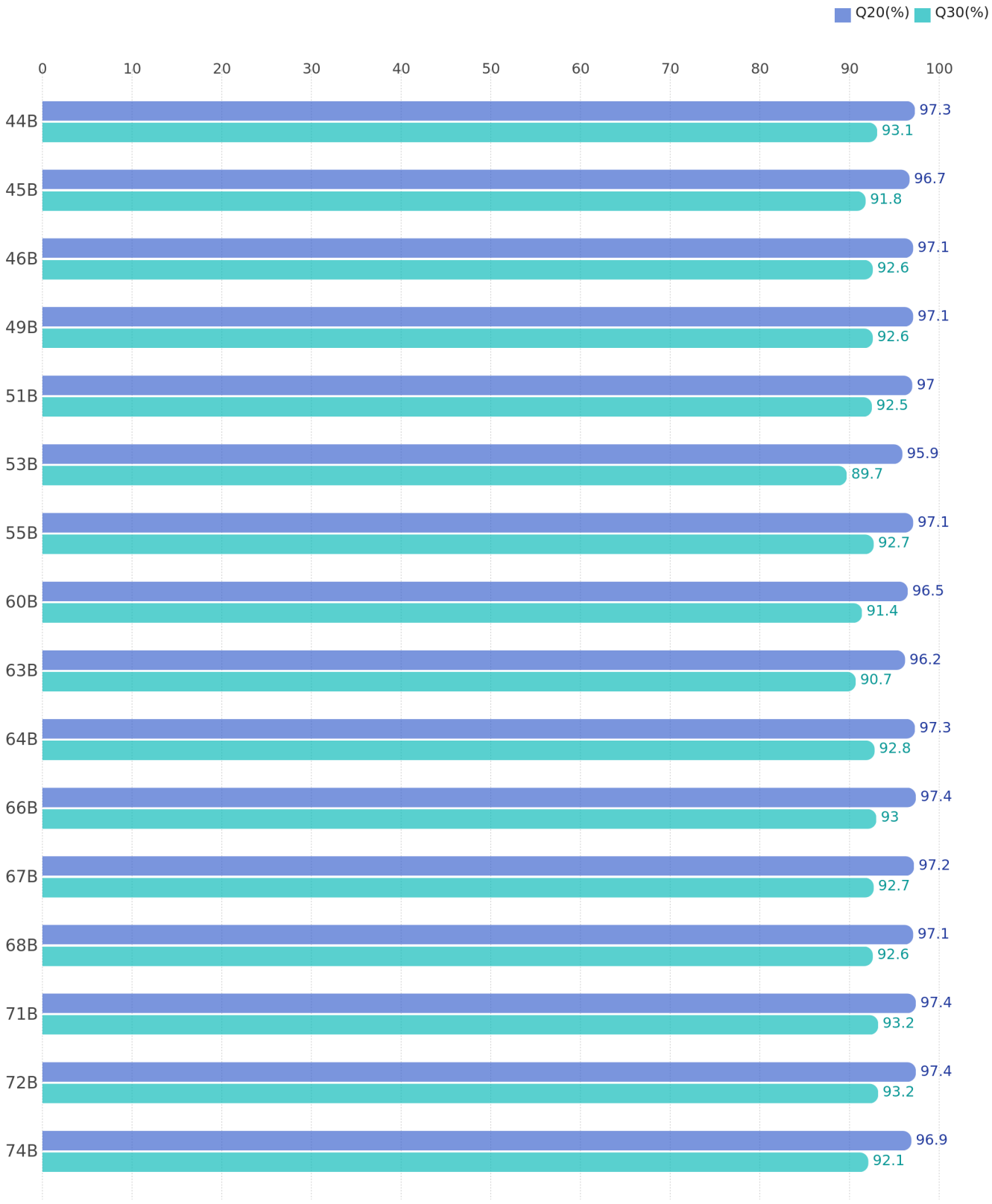
# Q20/Q30 (%)

* 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 | 2,783,633,675 | 6fedd1b2eae8343f5d3158bcebbc2c0c |
| 01B_2.fastq.gz | 2,768,219,140 | c29a3e61a87055d2b52c550b2457ccb6 |
| 02B_1.fastq.gz | 3,327,760,037 | 0b627f16925cf9fe7f243c1cf42d6b55 |
| 02B_2.fastq.gz | 3,328,007,943 | e04198f140dafe09dc23d007632ed71e |
| 03B_1.fastq.gz | 3,088,782,211 | 738b460aea7c72d26b4339f4010f0836 |
| 03B_2.fastq.gz | 3,040,957,800 | b6b5b0a9c172d96b6275bcd5f3b646d2 |
| 06B_1.fastq.gz | 3,052,325,250 | 31d75f4d0e054d4c86c0da862594e224 |
| 06B_2.fastq.gz | 3,063,952,630 | 7b3cd9980ea280d0c7b8947ec5d8aa7d |
| 08B_1.fastq.gz | 2,199,043,656 | 24cb8578366d1f5f4272c9c476de299c |
| 08B_2.fastq.gz | 2,190,122,769 | 09f051ec4e53159fbb69bc603b88d23f |
| 09B_1.fastq.gz | 2,424,752,075 | 8dc65df8b938972c14058e265e424909 |
| 09B_2.fastq.gz | 2,414,300,976 | 40be17a3abad1ab978596e6537afee31 |
| 13B_1.fastq.gz | 2,187,255,130 | 8805a3333ac34c19cf9dc7d38b246dc5 |
| 13B_2.fastq.gz | 2,153,812,866 | 4b2bf6c09f4362f1a3d839095f1dc814 |
| 17B_1.fastq.gz | 3,396,144,429 | b462bcc5e40ab8d25783bfb00267a37e |
| 17B_2.fastq.gz | 3,369,126,766 | e729ac7592cf45ac7bf904a78db8cfb6 |
| 21B_1.fastq.gz | 2,898,639,386 | bb77085d8990551375c2c4477b473069 |
| 21B_2.fastq.gz | 2,863,529,861 | bf2a635c91dacc32e08ad773ef8f3fb5 |
| 22B_1.fastq.gz | 2,782,352,576 | a8d5fe740943b2048efa3a731f274cf8 |
| 22B_2.fastq.gz | 2,752,601,065 | 6726f4ba4db5b4463c22bcd50d39b049 |
| 23B_1.fastq.gz | 3,225,696,771 | 2c3e7953f35655854efea640bb45b4f5 |
| 23B_2.fastq.gz | 2,935,000,455 | 586ed690d6f20cc82615a76e0da692e6 |
| 24B_1.fastq.gz | 906,209,418 | 8d48d2d22515567b42072e3b39244935 |
| 24B_2.fastq.gz | 701,645,423 | 6b551b3614b80a41bff3487f2bbe62bb |
| 31B_1.fastq.gz | 3,007,371,995 | 2e3bbef796be64494c19b594893b03bf |
| 31B_2.fastq.gz | 2,805,752,501 | e10fcb06d1741ccd59d1987744fd33a0 |
| 33B_1.fastq.gz | 1,044,165,177 | b4880f2b7a17ecdf13c2ed8db6ba375c |
| 33B_2.fastq.gz | 857,967,356 | b777a8b9214cd0306b53adab7d612a33 |
| 35B_1.fastq.gz | 788,546,365 | d11086a8709d38cec8c9ad47e757631d |
| 35B_2.fastq.gz | 670,834,013 | 01710283a30848c8c728bdb4683264fe |
| 36B_1.fastq.gz | 728,679,432 | bd5d5287d537eedbe669d71759c2b5a3 |
| 36B_2.fastq.gz | 586,542,373 | dc6af6661e21b0b114ba799fcc2b7dfb |

| File Name | File Size(byte) | md5sum |
|---|---|---|
| 44B_1.fastq.gz | 3,277,264,496 | ce0b909a7354d4b1c26447fb5e20b23a |
| 44B_2.fastq.gz | 3,267,960,980 | 1f3ce2bbc232d6a9ec615a944c43764a |
| 45B_1.fastq.gz | 3,166,828,317 | 79006753928d7cc3c8fa2ed4375d8074 |
| 45B_2.fastq.gz | 3,104,547,414 | 7ea905e73b9ac883b85969edd0d82828 |
| 46B_1.fastq.gz | 3,245,115,192 | 79c35e8666767866f064ba209bd68276 |
| 46B_2.fastq.gz | 3,150,579,274 | 739ff7a112276e29b5b7290e5717806e |
| 49B_1.fastq.gz | 2,845,388,714 | 758b88c47458d0dec527909ff9877a0c |
| 49B_2.fastq.gz | 2,859,070,880 | 98dddfe3baa5779d2a4b172767589f0f |
| 51B_1.fastq.gz | 3,344,533,749 | 6e84f8cf797b42a2eebf7e4d46bc227d |
| 51B_2.fastq.gz | 3,333,442,286 | f5871e2d3af8a8f87d8e3229e4f011cb |
| 53B_1.fastq.gz | 3,043,275,466 | c9ce5350e1ef0c22baa94816db38653c |
| 53B_2.fastq.gz | 2,933,564,244 | 543f527728677cebd99b3306b4c85c27 |
| 55B_1.fastq.gz | 3,269,222,765 | 65fca88ef809be97e6cd19a6d7651d9b |
| 55B_2.fastq.gz | 3,271,385,822 | af0aa7f72bc00c6f1b318595517e0a78 |
| 60B_1.fastq.gz | 3,061,321,321 | 20e6de2165601279640046b0086d9c80 |
| 60B_2.fastq.gz | 2,938,909,127 | 8412eb3fbd79a8de9ff0c6f24f68a3ed |
| 63B_1.fastq.gz | 2,952,918,473 | 613dcb49c1e8df2e458bf50a1fbe77f4 |
| 63B_2.fastq.gz | 2,860,204,752 | 1ff6f93f465e4ebdf8254e039e29f1e7 |
| 64B_1.fastq.gz | 2,127,599,635 | 03c0599d277f5e7950e360aa0caca2c1 |
| 64B_2.fastq.gz | 2,154,004,914 | dee3b9e4098965a5af60dea16316baa7 |
| 66B_1.fastq.gz | 2,289,702,135 | 4e030e08ee40d70e9db2e1f030f6a43a |
| 66B_2.fastq.gz | 2,249,116,844 | 2239f5ca9fbb39b0b2ad96472f0c0bd5 |
| 67B_1.fastq.gz | 2,977,652,190 | 60bfd59b26d81ace09caf891587aca8f |
| 67B_2.fastq.gz | 2,864,828,986 | 7dcecb43e08ec0a2ca00058edc32525f |
| 68B_1.fastq.gz | 3,533,853,523 | 361580141edcd36d7f8bb0bc299bba71 |
| 68B_2.fastq.gz | 3,535,521,463 | 0f68b4f14c0b4890ef324014bdf19a1a |
| 71B_1.fastq.gz | 2,440,282,467 | 088952fe569206921a39272bc12c1de5 |
| 71B_2.fastq.gz | 2,424,660,679 | 2eecdd9bac97c1324348a783828a0db8 |
| 72B_1.fastq.gz | 2,978,502,241 | ef6f95eeb868c33987293c15052b4f9a |
| 72B_2.fastq.gz | 2,954,388,292 | 0bde2c3921929ab7be2362682526c3a2 |
| 74B_1.fastq.gz | 2,790,450,639 | 3d8811cdaeec56dfd98ab9621d599ea6 |
| 74B_2.fastq.gz | 2,791,506,715 | 428fbb2f0f6d19d3cc0ddb6a81eb4c32 |

# 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

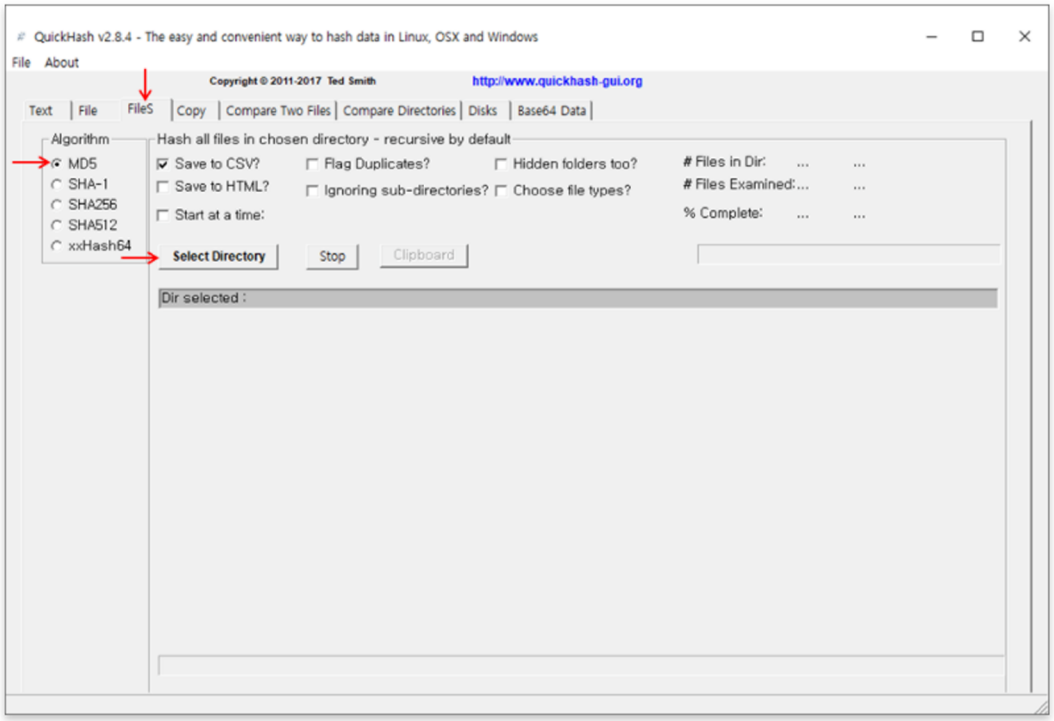
Windows does not provide a program for checking md5sum by default. An external program such as QuickHash-Windows can be used instead.

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?

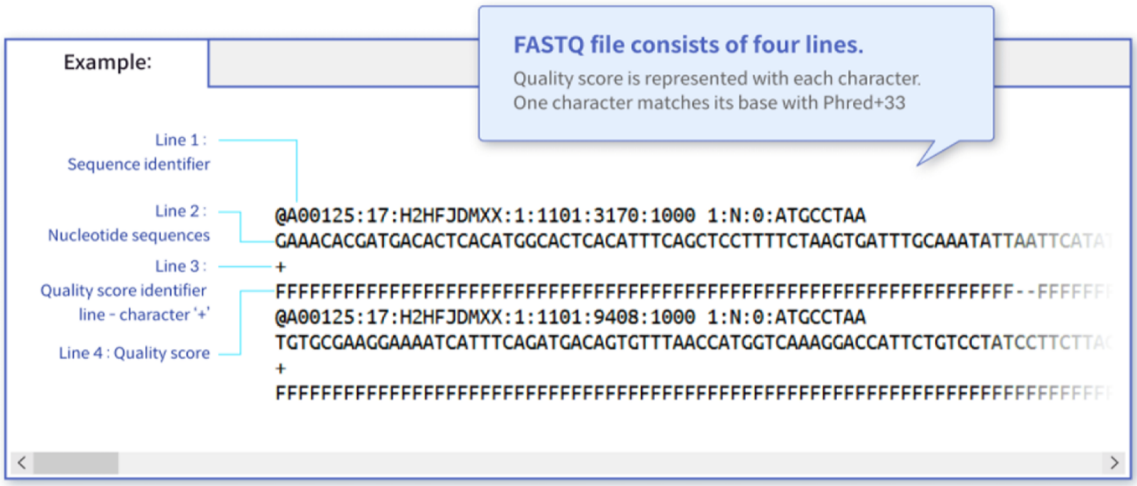
**A** Information on Illumina adapters can be found in this support document: Adapter Sequences Intro

# 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.<br><br>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