



Analysis of **PHRED** Score using Python

Koneenika Mallick
Krupali Bhayani
Nidhi Chaudhury
Guided by:
Dr. Jigar Mistry



OBJECTIVES

- Implement a Python-based method to identify low-quality reads
- Output results in a structured format (CSV)
- Understand Phred quality scores in sequencing data
- Highlight their importance in bioinformatics



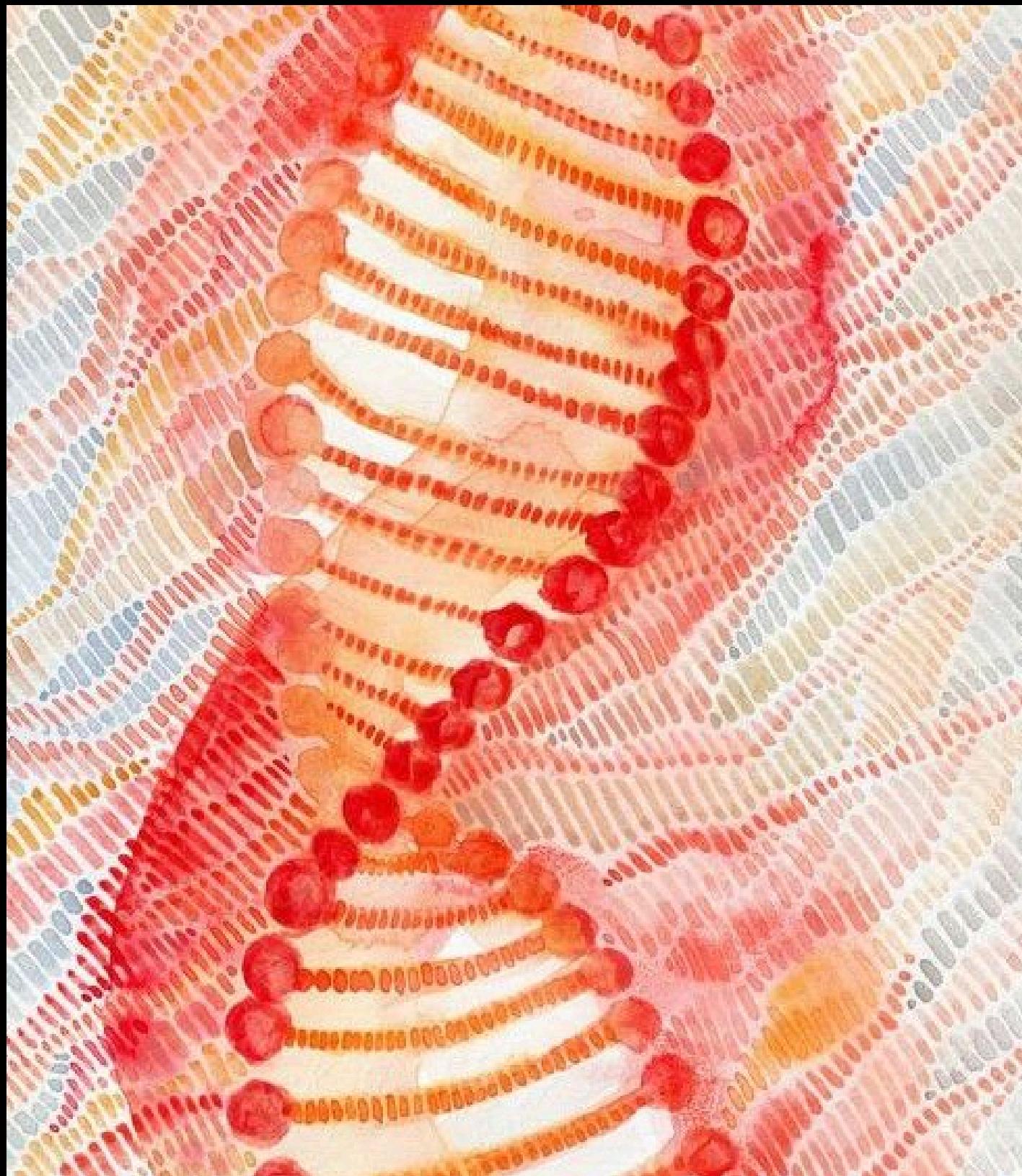
What is **PHRED** Score?



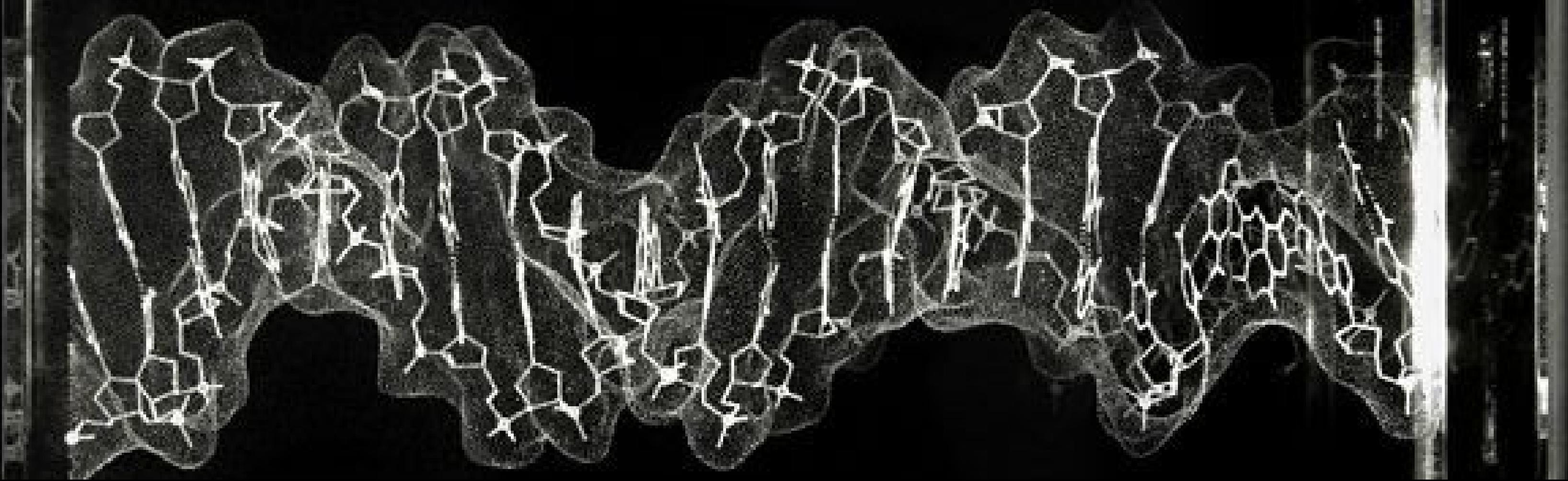
- • A numerical value that represents the quality of each nucleotide base call in DNA sequencing
- Indicates the probability of an incorrect base call



Importance of PHRED SCORE



- **Quality Control** – Filters out poor-quality bases/reads → ensures reliable data.
- **Better DNA Assembly** – Gives more weight to high-quality reads when combining sequences.
- **Accurate Variant Detection** – Reduces false positives in mutation analysis.
- **Universal Standard** – Used across all sequencing platforms, making results comparable.
- **Bioinformatics Pipelines** – Stored in FASTQ files → software uses scores for trimming, alignment, and analysis.
- **Reliable Genotyping** – Ensures confidence in calling genetic variants in populations and clinical studies



- General formula:

$$Q = -10 \times \log_{10}(P)$$

Where:

Q = Phred Quality Score

P = Probability that the base
call is incorrect

How is it
Calculated?



METHODOLOGY

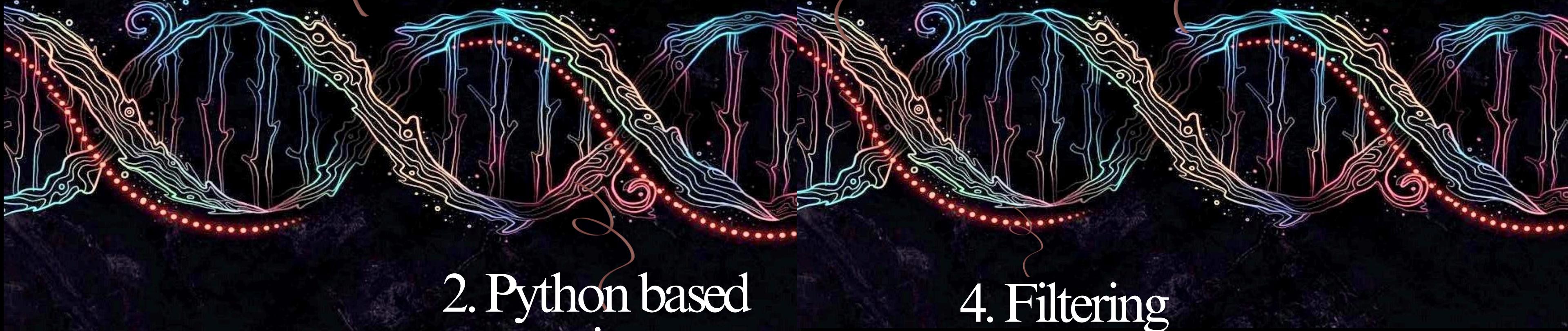
1. Input FASTQ file

3. Quality assessment

5. Output

2. Python based processing

4. Filtering



OUTPUT

The analysis produced a table of sequences that were identified as low-quality based on their mean Phred scores. Each record includes:

- Read ID: Unique identifier for each sequencing read.
- Sequence: The nucleotide sequence associated with that read.
- Mean Phred Score: Average quality score across the entire read.

Read_ID	Sequence	Mean_Phred_Score
@SRR34999326.1 TBGYX:08006:1	CCTACGGGAGGCAGCAGTGGGAATATTGACAATGGGGAAACCCGTATCCAGCCATGCCCGTGTGAAGAAGGCCTTATGGTTGTAAAGCACTTTAACCGAGGGAGGAGGACTCTA/	30
@SRR34999326.2 TBGYX:08028:1	CCTACGGGAGGCAGCAGTGGGATTCGGC	30
@SRR34999326.3 TBGYX:08031:1	CCTACGGGAGGCAGCAGTGGGAATAATTGGACAATTGGGGAAACCCGTATCCAGCCATGCCCGTGTGAAGAAGGCCTTATGGTTGTAAAGCACGTTAACCGAGGGAGGAGGC	30
@SRR34999326.4 TBGYX:08041:1	CCTACGGGAGGCAGCAGTGGGAATATTGACAATGGGGAAACCCGTATCCAGCCATGCCCGTGTGAAGAAGGCCTTATGGTTGTAAAGCACTTCATGCGAGGGAGGAAAGG/	30
@SRR34999326.5 TBGYX:07992:1	CCTACGGGAGGCAGCAGTGGGAATATTGACAATGGGGAAACCCGTATCCAGCCATGCCCGTGTGAAGAAGGCCTTATGGTTGTAAAGCACTTAACCGAGGGAGGAGGCGCTC	30
@SRR34999326.6 TBGYX:08025:1	CCTACGGGAGGCAGCAGTGGGAATATTGACAATGGGGAAACCCGTATCCAGCCATGCCCGTGTGAAGAAGGCCTTATGGTTGTAAAGCACTTAACCGAGGGAGGAGGACTCTG/	30
@SRR34999326.7 TBGYX:08028:1	CCTACGGGAGGCAGCAGTGGGAATATTGACAATGGGGAAACCCGTATCCAGCCATGCCCGTGTGAAGAAGGCCTTATGGTTGTAAAGCACTTAACCGAGGGAGGAGGCTACTG/	30
@SRR34999326.8 TBGYX:08029:1	CCCGGTACGGGAGGCAGCAGTAGGAAATCTCCGAATGGCGAAAGCCTGACGGAGCAACGCCCGTGAATGAAGGTCTCGGATCGTAAAACCTCTGTATTAGGAAGAACAAACG	30
@SRR34999326.9 TBGYX:08029:1	CCTACGGGAGGCAGCAGTGGGAATATTGACAATGGCGCAAGCCTGATCCAGCCATGCCCGTGTGAAGAAGGCCTTCCGGTTGTAAAGCACTTCAAGGAGGAAGCGTTGAG1	30
@SRR34999326.10 TBGYX:08017:	CCTACGGGAGGCAGCAGGCCCGTAAT	30
@SRR34999326.11 TBGYX:08023:	CCTACGGGAGGCAGCAGCCCCTACGTAT	30
@SRR34999326.12 TBGYX:08032:	CCTACGGGAGGCAGCAGTGGGAATATTGACAATGGGGAAACCCGTATCCAGCCATGCCCGTGTGAAGAAGGCCTTGTGGTTGTAAAGCACTTAACCGAGGGAGGAGGCTAC	30
@SRR34999326.13 TBGYX:08000:	CCTACGGGAGGCAGCAGTGGGAATCTTGACAATGGGGGAAGCCTGATGCAGCGACGCCCGTGAAGGGATGACGCCCTTCCGGTTCTGAAACCTCTTCCGACGGAGCAGC/	30
@SRR34999326.14 TBGYX:08010:	CCTACGGGAGGCAGCAGTGGGAATTCTCCGAATGGCGAAAGCCTGACGGAGCAATGCCCGTGGAGGTGGAAGGCCAACGGTCGCAACTCTTCTCGGAGAACAAATGACC	30
@SRR34999326.15 TBGYX:08017:	CCTACGGGAGGCAGCAGTGGGAATATTGACAATGGGGAAACCCGTATCCAGCCATGCCCGTGTGAAGAAGGCCTTGTAAAGCACTTAACCGAGGGAGGAGGCGCTCTAGGA	30
@SRR34999326.16 TBGYX:08021:	CCTACGGGAGGCAGCAGTAGGAAATCTCCGAATGGCGAAAGCCTGACGGAGCAACGCCCGTGAATGAAGGTCTCGGATCGTAAAACCTCTGTATTAGGAAGAACACGTGTAA/	30
@SRR34999326.17 TBGYX:08029:	CCCTAAGGCAGCAGTGGGAATATTGACAATGGGGAAACCCGTATCCAGCCATGCCCGTGTGAAGAAGGCCTTGTGTTGTAAAGCACTTAACCGAGGGAGGAGGCGCTCTAG	30
@SRR34999326.18 TBGYX:08024:	CCTACGGGAGGCAGCAGTGGGAATTGGACAATGGCGAAAGCCTGATCCAGCCATGCCCGTGTGAAGAAGGCCTTCCGGTTGTAAAGGACTTTGTCCGGGGAGGAATCCCCAGCC	30
@SRR34999326.19 TBGYX:08031:	CCCTCGGGGAGGCAGCAGTGGGGAAATTGGACAATGGGGAAACCCGTATCCAGCCATGCCCGTGTGAAGAAGGCCTTATGGTTGTAAAAAGCACTTAACCGAGGGAGGAGG	30
@SRR34999326.20 TBGYX:08032:	CCTACGGGAGGCAGCAGTGGGAATATTGACAATGGGGAAACCCGTATGCAGCCATGCCCGTGTGAAGAAGGCCTTGTGAA	30
@SRR34999326.21 TBGYX:08014:	CCTACGGGAGGCAGCAGTGGGAATATTGACAATGGGGAAACCCGTATCCAGCCATGCCCGTGTGAAGAAGGCCTTGTGTTGTAAAGCACTTAACCGAGGGAGGAGGCGCTC	30
@SRR34999326.22 TBGYX:08032:	CCTACGGGAGGCAGCAGTGGGAATATTGACAATGGGGAAACCCGTATCCAGCCATGCCCGTGTGAAGAAGGCCTTATGGTTGTAAAGCACTTAACCGAGGGAGGAGGCGCTGACTGAC	30
@SRR34999326.23 TBGYX:08034:	CCTACGGGAGGCAGCAGTGGGAATATTGACAATGGGGAAACCCGTATCCAGCCATGCCCGTGTGAAGAAGGCCTTGGTTGTAAAGCACTTAACCGAGGGAGGAGGCGCTCTAC	30
@SRR34999326.24 TBGYX:07995:	CCTACGGGAGGCAGCAGTGGGAATATTGACAATGGGGAAACCCGTATCCAGCCATGCCCGTGTGAAGAAGGCCTTATGGTTGTAAAGCACGTTAACCGAGGGAGGAGGCTG	30
@SRR34999326.25 TBGYX:08018:	CCTACGGGAGGCAGCAGTGGGAATTGGACAATGGCGAAAGCCTGATCCAGCCATGCCCGTGTCTGAAGAAGGCCTTCCGGTTGTAAAGGACTTTGTCCGGGGAGGAATCCCC	30
@SRR34999326.26 TBGYX:08033:	CCTACGGGAGGCAGCAGTGGGATTCGGCAATGGCGAAAGCCTGACGGAGCAAT	30

Thank
You

