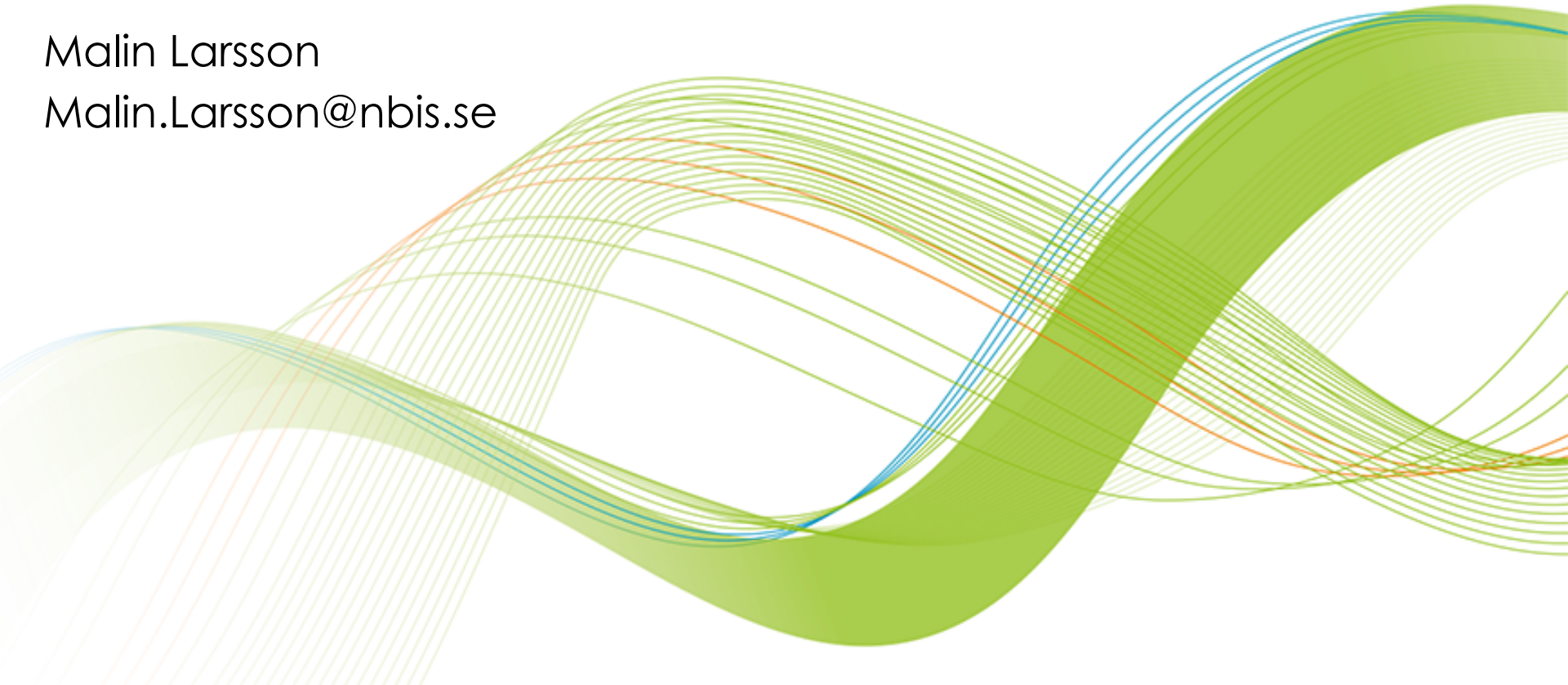
NBIS
NATIONAL BIOINFORMATICS INFRASTRUCTURE SWEDEN

SciLifeLab

# Quality Control of NGS data

Malin Larsson
Malin.Larsson@nbis.se



# Phred Quality Scores

```
+SEQ_ID
!''*(((((***+))%%%++)(%%%%).1***
```

A quality value Q is an integer representation of the probability *p* that the corresponding base call is incorrect.

$$Q = -10 \log_{10} P \quad \Longrightarrow \quad P = 10^{\frac{-Q}{10}}$$

| Phred Quality 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% |
| 50 | 1 in 100000 | 99.999% |

# ASCII encoded quality scores

Table 1 ASCII Characters Encoding Q-scores 0–40

| Symbol | ASCII Code | Q-Score | Symbol | ASCII Code | Q-Score | Symbol | ASCII Code | Q-Score |
|---|---|---|---|---|---|---|---|---|
| ! | 33 | 0 | / | 47 | 14 | = | 61 | 28 |
| " | 34 | 1 | 0 | 48 | 15 | > | 62 | 29 |
| # | 35 | 2 | 1 | 49 | 16 | ? | 63 | 30 |
| $ | 36 | 3 | 2 | 50 | 17 | @ | 64 | 31 |
| % | 37 | 4 | 3 | 51 | 18 | A | 65 | 32 |
| & | 38 | 5 | 4 | 52 | 19 | B | 66 | 33 |
| ' | 39 | 6 | 5 | 53 | 20 | C | 67 | 34 |
| ( | 40 | 7 | 6 | 54 | 21 | D | 68 | 35 |
| ) | 41 | 8 | 7 | 55 | 22 | E | 69 | 36 |
| * | 42 | 9 | 8 | 56 | 23 | F | 70 | 37 |
| + | 43 | 10 | 9 | 57 | 24 | G | 71 | 38 |
| , | 44 | 11 | : | 58 | 25 | H | 72 | 39 |
| - | 45 | 12 | ; | 59 | 26 | I | 73 | 40 |
| . | 46 | 13 | < | 60 | 27 | | | |

# FastQC

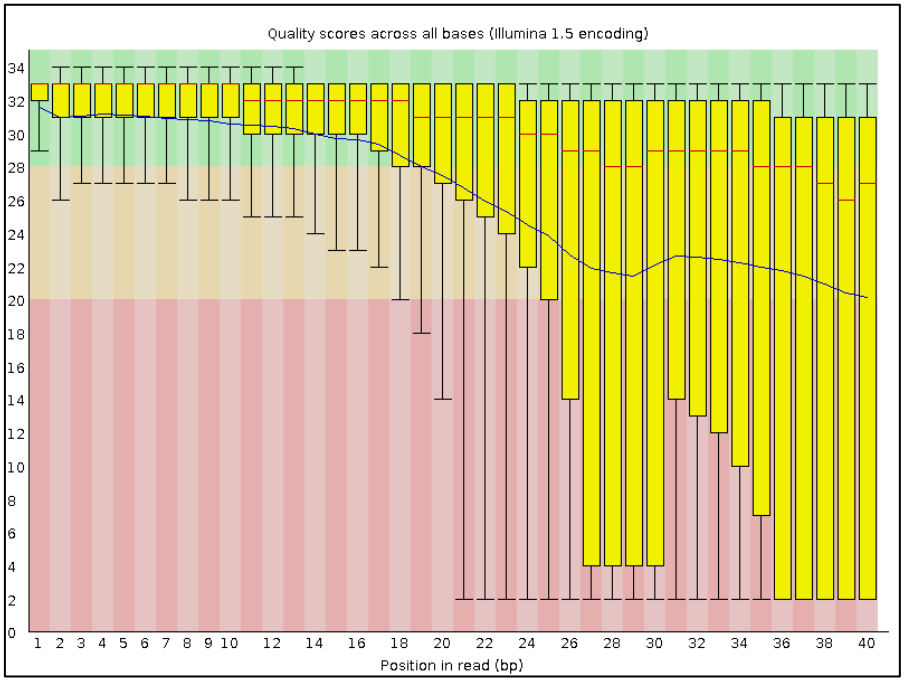
Babraham Institute

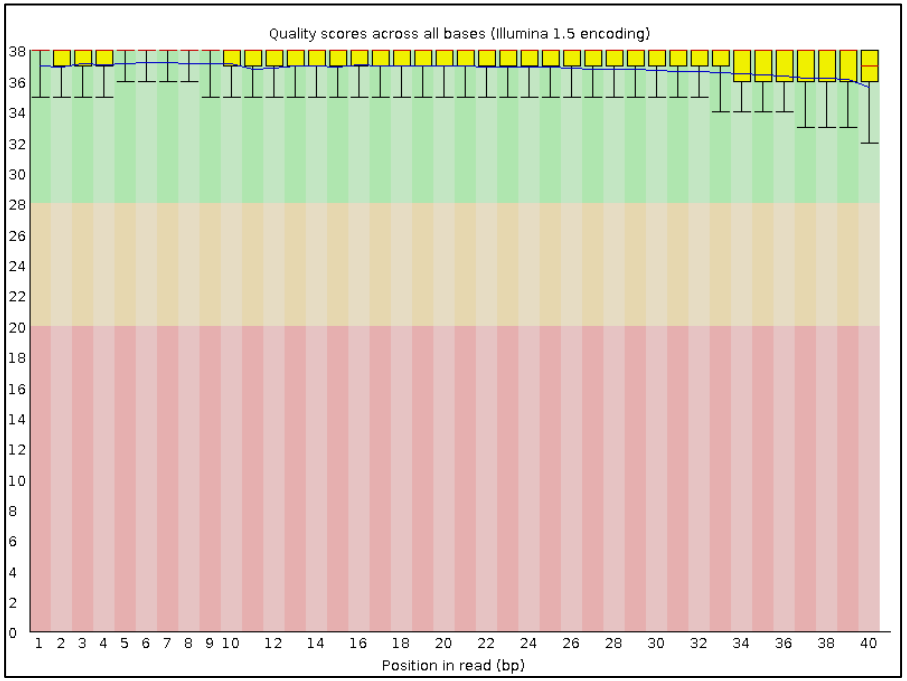
Babraham Bioinformatics

About | People | Services | Projects | Training | Publications

FastQC

Bad qualities:




Good qualities: