



# QCumber: A pipeline for automated QC of NGS data

Vivi Hue-Trang Lieu<sup>1</sup>, Piotr Wojciech Dabrowski<sup>1,\*</sup>, Bernhard Y. Renard<sup>1</sup>

## Introduction

Quality control of raw data is a critical step in any NGS data analysis. The potential of this step to influence all further analyses makes reproducibility and documentation especially important. While several pipelines for QC of NGS data exist, none specifically

addresses these points.

We present QCumber, a pipeline that is focused on performing the QC steps necessary in every NGS data analysis in an automated, reproducible and extensively documented way.

## Results

QCumber offers:

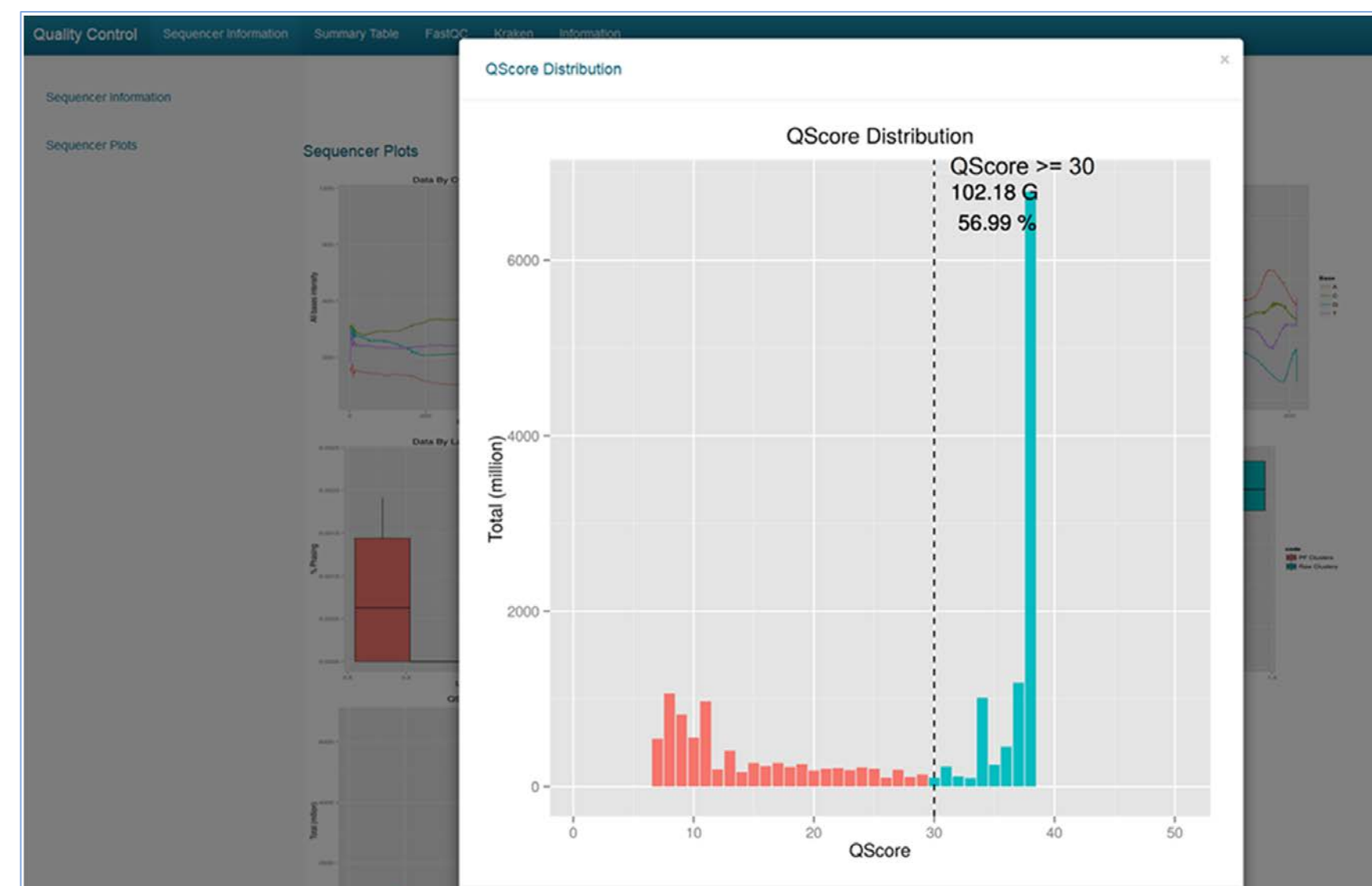
- Batch processing of samples
- Import of Illumina diagnostic data
- Trimming and visual QC
- Contaminant detection

Reports are intuitive and machine readable, allowing both manual data exploration and easy integration into external automated systems.

## Conclusion

QCumber is being actively employed in our bioinformatics service unit.

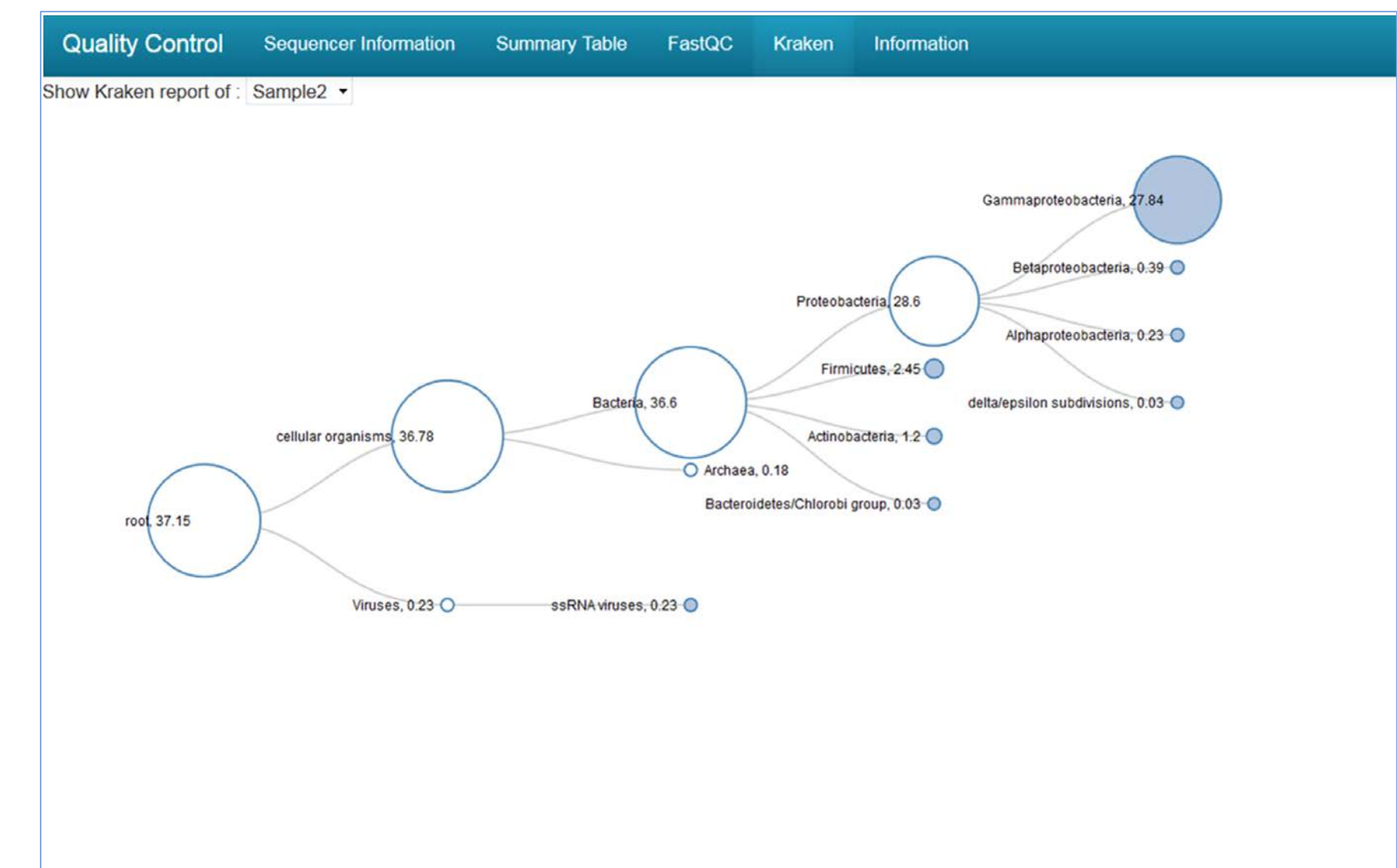
We believe that automated, comprehensive and mandatory documentation should be included in all pipelines used for data processing on the command-line and hope that QCumber is one step in this direction.



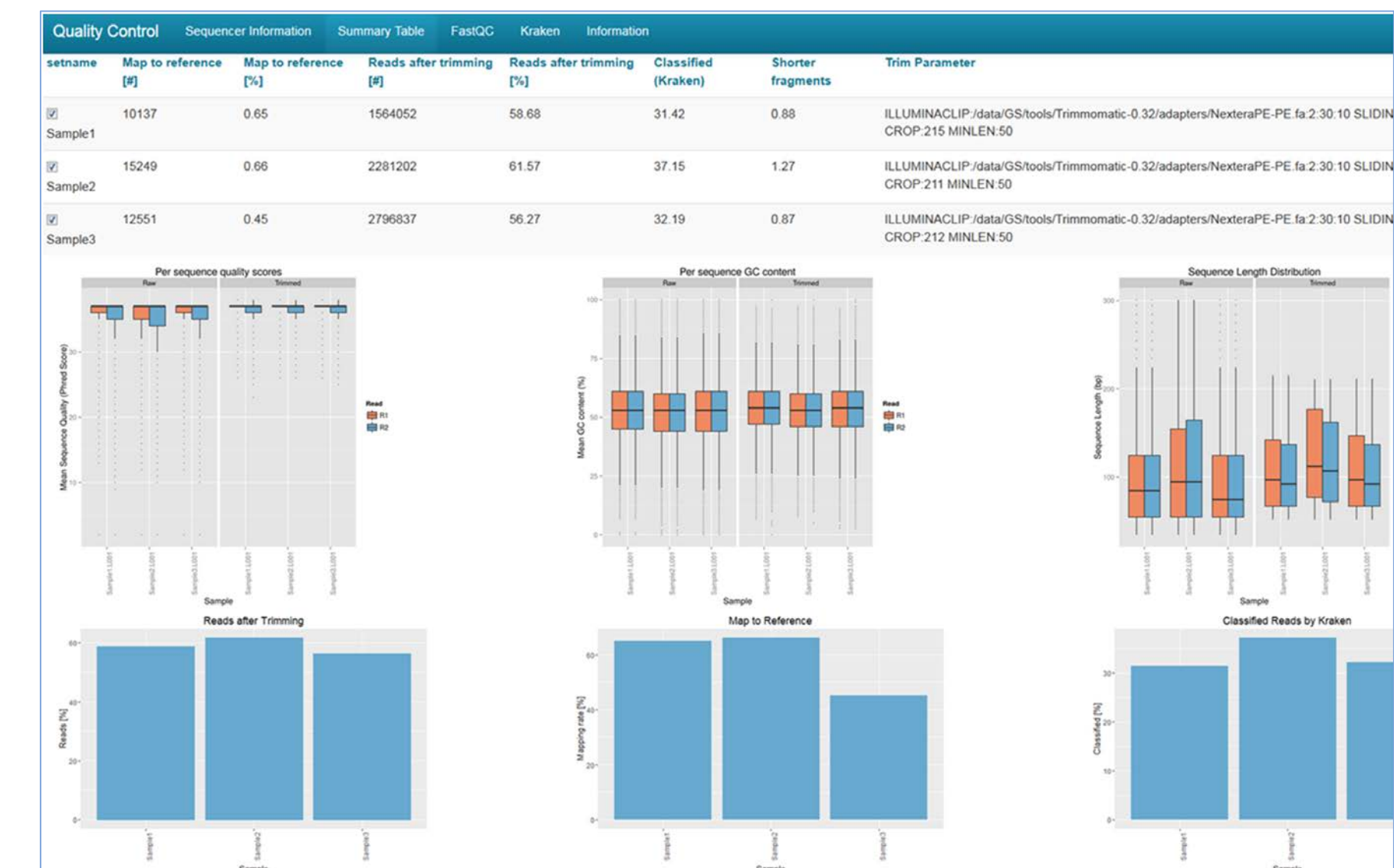
Illumina diagnostics



Comparison of pre- and post-trimming QC



Interactive KRAKEN visualization



Interactive batch report

<sup>1</sup>Research Group Bioinformatics (NG 4), Robert Koch Institute, Berlin, Germany

\*Contact: DabrowskiW@rki.de

Availability: <https://gitlab.com/RKIBioinformaticsPipelines/QCumber>