

Enhancing Data Integrity in Genome-wide Pooled Screens: The Role of oveQC in Quality Control

Jihyeob Mun*

†Supercomputing Application Center, Division of National Supercomputing, Korea Institute of Science and Technology Information

*Corresponding author: munji@kisti.re.kr

Genome-wide pooled screens have become an invaluable tool for functional genomics, enabling large-scale interrogation of gene function across various conditions. However, the complexity of datasets derived from next-generation sequencing (NGS) libraries, including those from Collecta, presents significant challenges in ensuring data integrity and reliability. In this study, we introduce oveQC, a novel quality control (QC) program specifically designed to address these challenges. By integrating multiple QC metrics, oveQC provides a comprehensive evaluation of pooled screen data, identifying potential issues that may compromise downstream analyses, such as off-target effects, sequencing biases, and sample contamination.

We applied oveQC to a large dataset generated using the Collecta library and demonstrated its effectiveness in detecting and mitigating quality-related issues. Our findings underscore the critical importance of robust QC procedures in genome-wide pooled screen experiments. Through systematic analysis, we reveal how overlooked quality issues can lead to inaccurate conclusions, potentially misleading further research. This study not only validates the utility of oveQC in enhancing the reliability of pooled screen data but also emphasizes the broader need for standardized QC practices in this field.

In conclusion, the integration of quality control measures, as exemplified by oveQC, is essential for ensuring the accuracy and reproducibility of genome-wide pooled screen analyses. This work serves as a foundation for future efforts to standardize QC protocols and improve the overall quality of data in functional genomics research.