

Quality Control Pipeline for Massively Parallel Reporter Assays (MPRAs)

Omer Ronen

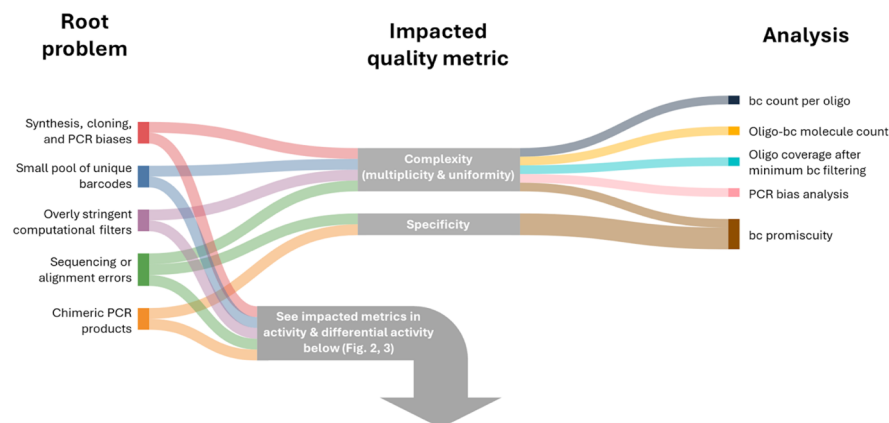
2025-12-28

Contents

Overview

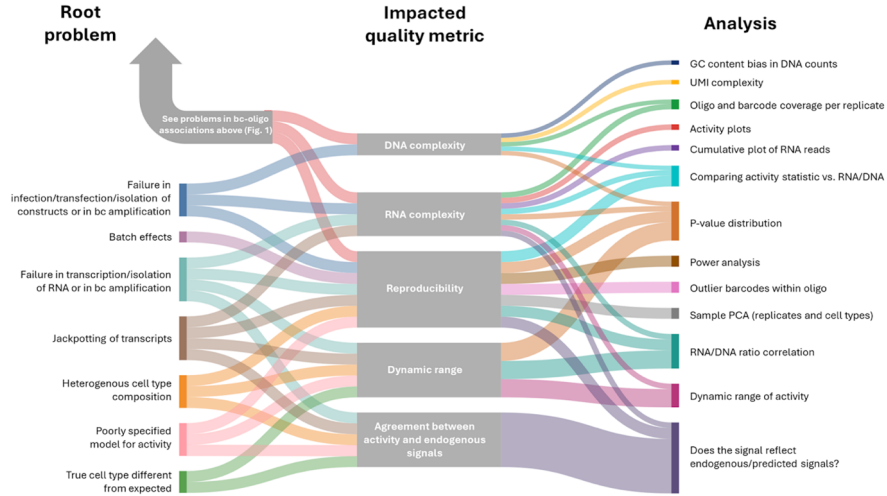
This pipeline is designed to help investigators evaluate the quality of their MPRA, quickly identify pitfalls, trace them to their source, and mitigate them. The scripts provided help ensure that the resulting MPRA data are suitable for robust statistical analysis and meaningful biological interpretation. This Bookdown accompanies our guide for best practices for MPRA, which outlines recommendations for study design and interpretation [REF]. The manuscript covers all key experimental and analytical steps, including library design, and estimation of activity differential activity. It then describes core problems that often compromise MPRA quality, illustrating how these issues manifest in the data, and offering practical strategies for correction and optimization. Because each issue can influence multiple quality metrics, and each metric may be affected by several issues, the relationships form a many-to-many network. The figures presented below map these interdependencies and connect them to recommended diagnostic analyses.

Usage



A scheme of root problems, the impacted quality metrics and analyses for the

cCRE-barcode association step.



Root problems, impacted quality metrics and recommended analyses for the RNA and DNA quantification step.

The quality control (QC) pipeline is organized into three chapters:

- (i) QC of the barcode association step
- (ii) QC of activity estimation
- (iii) QC of differential activity estimation

For each analysis, we provide an example of a successful and an unsuccessful dataset to illustrate how they manifest in the analysis.

We welcome questions, feedback, or suggestions. Please feel free to reach out at david.gokhman [at] weizmann.ac.il.

Scripts

All of these analyses are integrated into the quality control pipeline described in this resource, with scripts provided here: [\[link\]](#).

Chapter 1

A guide for running the analyses notebook

The QC pipeline has three main parts: Associations, Activity and Differential activity. For each part, there is a jupyter notebook file that enables you to run all the analyses that are presented in this book. Here we explain how to run these notebook files and what are the required inputs

1.1 Associations

First

1.1.1 Input

Input files

1.2 Activity

Second

1.2.1 Input

Input files

1.3 Differential Activity

Third

1.3.1 Input

Input files