Single-cell Morphology Quality Control (coSMicQC)

Dave Bunten[1], Jenna Tomkinson[1], Gregory Way[1]

[1]*Department of Biomedical Informatics, University of Colorado Anschutz Medical Campus*

Scan for coSMicQC!

Department of Biomedical Informatics
SCHOOL OF MEDICINE
UNIVERSITY OF COLORADO **ANSCHUTZ MEDICAL CAMPUS**

I. Erroneous outliers and analysis

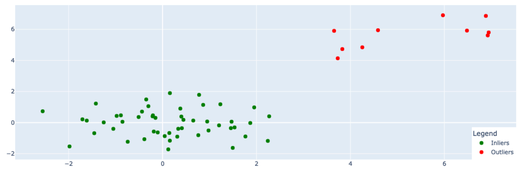


Figure 1: *Erroneous outlier anomalies can measure outside our expectations and impact analysis.*

Single-cell morphology data from high-throughput microscopy provide critical insights into disease mechanisms and therapeutic efficacy. However, segmentation errors during image analysis such as misidentifying cell compartments or artifacts as cells can lead to inaccurate single-cell measurements and erroneous anomalies within the data.

Researchers often resort to error-prone, bespoke filtering methods or aggregate data into bulk profiles to avoid discrepancies caused by anomaly outliers. These techniques fail to perform quality control on the data, often compromising the quality of single-cell profiles and impeding the potential for meaningful discoveries.

II. Single-cell quality control package

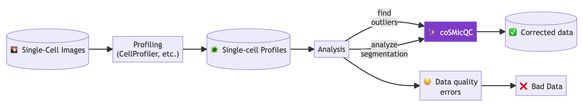


Figure 2: *coSMicQC enables high-quality data outcomes by checking for outliers.*

To address these challenges, we introduce ***coSMicQC (Single-cell Morphology Quality Control)***, an open source Python package designed to enhance the accuracy of single-cell morphology analysis. **`coSMicQC`** offers default and customizable thresholds for quality control, integrating seamlessly into both command line and Python API workflows.

`coSMicQC` features interactive visualizations that help users identify outlier distributions, and it introduces the ***CytoDataFrame*** — a novel data format that links single-cell measurements with their corresponding images and segmentation masks in real-time, enriching data analysis and interpretation.

III. Getting started with coSMicQC

☆ 1) ***Installation:***

```
# pip install from pypi
pip install coSMicQC

# install directly from source
pip install git+https://github.com/
WayScience/coSMicQC.git
```

`coSMicQC` may be installed from PyPI or source.

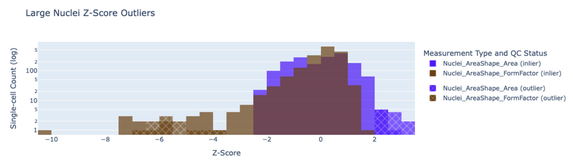
☆ 2) ***Finding outliers:***

```
import cosmicqc

# load a parquet file with single-cell
profile data and find outliers
scdf = cosmicqc.analyze.find_outliers(
    df="single-cell-profiles.parquet",
    metadata_columns=[
      "Metadata_ImageNumber",
      "Image_Metadata_Plate_x"
    ],
    feature_thresholds={
      "Nuclei_AreaShape_Area": -1
    },
)
```

```
Number of outliers: 328
Outliers Range:
Nuclei_AreaShape_Area Min: 734.0
Nuclei_AreaShape_Area Max: 1904.0
      Nuclei_AreaShape_Area  Metadata_ImageNumber Image_Metadata_Plate_x
23                    921.0                     2                Plate_2
28                    845.0                     2                Plate_2
29                   1024.0                     2                Plate_2
32                    787.0                     2                Plate_2
37                   1347.0                     2                Plate_2
...                     ...                   ...                    ...
```

☆ 3) ***Visualizing outliers:***



IV. Real-world impact

The effectiveness of coSMicQC is demonstrated through its application to the Joint Undertaking in Morphological Profiling (CPJUMP1) dataset. Researchers using coSMicQC have successfully identified and corrected technical outliers in single-cell profiles, leading to more accurate data interpretation and enhanced discovery potential. The package has been employed in various studies to refine single-cell morphology analysis and improve the reliability of research outcomes.

V. Future Steps

Moving forward, we aim to expand coSMicQC's capabilities by integrating additional quality control metrics and enhancing its compatibility with other data analysis platforms. We plan to incorporate feedback from the scientific community to further refine the tool and explore its application to other domains of microscopy and imaging research. Continued development will focus on improving user experience and extending the package's functionality to address emerging challenges in single-cell analysis.

VI. References

[Author(s) Last Name, Initial(s). (Year). Title of the reference. Journal Name, Volume(Issue), Page numbers. DOI] [Author(s) Last Name, Initial(s). (Year). Title of the reference. Conference Name, Location, Date.] [Software Documentation: coSMicQC. (Year). Title. URL]