

i. Erroneous outliers and analysis

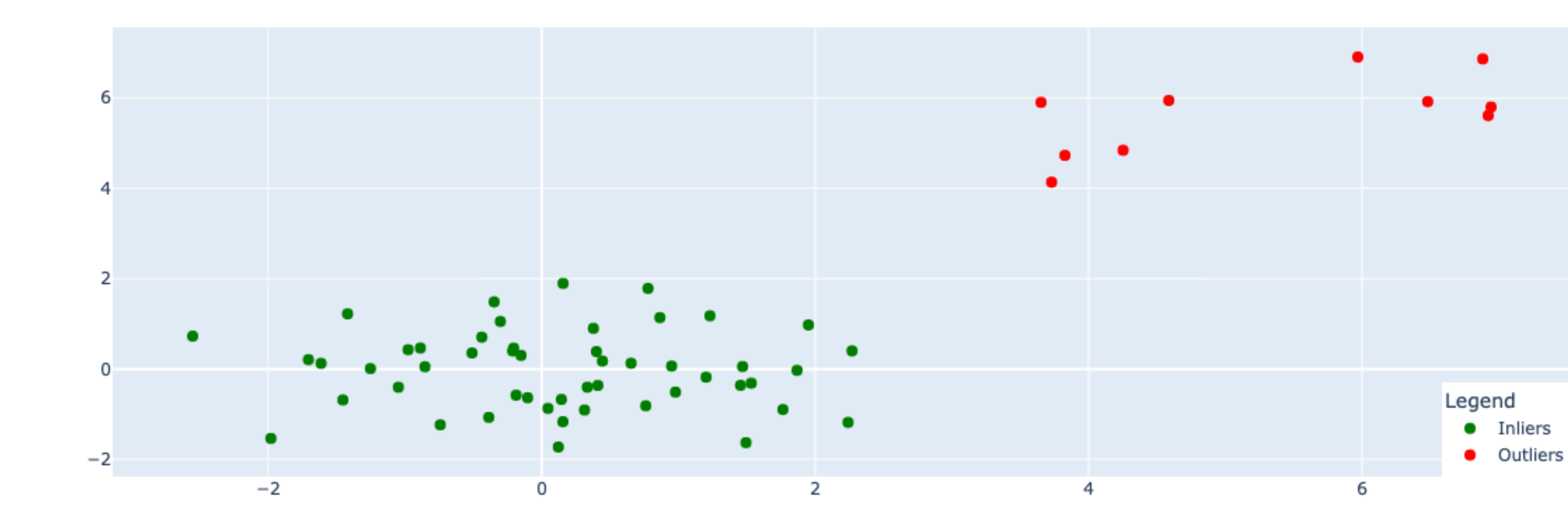


Figure 1: *Outlier anomalies measure outside our expectations.*

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. These inaccuracies emerge as ***anomalies***, data which do not conform to expectations, which complicate data interpretation and result in unreliable scientific conclusions.

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. coSMicQC: open source single-cell quality control

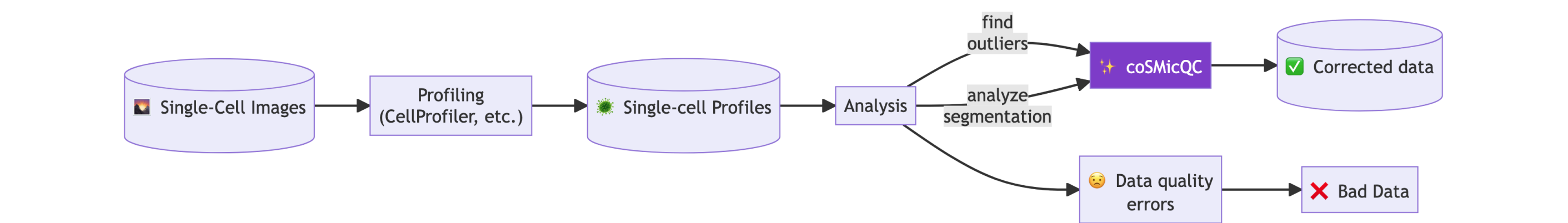


Figure 2: *coSMicQC enables high-quality data outcomes.*

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. When using **coSMicQC** through Jupyter notebooks it displays dynamic plots and labeled DataFrames for swift analysis iterations. Our goal is to create an open-science tool to study and eliminate erroneous data anomalies in a reproducible manner.

**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

See below for a quick guide on how to get started with **coSMicQC**.

1) Installation:

```
# pip install from pypi
pip install coSMicQC

# poetry install from pypi
poetry add coSMicQC

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

**coSMicQC** may be installed from PyPI or from source.

2) Finding 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]