

James Bonaffini^{1*}, Elizabeth Olesin^{1*}, Olivia Hwang², Stefani Nalle², Michael A. Meier¹, Monica Ge³, Bill O’Gorman¹, Mark McCreary³

*James Bonaffini and Elizabeth Olesin contributed equally to this work

¹Rancho Biosciences, LLC, Rancho Sante Fe, CA 92067, USA

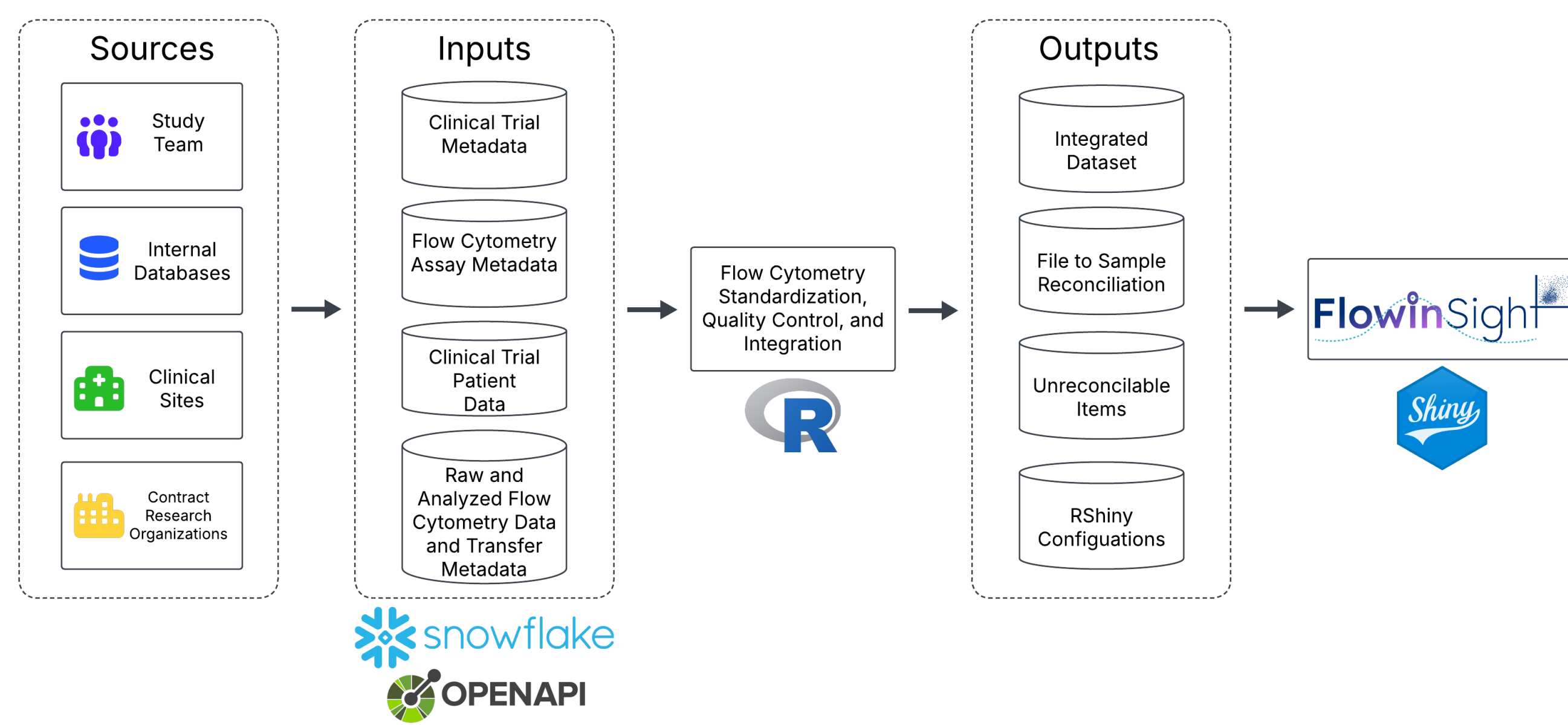
²Translational Medicine ODT, Genentech Inc., South San Francisco, CA 94080, USA

³Computational Catalysts, Computational Sciences Center of Excellence, Genentech Inc., South San Francisco, CA 94080, US

Introduction

Flow cytometry is a high-complexity assay used to assess therapeutic efficacy and immunological impact in exploratory, early phase clinical trials. The data delivered from Contract Research Organizations is often fragmented and in multiple formats including fcs files, pdf files, and analyzed tabular datasets. FlowInsight integration is a two-part solution to address the analytical and operational needs for research scientists and study managers in a clinical study setting. First, we developed an ETL pipeline that enables consistent and frequent data searching and integration across all onboarded studies. It searches for available data, standardizes them to the same format, reconciles sample assays across all sources, performs automatic QC checks, and outputs analysis-ready datasets. Second, we developed the FlowInsight RShiny application that enables scientific QC and analysis and operational data tracking. The FlowQC Pro module enables research scientists to analyze data longitudinally, view raw data files, and mark sample assays with QC status flags, comments, and actions. The CytoTracker module enables operational managers to view available samples by patient and visit and drill down into specific issues regarding assay completeness or missing files. Researchers and study managers may extract data issues from the FlowInsight app and communicate with vendors to take corrective action.

Flow Cytometry Data Integration and Standardization Pipeline



The Flow Cytometry Data Integration and Standardization Pipeline is a complex ETL script that extracts flow cytometry, patient, and study data. Basic standardization, integration, and QC is performed before loading final outputs into a location accessible by the FlowInsight RShiny application.

Pipeline Steps

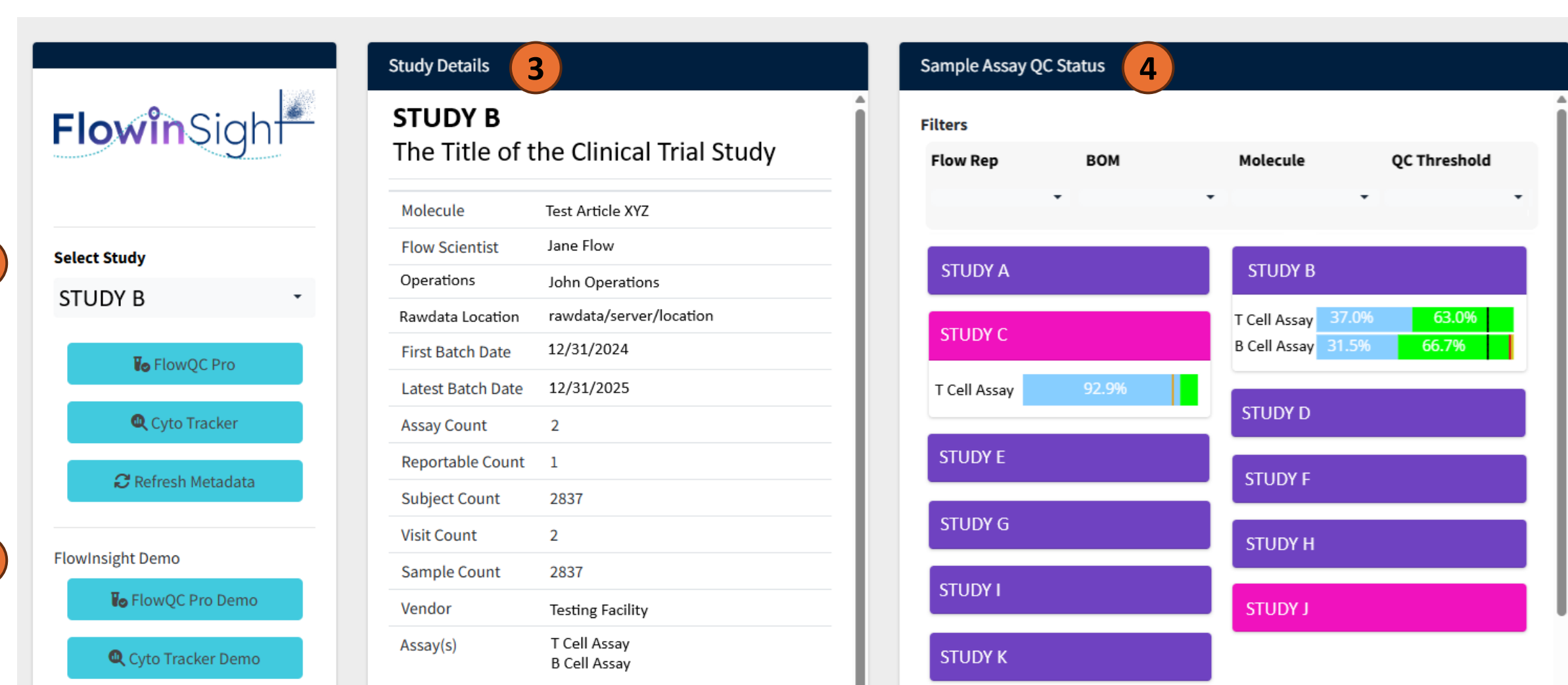
1. Extract study metadata, raw fcs and pdf flow data, tabular flow data, and clinical metadata into the environment.
2. Map raw and tabular data into a standard format, extracting key fields such as sample and assay.
3. Combine unique sample assays across sources, prioritizing the highest quality source for final output. Map raw data and tabular data to each other where they match.
4. Integrate with clinical metadata such as patient cohort and site.
5. Perform data validation to ensure all fields meet the expectations of research scientists.
6. Output final integrated table, reconciled sample assays, and app configurations to a database.

Pipeline Stages

1. Study Onboarding: Initial metadata loading (e.g. study assays, locations of data, vendors) is performed by programmatic searching of internal study databases and input from the study team.
2. Initial Study Integration: The pipeline is run interactively, outputting logs detailing the success of data standardization and integration. Issues are resolved either by updating the metadata loaded in step one or adjusting standardizing and mapping configurations. The FlowInsight RShiny application is checked to ensure data is properly available to view.
3. Study Maintenance: The study is scheduled to run non-interactively. A log can be viewed to assess standardization success in aggregate.

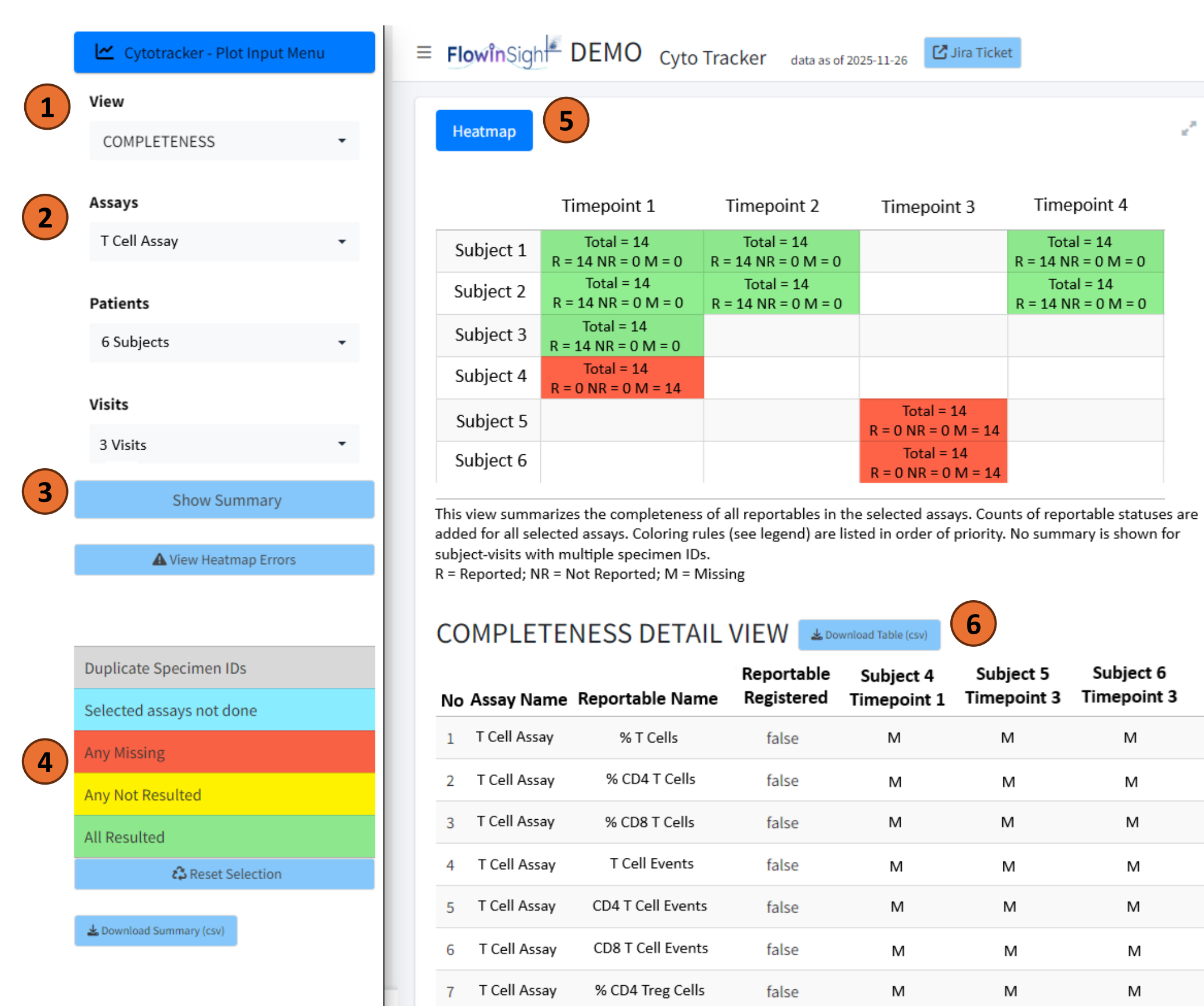
FlowInsight RShiny Application

FlowInsight Dashboard: High-Level Study Overview



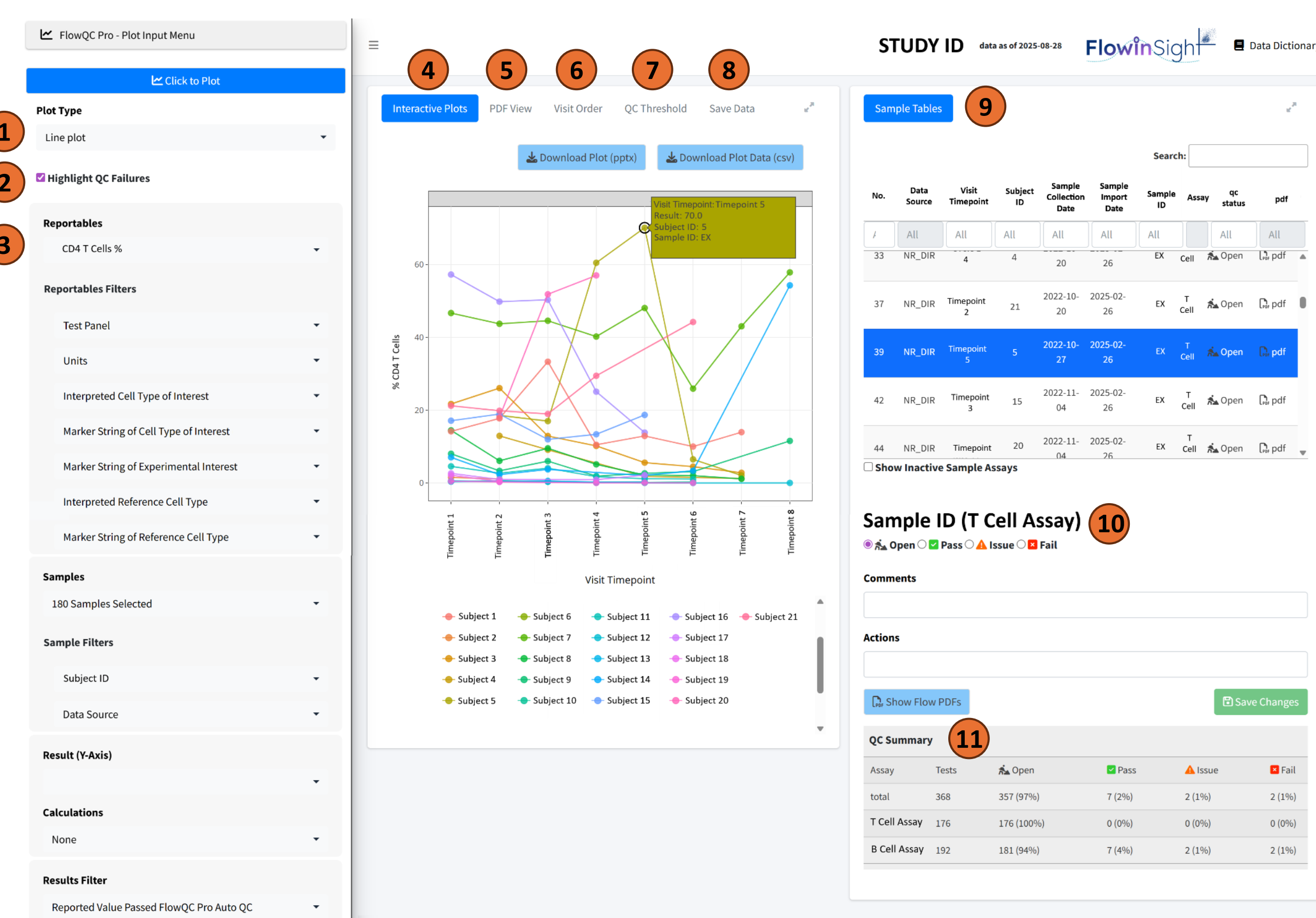
- 1 **Study Selection:** Study and Application Module Selection
- 2 **Module Demo:** Module Demo spaces containing mock data to allow individuals to test the application functionality
- 3 **Study Details:** Study information associated with a selected study
- 4 **Sample Assay QC Status:** Table with QC status charts and QC threshold lines. Studies with assays that do not meet QC threshold will highlight pink. Study filter functionality will populate table with studies that meet filter criteria

Cytotracker: Clinical Operations Data Tracker



- 1 **View:** Heatmap selection including views for completeness of reportables, duplicate sample IDs for a single subject/visit, turn around times, and number of files from vendor
- 2 **View Filters:** Allow user to filter for information of interest
- 3 **Heatmap Errors:** Populates unreconcilable information and duplicate specimen IDs
- 4 **Heatmap Legend:** Legend for heatmap colors. Selection populates Detail View
- 5 **Heatmap:** A heatmap table of subject-visit of selected corresponding View
- 6 **Detail View:** A breakdown to the reportable level of the selected values from the Heatmap or Heatmap Legend

FlowQC Pro: Data Visualization and QC Application



- 1 **Plot Selection:** User can define type of plot for data visualization
- 2 **Automatic QC:** Major data discrepancies can be identified programmatically and presented to the user for review or removed from visualization
- 3 **Customizable Filters:** Filters for reportable, sample, and result available to allow for dynamic data visualization. Additional trellis and color functions allow for additional complex visualization
- 4 **Interactive Plots:** Visualization of flow cytometry data based on user inputs
- 5 **PDF View:** Vendor-provided gating strategy file based on selected data point. Allows users to seamlessly review raw data files of interest
- 6 **Visit Order:** Allows user to define order of study visits to be visualized in Interactive Plots tab
- 7 **QC Threshold:** Allows user to define the required percentage of tests that must undergo QC for a given study. Visualized on Sample Assay QC Status of Dashboard Page
- 8 **Save Data:** Allows users to download integrated data table with QC comments for subsequent downstream use
- 9 **Sample Table:** Sample table that contains relevant sample metadata. Table is linked to the Interactive Plots for rapid insight into additional sample metadata
- 10 **User QC Input:** Allows users to input a QC status for a given data point as well as any comments or additional actions required
- 11 **QC Summary:** A table of the frequency of each QC status for a given assay within the study

Acknowledgements

Hailey Garma, Hanson Ho, Aaron Schroeder, Amelia Au-Yeung, Chaitanya Vala, Christina Cormier, and Yoonhee Choi

All trademarks belong to their respective owners

