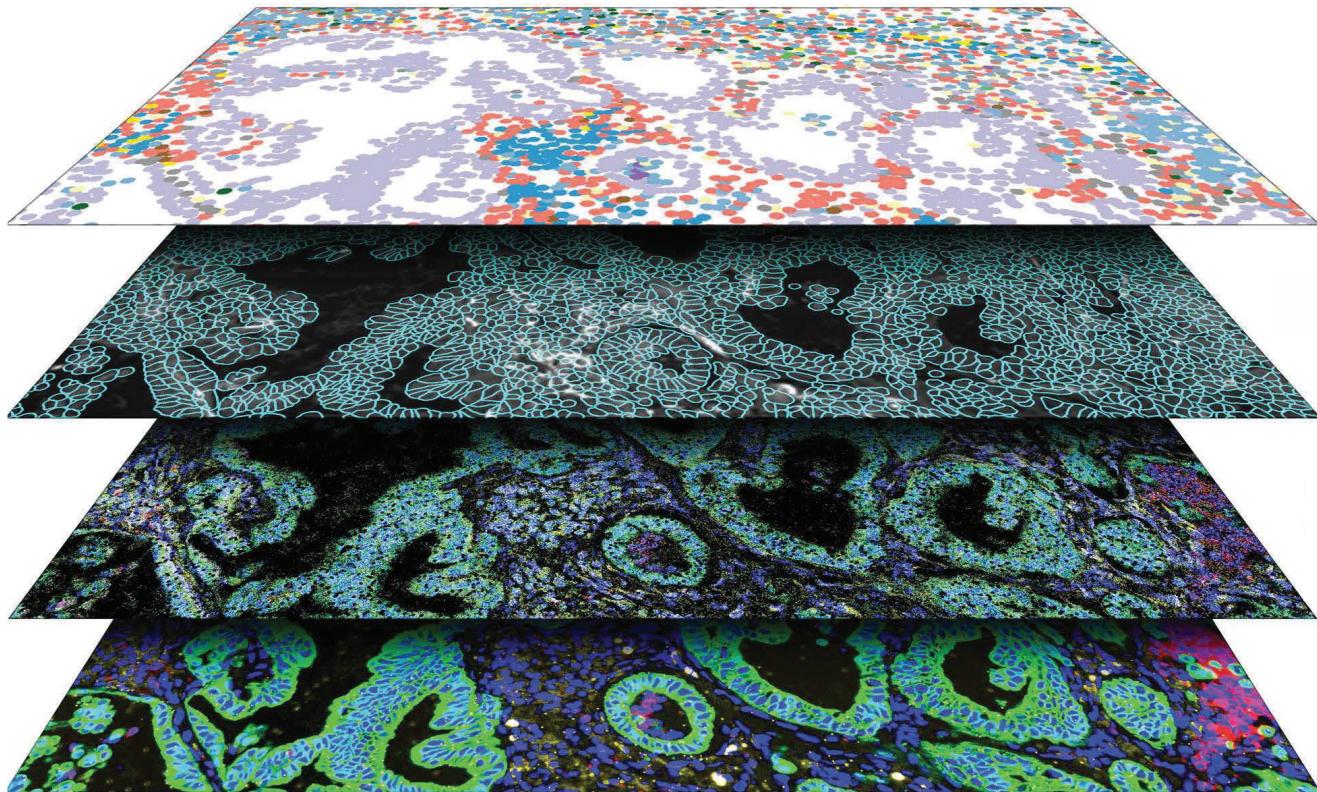


Data Analysis



MAN-10162-02 | Feb 2023

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

©2022-2023 NanoString Technologies, Inc. All Rights Reserved.

nanoString
nanostinger.com

NanoString, Inc.
530 Fairview Ave N
Seattle, Washington 98109

www.nanostring.com

T: 888.358.NANO (6266)
F: 206.378.6288
E: Support@nanostring.com

SalesContacts

United States: us.sales@nanostring.com

EMEA: europe.sales@nanostring.com

Asia Pacific & Japan: apac.sales@nanostring.com

Other Regions: info@nanostring.com



EU Authorized Representative

NanoString Technologies
Germany GmbH
Birketweg 31
80639 Munich
Germany

UK Authorized Representative

NanoString Technologies
Europe Limited
11th Floor Whitefriars
Lewins Mead
Bristol BS1 2NT
United Kingdom

Rights, License, and Trademarks

Use

Use for Research use only. Not for use in diagnostic procedures.

Intellectual Property Rights

This CosMx™ Spatial Molecular Imager (SMI) User Manual and its contents are the property of NanoString Technologies, Inc. (“NanoString”), and are intended solely for use by NanoString customers, for the purpose of operating the CosMx SMI System. The CosMx SMI System (including both its software and hardware components), this User Manual, and any other documentation provided to you by NanoString in connection therewith, are subject to patents, copyright, trade secret rights and other intellectual property rights owned by, or licensed to, NanoString. No part of the software or hardware may be reproduced, transmitted, transcribed, stored in a retrieval system, or translated into other languages without the prior written consent of NanoString. For a list of patents, see www.nanostring.com/company/patents.

Limited License

Subject to the terms and conditions of the CosMx SMI System contained in the product quotation, NanoString grants you a limited, non-exclusive, non-transferable, non-sublicensable, research use only license to use the proprietary CosMx SMI System only in accordance with the manual and other written instructions provided by NanoString. Except as expressly set forth in the terms and conditions, no right or license, whether express, implied or statutory, is granted by NanoString under any intellectual property right owned by, or licensed to, NanoString by virtue of the supply of the proprietary CosMx SMI System. Without limiting the foregoing, no right or license, whether express, implied or statutory, is granted by NanoString to use the CosMx SMI System with any third party product not supplied or licensed to you by NanoString or recommended for use by NanoString in a manual or other written instruction provided by NanoString.

Trademarks

NanoString, NanoString Technologies, the NanoString logo, CosMx, AtoMx, GeoMx, and nCounter are trademarks or registered trademarks of NanoString Technologies, Inc., in the United States and/or other countries. All other trademarks and/or service marks not owned by NanoString that appear in this document are the property of their respective owners.

Open Source Software Licenses

Visit <http://nanostring.com/cosmx-oss> for a list of open source software licenses used in CosMx Spatial Molecular Imaging.

Copyright

©2022-2023 NanoString Technologies, Inc. All rights reserved.

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

Table of Contents

CosMx SMI Data Analysis User Manual	1
Contacts	2
Rights, License, and Trademarks	3
Table of Contents	4
Conventions	6
CosMx SMI User Manuals and Resources	8
CosMx SMI Data is Managed in the AtoMx Spatial Informatics Platform	9
Introduction to AtoMx SIP	10
Connecting CosMx SMI to AtoMx SIP	10
Accessing AtoMx SIP	10
Orientation to AtoMx SIP	11
Create and Open a Study	12
Delete a Study	14
Orientation to CosMx SMI Data Analysis Suite	15
Manage Annotations	21
Create and Run a Pipeline	23
Custom Modules	25
Glossary of CosMx SMI Pipeline Modules	27
Initial Data	27
Quality Control - RNA	27
Quality Control - Protein	29
Normalization - RNA	30
Normalization - Protein	31
Principal Component Analysis (PCA) - RNA or Protein	32
UMAP - RNA or Protein	33
Cell Typing - RNA	35
Cell Typing - Protein	37

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

Table of Contents

Neighbor Network: Expression Space - RNA or Protein	37
Leiden Clustering - RNA or Protein	38
Identify Marker Genes - RNA or Protein	39
Neighborhood Analysis - RNA or Protein	39
Ligand-Receptor (LR) Analysis - RNA	40
Spatial Network - RNA or Protein	40
Cell Type Co-Localization - RNA or Protein	41
Spatial Expression Analysis - RNA or Protein	42
Pathway Analysis - RNA	43
Pre-Differential Expression - RNA	43
Differential Expression - RNA	43
Glossary of Data Visualizations	44
Study Statistics	44
XY Plot	44
Heatmap	45
Box Plot and Violin Plot	45
Histogram	46
PCA Plot	46
UMAP Plot	47
Volcano Plot	47
Flightpath Plot	48
Save a Visualization	49
Export Data	50
Export Output	53
Troubleshooting	56
Appendix: Literature References	57
Final page	58

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

Conventions

The following conventions are used throughout this manual and are described for your reference.

Bold text is typically used to highlight a specific button, keystroke, or menu option. It may also be used to highlight important text or terms.

Blue underlined text is typically used to highlight links and/or references to other sections of the manual. It may also be used to highlight references to other manuals and/or instructional material.

The gray box indicates general information that may be useful for improving assay performance. The notes may clarify other instructions or provide guidance to improve the efficiency of the assay work flow.



WARNING: This symbol indicates the potential for bodily injury or damage to the instrument if the instructions are not followed correctly. Always carefully read and follow the instructions accompanied by this symbol to avoid potential hazards.



IMPORTANT: This symbol indicates important information that is critical to ensure a successful assay. Following these instructions may help improve the quality of your data.

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

CosMx SMI Workflow Overview

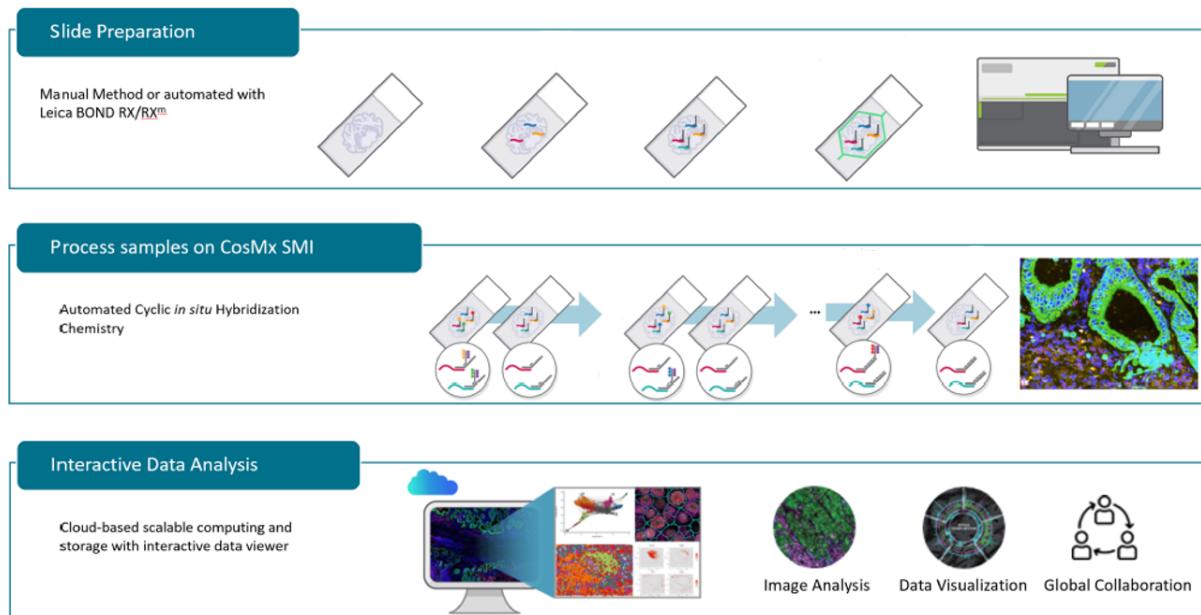


Figure 1: CosMx SMI workflow overview

Day 1: Slide Preparation. Prepare slides manually or using the BOND RX/RX^m fully automated IHC/ISH stainer from Leica Biosystems (BOND RX/RX^m).

Day 2: Process Slides on CosMx SMI. Complete assay and assemble the flow cells. Load assembled flow cells into the CosMx SMI instrument and enter flow cell/study information. Tissue is scanned to capture RNA or Protein readout and morphology imaging within user-designated fields of view (FOVs).

After run completion: Create a **Data Analysis** study in the AtoMx Spatial Informatics Platform (SIP) and perform quality-control checks, data analysis, and generate analysis plots.

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

CosMx SMI User Manuals and Resources

The CosMx SMI workflow is divided into the following user manuals:

Workflow Step 1	<u>CosMx SMI Manual Slide Preparation User Manual</u> MAN-10159
	<u>CosMx SMI Semi-Automated Slide Preparation User Manual</u> MAN-10160
Workflow Step 2	<u>CosMx SMI Instrument User Manual</u> MAN-10161
Workflow Step 3	<u>CosMx SMI Data Analysis User Manual</u> MAN-10162

User manuals and other documents can be found online in the NanoString University Document Library at <https://university.nanostring.com>.

Instrument and workflow training courses are also available in NanoString University.

For information about the AtoMx™ Spatial Informatics Platform, please refer to the [AtoMx Spatial Informatics Platform User Manual \(MAN-10170\)](#).

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

CosMx SMI Data is Managed in the AtoMx Spatial Informatics Platform

After primary processing on-instrument, data from NanoString's CosMx™ Spatial Molecular Imager (SMI) is automatically sent to the cloud-based AtoMx™ Spatial Informatics Platform (SIP) for further processing, storage, management, and collaboration. All CosMx SMI data analysis is performed within AtoMx SIP in the CosMx SMI Data Analysis Suite ([see Figure 2](#)).

This CosMx SMI Data Analysis User Manual outlines the steps to analyze data in the Data Analysis Suite using the pipeline orchestrator and easy-to-use modules.

Learn more about spatial data analysis through [NanoString University](#) and NanoString's [Publications page](#). Data analysis resources are also available through NanoString's [Spatial Data Analysis Service](#).

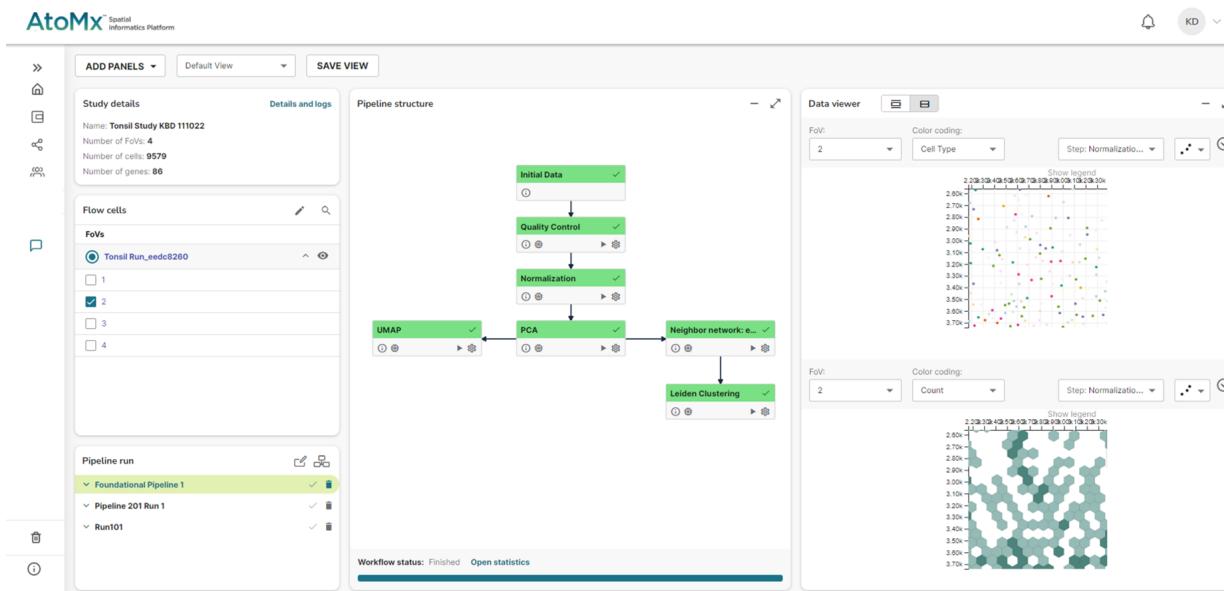


Figure 2: The AtoMx SIP cloud platform houses the CosMx SMI Data Analysis Suite

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

Introduction to AtoMx SIP

The AtoMx Spatial Informatics Platform is a cloud-based informatics platform that provides an integrated workspace with streamlined workflows to manage, analyze and share data from CosMx SMI and GeoMx® Digital Spatial Profiler (DSP). AtoMx SIP provides compute capabilities that scale with data sets and the complexity of the analysis pipeline. In addition, AtoMx SIP provides easy access to analysis methods for CosMx SMI through a pipeline orchestrator that enables users to create custom analysis pipelines.

Connecting CosMx SMI to AtoMx SIP

The CosMx SMI instrument is registered to AtoMx SIP during the instrument's installation or user training. Once registered, data transfer from the CosMx SMI instrument to the AtoMx SIP occurs automatically. Please refer to the [CosMx SMI Site Preparation Guide \(MAN-10171\)](#) and [CosMx SMI Instrument User Manual \(MAN-10161\)](#) for additional information and to address situations that require manual transfer or intervention.

Accessing AtoMx SIP

Log in to AtoMx SIP at <https://atomx.nanostring.com> or your organization's custom URL, if assigned. Please refer to the [AtoMx SIP User Manual \(MAN-10170\)](#) for more information about account set-up.

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

Orientation to AtoMx SIP

The AtoMx SIP Home page and key features are shown below ([see Figure 3](#)). For more details, please refer to the [AtoMx SIP User Manual \(MAN-10170\)](#).

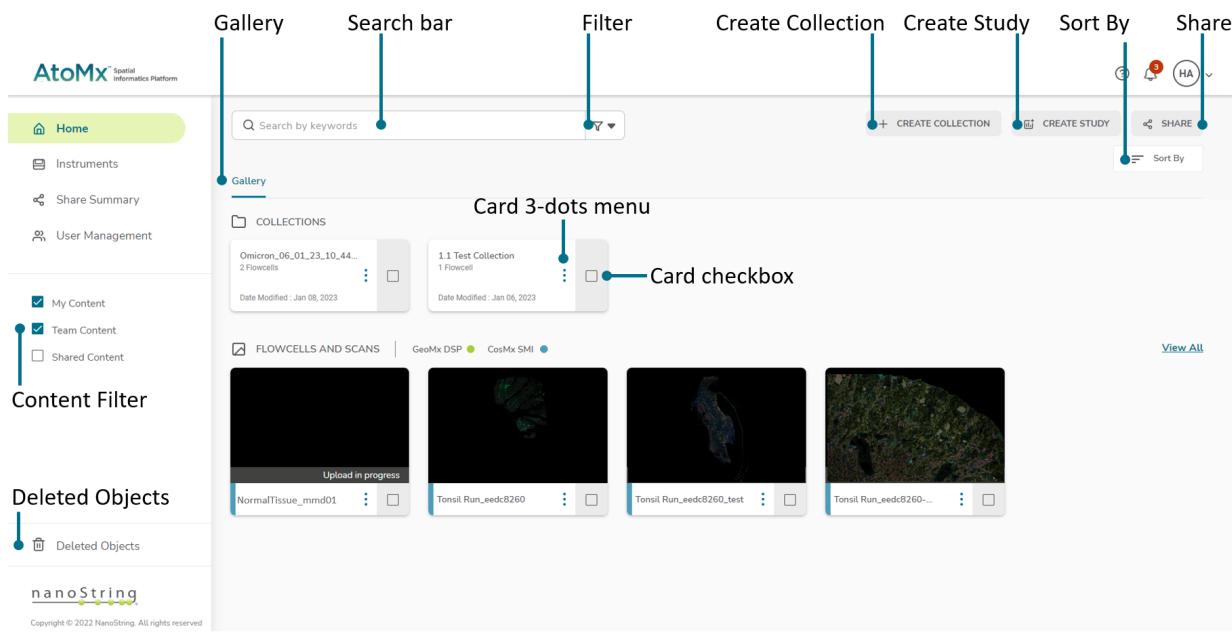


Figure 3: Home page with key components labeled

From the left menu, navigate within sections of the platform. Please note that some sections are not visible to AtoMx SIP Free Users.

The **gallery** is accessible through the **Home** tab, and displays the data collections, studies, CosMx SMI flow cells, and GeoMx DSP scans on individual **cards**. CosMx SMI objects are labeled in teal and GeoMx DSP objects in green. The most-recently accessed objects are displayed. Click on **View All** to view all objects.

Collections are groups of data objects (flow cells, scans, and studies). Individual objects can be part of more than one collection. Deleting a collection does not delete the individual objects in the collection.

Use the **search bar** at top to locate specific objects of interest using keywords, and click the **filter icon** in the search bar to open a filter pane.

Use the **Sort By** tool (top right) and the **Content Filters** (checkboxes at left) to locate objects of interest.

Click the **3 dots (⋮)** on a card for options including **View**, **Add to collection**, **View collaborators**, **View comments**, **View details**, and **More options (Delete)**.

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

Create and Open a Study

Studies can be created from data that has completed transfer (upload) from the instrument. Studies can be created from CosMx SMI flow cells or GeoMx DSP scans, but not both together.

To create a study:

1. From the AtoMx SIP gallery, click the **checkbox** on the card(s) of the flowcells of interest. (If the object is marked **Upload in progress**, it is not yet available for study creation.) The **Create Study** button in the top-right becomes active ([see Figure 4](#)).

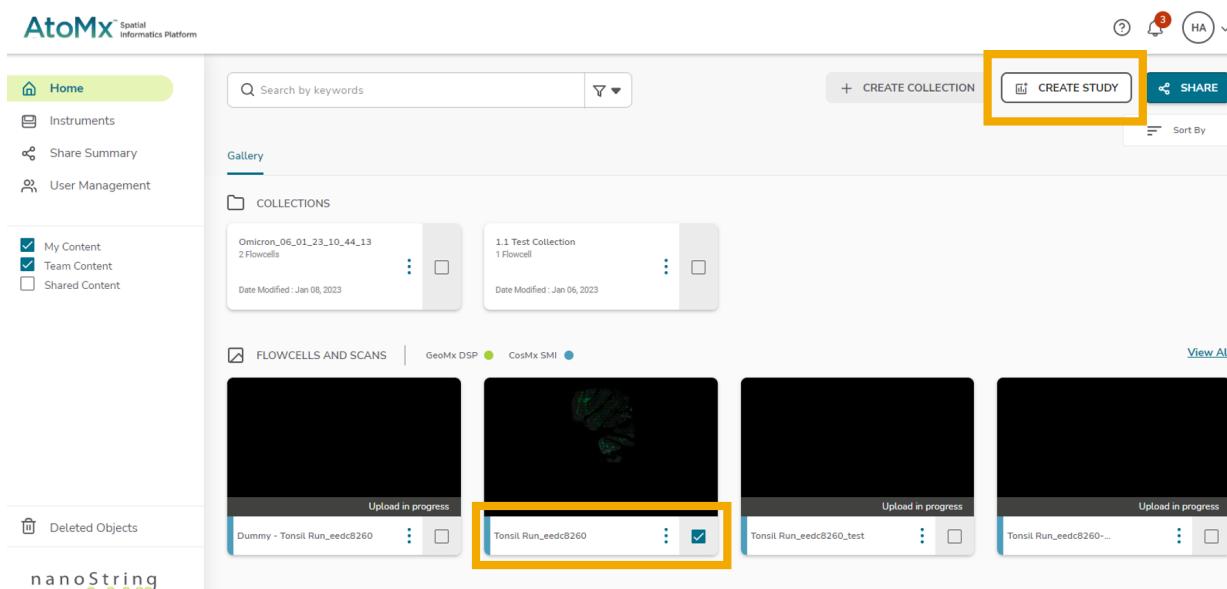


Figure 4: Select a flowcell to activate the Create Study button

2. Click **Create Study**.
3. In the **Create Study** window ([see Figure 5](#)), enter the **Study Name**, **Description**, and search **Tags** to annotate the study. Click **Create**.

During study creation, flowcell data is aggregated and organized to a data object. This process takes longer for studies with more flowcells and FOV (e.g. 15 min for a dataset of 4 FOV and 2.5 hr for a dataset of 153 FOV).

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

Create/Open a Study

The screenshot shows the 'Create Study' window. At the top right are 'CANCEL' and 'CREATE' buttons. The main area contains fields for 'Study Name' (with placeholder 'Enter Study Name'), 'Description' (with placeholder 'Enter Description'), and 'Tags' (with placeholder 'Enter Tags'). Below these is a section titled 'List of Flowcells/Scans' which lists 'Tonsil Run_eedc8260'. There is also a small circular icon with a question mark.

Figure 5: The Create Study window

To open this or any existing study, navigate to your studies in the Home gallery and click on a study to open it in the CosMx SMI Data Analysis Suite.

A **study creation log** records the software's steps in creating the study. If needed for troubleshooting, download the study creation log from the Pipeline Structure panel (prior to any pipeline run) or the Study Details Panel ([see Figure 6](#)) of the Data Analysis Suite.

The screenshot shows the 'Study details' panel. At the top are buttons for 'ADD PANELS', 'Default View', and 'SAVE VIEW'. The 'Details and logs' tab is selected and highlighted with a yellow box. The 'Study details' section shows: Name: KBD 120911 Study, Number of FoVs: 153, Number of cells: 1469685, Number of genes: 1000. The 'Flow cells' section lists 'FoVs' with 'Lymph RunA1118-S2' selected and checked. On the right, under 'Pipeline structure', it says 'No pipeline selected' and has 'RUN NEW PIPELINE' and 'DOWNLOAD LOGS' buttons, with 'DOWNLOAD LOGS' highlighted with a yellow box. Below that is a message 'No details available for this study'.

Figure 6: Download study creation logs

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

Delete a Study

Delete a Study

To delete a study:

1. Click the 3-dots menu (⋮) on the study's card in the gallery and select **More Options**, then **Delete** ([see Figure 7](#)).
2. A confirmation window opens to confirm deleting the study from the gallery. If confirmed, the study is moved to Deleted Objects, accessible from the menu at left.
3. To review deleted objects, click **Deleted Objects** ([see Figure 8](#)). Objects may be **restored** to the gallery or **permanently deleted**, depending on the permission level of the user, by checking the object's check box and selecting a button in the top right.

Deleting a collection does not delete the individual items in the collection.

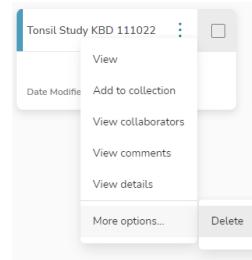


Figure 7: Delete a study

A screenshot of the 'Deleted Objects' folder. On the left, there is a sidebar with navigation links: Home, Instruments, Share Summary, User Management, My Content (checked), Team Content, and Shared Content. Below this is a 'Deleted Objects' section with a green button labeled 'Deleted Objects'. The main area shows a list of deleted study cards:

- 20220830 Study KAK (checkbox checked, delete button highlighted)
- 20220905 KAK Study
- Test Sept9 Peter
- 20220905 Study KAK 3

 At the top right of the list are 'DELETE' and 'RESTORE' buttons, both highlighted with orange boxes. There is also a 'Sort By' dropdown menu. A 'Search by keywords' bar is at the top center.

Figure 8: Deleted Objects folder and Delete/Restore buttons

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

Orientation to CosMx SMI Data Analysis Suite

CosMx SMI studies open in the SMI Data Analysis Suite ([see Figure 9](#)).

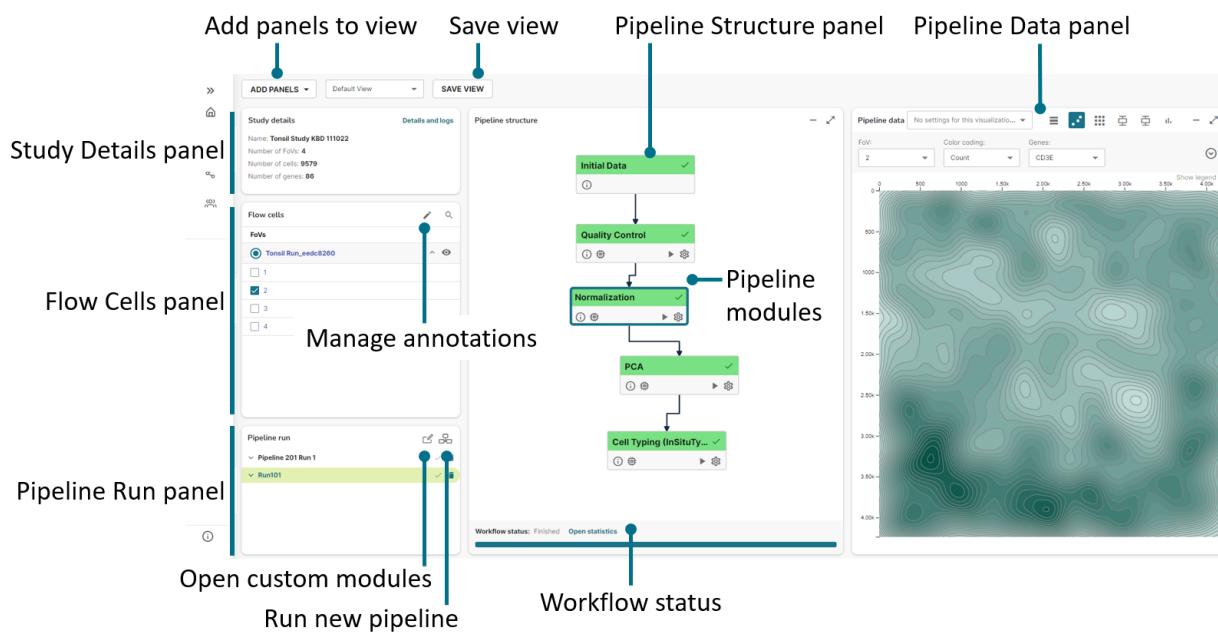


Figure 9: Orientation to the CosMx SMI Data Analysis Suite

Different **panels**, described in the following pages, can be displayed or hidden to provide a customized workspace. Choose **Default View** or **Default Data View** from the dropdown menu (see below), or customize by clicking **Add panels**. You may need to scroll to the right to view all displayed panels. Minimize or maximize a panel using controls in its top-right. Click and drag panels to arrange the view to your liking, then save your customized view by clicking **Save view**.

Default View / Default Data View

These are likely to be the most used views and are provided as presets in a dropdown menu:

- Default View: Study Details, Flow Cells, Pipeline Run, Pipeline Structure, and Pipeline Data panels.
- Default Data View: Image Viewer and Data Viewer panels.

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

Study Details Panel

Includes information about the study, including the number of Fields of View (FOV) analyzed and the number of cells and genes detected ([see Figure 10](#)).

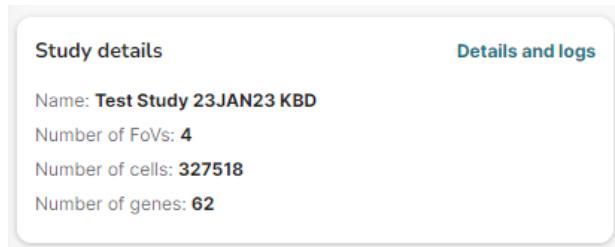


Figure 10: Study Details panel

Flow Cells Panel

Indicates which FOV is under analysis ([see Figure 11](#)). Click the **pencil** icon to open the Manage Annotations window (see [Manage Annotations on page 21](#)). Click the **Search** (magnifying glass) icon to search for a particular FOV by name. For a particular flow cell, click the **arrow** (carat) to view a list of the FOVs in the flow cell. **Uncheck** the box next to an FOV to de-select it from the Image Viewer panel. Click the **eye** icon to view the flow cell in the Image Viewer and FOV Viewer (see descriptions [on page 19](#)).

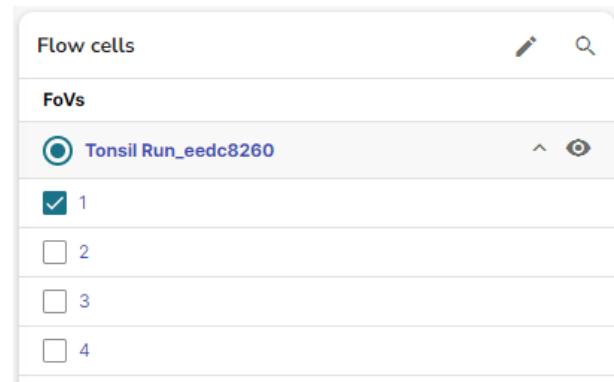


Figure 11: Flow Cells panel

Pipeline Run Panel

Displays executed pipelines ([see Figure 12](#)). Click the arrow (carat) to display modules comprising the pipelines. Click the pipeline icon to run a new pipeline (see [Create and Run a Pipeline on page 23](#)).

Click the custom modules icon to open Custom Modules (see [Custom Modules on page 25](#)).

Click the trash icon to delete the pipeline from the Pipeline Run panel.



Figure 12: Pipeline Run panel

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

Orientation to CosMx SMI Data Analysis Suite

Pipeline Structure Panel

Presents a diagram of the data analysis pipeline selected in the Pipeline Run panel. Individual pipeline modules are depicted as blocks ([see Figure 13](#)) with icons representing the following tools:

- Information about the module
- Show resource metrics (execution time, peak memory, average memory, peak CPU, average CPU) with option to download metrics and logs for this module.
- Mark for interactive run (for custom modules with live interaction; opens live terminal while module runs).
- Rerun step.
- Show settings (description, input parameters, output visualizations).
- Download files associated with module output.

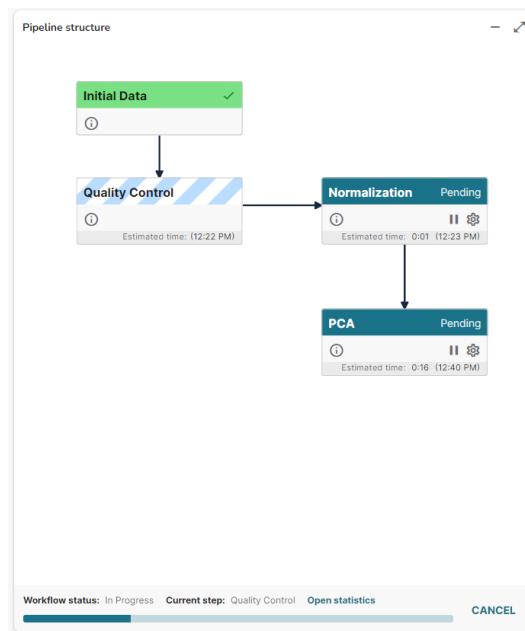


Figure 13: Pipeline Structure panel

When a workflow is running, its **status** is depicted at the bottom of the Pipeline Structure panel. Modules successfully completed are displayed in green, while modules in progress are striped and modules pending an upstream step are teal. Modules which failed are red.

For descriptions of each module, please see the [Glossary of CosMx SMI Pipeline Modules on page 27](#).

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

Pipeline Data Panel

Displays data visualizations for the module selected in the Pipeline Structure panel. Available visualization types vary depending on the module run. For example, the Normalization module enables Study Statistics, XY plot, Heatmap, Box plot, Violin plot, and Histogram ([see Figure 14](#)) while the PCA module enables Study Statistics and PCA plot. For descriptions of all visualization types and their customizable settings, please see the [Glossary of Data Visualizations on page 44](#).

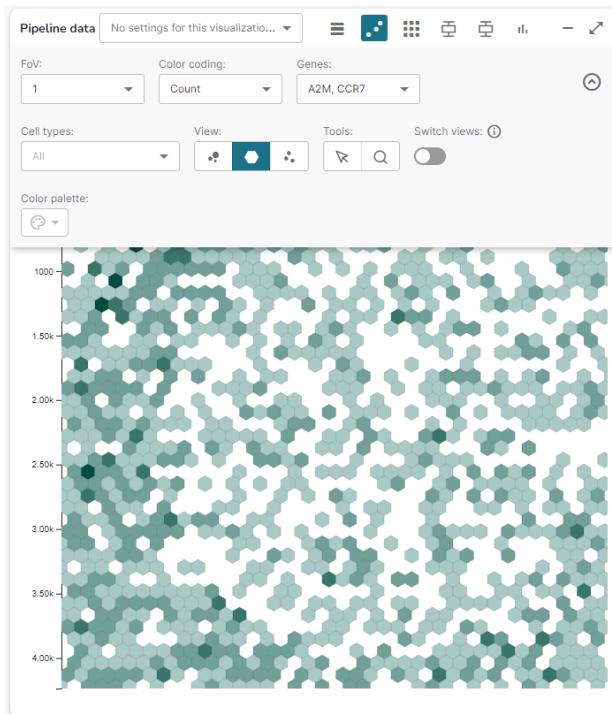


Figure 14: Pipeline Data panel - honeycomb plot example

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

Orientation to CosMx SMI Data Analysis Suite

Image Viewer Panel

Displays the tissue image and (when zoomed in) cell segmentation ([see Figure 15](#)), with the option to overlay data. FOVs are bordered by white boxes.

Toggle the **minimap** and channel **legend** display. Pan across the image by clicking and dragging in the minimap or main (larger) image. Use the **zoom** controls (+ -) or mouse scroll wheel to zoom.

Select the **flow cell** and **FOV(s)** to display from the dropdown menus. Jump between FOV using the left/right **arrows** (carats). Toggle **Show data points** to **on** to open additional customizable settings, including the pipeline step for which to show data points, coloring scheme, genes of interest, and cell types of interest. Under the **Color** dropdown menu, note that **Transcript** is not enabled unless the image is zoomed in sufficiently. Click the **color palette** to adjust the display colors.

Click the 3-bars menu in top left to open the **Image Viewer menu**, which includes tabs to further explore flow cell information, FOVs, image overlays, and render settings. Please refer to the [CosMx SMI Instrument User Manual \(MAN-10161\)](#) for more details.

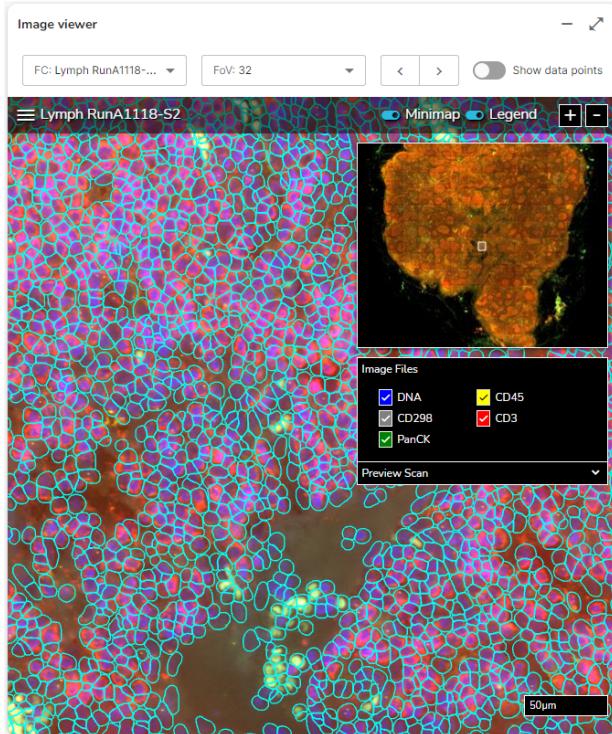


Figure 15: Image Viewer panel

FOV Editor Panel

View available flow cells and FOVs and search for them by name ([see Figure 16](#)).

Click the  icon to add flow cell or FOV attributes, which appear as new columns in the table. Click on a column's 3-dots menu to rename, delete, or sort by the column.

Alternatively, **export** the displayed FOV information and attributes by clicking the download icon. **Import** attributes from a .csv

file by clicking the upload icon and selecting the file from your computer. (You can download the attributes spreadsheet first, and use it as a template to fill in additional attributes for uploading.)

Click the copy icon, then select **Copy all** or **Copy only names**, to copy information to the clipboard (either all the information in the FOV Editor table, or FOV names alone).

FoV editor		
<input type="text" value="Search..."/>		
Flow Cells	FoVs	⋮
1 Tonsil Run_eedc8260	1	⋮
2 Tonsil Run_eedc8260	2	⋮
3 Tonsil Run_eedc8260	3	⋮
4 Tonsil Run_eedc8260	4	⋮

Figure 16: FOV Editor panel

Data Viewer Panel

Displays one or two data visualizations in separate panes ([see Figure 17](#)). Click the icons in the panel header to toggle between a single or paired visualization display. For each visualization, select the pipeline step to display from the dropdown menu. Customize the display by selecting from the available dropdown menus. Click the **arrow** (carat) to open additional customizable settings.

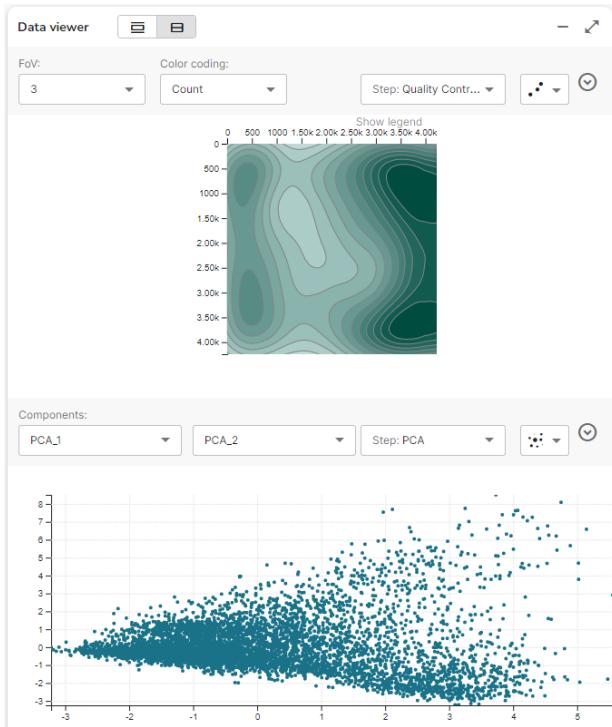


Figure 17: Data Viewer panel

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

Manage Annotations

Manage Annotations

From the Flow Cells panel ([see Figure 9](#)), click the **pencil** icon to open Manage Annotations.

The CosMx SMI Image Viewer is displayed on the left and the **FOV Editor** on the right ([see Figure 18](#)).

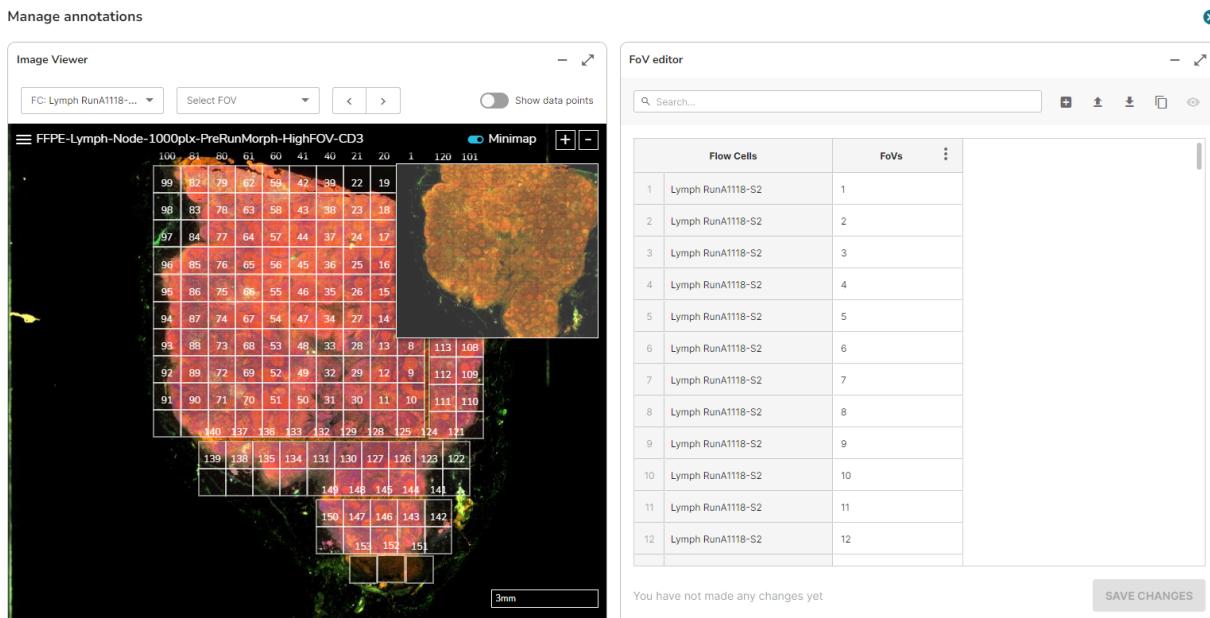


Figure 18: Manage annotations window

In the **Image Viewer** pane, select available Flow Cells from the **FC:** dropdown menu. FOVs are bordered by white boxes. To zoom in on a particular FOV, select it from the **Select FOV** dropdown menu. Use the **arrows** (carats) to switch the view between FOV. To return to a view of all FOV, click the **FOV:** dropdown menu and un-check the box for any individual FOV. You can also use the minimap and + - controls (or mouse scroll wheel) to adjust the view in the Image Viewer.

Toggle **Show data points** to **on** to open additional controls ([see Figure 19](#)), enabling you to select the **Step**, **Color pattern** (by cell type or expression level), **Gene**, or **Cell type**. Adjust the points opacity by clicking the color palette icon.

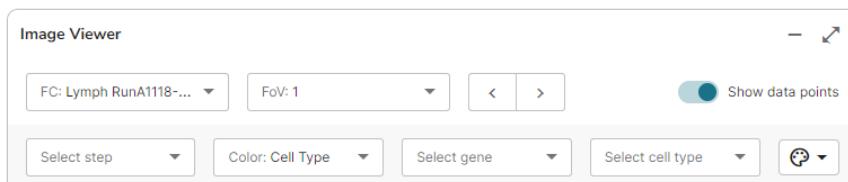


Figure 19: Manage Annotations Image Viewer with Show data points "on"

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

Manage Annotations

In the **FOV Editor** pane, view available flow cells and FOVs and search for them by name. Click the **+** icon to add flow cell or FOV attributes, which appear as additional columns in the table ([see Figure 20](#)). Click on a column's 3-dots menu to rename, delete, or sort by the column.

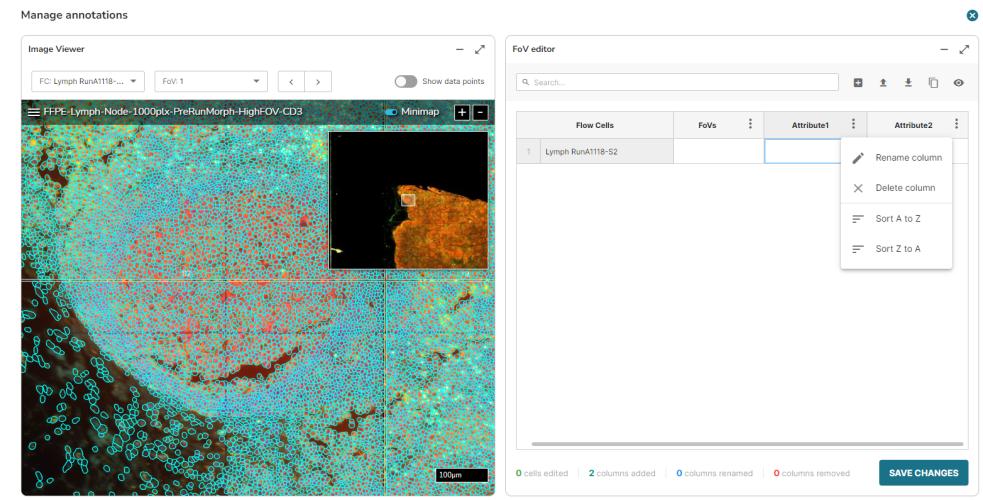


Figure 20: FOV Editor attributes columns and menu

Alternatively, **export** the displayed FOV information and attributes by clicking the download icon. **Import** attributes from a .csv file by clicking the upload icon and selecting the file from your computer. (You can download the attributes spreadsheet first, and use it as a template to fill in additional attributes for uploading.)

Click the copy icon, then select **Copy all** or **Copy only names**, to copy information to the clipboard (either all the information in the FOV Editor table, or FOV names alone).

Click **Save Changes** when done managing annotations to return to the Data Analysis suite. To exit without saving, click the **x** in the top right.

Annotations are applied only within the present study. In a future software release, flowcell annotations will be carried over to new studies.

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

Create and Run a Pipeline

Create and Run a Pipeline

From the Pipeline Run panel ([see Figure 9](#)), click the Pipeline icon  to open the Run Pipeline window ([see Figure 21](#)).

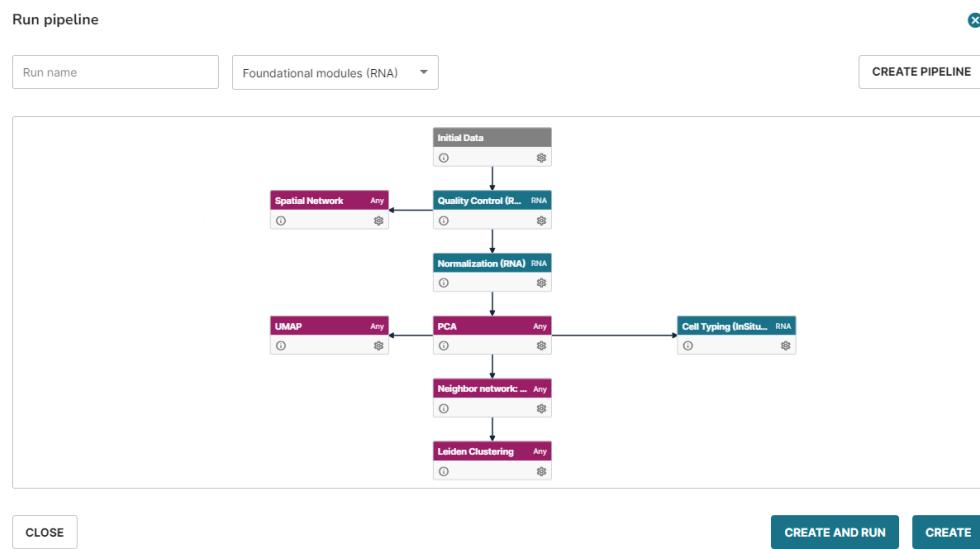


Figure 21: Run Pipeline window, displaying the NanoString foundational modules for RNA data.

Provide a **run name** and **select a defined pipeline or create a new pipeline**:

Select/Edit a Defined Pipeline

The dropdown menu displays previously built and saved data analysis pipelines. Select one, then optionally click the pencil icon to edit it before running.

Also available in the dropdown menu are the NanoString **foundational modules for RNA and Protein**. These pipelines consist of the modules **Quality Control (QC)**, **Normalization**, **Spatial Network**, **PCA**, **UMAP**, **Cell Typing**, **Neighbor Network: Expression Space**, and **Leiden Clustering** ([see Figure 21](#)). Read more about the modules in the [Glossary of CosMx SMI Pipeline Modules on page 27](#).

The foundational modules pipeline is a good place to start if you are new to spatial data analysis or are just beginning to explore your dataset. It may not suit all datasets and experimental design strategies. Neither the software nor this user manual intend to prescribe the "right" way to analyze your data, and analysis should ultimately be customized to the experimental design of your studies.

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

Create a New Pipeline

Click **Create Pipeline** to build a new data analysis pipeline. In the Create New Pipeline window ([see Figure 22](#)), enter a pipeline name and build your customized pipeline by dragging modules from the Toolbox into the gridded workspace.

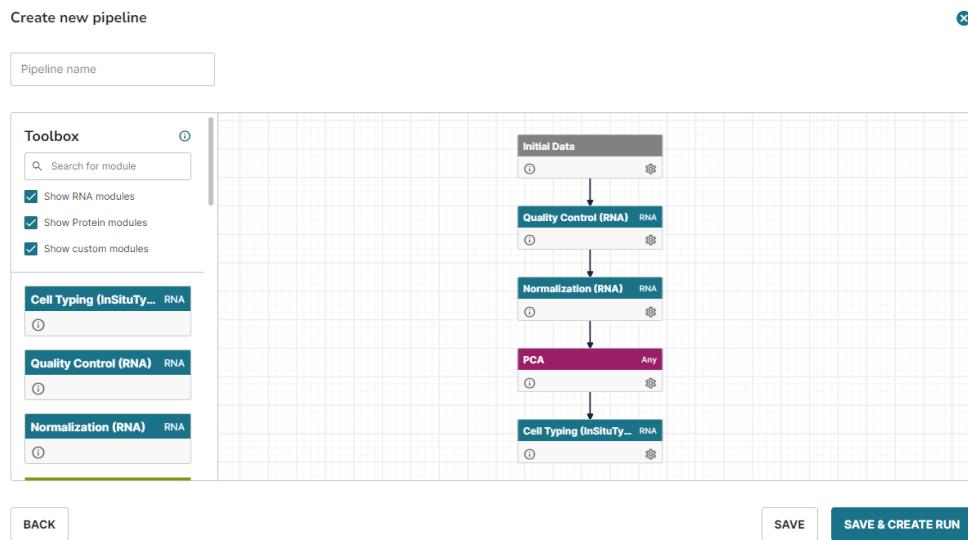


Figure 22: Create New Pipeline window

Click and drag the grid to pan around the workspace, and use the mouse scroll wheel to zoom .

Hover or click on a module in the workspace to summon arrows with which to create connectors between modules ([see Figure 23](#)): click the arrow, then the target module to draw a connecting arrow. If the module dependency prerequisites do not allow you to connect two modules, a notification appears in the top right. Review the module dependency prerequisites listed for each module in the [Glossary of CosMx SMI Pipeline Modules on page 27](#).

Hover over the **info** icon ⓘ to show information about the module, and click the **gear** icon ⚙ to show its settings.

Detailed descriptions of each module can be found in the [Glossary of CosMx SMI Pipeline Modules on page 27](#).

To remove a module or connector from the pipeline, click on it in the workspace, then click its red x.

Click **Save** to save the pipeline (enabling you to close the window or click **Back** without losing your work).

Once the pipeline is constructed, click **Save & Create Run**. Enter a name for the pipeline run. Back in the Pipeline Structure panel, click **Run All** to begin the run. Follow the progress of the pipeline run as described [on page 15](#).

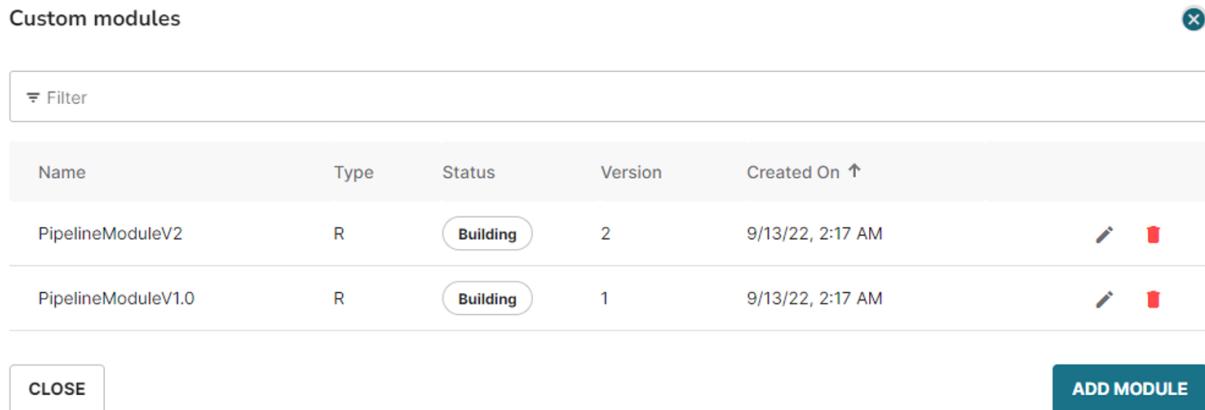


Figure 23: Arrows create links between modules

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

Custom Modules

From the Pipeline Run panel ([see Figure 9](#)), click the **custom modules** icon . The Custom Modules window opens ([see Figure 24](#)). **Edit** or **delete** existing custom modules with the pencil and trash icons, respectively.



Name	Type	Status	Version	Created On ↑
PipelineModuleV2	R	Building	2	9/13/22, 2:17 AM
PipelineModuleV1.0	R	Building	1	9/13/22, 2:17 AM

Figure 24: Custom Modules window with two custom modules loaded

To add a new custom module to the pipeline, click **Add Module**. The Add New Module window opens ([see Figure 25](#)). Enter information about the new module, including name (required), description, parameters, packages (click + to input package information), arguments (click + to input arguments), and/or entry point. Additionally, you can upload custom script files.

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

Custom Modules

Add new module

Module name *

Module Description

R Version

CPU Cores

RAM (GB)

Max run (h)

Arguments

Packages

Entry point

BACK

SAVE

Figure 25: Add new module window

Click **Save** to exit and add the defined module to the Custom Modules list, or exit without saving by clicking the **x** in the top right.

NOTE: If receiving an error such as "invalid multibyte character at line 3", check the R script for formatted quotation marks which are not accepted by the SMI Data Analysis software. Quotations marks may be " " but not “ ”.

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

Glossary of CosMx SMI Pipeline Modules

CosMx SMI data analysis pipelines are comprised of different modules to customize analysis according to the experimental design. **Modules do not necessarily need to be run in the order listed below — prerequisites are defined in each module's description.** See [Appendix: Literature References on page 57](#) for applicable literature references. *Please note that some modules are in development in v1.1.*

Analysis in the CosMx SMI Data Analysis Suite should be customized to the experimental design of your studies. Neither the software nor this user manual intend to prescribe the "right" way to analyze your data.

Initial Data

Prerequisite modules: None

Module description: CosMx SMI data is loaded into the pipeline orchestrator for downstream analysis.

Outputs and visualizations: Study statistics table

Quality Control - RNA

Prerequisite modules: Initial Data

Module description: This module flags unreliable negative probes, cells, FOVs, and target genes, as defined below. Removing those flagged negative probes, cells, FOVs, and target genes generates a filtered dataset to carry forward to downstream analysis.

- **Negative probe QC** flags negative control probes that appear to behave like outliers in the tissue. This helps prevent tissue-specific or sample-specific background effects from impacting other QC metrics that rely on the negative control probe values. Grubb's test is used, and negative probes that are designated outliers (according to p-value parameter set by the user; see below) are flagged and optionally removed.
- **Cell QC** flags cells with low or spurious signal based on the number of targets detected (more is better), fraction of probes which are negative controls (fewer is better), uniformity of count

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

distribution (higher complexity is preferred), and cell size (the top percentile may need to be removed as a QC of segmentation). See below for customizable parameters.

- **FOV QC** identifies FOVs that have generally low expression. Two approaches of FOV QC are available: the Mean method (default) flags FOV where the total count per cell averages below a threshold (default 100). The Quantile method flags FOV where a highly expressed gene (default 90th percentile) is below background (below the median of the negative probes of all cells in the FOV).
- **Target level QC** flags targets that appear to be below background across the dataset, based on probe distribution relative to negative control probes.

Customizable input parameters:

- **Negative probe QC:**
 - Grubb's test p-value (default: 0.01, range: 0-1) to flag outlier negative probes.
 - Remove QC flagged negative control probes (true/false).
- **Cell QC:**
 - Minimum number of counts per cell: recommend 20 for 1000-plex panel; 5 for 100-plex panel; must be >1. Increase the threshold to make QC more conservative.
 - Proportion of negative counts: flag cells where >10% (0.1; default) of the counts per cell are negative probes. Decrease the threshold to make QC more conservative.
 - Count distribution: flag cells where $(\text{total counts}) / (\text{number of detected genes}) \leq 1$ (default; numerical value must be ≥ 1). In other words, total counts must exceed the number of detected genes in the cell. Increase threshold to make QC more conservative.
 - Area outlier: Grubb's test p-value (default: 0.01, range 0-1) to flag outlier cells based on cell area. If megakaryocytes or other cell types known to have large cell area are in your sample, consider decreasing the p-value to make the outlier designation more stringent.
 - Remove QC flagged cells (true/false).
- **FOV QC:**
 - FOV QC method: mean (default) or quantile.
 - FOV count cutoff: only applies to Mean method. Default 100, must be ≥ 0 .
 - FOV QC quantile: only applies to Quantile method. Default 0.9 (90th percentile), range 0-1.

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

Glossary of CosMx SMI Pipeline Modules

- Quantile to negative cutoff: only applies to Quantile method. Designate the amount that the QC quantile must be greater than the negative probe median, in order to pass QC. Default 0; must be ≥ 0 .
- Remove QC flagged FOVs (true/false).
- **Target level QC:**
 - Negative control probe quantile cutoff: set the threshold at which to flag probes. A value of 0.5 (default) will flag probes with lower total counts than the median (50th percentile) of the negative control probes' counts. Range 0-1.
 - Detection over background p-value cutoff: set the p-value at which a target is deemed to be above background (0.01 default, range 0-1). A smaller p-value requires that the target is higher above background. A larger p-value allows the target to be closer to background.
 - Remove QC flagged targets (true/false).

Outputs and visualizations: study statistics table showing QC pass/fail metrics; XY plot, heatmap, box plot, violin plot, and histogram of QC-filtered data. See the [Glossary of Data Visualizations on page 44](#).

Explore your QC dataset:

- Look at QC'd data overlaid on the tissue image (include the **Image Viewer panel** in your Data Analysis Suite view, toggle **Show data points** to on, then select **Step: QC** from the dropdown menu). Select **Color: Expression** and under **Genes:** select a gene with known or expected expression in the tissue. Does it pass a 'sanity check' in terms of its expression in this region of the sample? Do visualization markers' expression match the tissue morphology?

Quality Control - Protein

Prerequisite modules: Initial Data

Module description: This module removes unreliable cells based on segmented cell area, negative probe expression, and high/low target expression. Cells with outlier Grubb's test p-values <0.05 for segmented area are flagged. Cells with mean negative probe values below the lower threshold or above the upper threshold (as defined in customizable parameters; see below) are flagged. Cells with overly high or low target expression (as defined in customizable parameters; see below) are flagged. All flagged cells are removed, generating a filtered dataset to carry forward to downstream analysis.

Customizable input parameters:

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

- Negative probe value thresholds: remove cells with negative probe mean below the lower threshold (default 2) or above the upper threshold (default 50).
- High expression proteins' percentage and expression percentile threshold: remove cells where at least 50% (0.5, default; range 0-1) of proteins are in at least the 90th percentile (0.9, default; range 0-1). To make QC more conservative, decrease percentage and/or decrease percentile.
- Low expression proteins' number and expression percentile threshold: remove cells where fewer than 3 (default; range 0-N where N is total number of proteins in panel) proteins are in at least the 50th percentile (0.5, default; range 0-1). To make QC more conservative, increase number and/or increase percentile.

Outputs and visualizations: study statistics table showing QC pass/fail metrics; XY plot, heatmap, box plot, violin plot, and histogram of QC-filtered data. See the [Glossary of Data Visualizations on page 44](#).

Explore your QC dataset:

- Look at QC'd data overlaid on the tissue image (include the **Image Viewer panel** in your Data Analysis Suite view, toggle **Show data points** to on, then select **Step: QC** from the dropdown menu). Select **Color: Expression** and under **Genes**: select a gene with known or expected expression in the tissue. Does it pass a 'sanity check' in terms of its expression in this region of the sample? Do visualization markers' expression match the tissue morphology?

Normalization - RNA

Prerequisite modules: Quality Control

Module description: Generates normalized expression data from raw counts. RNA normalization adjusts for factors to ensure that cell-specific total transcript abundance and distribution of counts (which may vary between some FOVs and between samples) does not influence downstream visualization and data analysis. Two per-cell normalization methods are available:

- Pearson's normalization (default) is based on the estimated mean and variance: $(\text{raw gene count in a cell} - \text{mean gene count in the cell}) / \text{standard deviation of gene counts in the cell}$.
- Total count normalization: $\text{raw gene count in a cell} / \text{total counts in the cell}$.

Customizable input parameters: Normalization method of choice (Pearson's residual or total count). For Pearson's residual method, select overdispersion value: default 100; must be ≥ 0 .

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

Glossary of CosMx SMI Pipeline Modules

This is the negative binomial theta parameter for Pearson residual normalization (see [Appendix: Literature References on page 57](#)).

Outputs and visualizations: study statistics, XY plot ([see Figure 26](#)), heatmap, box plot, violin plot, histogram. See the [Glossary of Data Visualizations on page 44](#).

Explore your normalized dataset:

- Evaluate the data for normalization bias: overlay normalized data on the tissue image (include the **Image Viewer panel** in your Data Analysis Suite view, toggle **Show data points** to on, then select **Step: Normalization** from the dropdown menu). Select **Color: Expression** and under **Genes**: select a housekeeping gene, expected to have even expression throughout the tissue. Do you observe expression bias across FOV?

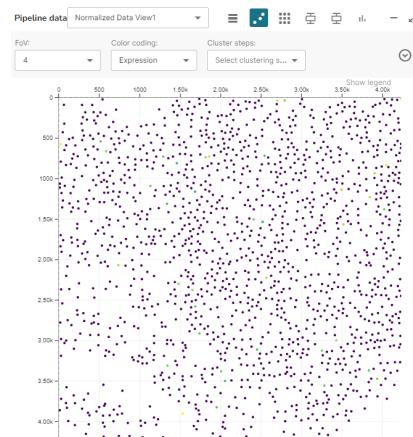


Figure 26: Normalized data in XY plot, colored by expression, helps evaluate normalization bias.

Normalized data is used as the input to generate heatmaps, violin plots, box plots, PCA, UMAP, and visualizing counts on tissue. It is **not** used as the input to differential gene expression or other modules of the SMI Data Analysis pipeline orchestrator, as those operations include a normalization function.

Normalization - Protein

Prerequisite modules: Quality Control (Protein)

Module description: Generates normalized protein data from raw mean fluorescence intensity (MFI) values. Protein normalization is based on the concepts of:

- Background subtraction, to ensure that cell-specific protein expression accounts for background observed in IgG isotype control antibodies;
- Total intensity scaling, to reduce the effect of technical artifacts such as shading or edge effects (see note below);
- arcsinh* transformation to improve visualization clarity and stabilize variance across the sample (see note below).

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

Total intensity scaling: Since protein data involves continuous intensities rather than counts, the total intensity for a cell refers to the *sum of (average intensity for each protein) in the cell*. This accounts for technical artifacts where certain parts of the image are brighter or dimmer. Total intensity scaling is essentially converting from an absolute intensity to a proportion. This proportion then gets scaled back up by the average (across cells) total intensity. This is similar to RNA normalization in which counts for a given gene in a cell are divided by total counts across all genes in a cell.

The *arcsinh* transformation is used to stabilize the variance, so that observations with higher intensity don't also have higher variance in that intensity. *Arcsinh* is a standard data transformation in modern flow cytometry comprised of linear scaling for values close to zero and logarithmic scaling for larger (negative and positive) values, with the transition between scales smoothed out. It brings protein data to a more "normal" distribution. Read more in [Finak, Perez, Weng et al \(2010\)](#) and [Folcarelli and van Staveren et al \(2021\)](#).

Customizable input parameters: Negative probe background subtraction (true/false); Total intensity normalization (true/false); Transformation (true/false).

Outputs and visualizations: study statistics, XY plot ([see Figure 26](#)), heatmap, box plot, violin plot, histogram. See the [Glossary of Data Visualizations on page 44](#).

Explore your normalized dataset:

- Evaluate the data for normalization bias: overlay normalized data on the tissue image (include the **Image Viewer panel** in your Data Analysis Suite view, toggle **Show data points** to on, then select **Step: Normalization** from the dropdown menu). Select **Color: Expression** and under **Genes**: select a housekeeping gene, expected to have even expression throughout the tissue. Do you observe expression bias across FOV?

Normalized data is used as the input to generate heatmaps, violin plots, box plots, PCA, UMAP, and visualizing counts on tissue. It is **not** used as the input to differential gene expression or other modules of the SMI Data Analysis pipeline orchestrator, as those operations include a normalization function.

Principal Component Analysis (PCA) - RNA or Protein

Prerequisite modules: Normalization

Module description: PCA provides an orthogonally constrained dimensional reduction analysis of the count data across all cells in the dataset. It produces output values (principal components, or PCs) representing axes of variation within the data, which are a combined value of weighted expression in a given cell. PCs are ordered by decreasing variation explained in the data. These

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

Glossary of CosMx SMI Pipeline Modules

can be used to better understand variation within a dataset, but are most commonly used in single-cell analysis as an input for the UMAP analysis.

Customizable input parameters: Principal components calculated (default 50; must be an integer ≥ 3).

Outputs and visualizations: PCA plot. See the [Glossary of Data Visualizations on page 44](#).

Explore your PCA dataset: While clustering may be scrutinized here, generally, the PCA dataset feeds directly into UMAP and clustering is evaluated there.

UMAP - RNA or Protein

Prerequisite modules: PCA

Module description: UMAP (Uniform Manifold Approximation and Projection for dimension reduction) provides a visualization of high-plex complex datasets in 2-dimensional space using a non-linear approach to estimate related groups of cells or features. This method is a common way of visualizing single-cell data to identify clusters of related cells which may be from the same lineage.

Customizable input parameters:

- Number of neighbors: the number of neighboring points used in local approximations of manifold structure. Default 30; range 5-50. Increase the value to preserve global structure at the loss of detailed local structure.
- Minimum distance: controls how tightly the embedding compresses points together. Default 0.3, range 0.001-0.5. Increase the value to more evenly distribute embedded points; decrease the value to allow the algorithm to optimize more accurately with regard to local structure.
- Spread: the effective scale of embedded points. In combination with minimum distance (above), this parameter determines how clustered the embedded points are. Default 1, range 0.5-10. Increase value to increase spread and reduce clustering.
- Distance metric: select cosine (default), euclidean, manhattan, or hamming, to determine the metric used to measure distance in the input space.

Read more about distance metrics in machine learning in [Ehsani and Drablos \(2020\)](#).

- Data fraction: set the fraction of PCA data to use as input. A value of 1 results in a standard UMAP, while values < 1 enable a UMAP projection with less computational burden (i.e., faster). Default 0.25, range 0.01-1.

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

Outputs and visualizations: UMAP plot (displaying data from all FOVs and flowcells in the study). See the [Glossary of Data Visualizations on page 44](#).

Explore your UMAP dataset:

- Evaluate the UMAP plot: include the **Pipeline Structure panel** and **Pipeline Data panel** in your Data Analysis Suite view, then select the **UMAP module** in the Pipeline Structure panel. Select different color schemes by clicking the arrow (carat) to expand customization options. Try coloring the UMAP plot by tissue annotation, expression of individual targets, cell type, or total cell transcript counts, to evaluate clusters.
- Compare the UMAP plot to the XY plot to see where clusters of cells defined in the UMAP exist in the tissue: include the **Data Viewer panel** in your Data Analysis Suite view, and select step: UMAP for one visualization and step: Normalization for the other ([see Figure 27](#)). Display the Normalization plot as a scatter plot and color code by cell type. Synchronize the color coding scheme to allow comparison of certain cell types in both visualizations.
- Examine patterns of co-expression between cell markers or targets with known behavior and the targets of interest to your experimental design.

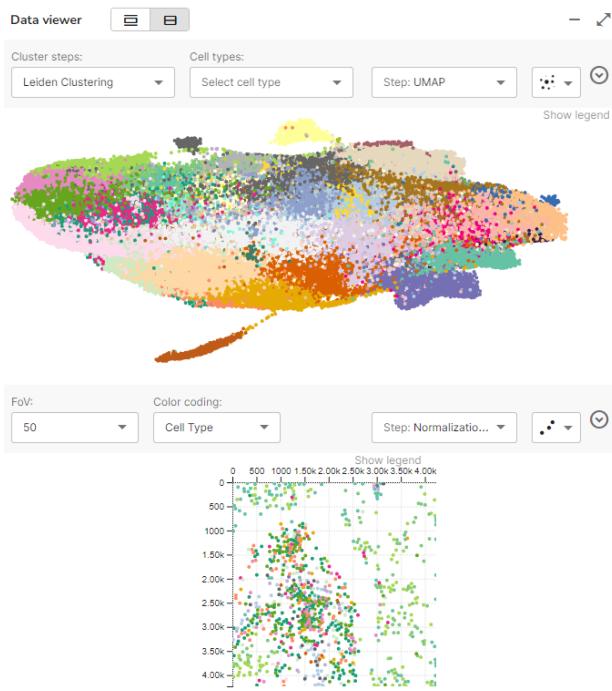


Figure 27: Use the Data Viewer panel to evaluate UMAP data (top) compared to expression data displayed in an XY plot (bottom).

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

Cell Typing - RNA

Prerequisite modules: PCA

Module description: This module uses the [InSituType algorithm](#) to identify and subset data based on cell types. Two methods are available:

- Supervised clustering: Cell type assignments are made based on a reference matrix specifying the average expression profile of each cell type. Use one of NanoString's reference matrices or provide your own (see instructions, below). A quality reference matrix will:
 - Include all the cell types present in your tissue. Granular cell types are preferred (e.g. separate profiles for "dendritic cell", "M1 macrophage", "M2 macrophage", etc), but broad cell types are accepted (e.g. a single "myeloid" profile).
 - Include most of the genes in your CosMx SMI panel.
 - Come from a robust dataset. A profile based on just 20 cells from a rare cell population will be inaccurate.
- Unsupervised clustering: Cell type clusters are determined by the software without a reference matrix input; then the user can assign cell type labels to clusters based on marker expression or other characteristic. The single argument in unsupervised clustering is the number of clusters to fit. If a range of values is specified, the algorithm will attempt to find the optimal value from within this range. Casual examination of the UMAP will suggest a reasonable range of cluster numbers; or rely on the default range of 10-20 clusters which works well in most settings. If the algorithm chooses a cluster number at the extreme of your range (e.g. 10 or 20 under the defaults), then consider expanding your cluster number range beyond that extreme and re-running the algorithm.

Customizable input parameters: Select Supervised or Unsupervised.

- If Supervised, set number of clusters to 0 and upload a reference matrix in .csv format (with genes in rows, cell types in columns, and expression values filling the matrix). Values should be untransformed linear scale, starting from 0. Scaling of columns does not matter. See gray box, below, for instructions to obtain a reference matrix.
- If Unsupervised, select number of clusters to generate (recommend 10-20).

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

To download a defined reference matrix, visit the NanoString Cell Profile Library at <https://github.com/Nanostring-Biostats/CellProfileLibrary> and download the appropriate .RData file. Create a "CellProfiles" directory in your RStudio working directory, if none exists. Run the following command in the RStudio Console — using file `ImmuneTumor_safeTME.RData` as an example:

```
>load("C:/Downloads/ImmuneTumor_safeTME.RData")
>write.csv(x = profile_matrix, file = "CellProfiles/ImmuneTumor_
safeTME.csv", row.names = TRUE, quote = FALSE)
```

Upload the resulting .csv file to the Cell Typing module in the CosMx SMI Data Analysis Suite.

Alternatively, derive your own matrix from an appropriate single-cell RNAseq (scRNA-seq) dataset. Use one of the defined reference matrices as a template file, to produce a .csv that will be recognized by the software.

Future software releases will allow upload of .RData files directly to the CosMx SMI Data Analysis Suite.

Outputs and visualizations: Visualizations such as XY plot and UMAP can be colored by the results of the Cell Typing module. To access the flightpath plot, click the image icon  on the successfully-executed Cell Typing module in the Pipeline Structure panel (*in development in v1.1*). See the [Glossary of Data Visualizations on page 44](#).

Explore your Cell Typing dataset: Once the cell type clusters are projected, evaluate each cluster's spatial distribution, expression profile, and immunofluorescence values. To do so, include the **Pipeline Structure panel**, **Pipeline Data panel**, and **Image Viewer panel** in your Data Analysis Suite view, then select the **Cell Typing module** in the Pipeline Structure panel. Scrutinize the cell typing results in the **Pipeline Data panel** and the tissue image in the **Image Viewer panel** with the following questions in mind:

- Should any cell types be merged, sub-clustered or deleted? Look at the output UMAP plot. Cell types that occupy disparate clusters on the UMAP are candidates for splitting into sub-clusters.
- Are cell types exhibiting the expected immunofluorescence results (e.g. do CD45 counts align with CD45+ immunofluorescence)? Can you identify known cell types based on tissue morphology?
- Supervised cell typing: are cell types correctly named? Do they have the expected spatial distribution (based on the XY plot)?
- Unsupervised clustering: what cell types do these clusters correspond to? Use their spatial distribution (in the XY plot) to assign cell types to the clusters.

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

Cell Typing - Protein

Prerequisite modules: Normalization

Module description: The Protein Cell Typing (CELESTA) algorithm performs cell typing by taking into account each cell's marker expression profile and, if necessary, spatial information. Cell typing calls are guided by a signature matrix that specifies the marker(s) known to have high/low expression for each cell type. A bimodal Gaussian mixture model is then fit to estimate the probability of each cell having "high expression" for each considered marker. When the probability is sufficiently high, a cell is considered an "anchor cell". When the probability is not sufficiently high to make a high-certainty cell type call, the algorithm also considers spatial information by taking into account the cell type calls of neighboring cells. These are considered "index cells". The probability thresholds can be tuned by updating the tuning parameter input file to increase (or decrease) the number of a given cell type by decreasing (or increasing) the high_expression_threshold_anchor or high_expression_threshold_index for anchor and index cells respectively.

Customizable input parameters: (*only default values available in v1.1*) Signature matrix.csv file, tuning parameter .csv file (use defaults or upload your own).

Outputs and visualizations: (*in development in v1.1*) Additional columns are added to the study metadata showing the results from each round of CELESTA cell typing (the final cell typing designations are shown in the column **final_cell_type**); XY plot colored by CELESTA cell type label; UMAP colored by CELESTA cell type label.

Explore your Cell Typing dataset: Please see the prompts above under **Cell Typing - RNA**.

Neighbor Network: Expression Space - RNA or Protein

Prerequisite modules: PCA

Module description: Constructs the KNN (k-Nearest Neighbor) graph based on the euclidean distance in PCA space, then constructs the SNN (Shared Nearest Network) graph with edge weights between any two cells based on the shared overlap in their local neighborhoods (Jaccard distance) and pruning of distant edges.

Customizable input parameters:

- Assay: select RNA or Normalized RNA data.
- Compute the shared nearest neighbor (true/false).

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

- Pruning value: the stringency of pruning the dataset. Edges with value \leq pruning threshold will be set to 0 and removed from the Shared Nearest Neighbor graph (default: 0.06, valid range: 0 (no pruning) to 1 (total pruning)).
- ANNOY (*Approximate Nearest Neighbors Oh Yeah* library) metric: euclidean (default), cosine, manhattan, or hamming.

Read more about distance metrics in machine learning in [Ehsani and Drablos \(2020\)](#).

Outputs and visualizations: None (the output dataset is used to run Leiden Clustering but is not visualized itself).

Leiden Clustering - RNA or Protein

Prerequisite modules: Neighbor network: expression space

Module description: Leiden clustering is an unsupervised clustering method that is used to identify groups of cells which are related based on how similar they are in a graph structure. Clusters are defined by moving cells to identify groups of cells that can be aggregated without changing the overall relationship of the graph and looking for unstable nodes which serve as bridges between related communities to help define the boundaries of different clusters. The resolution that you select will determine the overall number of clusters identified after running the algorithm, with lower numbers identifying fewer clusters, and higher numbers identifying more.

Customizable input parameters: Resolution (1-3 where 1 is low (fewer groups) and 3 is high (more groups); default 1), iterations (integers from 1-10, default 1).

Outputs and visualizations: Leiden cluster annotation is generated and included in study metadata. If UMAP has been run, the default visualization of the UMAP plot will be to color by Leiden clusters ([see Figure 28](#)). Leiden clusters can also be used to color an XY plot. See the [Glossary of Data Visualizations on page 44](#).

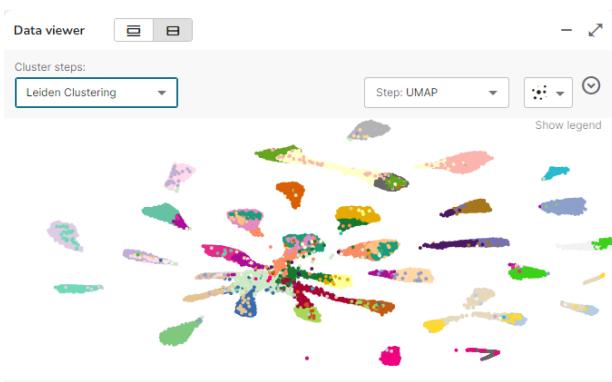


Figure 28: UMAP with color coding by Leiden clustering

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

Identify Marker Genes - RNA or Protein

Prerequisite modules: Cell Typing, Leiden Clustering, Neighborhood Analysis, or other modules that add cell-level factor annotations.

Module description: This module identifies markers associated with each cell type or cluster previously identified in the dataset. It looks for genes that are expressed above background consistently, but also most specifically restricted to each cell type or cluster within the dataset. The module acts on each gene independently.

Customizable input parameters: (*only default values available in v1.1*) ID markers for Cell Typing, Leiden Clustering, or Neighborhood Analysis (depending on which prerequisite module has been run). Minimum markers per cell type: default 5, range 1-30.

Outputs and visualizations: (*in development in v1.1*) Results matrix which consists of *genes x cell types* (one value for each cell type/gene pair; values are average estimated value of gene within all cells matching the ID); heatmap of marker genes vs. cell types scaled across cell types such that the heatmap value is the z-score of expression across all cell types for a given gene. See the [Glossary of Data Visualizations on page 44](#).

Explore your identified marker genes:

- Do well-characterized cell type markers appear to be expressed only in their canonical cell types?

Neighborhood Analysis - RNA or Protein

Prerequisite modules: Cell Typing or Leiden clustering

Module description: This module identifies distinct cellular neighborhood clusters (niches) based on cell type composition and XY coordinates. This module helps define the structural composition of a tissue by looking for regional differences in cell type composition. Niches can be repeated structures that are frequently found within a tissue but which are not contiguous (e.g. glomeruli in the kidney, germinal centers in the lymph node) or which are physically connected across a tissue (e.g. epithelial layer in the colon).

Customizable input parameters:

- Cell type annotation
- Number of neighborhoods (clusters) desired: default: 10, range: integers ≥ 3 .

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

- Select either **Radius** (μm) to capture nearest neighbors in space (default: 50 μm , range: 20-300 μm) or input a **number of neighboring cells** to indicate the number of nearest neighbors to evaluate (range: 10-500).

Outputs and visualizations: There is not a distinct visualization for this module, but other visualizations like XY plot or UMAP can be colored by the Neighborhood Analysis results. See the [Glossary of Data Visualizations on page 44](#).

Ligand-Receptor (LR) Analysis - RNA

Prerequisite modules: Leiden Clustering or Cell Typing - RNA

Module description: Scores pairs of cells and individual cells for ligand-receptor signaling. Ligand-receptor target expression in adjacent cells is used to calculate a co-expression score. A test is then performed to determine if the overall average of the scores for each ligated-receptor pair is enriched by the spatial arrangement of cells. Specific cell types can be defined for the analysis. Note that a pipeline that includes LR Analysis will pause prior to this module to allow the user to designate the Leiden Clustering or Cell Typing data as the input to LR Analysis.

Customizable input parameters: Ligand expressing cell type(s); Receptor expressing cell type (s); Receptor expressing cell type permutations (default 100; must be an integer >0); Calculation method (directional or non-directional). (*Directional* counts L₁:R₁ as distinct from R₁:L₁ (two pairs) whereas *non-directional* counts those as one pair).

Outputs and visualizations:

- Module output is a results matrix of average LR score, significance of spatial enrichment for all ligand-receptor pairs in the selected cell types.
- Heatmap, with each LR pair on the y-axis and flow cells names on the x-axis. Significant enrichment scores are colored distinctly from insignificant enrichment scores.

See the [Glossary of Data Visualizations on page 44](#).

Spatial Network - RNA or Protein

Prerequisite modules: Initial Data

Module description: Creates a network or graph structure of the physical distribution of cells. Cells are converted to nodes in the graph, and connections between cells (e.g. nearest

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

Glossary of CosMx SMI Pipeline Modules

neighbors) are represented as edges. The network can be built in one of three ways: radius-based (all cells connected within a given radius), nearest neighbors, or Delaunay triangulation.

Customizable input parameters: (*only default values available in v1.1*)

- Method of building the network (distance (default), nearest neighbors, or Delaunay)
 - If "distance" method, input radius (μm): radius to select cells to create edges (default: 20 μm / 0.02 mm, range: 10 μm -100 μm / 0.01-0.1 mm).
 - If "nearest" method, input number of nearest neighbors (cells) to evaluate (default: 5, range: 1 - 50, integers only).
- Customized graph name (as alternative to default "Spatial Distances").

Outputs and visualizations: An adjacency matrix with dimensions *number of cells x number of cells*. Each edge is recorded in the matrix as the euclidean distance between the cells. No visualizations are available, but this module's output can serve as the input to existing and custom modules.

Cell Type Co-Localization - RNA or Protein

Prerequisite modules: Cell Typing, Leiden Clustering

Module description: Examine the tendency of different cell types to be located near each other, based on module input of cell type annotations or cluster annotations and spatial coordinates for cells.

Customizable input parameters: (*only default values available in v1.1*) Cell type or cluster (select from cell types already defined in dataset), Radius (μm) within which to evaluate neighbor cell types (default 100 μm ; range 0-300 μm), Stratify results (stratify results across a metadata variable separating the tissue space across slides or FOVs).

Outputs and visualizations: (*in development in v1.1*)

- Dataframe for each cell type permutation (e.g. B-cell to fibroblast; n choose 2) with theoretical and observed K function value across range of input radii.
- Heatmap showing net difference across input radii between theoretical and observed K-function value, positive in red indicating clustering, negative in blue indicating separation.

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

Spatial Expression Analysis - RNA or Protein

Prerequisite modules: Normalization

Module description: Identify genes or proteins with spatially dependent expression patterns. This module identifies genes which have a spatial distribution that is non-uniform throughout a tissue, and which may be associated with specific tissue structures, microenvironment niches, or cell types. The module also measures associated spatial expression between genes which can be used to group genes into different spatial expression patterns. The two statistics calculated related to spatial expression patterns are Moran's I and Lee's L.

This module does not assume any specific relationship between structures in the tissue. Genes with significant count values should be visualized to determine how they are related to the tissue morphology.

Customizable input parameters: (*only default values available in v1.1*) Neighborhood size (number of cells in spatial network: default 10; must be an integer >0); Monte-Carlo permutations (default 100; must be an integer >0); Cell types of interest (optional; select cell type(s) in which to look for differential expression).

Outputs and visualizations: (*in development in v1.1*)

- Results table with Moran's I values for each gene and results from the Monte-Carlo test for significance of each I value.
- Lee's L association matrix with gene by gene measures of spatial association.
- Volcano plot of Moran's I against - log p-values for each gene, with x-axis "Spatial Autocorrelation, Moran's I" spanning "unstructured" to "patterned".
- Heatmap of Lee's L values with x- and y-axes clustered
- Spatial expression plot depicting the average expression of genes grouped by spatial association (Lee's L).
 - Alternate spatial graph: dropdown menu showing all genes which reach a specific significance level (Moran's I P-value). Change the significance threshold to limit which genes are graphed (default = 0.05).

See the [Glossary of Data Visualizations on page 44](#).

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

Pathway Analysis - RNA

Prerequisite modules: Normalization

Module description: Score cells for pathway/gene set activity. Pathway analysis is calculated on a per-cell basis using gene sets of pre-defined pathways. Relative enrichment of a pathway is calculated using the genes within each pathway. Gene sets which do not have sufficient coverage are excluded from analysis.

Customizable input parameters: (*only default values available in v1.1*)

Outputs and visualizations: (*in development in v1.1*) Matrix (*gene sets x cells*) with estimated gene set score of each cell, added to meta data of the study); UMAP colored by pathway score.

Pre-Differential Expression - RNA

Prerequisite modules: Leiden Clustering or Cell Typing - RNA

Module description: Preparatory step for differential expression (DE) analysis module to calculate neighboring cells' expression.

Customizable input parameters: (*in development in v1.1*)

Outputs and visualizations: None; provides input for Differential Expression module.

Differential Expression - RNA

Prerequisite modules: Pre-Differential Expression

Module description: Estimates and summarizes generalized linear (mixed) models for single cell expression. Control for the expression of neighboring cells on a tissue by including "neighboring cell expression" of the analyzed gene as a fixed-effect control variable in the DE model. (Controlling for expression in neighboring cells is motivated by the observation that cells in close proximity on a tissue are not independent, and comparisons of DE between groups may be affected by cell-segmentation and cell-type uncertainty.)

Customizable input parameters: (*only default values available in v1.1*)

Outputs and visualizations: (*in development in v1.1*) Data table summarizing the results; heatmap; volcano plot (log₂ fold changes (x) against -log₁₀(p-values) (y)). To access the volcano plot, click the image icon  on the successfully-executed DE module in the Pipeline Structure panel. See the [Glossary of Data Visualizations on page 44](#).

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

Glossary of Data Visualizations

Data visualizations are listed in order of their display in the user interface.

Study Statistics

Lists Number of FOV, Mean transcript per cell, Number of cells, 10th percentile transcript per cell, 90th percentile transcript per cell, and Mean Negprobe counts per cell ([see Figure 29](#)). Click the **arrow** (carat) to expand the list of FOV in the selected flow cell.

Flow cell name	Number of FoVs	Mean transcript per cell	Number of cells	10th percentile transcript per cell	90th percentile transcript per cell	Mean Negprobe counts per cell
1_Tonsil_B16_Try2	11	15.25216	34597	2	33	1.18496
FoV Name		Mean transcript per cell	Number of cells	10th percentile transcript per cell	90th percentile transcript per cell	Mean Negprobe counts per cell
10		11.82404	3944	2	24	1.1954
11		10.70387	2977	3	21	1.10169
1		16.79605	3193	3	32	1.25984
2		21.1948	3462	4	42	1.20548
3		23.54222	3743	6	47	1.1875

Figure 29: Study statistics in Pipeline Data panel

XY Plot

Displays data output of the selected module in X,Y space ([see Figure 30](#)).

Select the **FOV** for data visualization from the dropdown menu. Choose from available **color** options in the second dropdown menu (note that **Transcript** is not enabled unless the image is zoomed in sufficiently). Select **genes of interest** for visualization from the third dropdown menu. Click the **arrow** (carat) to expand the menu of visualization options, including **cell type selector**; **density**, **honeycomb**, or **scatter** view; and tools such as **area selection** and **zoom**. Optionally, enable automatic scatter view when the number of points displayed is less than 10,000. Choose from available **color palettes** from the dropdown menu.

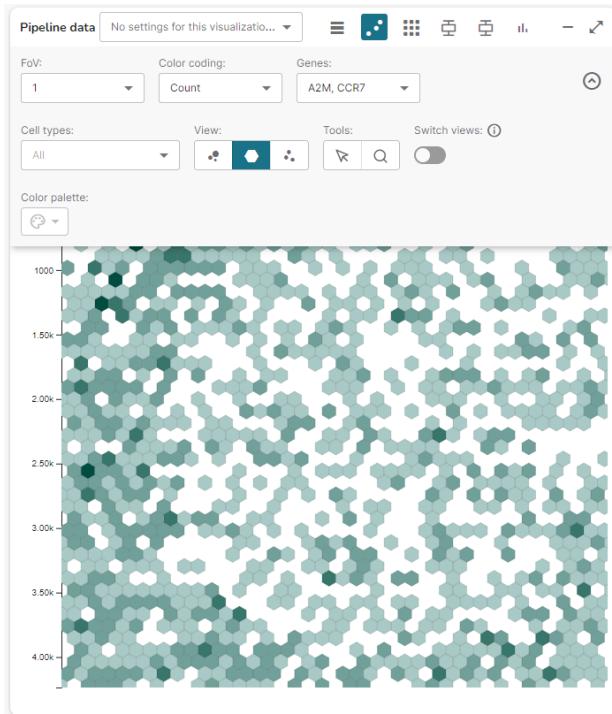


Figure 30: XY plot with honeycomb view in Pipeline Data pane

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

Glossary of Data Visualizations

Heatmap

Displays data output of the selected module as a heatmap, sorting by FOV and targets.

Select the **FOV** for data visualization from the dropdown menu. Use the buttons in the heatmap header to select **Linear** or **Log2** scaled data; display the row and/or column names; and choose the fit of the data displayed in the panel. Click the **arrow** (carat) for more options, including **Zoom**, **Save**, and adjustments to the **color palette**.

Box Plot and Violin Plot

Displays data output of the selected module as a box-and-whisker or violin plot ([see Figure 31](#)).

Select **FOV(s)** for data visualization from the dropdown menu. Use the buttons in the box plot header to select **Linear**, **Log2**, or **Log10** scaled data; and hide/display points. Click the arrow (carat) for more options, including a toggle between box, violin, or combination display; minimum expression value threshold; and custom plot title and axis names. To export box plot data, click the Save icon  in the top right of the chart.

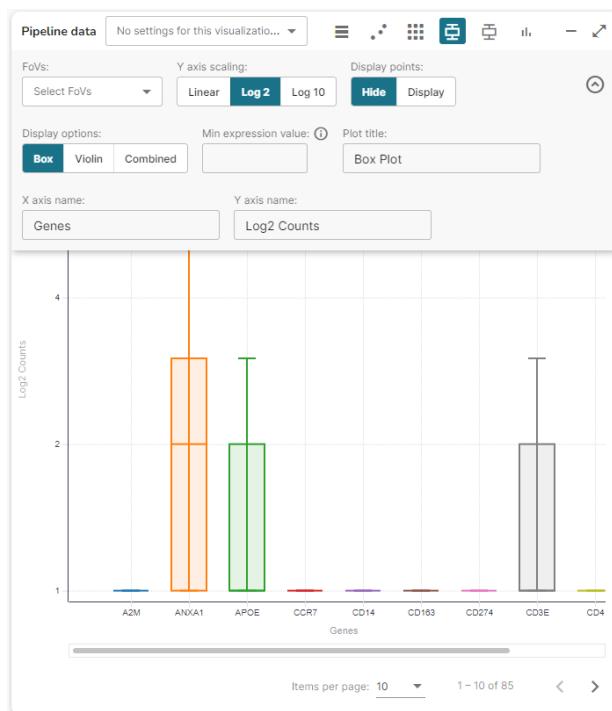


Figure 31: Pipeline Data panel - box plot

Histogram

Displays the Number of Cells (y-axis) with a particular Counts per Cell value (x-axis) ([see Figure 32](#)). Select genes of interest from the dropdown menu. If cell typing has been performed, select certain cell types under the second dropdown menu. If desired, adjust the bins number (how many categories the x-axis data is sorted into). Choose between Linear, Log 2, or Log 10 scaling. Click the arrow (carat) to rename the axes and change the bin color and opacity.

Click the Save icon  in the top right of the chart to export histogram data.

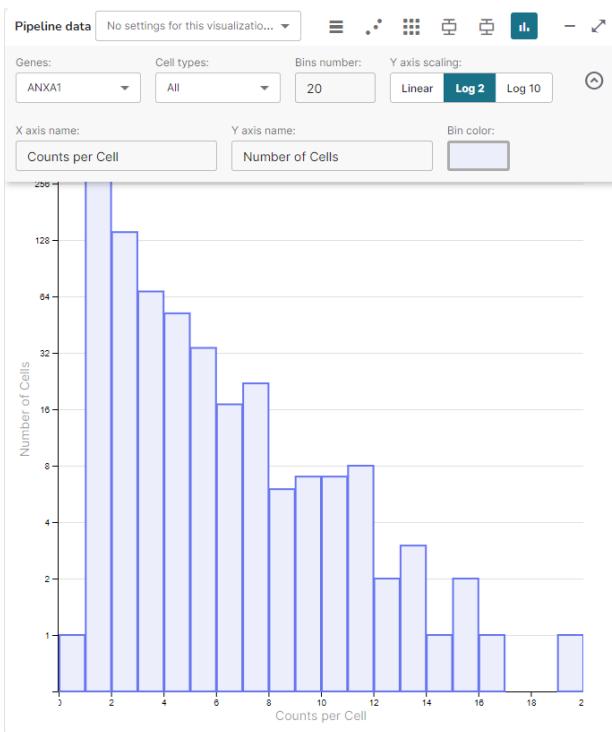


Figure 32: Pipeline Data panel - histogram

PCA Plot

Displays the 2D representation of Principal Component Analysis as a scatter plot with default axes Principal Component 1 (PCA_1) and Principal Component 2 (PCA_2) ([see Figure 33](#)). Select alternative axes from the Components dropdown menu. If cell typing has been performed, select particular cell types from the third dropdown menu.

Click the arrow (carat) to access the selection tools.

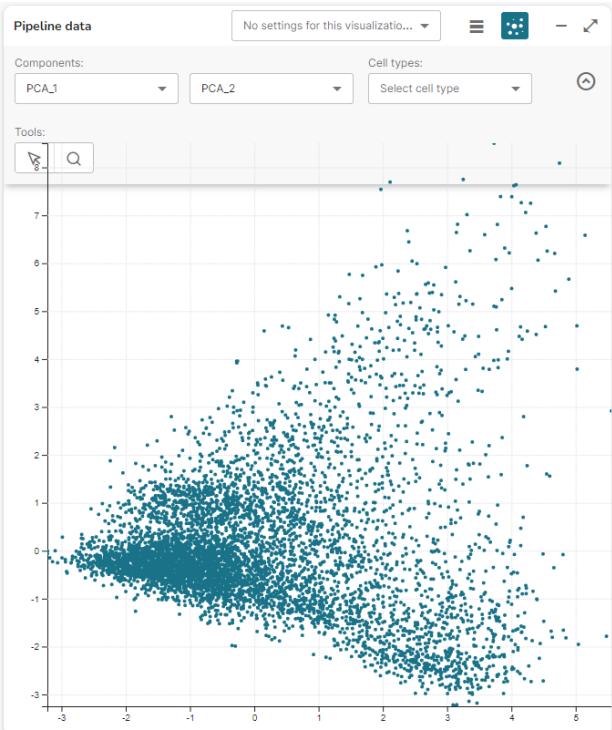


Figure 33: Pipeline Data panel - PCA plot

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

Glossary of Data Visualizations

UMAP Plot

Displays the UMAP analysis for all FOVs and flowcells in the study as a scatter plot ([see Figure 34](#)). Select **cell types/clusters** to plot from the dropdown menu. Use the pointer tool (arrow) to select a single cluster. To select multiple clusters, click on cluster names in the legend. Use the zoom tool (magnifying glass) to zoom in on a particular cluster.

Toggle **Enable selection** to **on** to allow the selection of data point(s) in the graph using a lasso, square, circle, or rectangle annotation tool.

Choose between methods of data reduction in the display: normalization or saturation (normalization normalizes data in each tile of the display; saturation sets a maximum number of dots per tile).

Click the arrow (carat) to access additional visualization settings, including Tile Count (number of tiles in the display), Tile Capacity, Max Data Points, Points size, and Points Transparency.

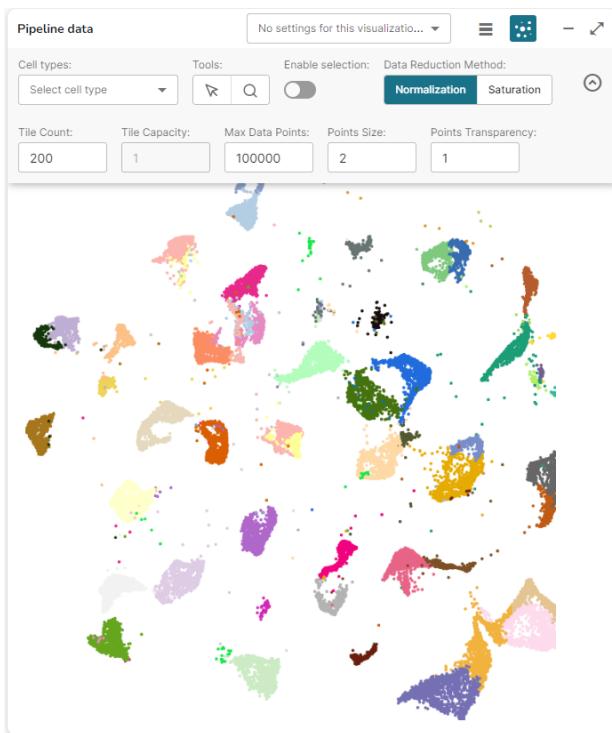


Figure 34: Pipeline Data panel - UMAP

Volcano Plot

Displays the results of differential expression analysis by plotting log2 fold changes on the x-axis against -log10(p-values) on the y-axis.

Access the volcano plot visualization as a .PNG file after successful execution of the Differential Expression module. Click on the image icon on the module in the Pipeline Structure panel to download the file.

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

Flightpath Plot

Displays cells in groups as a function of their probability of being a particular cell type. In this plot, each cell type is given a centroid, and placed near other cell types with similar profiles. Then, each individual cell is placed based on its probability of belonging to each centroid. For example, cells with 100% confidence are placed directly atop their centroid, and a cell with 50% confidence in two cell types will be placed directly between their centroids. Flightpath plots illustrate the tendency of different cell types to be confused with each other.

Access the flightpath plot as a .PNG file after successful execution of the Cell Typing - RNA module. Click on the image icon  on the module in the Pipeline Structure panel to download the file.

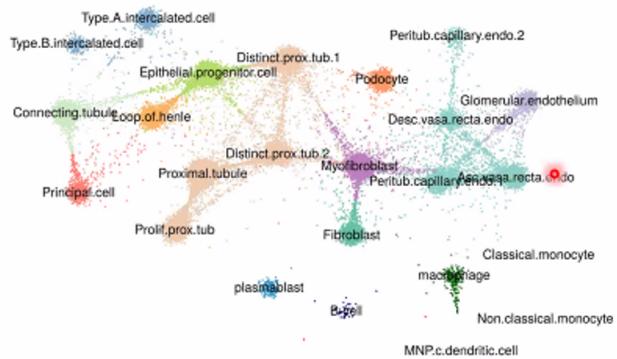


Figure 35: Example of a flightpath plot. Dots between centroids represent cells with ambiguous identity.

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

Save a Visualization

If you modify a visualization's default settings, then navigate away from the visualization, the software will prompt you to save the visualization settings ([see Figure 36](#)).

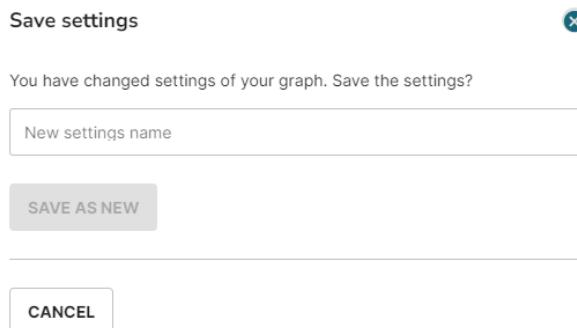


Figure 36: Save visualization settings prompt

Once saved, the visualization is available from the dropdown menu at the top of the Pipeline Data panel.

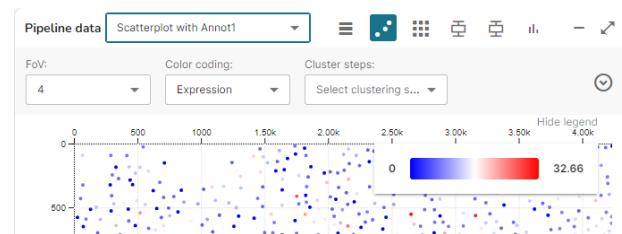


Figure 37: Saved visualization settings available in dropdown menu

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

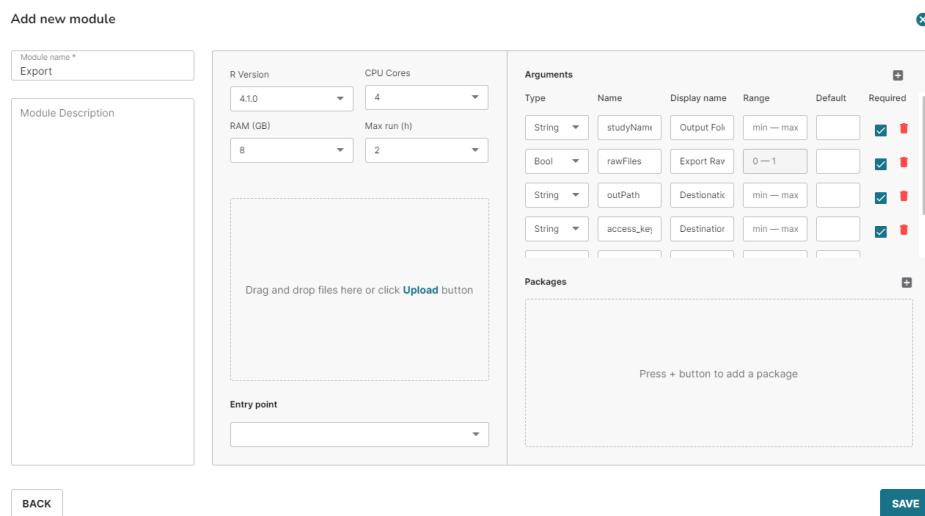
Export Data

The CosMx SMI data export module copies data from the AtoMx SIP cloud platform to a user-specified AWS S3 bucket. From there, users can further move or manage their data. NanoString is developing **vignettes** (instructive tutorials) to walk through CosMx SMI export operations. Please consult with your Product Applications Scientist for additional resources.

Prior to exporting CosMx SMI data, set up an AWS S3 bucket following the instructions at [Getting Started with Amazon S3](#) or speak with your Informatics or IT teams for support through your institution. AWS Free S3 5 GB plan will suffice for the transfer and storage of most studies (excluding raw data files). NanoString is gathering data to inform on the best plan options for different use cases.

To export CosMx SMI data,

1. Open the study of interest.
2. If you have already used the Export module to export data, then the script is already loaded in the software; skip to step 3. To load the Export module for the first time,
 - a. Click the **custom modules** icon  in the Pipeline Run panel to open Custom Modules.
 - b. Click **Add Module** to open the Add new module window.
 - c. Name the module **Export**. Upload the **Export script** obtained from NanoString.
 - d. Select the maximum available **RAM (GB)** and **Max run (h)** from the dropdown menus ([see Figure 38](#)).



The screenshot shows the 'Add new module' dialog box. In the 'Module name' field, 'Export' is entered. The 'Module Description' field is empty. Under 'R Version', '4.1.0' is selected. 'CPU Cores' is set to 4. 'RAM (GB)' is set to 8, and 'Max run (h)' is set to 2. The 'Arguments' section contains four entries: 'studyName' (Type: String, Default: checked, Required: checked), 'rawFiles' (Type: Bool, Default: checked, Required: checked), 'outPath' (Type: String, Default: checked, Required: checked), and 'access_key' (Type: String, Default: checked, Required: checked). The 'Packages' section is empty. At the bottom are 'BACK' and 'SAVE' buttons.

Figure 38: Loading the Export module into Custom Modules; maximize RAM and Max run time

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

Export Data

- e. In the section **Arguments**, click the **+** button to add the following arguments. Ensure the display names are spelled exactly as follows:

- [string] studyName - Output Folder Name
- [boolean] rawFiles - Export Raw Files
- [string] outPath - Destination S3 file path
- [string] access_key - Destination AWS access key
- [string] secret_key - Destination AWS secret key
- [string] s3Region - Destination AWS region

- f. Click **Save**. The Export module should now appear in the list of Custom Modules. Close the Custom Modules window.
3. Click the Pipeline icon  in the Pipeline Run panel to open the Run Pipeline window.
4. To incorporate the Export module to an existing pipeline, select the pipeline of choice from the Pipeline dropdown menu, then click the pencil icon to edit it. To create a new pipeline with the Export module, click **Create Pipeline** to open the Create New Pipeline window.
5. Drag the Export module from the list at left to the gridded workspace and connect it to an existing module. The Export module will export all data upstream of it in the pipeline. [Figure 39](#) shows how to export the Initial Dataset.

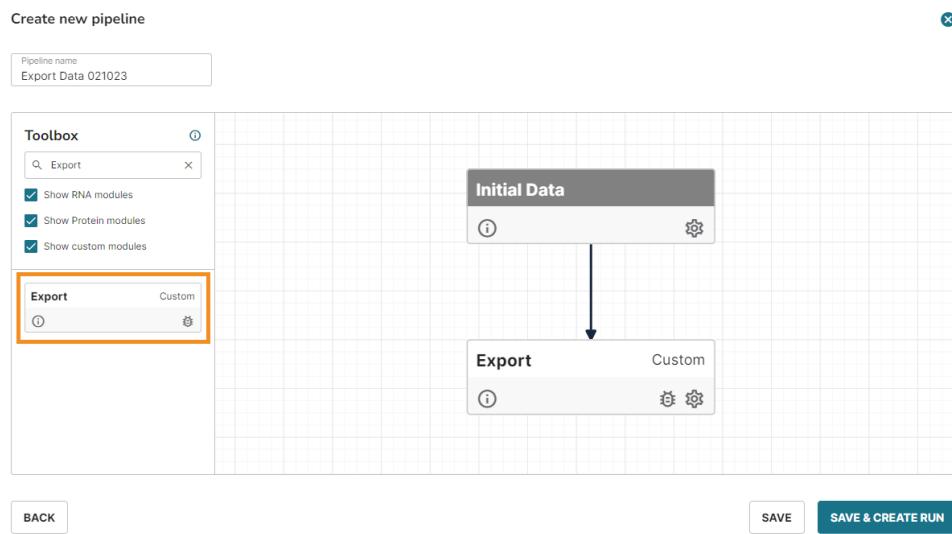


Figure 39: Building a pipeline with the Export module

6. Click the gear icon on the Export module to input the parameters for exporting:

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

Export Data

- a. Select Version 1 from the Module Version dropdown menu.
 - b. Output Folder Name: create a name for the output folder, which will appear within the destination folder of the S3 bucket.
 - c. (Optional) Check the box to include raw files along with the Seurat object and TileDB array (see [Export Output on page 53](#)). Including raw files will significantly increase the file size.
NOTE: if already checked, please uncheck, then check the box to include raw files.
 - d. Destination S3 file path: the address of the destination folder in the S3 bucket. (The output folder named in step 2b will be created within it.) From the S3 console, create a destination folder in your S3 bucket. Select the folder using the check mark on the left, then click **Copy S3 URI**. Paste this URI into the field 'Destination S3 file path' in AtoMx SIP. The format is S3://...
 - e. Destination AWS access key: if not already done, create an access key to your S3 bucket from AWS Identity and Access Management (IAM), accessible by clicking on the user icon in the top right of the AWS S3 console then selecting **Security credentials**. Follow prompts to **Create Access Key**. Once created, save the keys in a secure location for future reference.
 - f. Destination AWS secret key: provide the secret key created with the access key in the previous step.
 - g. Destination AWS region: enter the region name of your bucket as it appears in your Amazon S3 console under the tab **Properties**, in the format us-west-1.
 - h. Click **Save** or **Save & Create Run**.
7. Back in the Pipeline Structure panel, click **Run All**. For a study of 4 FOV (1000 plex panel) without raw files, export should take 5-10 min.
 8. When the Export run is completed, check the contents of the S3 bucket from the S3 console to confirm the transfer of files. Files can be downloaded using the S3 console; folders must be downloaded using command line interface (CLI). Downloading from S3 (file egress) incurs costs according to AWS pricing structure. Please refer to AWS documentation for support managing your data after its export from AtoMx SIP.

AWS provides comprehensive user documentation. See, for example:

- <https://docs.aws.amazon.com/AmazonS3/latest/userguide/creating-buckets-s3.html>
- <https://aws.amazon.com/s3/pricing/>

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

Export Data

Export Output

The output folder structure is shown here ([see Figure 40](#)).

- Seurat object
- TileDB array
- Raw files (if included with export)

Name
3da5b442-ac8a-483b-a818-4b9e995acba9_seuratObject.RDS
3da5b442-ac8a-483b-a818-4b9e995acba9_TileDB/
SO/

Figure 40: Example export output folder structure

The following computational packages are required to interact with the exported data.

- Seurat: R Toolkit for Single Cell Genomics. Install in RStudio using `install.packages ("Seurat")`
- tiledbsc: an R implementation of the Stack of Matrices, Annotated (SOMA). Install in RStudio using `remotes::install_github("tiledb-inc/tiledbsc", force = TRUE)`
- tiledbr: an R interface to the storage engine of TileDB. Install in RStudio using `remotes::install_github("TileDB-Inc/TileDB-R", force = TRUE)`

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

The following are the key files for **RNA analysis** outside of the CosMx SMI Data Analysis suite:

Flowcell Folder/Logs/SpatialProfiling_[sequence_name].fovs - *cell coordinates for study, slide number matches SlideNum in next folder*

Flowcell Folder/[Flowcell]_[SlideNum]/CellStatsDir/CellComposite/CellComposite_FOV[FOV].jpg - *composite colored image for each FOV*

Flowcell Folder/[Flowcell]_[SlideNum]/CellStatsDir/CellOverlay/CellOverlay_[FOV].jpg - *B&W image with cell segmentation overlays for each FOV*

Flowcell Folder/[Flowcell]_[SlideNum]/CellStatsDir/FOV[FOV]/CellLabels_F[FOV].TIF - *cell segmentation labels; pixel intensity values correspond with the unique cell_id*

Flowcell Folder/[Flowcell]_[SlideNum]/CellStatsDir/FOV[FOV]/CompartmentLabels_F[FOV].TIF - *cell subcellular compartment labels; pixel intensity values correspond with the identified compartment label; nuclear = 1, membrane = 2, cytoplasmic = 3.*

Flowcell Folder/[Flowcell]_[SlideNum]/CellStatsDir/FOV[FOV]/[run_name]_[sequence_name]_S[slot]_Cell_Stats_F[FOV].csv - *cell info from image like area, centerX, Y coordinates and fluorescence values*

Flowcell Folder/[Flowcell]_[SlideNum]/CellStatsDir/Morphology2D/[sequence_name]_S[slot]_C[cycle]_P[pool]_N[spot]_F[FOV].TIF - *stacked OME-TIFF of fluorescence values*

Flowcell Folder/[Flowcell]_[SlideNum]/CellStatsDir/Morphology3D/FOV[FOV][sequence_name]_S[slot]_C[cycle]_P[pool]_N[spot]_F[FOV]_Z[z].TIF - *stacked OME-TIFF of fluorescence values at each Z plane*

Flowcell Folder/[Flowcell]_[SlideNum]/CellStatsDir/RnD/Run_[GUID]_[sequence_name]_S[slot]_Summary_F[FOV].csv - *FOV summary statistics*

Flowcell Folder/[Flowcell]_[SlideNum]/RunSummary/Run_[GUID]_[date]_S[slot]_[instrument_name]_ExptConfig.txt - *instrument config - includes pixel to nm ratio*

Flowcell Folder/[Flowcell]_[SlideNum]/AnalysisResults/[processing_id]/FOV[FOV]/FOV[FOV]_Analysis_Summary.txt - *limits of detection*

Flowcell Folder/[Flowcell]_[SlideNum]/AnalysisResults/[processing_id]/FOV[FOV]/Run_[GUID]_FOV[FOV]_complete_code_cell_target_call_coord.csv - *target coordinates and counts per cell; other coordinate files are intermediate files*

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

Export Data

The following are the key files for **protein analysis** outside of the CosMx SMI Data Analysis Suite:

Flowcell Folder/Logs/SpatialProfiling_[sequence_name].fovs - *cell coordinates for study, slide number matches SlideNum in next folder*

Flowcell Folder/[Flowcell]_[SlideNum]/plex_[processing_id].csv - *target probes metadata*

Flowcell Folder/[Flowcell]_[SlideNum]/CellStatsDir/CellComposite/CellComposite_FOV[FOV].jpg - *composite colored image for each FOV*

Flowcell Folder/[Flowcell]_[SlideNum]/CellStatsDir/CellOverlay/CellOverlay_[FOV].jpg - *B&W image with cell segmentation overlays for each FOV*

Flowcell Folder/[Flowcell]_[SlideNum]/CellStatsDir/FOV[FOV]/CellLabels_F[FOV].TIF - *cell segmentation labels; pixel intensity values correspond with the unique cell_id*

Flowcell Folder/[Flowcell]_[SlideNum]/CellStatsDir/FOV[FOV]/CompartmentLabels_F[FOV].TIF - *cell subcellular compartment labels; pixel intensity values correspond with the identified compartment label; nuclear = 1, membrane = 2, cytoplasmic = 3*

Flowcell Folder/[Flowcell]_[SlideNum]/CellStatsDir/FOV[FOV]/[run_name]_[sequence_name]_S[slot]_Cell_Stats_F[FOV].csv - *cell info from image like area, centerX, Y coordinates and fluorescence values*

Flowcell Folder/[Flowcell]_[SlideNum]/CellStatsDir/Morphology2D/[sequence_name]_S[slot]_C[cycle]_P[pool]_N[spot]_F[FOV].TIF - *stacked OME-TIFF of fluorescence values*

Flowcell Folder/[Flowcell]_[SlideNum]/CellStatsDir/Morphology3D/FOV[FOV]/[sequence_name]_S[slot]_C[cycle]_P[pool]_N[spot]_F[FOV]_Z[z].TIF - *stacked OME-TIFF of fluorescence values at each Z plane*

Flowcell Folder/[Flowcell]_[SlideNum]/CellStatsDir/RnD/Run_[GUID]_[sequence_name]_S[slot]_Summary_F[FOV].csv - *FOV summary statistics*

Flowcell Folder/[Flowcell]_[SlideNum]/RunSummary/Run_[GUID]_[date]_S[slot]_[instrument_name]_ExptConfig.txt - *instrument config - includes the pixel to nm ratio*

Flowcell Folder/[Flowcell]_[SlideNum]/ProteinDir/FOV[FOV]/ - *files are used to generate ProteinImages, ProteinMasks, & Protein Stats in AnalysisResults/ folder*

Flowcell Folder/[Flowcell]_[SlideNum]/AnalysisResults/[processing_id]/FOV[FOV]/PerCellStats/[sequence_name]_[slot]C001_F[FOV][probe_id]_perCell_1ChStats.csv - *counts per probe_id per cell; 1 file per protein in panel; sum fluorescence is considered count*

Flowcell Folder/[Flowcell]_[SlideNum]/AnalysisResults/[processing_id]/FOV[FOV]/ProteinImages/[sequence_name]_[slot]C001_F[FOV][probe_id].TIF - *OME-TIFF with decoded fluorescence intensity of protein target across FOV*

Flowcell Folder/[Flowcell]_[SlideNum]/AnalysisResults/[processing_id]/FOV[FOV]/ProteinMasks/[sequence_name]_[slot]C001_F[FOV][probe_id]_Mask.TIF - *OME-TIFF displaying area of target; 255 = present, 0 = absent*

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

Troubleshooting

If receiving the message "Something went wrong" when accessing a data analysis study, log out of AtoMx SIP and log back in.

It may be necessary to troubleshoot the failed creation of a study or the failed execution of a pipeline module:

To download logs for the creation of the study, click on **Details and logs** in the Study Details panel ([see Figure 41](#)).

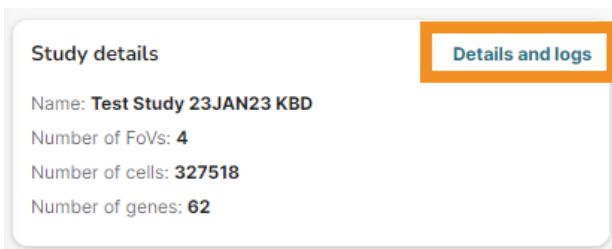


Figure 41: Details and logs link in the Study Details panel.

To download the log files for an individual module (that ran successfully or failed), click on the metrics icon  on the module block in the Pipeline Structure panel, then click the download icon ([see Figure 42](#)).

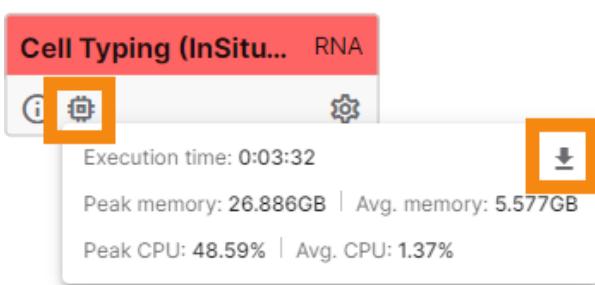


Figure 42: The metrics icon opens module metrics and the option to download log files

Log files are downloaded as zipped files to your local Downloads folder. Once unzipped, they can be opened in a text editor such as Notepad.

Contact Support@nanostring.com for assistance with troubleshooting in the CosMx SMI Data Analysis Suite.

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

Appendix: Literature References

Appendix: Literature References

The following references provide additional information on the modules of the CosMx SMI Data Analysis Suite v1.1.

Quality Control	https://www.itl.nist.gov/div898/handbook/eda/section3/eda35h1.htm
Normalization	https://scanpy-tutorials.readthedocs.io/en/latest/tutorial_pearson_residuals.html
	https://genomebiology.biomedcentral.com/articles/10.1186/s13059-021-02451-7
UMAP	https://pubmed.ncbi.nlm.nih.gov/30531897/
Cell Typing - RNA	https://www.biorxiv.org/content/10.1101/2022.10.19.512902v1.full
Cell Typing - Protein	https://doi.org/10.1038/s41592-022-01498-z
Neighborhood Analysis	https://pubmed.ncbi.nlm.nih.gov/32763154/
	https://pubmed.ncbi.nlm.nih.gov/27818791/
Leiden Clustering	https://www.nature.com/articles/s41598-019-41695-z
Spatial Expression Analysis	https://link.springer.com/article/10.1007/s101090100064
Differential Expression	https://github.com/glmmTMB/glmmTMB
	https://github.com/rvlenth/emmeans
Cell Type Co-Localization	https://www.jstor.org/stable/2984796
	https://book.spatstat.org/
Signaling Pathways	https://doi.org/10.1038/nmeth.4463.02200317
	https://github.com/aertslab/AUCell
	https://bioconductor.org/packages/release/bioc/html/AUCell.html

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.



NanoString Technologies, Inc.
530 Fairview Ave North
Seattle, Washington 98109 USA
www.nanostring.com

CONTACT US
info@nanostring.com
Tel: +1 888 358 6266
Fax: +1 206 378 6288

SALES CONTACTS
United States: us.sales@nanostring.com
EMEA: europe.sales@nanostring.com
Asia Pacific & Japan: apac.sales@nanostring.com
Other regions: info@nanostring.com

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.