2083 Cloud Computing Meets Spatial Biology: AtoMx Spatial Informatics Platform, a Powerful New Approach to Data Analysis

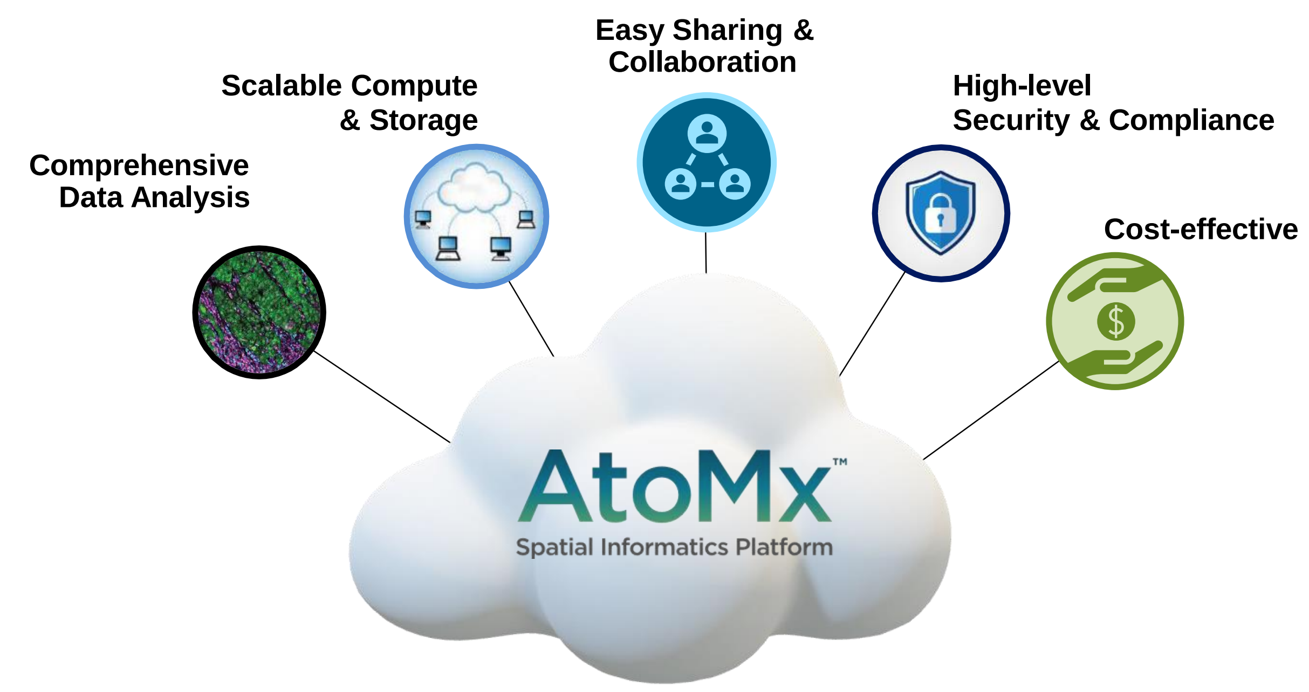
Jason W Reeves, David Henderson, Matt Crow, Emily Killingbeck, Alyssa Rosenbloom, Maddy Griswold, Nicole Ortogero, April Munn, Kenlyn Terai, Joachim Schmid
Nanostring Technologies, Inc. Seattle WA

nanoString
NanoString Technologies
530 Fairview Avenue North, Seattle, WA
98109

Abstract

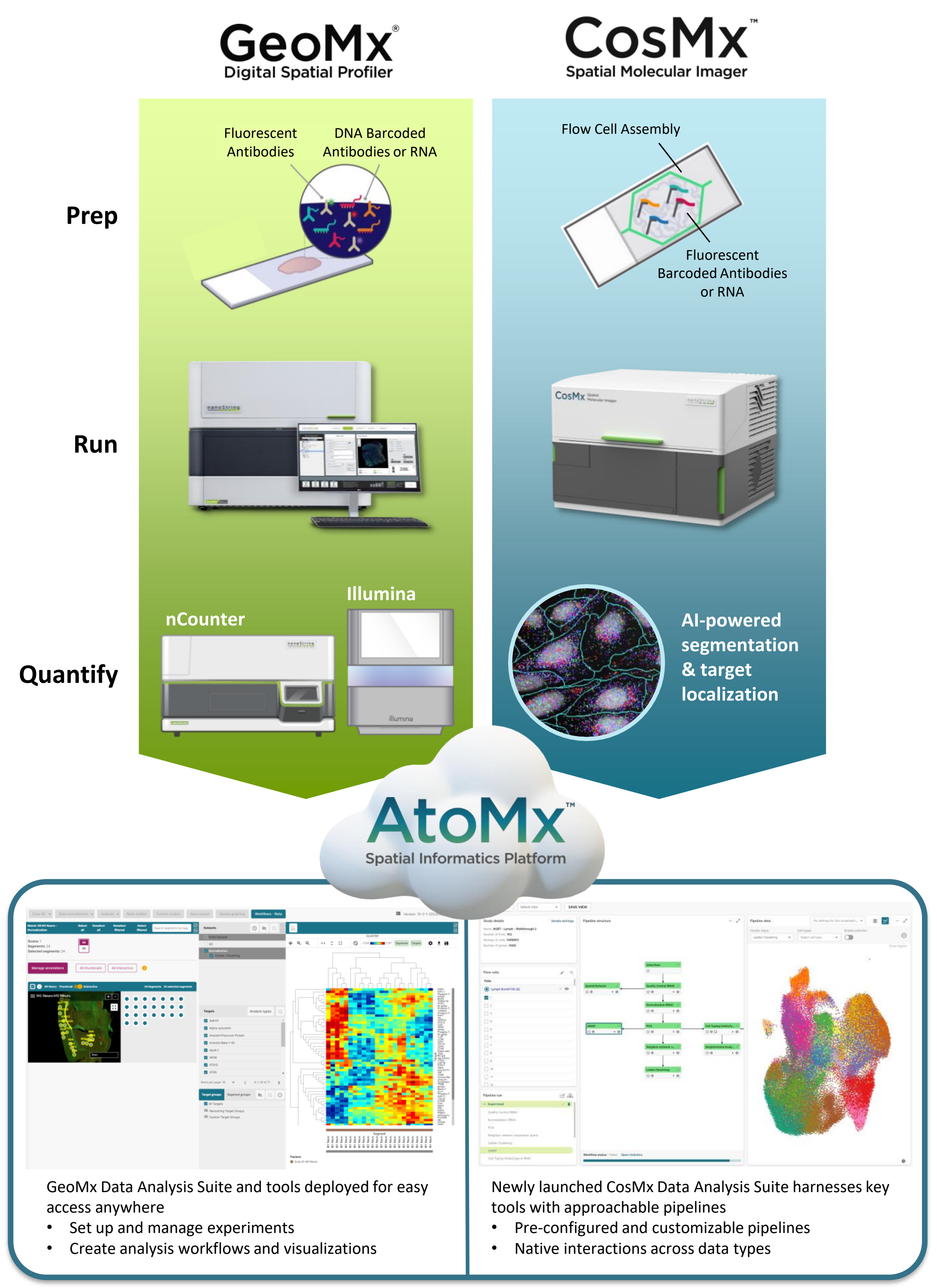
Spatial biology is revolutionizing the way in which we extract biological insights from precious clinical samples, allowing profiling of thousands of RNA species and ever higher numbers of antibodies to be analyzed from a single tissue slide. However, with new 'omics capabilities, scaling the analytical and infrastructure requirements for these high-content datasets is more critical than ever, as these datasets contain billions of data points and hundreds of images that require more extensive storage and analysis capabilities. To enable scientists to leverage spatial 'omics instruments, we have launched the AtoMx™ Spatial Informatics Platform, which will enable analytics across NanoString's spatial profiling and imaging solutions. The AtoMx platform was designed to specifically host, ingest, and manage data generated from both the GeoMx® Digital Spatial Profiler and the CosMx™ Spatial Molecular Imager. Data generated by either of these platforms is analyzed, stored, and managed in cloud or private HPC environments. Within the AtoMx platform, we provide native, comprehensive analysis suites specifically designed to integrate the high content, multiplexed imaging from either platform with the molecular readouts. Machine learning methods in key analysis modules allow AtoMx to scale to the billions of data points collected while still maintaining lower compute times. The AtoMx platform is built to conform to open data standards and will support customization through custom analysis scripts, APIs for integration to external software, and a community standard data model that facilitates the integration of tools in open-source repositories. With the CosMx platform we have been able to demonstrate a 75x decrease in data processing times compared to locally resourced servers for equivalent data analysis requests by utilizing the scaling tools a cloud service provides such as scaling by compute type for greater RAM access or multiple, parallel compute instances. The platform is designed to facilitate data sharing and collaborative *in silico* analysis of spatial data, with direct permissions management and compliance with regional privacy requirements. The AtoMx Spatial Informatics Platform provides an extensible, secure management and analysis solution, capable of processing transcriptome-wide expression or high plex protein profiling from the GeoMx Digital Spatial Profiler as well as subcellular spatial 'omics data

AtoMx™ Spatial Informatics Platform Overview



AtoMx provides a complete spatial multi-omics data analysis solution for all NanoString spatial biology platforms. Launched as a cloud-first ecosystem, AtoMx drives study management, data analysis, and collaborative suites in a secure, cost-effective environment to remove IT barriers from scientific researchers

Data Generation & Analysis Workflow



GeoMx Data Analysis Suite and tools deployed for easy access anywhere

- Set up and manage experiments
- Create analysis workflows and visualizations

Newly launched CosMx Data Analysis Suite harnesses key tools with approachable pipelines

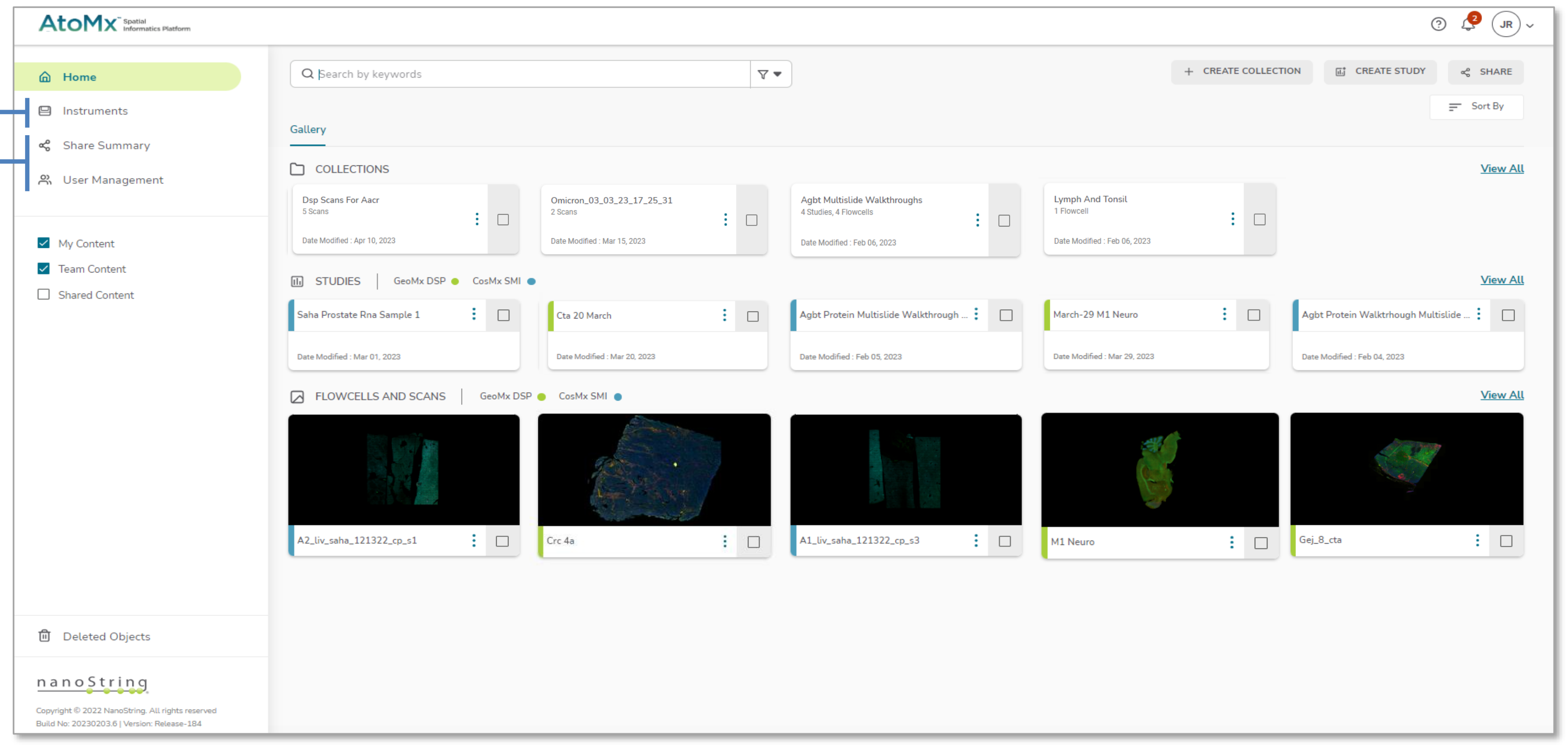
- Pre-configured and customizable pipelines
- Native interactions across data types

AtoMx™ Spatial Informatics Platform Stores and Manages NanoString Spatial Multi-omics Data

Instrument Management

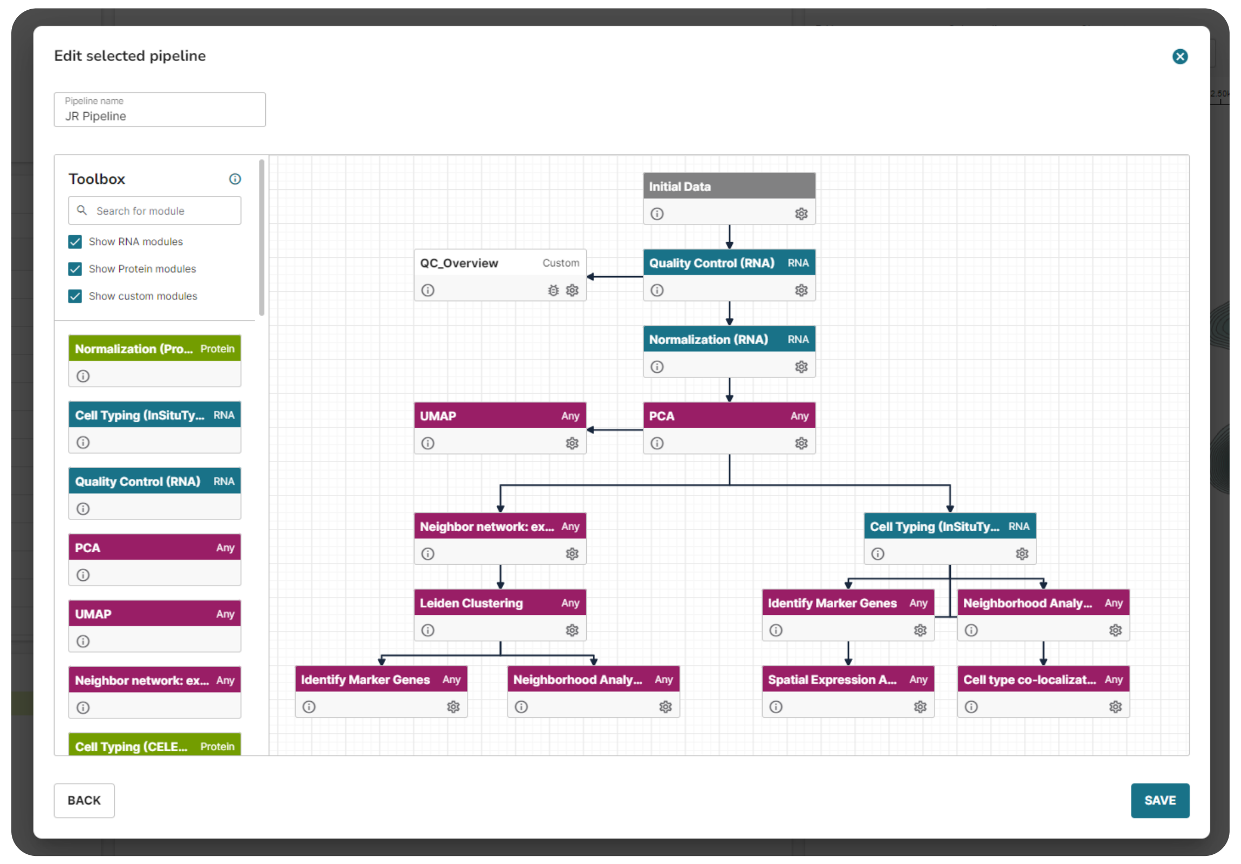
Security Controls & Collaboration Management

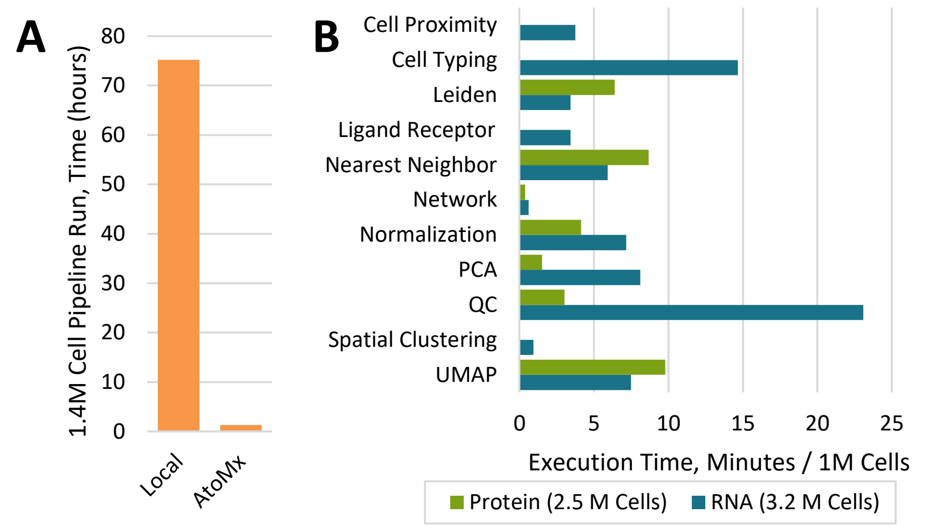
Control who has access to your GeoMx and CosMx data with instrument, security, and collaboration management tools



Flowcells & Scans	Data generated when running CosMx (flowcells) or GeoMx (Scans) projects. Contains images, metadata about reagents used, ROI placement (GeoMx) and cell segmentation (CosMx) results
Studies	Analysis of individual or multi-slide studies, with native tools for analysis of both GeoMx and CosMx platforms
Collections	Groups of studies, flowcells, or scans which can be used to track projects over time and share multiple components with collaborators in an easy-to-access bundle

Robust Library of Analysis Tools for Any Multi-omics Project within a Scalable Compute Environment





Compute Performance Optimized to Scale with Studies

A) Benchmarking of local vs AtoMx deployment of pipeline on 1.4M cell study with CosMx RNA 1000-plex Universal Cell Characterization Panel

B) Benchmarking results of analysis modules operating on either:

- 3.2M cells with 1000-plex CosMx Universal Cell Characterization RNA Panel (teal)
- 2.5M cells with 68-plex CosMx Human Immuno-Oncology Protein Panel (green)

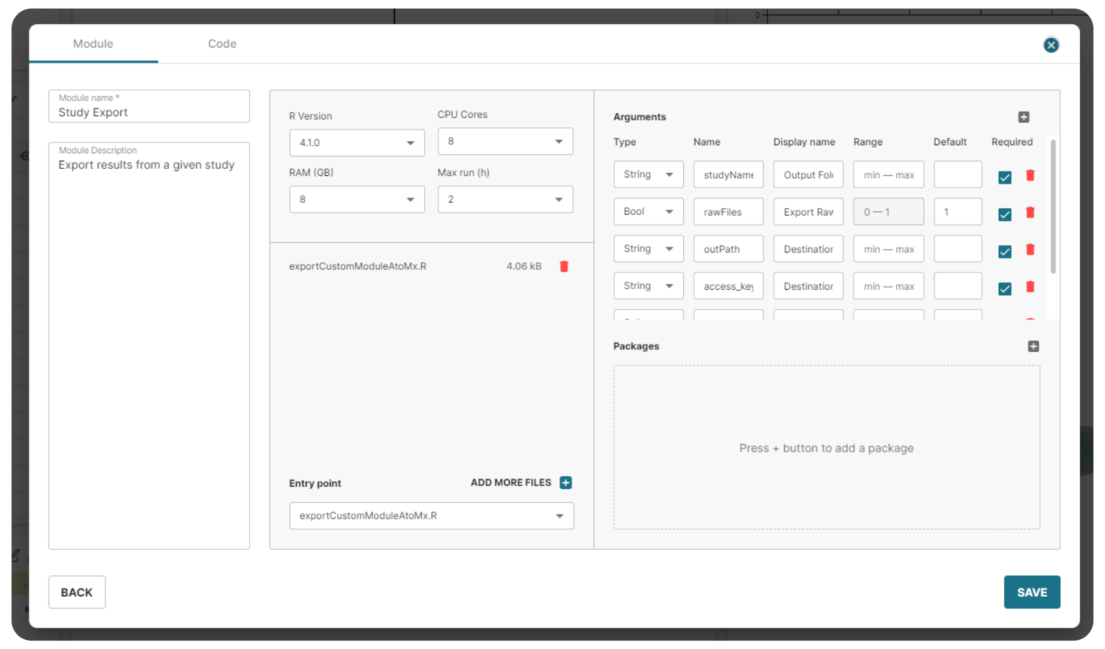
Each module timed on standard deployment AtoMx instance for performance capture.

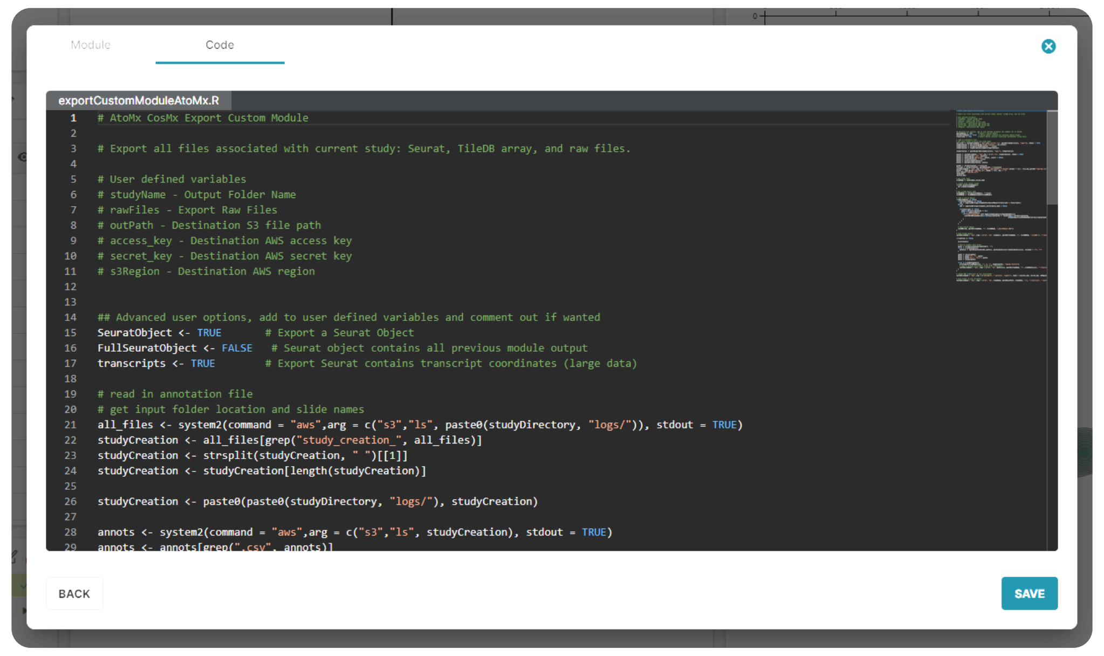
CosMx Analysis Pipeline & Analysis Modules

Analysis modules come pre-loaded with AtoMx so that users can hit the ground running with the appropriate analysis functionality to get the most out of their spatial 'omics projects. The table on the right highlights some of the key modules available to users. Pipelines also come preconfigured, though the user can easily create or modify any pipeline in the library (example shown above).

Foundational Modules	Spatial Analysis Modules	Pre-configured Pipelines
Basic modules like QC, normalization	Modules to create novel insights/ visualizations	Pipelines to address specific questions in spatial biology
• QC • Normalization • UMAP • Cell typing • PCA • Spatial clustering • Cell Segmentation	• Ligand/ Receptor • Nearest neighbor • Spatial network • Cell proximity analysis • Differential expression	• RNA Foundational Modules • RNA Comprehensive Modules • Protein Foundational Modules • Protein Comprehensive Modules

Scripting Integration for Customizing Analysis Pipelines with Open-source Methods

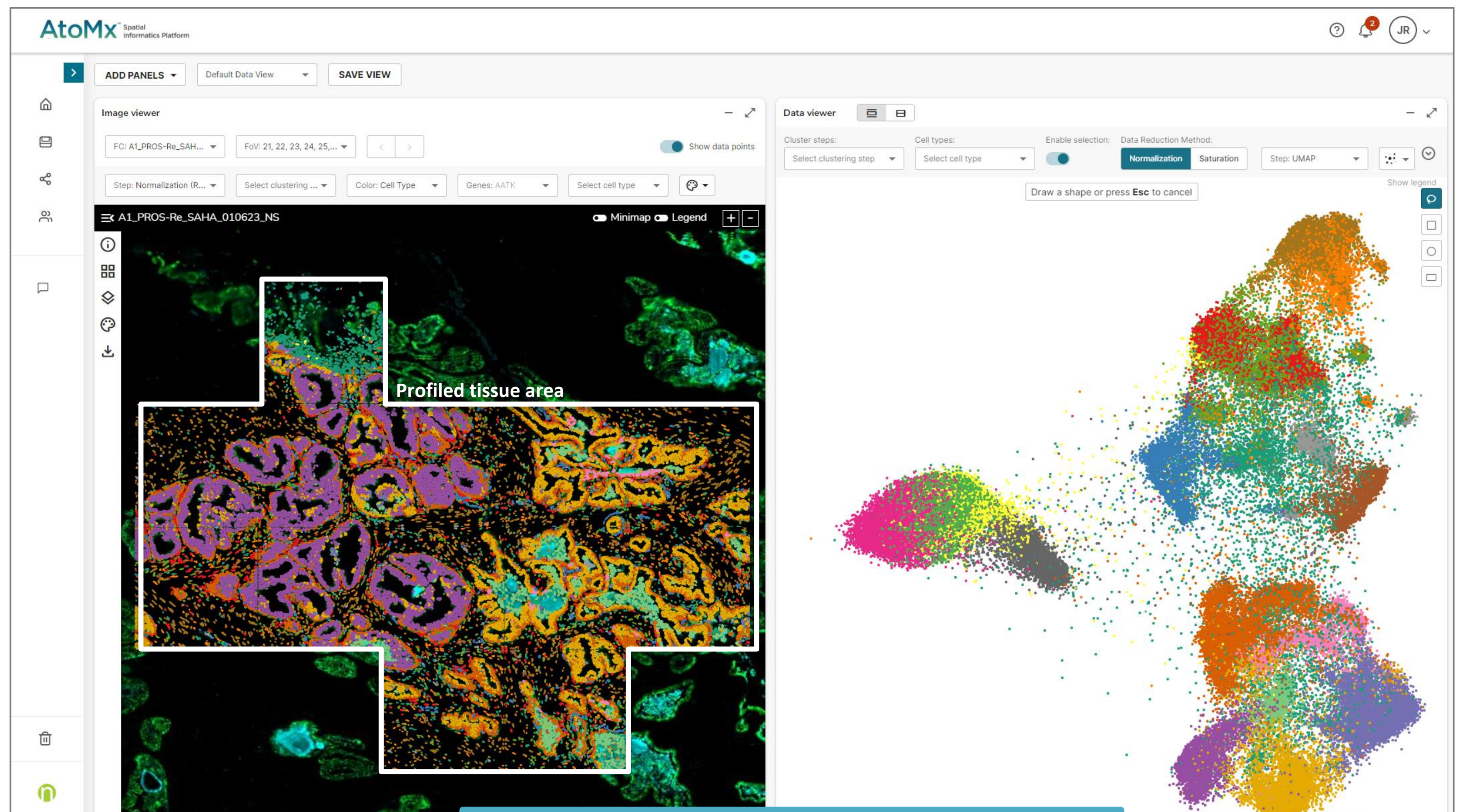


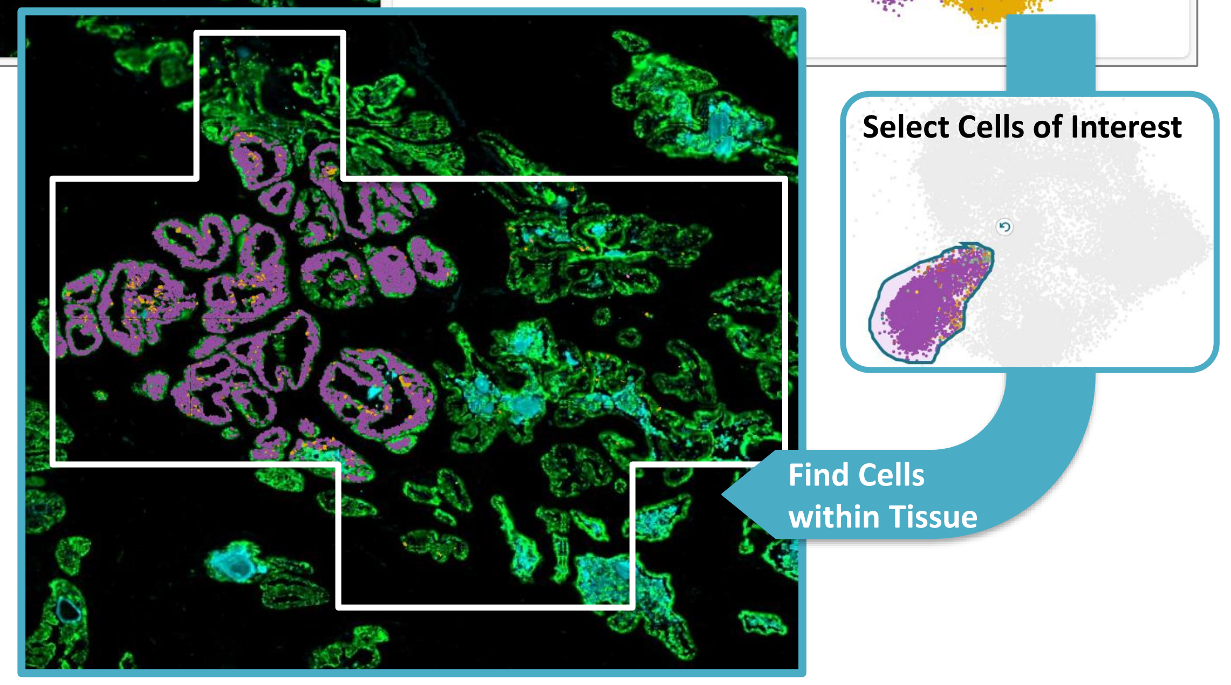


```
# AtoMx CosMx Export Custom Module

# Export all files associated with current study: Seurat, TileDB array, and raw files.

# User defined variables
# studyName - Output Folder Name
# rawFiles - Export Raw Files
# outPath - Destination S3 file path
# access_key - Destination AWS access key
# secret_key - Destination AWS secret key
# s3Region - Destination AWS region


## Advanced user options, add to user defined variables and comment out if wanted
SeuratObject <- TRUE       # Export a Seurat Object
FullSeuratObject <- FALSE  # Seurat object contains all previous module output
transcripts <- TRUE        # Export Seurat contains transcript coordinates (large data)

# read in annotation file
# get input folder location and slide names
all_files <- system2(command = "aws",arg = c("s3","ls", paste0(studyDirectory, "logs/")), stdout = TRUE)
studyCreation <- all_files[grep("study_creation_", all_files)]
studyCreation <- strsplit(studyCreation, " ")[[1]]
studyCreation <- studyCreation[length(studyCreation)]

studyCreation <- paste0(paste0(studyDirectory, "logs/"), studyCreation)

annots <- system2(command = "aws",arg = c("s3","ls", studyCreation), stdout = TRUE)
annots <- annots[grep(".csv", annots)]
```

AtoMx features a native command line interface so that users can build in new analysis modules. At launch AtoMx features support for R libraries from Bioconductor, CRAN, and other key packages, with controls for memory & RAM utilization, runtime allowance, and parameters that are passed to the code within the system. All code is editable within AtoMx custom scripting environment. Additional language support will be released with future versions of the software.

Native Interaction Tools Enable Discovery in Spatial Multi-omics Studies



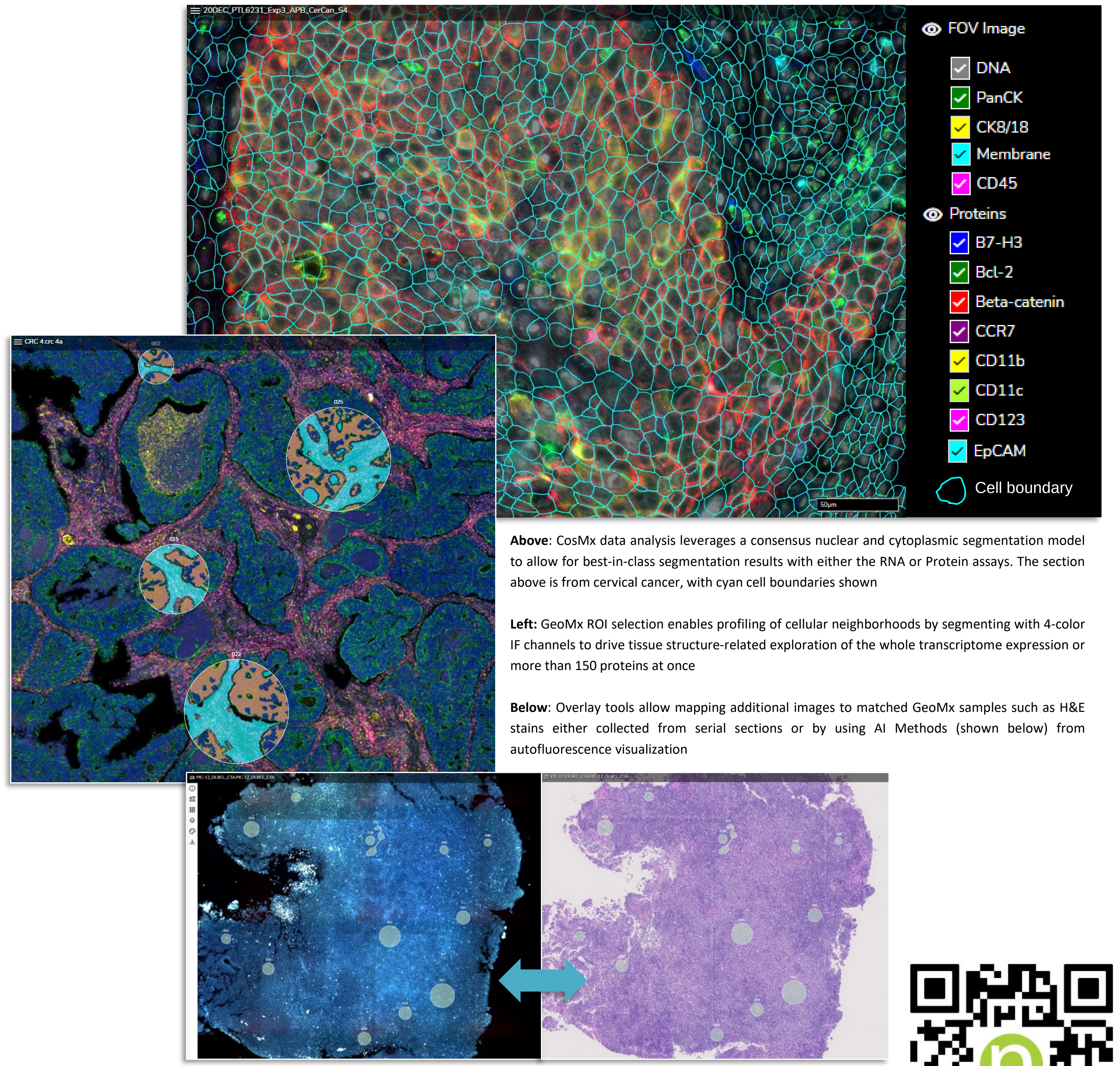


The CosMx Data Analysis Suite natively links all cells and images together so that you can visualize any results from an analysis in real-time

Interactive tools ensure that scientists of all backgrounds can approach and work with their data, regardless of the complexity of the methods deployed

Shown here Leiden clusters related to inactive (purple) luminal epithelial cells within the prostate have been selected to reveal the physical location of these cells in their spatial neighborhood

AI-driven Segmentation and ROI Selection Tools Enabled Across NanoString Platforms



Above: CosMx data analysis leverages a consensus nuclear and cytoplasmic segmentation model to allow for best-in-class segmentation results with either the RNA or Protein assays. The section above is from cervical cancer, with cyan cell boundaries shown

Left: GeoMx ROI selection enables profiling of cellular neighborhoods by segmenting with 4-color IF channels to drive tissue structure-related exploration of the whole transcriptome expression or more than 150 proteins at once

Below: Overlay tools allow mapping additional images to matched GeoMx samples such as H&E stains either collected from serial sections or by using AI Methods (shown below) from autofluorescence visualization

www.nanostring.com | info@nanostring.com

Visit our booth to see a live demo