WEHI's SODA-Hub Analysis Environment: Open-Source Workflow Infrastructure for Spatial Omics

Marek Cmero[1,2], Michael McKay[1], Michael Milton[1], Pradeep Rajasekhar[1,2], Lachlan Whitehead[1,2], Emma Watson[1,2], Patrick Lam[1], Julie Iskander[1,2], Rory Bowden[1,2], Marie-Liesse Asselin-Labat[1,2], Kelly Rogers[1,2], Tony Papenfuss[1,2]

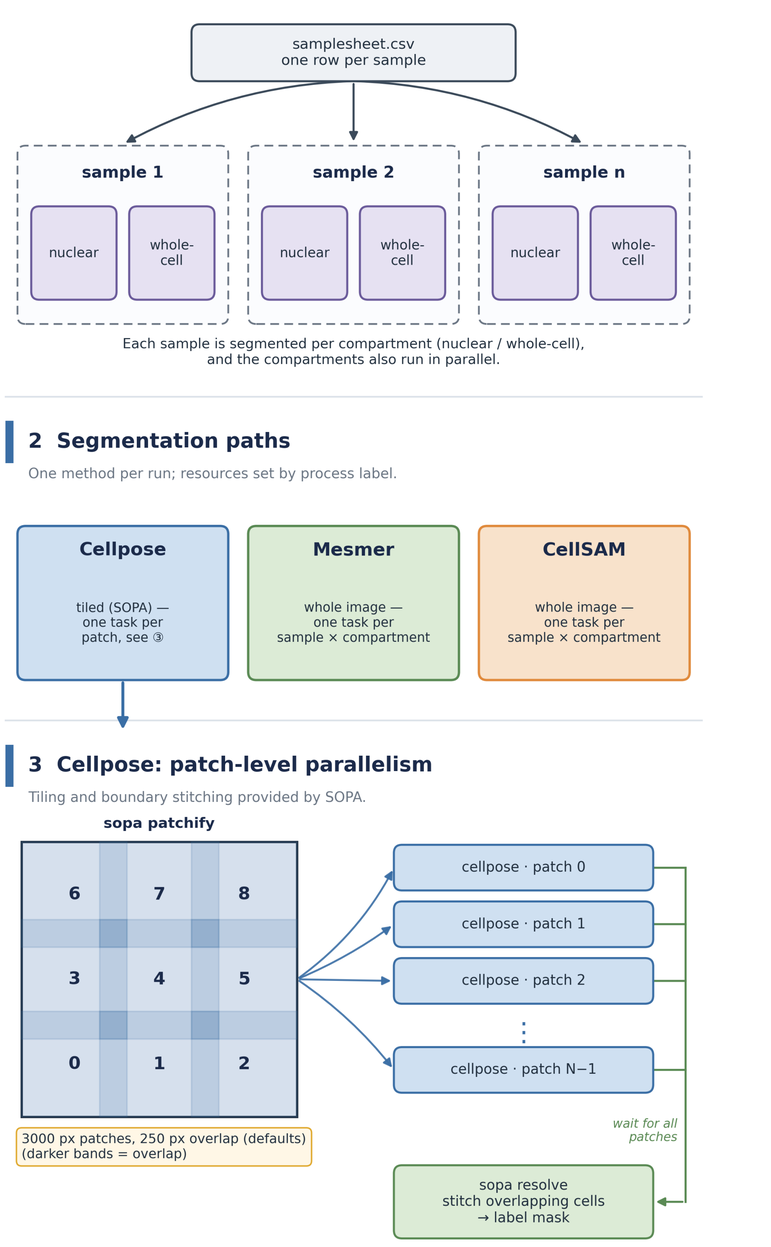
1. Walter and Eliza Hall Institute of Medical Research, Melbourne; **2.** Department of Medical Biology, The University of Melbourne

Background

- Spatial omics technologies are growing in use in biomedical research, allowing investigation into tissues in unprecedented detail.
- Large data volumes and multimodal nature of spatial technologies presents challenges in data management and processing.
- To address this, WEHI has established the Spatial Omics Data Analytics (SODA) Hub to provide a streamlined environment for spatial analysis, from data ingestion to analysis pipelines.
- The SODA-Hub is a collaborative effort between WEHI's Bioinformatics Division, Research Computing Platform, Bioimage Analysis Core, Advanced Genomics Facility and spatial researchers.

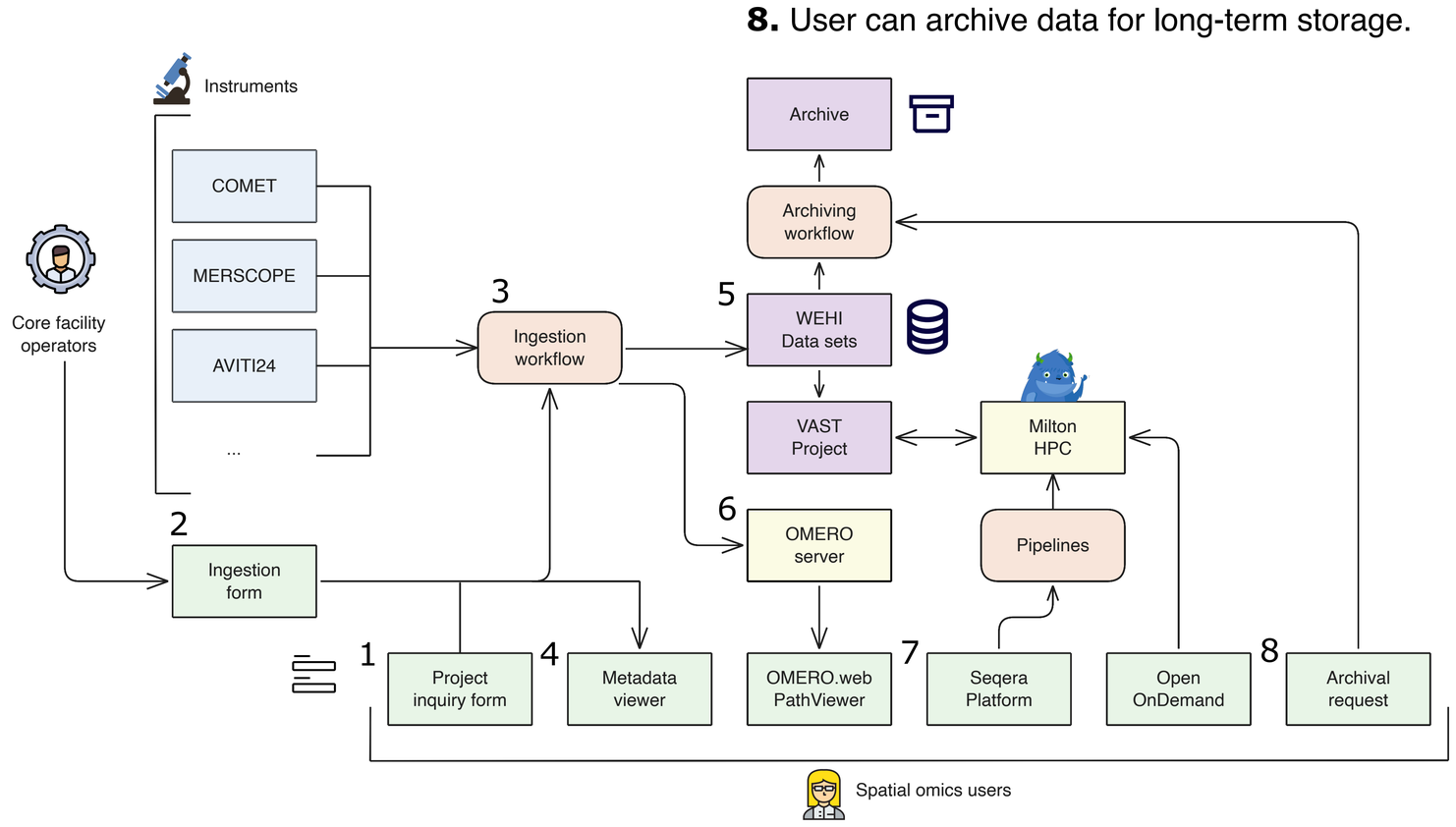
Pipelines and tools for spatial analysis

- To streamline cell segmentation, we developed the spatialvpt pipeline for the MERSCOPE, and the sp_segment pipeline for various spatial proteomics platforms. These pipelines are:
  - Open source: for wider use in the scientific community.
  - *User friendly*: CLI not needed when running via Seqera Platform.
  - *Portable*: dependencies handled.
  - *Flexible*: configurable and customisable.
  - *Traceable & interpretable*: detailed run & QC reports.
  - *Scalable*: patch large images and parallelise (below).



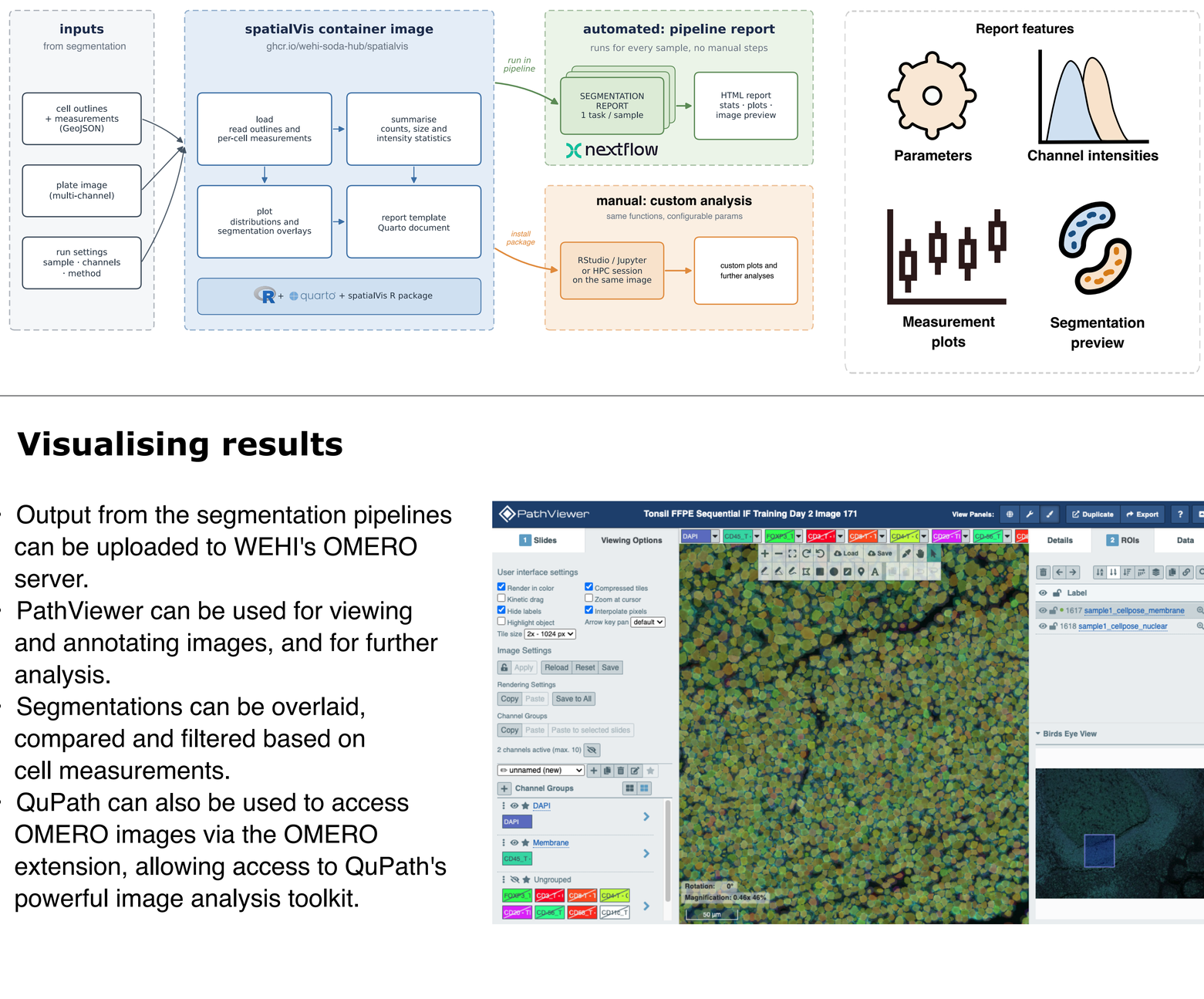
Spatial Data Flows

1. Spatial projects are captured in an inquiry form on Monday.com.
2. Core facility operator runs experiment on instrument and submits ingestion form.
3. Ingestion workflow triggers, transferring run to staging area and performing preprocessing.
4. User verifies metadata via metadata viewer.
5. Data and metadata are moved into data repository, accessible via VAST projects.
6. Images are imported in-place to OMERO server, accessible via OMERO.web/PathViewer.
7. Pipelines can be run via Seqera Platform on the local HPC. Data can be visualised on VMs using Open OnDemand.
8. User can archive data for long-term storage.



Report Generation

- The sp_segment pipeline generates an automated report to aid in quality control and data provenance of cell segmentation.
- The report captures parameters for each sample run, plots cell measurements, intensity distributions and segmentation previews.
- The spatialVis R package underpins the report generation and can be manually run for custom plots and analyses.



Visualising results

- Output from the segmentation pipelines can be uploaded to WEHI's OMERO server.
- PathViewer can be used for viewing and annotating images, and for further analysis.
- Segmentations can be overlaid, compared and filtered based on cell measurements.
- QuPath can also be used to access OMERO images via the OMERO extension, allowing access to QuPath's powerful image analysis toolkit.

Generously funded by the Kinghorn Foundation
Contact: soda@wehi.edu.au
Web: wehi-soda-hub.github.io
Code: github.com/WEHI-SODA-Hub