INTRODUCING A SOFTWARE PACKAGE FOR INTEGRATED VISUALIZATION AND DATA ANALYSIS OF SPATIAL OMICS DATASETS

W. Zhang[1], P. T. Tran[1], F. Wee[2], W. Yim[2], C. R. Joseph[2], J.C.T. Lim[2], L. Y. Chong[2], Z. W. Neo[2], M. C. Lau[3,4], J. P. S. Yeong[2,5,6], A. Ly[1], N. Verbeeck[1], M. Claesen[1], N. H. Patterson[1]

[1] Aspect Analytics NV, Genk, Limburg, Belgium
[2] Institute of Molecular and Cell Biology (IMCB), A*STAR, Singapore
[3] Bioinformatics Institute (BII), A*STAR, Singapore
[4] Singapore Immunology Network (SIgN), A*STAR, Singapore
[5] Immunology & Serology Section, Department of Microbiology, Division of Pathology, Singapore General Hospital, Singapore
[6] Department of Anatomical Pathology, Division of Pathology, Singapore General Hospital, Singapore

Introduction

Spatial omics enables detailed molecular analyses directly from tissue, providing novel insights into the tumor immune microenvironment (TIME). As the field matures, users are increasingly conducting tissue cohort analysis and/or combining different omics readouts to obtain a holistic view of TIME biology. However, spatial omics data analysis presents specific bioinformatics challenges. In spatial omics, data is typically acquired at different spatial resolutions using a variety of platforms, while data volumes and batch effects are issues for cohort studies. We present Weave®, a cloud- based software that addresses the need for spatial omics bioinformatics solutions, enabling efficient integration and joint visualization of different assays, as demonstrated via two different use-cases.

Experimental Data Collection

Spatial multi-omics

Human lung cancer tissue sections were analysed with spatial transcriptomics using a human cancer panel targeting 289 genes per the recommended workflow (Xenium, 10X Genomics), followed by multiplexed immunofluorescence using a 40-antibody panel (COMET, Lunaphore), and then H&E staining that was subsequently digitized using an Axioscan 7 (Zeiss). Cell segmentation was performed in two ways: first, using DAPI-based nuclear expansion from 10X Genomics on the Xenium data [1]; and second, employing CellSAM, an advanced deep learning-based method that incorporates both nuclear (DAPI) and membrane (PanCK) markers from the COMET dataset. Correlation between gene expression (transcript count) and protein expression (mean immunofluorescence intensity) captured on the aforementioned segmentation boundaries was assessed using SciPy [3].

Cohort study

Fresh frozen EBV-associated cancer tissue that received anti-PD-1/PD-L1 immunotherapy treatment (responder n = 2, non-responder n = 6) were measured via spatial transcriptomics following recommended protocols (Visium; version 1, 10X Genomics). Pathology annotation was performed in QuPath on the digitized post- Visium H&E-stained sections. Batch correction was conducted using different algorithms [4-5].

Results

Reading direction

Spatial data registration and integration

Our approach for spatial multi-omics data analysis consists of multiple steps. Prior to data analysis, metadata labeling of individual section data informs the creation of stacks and helps identify sequences of tissue sections where stack creation is feasible. Accurate, non-rigid image registration is then used to create a single coordinate system across different measurements that can account for different sections or variations in measurement region size. For data that are generated from the same section, Weave's co-registration tool can be accurate to the single-cell level (Figure 1).

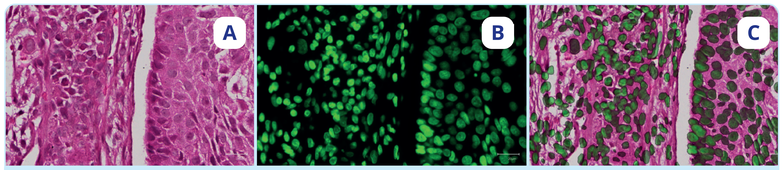

Fig. 1: Example of co-registration accuracy **A:** H&E of lung cancer biopsy **B:** DAPI channel from COMET dataset pseudo colored in green/DAPI channel from Xenium dataset pseudo colored in yellow. **C:** Overlay of the two images. Scale bar is 20 µm.

Joint visualization of integrated spatial multi-omics datasets in a single software

Communication and visualization of data is via web-based Weave reports, which can be shared via URLs. Figure 2 shows two examples of Weave reports displaying the integrated Xenium, COMET and H&E datasets for lung cancer samples, and data analysis results (cell segmentation, clustering results). The Xenium and COMET datasets with their respective H&E images and data analysis results are overlaid in a single view in the Spatial panel. Control of the visualization of the different modalities is via individual dropdown menus in the Spatial Layers. Xenium results can be either be visualized as cells or cell contours that expressed a gene via the Gene List panel or as transcript spots or density maps via the Transcript spot panel. The Spatial panel and relevant data plots (e.g. UMAP scatterplot) can be interactively controlled, allowing for zooming, panning, and selection of interesting features.

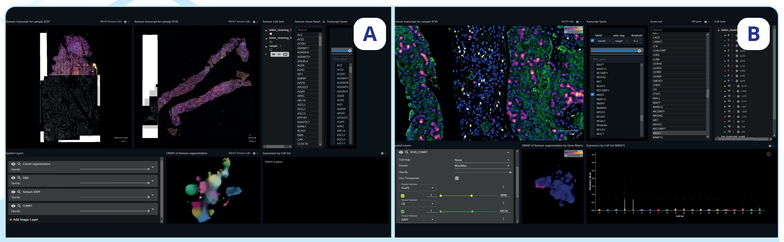

Fig. 2: Joint visualization of Xenium, COMET and H&E data in a Weave report. **A:** Screenshot of the starting page for a report featuring two lung cancer samples. **B:** Demonstration of gene-protein pairs. Xenium MKI67 in magma density plot, COMET CK in green, COMET FoxP3 in blue, overlaid onto the Xenium DAPI channel.

Correlation analysis of gene transcript and protein expression patterns.

From the 289 genes targeted with the Xenium, 16 corresponding proteins were targeted using COMET. We conducted correlation analysis to identify which transcript-protein pairs have similar spatial expression, and if this correlation was affected by cell segmentation approach. Table 1 shows the results for the transcript-protein pairs. Some pairs had high correlation, regardless of cell segmentation (e.g. CDH1/E-cadherin), while correlation of a protein markers which refers protein complexes that derive from different genes yield variable result (e.g. CD3).

Validation of the correlation between COMET-Xenium signal will be addressed in a follow-up study.

xenium	comet	correlation (comet segmentation)	correlation (xenium segmentation)
MSA4A1	CD20	0.53	0.50
FOXP3	FoxP3	0.52	0.45
CD3D + CD3E	CD3	0.51	0.44
CDH1	E-Cad	0.50	0.62
CD34	CD34	0.42	0.44
CD3E	CD3	0.41	0.35
CD3D	CD3	0.41	0.34
CD14	CD14	0.40	0.44
CD8A	CD8	0.39	0.34
CD8A + CD8B	CD8	0.38	0.33
FCGR3A	CD16	0.38	0.47
CD68	CD68	0.37	0.42
KRT15 + KRT7	CK	0.34	0.38
KRT15	CK	0.34	0.44
MKI67	KI67	0.33	0.33
CD38	CD38	0.31	0.30
CD163	CD163	0.31	0.37

Table 1: Correlation of gene transcript-protein expression when using COMET-based cell segmentation and Xenium-based cell segmentation.

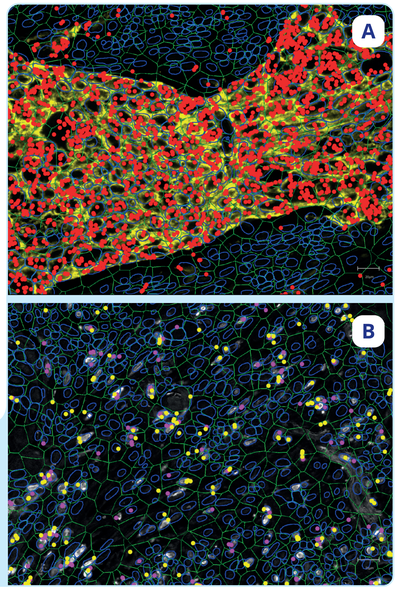

Fig. 3: Joint visualization of transcript-protein pairs with different cell segmentation.

A: CDH1 transcripts (red spots) and corresponding protein E-cadherin (yellow) had high spatial correlation, regardless of whether cell segmentation approach was via Xenium (green contours) or COMET (blue contours).

B: Protein complexes, e.g. CD3 (white) have more variable correlations with respective genes CD3D (fuschia spots) and CDE (yellow spots), regardless of cell segmentation approach. Scale bar is 20 µm.

Multi-sample viewer for Visium datasets

The Visium Spatial Transcriptomics assay (VST) has been widely adopted as it maps quantitative gene expression data to their original location in tissue sections [1]. This has produced demand for software where multiple VST datasets can be collectively viewed and analyzed (Figure 4). Weave allows simultaneous viewing of a cohort of VST measurements, including integration of complementary data such as bioinformatics results (Figure 5). Individual panels can be resized or repositioned according to user preference, e.g. re-arranging sample grid or resizing the spatial panel in order to concentrate on a specific sample.

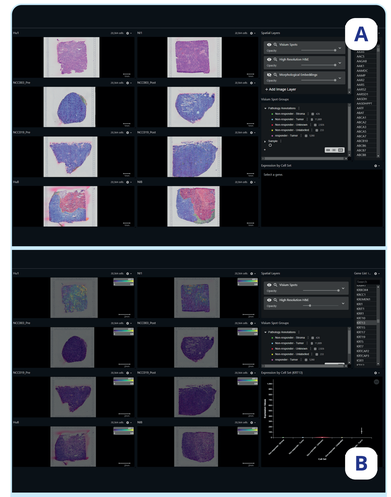

Fig. 4: A: The starting view of the software showing all eight samples, with the Visium spots overlaid onto respective H&E images. In this view, the color of the Visium spots correspond to regions annotated by a pathologist. Key to the regions shown in 'Visium Spot Groups' panel (arrow). **B:** Intensities for gene expression can be viewed collectively such as for KRT13 which is highest in responder samples.

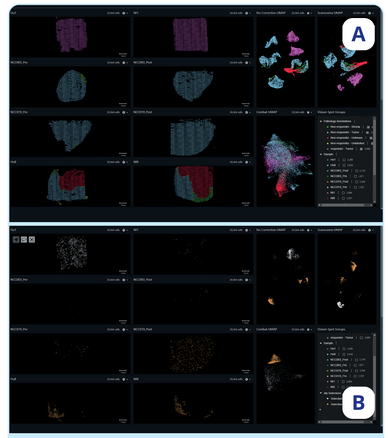

Fig. 5: Example of integration of complementary data – batch correction scatterplots. **A:** Different batch correction algorithms were applied to the datasets. The resultant UMAP scatterplot for each approach can be integrated for interactive investigation. For example, Scanorama shows two distinct responder tumor populations (in pink, white arrow), while COMBAT indicates what might be a population in non-responder tumour (white asterisk). **B:** Lassoing the populations in the scatterplots shows their locations in the samples - the smaller responder tumor population is only in one sample (yellow arrow), while spots belonging to the non-responder tumour subpopulation are mostly in two specific samples (yellow asterisks).

References
[1] https://www.10xgenomics.com/support/software/xenium-onboard-analysis/latest/algorithms-overview/segmentation#seg-nucleus-expansion
[2] Israel, Uriah, et al. "A foundation model for cell segmentation." bioRxiv (2023)
[3] https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.pearsonr.html
[4] Johnson WE, et al. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics. 2007 Jan;8(1):118-27.
[5] Hie B, et al. Efficient integration of heterogeneous single-cell transcriptomes using Scanorama. Nat Biotechnol. 2019 Jun;37(6):685-691.

AGENTSCHAP INNOVEREN & ONDERNEMEN · Vlaanderen is ondernemen · Horizon · FLANDERS INVESTMENT & TRADE · Flanders State of the Art

Conclusion

- As spatial omics is increasingly used to investigate TIME biology, Weave® software addresses the need for spatial omics and spatial multi-omics bioinformatics solutions, as shown via two use cases.
- Overlays of multiple spatial omics datasets from different technologies and vendors can be integrated, enabling direct visual comparison with interactive browsing of full resolution images and data analysis results.
- Simultaneous viewing of tissue cohorts, as demonstrated with VST datasets, with complementary data analysis results allows for collective data interpretation.