Q&A Hybrid Day 2 (online)

Talk 1 - From Cells to Space: Spatial Transcriptomics Technologies - Julieta Aprea

Q: Is this method also recommended for Coral samples tissues? (In regard to their hard tissues and normally we freeze them?)

A: Unluckily I do not have experience with hard tissues. But I would highly recommend you to contact the support of the different companies and they will guide you on the tissues they have tested and any recommendations they might have.

Q: In Visium HD spatial transcriptomics, what are the potential drawbacks or complications of having overly strong or saturated signal intensities in certain regions? Could this affect downstream data analysis or interpretation, such as clustering, normalization, or spatial pattern identification?

A: In VisiumHD panel based approach, since it is panel based, the panel is already titrated, which means that highly expressed genes will get less probes per gene. In the case of mRNA capture I could not tell you as it has not been released yet.

Visium HD can handle tissues with high levels of heterogeneity in terms of tissue architecture. For example in tumour you would see higher levels of UMIs/Features in TME compared to fibrotic regions. You can nicely correlate the tissue morphology with the transcriptomic layer. From the downstream analysis, the bins (or cells after segmentation) are library-normalized, accounting for the variation in the counts between bins. So this is usually taken care of in Seurat or squidpy. I hope this answers your question :)

Q: If the average size of a cell is 10µm, why do we need a resolution higher than 8µm spot diameter?

A: Cells come in different sizes and we give the user an option to change the bin size to account for it. We have a new version of Space Ranger coming, which will use the H&E image and create cell masks based on this, to account for different cell sizes.

Q: Maybe I missed this, but why do we have a low mRNA capture efficiency with the fine-grained (2um) sequencing-based spatial technologies?

A: Here the biology of the samples is important as there might not be so many transcripts present in this limited area that would allow meaningful analysis. In addition, given the

limited physical area that each spatial barcode has, then there are not as many capture oligos for getting the mRNA.

Q: Regarding Visium HD, any idea when it would be available for porcine ?

A: The Visium HD has been upgraded for 3' and was released a few days ago. This should allow it to be species agnostic as long as you have a good reference genome.

Q: In imaging-based technologies, is there any information on "yield"/signal strength per round of hybridization?
E.g. would I get different signals of Gene A according to the hybridization cycle in which it is amplified?

A: All of the transcripts are measured multiple times over several cycles. So in theory there should not be a difference

Q: For image-based analysis, are the probes also limited to human and mouse?

A: With Xenium we can design probes for Xenium v1 for other species than human. Xenium Prime is currently limited to human and mouse, but not Xenium v1.

FOLLOW UP Q: As a follow up, using these technologies in the context of infection, for example, is it possible to have probes for host & pathogen?

A: Yes - we can create probes for bacterial transcripts and detect them in the tissue of a host.

Q: Hi, is there a possibility to do a sequencing based and imaging based approach on the same tissue? So for example first doing Visium and then Xenium? I know I could do a 4-plex IF staining after Visium but what if I have 20-30 targets I would like to check in an imaging based approach?

A:Xenium is non-destructive and you can perform additional experiments afterwards. We have some guidance on our website on running Visium HD post Xenium.

Talk 2 - Introduction to spatial RNA sequencing data - Roman Laddach

Q: More of a comment, again in an infection with intracellular pathogens, having an increased resolution, can also include some patterns from the pathogen, no?

A: Yes - the transcript resolution for Xenium in x/y coordinates is <30nm, and for the z stack <100nm. You should be able to detect pathogens

Q: Is there a way to check the correct binding of the barcodes?

A: For both Visium HD and Xenium we provide the full information about the probes, which you can double check against your reference genome.

Q: Are there any analysis variations in using Space Ranger or usnig the Cloud Analysis?

A: No - there Space Ranger on the 10x Analysis Cloud should work in the same way as on the local installation (i.e. HPC). What you may find is it may be faster on the Cloud.

Q: How does it help doing nuclei segmentation if you might still get a nuclei spatially located across different spots?

A: The probability of nuclei being within 2um bin should be very low, so the nuclei segmentation should help with the downstream analysis.

Q: Is it mandatory to perform cell segmentation on Visium HD data since its sequencing-based?

A: It is not mandatory, but it will help with your analysis. Without the segmentation you may find bins (8um) which have genes from multiple cell populations. The segmentation should help in those instances and result in spatially separated clusters.

Q: I heard that Xenium is used for wheat (plant) tissue and the results are published! Did I hear you right?

A: Here is an example of Xenium run on maze -
https://www.nature.com/articles/s41467-024-55803-9
“Multiplexed transcriptomic analyzes of the plant embryonic hourglass”

Q: Is there any alternative for 18s rRNA for interior RNA in multimodal segmentation?

A: Not at the moment - it is manufactured together with the other stains. If you have some suggestions about it or would like to discuss your use case where 18s rRNA is not suitable for your project, can you send an email to support@10xgenomics.com? I would love to hear more about your project.

Q: Why are cell membrane or cell surface stained images not used for cell segmentation in Xenium technology?

A: If you use the multimodal kit, then the cell membrane and internal stain are used for cell segmentation. If you do not use the kit, then the segmentation will be based on the DAPI stain and expansion.

Q: Hi, do you have an idea, why most of the published literature used only FFPE for visium HD? Another question: does it affect the quality of data if we stored fresh frozen tissues in -80 C for some time and not directly send it? Is there any time limit for the storage?

A: I think there is a high number of tissue blocks embedded in FFPE as well as FFPE preserves tissue morphology better than FF. That would be the main reason I would assume. I do not have the answer for the storage question - I am a dry lab person. Can you send an email to support@10xgenomics.com with your question and we will be able to follow up.

Q: Thanks for sharing about the upcoming version that includes cell type annotation that sounds really useful. I was wondering, for Visium HD data that we've already processed and explored in the Loupe Browser, will it be possible to apply this new feature directly, or would we need to rerun the analysis from the raw data?

A: You will need to re-run your analysis with the new Space Ranger to and have the high-res H&E image to benefit from the new cell segmentation.

Q: Hi, the Xenium Ranger, could it be used with Macbook Pro M3 max processor?

A: Unfortunately no - all of our ranger pipelines are only suitable for standard Linux installation on an AMD or Intel processor. I have seen a few requests to enable them on ARM processors - we have a Software Feature Request for this to be considered in the future.

Q: So far, all the sample examples are mostly mammal-oriented but, what about plants? Can we detect transcripts in the cytoplasm without removing the cell wall?

A: Yes - I shared the maze paper link. Hopefully this will be useful for you!Hi there! Here is an example of Xenium run on maze - https://www.nature.com/articles/s41467-024-55803-9

“Multiplexed transcriptomic analyzes of the plant embryonic hourglass”

Q: Hi, can I also detect ncRNAs?

A: As long as we can create a probe then it should not be a problem. In general we require a sequence of around 80 bp long to create a set of probes. Additionally, the probes have to match certain constraints (i.e. melting point). That would be a part of the Advanced Panel Design.

Q: Could you please share a paper or link regarding guidelines for cell segmentation in visium HD? ex: when it should be performed, what are the warnings?

A: We have an analysis guide here:
https://www.10xgenomics.com/analysis-guides/segmentation-visium-hd

Q: My question regarding 18s is, again in a context where you have intracellular pathogens, there could be constraints in the cell segmentation process using common markers between pathogen and host, such as 18s rRNA. Is there any information on cell segmentation efficiency in a scenario like that?

A: We do not have anything public for this, but I am sure it was considered during the R&D stage. The segmentation kit focuses on “normal” cells and if you have a pathogen, you would need a different approach to identify it - like specific probes.

Q: I see you addressed a question from the colleagues related to plant research, that’s nice to see that there are more plant-people representatives members! I would be nice to have the paper links from wheat and other plant species published works, thanks

A: Here is a link to a paper for Xenium and maze:
https://www.nature.com/articles/s41467-024-55803-9

Q: Is the protein panel available only for Xenium? I am asking this because I know a project that uses multiplex imaging, that could be interested in performing spatial and this combination looked very interesting!

A: The protein panel is available for Xenium with 28 predefined immuno-oncology proteins. Xenium is non-destructive and you can run other assays afterwards on the tissue. The only aspect will be to combine the data across modalities - and there are multiple tools for this (i.e. SpatialData, VoltRon).

Talk 3 - Introduction to spatial analysis - Fabian Rost

Q: Hi, what is the difference between the high and low resolution images in the output for visium? Why do most of the analysis tutorials use the lower quality image?

A: Not exactly sure how they are created, probably Julieta would know. The reason to use low resolútion images in the tutorials, I think, is just runtime. Plotting the high-resolution image multiple times can be quite time consuming. Also, if you just do small plots, low-resolution might just be sufficient.

Q: Considering spatial data have additional information (location) that a single cell does not, are there any ideas/approaches to assess spatial bias and, if yes, try to normalize based on that?

A: I'm afraid I do not understand the question, I would need more content on what you mean with spatial bias.

Q: I have a question regarding coding in R while doing spatial analysis. If I have several animals, and wanna integrate the samples in order to do PCA, clustering and so on, should I use the function "IntegrateData" or "IntegrateLayers"?

A: I am not a Seurat person, but I took a quick look. I think both work. IntegrateData seems to be part of Seurat's first approach to do data integration. See https://satijalab.org/seurat/articles/integration_rpca#performing-integration-on-datasets-normalized-with-sctransform

IntegrateLayers seems to be a wrapper for different integration methods: https://satijalab.org/seurat/articles/seurat5_integration#perform-streamlined-one-line-integrative-analysis

Q: Hi, There are many methods regarding detecting spatial domains for visium (Banksy, deep learning methods like GraphST), how can I determine which method is the best for my data?

A: I am not aware of a comprehensive benchmark. I think you have to try it yourself.

Q: Banksy has many parameters and it is hard to detect these parameters (like k-neighbours and resolution), is there any recommendation regarding this issue?

A: I'm afraid there is not much. You can check out their paper and maybe the Seurat vignette (https://satijalab.org/seurat/articles/visiumhd_analysis_vignette#identifying-spatially-defined-tissue-domains)

Q: Taking the example of the immune microenvironment in a tumor tissue, for example, glioblastoma and spatial analysis of the same. As macrophages can phagocytose tumor cells, is it possible that due to the RNA signatures, macrophages might get recognized incorrectly as tumor cells or vice-versa? If so, how do we deal with this issue?

A: It depends...Image-based methods: depends on the segmentation. If you segment based on DAPI/cell membrane, you might see tumor/macrophage doublets. If you segment based on the transcriptome (Baysor etc), these might be called separate cells. I think there is no way to tell. You need to try and then examine your data. I do not think that there is a simple answer.

Q: How can we do QC for a whole brain section, where there many regions are and there could be differences in mitochondrial percentages between regions. and choosing a specific percentage for the whole brain is maybe not appropriate?

A: You could start without QC and then check for clusters that you cannot make sense of. Then kick out those bins/cells, and analyse the rest. Katrin suggested a similar procedure for scRNAseq yesterday.

Q: Do you have an idea, which method is the best for finer domain segmentation in visium v1 (detecting substantia nigra in sagittal mice brain image)? How to choose the best deconvolution method for my data?

A: According to https://www.sc-best-practices.org/spatial/deconvolution.html#key-takeaways, Cell2Location performs best. But this is not really segmentation, but cell type deconvolution. You need scRNAseq data for this. For real higher resolution you would need a different technology.

Q: As a curiosity, is there any approach to rank ligand-receptor pairs based on ligand diffusion, proximity effect, etc. that can allow for "prioritizing" checking these pairs in a neighborhood analysis?

A: I am not aware of a method that takes these things into account. I also imagine that ligand diffusion for morphogens can be highly tissue dependent. Not sure what proximity effect means though. Some tools take into account that a ligand has to form a dimer with another molecule to be functional. I think checking out the liana paper would be a good starting point to check all the methods they mention there for the details of the assumption.

Q: You skipped the denoising step — how essential is it in this context?

A: I think it is not essential, but it can help getting cleaner cell types. Check out https://docs.scvi-tools.org/en/1.3.0/user_guide/models/resolvi.html

Q: When integrating with single cell sequencing: let's say I have cortex domain within my mice samples and I want to integrate it with a single cell sequencing reference of the cortex. The only available references are parts of the cortex and not the whole cortex. Is there any idea how I can solve this problem? Could I combine all matrices of these references in one reference? If yes, should I do any normalization with/without scaling before integrating this combined reference into my data? And in general should I use normalized single cell sequencing for integration or raw data?

A: Normalised vs raw data: depends on the concrete tool that you use.
I think most tools do not directly address batch effects.
In general, it is best if both datasets match as good as possible. If there are different regions of the brain and different scRNAseq datasets, you could try to crop out the very same regions in your spatial data, and then integrate the matching scRNAseq dataset.

Q: Thanks for the great talk! I was wondering, when using Cell2location in complex tumor tissues, how do you deal with the challenge of capturing rare or reprogrammed cell types that might not be well represented in the reference scRNA-seq data? Do you combine it with any other spatial methods to improve resolution

A: Tough question. If you do not have the cell type in your scRNAseq data, there is not much that you can do. Even with Visium HD it can be hard to detect rare cell types. Check out this paper: https://www.nature.com/articles/s41586-024-07563-1, they could only find a rare cell type in Visium HD, because they imputed these cells from a scRNAseq dataset. This might also be presented tomorrow by Heather. I believe imaging based methods could give you the necessary spatial resolution, but you would depend on the correct probes in your panel to identify your cells. Spatially aware clustering (like banksy) might also help to identify rare cell types. And of course you could just increase the number of cells in your scRNAseq dataset, to finally see those cells.

Q: I just want to make sure that I understood correctly: is the mixture clusters issue due to the fact that we are clustering spots instead of cells?

A: Exactly, that would explain it for spot-based methods. For image-based methods we could see these artifacts because of incorrect cell segmentation. ResolVI can also help in denoising the UMAP.

Talk 4 - Integrating single cells into a spatial context in the era of personalized medicine - Joana Bernardes

Q: Do you have clinical data, like calprotectin levels, or even more general markers of inflammation, that you could correlate with the scRNASeq?

A: Always! Do because our datasets are sparse as we deal with human samples. We make sure to have a lot of clinical information; calprotectin, IL6, CRP, disease activity scores etc. This is how I leverage studying some cell-types over others

Q: How is remission vs non-remission determined, from the transcriptomic data? Are there biomarkers that can separate these cases from your data?

A: So Remission and no remission is attributed by the clinicians. So it is a mixture of self-assessment questionnaire, physician assessment and endoscopy score (add pathology scores sometimes). We are now working on getting a molecular remission score (from the flare dataset)

Q: Thank you for the insightful talk. I was particularly curious about the integration part. Could you elaborate a bit more on the CODEX–scRNA–scATAC-seq approach?

A: This is a bit of a longer discussion. As I am still trying to apply some of the methodology myself. So single to spatial is straight forward. spatial to single not so much and this what I am working on right now. I am happy to share the paper https://www.nature.com/articles/s41587-023-01935-0 and we can continue our discussion maybe with email.

Q: Enrichment data - doing almost an enrichment analysis per participant (not sure if possible) to see if enrichment of a certain GO term, KEGG pathway, etc.. Thank you for the response!

A: Yes, so we try already on most comparisons, enrichment by gene group. So we favour first know genes with protein ID, and then we go GO term/Hallmark term (enrichment term) to try to disentangle if true for all patients or if there is one sample driving the signal.

Q: (Came up after the talk) Thanks Joana!! As a follow up - have you integrated the clinical data with single cell, to correlate features from both sources? You could check if you get a similar dynamic that you found in literature (e.g. for calprotectin)?

A: Yes! I always keep minimum clinical info in the metadata. I have correlated clinical data with: 1. cell proportion, 2. coexpression patterns and 3. gene groups

Talk 5 - What spatial omics reveals that other omics might miss - Nicolas Casadei

Q: Are there tissue atlases for other species like mice and rabbits?

A: In the database https://db.cngb.org/stomics/ are relatively many experiments in human, but also mouse and non reference organisms.

Q: Sorry to go slightly off topic, I was just interested about the findings of your team on Alzheimer's, since Lesné's research on amyloid plaques and AD allegedly had fabricated results (which does not necessarily mean they are incorrect and the theory seems solid in many regards). Were you able to see this connection (amyloid plaques and expression characteristic of AD) when you were analysing spatial expression in affected brain cells?

A: My opinion is that the evaluation of fabricated data might not be easy to detect, but if other scientists showed their concerns and you would agree that the concerns are justified, you should not use these data in your research. That being said, our team is working on the identification of transcripts changed in the microenvironment of plaques or protein aggregate. Our main issue is that the classical neuropathological staining is not compatible with the spatial transcriptomics kits we are using. We are still working on this topic, so no more information I can provide you.

Q: In SpaTalk, is information about ligand diffusion, receptor specificity, etc. used during the process?

A: Information about ligand and receptor are from public databases, these databases do not have systematic models for a specific receptor activation or other more complex cell signaling particularity. I am not aware of a computing method or statistical method to measure and correct such parameters.

Q: Could you please elaborate a bit more on the knowledge base that Spatalk uses to get information of all possible ligand receptor interaction pairs.

A: 1) The publication https://www.nature.com/articles/s41467-022-32111-8#Sec10 mentions KEGG / Reactome and CellTalkDB

Q: Does Spatalk preclude the interaction between Ligand and receptors beyond a certain distance. If yes, what is this threshold?

A: The distance is a factor in the weight matrix. I am not sure how simple it would be to set for a certain pathway certain weight for distance. Sounds to me very interesting but difficult to characterize and implement.

Q: Did I understand correctly, that if I want to use SpaTalk I need the scRNAseq dataset from the same tissue piece I'm doing the ST on? Or is it enough if it's a reference scRNAseq dataset from the tissue type? So for example if I have a tumor sample can I just use a publicly available reference dataset for that tumor type or does it have to be from the same tumor tissue?

A: The input should be the matrix space / gene expression. Single cell integration in spatial would help as many of the transcription factors and signaling genes are low expressed. But single cell are optional.

Q: Has any assessment been done to determine if, by introducing more data collected via omics, there is any effect on the transcriptomic part, both in efficiency or expression pattern?

A: Excellent question, the more the samples are handled, the less good the results are. Our experience with protein barcode + RNA is good and does not reduce much the quality of the transcriptomics.

Q: Thanks Nicolai, the question was indeed if such information is available in the databases of ligand-receptors and, if so, if it was used in any way. But a very important reminder on transgenic experiments, especially with the receptors!

A: Yes, we forgot too often that changing genetics and environment deviate our data set of the reference / physiological condition

Talk 6 - Latest trends in bio-image analysis - Johannes Soltwedel

Q: Very naive question but, can weak labelling be used for diagnostic, e.g. disease vs not disease?

A: Yes! That's exactly how it can be used :) It's probably worth putting some thought into what reasonable categories to predict are, but in principle, the answer is yes.

Q: Thank you for the great presentation. I was wondering if Cellpose-SAM is ready to be installed and used by users now. How computationally intensive is it? Would it be suitable for daily analyses on a standard lab computer?

A: To my knowledge, it can readily be installed. Since it has SAM as a backend, it would require a (reasonably dimensioned) NVIDIA graphics card in order to work, though. I'm not sure whether standard lab computers have these, tbh.

I think plugins providing cellpose-sam should offer some functionality for tile-wise application, because that's probably the only way to make it work for large-scale image data.

Q: When one is comparing images, is batch effect (e.g. collection in different times) relevant in processes like segmentation?

A: Yes, absolutely! In time-lapse images, for instance, you'd have effects like bleaching that diminishes the intensity of stained objects over time. In different batches, different chemical lots may be used, all of which can lead to heterogeneity inside the data. A good way to examine batch effects in measured features is typically PCA.

Q: Since you talked about containers, to your knowledge, is there a repository for containers in the context of biological data analysis?

A: Good question. I think the bio-image model zoo is moving in this direction, but as of this day it's probably easiest to upload containers on the docker hub. Otherwise I think it's possible to set up singularity repositories on an institutional level, but maybe the docker hub already serves most purposes.

Q: As a follow-up, when one has several clusters and several cells segmented, UMAPs tend to be used instead of PCA's. Can UMAPs be used to identify these batch effects, or should we use the good old PCA's?

A: I would expect batch effects to go missing in a UMAP. Depending on the parameter set in the UMAP dim. reduction, you *may* still pick it up if you are looking for clusters on a global scale (look for the min_cluster_size or n_neighbors parameters) but generally I'd rather use them side by side for what they are best at.

Talk 8 - Single-cell and spatial atlases across human organs - Simon Koplev

Q: This sounds a bit philosophical but, I am wondering what one can separate cell types from cell status? From a practical standpoint, do/can we do this bioinformatically?

A: Good question. I think of cell types as stable configuration and states as more transitory and reversible cellular states. One could look at predicted trajectories to get at this. But, it may be out of scope for single-cell RNA-seq alone to address this computationally.

Q: For the ligand-receptor database for the glial cell - pacemaker comparison, you said you "complemented" the database for GPCR. Was this done manually? or was any automatic tool/approach used?

A: It was largely curated manually from the literature. The additions should be part of the newer versions of CellPhoneDB.

Q: At which stage of development is the hormone cell atlas?

A: If you mean by human development, the data and scope of this project is based on adult and pediatric samples. If how close to getting published, I'm not sure. It is relatively far advanced -- so possibly this year.