PDAC

Comparative Analysis of Cell Types in Pancreatic Ductal Adenocarcinoma Using Digital Spatial Profiling Gene Expression Data

If you end up joining us for your summer research project, you will be working on understanding immune-cancer interactions in pancreatic cancer using a novel data type called digital spatial profiling. To give you a small sense for what it is like to work with these data I have collected a small dataset for you to play around with. This dataset is from an existing paper (Carpenter et al., 2023) where the authors tried characterizing the early state of pancreatic cancer. We have a much bigger dataset that was created with the same technology to significantly expand on this work for your actual summer project.

This assignment is designed to be hard but doable. There will be some questions that you will likely not have an answer to, that is ok, just do you best. I will be sure to give you feedback and expand on anything you are curious about. You are also welcome to email me with any questions or ask for help whenever you feel like you need it. Think of this as a way to get a sense for the supervision and help you will get if you do your summer project with us.

Some background:
Pancreatic Ductal Adenocarcinoma (PDAC) is a lethal disease with a five-year survival rate of less than 10%. The pancreas plays an important part in digestion. It creates digestive enzymes that are transported from the inside of the pancreas to intestine through ducts. To accomplish these functions, it contains several important cell types including:

- Epithelial cells: these line the ducts through which the digestive enzymes flow
- Acinar cells: these synthesize, store and secrete digestive enzymes
- Endocrine cells: these create hormones (e.g. insulin and glucagon)
- Immune cells: these surveil the environment for infections and cancer

Understanding the intricacies of PDAC cell types and the cells cancer originates from (cell type of origin) is imperative for advancing treatment strategies. However, the cell type of origin for PDAC is still unknown. Two hypotheses for the cell type of origin have been proposed:

2 possible hypotheses

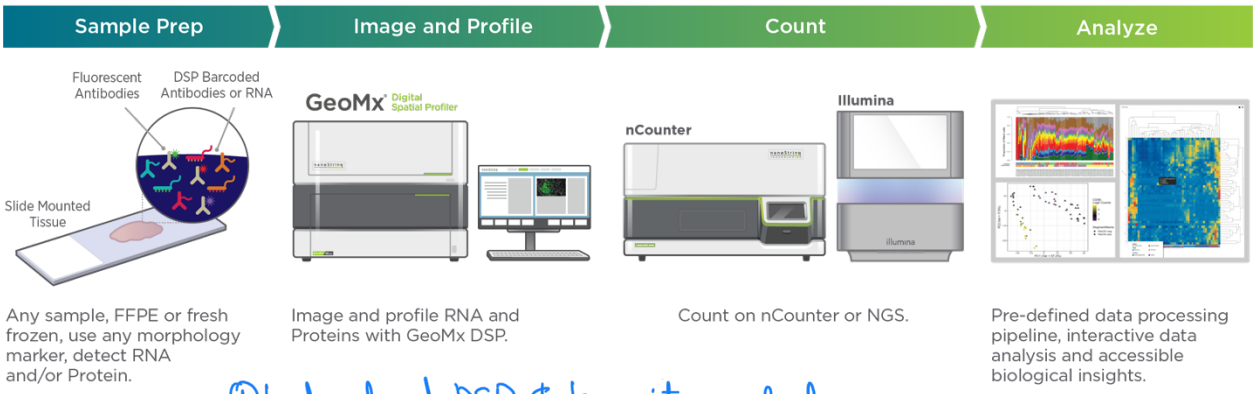
- Acinar to ductal metaplasia: in this process acinar cells transform into ductal cells that then end up forming PDAC
- Directly from ductal cells: this would involve a direct transformation of ductal cells, however, there is some controversy as to whether these cell types even exist in the pancreas

In addition, PDAC is thought to start as pancreatic interepithelial neoplasia (PanIN), these are common in the general population and are thought to be the precursor for PDAC. Once cancer progresses from PanIN to something more malignant it can further be classified into distinct subgroups. Glandular tumours for instance look more like normal tissue, while poorly differentiated tumours look less like normal tissue.

Starts as PanIN (precursor for PDAC)

The technology: Digital Spatial Profiling (DSP)
This is a sequencing platform built by a company called Nanostring. It allows us to take a biopsy from a patient, mount it on a slide and select very specific regions of interest. These could be regions with a lot (or very few) immune cells, regions with lots of cancer, regions where cancer used to be or many other regions. The selected regions are then sequenced using a genome wide expression panel – this means that we know the expression of every gene at every region of interest.

Here is a brief schematic for how this works, if you are interested in the details you can take a look here: https://nanostring.com/products/geomx-digital-spatial-profiler/geomx-dsp-overview/



① Understand DSP & how its applied.

The dataset
Carpenter et al. identified PanIN, normal ductal, acinar, as well as poorly/glandular tumour regions and then applied DSP to each of these regions from several patients. Based on the gene expression profiles of these regions they defined signatures that can differentiate between these cells and used them to find these cell types in single cell sequencing data (look at figure 6E if you are interested, though the details don't matter for our purposes).

I have processed the data for you. In the tsv file, the first column is the gene name, the remaining columns are samples and contain the patient ID. For example, 56855-PanIN_Tumor is a PanIN tumour sample from patient number 56855. Each cell in the matrix corresponds to the expression of a gene in a sample.

gene name, sample, patient id

Your research question
How different are the PanIN, normal, acinar and tumour cell types to each other? Are there any batch effects or patient specific effects?

There are many ways you could answer this question, what you do is up to you. However, here is one suggestion for something you could do that will answer the question:

1. Run principal component analysis (PCA) over the entire matrix and visualize the first set of PCs
2. What happens when you colour this plot by cell type?
3. Which cell types are the most different from each other?
4. Where do normal ductal, PanIN, poorly and glandular cell types lie relative to each other? Given what I have told you about these cell types, does this match what you expected?
5. What next steps would you take to further understand this dataset?

Submitting your findings

Please send me the output of your analysis in an easily accessible format (word, google docs, pdf, jupyter notebook, html are all acceptable) as well as any code you created to answer these questions.

Carpenter, E. S., Elhossiny, A. M., Kadiyala, P., Li, J., McGue, J., Griffith, B. D., Zhang, Y., Edwards, J., Nelson, S., Lima, F., Donahue, K. L., Du, W., Bischoff, A. C., Alomari, D., Watkoske, H. R., Mattea, M., The, S., Espinoza, C. E., Barrett, M., … Pasca di Magliano, M. (2023). Analysis of Donor Pancreata Defines the Transcriptomic Signature and Microenvironment of Early Neoplastic Lesions. *Cancer Discovery*, *13*(6), 1324–1345.