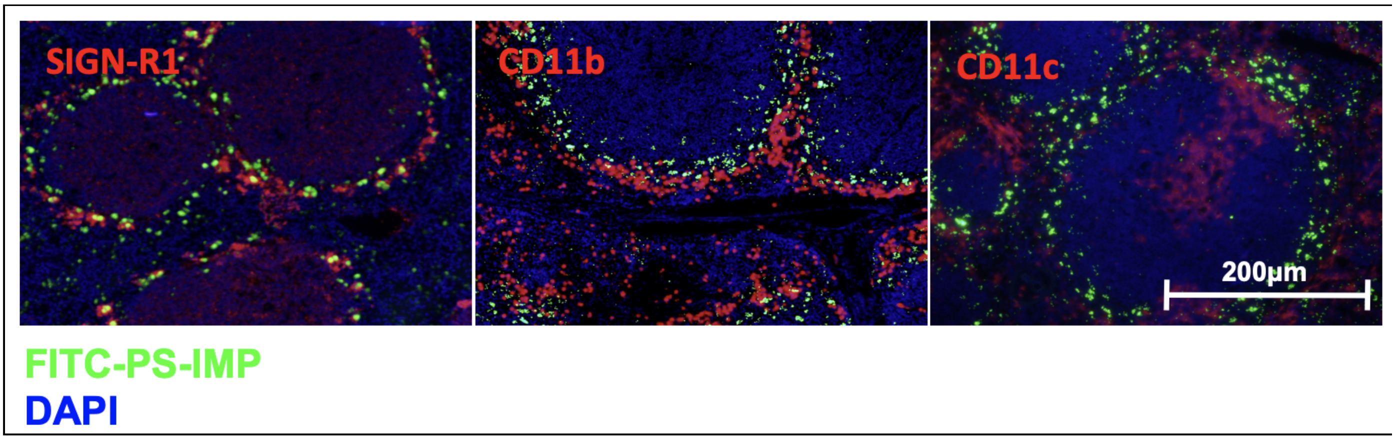
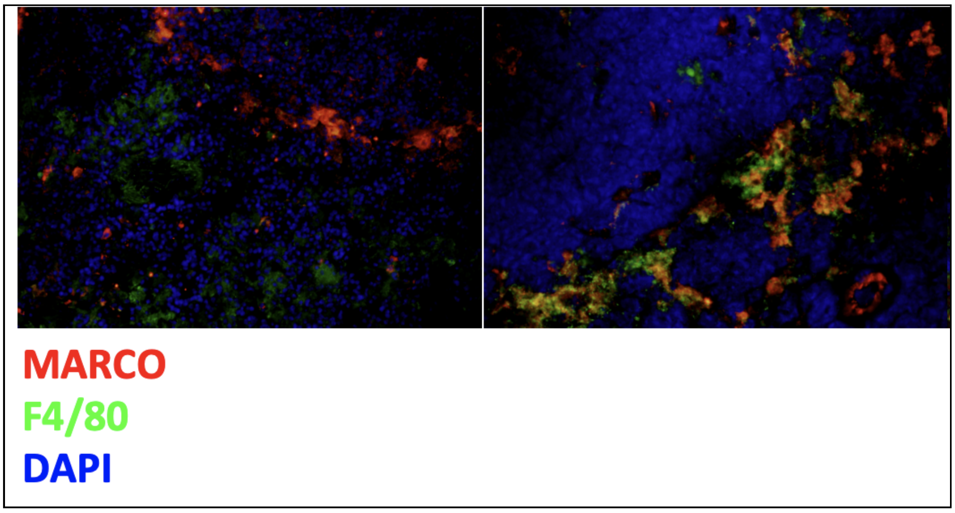
Video: overview of spatial analysis: part of 'Integration, exploration, and analysis of high-dimensional single-cell cytometry data using Spectre' - Oz Single Cell 2020, Computational biology.
PDF: overview of segmentation and spatial analysis steps.

Introduction

Traditional image/microscopy analysis

Traditionally, microscopy analysis is fairly qualitative – i.e. we are looking the picture, and making general statements about the kinds of things we can see. When it's quantitative, it's really about the amount of a certain marker being present in a certain area, or the 'co-localisation' of markers in specific spots on a cell/tissue. When 'single cells' are sought to be analysed, this is often in the form of identifying the 'position' of each cell, and where it is relative to an area etc, rather than looking at all of the markers expressed on that cell. We also split our panels into sets of 3-4 markers.



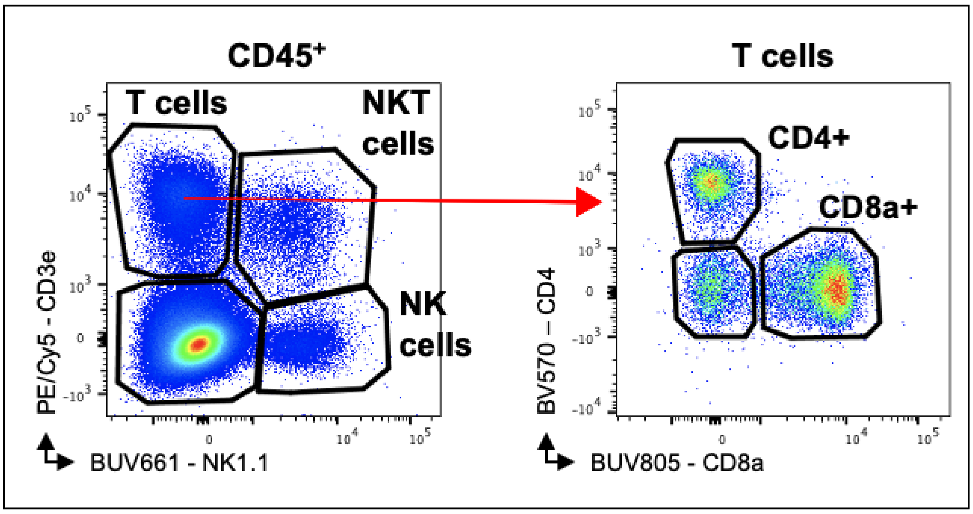


Traditional cytometry analysis

Cytometry data, when it comes down to it, is pretty simple. On a table, each row is a cell, each column is a marker, and the values represent the amount of that marker on each cell. So here we really deal with 'poplulations' of cells – how many there are, how much of different markers they express. Obviously in this data, we have no spatial information.

	Level of 'CD3' protein	**Level of 'CD4' protein**	**Level of 'CD8' protein**	**Level of 'NK1.1' protein**	**Summary of positive markers**	**What immune population is this?**
Cell #1	100,000	850,094	534	346	**CD3+CD4+**	*CD4+ T cell*

	Level of 'CD3' protein	Level of 'CD4' protein	Level of 'CD8' protein	Level of 'NK1.1' protein	Summary of positive markers	What immune population is this?
Cell #2	900,000	1424	991,242	128	**CD3+CD8+**	*CD8+ T cell*
Cell #3	860,523	849	420	242	**CD3+**	*Double negative T cell*
Cell #4	872,049	125	235	952,284	**CD3+NK1.1+**	*NKT cell*
Cell #5	457	157	312	892,401	**NK1.1+**	*NK cell*



Bringing the two together: high-dimensional tissue cytometry/imaging technologies

The promise of 'high-dimensional' imaging technologies is that we can now analyse single cell phenotypes, and incorporate spatial data into that process. **Imaging Mass Cytometry (IMC)** is one of these technologies that allows us to measure ~40 markers on a tissue section. There are other competing technologies as well.

Image analysis to cellular analysis via cell segmentation

Bringing 'cells' and 'space' together through cell segmentation: what it would ideally look like

The overall idea, is that we can use all this marker expression information to identify cells, segment them (i.e. give them a boundary), analyse what populations are there (how may cells, what they express), and then link those populations to the spatial information (e.g. whether a certain population of cell is enriched in certain areas etc). This can lead us to predictive metrics: e.g. let's consider a tumour – the presence of activated immune cells in the tumour tissue likely suggests that the immune response is doing OK at fighting the cancer. If in another patient, these cells all exhibit a 'regulatory' phenotype, it potentially indicates that the tumour is suppressing the function of the immune cells, and will start growing out of control. As such – if we know how these things correlate with patient outcomes – we can now use high-dimensional imaging to predict response to disease by incorporating both cellular and spatial data, which could probably not been predicted from cytometry data or spatial data alone.

One tool that does this is 'HistoCat' (https://www.nature.com/articles/nmeth.4391) – with the analysis process summarised here:

Figure 1: Multiscale analysis of the tissue ecosystem.

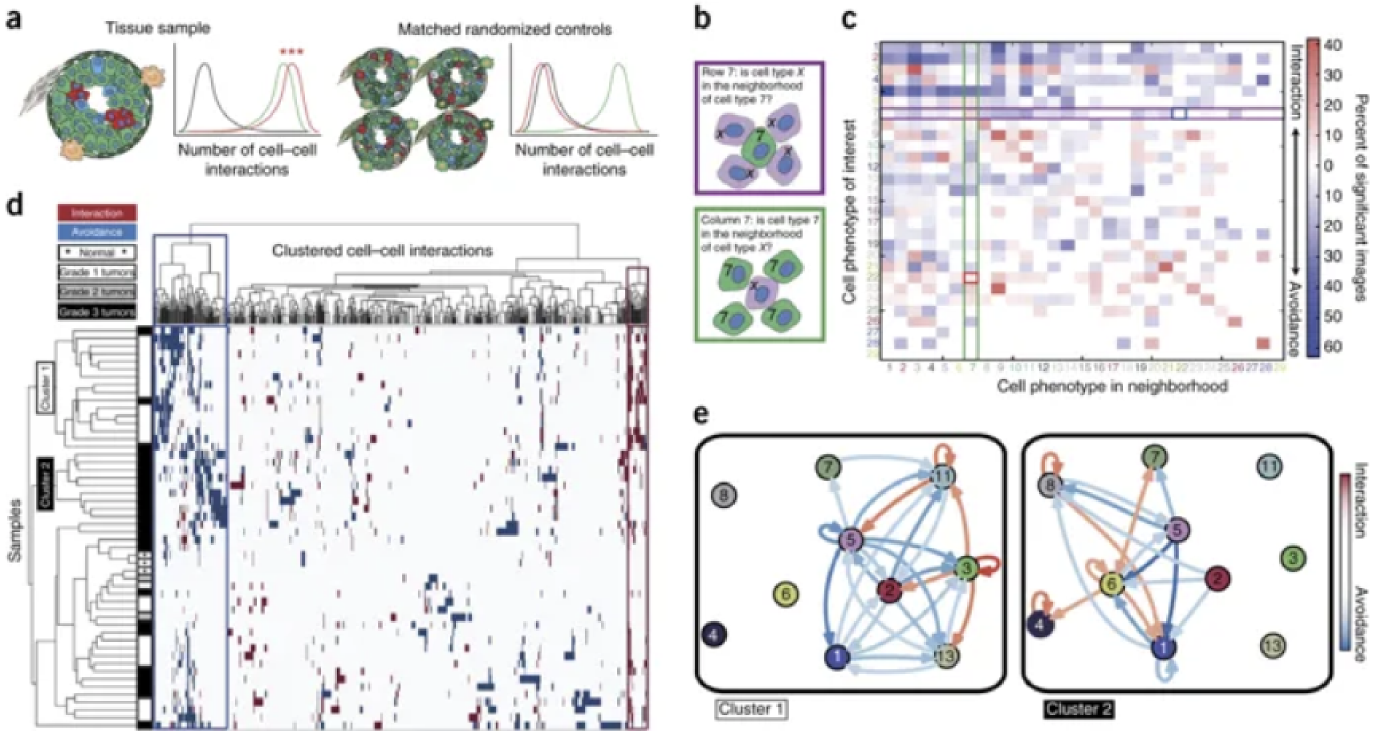
*(**a**) Visualization of images, (**b**) cytometry analysis, and (**c**) analysis of neighbors and cellular interaction networks facilitate 'round-trip' analysis through layers of information. (**d**) Experimental cohorts can be compared and contrasted using molecular, cellular, and spatial signatures.*

Specific challenges: how do we do spatial analysis?

One the challenges here is about how do we do 'spatial' analysis. One option is something like regional occupancy – we can train a classifier to see 'tumour' or 'normal' tissue etc, and then ask which cells are located where. Alternatively, we could draw these regions manually. Another approach is to do this more mathematically, like working out the average distances between certain cell types and specific tissue areas or features.

Yet another approach is 'neighbourhood' analysis, which is used in HistoCat. See the following figure:

Figure 3: Neighborhood analysis of breast cancer cell phenotypes.



*(**a**) Schematic of neighbor analysis. Number of interactions between abundant green cells (green line), between rare clustered red cells (red line), and between abundant green cells and rare red cells (black line). (**b**) Schematic depicting directional aspects of neighbor interactions visualized in the heatmap. Rows visualize the significance of all cell types surrounding a cell type of interest. Columns visualize the significance of the cell type of interest surrounding other cell types. White represents an interaction prevalence of less than 10%. (**c**) All interactions present in 49 breast tumor images and three matched normal tissue images are represented as a heatmap in which the cell type in the row is significantly neighbored (red) or avoided (blue) by the cell type in the column. Significance was determined by permutation test (P < 0.01). Highlighted squares indicate an example of a directional interaction. (**d**) Agglomerative clustering of all samples and cell–cell interactions according to the presence of significant (P < 0.01) phenotype interaction (red) or avoidance (blue). White represents interactions that are not present or not significant. (**e**) Cell social interaction network graphs representing the interactions of PhenoGraph-defined cell phenotypes in cluster 1 and cluster 2 tumors. Circle color corresponds to PhenoGraph cluster. Red arrows indicate interaction and blue arrows avoidance, and intensities of the line color indicate significance.*

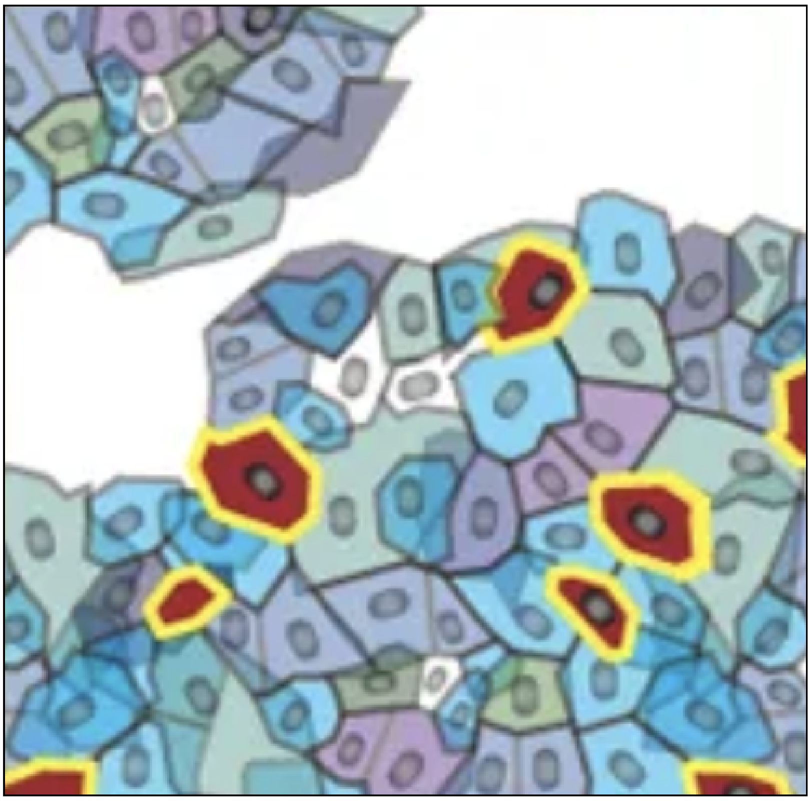
The idea here is to identify populations of cells (using clustering in this case), then generate neighbourhoods that are comprised of combinations of these cells. Then you can look at the likelihood of a specific cell type appearing in a certain neighbourhood, and whether this correlates with something like patient outcomes. Unfortunately, the tricky thing here, is it's super hard to validate the 'neighbourhoods' – 'where are they?', 'are they just noise?' are questions one might ask in validation, but finding answers (at least with current tools) is extremely difficult. Therefore, while mathematically cool, is a hard tool to use because it is difficult to validate.

The additional challenge that is relevant here is that sample to sample (image to image) variation is huge, much bigger than we see in 'normal cytometry' data. If segmentation in different images is done slightly differently, the more or less of the markers get incorporated into that cell, and the corresponding average or integrated levels per 'cell' can be quite different.

Challenges in cell segmentation

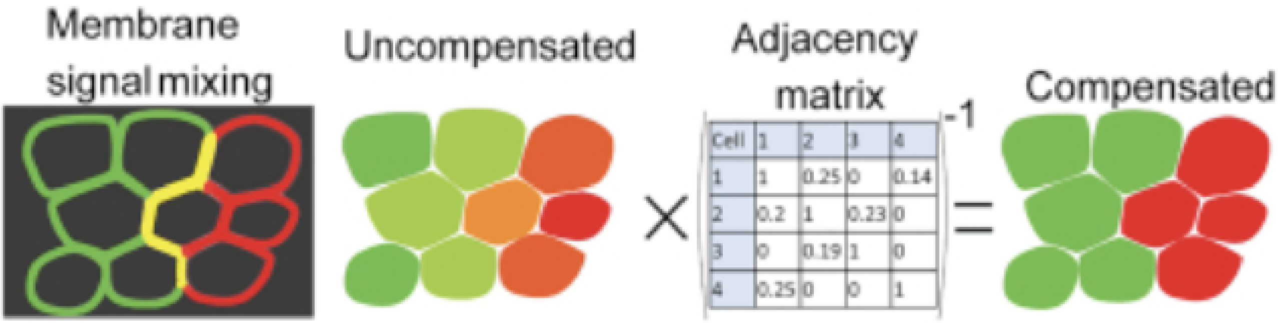
Specific challenges: each 'cell' contains an overlap of nearby cells – why 'cells' aren't the same as 'cells' in cytometry data – and what this means for analysis

Probably the biggest challenge that relates to this whole process, and in truth potentially changes the entire way we have to think about single cells in this context, is cell segmentation. In it's simplest form, segmentation is pretty straightforward – on a 2D image, cells sit next to each other, so you can draw a boundary around each one, and whatever signal appears within that boundary is assigned to that cell. However, the reality is not that simple. Many cell borders are essentially squished together, so you can actually distinguish them. Moreover, in some cases, cells do in fact 'overlap':



This is because the 'thickness' of a tissue section is often between 5 - 7 microns (about half the width of a cell). So in the same tissue section you can have two cells that are mixed together when they are imaged. As a result, each cell, and the signal within that cell border is not actually just that cell, so the expression pattern on the cell is essentially a profile of that cell, and the ones around it.

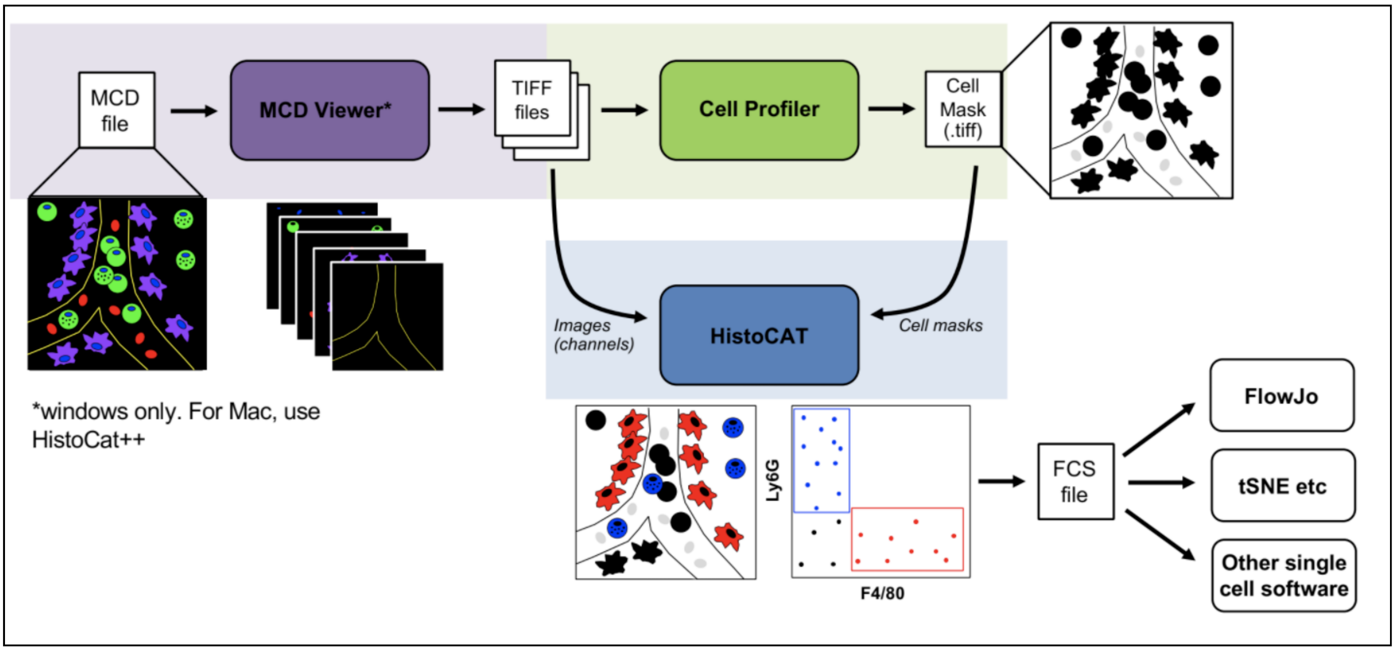
One approach here is to do what the guys who developed CODEX (an IMC competitor) did – 'compensate' the signals spatially (https://www.ncbi.nlm.nih.gov/pubmed/30078711). Essentially this looks at the expression profile of a cell, and looks at the profile of neighbouring cells. A 'spillover matrix' (called an adjacency matrix) is then created, whether the signal coefficients can be used to correct the signal (see image). Unfortunately there are limitations here, as I think in their version cells must be circular, but the idea is solid.

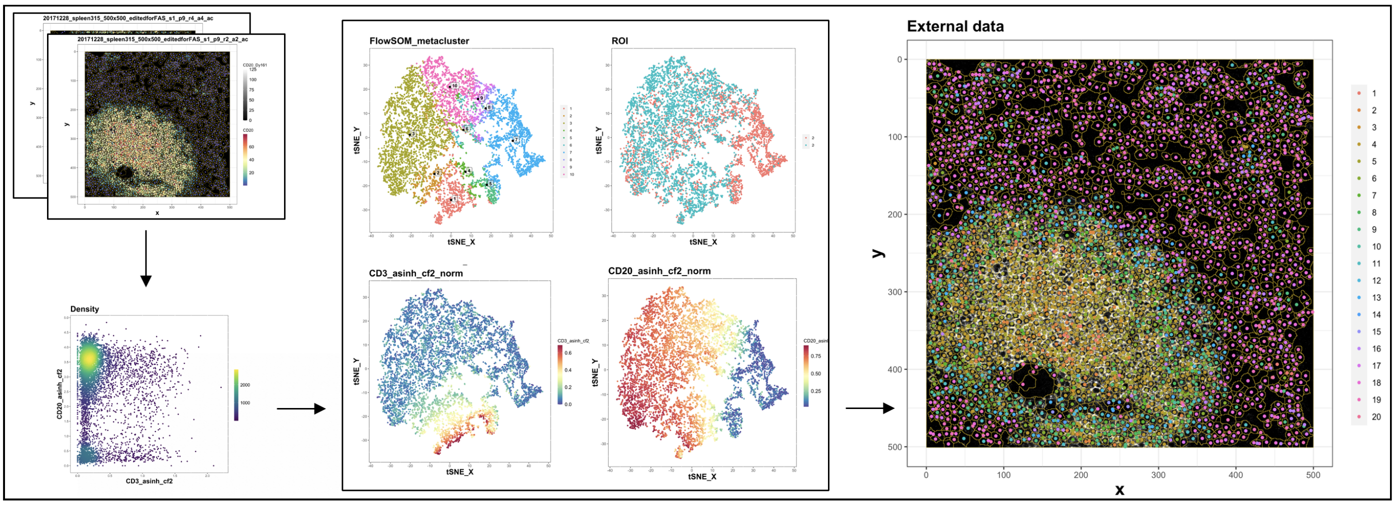


Cell	1	2	3	4
1	1	0.25	0	0.14
2	0.2	1	0.23	0
3	0	0.19	1	0
4	0.25	0	0	1

Alternatively, we can use a classifier to identify the types of cells before we even get to the proper analysis, and then forget about their actual expression patterns. E.g. if I have segmentation of all my cells, and then I train a classifier on the, say, 10 different subsets I was looking for, classifiers could generally pull them apart. However, this means I have to decide on exactly what cells I am looking for ahead of time, and I would also have to discard the rest of the expression data on those cells (markers that indicate inflammatory status, activation, etc). This would be kind of like what the Seurat package does (https://satijalab.org/seurat/v3.1/spatial_vignette.html).

How does segmentation and spatial analysis work


MCD file
MCD Viewer*
TIFF files
Cell Profiler
Cell Mask (.tiff)
Images (channels)
HistoCAT
Cell masks
*windows only. For Mac, use HistoCat++
Ly6G
F4/80
FCS file
FlowJo
tSNE etc
Other single cell software


20171228_spleen315_500x500_editedforFAS_s1_p9_r4_a4_ac
20171228_spleen315_500x500_editedforFAS_s1_p9_r2_a2_ac
x
y
Density
CD3_asinh_cf2
CD20_asinh_cf2
FlowSOM_metacluster
ROI
CD3_asinh_cf2_norm
CD20_asinh_cf2_norm
tSNE_X
tSNE_Y
External data
x
y