John Lin
COMP790-166
2022-04-18
Reading Question #2
Paper: *In situ* functional cell phenotyping reveals microdomain networks in colorectal cancer recurrence

1. Furman et al. present LEAPH ("learning algorithm for identifying cell phenotypes"), which is an unsupervised machine learning approach to describe cellular functional phenotypes (FP) using immunofluorescence imaging (in this paper, this is characterized as hyperplexed imaging datasets, ie visualization and identification of various biomarkers). By more finely capturing various cell phenotypes in conjunction with spatial information, the authors propose this approach is better suited to characterizing microdomains, which can reveal more insights underlying the mechanisms governing cancer growth/suppression.

2. The main contributions of these authors were:

- Development of LEAPH. As alluded to above, the algorithm uses "probabilistic clustering and spatial regularization" to identify FPs along a "continuum of phenotypes", which contrasts against more traditional hard clustering and discretization of cell states
- Discovery of two microdomains (specific to tumor-promoting/suppression colorectal cancer (CRC) FPs) based on pointwise mutual information (PMI) from clinical data and immunofluorescence imaging
- Characterization of network biology ie analysis of correlations between biomarker pairs specific to each microdomain
- Demonstration of using LEAPH to identify an "optimal" set of biomarkers necessary to capture phenotypic diversity via virtual simulation and iterative, non-destructive imaging

3. Furman et al. go through the methodology of LEAPH, showing that LEAPH creates a tree hierarchy of FPs. Cells are found to have either specialized (single FP), transitional (two FPs), or multi-transitional (more than two FPs) cell states. They elaborate that the "recursive probabilistic clustering" is based on mixture factor analyzers (MFA) and "spatial regularization" helps to filter non-specialized cells that are false positives. Using LEAPH experimentally with a CRC tissue sample dataset (n=213, ~500,000 cells), the authors found 13 FPs. Interestingly, they show a small fraction of cells were non-specialized and shared between FP2 and FP4. Additionally, there was a larger population of these cells in the NED-8 year (no evidence of disease after 8 years) cohort versus the REC-3 year (recurrence of disease after 3 years) cohort. Upon further analysis of biomarkers, the authors propose that FP4 may be associated with protumorigenic cells. Also, the authors looked at the distribution of PMI values amongst different FP pairs to determine significant differences in spatial co-occurrence between the NED- 8 year and REC- 3 year cohorts. They found eight FPs with PMI distributions higher in the REC-3 year cohort and 1 FP with PMI distribution higher in the NED-8 year cohort. Analysis of these FPs resulted in the characterization of two microdomains alluded to earlier.

4. There is some discussion of previous work, but no direct comparison to other baselines. The novelty of LEAPH seems to lie in its use of spatial information to characterize FPs. Some other unique features of LEAPH include its use of probabilistic (versus hard) clustering. Where pciSeq and Harmony also use probabilistic clustering, the authors explain their use cases are different. pciSeq tries to assign reads to cells and then cells to cell types, whereas Harmony tries to estimate cell types amongst various datasets. Additionally, LEAPH is an unsupervised method. I don't believe we can say the evaluation was objective per se, since there was no direct comparison to other methods. However, LEAPH seems to reveal interesting results from the analysis of the CRC dataset.

5. There are interesting results in the context of their CRC findings. However, as mentioned above, there was no direct comparison to other methodologies. The authors analyze how FP identity accuracy changes with different subsets of biomarkers. Per Figure S4a, to define ground truth labels, they use the assignment associated with maximum FP ownership probabilities. They find that just two biomarkers are needed to achieve more than 97% accuracy in dissecting epithelial-stromal cells. However, subtyping further requires more biomarkers as expected. Additionally, per Figure S4c, they compare NS-LEAPH (non-spatial) vs LEAPH to determine the effect of spatial regularization. While 51% of cells have a specialized cell state from both methods, analysis of the remaining cells shows that spatial information decreases entropy distributions. This implies that spatial information helps define cell states. Given this, the authors try to evaluate various novel aspects of LEAPH using their own methods. However, it would be interesting to see how similar pre-existing methods compare against the characterization of microdomains as they did with the CRC dataset. Additionally, further experimental validation of their CRC results would help solidify these findings.

6. Furman et al. demonstrate the generalizability by applying LEAPH to a breast cancer dataset. However, it seems that LEAPH relies on spatial information specific to immunofluorescent images of solid tumors. One follow up idea would be to see how this could expand this to blood cancers. The author acknowledge that it may be possible to extend LEAPH to single cell transcriptomics ie scRNA-seq, using a non-spatial version of LEAPH. However, are there other physical properties of blood cancers that could be incorporated into LEAPH? Perhaps there are different fluid mechanics between different types of cells, which might also help inform FPs derived from LEAPH? Investigating other features of different cancer types may increase the generalizability of LEAPH.