Data-Guided Modeling of Cancer and Tumor Heterogeneity

SIVASOMASUNDARI (SUNDARI) ARUNARASU, *EMORY UNIVERSITY*

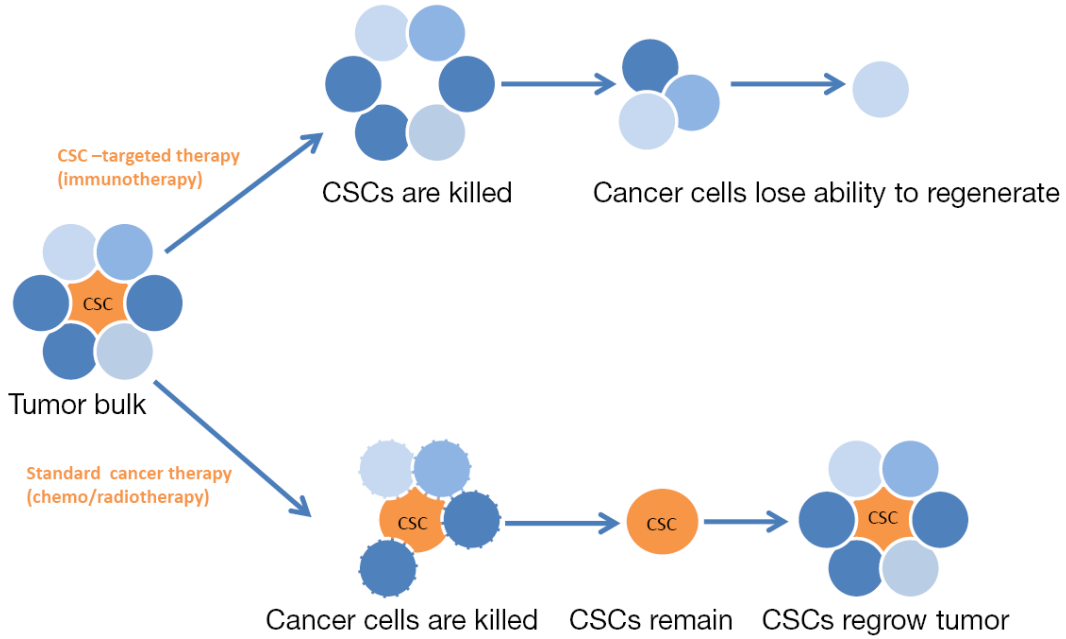
DR. SUBHAJYOTI DE, *RUTGERS CANCER INSTITUTE OF NEW JERSEY*

Background

- Intra-tumor heterogeneity: variation among tumor cells
  - Sources: genetic (mutations), nongenetic (epigenetic and transcriptomic changes), microenvironment
- Cancer stem cell (CSC) model: intra-tumor heterogeneity at the nongenetic level
  - Certain cells are characteristic of regular stem cells: have the ability to self-renew OR differentiate into tumor cells
  - Thought that CSCs originate from either:
    - Regular tissue cells that become mutated and turn into tumor cells
    - Tumor cells that become dedifferentiated, leading to stem cell characteristics
  - Stemness: thought to be a dynamic characteristic of tumor cells

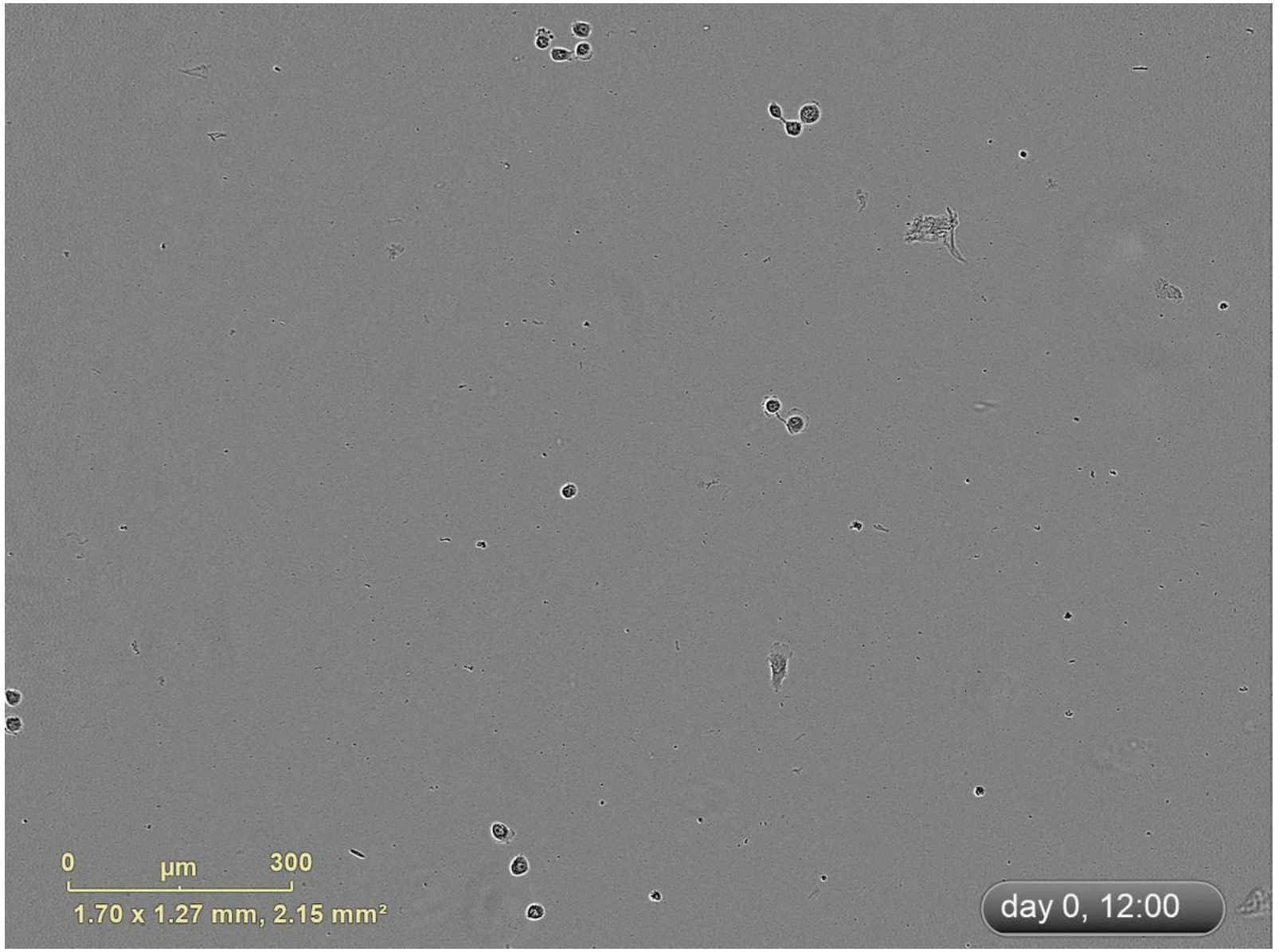
Motivation

- Better understanding of CSC model: help create more effective and targeted cancer treatment strategies
  - If CSCs can be identified by genetic and/or phenotypic characteristics, treatments can target these cells and prevent regrowth



Motivation

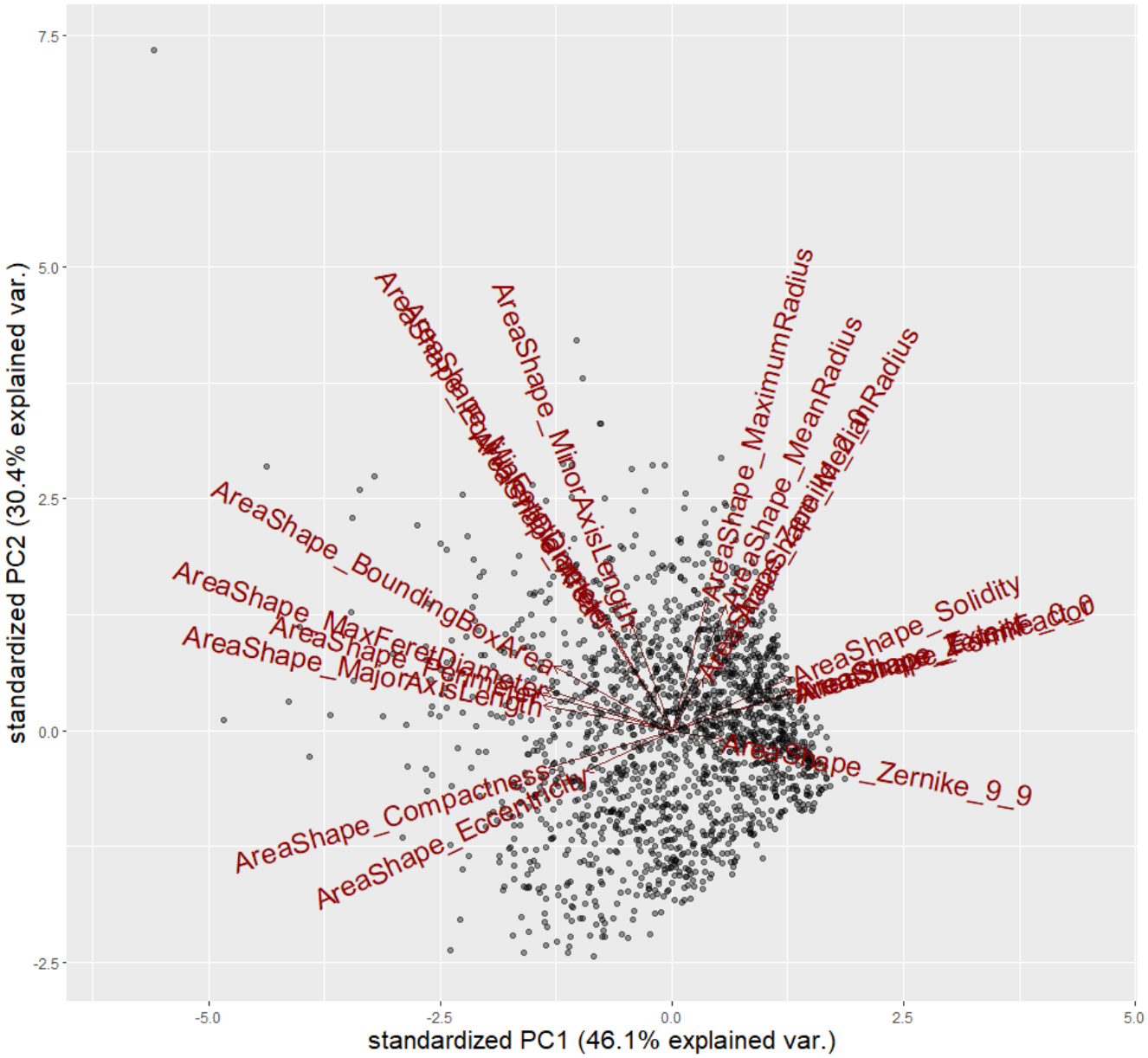
- Goals: to examine the phenotypes of cells and colonies during tumor growth
  - To understand characteristics of stem cells VS regular tumor cells
  - Identify features to classify cells into these two states
- Biological classification of tumor cells:
  - Holoclones: more stem-like
    - Divide at a faster rate (maintains more circular shape)
    - More compact colonies
  - Meroclones: more differentiated
    - More irregularly shaped cells and colonies



Growth of tumor cells from T24 cell line was imaged at 2-hour intervals over several days

Principal Component Analysis (PCA)

- Images were then analyzed using CellProfiler: free open-source software used to quantitatively measure phenotypes from many images of cells
  - Large number of variables (83) were generated by CellProfiler
  - Many features redundant and/or highly correlated (area, perimeter, diameter, min/max radius)
- PCA: method for reducing dimensionality of large data sets
  - Used to determine the most significant variables from the CellProfiler data (which features contributed most to the first/second principal components)
- Five images spaced over course of tumor growth were used for PCA (performed in R)

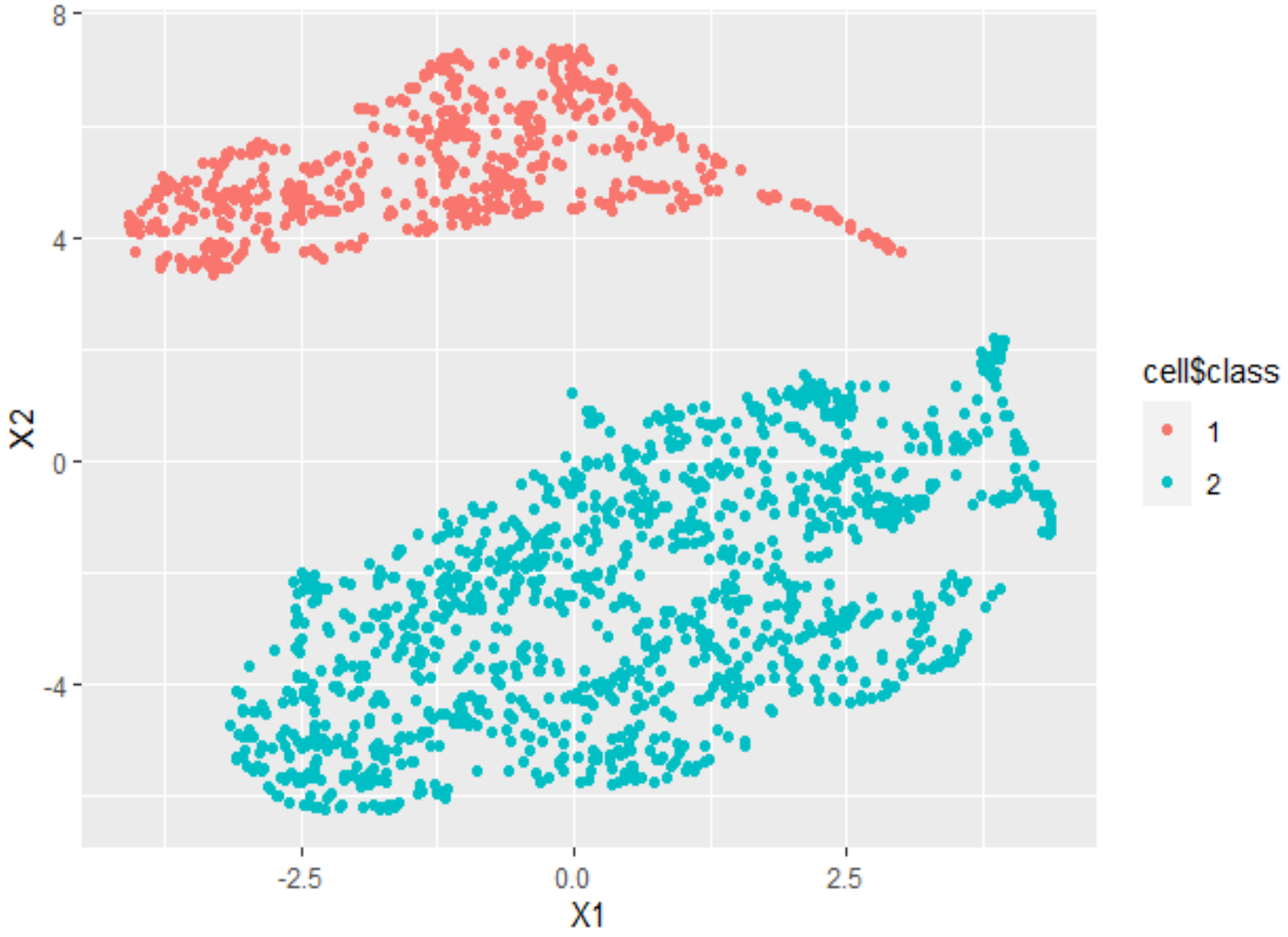


Variables most positively and negatively correlated with first principal component (PC1):

Perimeter	MedianRadius
-0.31281924	0.15771155
MajorAxisLength	Solidity
-0.30606363	0.28742077
MaxFeretDiameter	Zernike_0_0
-0.30123261	0.29023546
Compactness	Extent
-0.29473489	0.29390289
BoundingBoxArea	FormFactor
-0.28178664	0.30619106

UMAP analysis

- UMAP (Uniform Manifold Approximation and Projection): method for dimensional reduction, non-parametric clustering scheme
- Used to separate data into groups based on similarities in certain variables
- Analysis conducted including the variables found to be most important from PCA
  - Excluded all granularity variables: associated strongly with image number (number of cells/objects in an image)
- UMAP first applied on the 5 sample images, then on each set (3) of image data corresponding to a quarter of a well in 24-well plate



UMAP results of 5 images over tumor growth

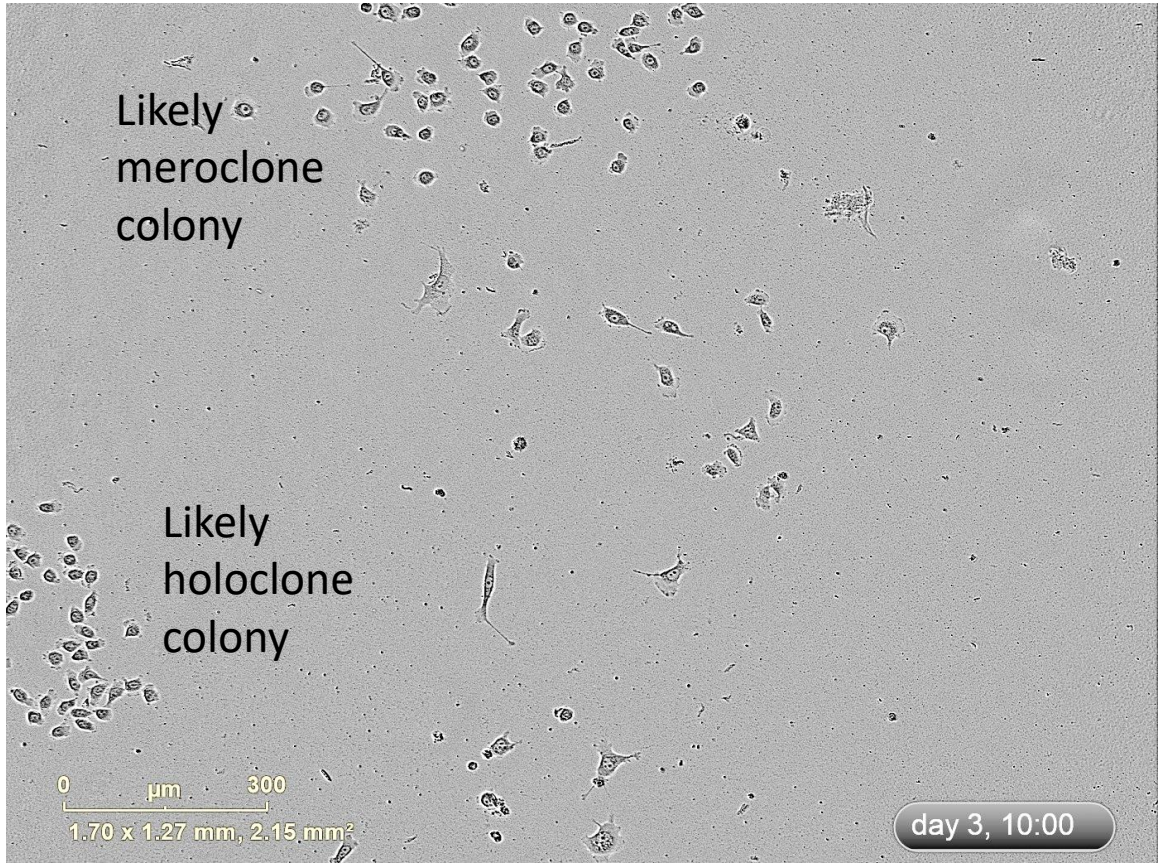
Correlation of X2 and variables used in UMAP

-0.06903	Area
-0.07567	MinFeretDiameter
-0.08491	EquivalentDiameter
-0.20019	MinorAxisLength
-0.26611	Zernike_2_0
-0.28286	Zernike_9_9
0.316292	BoundingBoxArea
0.454833	MaxFeretDiameter
0.515854	MajorAxisLength
0.517108	Perimeter
0.640168	Eccentricity
-0.68815	MaximumRadius
0.752384	Compactness
-0.77909	MeanRadius
-0.79557	Extent
-0.81502	Zernike_0_0
-0.8336	Solidity
-0.84428	MedianRadius
-0.86534	FormFactor

Tumor Cell Classification

- Classifying cells into groups based on UMAP:
  - FormFactor: 1 for a perfectly circular object ($4*\pi*Area/Perimeter^2$)
  - MedianRadius
  - Solidity: proportion of pixels in the convex hull also in the object itself
  - Zernike_0_0: measure of shape, correlated with other variables
  - Extent: proportion of pixels in the bounding box located in the actual object, equals 1 for more circular objects
- Class 1: more meroclone-like
- Class 2: more holoclone-like, higher “stemness”
  - Positively related to measures above (higher FormFactor, Extent, etc.)

UMAP VS Manual Analysis



Cell State Transition Rate

- Split images into different number of sections (grids), assign cells
  - Exclude first and last 20 images from each set
  - Exclude overcrowded or sparse grids
- Each cell at t=2 comes from either:
  - Sustained (directly from t=1)
  - Homologous growth (from same class)
  - Heterologous growth (other cell type)
- Columns (C1_, C2_): source (t=1)
- Rows (_C1, _C2): destination (t=2)

Grids (#)		_C1	_C2
2	C1_	0.3469388	-1.693878
	C2_	0.5748299	2.683673
3	C1_	0.2033898	0.1694915
	C2_	0.5462842	1.2372881
4	C1_	1.0004606	0.7148779
	C2_	-0.1303547	0.6895440
5	C1_	0.5373449	0.343011
	C2_	0.2013580	0.921564

Cell state transition rates based on different grid sizes for image set C1

Cell State Transition Rate

- Class 1 cells (source) do not show a strong trend of transition rates
- Class 2 cells (source) generally much more likely to become (or maintain as) C2 cells than transition to C1 cells
  - Seems to make biological sense since C2 cells (more holoclone or stemlike) can self-renew (maintain C2 state) or differentiate into C1 cells
  - For a tumor to keep growing, however, holoclones should be more likely to self-renew, in order to create more CSCs and ensure tumor survival

Grids (#)		_C1	_C2
2	C1_	1.0	0.000000
	C2_	-0.2	1.266667
3	C1_	0.1524164	-0.2453532
	C2_	0.1449814	1.2788104
4	C1_	0.23785595	0.6566164
	C2_	0.05695142	1.0586265
5	C1_	0.56097561	0.4390244
	C2_	-0.02926829	1.0292683

Cell state transition rates based on different grid sizes for image set C3

Summary

- Overall, PCA and UMAP analysis in R was used to classify tumor cells into two different classes based on phenotypic characteristics
  - Most significant characteristics were related to cell shape and irregularity
  - Thought to correspond with more holoclone (stemlike) and meroclone cells
  - This classification can be used to analyze the colony makeup of other tumors
  - Might help determine the likelihood that a tumor will regenerate after treatment, based on proportion/density of each cell type
    - Could also guide possible treatments to target specifically holoclone type cells and prevent regrowth

References

Biswas A, De S. Drivers of dynamic intratumor heterogeneity and phenotypic plasticity. Am J Physiol *Cell Physiol* 320: C750-C760, 2021. doi:10.1152/ajpcell.00575.2020.

Kreso A, Dick JE. Evolution of the cancer stem cell model. *Cell Stem Cell 14*: 275–291, 2014. doi:10.1016/j.stem.2014.02.006.

Prasetyanti PR, Medema JP. Intra-Tumor Heterogeneity From a Cancer Stem Cell Perspective. *Mol Cancer* 16(1): 41, 2017. doi:10.1186/s12943-017-0600-4.

Sharma A, Merritt E, Hu X, Cruz A, Jiang C, Sarkodie H, et al. Non-genetic intra-tumor heterogeneity is a major predictor of phenotypic heterogeneity and ongoing evolutionary dynamics in lung tumors. *Cell Rep* 29(8): 2164–2174, 2019. doi:10.1016/j.celrep.2019.10.045.

Marusyk A, Almendro V, Polyak K. Intra-tumour heterogeneity: a looking glass for cancer? *Nat Rev Cancer* 12: 323–334, 2012. doi:10.1038/nrc3261.

Brock A, Chang H, Huang S. Non-genetic heterogeneity—a mutation-independent driving force for the somatic evolution of tumours. *Nat Rev Genet* 10: 336–342, 2009. doi:10.1038/nrg2556.

Chen C, Qi XJ, Cao YW, et al. Bladder Tumor Heterogeneity: The Impact on Clinical Treatment. *Urol Int* 95: 1-8, 2015. doi:10.1159/000370165.

Acknowledgements

Thank you very much to:

- NSF grant CCF-1852215
- DIMACS REU program
- My mentor: Dr. Subhajyoti De and Antara Biswas (Postdoc)