Using Machine Learning to Predict the Risk of Either Having an Aggressive Form of Prostate Cancer (PCa) or Lower Grade PCa/Benign Prostatic Hyperplasia (BPH) Based Upon the Flow Cytometry Immunophenotyping of Myeloid-derived Suppressor Cells (MDSCs) and Lymphocyte Cell Populations

George A. Dominguez[1], John Roop[1], Alexander Polo[1], Anthony Campisi[1], Dmitry I. Gabrilovich[2], and Amit Kumar[1]

[1]Anixa Biosciences, Inc., San Jose, CA; [2]The Wistar Institute, Philadelphia, PA

Flow Cytometry Based Immunophenotyping

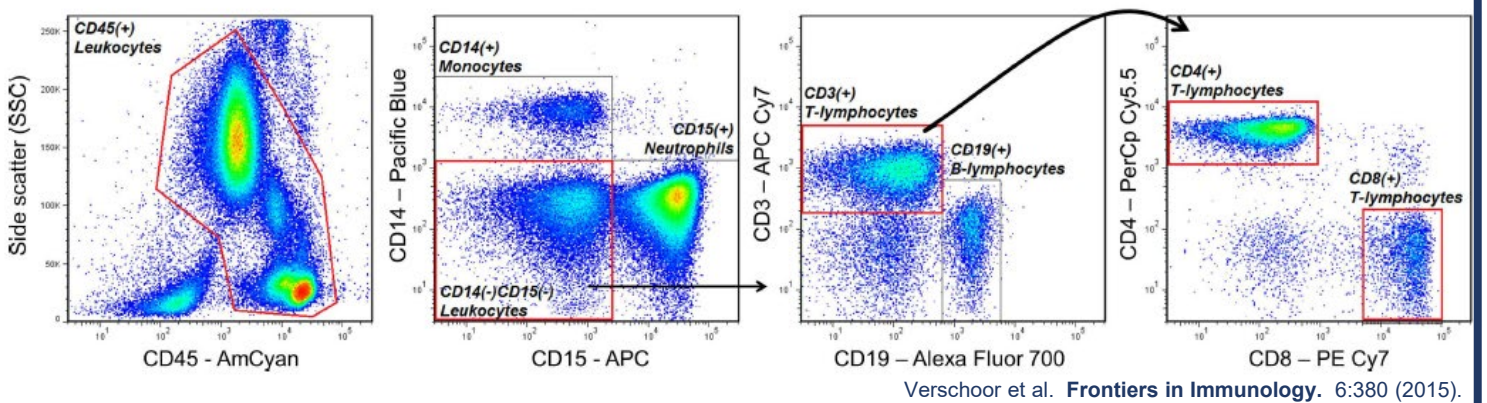
Flow Cytometry

- ✓ 50+ years in use
- ✓ A foundation of immunology and tumor immunology research

Immunophenotyping

- ✓ Know what you are looking for
- ✓ Manual gating = Boolean logic
- ✓ Operator experience
- ✓ More parameters (N) = more 2D unique relationships (R)
  - ✓ N = 2; R = 1
  - ✓ N = 3; R = 3
  - ✓ N = 4; R = 6
  - ✓ N = 5; R = 10
- ✓ Higher dimensions difficult for us to understand

- ✓ Real time analysis for thousands of cells per second
- ✓ Multiple cell markers can be used
- ✓ Cost Effective



Verschoor et al. **Frontiers in Immunology.** 6:380 (2015).

N = 7; R = 21

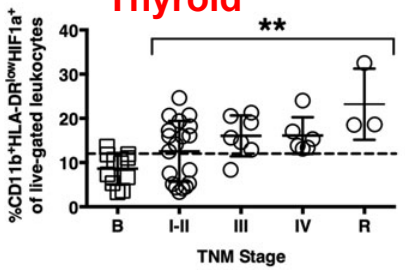
Identified 7 populations

A lot of information is discarded during analysis…

Myeloid-derived Suppressor Cells – What about them?

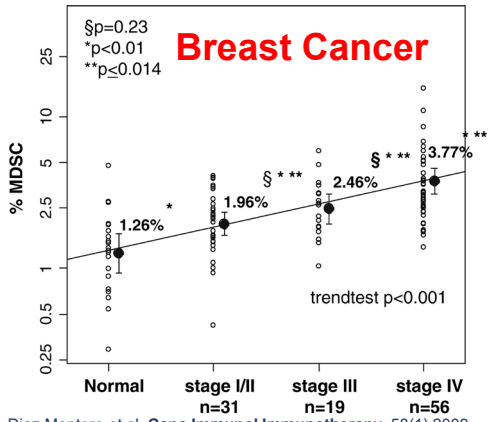
Indicative of Solid Tumors and Severity

Thyroid



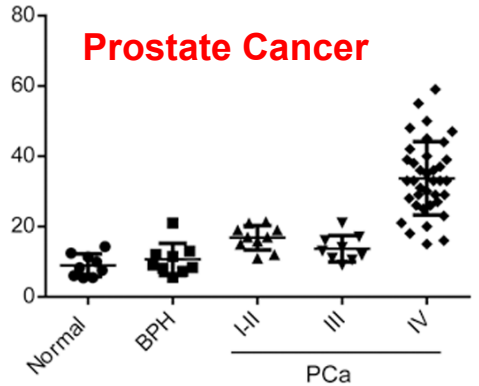
Angell et al. **Thyroid** 26(3) 2016.

Breast Cancer



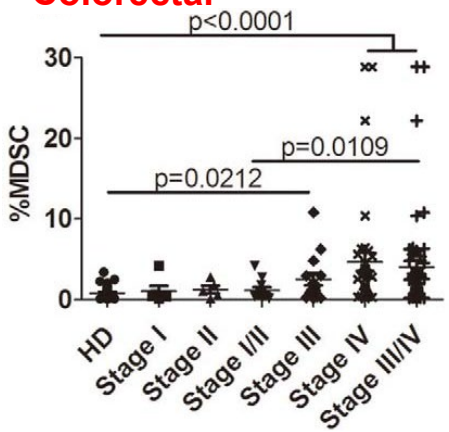
Diaz-Montero et al. **Canc Immunol Immunotherapy** 58(1) 2009.

Prostate Cancer



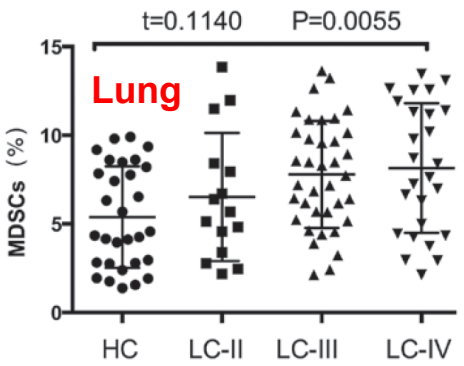
Chi et al. **Int J Clin Exp Med** 7(10) 2014.

Colorectal



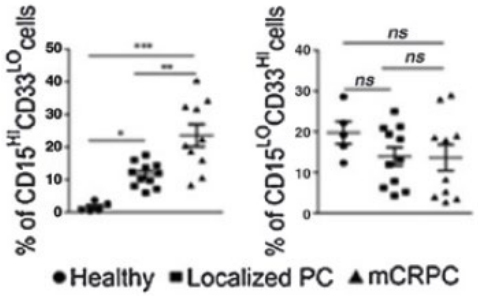
Zhang et al. **PLoS ONE** 8(2) 2013.

Lung



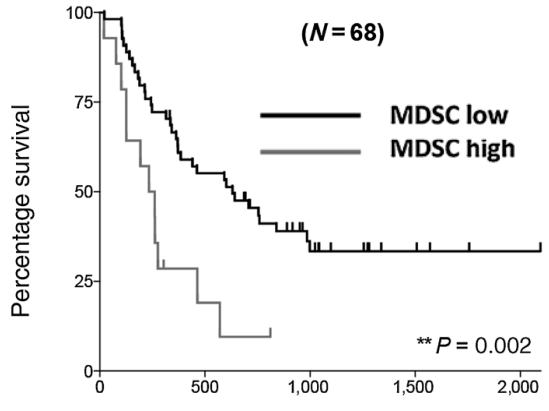
Chen et al. **Oncology Letters** 14 2017.

Prostate Cancer



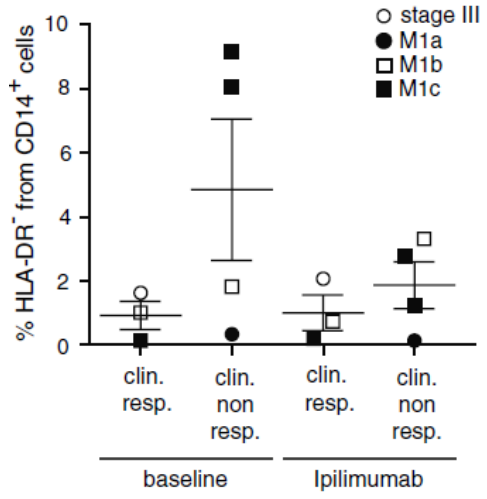
Hossain et al. **Clin Cancer Res** 21(16) 2015.

Predictors of Immunotherapy Response?



Kitano et al. **Cancer Immunol Res** 2(8) 2014.

CTLA-4 Responses in Late Stage Melanoma



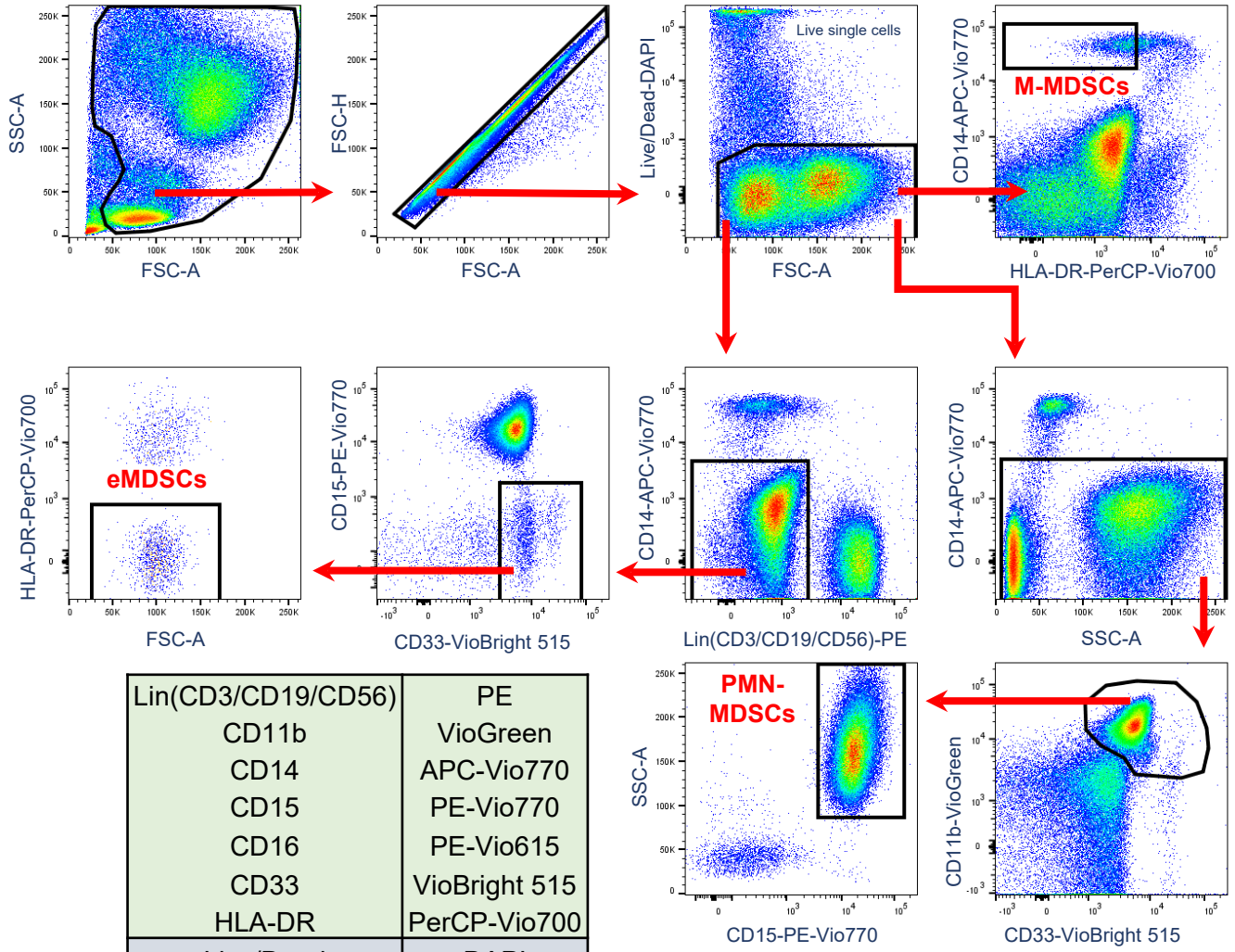
Meyer et al. **Cancer Immunol Immunother** 63:247-257 2014.

First Question

Can we use MDSCs and other leukocytes as a predictor for higher grade prostate cancer (PCa) and distinguish them from benign prostatic hyperplasia (BPH)/lower grade PCa?

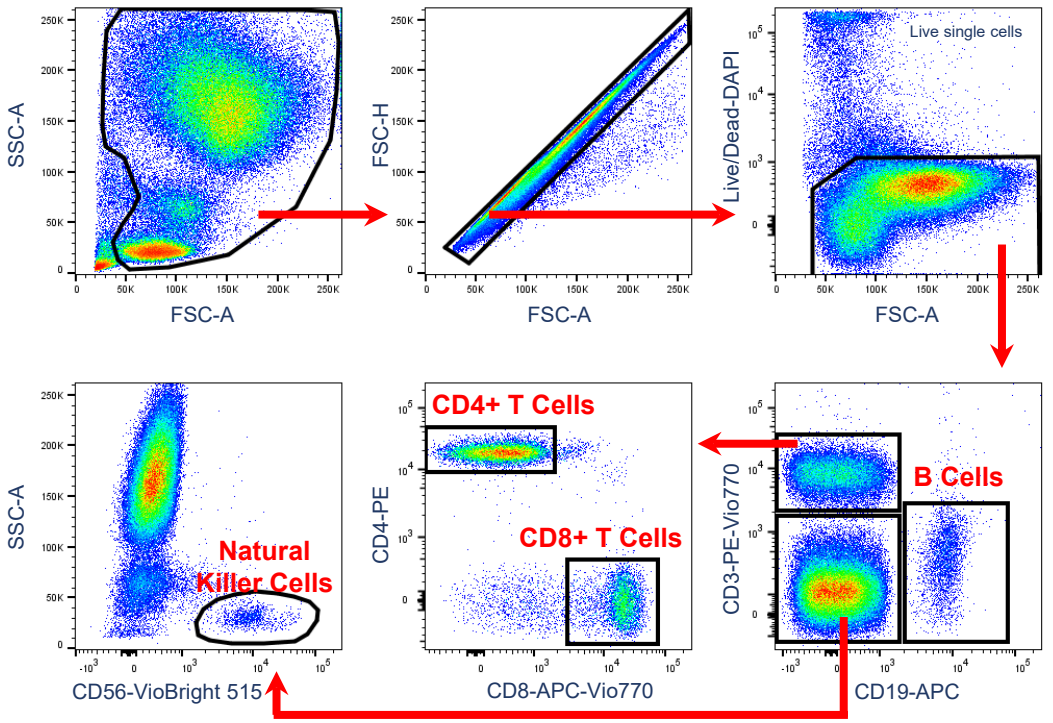
Myeloid and Lymphocyte Immunophenotyping

Myeloid Panel (10 Markers)



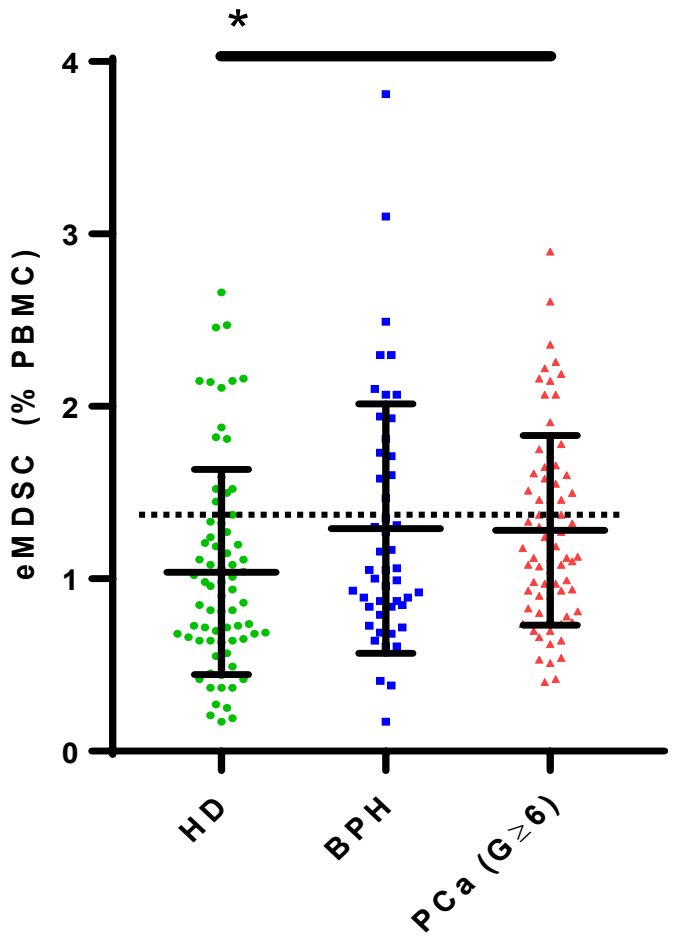
Lin(CD3/CD19/CD56)	PE
CD11b	VioGreen
CD14	APC-Vio770
CD15	PE-Vio770
CD16	PE-Vio615
CD33	VioBright 515
HLA-DR	PerCP-Vio700
Live/Dead	DAPI

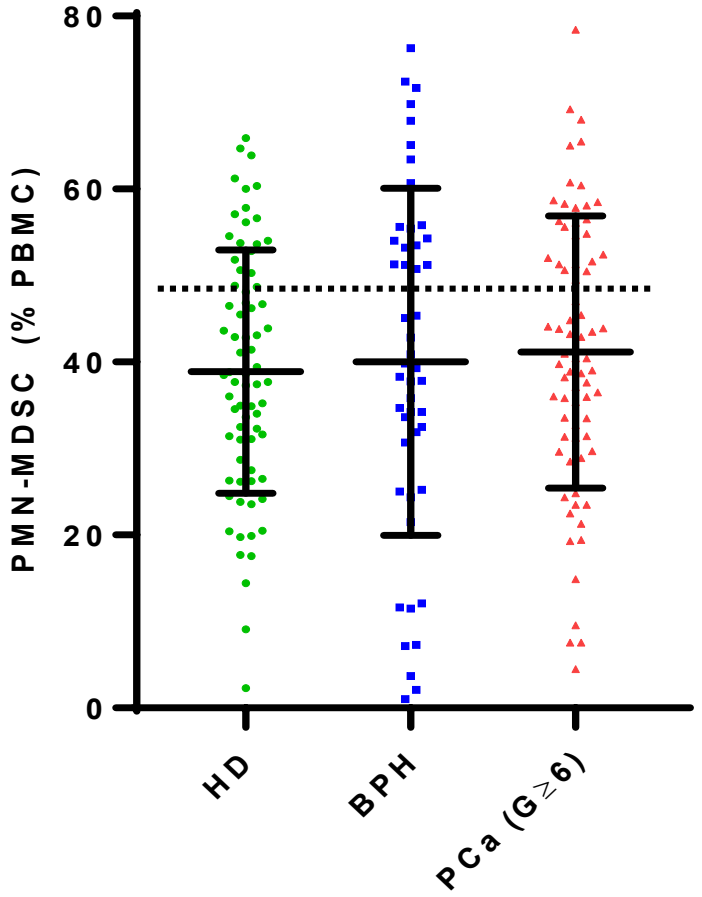
Lymphoid Panel (5 Markers)

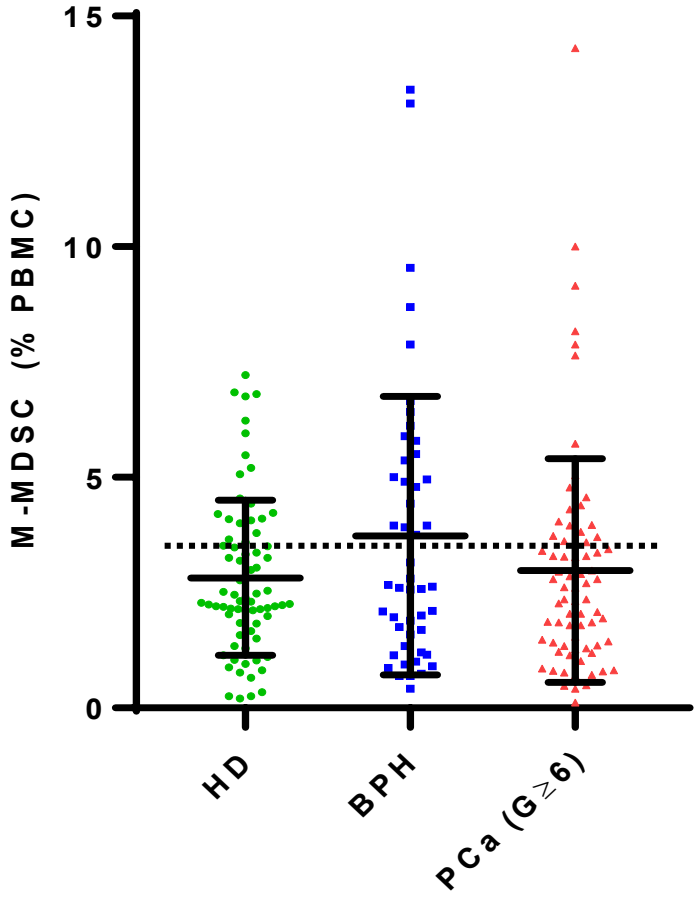


CD3	PE-Vio770
CD4	PE
CD8	APC-Vio770
CD19	APC
CD56	VioBright 515
Live/Dead	DAPI

Traditional Gating: Manual Counting - MDSCs







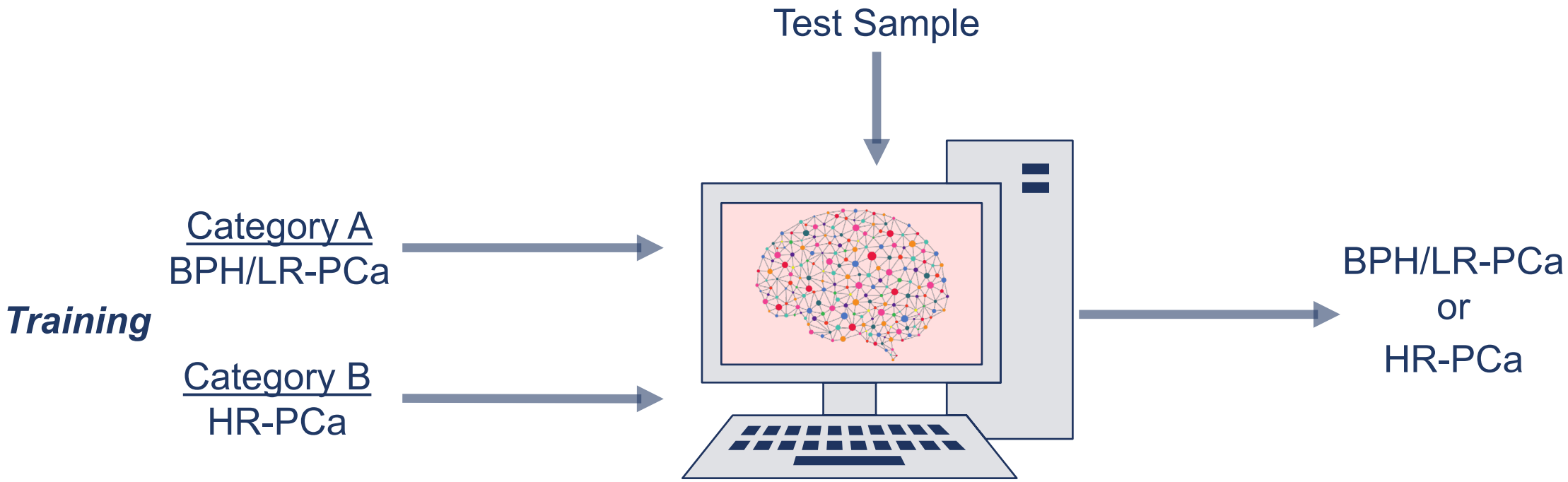
Simple cell counts can provide information about trends, but can only categorize some subjects

*The dotted line represents the 75th percentile for HD values (75% specificity)

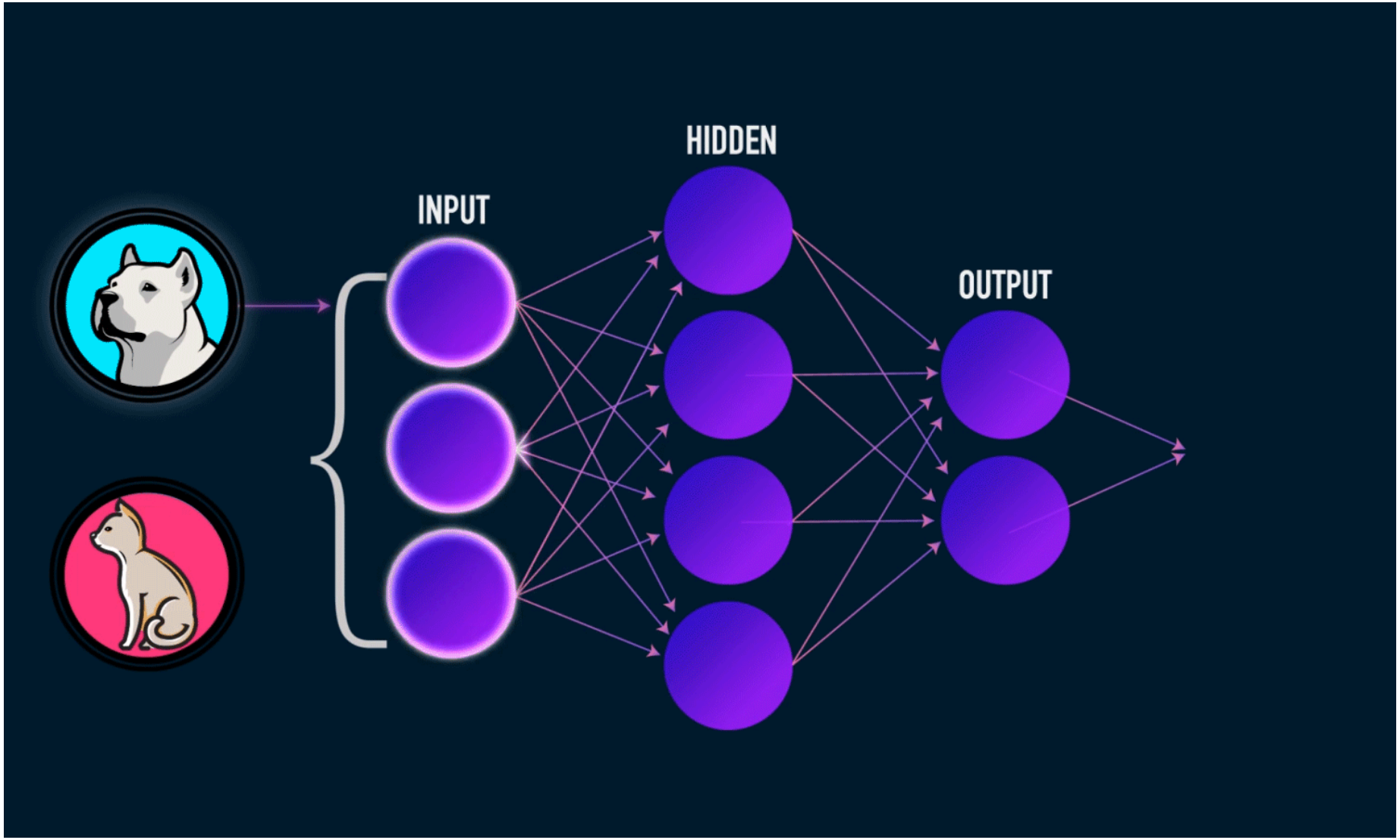
Healthy Donor (HD) n = 73
Benign Prostatic Hyperplasia (BPH) n = 48
Prostate Cancer (PCa) n = 73

Next Question

Can we use machine learning (neural networks) to analyze the flow cytometry data to categorize patients?

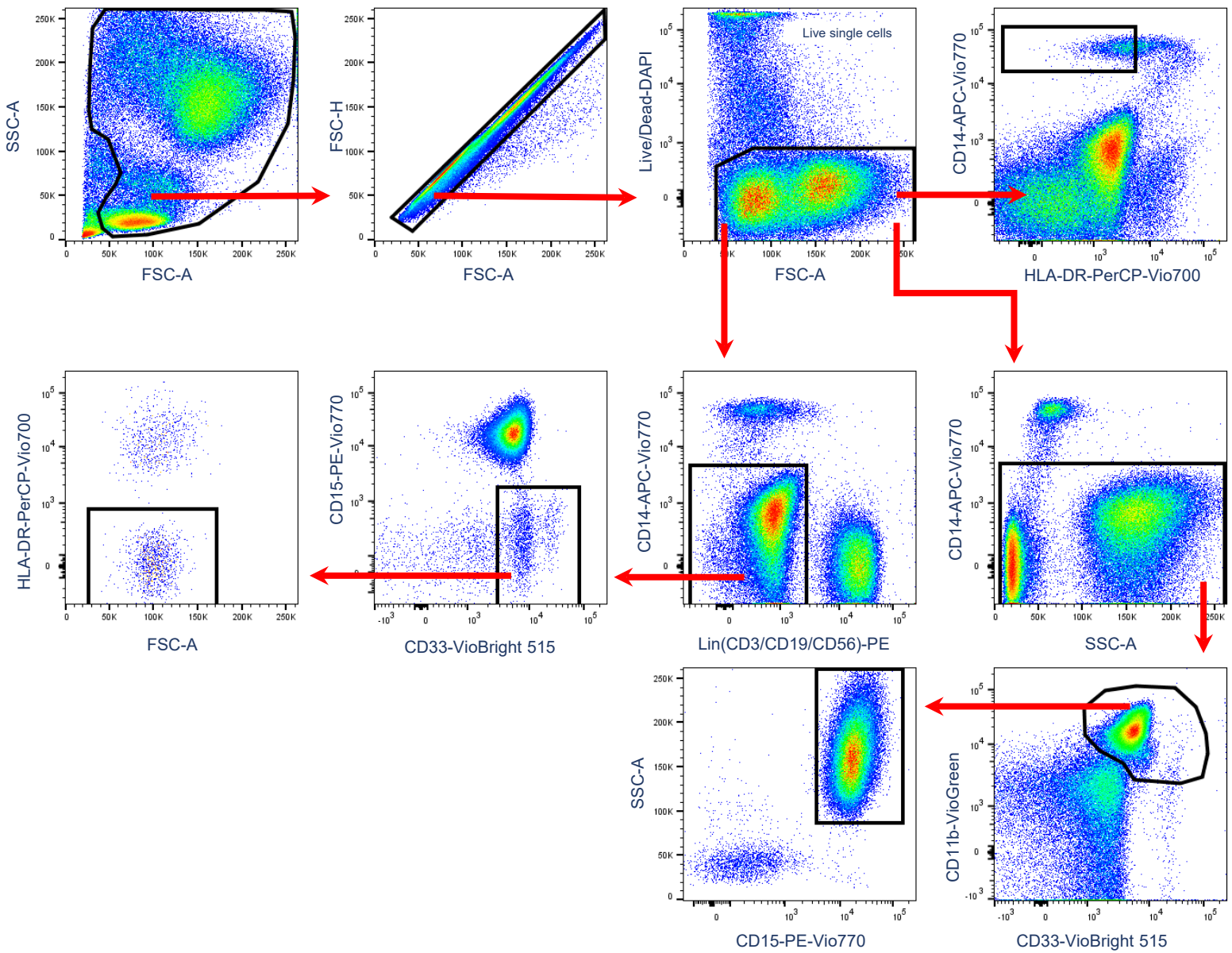


Neural Network – A Type of AI

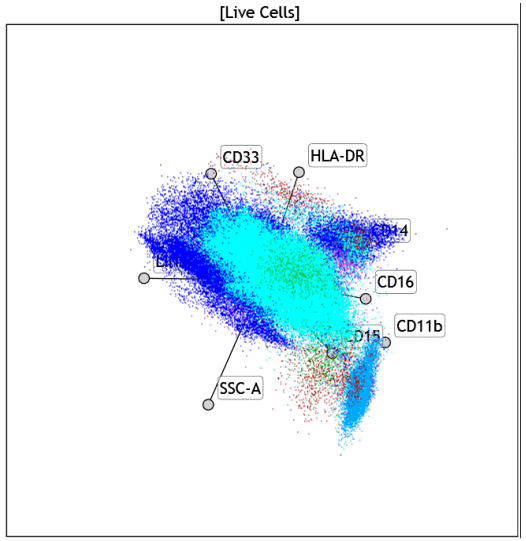
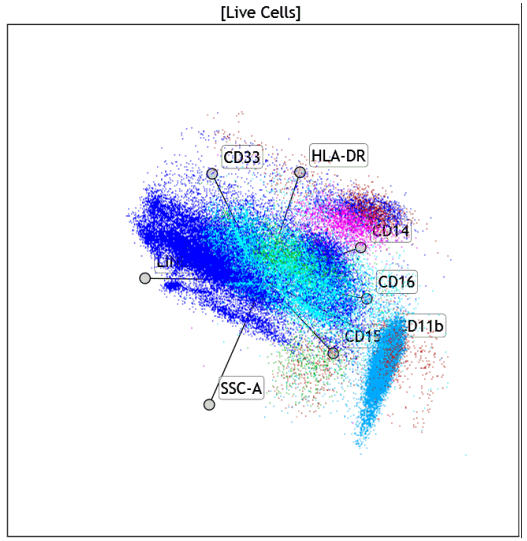
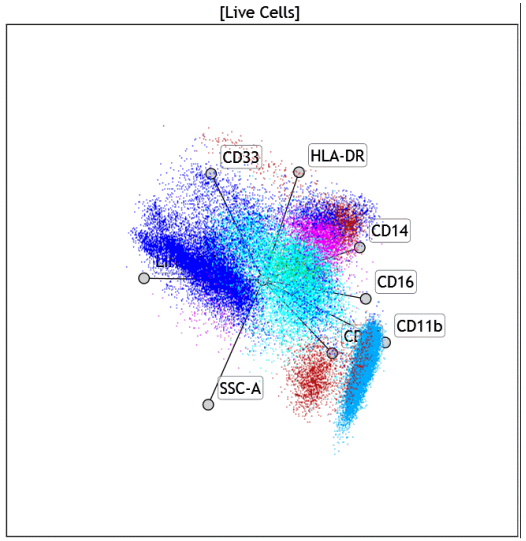
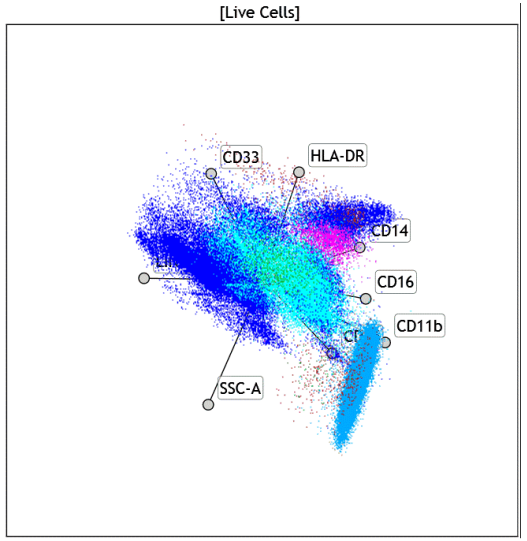
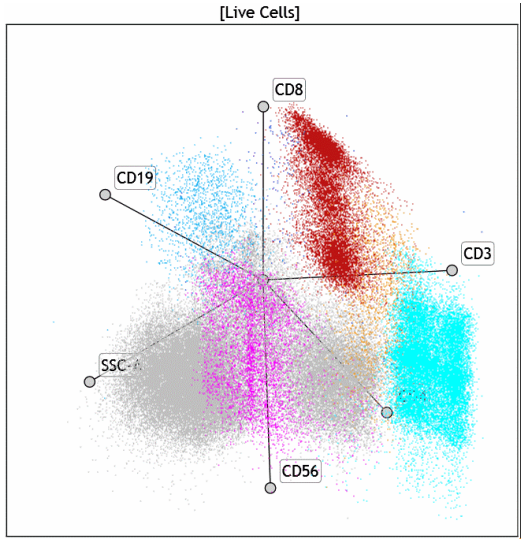
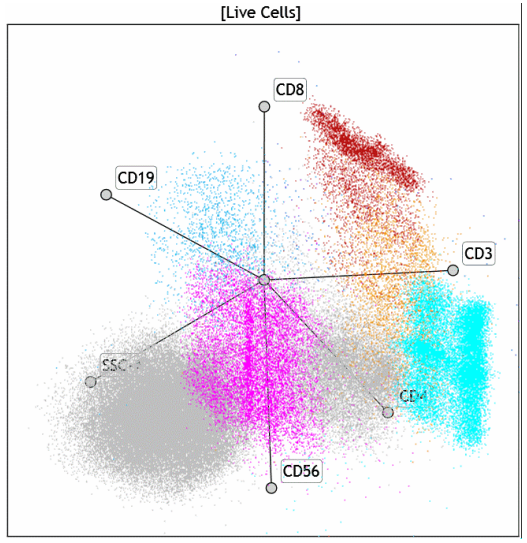
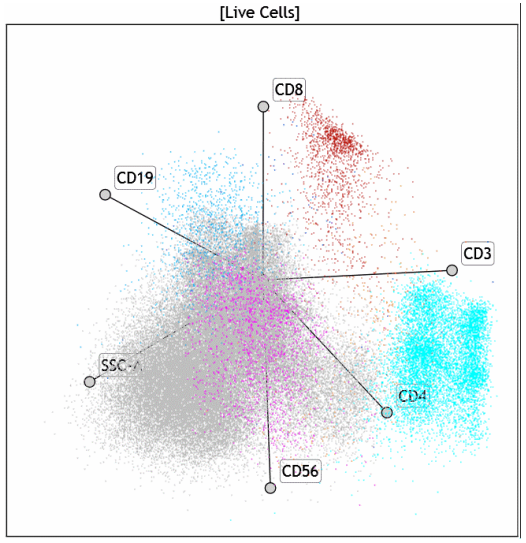
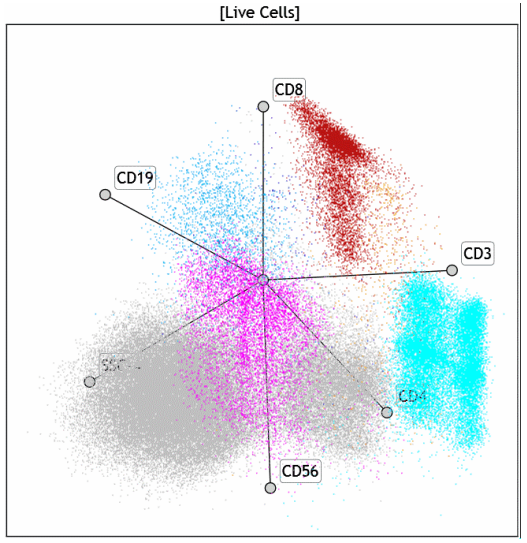


Source: https://www.analyticsindiamag.com/how-to-create-your-first-artificial-neural-network-in-python/

Complex Relationships – More Than Just Gating

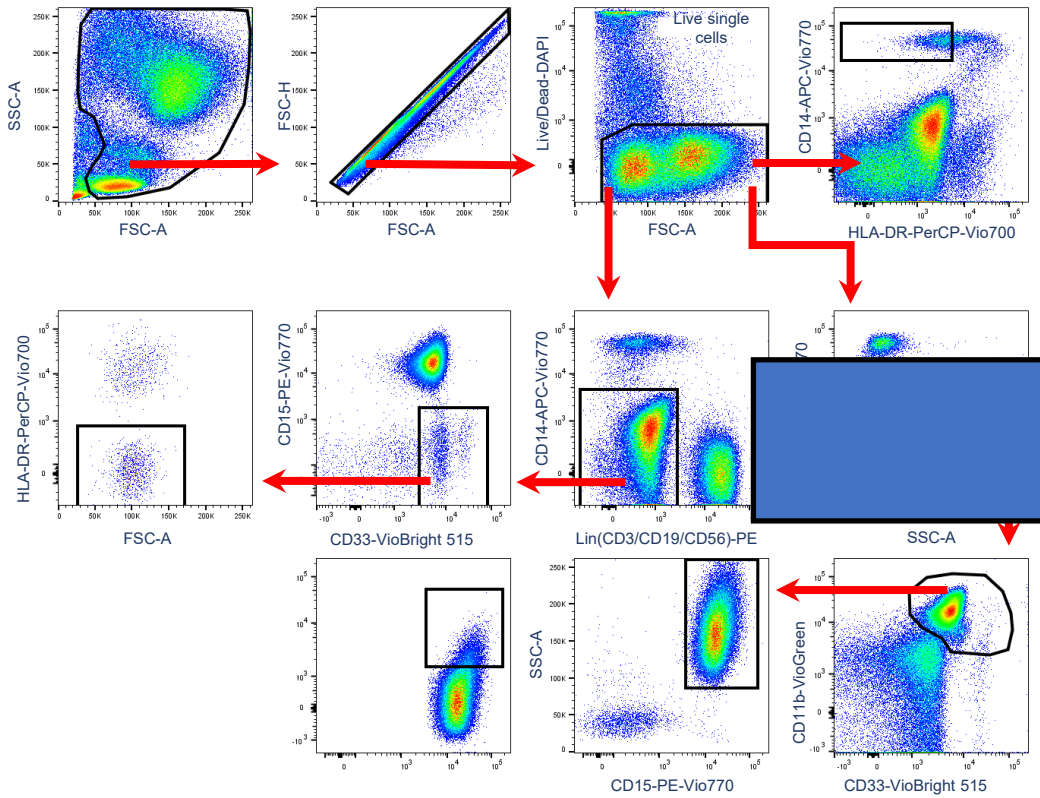


Complex Relationships – More Than Just Gating

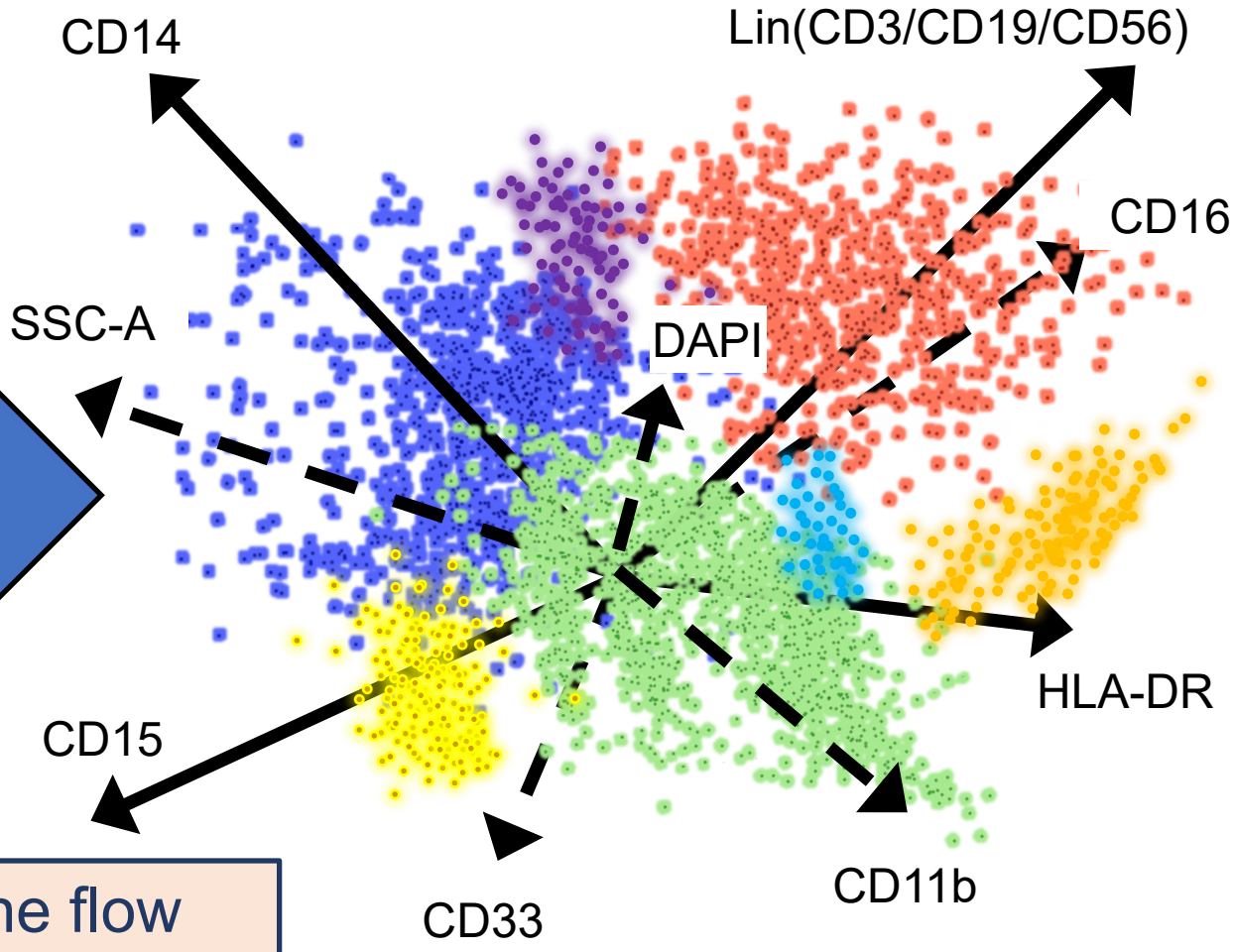
	HDM	BPH	PCa (G=6)	PCa (G≥7)
Myeloid Panel				
Lymphocyte Panel				

Data Transformation

FCS File → CSV File (Event counts for each channel)



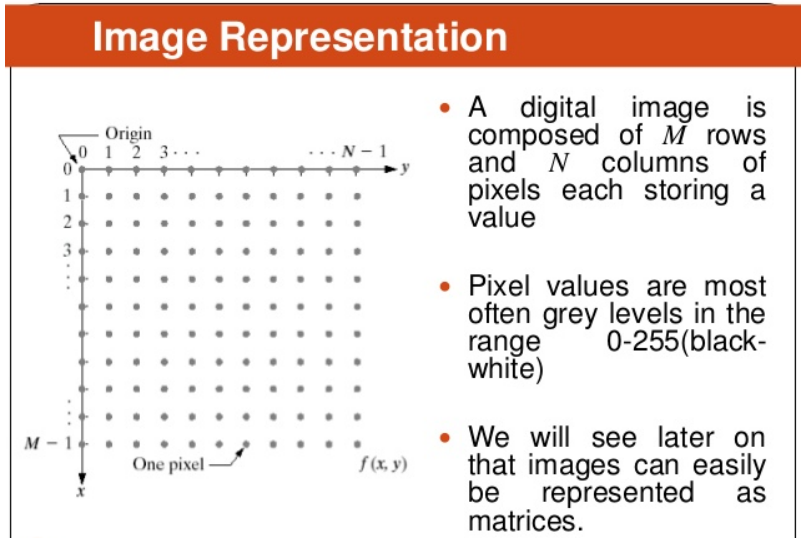
Multidimensional Space

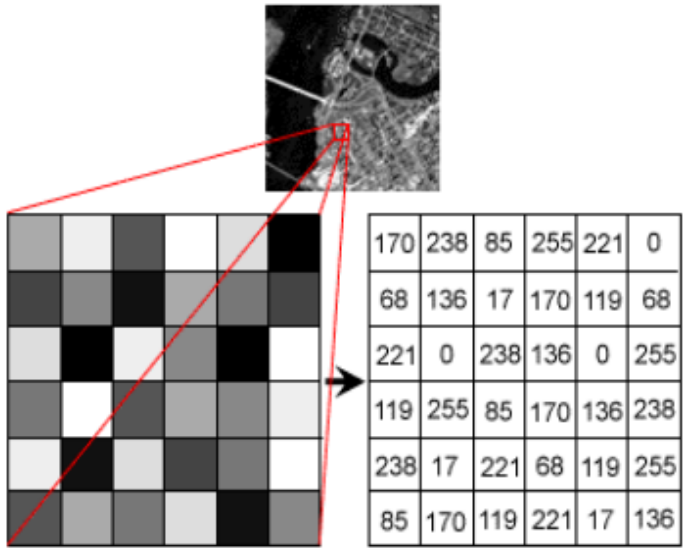


How do we effectively translate the flow cytometry data into a format for NN analysis?

Data Transformation – What is the Common Feature?

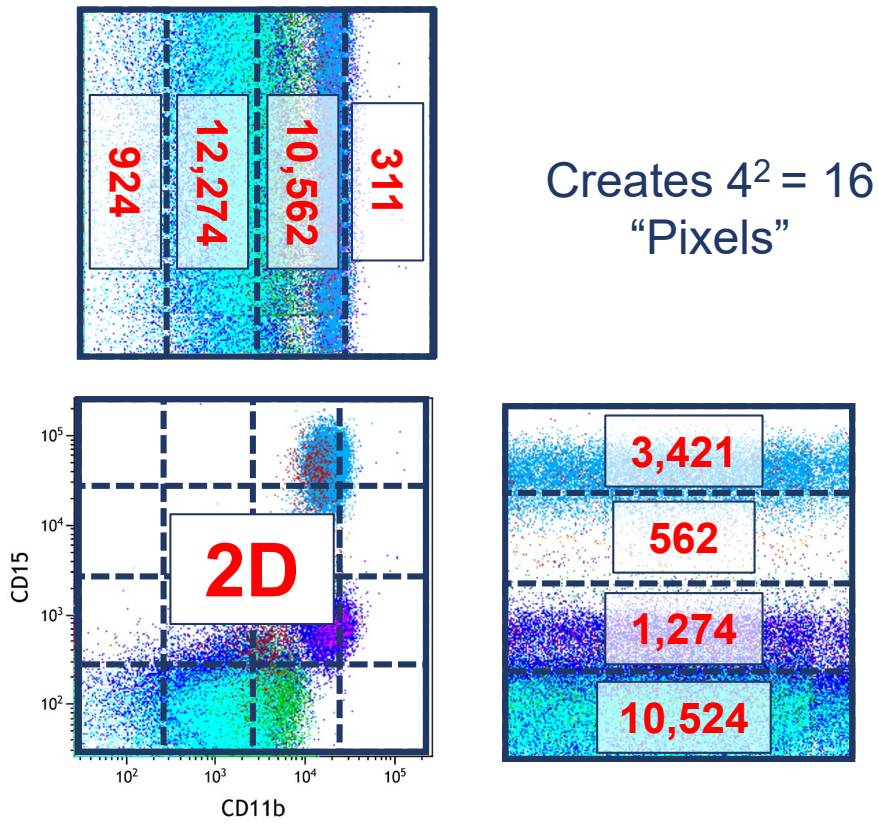
- Samples to be used by a neural network must have features in common across all samples
- For images, every sample has some specific pixel value for each pixel space
  - The pixels are the common features across all images
- But there is no common feature *for the entire sample set* in flow cytometry data
  - An immunofluorescence channel value on a specific event is a common feature across all cells, not between samples





Hypervoxel Generation – The Inputs

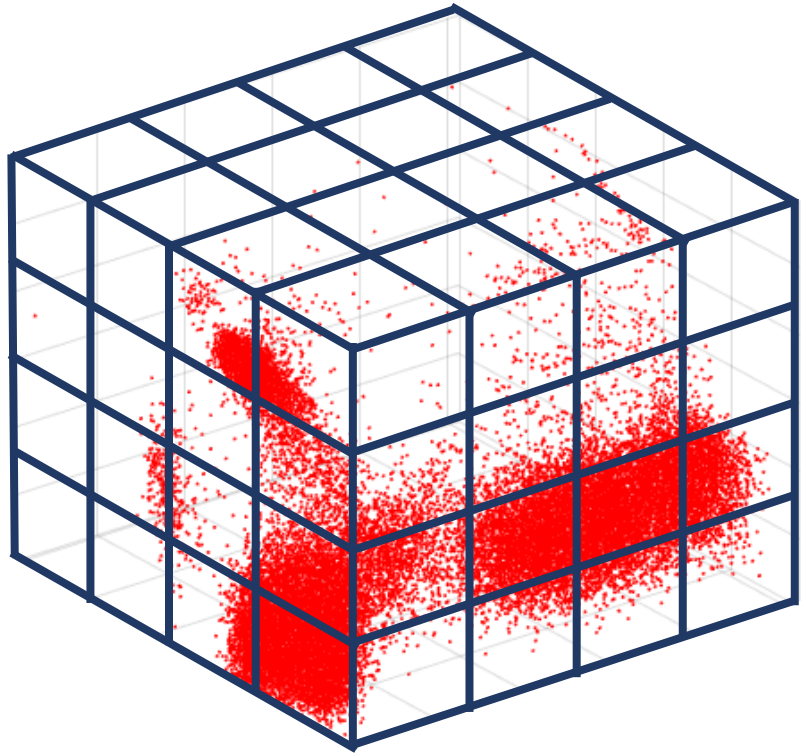
2D: Pixel



Creates $4^2 = 16$ "Pixels"

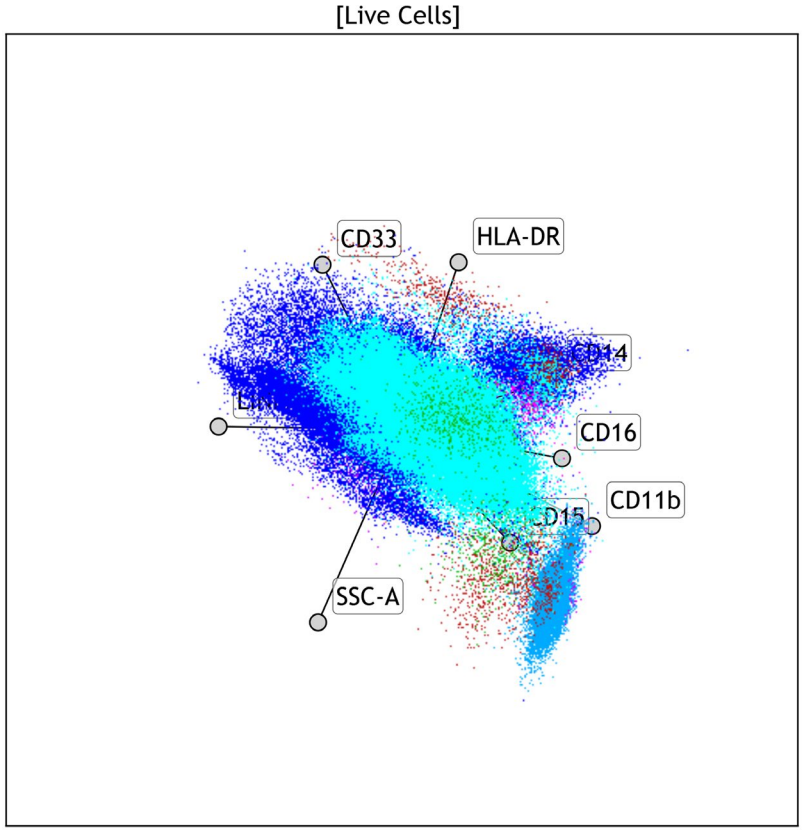
3D: Voxel

Creates $4^3 = 64$ Voxels



Hypervoxel Generation – The Inputs

Multidimensions (>3D): Hypervoxel



8 Markers
+
DAPI

$4^9 = 262{,}144$
Unique Hypervoxel Spaces

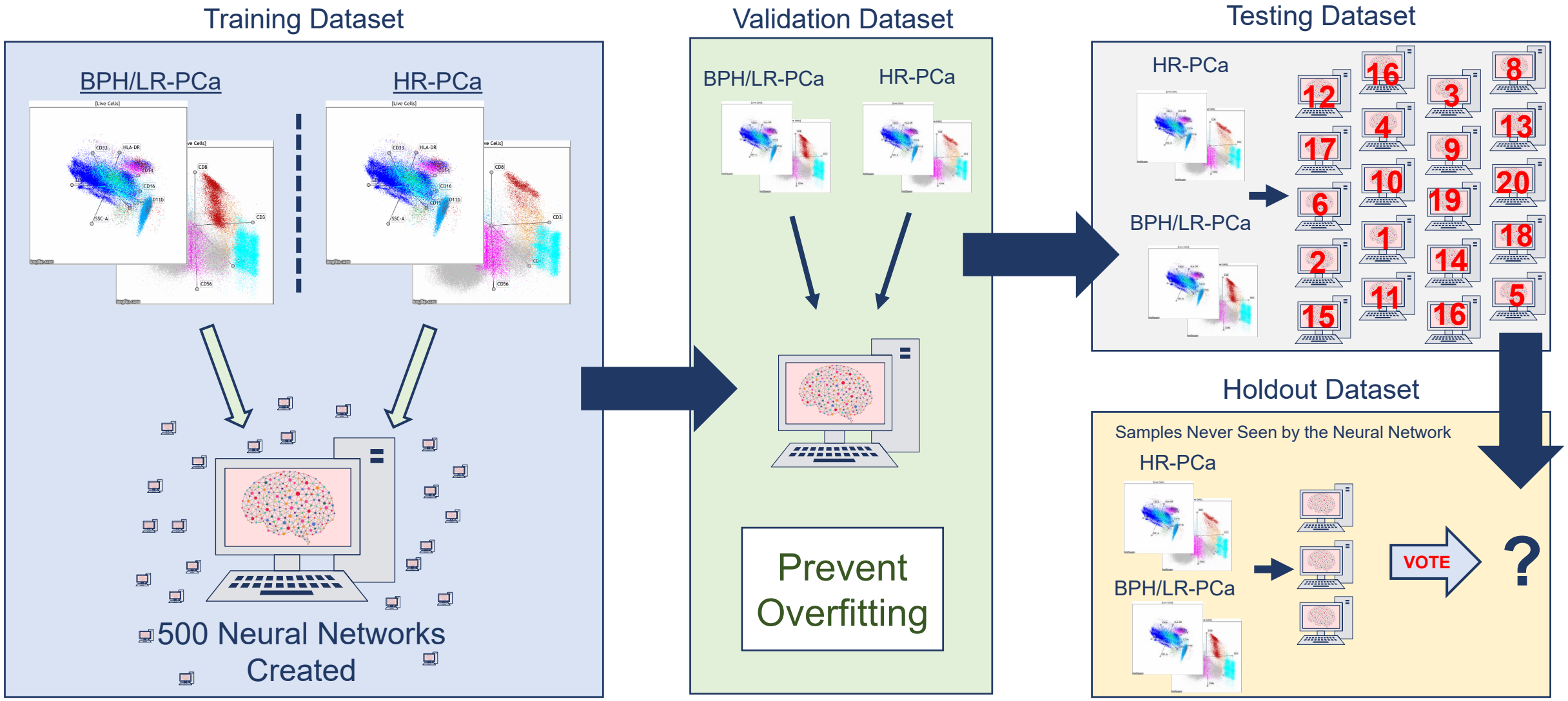
What is the input for the network?

- Each hypervoxel is a unique address
- Each event has channel values that give it an address to a hypervoxel
- The number of events that fall into that address is counted to be the input to the network
    - array of counts

Dataset Generation

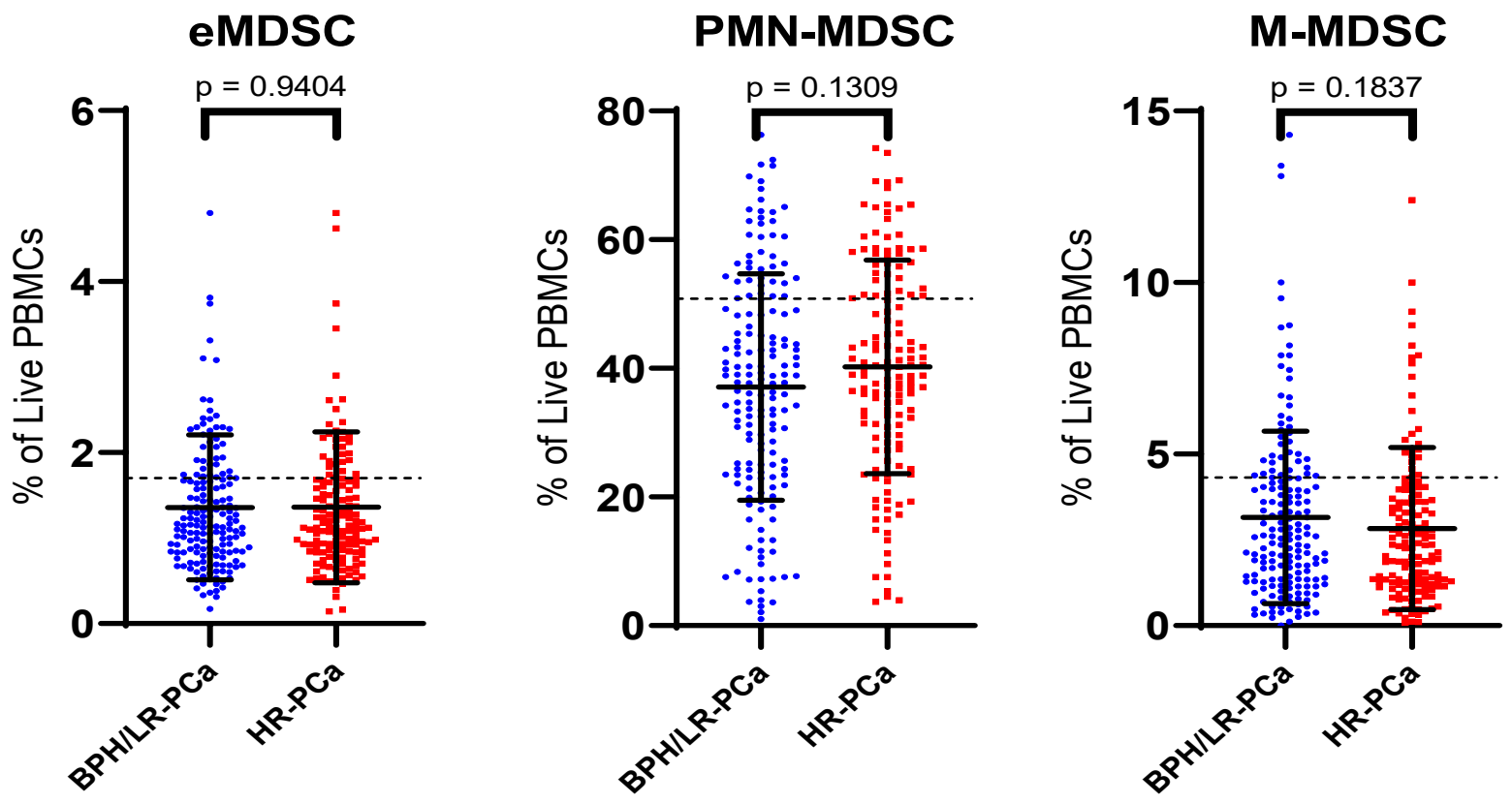
- 100,000 events are collected per sample
- Create sibling samples – variations of 50,000 events
- Per sample – 20 sibling samples are created

Neural Network – Binary Classifier



Clinical Characteristics and Manual Counting

Characteristic	PCa	BPH
Total	145	115
Median Age	68	63
Age Range	42 – 86	40 – 81
Gleason Score		
6	55	
7 (3+4)	31	
7 (4+3)	31	
>8	28	
T-Stage (AJCC 8th Ed.)		
T1c	117	
T2a	6	
T2c	1	
Unknown	21	



Still…simple cell counts can provide information about trends, but not really categorize subjects

Clinical Application: Confirmatory Testing

		Actual	
		HR-PCa	BPH/LR-PCa
Test	HR-PCa	46	33
	BPH/LR-PCa	3	24
Sens. (%) (95%CI)		93.88	(83.13 to 98.72)
Spec. (%) (95%CI)		42.11	(29.14 to 55.92)
PPV (%) (95%CI)		58.23	(52.48 to 63.76)
NPV (%) (95%CI)		88.89	(71.94 to 96.15)
Acc. (%) (95%CI)		66.04	(56.20 to 74.96)

Population: Men who were going to prostate biopsy

Samples used for training:

41 HR-PCa
113 BPH/LR-PCa

Holdout Samples for Testing:

49 HR-PCa
57 BPH/LR-PCa

- Potentially reduce the number of unnecessary prostate biopsies?

Conclusions

- We demonstrated that we can combine machine learning with flow cytometry data
  - Pros: Objective, Large Amounts of Data, More Data Inclusion, Simple Assay
  - Cons: Supervised Learning, Requires Large Amounts of Data for Training, Difficult to Determine Most Important Relationship – "Black Box" Idea
- We have applied this technique to be a binary classifier
  - distinguish between BPH/PCa in a small number of samples
- This can used for other classifications and tumor types
  - Breast Cancer: Early Stage vs non-Tumor Bearing
  - Breast Cancer: Early Stage vs DCIS
  - Prostate Cancer: Advanced Stage vs non-Tumor Bearing

Future Work

- Identify the critical relationships between cell populations that are used to make the classifications → hypervoxels
- Can this technique be applied to other flow cytometry data sets with different cancers or more cell populations? CyTOF? (retrospective analysis)
- As more and more data sets are generated with more and more information, can this help answer the question: What do we do with all of this data? What are we really trying to answer?
- Can this be used for answering other questions?
  - tumor recurrence, treatment and/or immunotherapy responses
    - Collaborative projects

Thank You!

Anixa Biosciences

San Jose

Amit Kumar, PhD

John Roop

Anthony Campisi

Philadelphia

Alexander Polo

The Wistar Institute

Dmitry I. Gabrilovich, MD, PhD

Flow Cytometry Core Facility

Jeffrey Faust

Clinical Collaborators

New Jersey Urology, LLC

Adam Perzin, MD

Renee Haney, CCRC

Brianna Florentine, CRC

Jennifer Pilallis, CRC

MD Anderson Cancer Center at Cooper

Robert Somer, MD

Mary Schafer

Questions?