Basic Data Analysis, Gating, and Statistics in Flow Cytometry

Alex Henkel

Associate Instrumentation Specialist

amhenkel@wisc.edu

uwflow@uwcarbone.wisc.edu

Outline

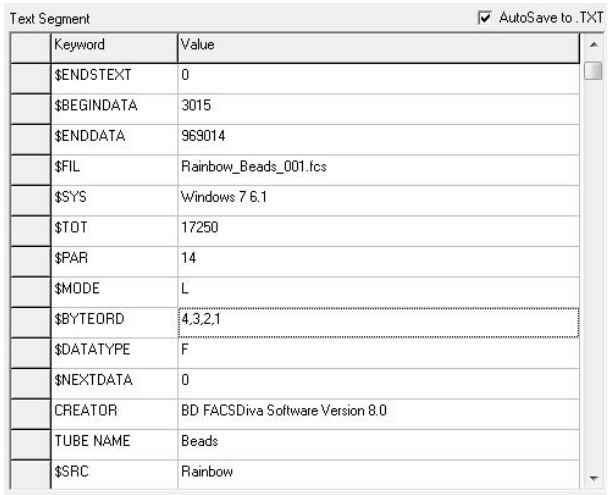
- What is an FCS File?
- Visualization and Scaling
- Gating and Controls
- Basic Statistics in Flow Data
- The Process

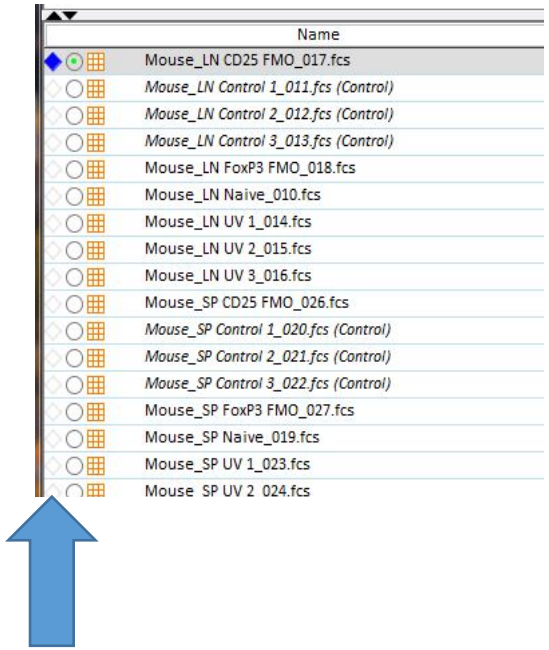
The FCS File – Flow Cytometry Standard File

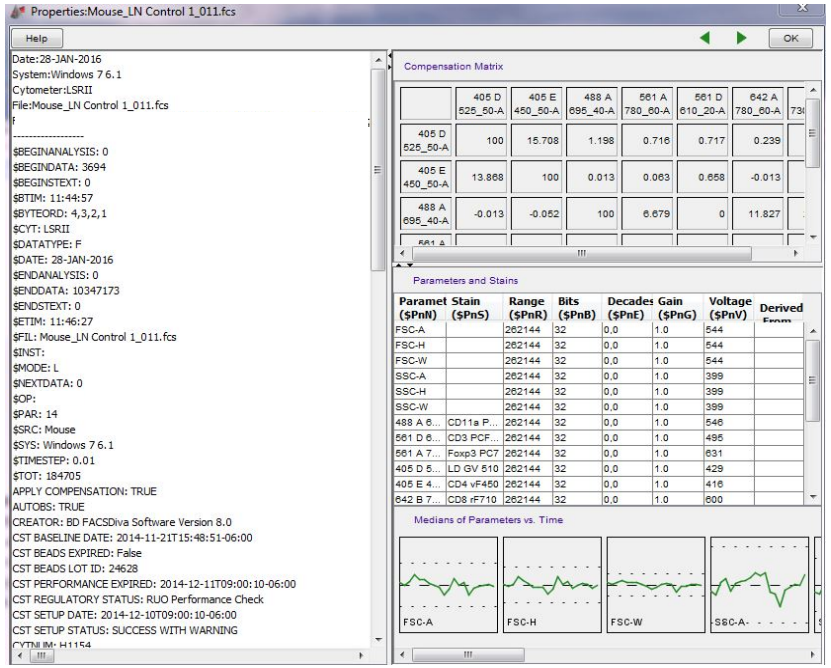
Parameter	2	3	4	5	6	7	8	9	10
Name	FSC-H	FSC-W	SSC-A	SSC-H	SSC-W	488 A 695/40	561 D 610/20	561 A 780/60	405 D 525/50
String									
Range	262144	262144	262144	262144	262144	262144	262144	262144	262144
Bits	32	32	32	32	32	32	32	32	32
1	136813.00	72175.84	100686.88	88489.00	74569.90	9000.32	15158.64	3904.32	15887.52
2	136392.00	72013.31	98798.56	86540.00	74819.30	8873.76	15250.76	3290.84	16505.28
3	136761.00	72370.88	99523.20	86557.00	75353.27	8064.00	14991.06	3386.88	16967.61
4	136288.00	72599.80	102604.32	90900.00	73974.45	8835.68	13644.54	2762.62	16658.73
5	136493.00	72679.64	102473.28	90521.00	74189.29	8384.32	14253.12	3108.56	15756.84
6	137809.00	71919.16	102944.80	91413.00	73803.40	8487.36	14612.78	3410.40	17198.28
7	136451.00	72557.54	97821.92	85559.00	74929.08	8755.04	14392.28	3188.92	15426.18
8	131217.00	73191.84	87923.36	77317.00	74526.24	7935.20	13974.80	3613.26	14421.33
9	137763.00	72796.29	98178.08	86391.00	74477.66	8622.88	14378.56	2479.40	15633.09
10	138296.00	71846.40	101661.28	90380.00	73716.24	8266.72	15017.52	3336.90	17202.24
11	138825.00	72279.32	108109.12	94829.00	74713.85	9089.92	14271.74	3181.08	16338.96
12	137084.00	71431.94	97625.92	86136.00	74278.02	9569.28	15296.82	3063.48	17455.68
13	136099.00	72974.48	94170.72	81575.00	75655.19	8542.24	16242.52	3597.58	17812.08
14	138561.00	72564.37	100914.24	87989.00	75162.98	8611.68	15122.38	3232.04	17161.65
15	120599.00	73751.43	128493.12	112851.00	74619.85	13147.68	22317.54	4793.18	23744.16
16	132646.00	73071.72	91556.64	79586.00	75393.36	7881.44	14176.68	3687.74	14792.58

- Comprised of a text segment and data segment.
- FCS Files are in a list mode data format.
- Rows = Events
- Columns = Parameters
  - H,A,W each get their own column.
- Annotate data before recording.

The FCS File - Header





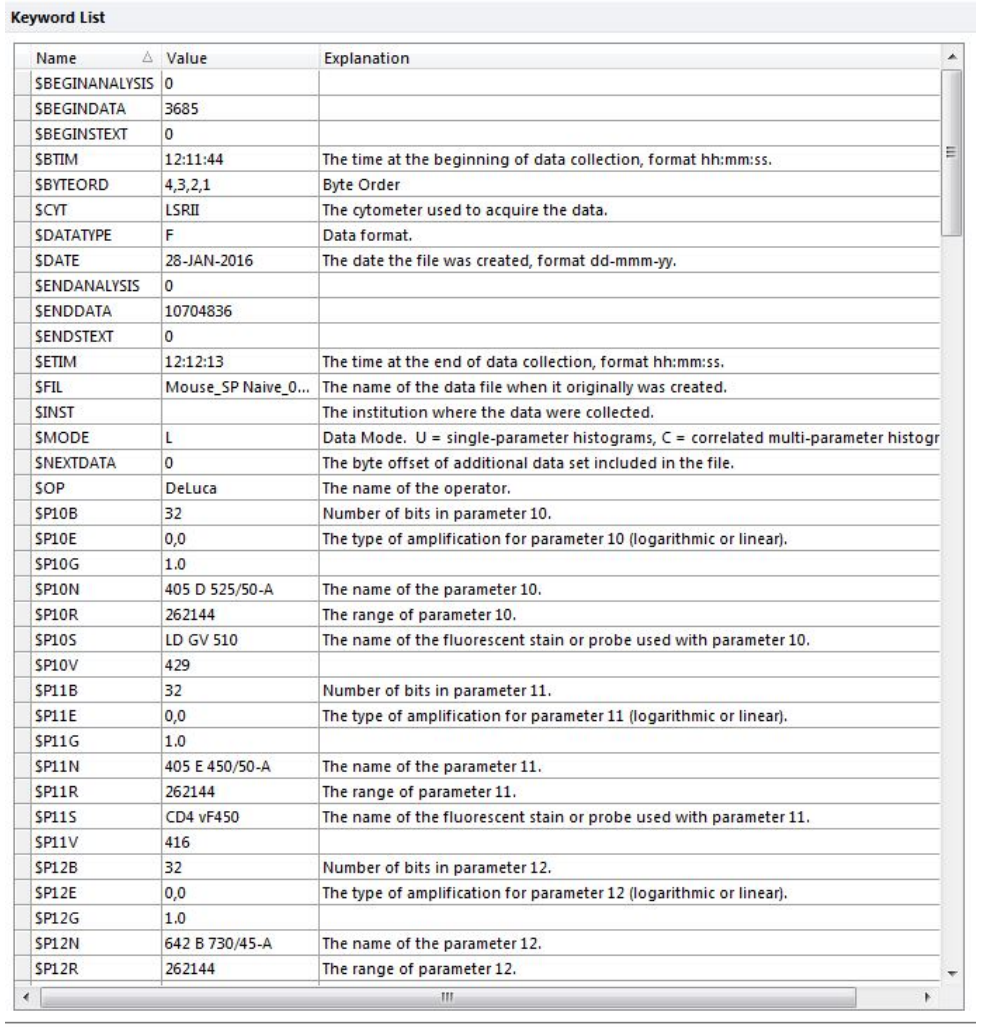


- Header shows information about the data collection.
- Stores all the metadata from the instrument, including any added labels.

- Double clicking on the diamond next to a file in FlowJo will open the file information, including header information.

Keywords

- All the header information can be extracted as keywords in data analysis software, allowing you to group, or sort files by keyword.
- Keywords can be added to tables with sample statistics.
- Keywords can be metadata.

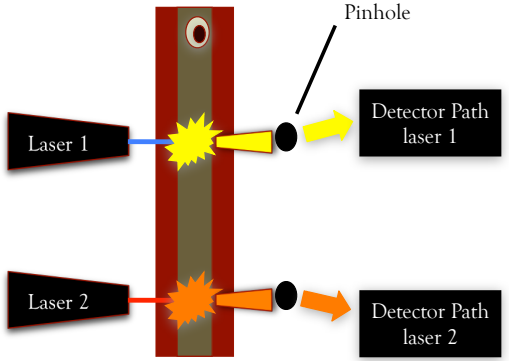


Keyword List

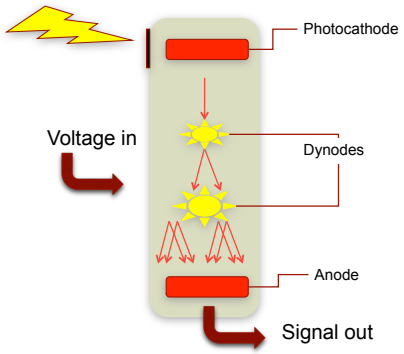
Name	Value	Explanation
$BEGINANALYSIS	0	
$BEGINDATA	3685	
$BEGINSTEXT	0	
$BTIM	12:11:44	The time at the beginning of data collection, format hh:mm:ss.
$BYTEORD	4,3,2,1	Byte Order
$CYT	LSRII	The cytometer used to acquire the data.
$DATATYPE	F	Data format.
$DATE	28-JAN-2016	The date the file was created, format dd-mmm-yy.
$ENDANALYSIS	0	
$ENDDATA	10704836	
$ENDSTEXT	0	
$ETIM	12:12:13	The time at the end of data collection, format hh:mm:ss.
$FIL	Mouse_SP Naive_0...	The name of the data file when it originally was created.
$INST		The institution where the data were collected.
$MODE	L	Data Mode. U = single-parameter histograms, C = correlated multi-parameter histogr
$NEXTDATA	0	The byte offset of additional data set included in the file.
$OP	DeLuca	The name of the operator.
$P10B	32	Number of bits in parameter 10.
$P10E	0,0	The type of amplification for parameter 10 (logarithmic or linear).
$P10G	1.0	
$P10N	405 D 525/50-A	The name of the parameter 10.
$P10R	262144	The range of parameter 10.
$P10S	LD GV 510	The name of the fluorescent stain or probe used with parameter 10.
$P10V	429	
$P11B	32	Number of bits in parameter 11.
$P11E	0,0	The type of amplification for parameter 11 (logarithmic or linear).
$P11G	1.0	
$P11N	405 E 450/50-A	The name of the parameter 11.
$P11R	262144	The range of parameter 11.
$P11S	CD4 vF450	The name of the fluorescent stain or probe used with parameter 11.
$P11V	416	
$P12B	32	Number of bits in parameter 12.
$P12E	0,0	The type of amplification for parameter 12 (logarithmic or linear).
$P12G	1.0	
$P12N	642 B 730/45-A	The name of the parameter 12.
$P12R	262144	The range of parameter 12.

How does the cytometer generate values?

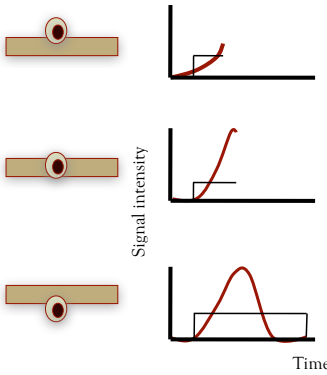
- Detection of light from scatter or fluorescence in cells is converted into a voltage pulse.
- Each pulse generated by an object passing by the laser has a height, area and width measurement.
- Values for H, A, and W from each parameter stored in a listmode data format in FCS file.



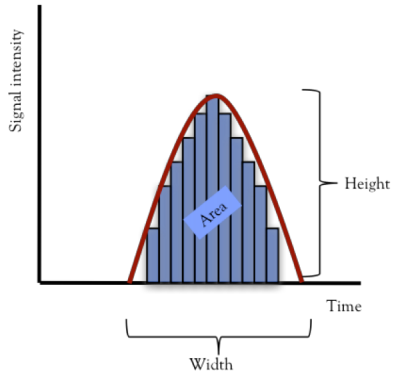
1. Emission



2. Detection



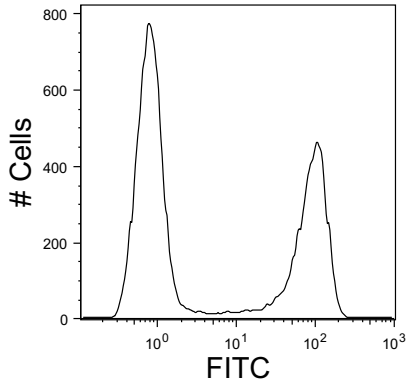
3. Converted to Voltage



4. Measured

Event	Time	FSC	SSC	FITC	PE	APC
1	0	100	500	10	650	4
2	0	110	505	700	700	6
3	0	90	480	720	670	10
4	0	95	490	700	720	15
5	0	12	76	15	15	13
6	0	120	600	14	810	785
7	0	108	530	16	595	18
8	0	117	654	12	720	12
9	1	54	276	378	576	18
10	1	193	803	690	912	790

5. File Generated



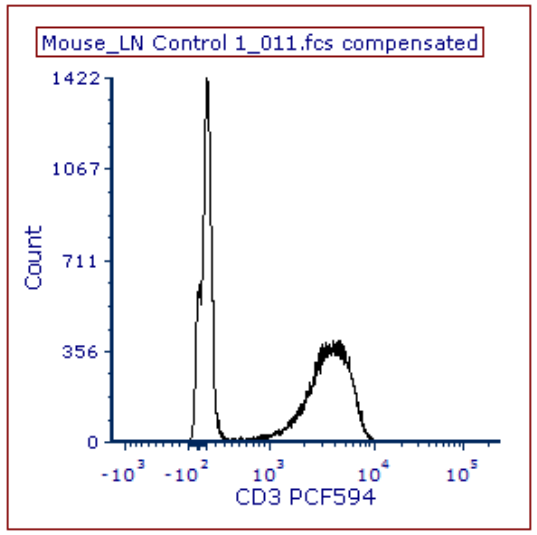
6. Plotted

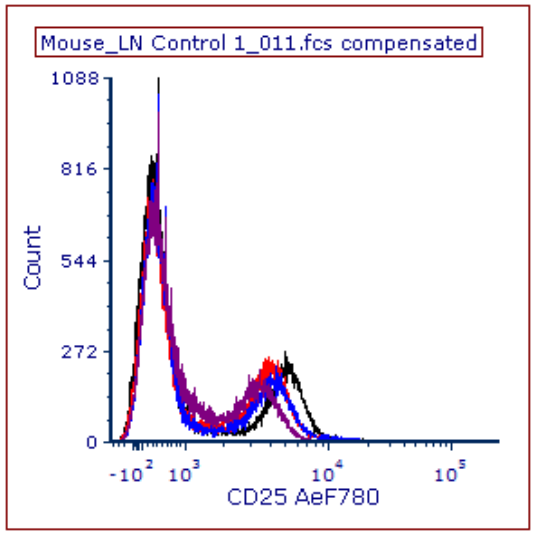
From Excyte Expert Cytometry

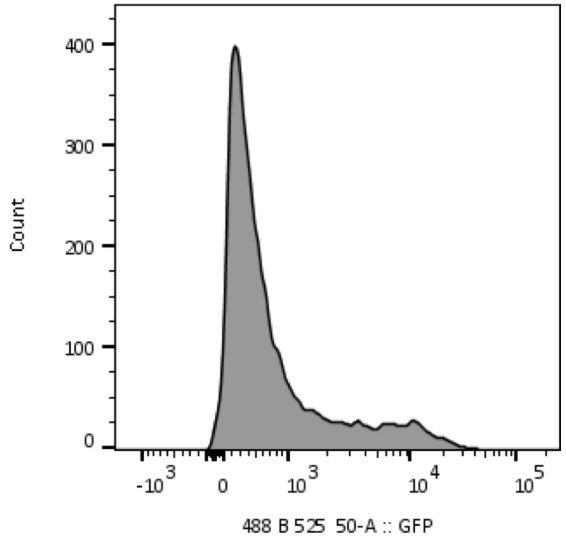
Visualization

- We can create several types of plots with the generated data.
- Histograms, dot plots, density plots, contour plots.
- We want to display the data in a way that relates to our hypothesis.

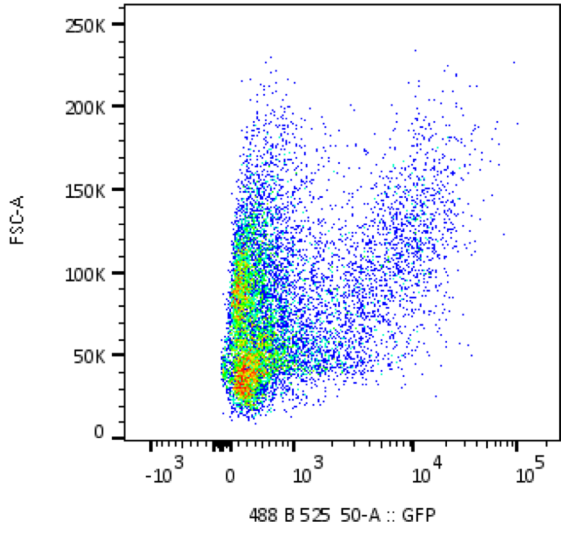
Histograms



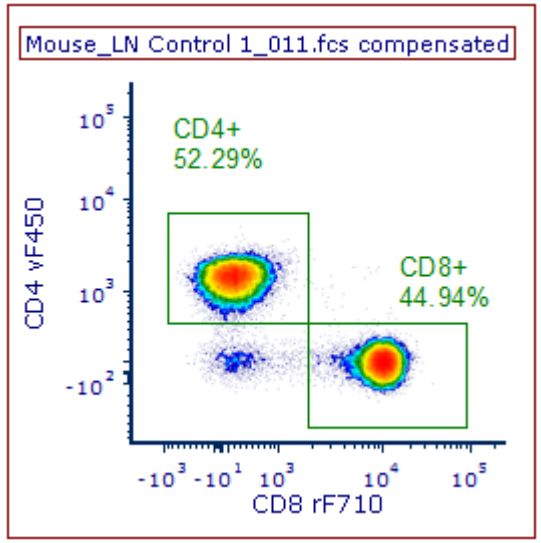


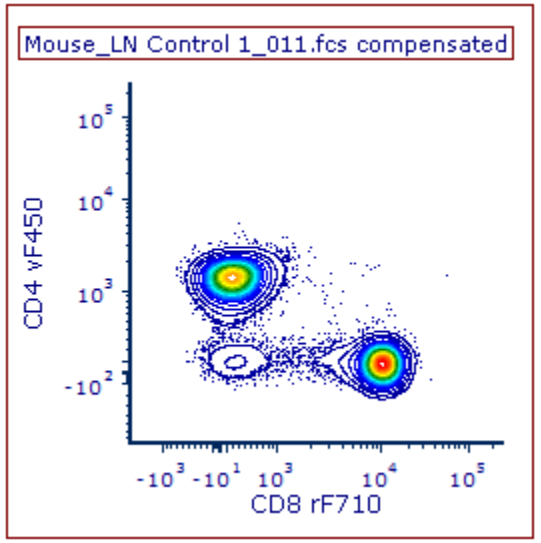


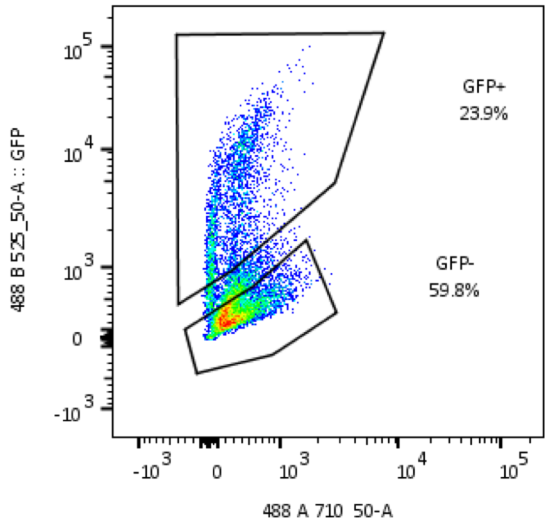
- Shows the distribution of values for a specific parameter.
- Cannot see the relationship between two populations.
- Can miss sub-populations that have similar values in one parameter.



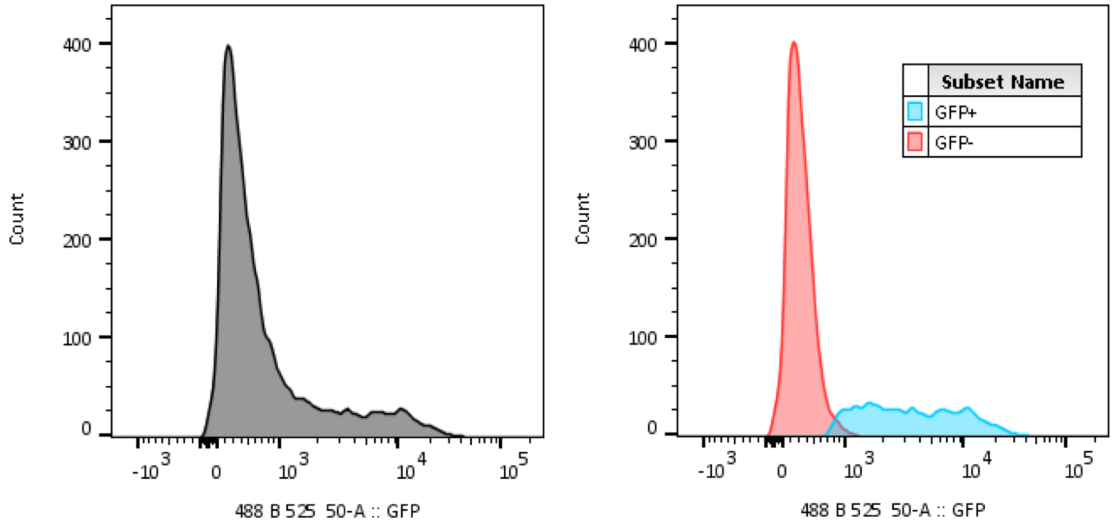
Bi-variate Dot Plots





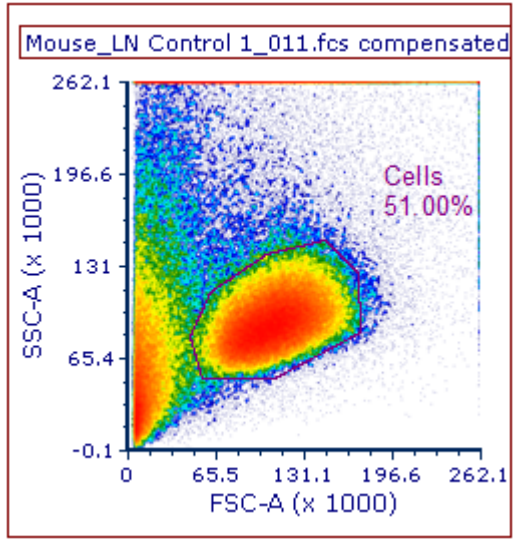


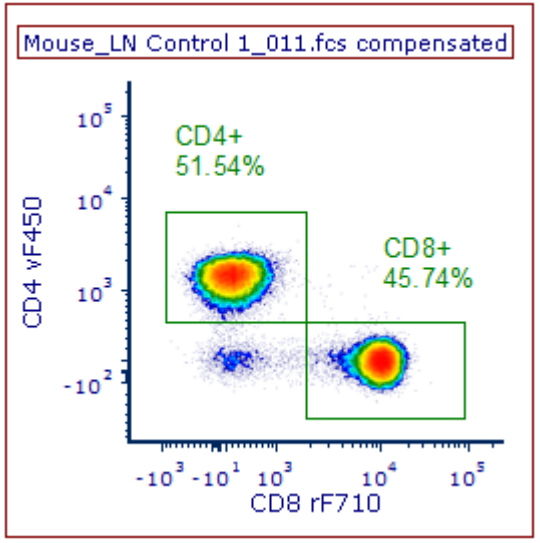
- Can see relationships between markers.
- Can see sub-populations from distinctions in two dimensions instead of one.



Scaling Considerations

- Forward and Side scatter almost always displayed in **linear scale**.
  - Exception for very small things like bacteria, extracellular vesicles, and nuclei.
  - Use linear for small dynamic range.
- Fluorescence parameters almost always in **log scale**.
  - Exception for low signal increase. Ex. Cell cycle assay will only have a two-fold increase in fluorescent intensity.
  - Use log for large dynamic range.

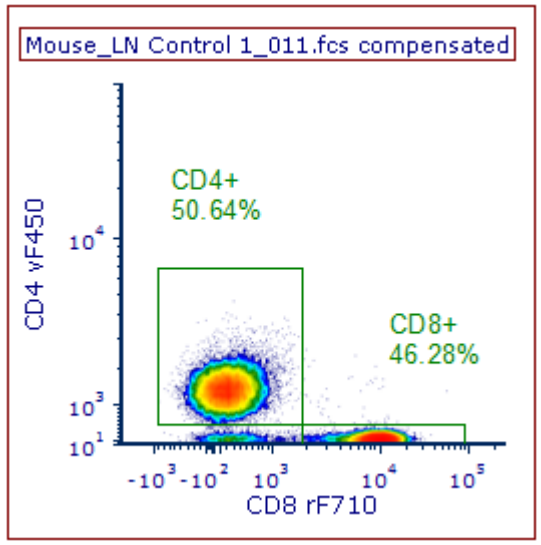




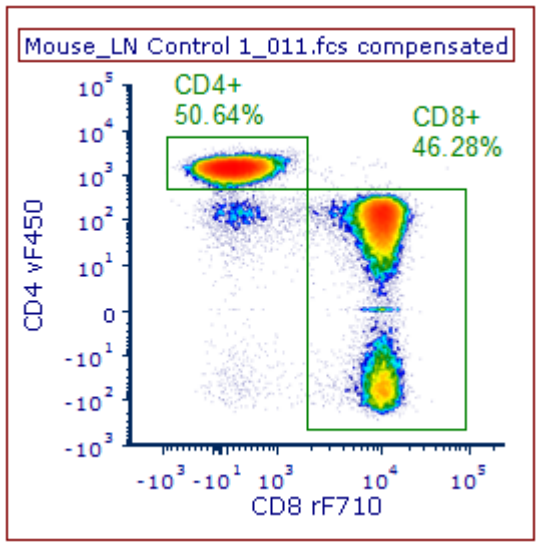
Scaling Plots to Display Data

- Sometimes plots don't display the data in the best way.
- Changing the scaling does not change the values, just the display of the data.
- Linear, Log, Biexponential, Hyperlog.

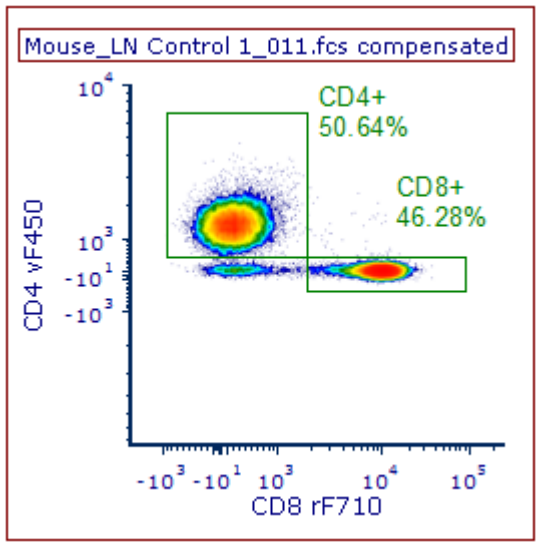
Log



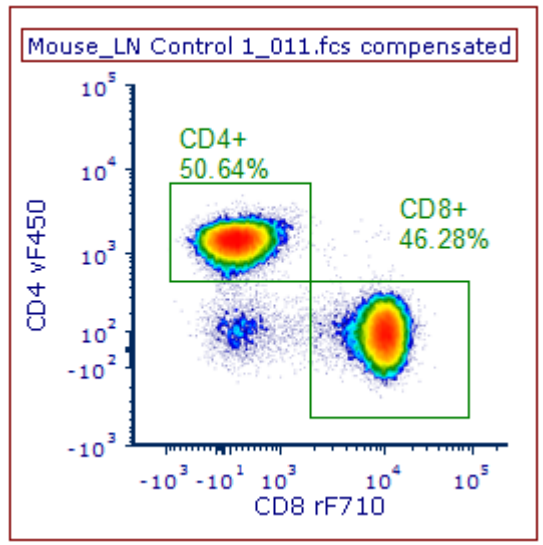
Log w/ Negative



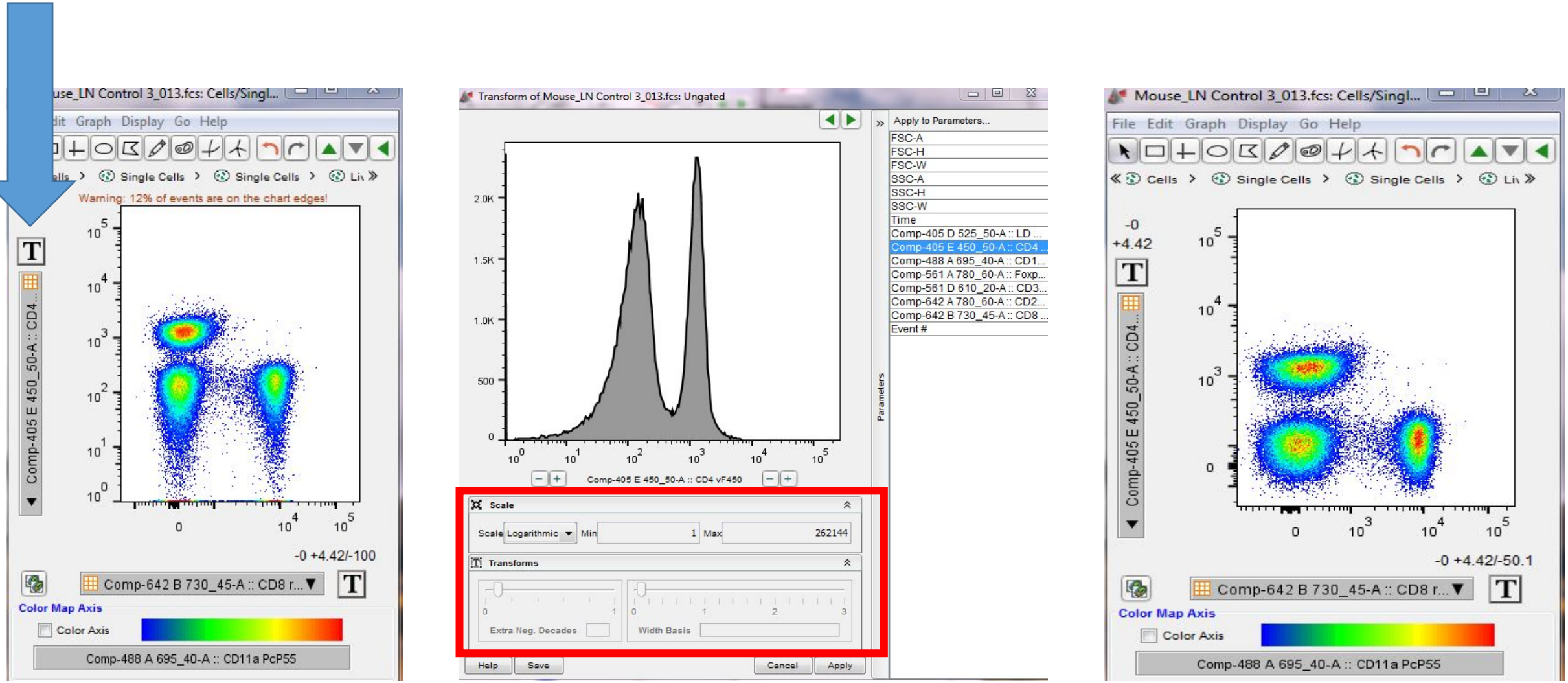
Biexponential – Extra negative decades



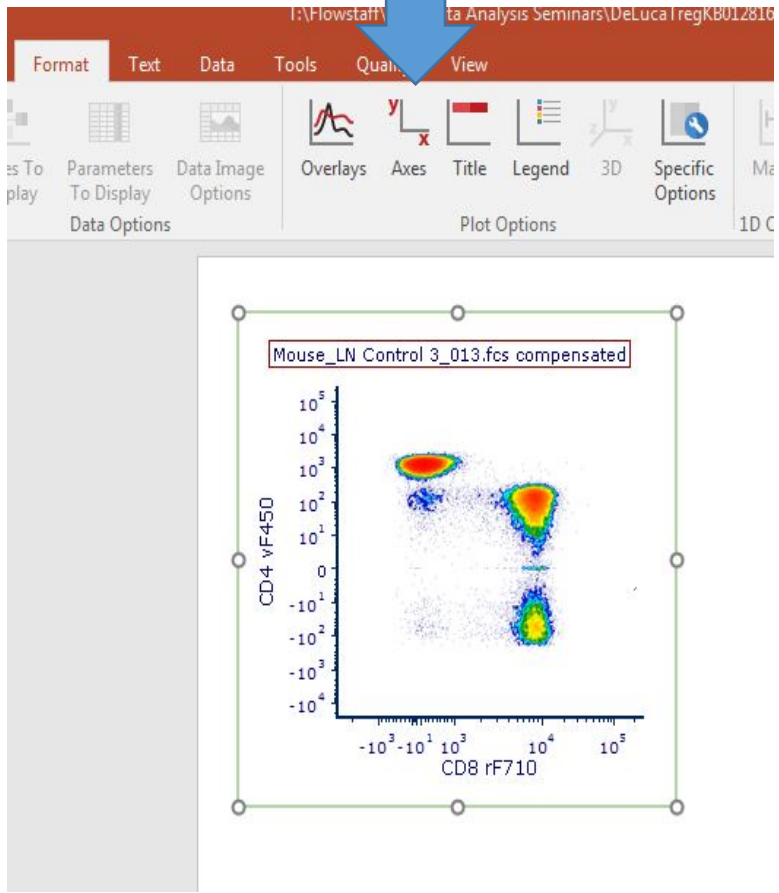
Biexponential – Proper extra negative decades

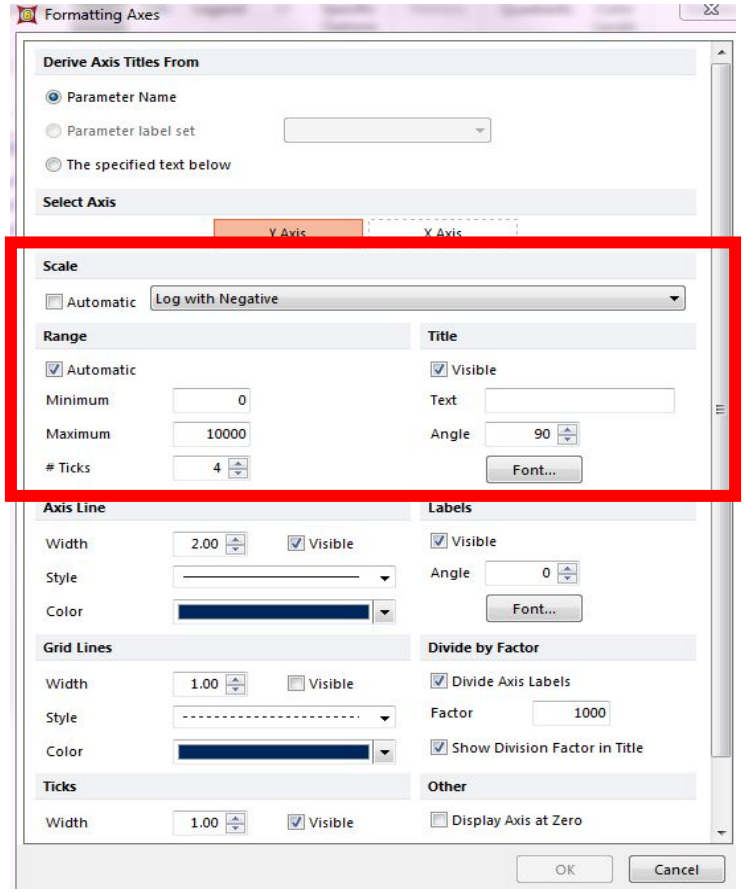


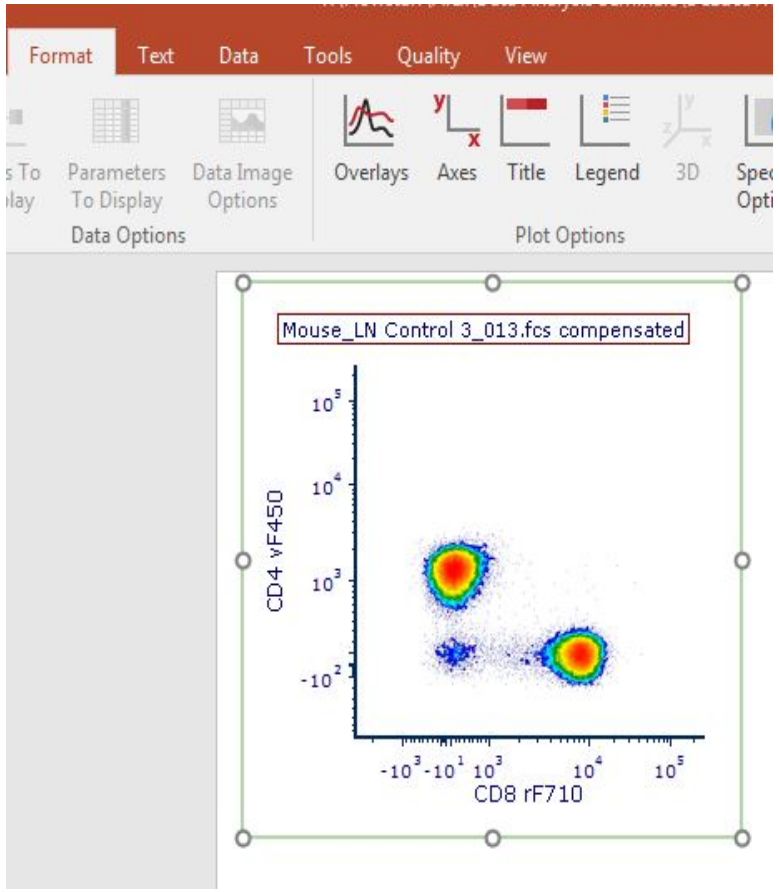
Adjusting Scaling in FlowJo



Adjusting Scaling in FCS Express

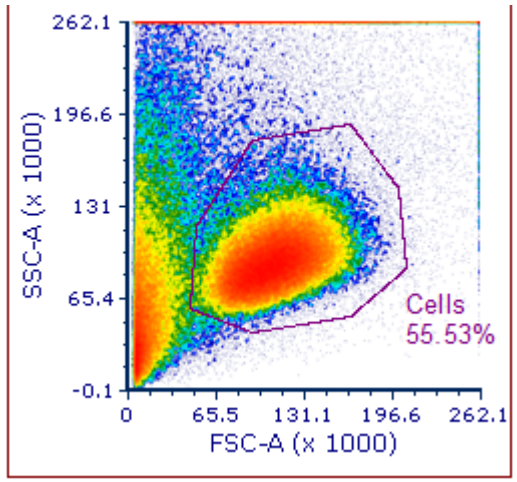


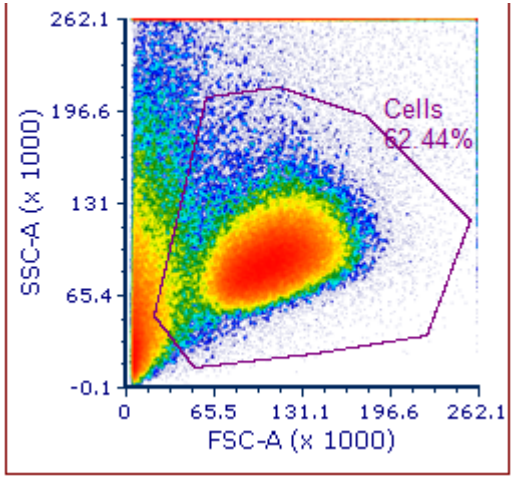


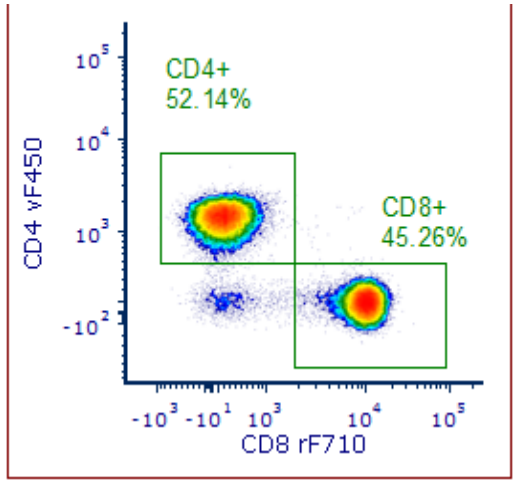


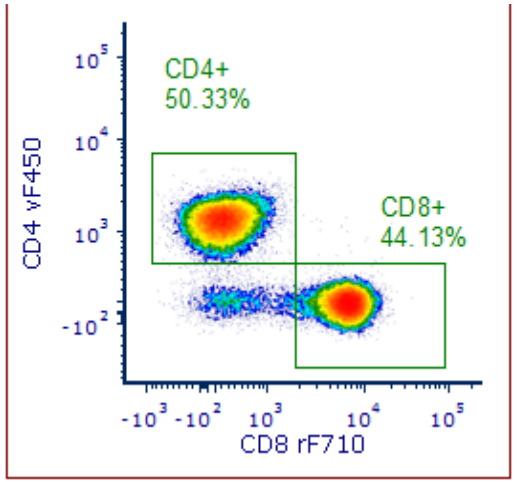
Basic Gating Considerations

- Gating on cells only
  - Exclude debris.
- Doublet Discrimination
  - Removes events that are two cells stuck together.
- Live/Dead gating
  - Dead cells soak up antibody.
- Each of these things lead to double positive or strange events!

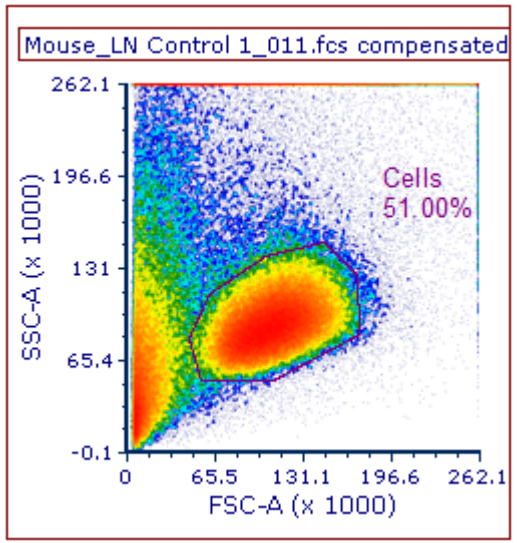


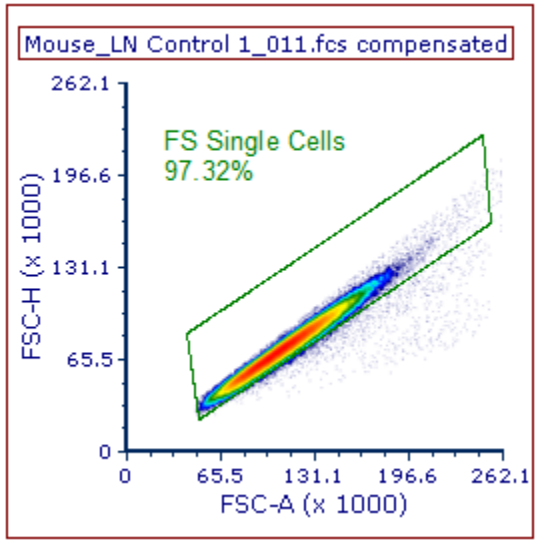


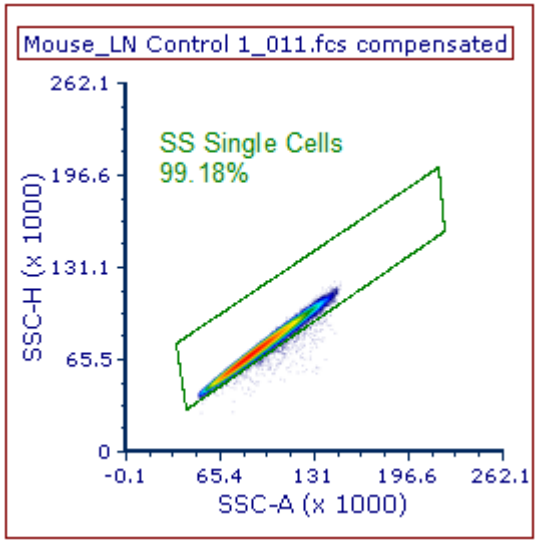


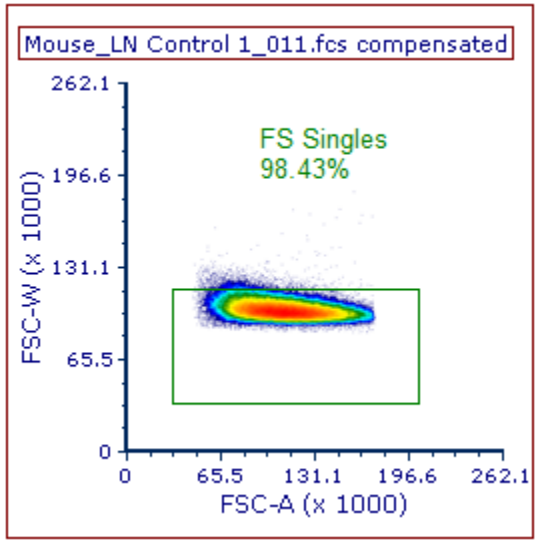


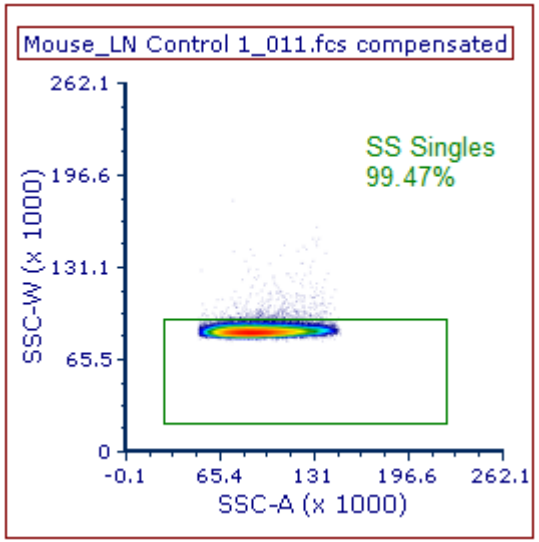
Doublet Discrimination





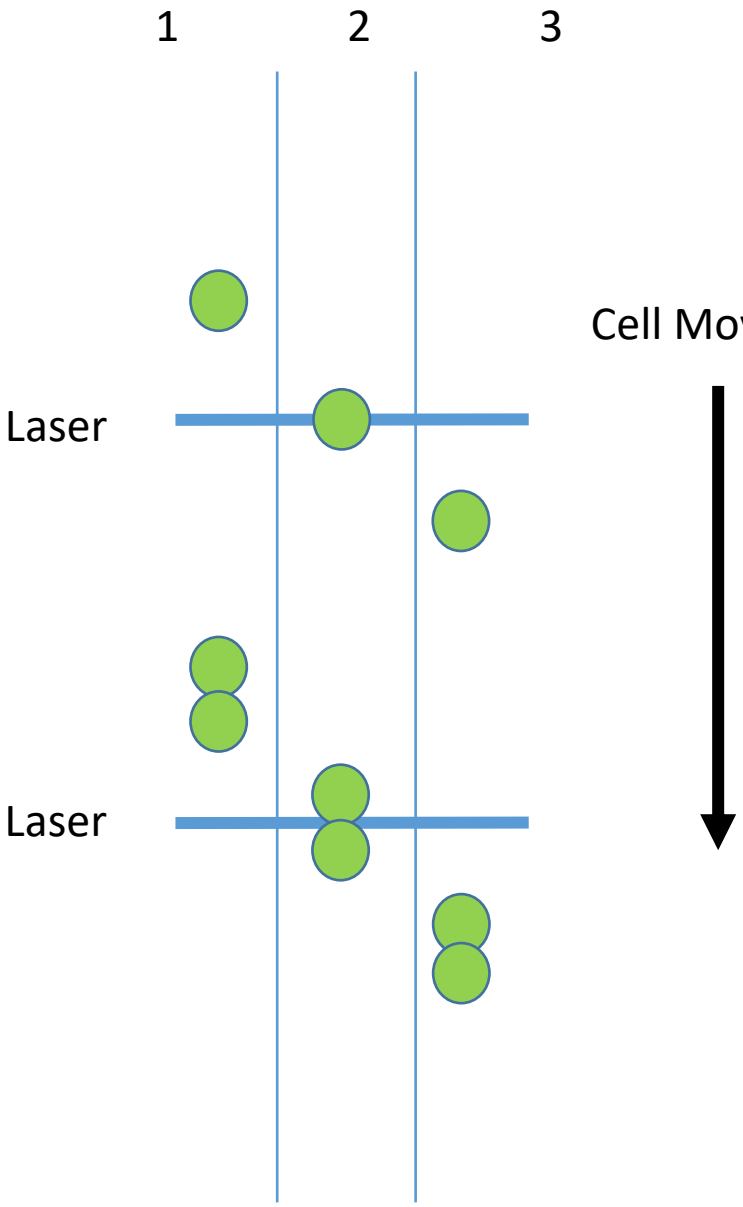


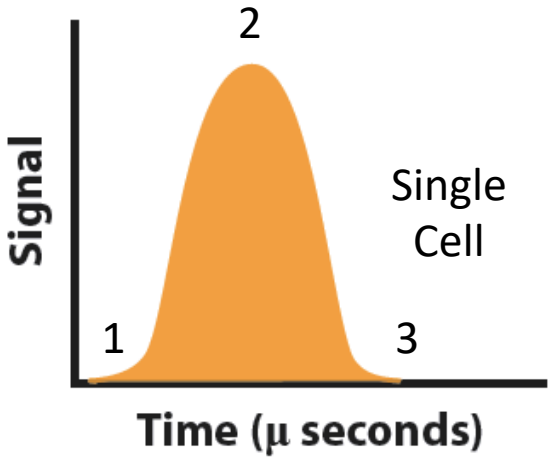


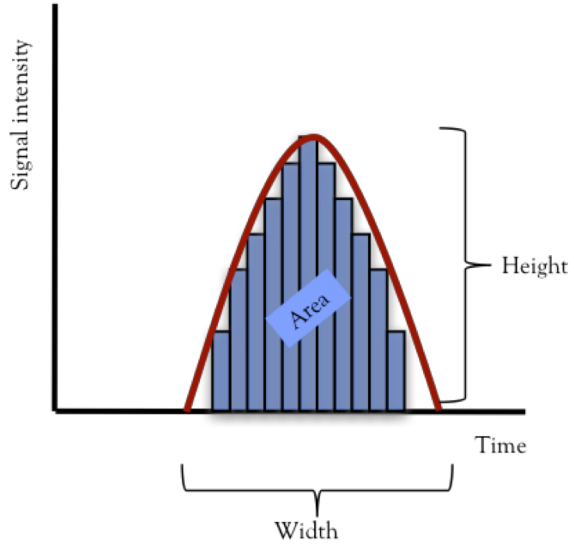


- Helps remove events that are two or more cells stuck together.
- Reduces contribution to false positive or double positive events.

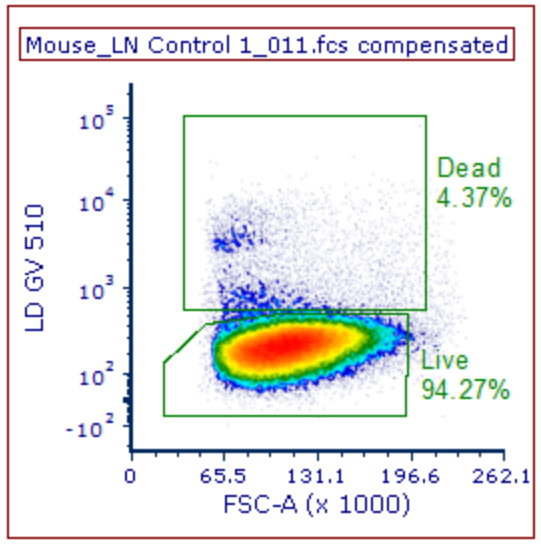
Why Doublet Discrimination Works

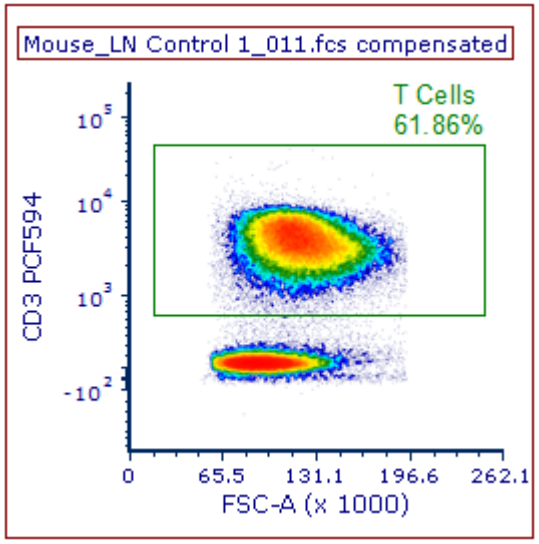




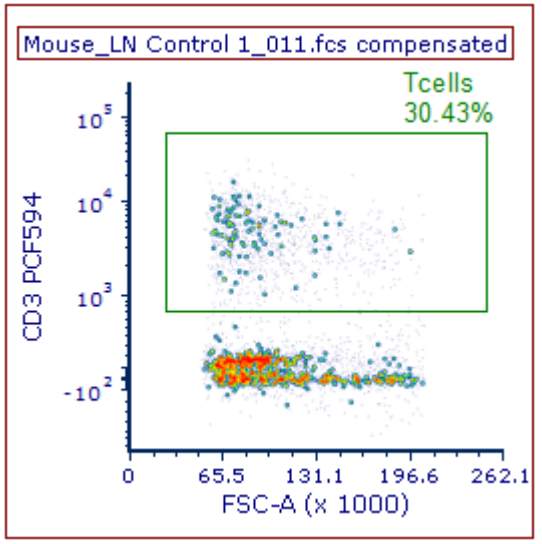


Live/Dead Gating





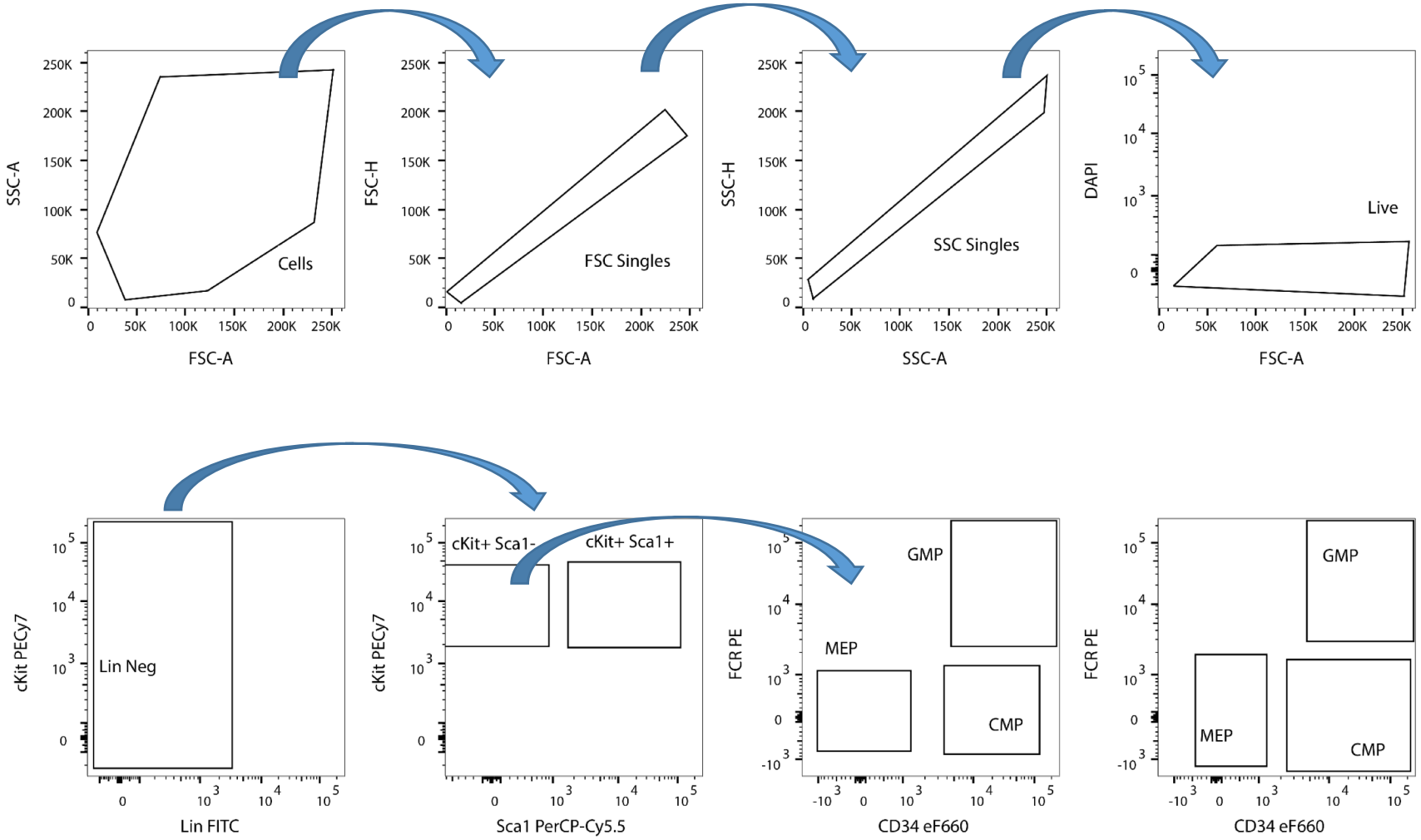
From Live gate



From Dead gate

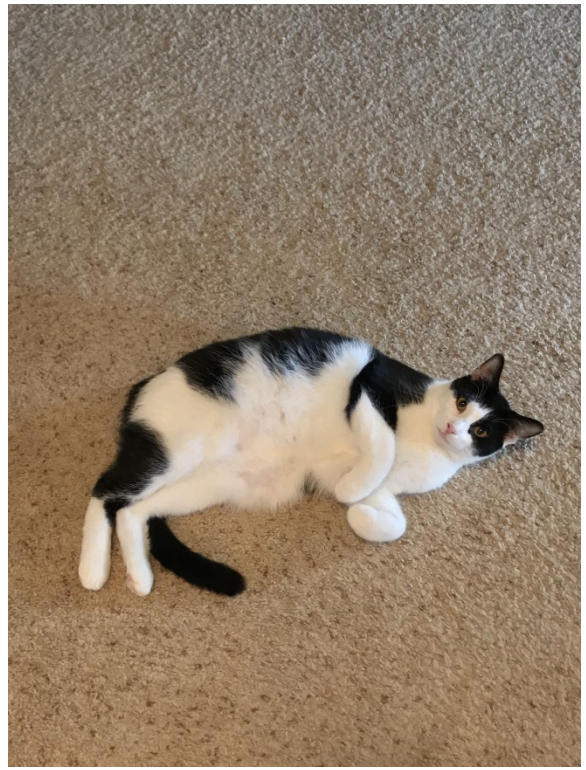
Gating Strategies

- General gating strategy
  - Doesn't have to be in this order.

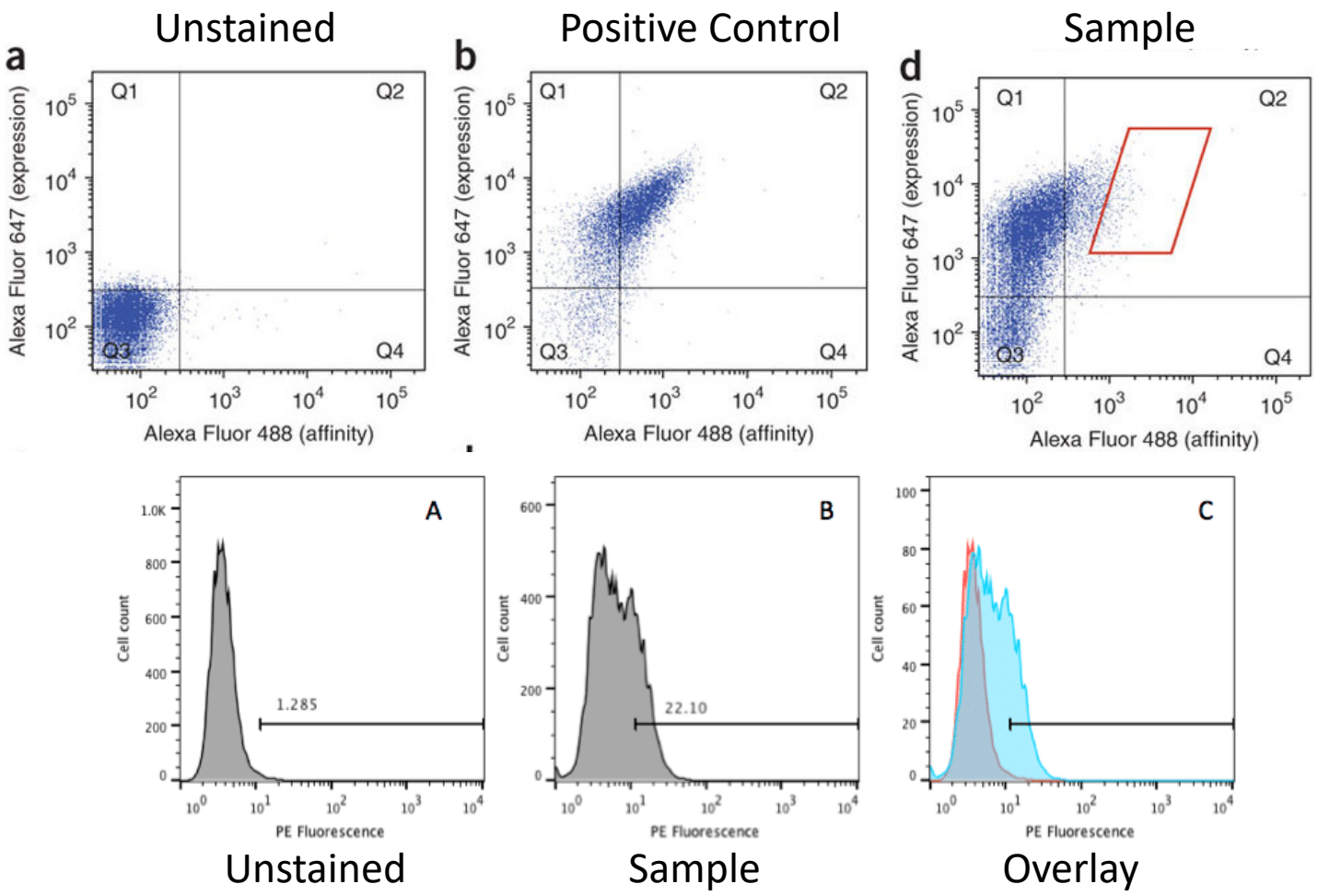


Common Gating Controls

- Fluorescence minus one (FMO) Control
  - Gating control, shows the background and contributions from neighboring fluorescence spillover.
- Positive Control
  - Standardize gating procedure and observe staining profile.
  - Treated to induce positivity.
- Biological Controls
  - Stim vs. Unstim, T0 vs. Time course, Treated vs. Untreated.
  - Any control you need to prove your hypothesis.
- Unstained control
  - To evaluate inherent background and autofluorescence.



Negative and Positive controls slide

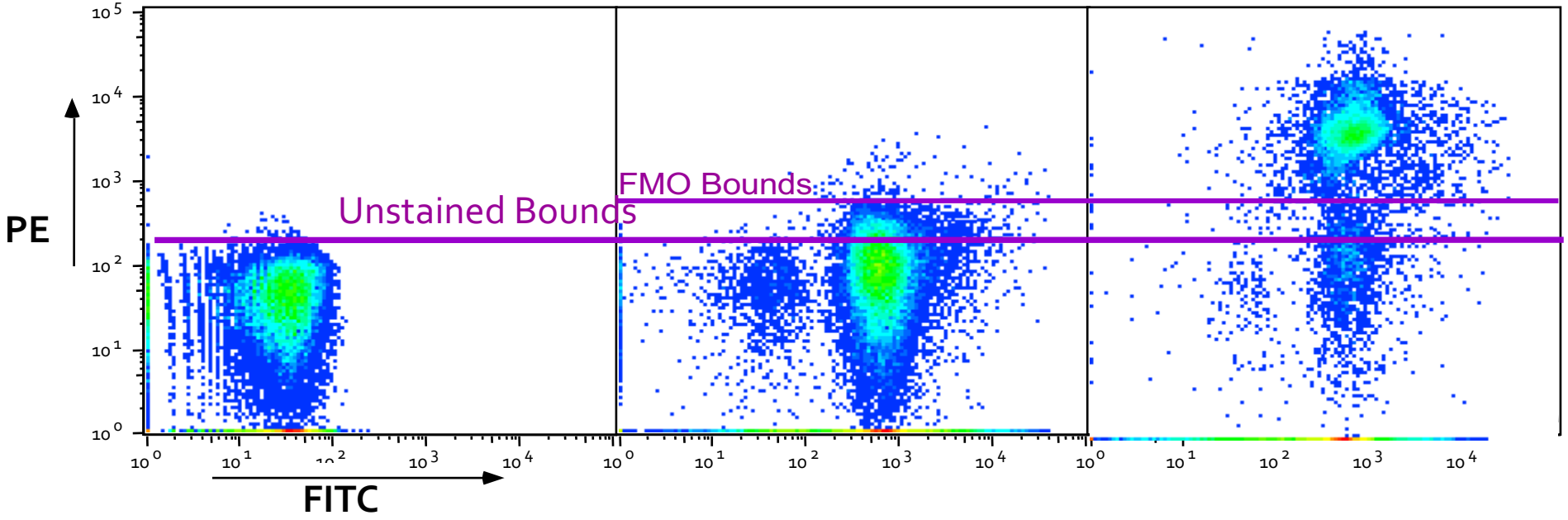


Courtesy of BiteSize Bio

- Positive controls
  - Biological control to assess what the signal looks like when the antigen of interest is present.
  - Useful for rare positive populations or when antigen expression is variable between samples.
- Unstained/Negative control
  - Helps make gating decisions.
    - Visualizing autofluorescence.

FMOs

	Unstained Control	FMO Control	Fully Stained
FITC	**–**	**CD3**	**CD3**
PE	**–**	**–**	**CD4**
Cy5PE	**–**	**CD8**	**CD8**



From Excyte Expert Cytometry (Courtesy of M. Roederer, Ph.D, NIH Vaccine Center)

Why **NOT** Isotype Controls?

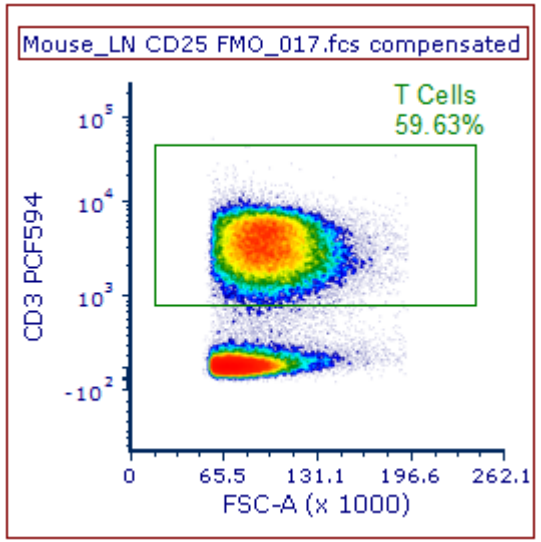
- Nearly impossible to determine if the isotype antibody has the same number of average fluorophores attached per Ab as experimental Ab.
- Different antibody than test sample, different binding properties.
- Maecker HT, and Trotter J. Flow cytometry controls, instrument setup and determination of positivity. Cytometry Part A 2006; 69A:1037–1042
  - Flow Server (T drive) > Flowdata > Flow Resources > Flow References > Isotypes.
- Can use isotype control to test how well the blocking buffer worked.

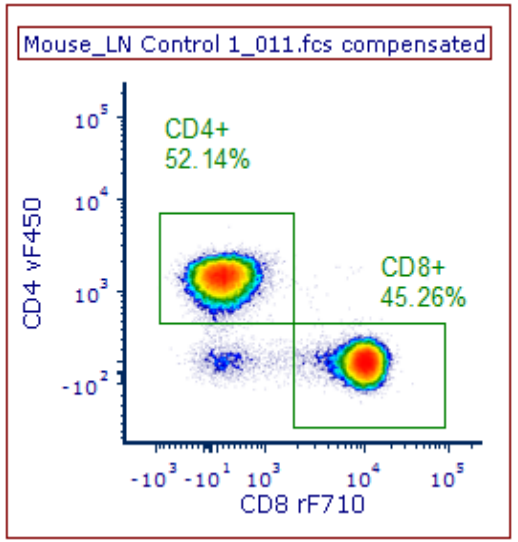
Basic Statistics in Flow Cytometry

- Typically described using frequencies and fluorescence intensity.
- Frequency
  - Number of events in the target population within a larger population.
- MFI (Median Fluorescence Intensity)
  - NOT mean. Mean is subject to outliers, median is less affected.
- Statistical modeling (Following Seminars)
  - DNA Cell Cycle analysis
  - Proliferation analysis
- Absolute counts
  - Volumetric based acquisition cytometer or counting beads spiked in sample at known concentration. (Counting beads tend to be problematic)

Frequency

- Ex. Number of CD4+ cells in a population of Live, single, CD3+ positive cells.
- Used to analyze presence of antigen/marker.





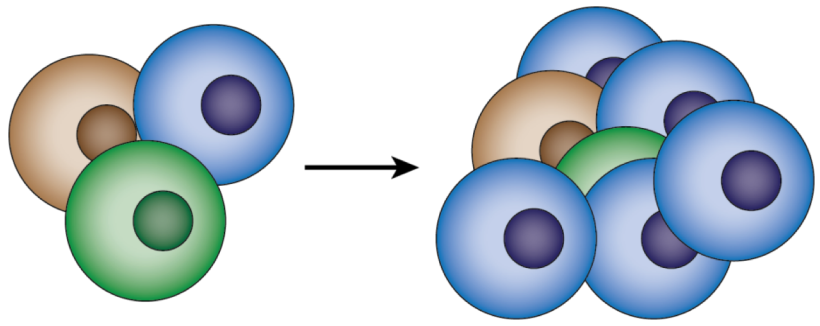
From Live Single Cells

49.81% of CD3+ cells are CD4+
44.86% of CD3+ cells are CD8+

Frequency Hypotheses

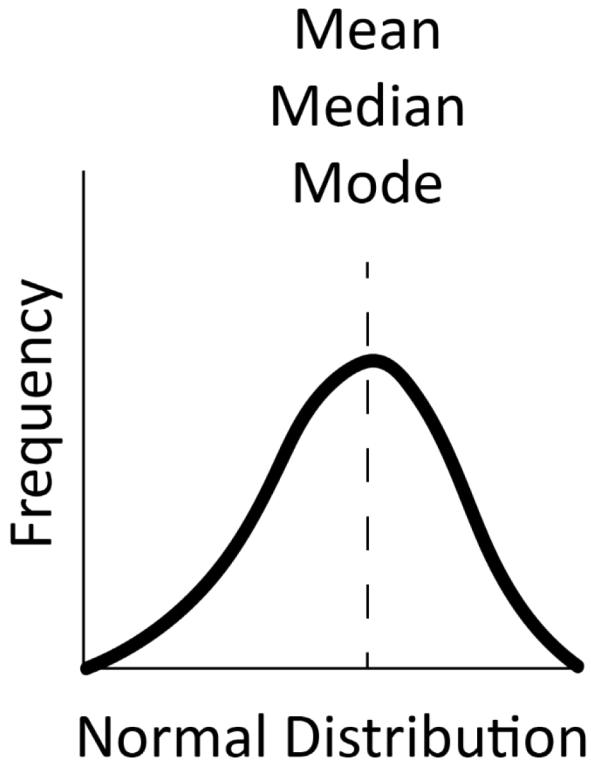
Treatment increases numbers of cell type Y

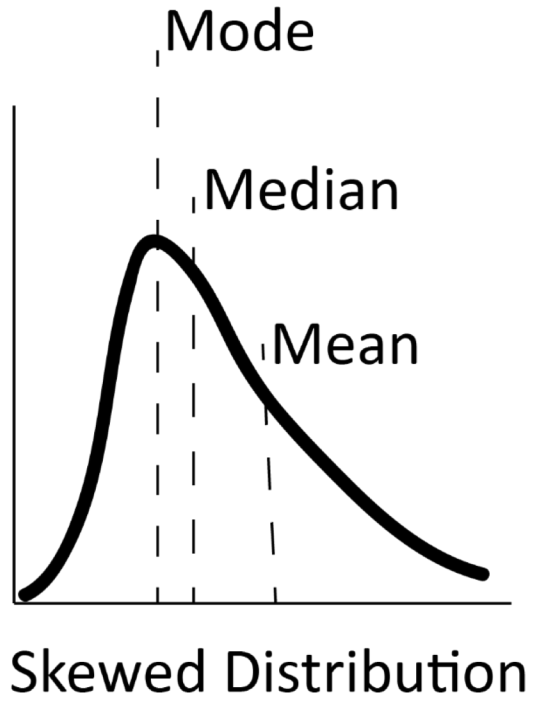
Cell type Y is more prevalent in disease state A



Report % positive to evaluate changes in composition of cell populations

Central Tendency



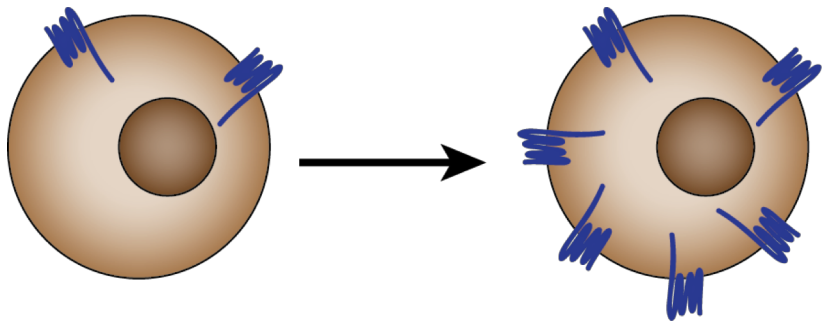


Most flow cytometry data is displayed on a Logarithmic scale –
What looks symmetrical is actually skewed!

Median Fluorescence Intensity

Treatment increases expression of protein X

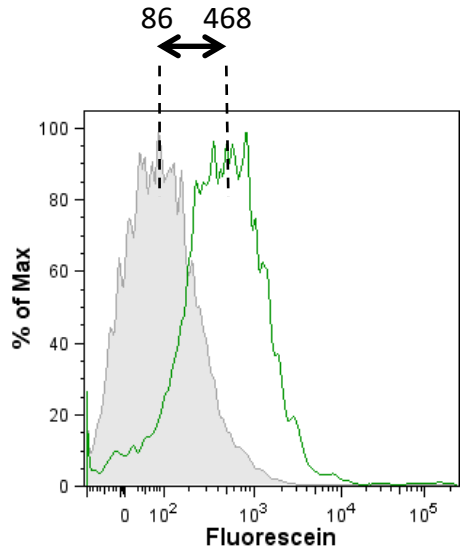
Protein X is upregulated in disease state Y



- Use the MFI to assess *levels* of target protein expression
  - Median for logarithmic data.
  - Mean is ok for linear only.
- Standardizing your assay is critical
  - Reference Standard for PMT sensitivity (Rainbow Beads).
  - Can compare samples run on different days.
  - https://cancer.wisc.edu/research/wp-content/uploads/2017/03/Flow_TechNotes_Rainbow-Standard-Tech-Note_20170918.pdf
- Fold increase?

Fold-change in MFI

- Used in comparison of expression level of antigen/marker between samples.
  - Fold-change in MFI = MFI(sample)/MFI(control)
  - Can compare fold-change in MFI between treatments/samples.



Control MFI = 86
Experimental MFI = 468
Fold-change in MFI = 468/86 = 5.44

Caution:

- In order to use Fold-change in MFI, need to be aware of potential skewing of data due to log scale.
- Small changes in negative can translate into large changes in the fold.

The Data Analysis Process

1. Have a specific Hypothesis. ASK A SPECIFIC QUESTION!
    1. Need to know which statistics you are after.
2. Gate on live, single cells and use controls to gate each fluorescent parameter.
3. Gather statistics from plots and gates.
4. Perform Analyses.

Acknowledgements

- Thank you to the DeLuca Lab for data used in this presentation.

- Excyte Expert Cytometry for graphics used in this presentation.

Mark your calendars for upcoming UWCCC Flow Lab Seminars!

Rigor and Reproducibility in Flow Cytometry
Friday, December 14, 2018
10am, WIMR 7001A

Overview of Computational Data Analysis Platforms for Flow Cytometry
Friday, January 11, 2019
10am, WIMR 7001A

Flow Cytometry – Compensation with Confidence
Friday February 1, 2019
10am, WIMR 7001A

Flow Cytometry Current Best Practices for PIs
Thursday, February 14, 2019
7:30am, WIMR 7170

Multicolor Panel Design for Flow Cytometry
Tuesday, March 5, 2019
2pm, WIMR 7001A

Data Analysis with Alex II
Tuesday, March 7, 2019
10am, WIMR 7170

Data Analysis with Alex III
Wednesday, May 16, 2019
10am, WIMR 7170