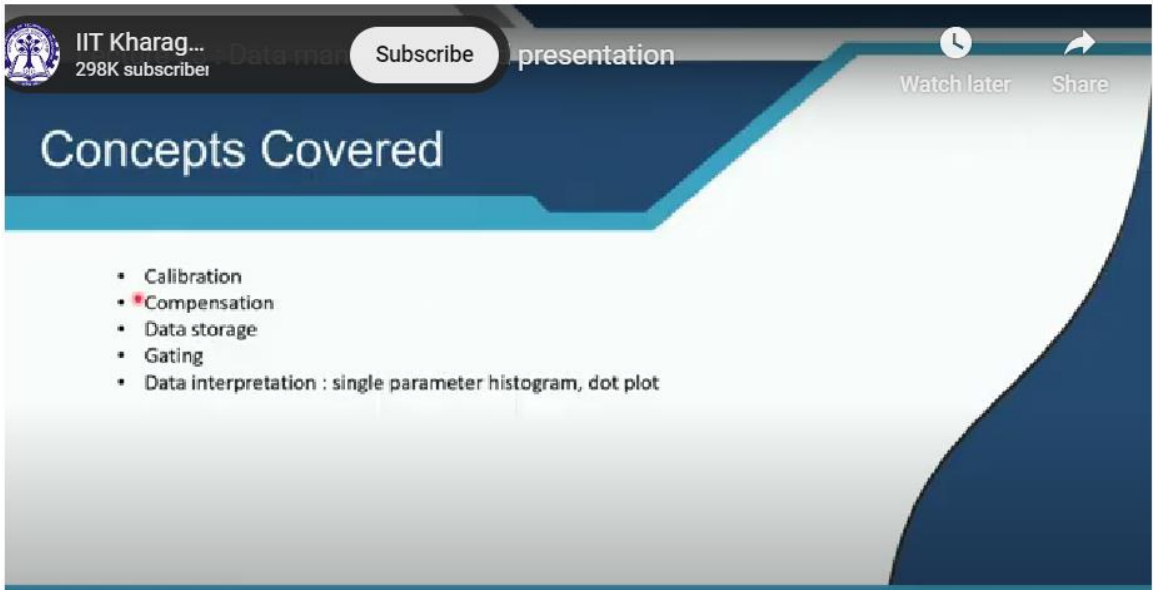
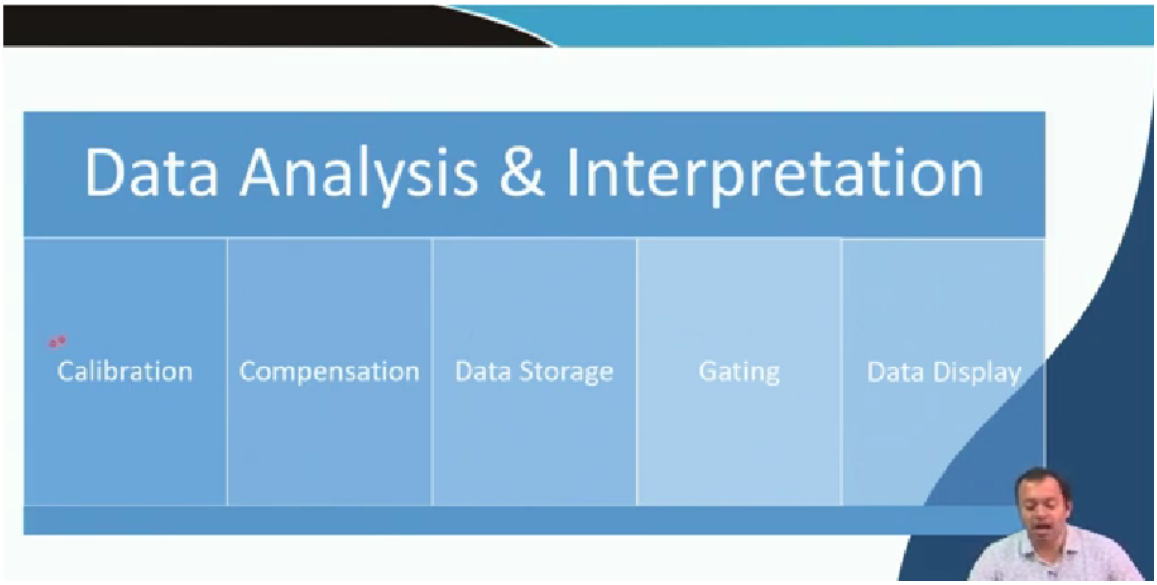
Nanobiophotonics: Touching Our Daily Life
Professor. Basudev Lahiri
Department of Electronics and Electrical Communication Engineering
Indian Institute of Technology, Kharagpur
Lecture No. 23
Data manipulation and presentation

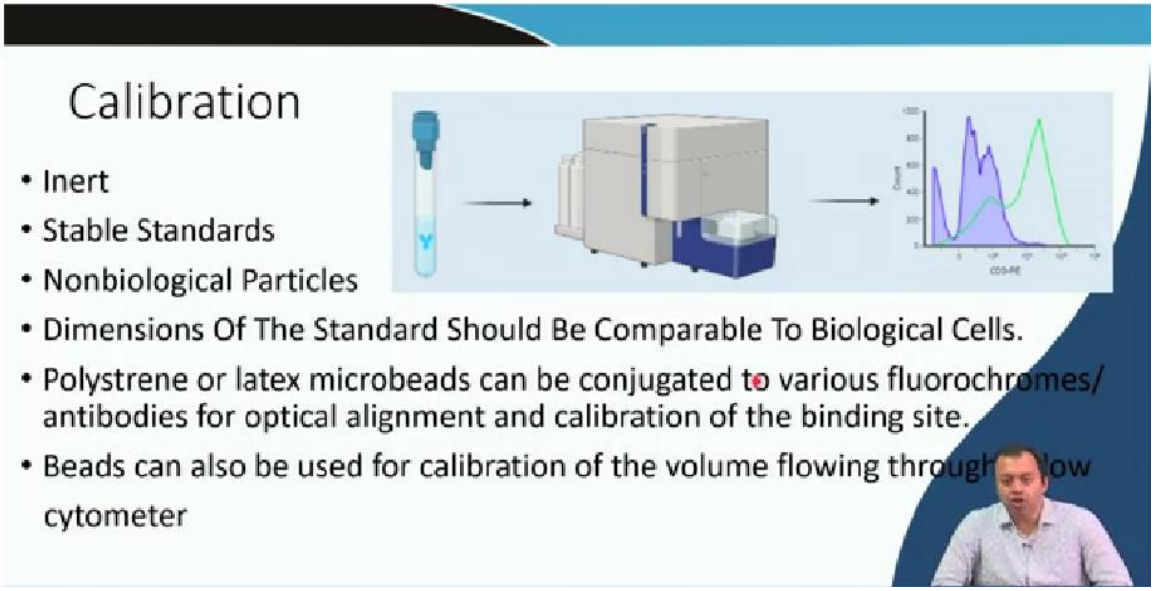
Welcome back. We were discussing flow cytometry, part of the module of biophotonics for disease diagnosis. And in today's lecture, this is going to be a short lecture, we will be discussing data manipulation. How do you present the data, you analyze the data, you manipulate the data and you finally get a clearer picture of what exactly is or how the cells needs to be sorted, how the cells need to be separated.



So, these are going to be the concepts that I will be covering in today's lecture.



So the data analysis and interpretation can be divided into these five different sets calibration, compensation, data storage, gating and finally the display. So, at the end of the day, these things are what is behind the data display. You will be mostly looking at the data which is displayed and from that you will be coming to your own interpretation regarding the total number of good versus bad cells, good, bad, medium cells whereas these four categories will be something that is supporting it from behind. The first and foremost is calibration. In calibration process what we do, we do not put any biological cell or biological structure to begin with.



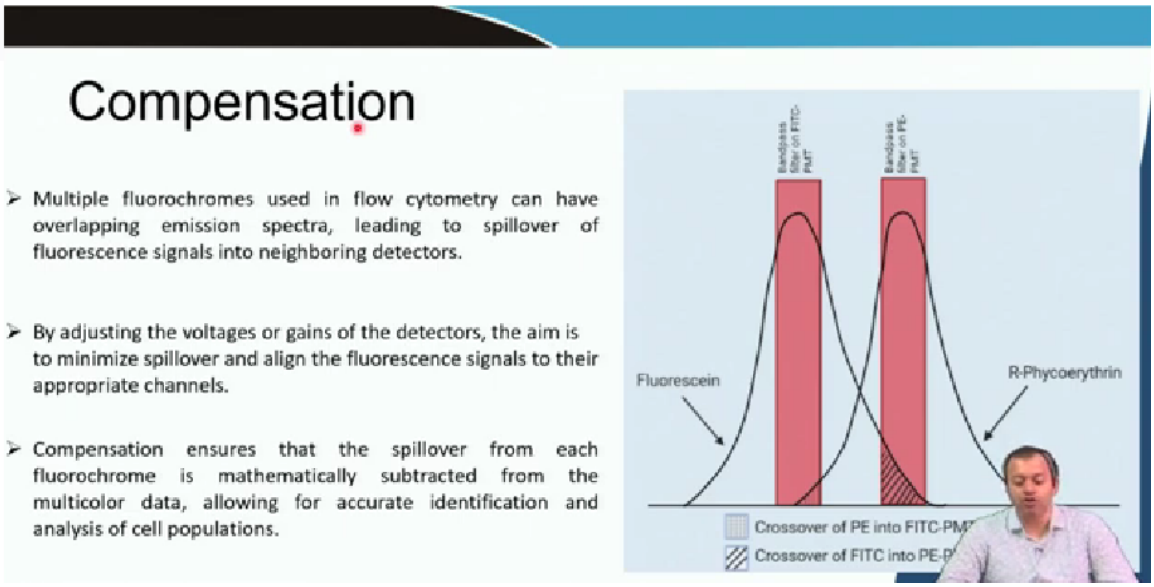
We put some kind of a non-biological inert particle such as polystyrene spheres or silica spheres, they are covered with some kind of a fluorescence particle, fluorescence material, fluorophore and the size of these particles will be similar to the size of the cells, 1 to 10 microns or if you are trying with other different kinds of cells then your cellular particle,

these inert particles will be of similar size. We then perform the fluid dynamics based on these cellular structure. They will be mimicking the property of the cell, the density, the size, we will try to make it as close as possible to cells as close as possible. Obviously, your polystyrene beads or silica sphere will not be you know 100 percent match to any cellular structure, but we start with that. We start with that, we calibrate, we understand, we measure and we control the fluid flow trying to see that if you can maintain the laminar flow, maintain the cylindrical concentric cylindrical column, we can measure one bead at a time.

The dimensions of the standard should be comparable to biological cells. They are coated with fluorophore, they are coated with this fluorophore and then you pass laser light through it and you see the measurement, you calibrate, you focus the laser light, make sure that there is a elliptical beam, make sure that the focus of the laser light is exactly at the center or more or less an elliptical beam covering the center and then the PMTs or the detectors are put in such a manner that all the light that is coming through is all the light that is scattered is then individually measured and you do the calibration. Polystyrene or latex microbeads can be conjugated to various fluorochromes antibodies for optical alignment. So, basically what you are doing with this non biological inert particles is alignment, optical alignment, what should be the focus of the laser light, how close it should be to the tube, how far away should be the detector, at what angle should the detectors be present, what should be the sensitivity, the amplification value of the detector, how you are trying to convert or how do you prevent any additional external electromagnetic field, external electromagnetic light to pass through it, what kind of optical filters that you are going to use, all of those things all of those things are to be given are to be done in this calibration process. Beads can also be used for calibration of the volume flowing through the flow cytometer as I said the total number of polystyrene beads that needs to pass through one at a time through the center by changing the concentration, by changing the velocity, by changing the differential pressure of the sheath liquid, all of those things are calibrated using inert non biological particle mimicking the biological cells.

At the very first instance you do not put a biological cell which is costly you might not know what you are looking for, but you calibrate, you standardized, you normalize the entire process using non biological inert particles which do not have any biological response of their own which will simply scattered their light accordingly and that will form the reference that will form the 0 value with respect to which you will calculate the scattered light coming out of cells. So this is fundamentally the most important part, if the calibration is wrong, if your reference point is wrong, if your 0 0 0 point is wrong in a three dimensional Cartesian coordinates all your measurement fails. So this is how you

normalize, you rationalize, you standardize the entire measurement system and then comes the compensation.



Multiple fluorochromes used in flow cytometry can have overlapping emission spectra leading to spillover. So what does that mean? So suppose you are sending a blue light, it is illuminating a fluorochrome that is attached with a cell that fluorochrome absorb that fluorophore absorb the blue light and emit red light.

Yeah that is the standard procedure. Now understand that there are two liquids, the sheet liquid and the core liquid. At the same time the 100% amount of blue light will not be absorbed and converted into red light. You will not have 100% quantum efficiency. At the same time what you are trying to see how much of red light is in the output is produced, the intensity of the red light.

So chances are a substantial amount of blue light still exist which is not absorbed or which has been absorbed differently or which has been scattered differently by the sheet liquid and the output is a combination of red light, blue light, any other kind of light. A substantial amount of blue input light will not be absorbed and will be present in the output causing some kind of a problem. So what do you do? You put an optical filter, a bandpass filter. What is a bandpass filter? Biologics and electronics guys are rolling their eyes. I know stating that everybody know what a bandpass filter is.

Biology may or may not know it. Few days ago you did not know what central dogma is. Every single biologist know what a central dogma is. So do not laugh or do not make fun of bandpass filter. Bandpass filter allows a particular band of light to pass through.

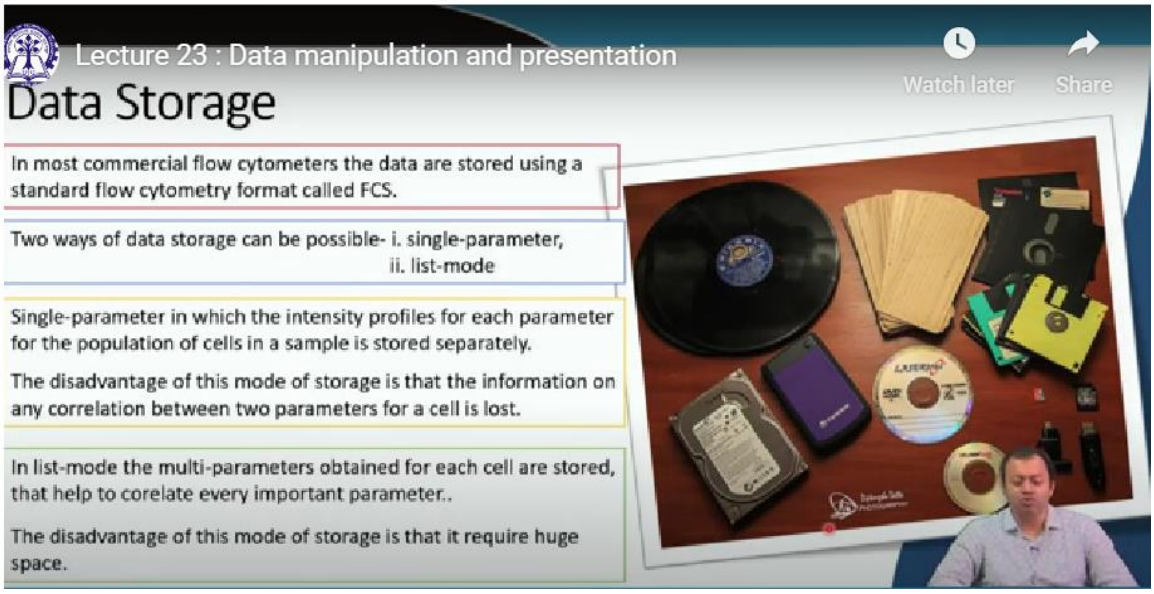
It can be used in electronics as well, but in optically I mean a particular band of light is allowed to pass through. So you have visible spectrum, VIBGYOR surrounded by straddle between ultraviolet and infrared. So a bandpass filter like a prism mostly allows the 7 colors of visible spectrum to pass through. Rest of them can be blocked. Prism can also allow other things to pass through, but I am giving you an example.

So a bandpass filter is allowing a specific set of wavelengths to pass through depending on the wavelength that is the range of your band. You can allow only red light to pass through, only red light to pass through or you allow red and green to pass through, blue and lower wavelengths all of them are simply cut off. So compensation has to be done. You need to put filters, you need to put bandpass filters and thereby you try to figure out a specific group of frequencies, group of wavelength of light of your interest coming from the total output, the total scattering that has happened because of the blue laser light or green laser light that is illuminating your sample. At the same time if you want to do a complex measurement, you are not only trying to see just the size of the cell, but you want to see what are the concentration of different proteins.

There are different types of proteins inside the cell, protein A, protein B, protein C, protein D. You have marked, you have leveled each one of them with specific antibodies connected with specific fluorophore emitting specific wavelengths. Protein A is connected with antibody A connected with fluorophore which emits red light. Next is green light. Next is yellow light.

Next is orange light. And then a simple blue light is in the input. You need to understand the concentration of these four different lights at the output. You do not want a combination, a mixture of all four light together. You do not want blue light to come into effect.

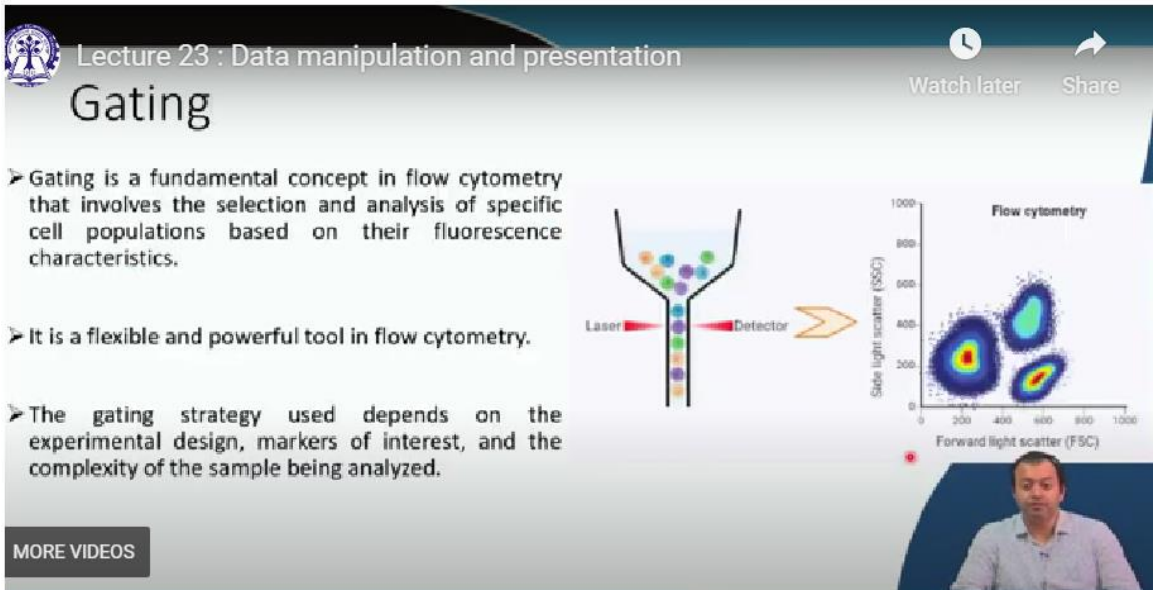
You do not need all these other noises electromagnetism to come through. So thereby you need to have the, you adjust the voltages of the gain of the detector. You mathematically subtract, you do some kind of a digital calculation, you extract all the information out, then you take one information at a time subtract one with respect to other the background subtraction method or you use a band pass filter to analyze and identify individual cell population. This is also very very important because at the end of the day if you are unable to do compensation, if you are unable to mathematically subtract or do this measurement, then you have a very very noisy data where you are unable to differentiate between what is present in what amount.



Then of course there is data storage, the most commercial flow cytometer the data is stored using a standard flow cytometry format called FCS.

Two ways of data storage is possible, single parameter or list mode. Single parameter is in which the intensity profiles of each parameter for the population of cell sample is stored separately. In the list mode the multi parameters obtained from each cells are stored, help to correlate. So, list mode have several different parameters, size, granularity, all of different how much amount of a particular protein A, B and C, single parameter as the name suggests big size small size. The disadvantage is that the information on any correlation between two parameters for a cell is loss.

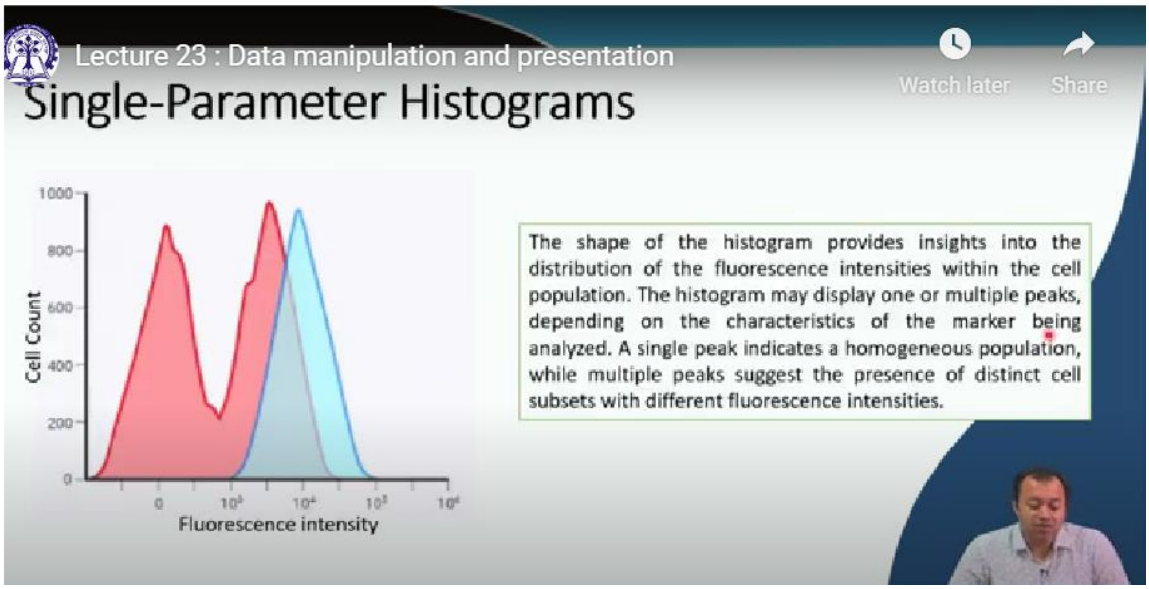
It is big because this protein is less or big because this granularity has formed that is not present. The disadvantage of list mode on the other hand is it is required huge huge amount of space, external hard drives and what not.



Finally we get into the gating part. Gating is a fundamental concept in flow cytometry that involves the selection and analysis of specific cell population based on their fluorescence characteristics. It is flexible and powerful tool.

The gating strategy used depends on the experimental design, markets of interest and the complexity. Basically the overall software which makes this sort of measurement. Forward scattered light, side scattered light, total number of cell count. How many cell count which has given is proper arrangement. So, the laser light is illuminating the cell, you have measured both the side scattered, you have measured the forward light scattered, you have plotted it, you have plotted it for every hit.

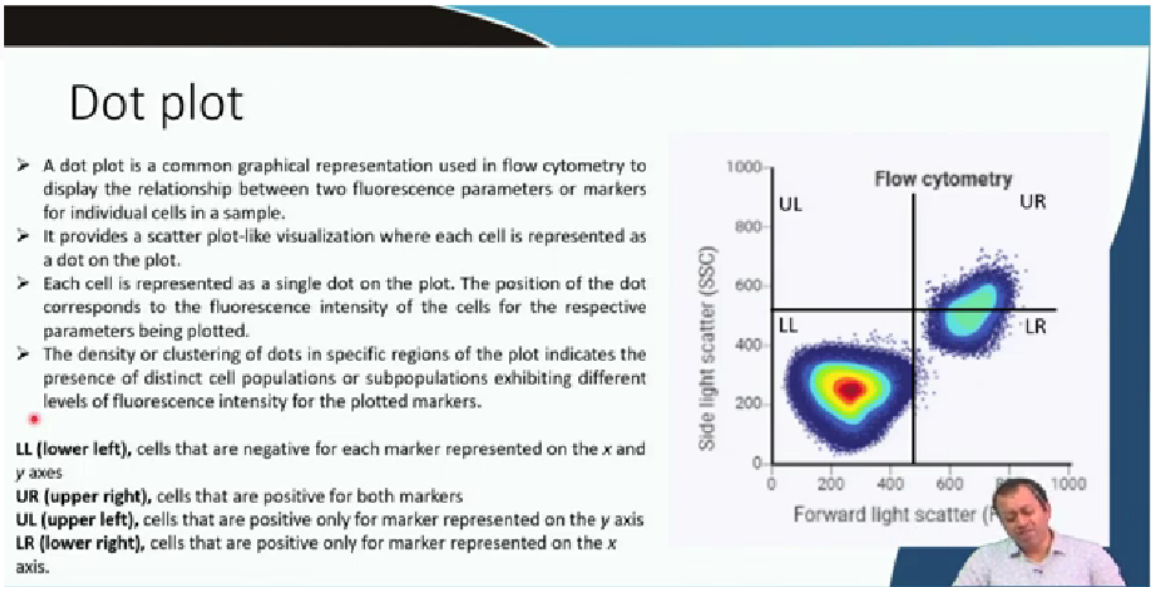
Every time a particular cell is illuminated you are measuring two different parameters, the forward scattered and the side scattered of the same particular measurement trying to figure out what exactly is going on and you are putting a colorful plot where there should be a color chart which determines a particular protein, intensity of this protein of this many numbers are present here as compared to the other measurements.



How do you interpret the data? Cell count versus fluorescence intensity, this is the easiest form. What is the intensity and how many total number of cells produce that intensity? So for example, red light, so if this is the calibrated value, the reference value and if is your original value then you can have 1000 number of cells giving an intensity of 10^4. The shape of the histogram provides insight onto the distribution, the population distribution of the fluorescence intensity within the cell population. The histogram may display one or multiple peaks depending on the characteristics of the marker being analyzed.

A single peak indicates a homogeneous population while multiple peaks suggest that the presence of distinct cell subsets with different fluorescence intensities. It is self-explanatory, I do not have to analyze or explain this much. Total number of cells at what particular intensity they are fluorescence. This cell give this fluorescence value, another cell next cell give this fluorescence value, next cell give this fluorescence value, so on and so forth. You make a chart of total number of cells versus total number of fluorescence value.

From that you interpret whether the cell is good or bad. Cell peak gives homogeneous data, most of them are good, most of them are sending normal information. Multiple peaks have multiple different types of cells in homogeneous, so several things might have gone wrong with it or the cell structure that you have extracted from the human might have been contaminated, you have dust particles, bacterias or some other kind of pathogens are presence.



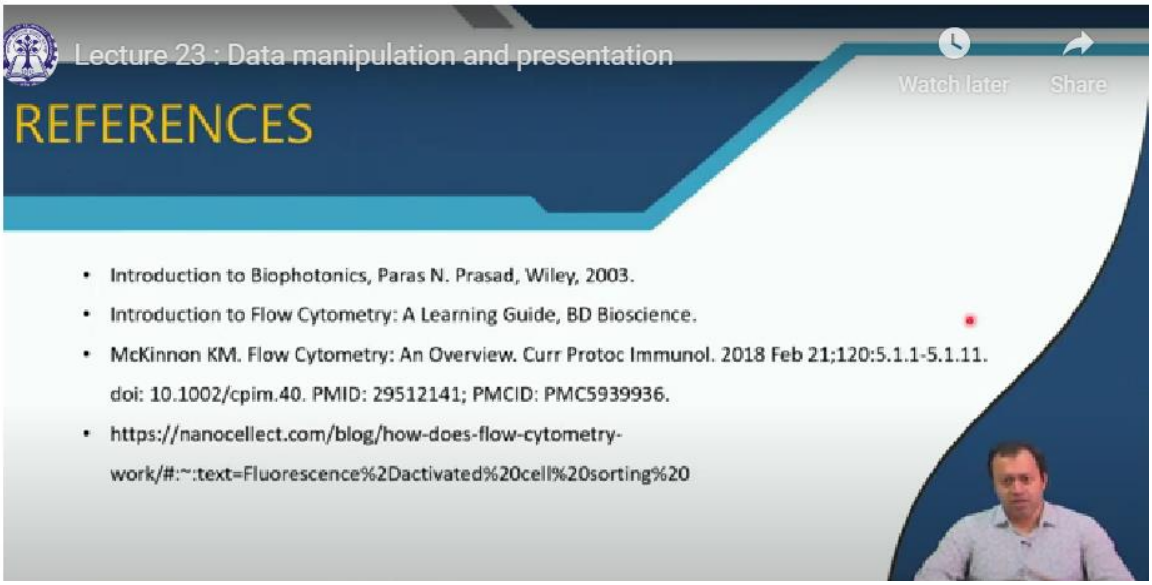
Another thing is dot plot. Dot plot is very similar to the one which I showed before, it is just divide the forward light scatter versus the side light scattered into four coordinates, lower left, upper right, upper left and lower left.

Cells that are negative for each marker represented by x and y axis, cells that are positive for both markers, cells that are positive for only one type of marker, cells that are positive only for another type of marker. So, you have put multiple at least two different type of fluorophore in trying to understand two different type of proteins in the same cell or two different types of anti-J pathogens in the same cell. A dot plot is common graphical representation used in flow cytometry to display the relationship between two fluorescence parameters or markers for individual cells, two different proteins or two different antibodies or two different types of pathogens. It provides a scatter plot like visualization where each cell is represented as a dot on the plot, each cell is represented by, so how many cells you can understand. Dot in the plot, each cell is represented as single dot, the position of the dot corresponds to the fluorescence intensity of the cell for respective parameters being plotted and the density of the clustering of the dots in specific region of the plot indicates that the presence of distinct cell population or sub population.

So, this is mostly dense in the LL, LL stands for cells that are negative for each marker represented on the x and y axis. So, most of your cells that you have measured are negative for both of the biomarker. I do not know what they were trying to measure, but most of the cells are in the LL coordinate value meaning most of the cell are not having the biomarker that you are looking for. Probably you are looking for a particular disease or two different types of disease, most of the cells are not here. UR and LR stands for positive for both markers and LR stands for cells which are positive only for another type of marker.

So, from the total plethora, total amount of cell that you have extracted you have got, a majority of them are negative for both, a small portion is positive for both and similar small portion is positive for just one. So, that is it. This overall information gives you that the biomarkers are missing in majority of cell. So, this can be used to determine so called the viral load. Say you are looking for a particular antibody, you are trying to look for a particular pathogen and thereby the antibody released by this pathogen because of the presence of this pathogen.

So, how much of the cells are actually releasing this antibody or how much of the particular amount of cells have been infected by a particular virus can overall help us determine viral load and by measuring the viral load we can thereby determine what dose of medicine needs to be prescribed. 2 tablets per day or 3 tablets per day or you do not need any tablet you need direct injection that is determined by how much your immune system has been compromised. So, this determines this versus this total negative versus some amount of positive overall determines how much of the cells have had showing a particular set of markers, group of markers that you are trying to analyze. So that is basically it.



In the next two classes I will be giving you specific examples of specific diseases that has been detected by cell cytometry and not just specific diseases we will also see that if we can send lasers instead of fluorescence can we do other type of scattering spectroscopy for example Raman scattering spectroscopy.

So, please go through my references and

these are my conclusions that you can read at your own leisure time and I will see you in the next class. Thank you very much.