flowPloidy: Flow Cytometry Histograms

Tyler Smith

2019-03-20

Introduction

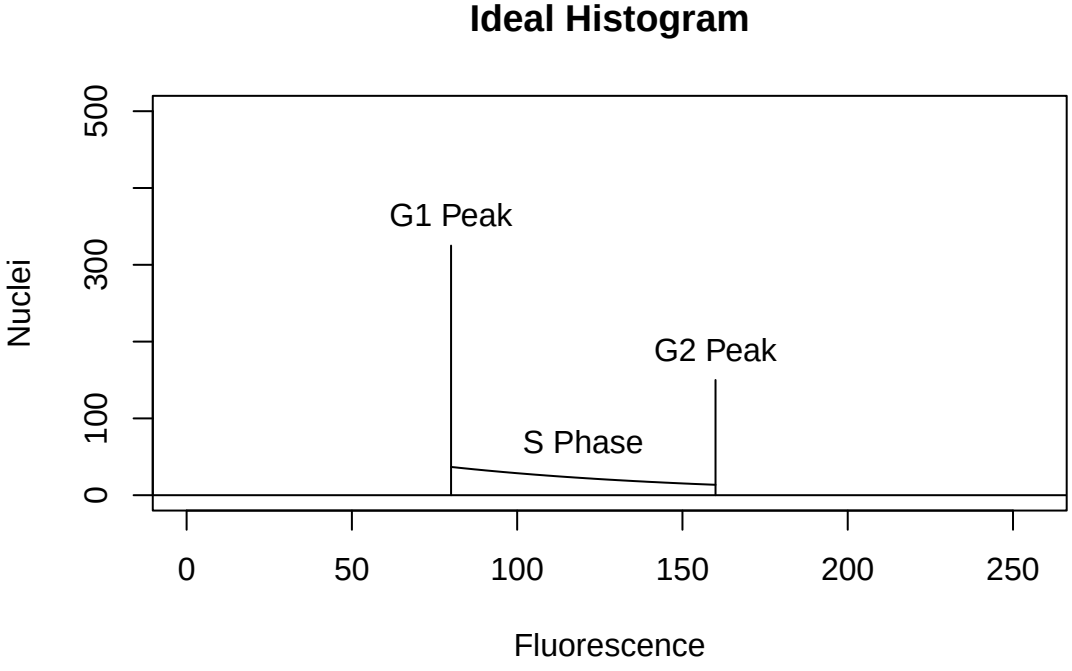
The application of flow cytometry (FCM) to ploidy assessment is based on a simple concept. Cell nuclei are stained with DNA-binding fluorochromes, and the amount of DNA present is determined by measuring the fluorescence emitted when the nuclei are excited with light of a particular wavelength. Greater fluorescence means more DNA in the nucleus. That is, a diploid nucleus should produce half the fluorescence of a tetraploid nucleus, and a hexaploid nucleus should produce three times more fluorescence than the diploid.

Histogram Basics

Applying this concept is complicated by several factors:

- nuclear DNA content varies over the cell cycle
- secondary metabolites may interfere with fluorochrome staining, or may themselves fluoresce
- tissue preparations are usually contaminated with non-nuclear cellular components, damaged nuclei, and other debris that fluoresce to greater or lesser extent
- the accuracy of flow cytometers is limited, and values tend to shift or drift over the course of a run

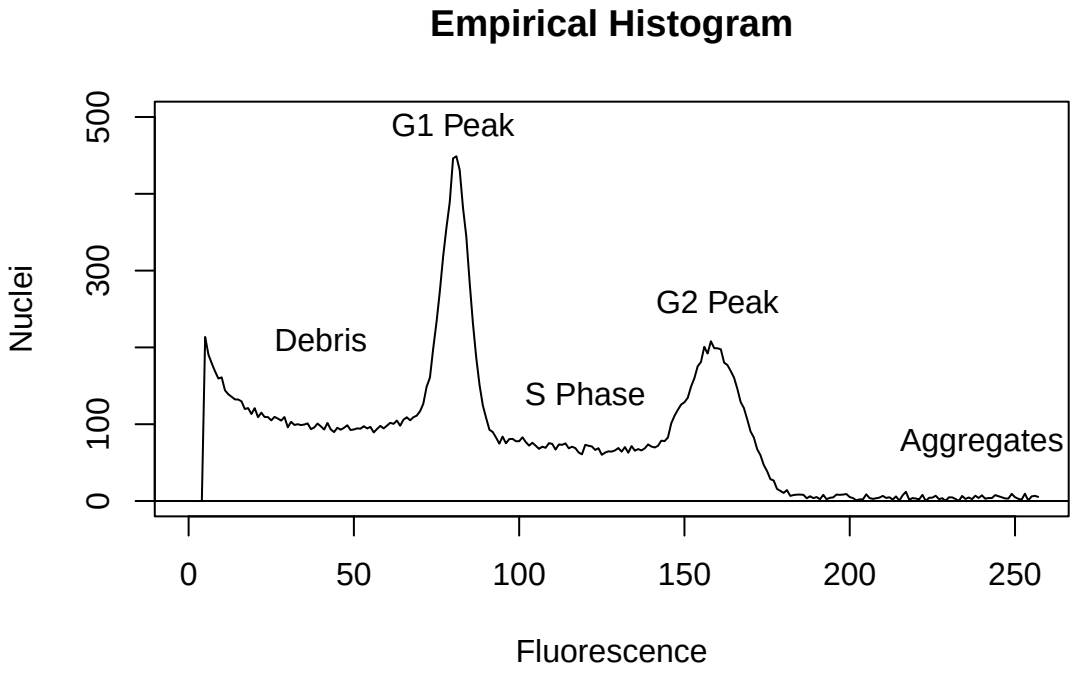
We can illustrate these challenges by comparing an ideal FCM DNA histogram with simulated experimental data.



In the ideal case, we have two clear, accurate cell groups, G1 and G2. The G1 ("gap 1") group represents cells in their 'normal' state. That is, they have the usual diploid DNA complement. The G2 ("gap 2") group represents cells that have duplicated their DNA, but haven't yet undergone mitosis. Consequently, they have twice the diploid DNA complement. In between the two groups is the S phase ("synthesis") region. This area of the plot represents cells that are in the process of duplicating their DNA, in preparation for mitosis.

(N.B. in the flow cytometry literature, clusters of cells, which I've called "groups" above, are often referred to as cell "populations")

If our histogram were this clean, we could easily determine the DNA content of our sample by comparing it to a known standard (see below). In practice, you'll never see a histogram that looks like this. A more realistic example follows:

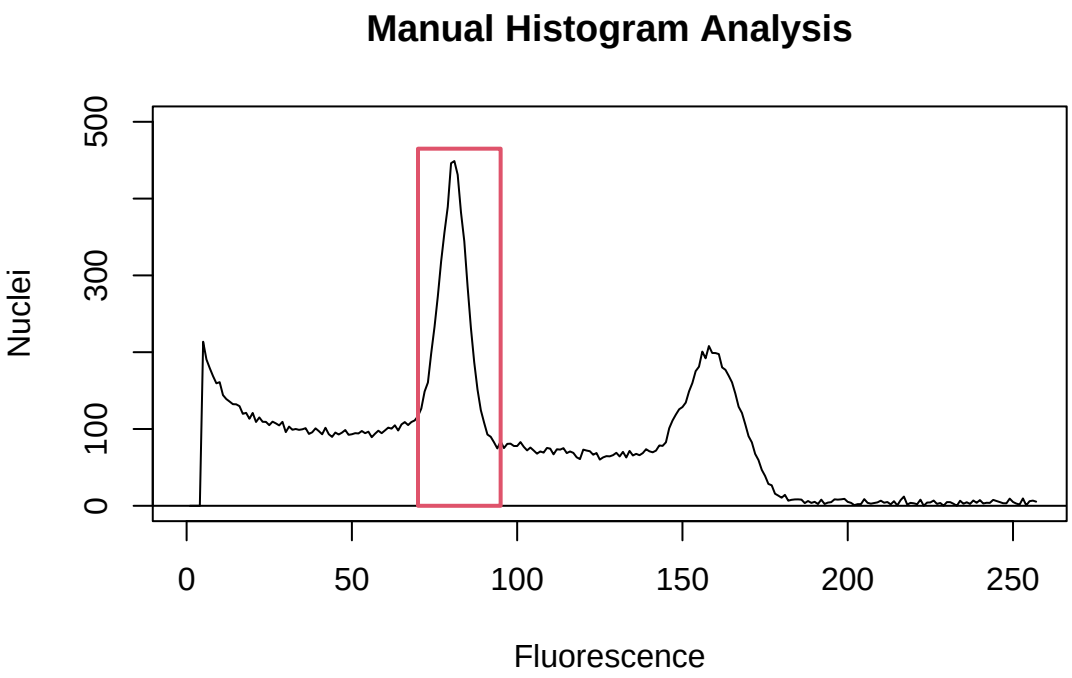


We can still see the G1 and G2 groups, but they aren't exact values. Instead we see that the fluorescence values for each group are distributed around a central peak. Similarly, the S phase is not a clean polygon. The jagged edges of the region are due in part to instrument error: we can't capture DNA content perfectly. In addition, some of that noise is "real". That is, the cells in our tissue sample are not proceeding through the cell cycle in a perfectly regular manner. Even if we could measure the DNA without error, the S-phase region wouldn't necessarily be a perfectly smooth region in the histogram.

In addition to the G1, G2, and S phase regions from our ideal histogram, real histograms also contain debris. This is a mixture of cellular debris, damaged nuclei, and other contaminants. The dyes we use bind best to DNA, so most of the debris, which contains relatively little DNA, is found on the left/lower fluorescence side of the histogram. We may also see aggregates at the right side of the plot. These are nuclei that have stuck together in the flow cytometer, and so appear as higher-ploidy nuclei.

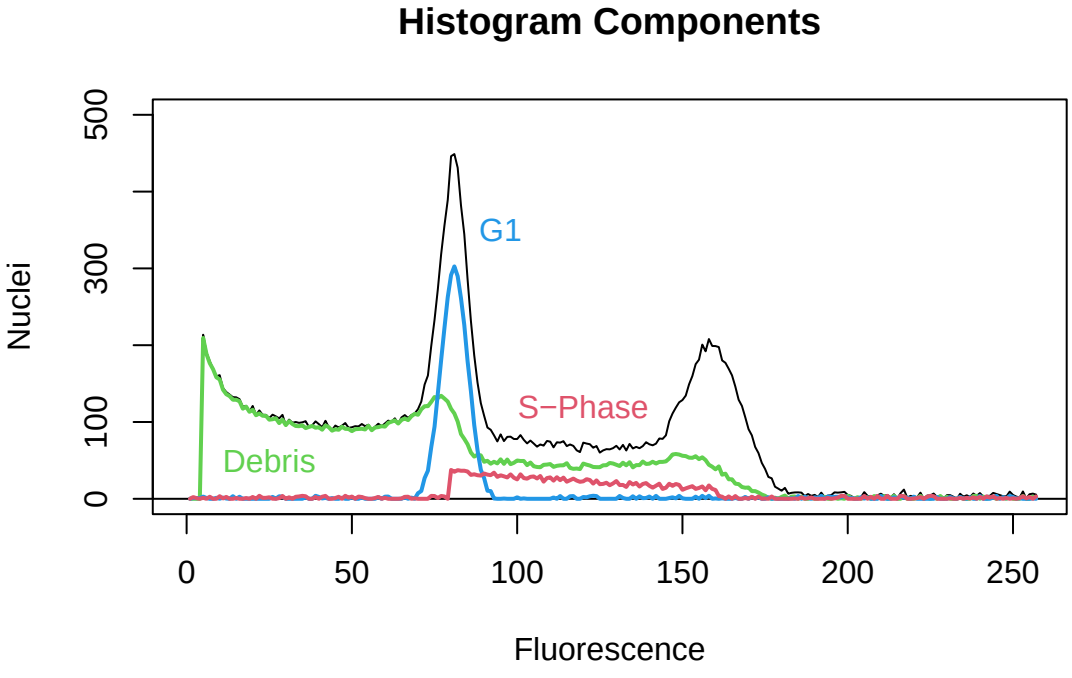
Histogram Analysis

Given that we have a noisy histogram, how do we extract the values of interest from the data? There are two main approaches. The first, perhaps more intuitive, approach requires us to manually draw boxes around the peaks we're interested in, and have the computer calculate the parameters of interest for us. For the previous example, that would look like this:



There are two main problems with this approach. First, the placement of the box around a peak is subjective. One of the most important quality-checks for ploidy analysis is the coefficient of variation (CV) for a peak. We can easily lower the CV just by drawing a narrower box around our data. Consequently, our confidence in our results is undermined by not having an objective and repeatable method for setting the limits around a peak.

The second problem is that we are not accounting for the different types of cells that contribute to each peak. What we see as the G1 peak also contains debris and S-phase nuclei:

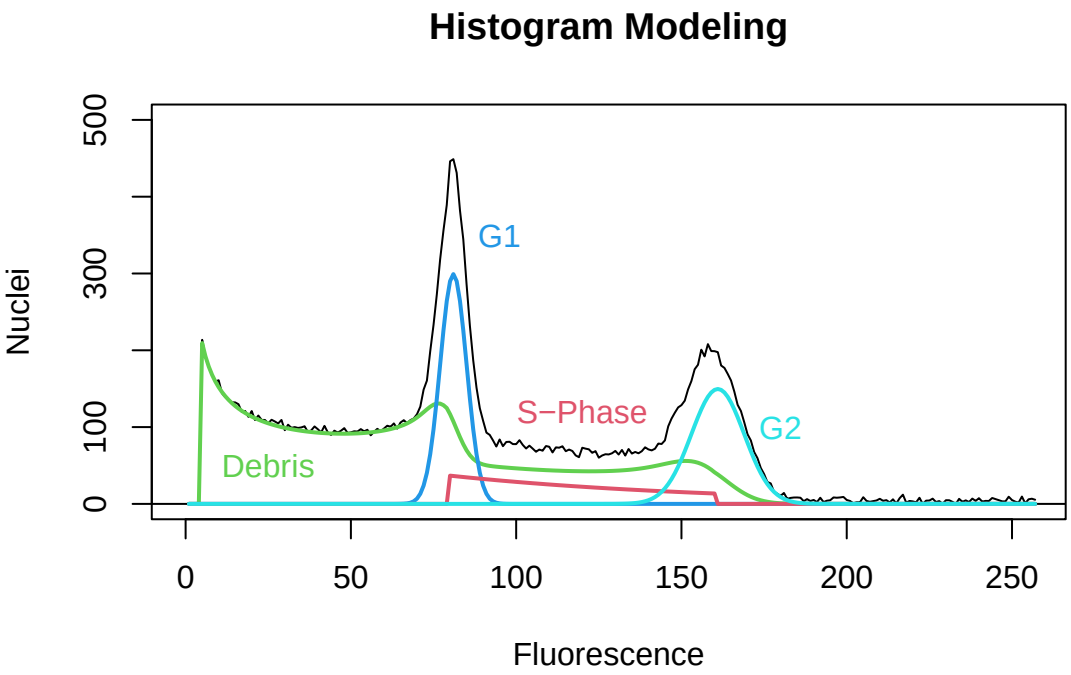


When we draw a box around the G1 peak, we are also including debris and s-phase cells, which distorts the true distribution of the G1 nuclei DNA content.

The second approach to FCM histogram analysis uses non-linear regression to distinguish the different

cell populations that contribute to the total histogram. The G1 and G2 peaks are modeled as normal curves. A variety of phenomenological and mechanistic functions are available to model the debris and S-phase components. See Bagwell (1993) and Watson (1992) for a more detailed presentation of the modeling process.

Using the modeling approach, the analysis of our example data looks like this:



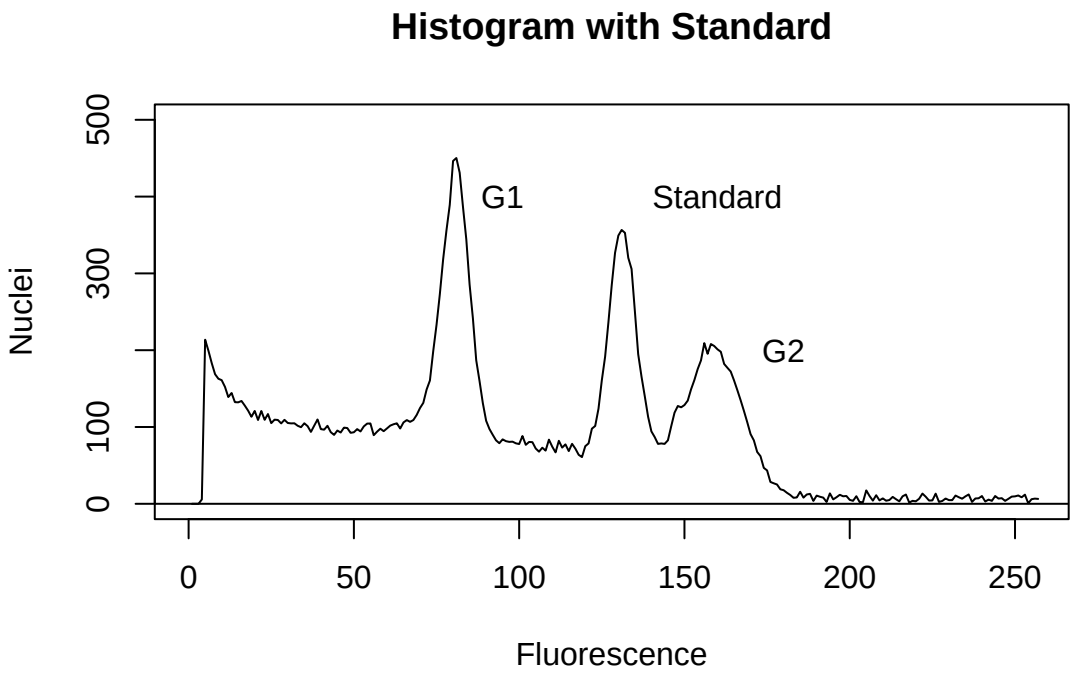
This is a simulated example, but it illustrates the modeling process very clearly. We can extract the values of interest directly from the model components. The regression results include the mean value for the G1 peak (80), as well as its CV (5%). These values have taken into account the contribution of the debris curve and the s-phase nuclei, and we'll get the same results every time - they don't depend on where we decide to draw the box around each peak.

Standards

To simplify the presentation of histograms in the previous sections, I omitted the internal standards used in FCM analysis. This is a critical component of ploidy analysis, because the fluorescence values for a given quantity of DNA are not fixed. The measured values vary based on the tissue preparation and the flow cytometer. Even for the same tissue prep run on the same flow cytometer, the fluorescence values can vary over the course of a single day.

If we were to re-run the sample that yielded a mean value of 80 in our example, we might get a different result. Which means we can't interpret the raw fluorescence values. Instead, we include a tissue standard in our preparations, so we can compare our sample peak to the known value. While the absolute fluorescence values may shift over time, the *ratio* between our sample mean and the standard mean will stay the same.

Adding a standard to our example data, we get a histogram that looks like this:



To analyze this histogram, we add another normal curve to account for the standard sample. That allows us to compare our sample mean (80) to the standard mean (130). If the standard was *Solanum lycopersicum*, with a known 2C value of 1.96pg, then our sample has a 2C value of $\frac{80}{130} \times 1.96pg = 1.21pg$.

References

Bagwell, C. Bruce. 1993. "Theoretical Aspects of Flow Cytometry Data Analysis." In *Clinical Flow Cytometry: Principles and Applications*, edited by Kennth D. Bauer, Ricardo E. Duque, and T. Vincent Shankey, 41–61. Williams & Wilkins.

Watson, James V. 1992. *Flow Cytometry Data Analaysis*. Cambridge University Press.