Multi-parametric Cell Cycle Analysis: A comparison of transition state-related cluster and probability state analyses

Christopher M. Bray[1], C. Bruce Bagwell[1], R. Michael Sramkoski[2], Tammy Stefan[2], James W. Jacobberger[2]

1 Verity Software House; 2 Case Comprehensive Cancer Center, Case Western Reserve University

Program Number: 136

Abstract

Cytometry-based cell cycle analyses rely on identification of data clusters identified by the level of expression of biochemicals and the frequency at which cells with specific levels occur within cell populations. The simplest case is DNA content analysis of euploid cells (G1 = 2C DNA, G2 and M = 4C DNA, and S = every level in between). This case can be extended in an informative manner to include cell cycle markers that oscillate within one cell cycle. If the markers are chosen appropriately, the cell cycle becomes a closed loop through n-dimensional data space. The cell cycle programmed expression profile for each marker or ancillary markers can be extracted using the cell frequency as a surrogate for time. This is referred to as transition state-related cluster analysis (TSCA).

A Probability State Model can be used to re-order data from a cytometric assay according to the expected progression through a biological process. This re-ordering can follow the time component of a process but is not bound by it – e.g., the model can run backwards or forward through the process, even though the process can only move forward through time. In probability state modeling as practiced here, the length of each ordered data segment is proportional to the probability of finding the cells at a specific level of the biological marker. Thus, expression profiles and probability models are conceptually equivalent.

Here we compare transition state analysis to probability state modeling for the cell cycle by comparing expression profiles for DNA content, cyclin A2, cyclin B1, and phoshpo-S10-histone H3 to probability state models for the same markers. Two approaches are compared. In the first approach, the probability state model is defined based on verbal or cartoon descriptions of expression profiles. In the second approach the expression profiles from the transition state analysis are used to create the probability state model.

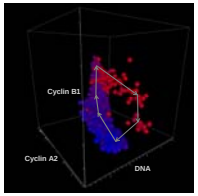
① Introduction

The purpose of cell cycle analysis is to (1) determine the fractions of cells in specific compartments (e.g., phases, checkpoints, mitotic stages), (2) determine the level of expression of a biomarker in specific compartments, and/or (3) determine the continuous, programmed expression of bio-markers as a function of the cell cycle. Here, our objective is to obtain the third by two different approaches. Cytometry measurements on asynchronous populations are unique in that they capture cells at all states within the lifetime (cell cycle) of an average (modal) cell. Because of this, we can achieve goal three from single samples. This is dependent on the presence of a set of measurements that allow us to follow the cell cycle through n-dimensional space in an unambiguous manner. **Figure 1** shows this concept.

Materials & Methods

We stained three samples of exponentially growing Molt4 cells for DNA content, cyclin A2, cyclin B1, and phospho-S10-histone H3 (PHH3) using DAPI and conjugated antibodies (phycoerythrin, Alexa Fluor 647, and Alexa Fluor 488, respectively). Data were acquired on a BD Biosciences LSR II with standard filter set up. Data were preprocessed in all cases (compensation, aggregate removal, background subtraction) with WinList (Verity Software House, VSH). Expression profile extraction and plotting was performed with a combination WinList, Excel (Microsoft), and Prism (Graphpad, Inc.) Probability state modeling was performed with GemStone (VSH). The cyclin A2 antibody was a gift from Vince Shankey (Beckman Coulter); the cyclin B1 antibody was GNS1 (BD Biosciences), which was conjugated with a kit from Molecular Probes; the PHH3 antibody was from Cell Signaling Technology, Inc.

Figure 1:
OpenGL rendering, from WinList, of the four parameter data, with manually placed arrows showing continuous progression of the four measurements within the data space. Three parameters were used to position the events in space. The fourth, PHH3, was color coded, progressing from blue (lowest values) to red (highest values).



② Contiguous Region Expression Profile Extraction

In this approach, regions are set contiguously along the back-bone of the four-dimensional data trajectory, using bivariate view. For this specific approach, 3 views are required (Fig. 2 B-D). However, we used an additional view (Fig. 2A) to take full advantage of the high level expression of cyclin B1 in G2 phase.

Figure 2:
39 Regions were set to obtain measurements of the median values for cyclin B1, cyclin A2, and PHH3 and the number of resident cells in each region. The regions are numbered consecutively from 11-50 as we move through the cell cycle A: Interphase; B: entire cell cycle; C: mitosis; D: cells gated from region 48 (region 48 was not used to calculate PHH3 values.

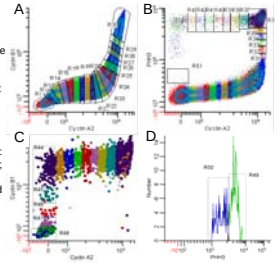


Figure 3 shows the normalized expression for the sample shown above. Each circle represents a region, except the t=0 and 1, which are duplications of first and last measured values. An exception is PHH3, in which the 0 value is ½ the last measured value (to account for cell division). The expression of cyclins A2 and B1 at t=0 and 1 were set to 0.

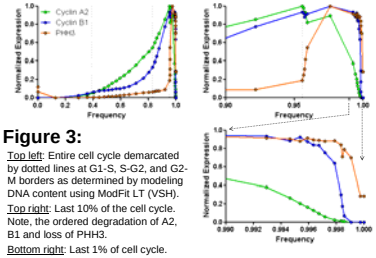
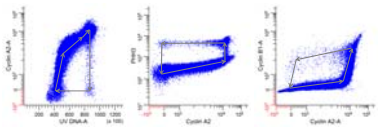


Figure 3:
Top left: Entire cell cycle demarcated by dotted lines at G1-S, S-G2, and G2-M borders as determined by modeling DNA content using ModFit LT (VSH).
Top right: Last 10% of the cell cycle. Note, the ordered degradation of A2, B1 and loss of PHH3.
Bottom right: Last 1% of cell cycle.

The average %CV for expression measurements at each point along the profile was less than 3% except for late measurements of cyclin A2 (most subject to compensation induced error). The average %CV for frequency was less than 1%. (N=3)

Figure 4: Dot-plots with manually placed arrows approximating the expression curves shown in Figure 3.



③ Probability State Model Analysis 1

In this analysis, the Probability State Model (PSM) was based on the expected expression of the various markers through the course of the cell cycle.

The expected expression was that cyclin A2 and cyclin B1 increased at varying rates as a function of cell cycle progression before reaching their maximum values at the onset of mitosis. PHH3 increased slightly until the onset of mitosis and then increased dramatically to its maximum value. The expression of cyclin A2 decreased before cyclin B1, and cyclin B1 decreased before the end of the cell cycle.

The model created by this method is shown in **Figure 5**. The pre-mitotic (Figure 5A) and the mitotic (Figure 5B) events are modeled as two progressions in order to have more detail in the mitotic range.

The PSM's reduced chi-squared (RCS) value quantifies how well a model is fitting the data. For the data shown in **Figure 5**, the RCS values for the two progressions were 2.0 and 1.8 respectively, where 1.0 is the average RCS value for data appropriately modeled.

The PSM shown below fits all three of the replicates in this experiment quite well. When the PSM was optimized for each replicate the quality of the fit, based on the RCS, was equivalent to the data shown below.

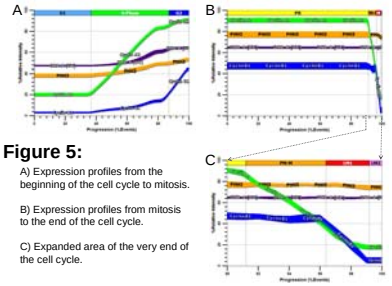
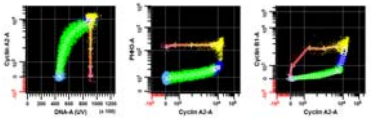


Figure 5:
A) Expression profiles from the beginning of the cell cycle to mitosis.

B) Expression profiles from mitosis to the end of the cell cycle.

C) Expanded area of the very end of the cell cycle.

Figure 6: Dot-plots showing the model's progression in two-dimensional projections of the four-dimensional data. The color of the dots and the location of the progression arrows are automatically determined by the model.



④ Probability State Model Analysis 2

In the second analysis, the PSM was based more directly on the expression of the markers determined by the TSCA (see section 2). The model created by this method is summarized in **Figure 7**. The profile shapes in these overlay graphics are accurate representations of the profiles in section 2. One difference in the profile shape is that the profiles below are shown using a log-like transformation and **Figure 3** shows the linear curves. As shown, the model created in this manner is very similar in expression to the model created in section 2. For the data shown below the RCS value for the two cell types were 15.73 and 5.11.

In this analysis we expected that, if the parameter expression shapes generated by the TSCA are correct, the PSM based on those shapes would distribute the events uniformly through the progression. **Figure 8** shows the frequency distributions for the two progressions. We can see that the frequency distributions are perturbed, indicating that the model created in this manner is not perfectly fitting the data.

One reason that this PSM may not describe the data as well as the PSM in section 3 could be due to slight differences in the pre-processing gates used here and in TSCA. Further, we are modeling only 20 out of the 39 points along each curve found in section 2. Lastly, and likely most importantly, the PSM is accounting for measurement variability where TSCA is not.

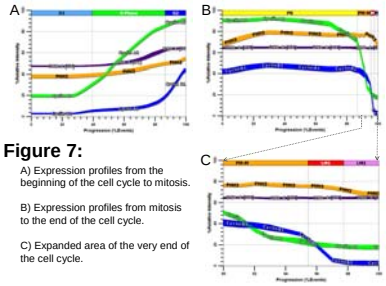
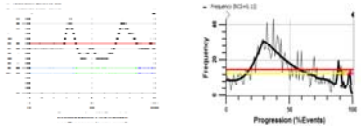


Figure 7:
A) Expression profiles from the beginning of the cell cycle to mitosis.

B) Expression profiles from mitosis to the end of the cell cycle.

C) Expanded area of the very end of the cell cycle.

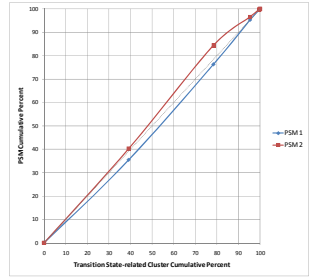
Figure 8: Slightly perturbed frequency distributions for the two progressions.



⑤ Comparison of Analyses

One preliminary approach to quantitatively compare the two PSM analyses with region-based extraction of profiles is to plot the cumulative percent of events found at each stage of the cell cycle, one method vs. the other. **Figure 9** shows that all three methods are correlated. Interestingly, if we take the region extraction method as the truth, the PSM 2 effort faired less well than the original effort, working from a verbal description (the cartoon version).

Figure 9: Comparison of cumulative percent of cell cycle progression found by the three analyses. The dashed diagonal line represents a perfect correlation.



Conclusion

We found that analysis by PSM resulted in expression profiles similar to TSCA. Cyclin A2 rises faster in S compared to cyclin B1; cyclin A2 degrades before cyclin B1; PHH3 rises abruptly at the onset of mitosis. Further, we found that analysis by both PSMs resulted in classification of the cell cycle that was approximately the same as TSCA (Figure 9). This suggests that the PSM 1 approach is sufficient to provide a good compartment or transition state analysis of the cell cycle based on these measurements. At least in this study, there doesn't seem to be much gained by using the locally detailed information from the region-based method when constructing the PSM.

We have not quantitatively compared the expression profiles from the TSCA and PSM analyses. However, it is clear from **Figures 3**, **5**, and **7** that the two approaches differ, particularly for cyclin A2. E.g. at frequency 0.8, in **Figure 3**, cyclin A2 is increasing significantly while at state 80, in **Figures 5A** & **7A**, cyclin A2 has become asymptotic. The profile shapes for cyclin B1 and PHH3 seem more concordant among the three methods.

There are some advantages of the PSM approach. A PSM accounts for overlaps between the stages of the cell cycle. Directly describing the profile shapes is simpler than using the numerous gates need for the TSCA. The PSM allows the modeler to easily account for measurement variability. Further, the PSM gives feedback as to how well the model is fitting the data via the RCS value. For a more detailed description of PSM see poster P96, program number 182.