MultiCycle AV Cell Cycle Data Display Explained

OVERVIEW

After a MultiCycle cell cycle analysis has been performed on a DNA histogram in FCS Express, the operator is presented with a graphical display of the data which contains multiple shaded areas corresponding to the debris and the different compartment of the cell cycle. If the mouse is held over the DNA histogram and right button of the mouse is depressed, a menu appears with multiple options. About a third of the way down on the menu is the "Statistics" option. There are four choices on this drop down menu of which three apply to cell cycle analysis. The cell cycle statistics selectable are: DNA Cycle Statistics, DNA Model Summary Statistics, and DNA Experiment Statistics. Here is an explanation of what is contained in each data box.

DNA CYCLE STATISTICS

Cycle	G1 Mean	G1 CV	%G1	G2 Mean	G2 CV	%G2	%S	G2/G1	%Total	B.A.D.
Diploid	83.10	6.33	86.98	160.43	5.95	5.57	7.45	1.93	100.00	17.98

This box contains the "meat" of the cell cycle analysis. The percentages and CV's of the different compartments of the cell cycle analysis and the Background Aggregates and Debris measurement. If this doesn't make sense, then go back and read the introductory chapter, "Introduction to Cell Cycle Analysis"

Note: If your default DNA Cycle Statistics box contains more or less columns of information than in the above example, move your mouse over the DNA Cycle Statistics box, right click it and select the information displayed above from the drop down menu. Once you have the information to your liking, place your mouse cursor over the DNA graph, right click again and click on Set As Default. FCS Express will then remember the correct columns of information the next time you do an analysis.

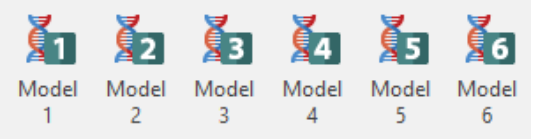
DNA MODEL SUMMARY STATISTICS

Model	Dip %G2	Dip %S	Dip G2/G1	Dip %	Dip BAD	BAD	Debris	Chis
1) SL S0	5.57	7.45	1.93	100.00	17.98	17.98	51.20	1.65
2) SL CL S0	4.18	4.49	1.89	100.00	22.27	22.27	50.03	1.45
3) SL G21 S0	4.53	8.94	1.97	100.00	18.22	18.22	51.61	1.77
4) SL CL G21 S0	4.30	8.67	1.97	100.00	18.82	18.82	51.52	1.78
5) SL S1	7.36	7.17	1.91	100.00	18.16	18.16	51.54	1.55
6) SL CV S1	6.98	7.32	1.91	100.00	18.12	18.12	51.45	1.55

The **DNA Model Summary Statistics** box adds information to the analysis that permits different cell cycle models to be rapidly compared so the reliability of the S and G2 phase estimates can be objectively assessed. Six different cell cycle models are automatically fit. The variability in results is one aid to assessing confidence in S and G2 phase estimates. This feature allows much faster and more convenient display of the results of fitting a histogram using different variations of the fitting model. The results of histogram fitting using models with and without aggregation, CV constraints, G2/G1 ratio constraints, and a first order (tilted) S phase polynomial are all available for viewing and comparison.

The first column in the **DNA Model Summary Statistics** box shows the different constraints added to the basic cell cycle analysis. Remember, MultiCycle AV always uses the Polynomial S-fit model of Dean and Jett. The first column also corresponds to Model 1, Model 2, Model 3, Model 4, Model 5 & Model 6 buttons on the top menu of FCS Express when the MultiCycle tab is selected. Together these six models are used to calculate the intermodel and intramodel fitting variation. The individual numbers from these calculation are not displayed in the FCS Express version of MultiCycle AV but the results of the calculations are displayed in the **DNA Experiment Statistics** Box mentioned in the last section of this paper (See Appendix for explanation of math). All six models are always fit, however, which model statistics are displayed in the **DNA Cycle Statistics** box mention above is selected by the Model 1-6 buttons on the top menu.

Fitting Models 1-6



The six models corresponding to the 1-6 Model Buttons are described below:

- **SL S0**: Sliced nucleus background modeling, with zero order S phase, but without other options.

- **SL CL S0**: Sliced nucleus background modeling, Clumping compensation with zero order S phase.

- **SL G21 S0**: Sliced nucleus background modeling with zero order S phase (i.e. as **SL S0**), plus constraint of the G2/G1 ratio. For one cell cycle fitting models, the G2/G1 ratio is the current default stored within MultiCycle AV.

** Note: the default G2/G1 ratio constraint value is the last value set by the user. If this value needs to be changed, used the **[G2/G1 Fixed]** option on the fitting option menu.*

For two cell cycle models, the G2/G1 ratio of the diploid cycle is constrained to be equal to the aneuploid G2/G1 ratio that resulted from the user-chosen fitting model.

For three cell cycle models, the G2/G1 ratio of all three cell cycles is constrained to be equal to the average of the two aneuploid G2/G1 ratios that resulted from the user-chosen fitting model.

- **SL CL G21 S0**: Sliced nucleus modeling with zero order S phase, plus aggregation modeling, and addition of a G2/G1 constraint. The G2/G1 constraint is the same as described in the paragraph above, except for 1 cell cycle models: for histograms with 1 cell cycle, the G2/G1 ratio is constrained to be identical to the value of the G2/G1 ratio resulting from the initial user selected fitting model.

The G2/G1 constraint is utilized in order to provide a conservative model of aggregation: without this constraint, aggregation modeling can occasionally predict excess aggregation and small G2 peaks with abnormal G2/G1 ratios. The G2/G1 constraint prevents this from occurring.

- **SL S1**: A first order (tilted or trapezoidal) S phase polynomial is used with sliced nucleus background modeling.

- **SL CV S1**: Sliced nucleus modeling, with first order S phase, plus the CVs of all peaks are constrained.

DNA model summary statistics box Terms

Dip %G2-Percent Cells in G2-phase

Dip %S-Percent Cells in S-phase

Dip G2/G1-Diploid G2/G1 ratio

Dip %-Percent Cells in Diploid Cell Cycle

Dip BAD-Percent Background Aggregates & Debris in Diploid Cycle

BAD-Percent Background, Aggregates & Debris in all Cycles

Debris-Total Percent Debris in Histogram

Chis-Overall Chi-Square of Cell Cycle Analysis

DNA EXPERIMENT STATISTICS

The third box that can display information about the cell cycle analysis is the DNA Experiment Statistics. Inside this box are two sections: Interpretation and Experiment Statistics.

The Interpretation section will show a concise statement of ploidy, the results of S and G2 phase calculations and the predicted degree of confidence.

When two or more cell cycles have been fit, the interpretation describes the aneuploid cell cycle(s), unless there is extensive overlap of cell cycles, in which case the average S and G2 phase is described. When two or three aneuploid populations extensively overlap, the Interpretation page will show the average aneuploid S phase and its confidence estimate.

The interpretation of estimated confidence is based on the percent aneuploid cells (if present), the background aggregates and debris (B.A.D.), and the intra- and intermodel confidence intervals. MultiCycle AV examines the number of negative factors that are present, and the magnitude of how unfavorable each factor is, and synthesizes this data into an estimate of the overall level of confidence. The statement of confidence spans the range of good, fair, poor and very poor.

When S phase confidence is less than good, the Interpretation will also provide a note indicating which factor(s) (percent aneuploid cells, B.A.D., intra- and intermodel confidence) are contributing to the reduced confidence.

**** Note: The histogram interpretation and confidence information should be used only as a guideline to begin detailed human analysis of the histogram. These guidelines should NOT be used for histogram interpretation without careful operator evaluation and confirmation or rejection.***

APPENDIX

OBJECTIVE INDICATORS OF CURVE-FITTING RELIABILITY

The chi square statistic ($\chi^2 = \sum \frac{(yfit_i - ydata_i)^2}{\sigma_i^2}$, where σ_i^2 is the uncertainty of each data point $ydata_i$) or the reduced chi square ($\chi_v^2 = \sum \frac{\chi^2}{degrees\ of\ freedom}$) are measures of the deviation of the fitting function from the data, i.e., measures of the goodness of fit. The χ^2 statistic may thus be useful to indicate the extent to which the fitting model matches the histogram distribution, lower values of χ^2 being better. The problem with presuming that the fit with the lowest χ^2 is the best is that the fitting model may be incorrect for that histogram or the model may be performing an excellent fit to artifacts in the data (e.g. peaks with skews or shoulders). Thus, human judgment of the fit is essential, although the χ^2 may be used to aid this judgment. In addition, the user should also bear in mind that the χ^2 statistic also is affected by the number of cells acquired in the histogram and by the relative proportion of cell cycle vs. background distribution represented in the histogram.

INTRA-MODEL CONFIDENCE INTERVALS

The nonlinear least squares method of fitting is a process of successive approximations, and the result of the fitting is not necessarily an exact answer. As described in "The Introduction to Cell Cycle Analysis paper", the nonlinear least squares technique optimizes the fit by adjusting each fitting parameter to minimize the value of χ^2. If we imagine a "cell cycle" model with the fitting dependent just on two parameters called "North" and "West", and if we imagine the values of the χ^2 obtained for each combination of "North" and West as plotted on a vertical scale, then optimizing the fit is like searching for combinations of "North" and "West" that lead to the lowest depression in a valley surrounded by hills. When you get towards the bottom of the valley, it may have steep sides and a pool at the bottom to show the lowest point, or it may be flatish and bumpy on the bottom. If the latter, then what spot you call bottom may depend on your starting point and how you walked into the valley, and you may not arrive at the same "bottom" point twice! This condition obviously results in greater uncertainty in the "correct" combination of "North" and "West". Some cytometrists are surprised to find that there is not one exact answer to the question of what is the S phase fraction of a histogram. This may become especially apparent as different results are obtained from successive analyses of the same histogram when the initial approximations (CV estimates, estimated peak means, etc.) vary.

One way to estimate the confidence in finding the same "best" fit is to

perform the fitting many different times, starting from different places each time. This would be useful, but tedious. Fortunately, the statistical process used by the nonlinear least-square fitting can provide a similar estimate, based on the "steepness" of the bottom of the valley. Stated mathematically, the rate of change in χ^2 as the parameter is adjusted – the degree of curvature of the multidimensional surface described by the values of χ^2 – is an indicator of the uncertainty in the fitting parameters. If the χ^2 changes rapidly as a parameter is altered, then the optimized fitted value for that parameter is likely to be narrowly constrained near its "correct" value. This uncertainty is estimated from the matrices used in the least squares fitting, specifically, the "error matrix," which is itself derived from the matrix that describes the curvature of the χ^2 multidimensional surface (Bevington, 1969). We call the estimate based on this process the "intramodel" error estimate (to distinguish it from the "intermodel estimate", described subsequently), and this value is very useful to indicate a confidence interval for the cell cycle measurements.

The S phase intramodel confidence interval will be wider in histograms in which the S phase is highly overlapping with other cell cycle compartments. Higher levels of debris and aggregates overlapping the S phase will also results in wide intramodel confidence intervals. Another circumstance in which the S phase intramodel confidence intervals will be wide is if the histogram cell number is low, especially in the region of S phase.

Table 4.1 shows an example of error estimates calculated for fitting of the histogram shown in Figure 4.1. The intramodel error range is defined as a 95% (±2 S.D.) confidence interval. This example indicates that the diploid S phase estimate from the analysis shown in Figure 4.1 is unreliable. This is most likely due to the large component of sliced nucleus debris that overlaps the diploid S. In contrast, the aneuploid S phase estimate has a narrow intramodel confidence interval.

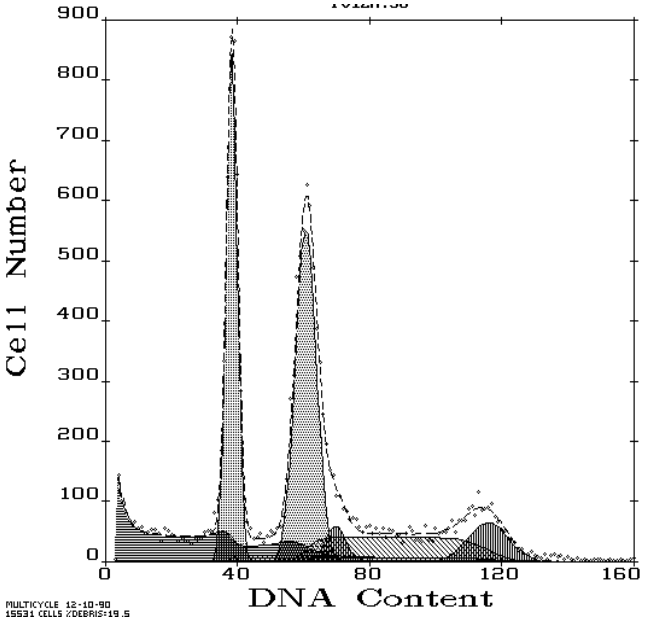


Table & Figure 4.1*. Intramodel and intermodel confidence intervals (C.I.) from analysis of the histogram shown above.*

Model	Diploid S	Diploid G2	Aneuploid S	Aneuploid G2	χ^2_v
(1) Sliced Nucleus, zero order S phase	8.1	9.3	28.6	13.1	1.4
Intramodel 95% C.I.	**0-16.8**	**6.8-11.7**	**23.6-33.2**	**11.9-14.3**	
(2) #1 + G2/G1 ratio constraint	9.7	6.3	27.5	13	1.5
#2 plus aggregation model	10.1	6.2	26.0	12.1	1.2
#1 plus all CVs equal	8.1	15.4	31.7	12.7	1.6
#1 plus first order S phase	5.3	9.6	27.3	14.4	1.4
Intermodel C.I.	**5.3-10.1**	**6.2- 15.4**	**26.0-31.7**	**12.1-14.4**	

INTER-MODEL CONFIDENCE INTERVALS

The intramodel error estimates still do not address the question of whether the model used is appropriate for the histogram. The extent to which the cell cycle parameters vary depending upon the model chosen is another indicator of the range of uncertainty in the estimates. This "intermodel" error estimate is illustrated in Table 4.1. In this example, the intramodel and intermodel techniques yield similar indications of reliability and, in the case of G2 estimates, they agree very closely. In other histograms, the two methods may yield different confidence estimates.

The intramodel and intermodel confidence limits can be used together to give a useful estimate of the confidence in S and G2 phase measurements. In general, if either confidence interval is wide, then that parameter is more likely to be approximate and subject to increased error. In some cases, examination of one of the automated model fittings may suggest that fitting is unrealistic; in this situation the S and G2 estimates from that model may be dropped from the intermodel comparisons.

The novel and important feature of this innovation in MultiCycle AV is that through the combined usage of the intramodel and intermodel error analyses, B.A.D, and percent aneuploid cells, the user can now obtain a systematic and comprehensive approach to the estimation of reliability of S and G2 phase fractions from DNA cell cycle analysis.

CONFIDENCE ESTIMATION AND IMPROVED PROGNOSTIC STRENGTH OF DNA ANALYSIS

Guidelines for rejection of histograms based on "poor" quality, such as those described above, are based upon a consensus of experience, but there is minimal objective published data. Existing databases of DNA Content histograms and the associated clinical outcome can be used to objectively define improved reporting methods, including tests of criteria for histogram rejection. In such an analysis, "poor" histogram quality should be associated with poor prognostic strength, and removing these less reliable measurements should improve the prognostic strength of the remaining cases in the study.

An analysis of predictors of poor reliability of S phase measurements is illustrated in Table 4.2. Consistent with the DNA Cytometry Consensus Conference Guidelines, rejection of cases with low proportions of aneuploid cells and high B.A.D. was beneficial, although the exact rejection criteria were slightly different than suggested by the Consensus Conference (more stringent for percent aneuploid cells, less stringent for B.A.D.). The MultiCycle intramodel error was also beneficial. In addition, comparison of the intermodel

range of S phase values (e.g. with and without aggregation modeling, zero vs. first order S phase models, with vs. without G2/G1 ratio constraints) was a moderately useful predictor of reliability. These four predictors are combined into an overall estimate of histogram reliability in MultiCycle AV.

Table 4.2. *Predictors of S phase reliability: prognostic strength of the cases remaining after rejecting subsets of histograms*[1].

Cases Removed	Percent cases removed	Hazard (Relative Risk)	Significance (p)
None	0	2.10	.0005
% Aneuploid cells <30% % Aneuploid cells <40%	14 21	2.54 2.82	.00002 .000005
B.A.D. > 35% B.A.D. > 28%	10 20	2.40 2.54	.00004 .00002
Intra-model error 80th percentile Intra-model error 70th percentile	20 30	2.50 2.64	.0002 .0002
Inter-model error 90th percentile Inter-model error 80th percentile	10 20	2.18 2.32	.0004 .0002
CV >8.6 CV >7.5	10 20	2.00 1.94	.002 .006
Chi Square >2.0 Chi Square >1.7	10 20	1.93 2.05	.005 .003

[1] Breast cancer, all stages, surgical treatment only. Cox multivariate analysis. Modified after Barlow et al., 1995.

In contrast, the analysis shown in Table 4.2 failed to demonstrate that either elevated CV or higher chi square of the least squares fit was useful as a criterion for histogram rejection. The absence of utility of CV is surprising, but may indicate that other measures of histogram quality are more important. The chi square is affected by a large number of variables, not all related to goodness of the fit; these include the number of cells acquired in the histogram and the endpoints of the analysis region used within the histogram.

MultiCycle AV has been made to utilize these objective parameters in a "friendly" approach to confidence estimation in cell cycle analysis.