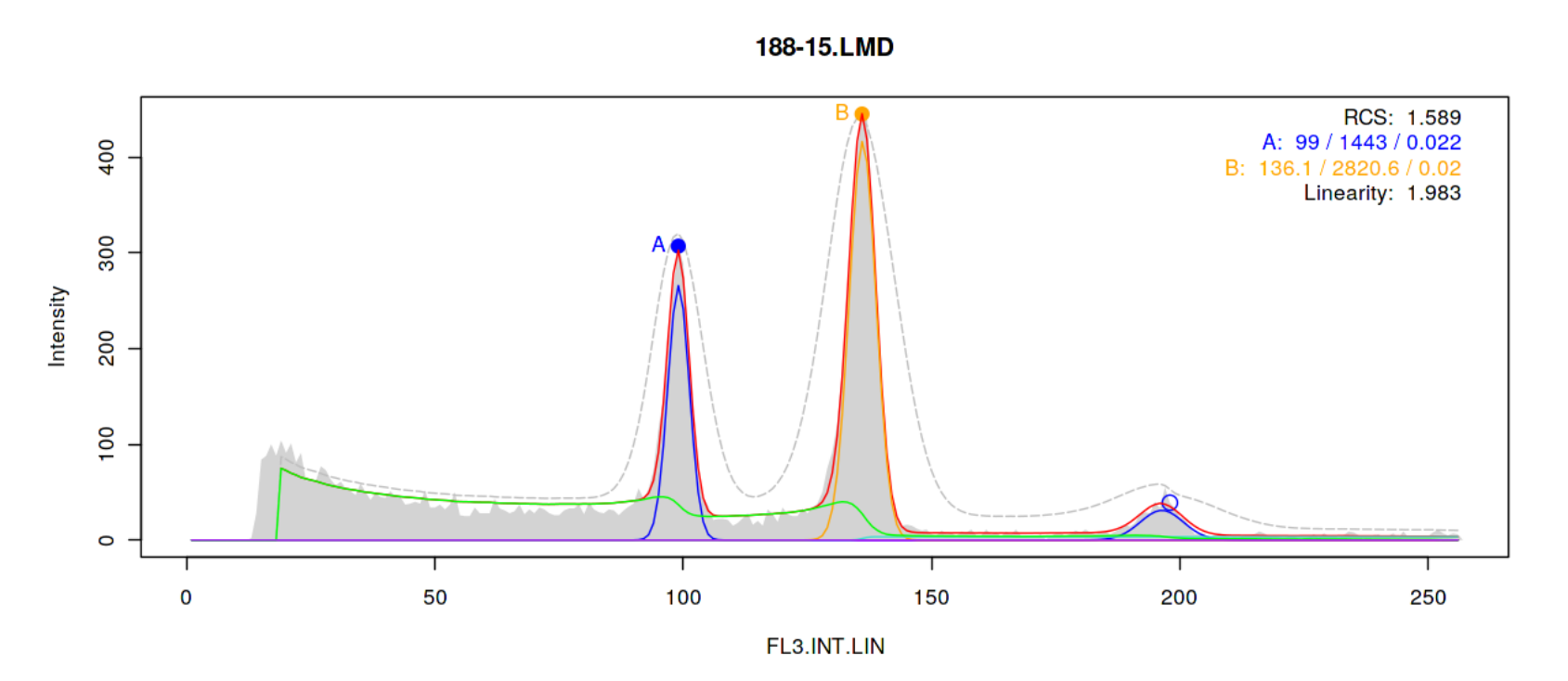
flowPloidy: Determining Genome Size and Ploidy from Flow Cytometry Histograms in R



Tyler Smith[1], Paul Kron[2] and Sara Martin[1]

[1]Agriculture and Agri-Food Canada, [2]University of Guelph

Motivation

Evolutionary biologists working with polyploid taxa

Population screening:

- large sample sizes
- poor tissue quality

Genome size assessment:

- repeat measurements of fresh/greenhouse tissue
- high precision required

Need **high-throughput**, **high-precision** genome size estimates

Histogram Analysis

Tissue preparation

Chop and stain tissue

- solution contains 1000s of individual nuclei (and debris)

Flow Cytometer

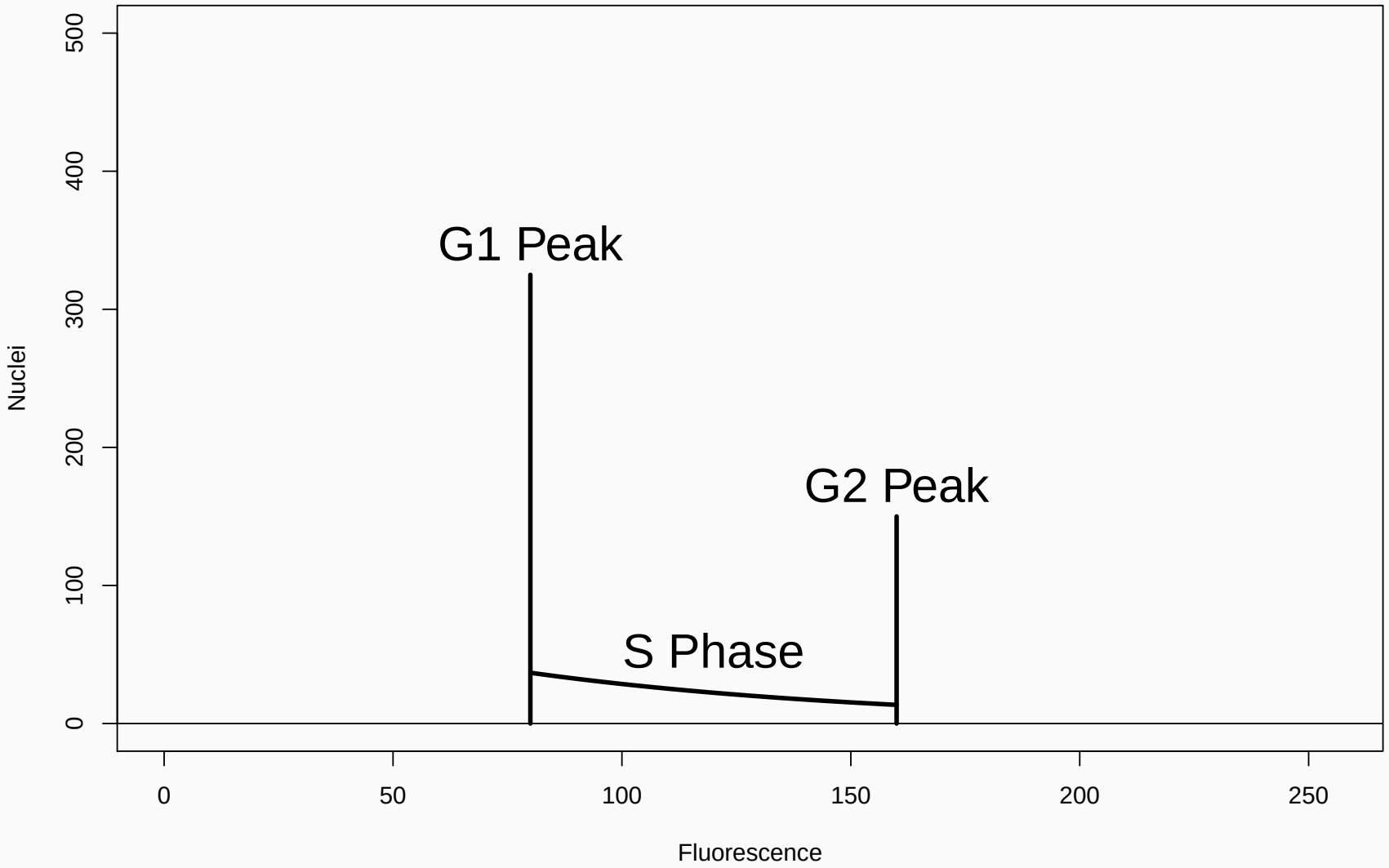
Measure the fluorescence properties of each nuclei:

- size (forward scatter)
- DNA content (fluorescence)
- granularity (side scatter)

Histogram Construction

- aggregate data into bins (256, 512, 1024)

Ideal Histogram



Histogram Components

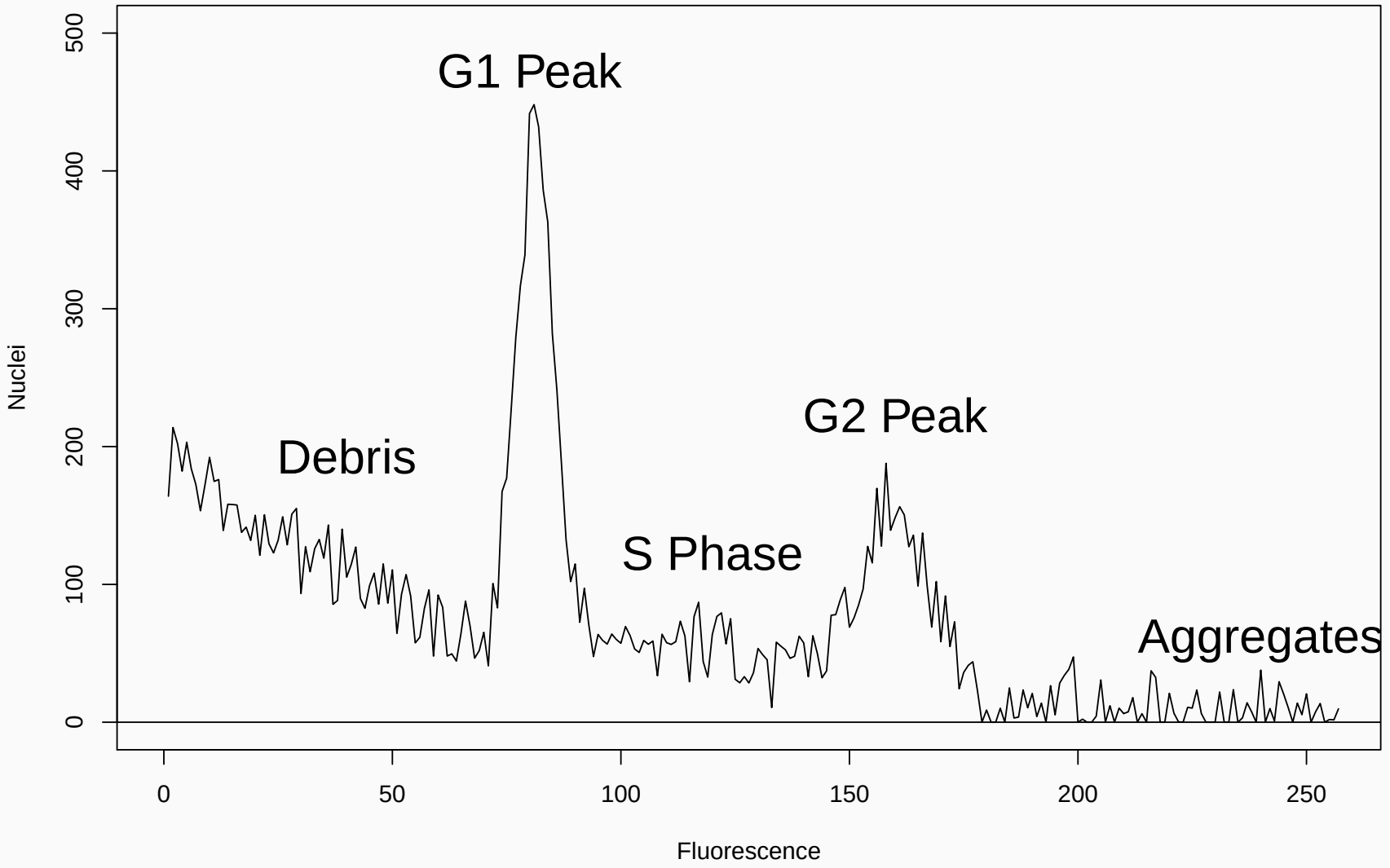
G1 peak Gap 1 diploid cells

G2 peak Gap 2 tetraploid cells (pre-division)

S Phase Synthesis cells (actively duplicating DNA)

(**Co-Chopped Standard Peak** ignored for now)

Empirical Histogram



Histogram Components

G1 peak Gap 1 diploid cells

G2 peak Gap 2 tetraploid cells (pre-division)

S Phase Synthesis cells (actively duplicating DNA)

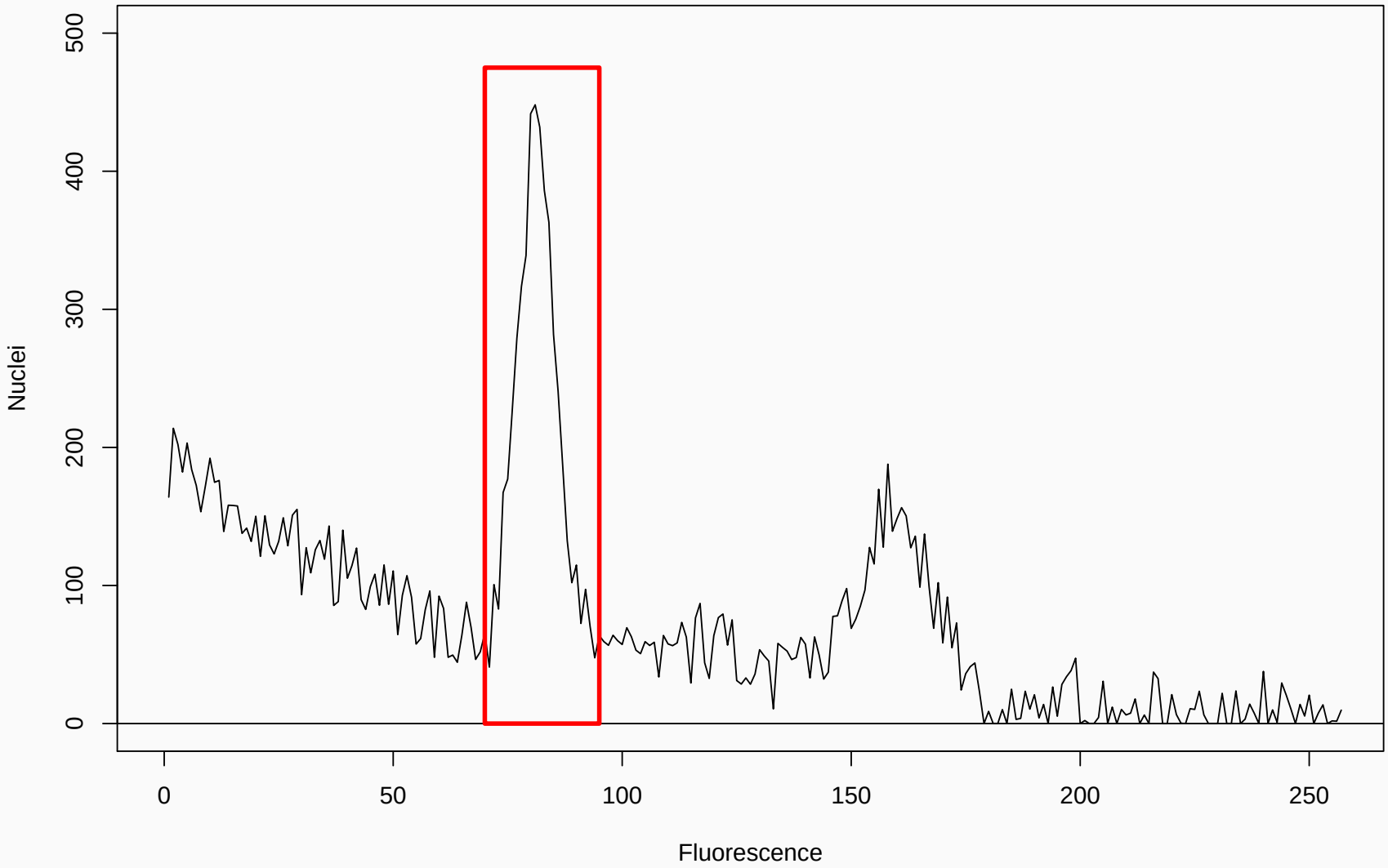
Debris Damaged nuclei, cell components, contaminants

Aggregates clusters of two or more nuclei stuck together

NOISE Measurement error, capriciousness of life

(**Co-Chopped Standard Peak** ignored for now)

Manual Histogram Analysis



Manual Histogram Analysis

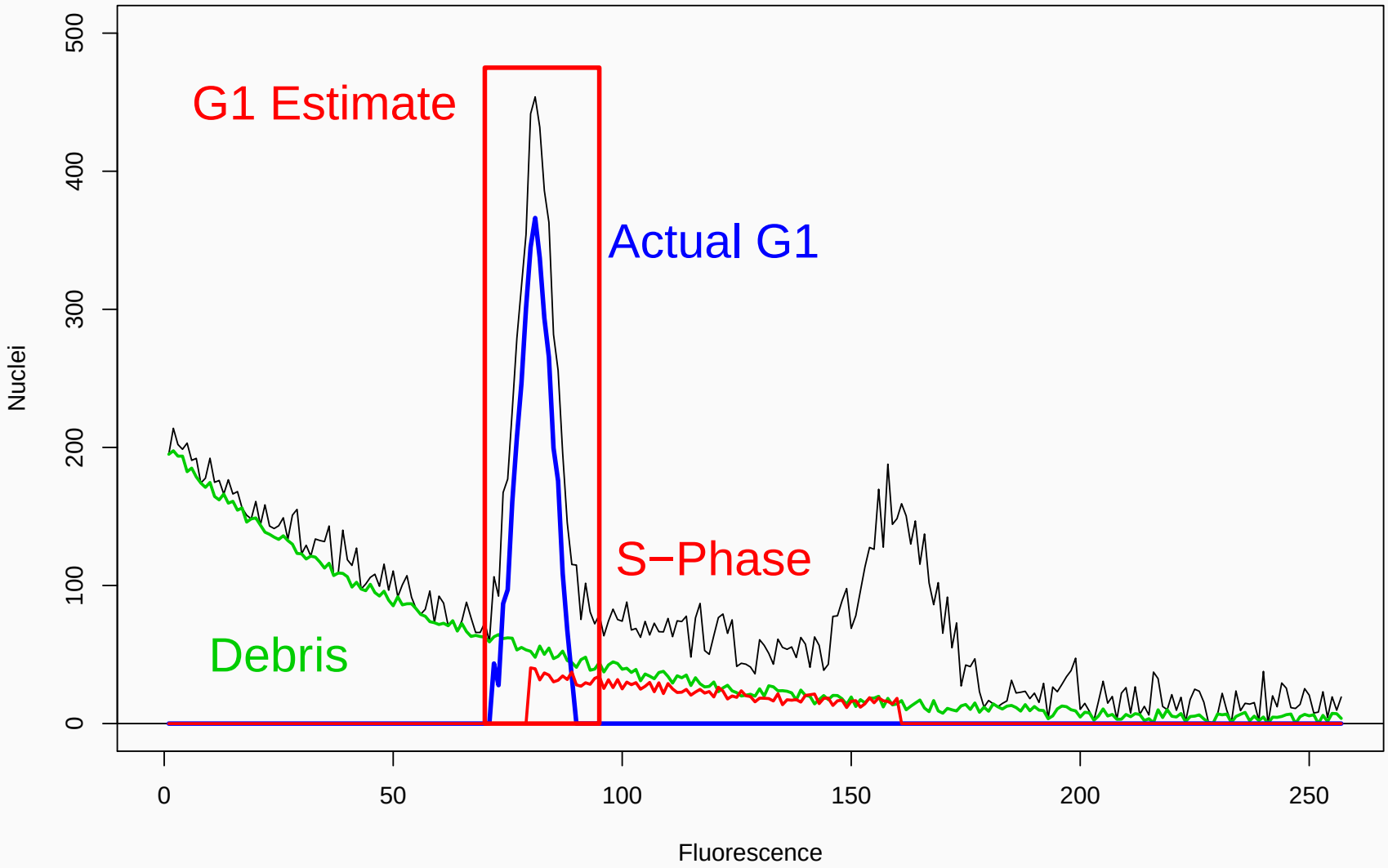
Advantages

- Intuitive
- Several Programs Available
- Can be done ad-hoc in R

Disadvantages

- Subjective
  - CV estimate depends on user
- Doesn't account for overlapping components
  - G1 cell count estimate inflated by debris and S-phase

Overlapping Histogram Components



Non-linear Regression Histogram Analysis

Model histogram components using mathematical functions:

- G1 and G2 peaks fit as Normal curves
- Debris and aggregates fit using theoretical models

Advantages

- objective
- estimates taken directly from the data

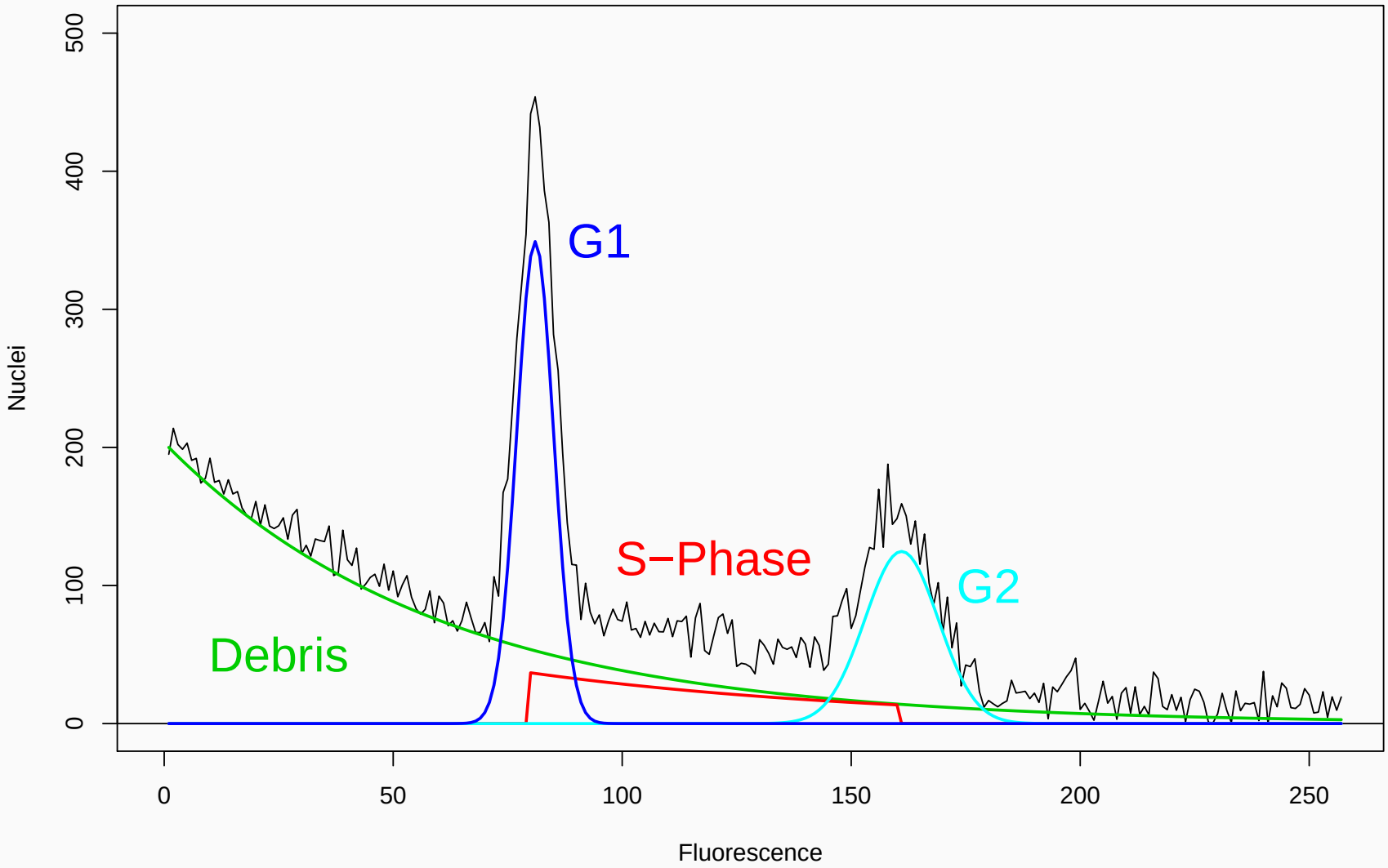
Disadvantages

- availability (few programs, expensive licenses)
- conceptually complex

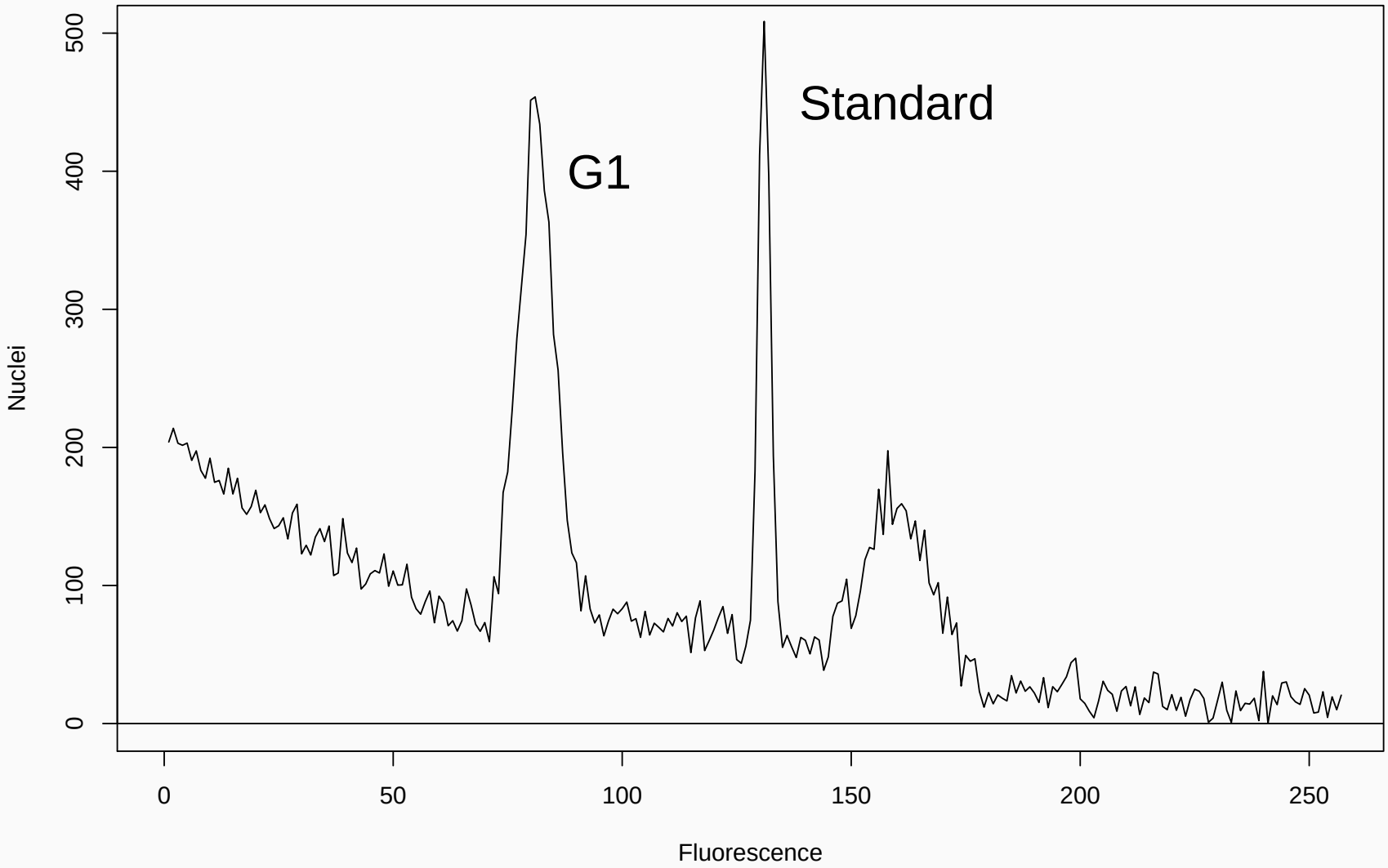
Source

Bagwell, C. B. (1993). Chapt. 3 *In* K. D. Bauer et al., Clinical flow cytometry: principles and applications. Williams & Wilkins.

Non-linear Regression Histogram Analysis



Co-Chopped Standard



flowPloidy - Motivation

Issues with ModFit

- Cost and accessibility
- Functionality (too much and too little)

flowPloidy Goals

- Streamline our workflow, integrate with R
- Increase our understanding of histogram analysis
- Platform for developing best practices

Free Software

Open access for everyone in our lab and yours

You can learn from my work

You can improve and extend my work

Bioconductor

Builds on existing tools for flow cytometry in R

Installation is simple:

```
source("https://bioconductor.org/biocLite.R")
biocLite("flowPloidy")
biocLite("flowPloidyData")                    # for examples
```

Loading Data

```
library(flowPloidy)
library(flowPloidyData)                    # for examples
batch1 <- batchFlowHist(flowPloidyFiles,
                        channel = "FL3.INT.LIN")

batch1 <- browseFlowHist(batch1)
```

Reviewing Histograms

File **1** of 14

Exit Prev Next

Samples 2

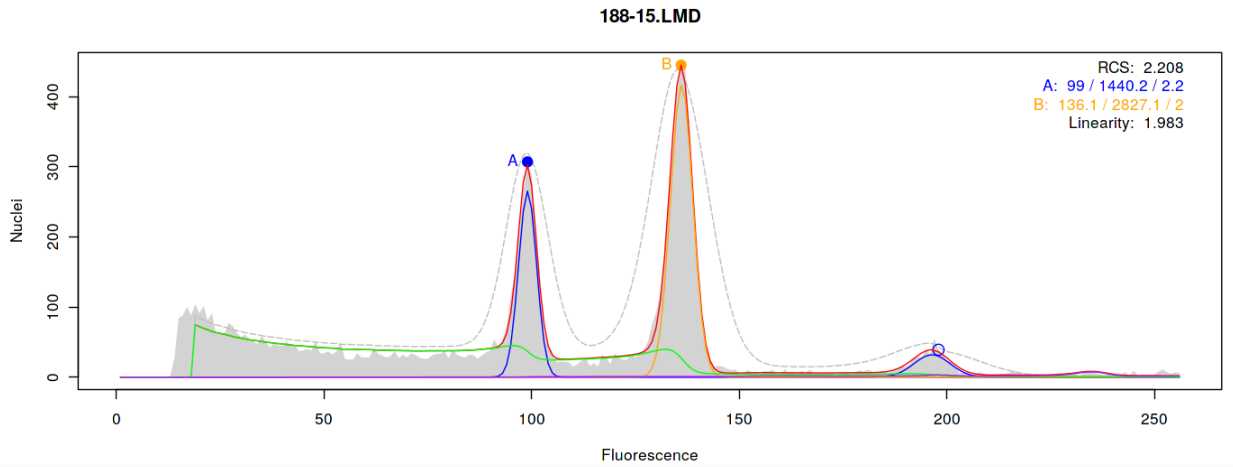
Peak A

Standard Value 0

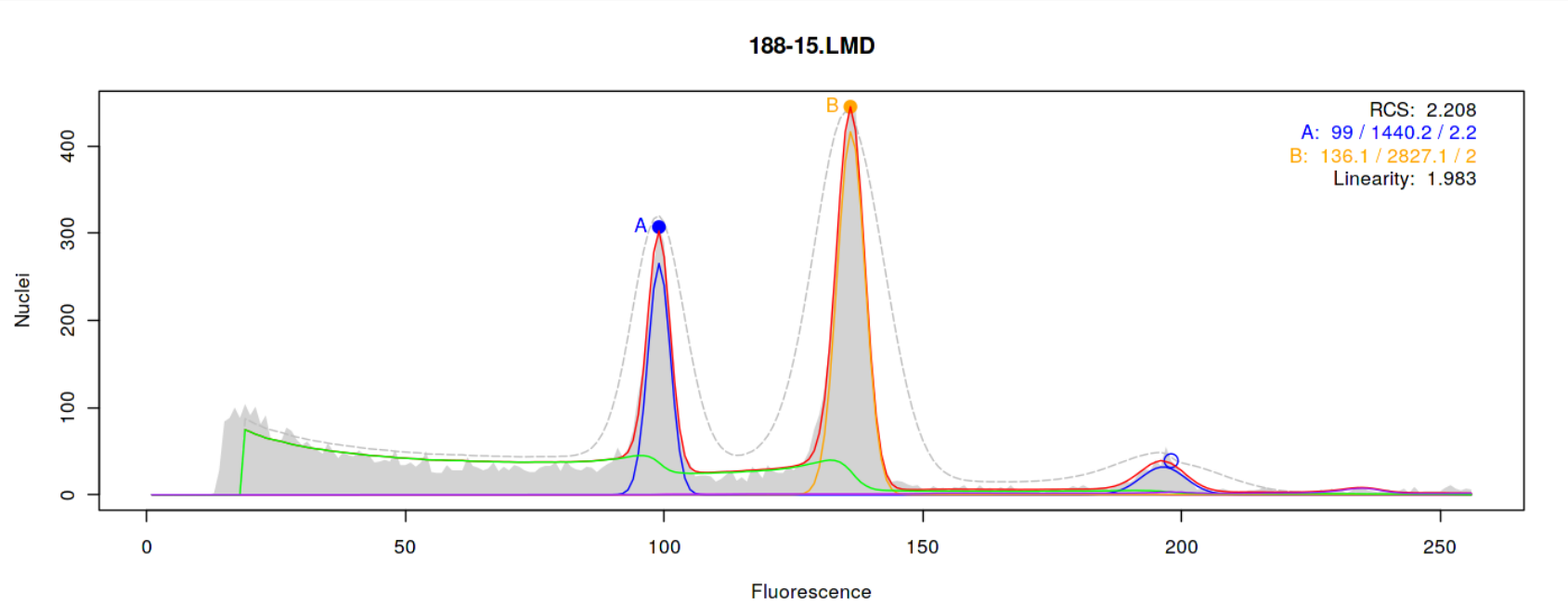
Standard Peak X

Linearity Variable

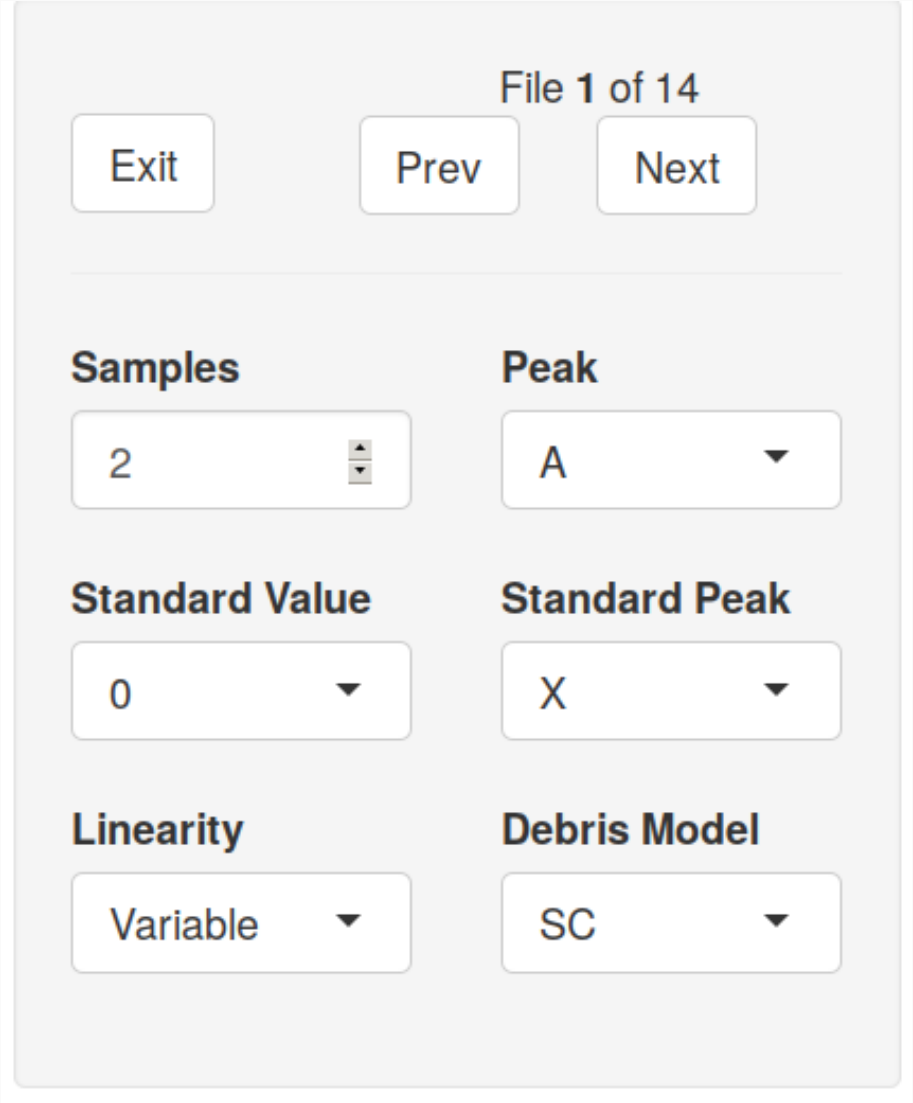
Debris Model SC



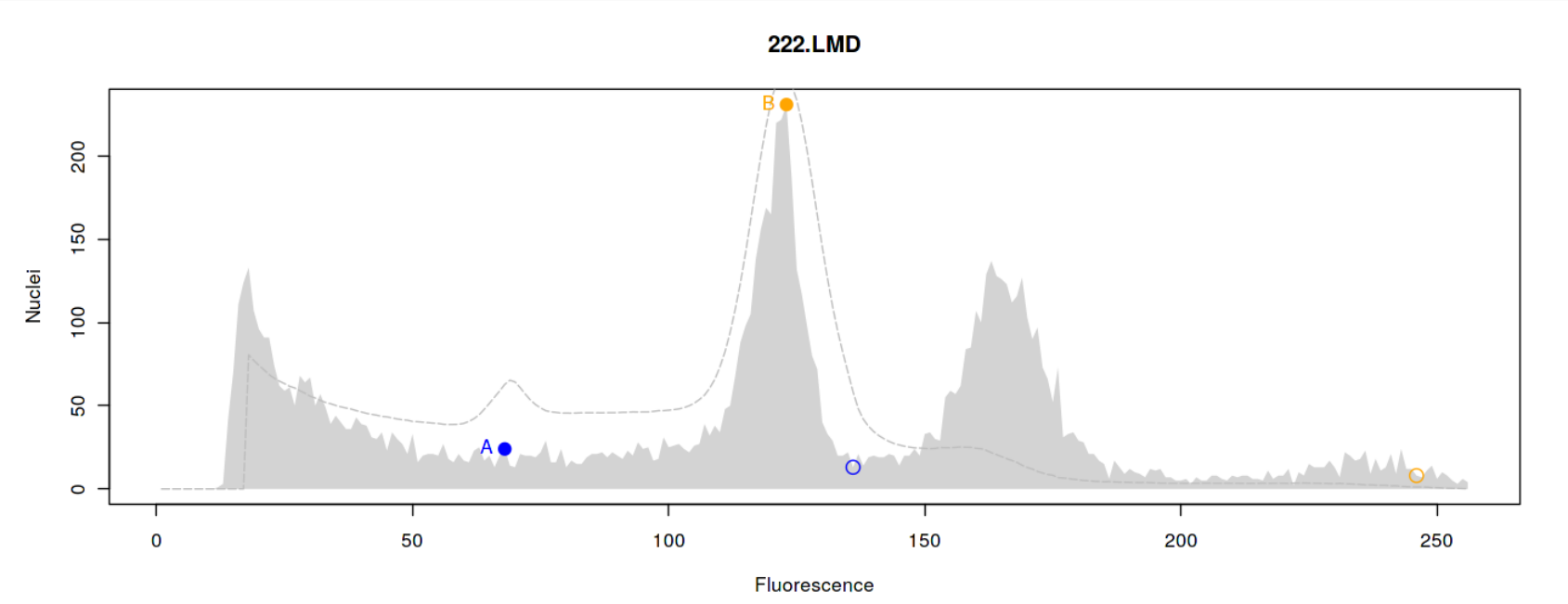
Reviewing Histograms



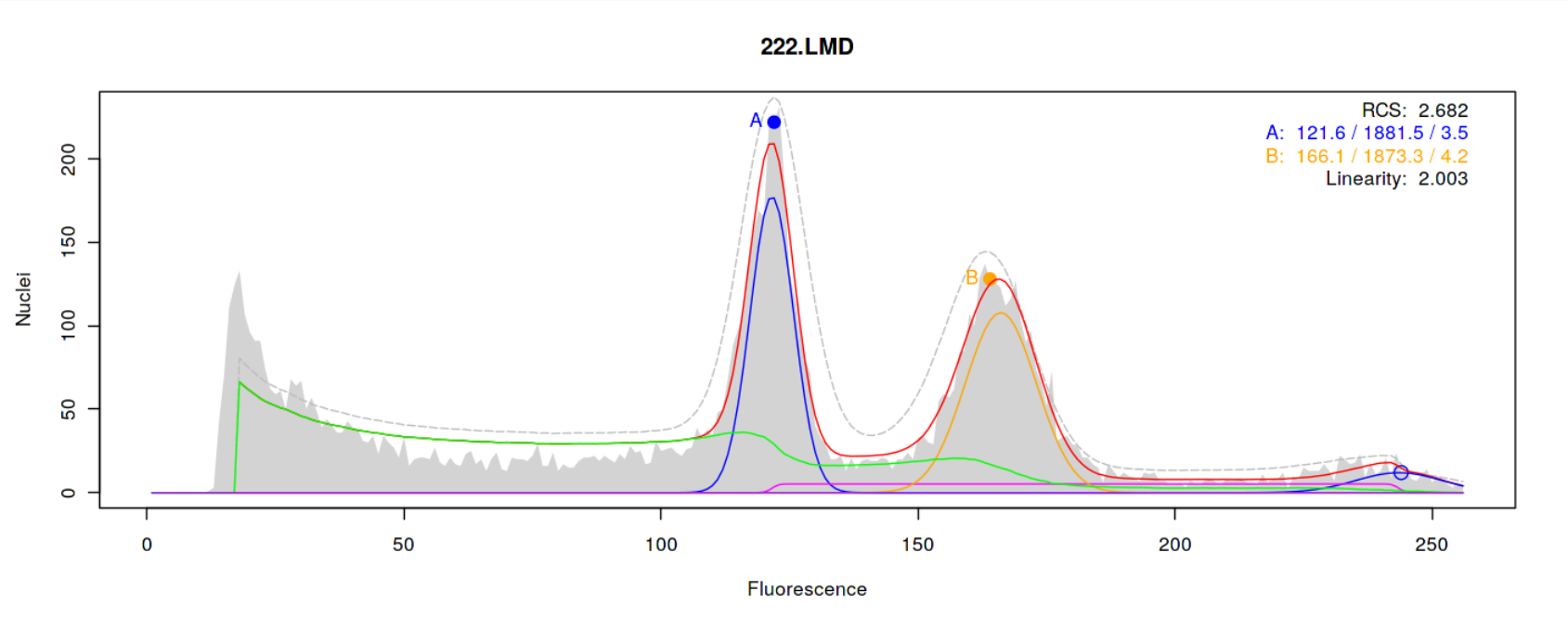
Reviewing Histograms



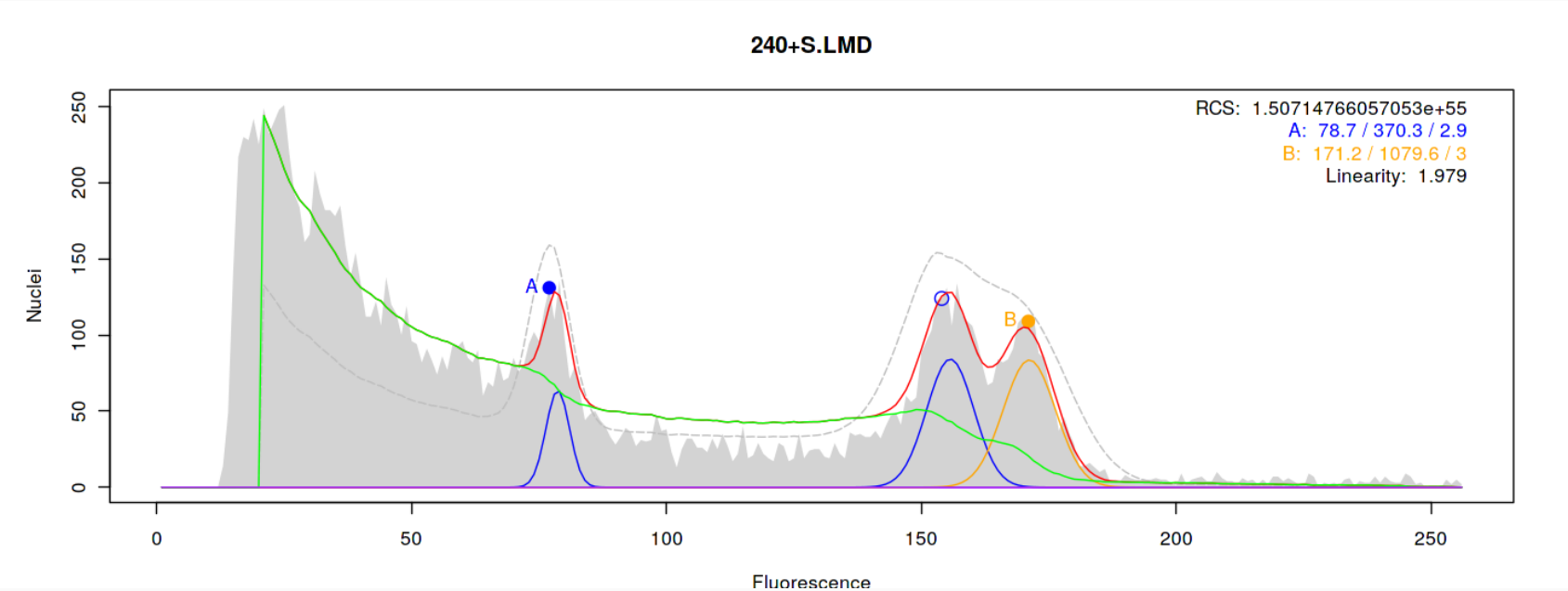
Correcting Histograms



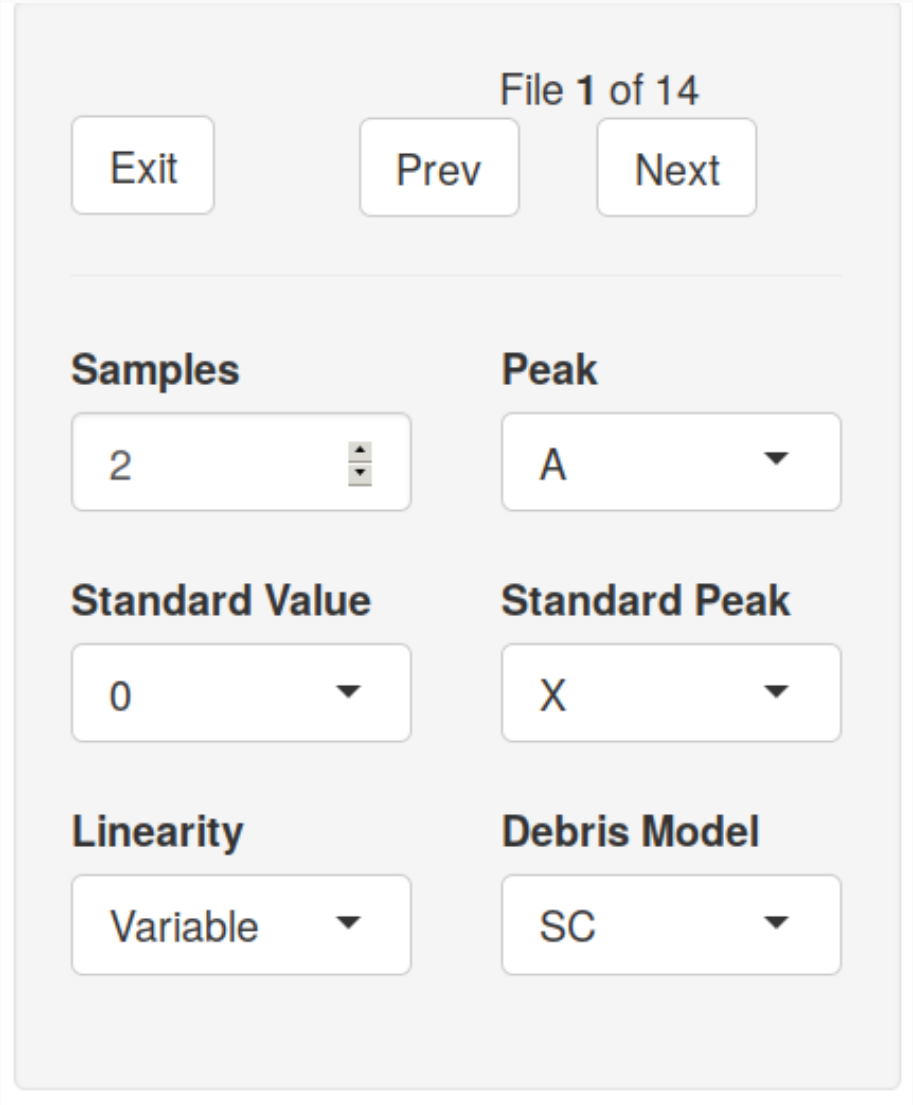
Correcting Histograms



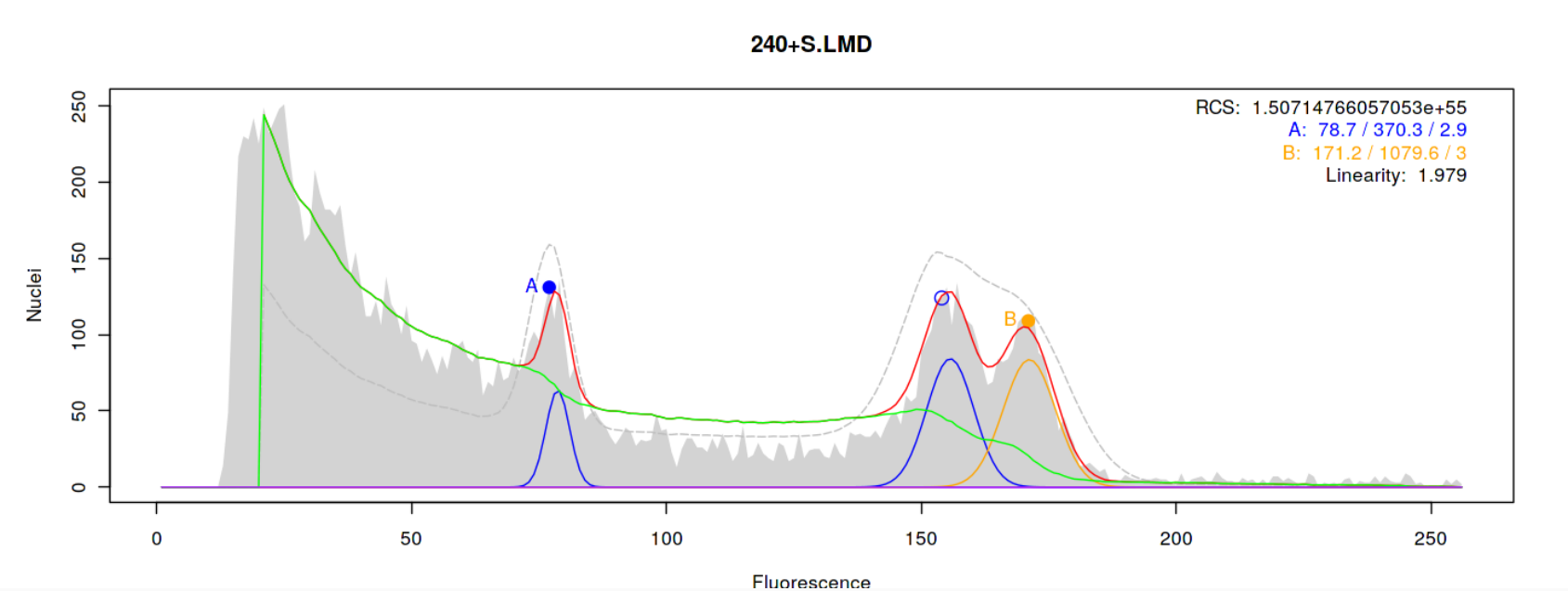
Changing Model Components: Debris

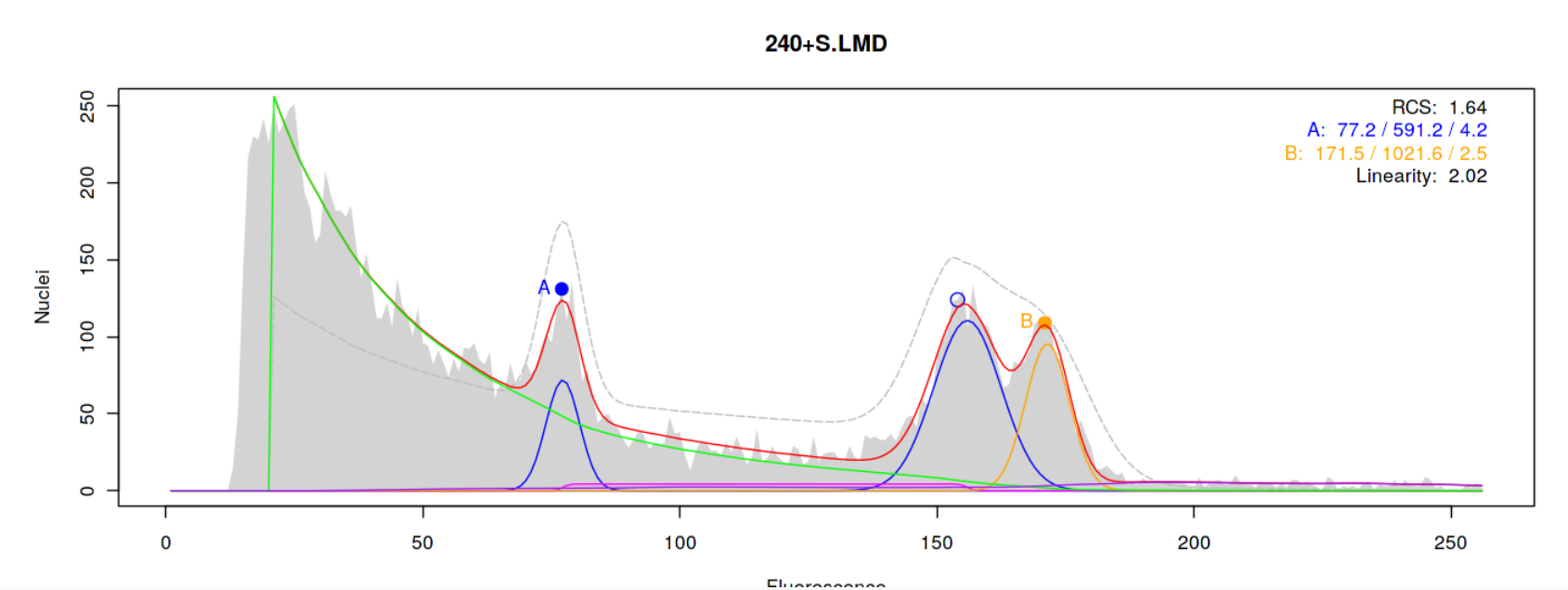


Changing Model Components: Debris



Changing Model Components: Debris





Changing Model Components: Debris

One size doesn't fit all

Depends on:

- species
- preparation method
- individual sample quality

Response

Make switching components quick and easy

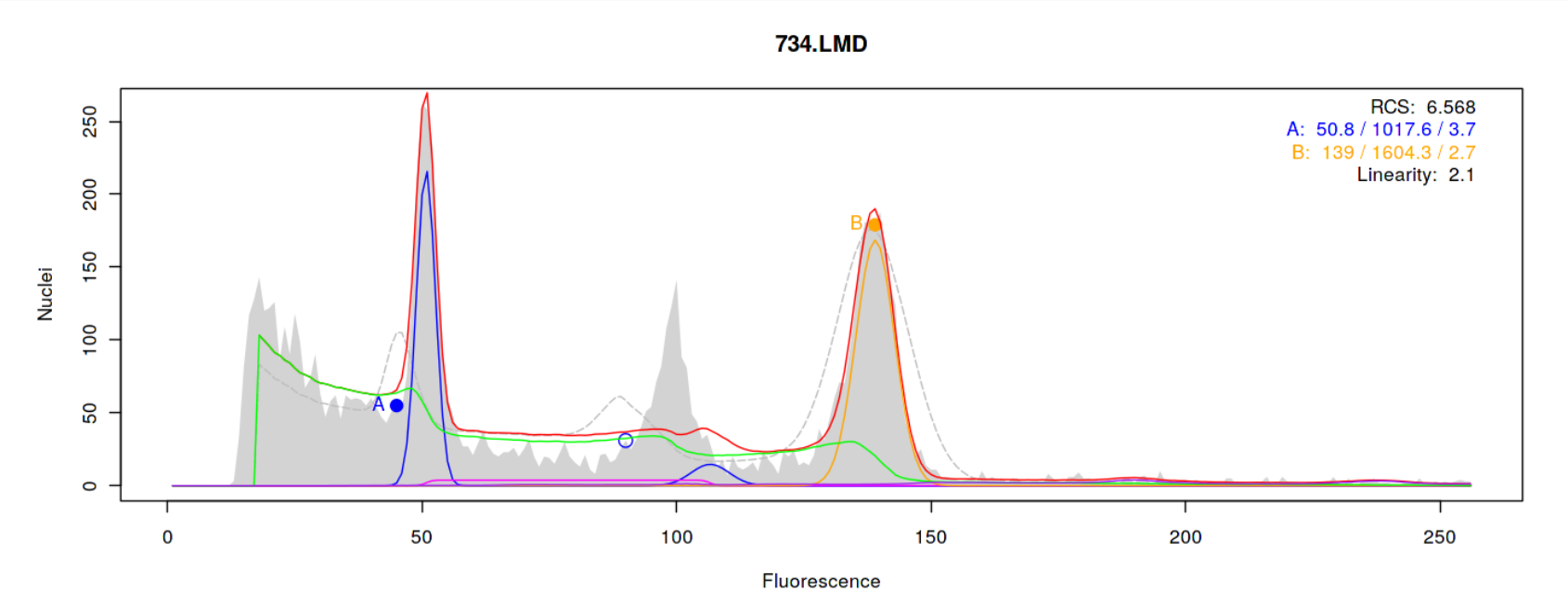
Impact on parameter estimates is usually small

RCS provides an objective basis for choice

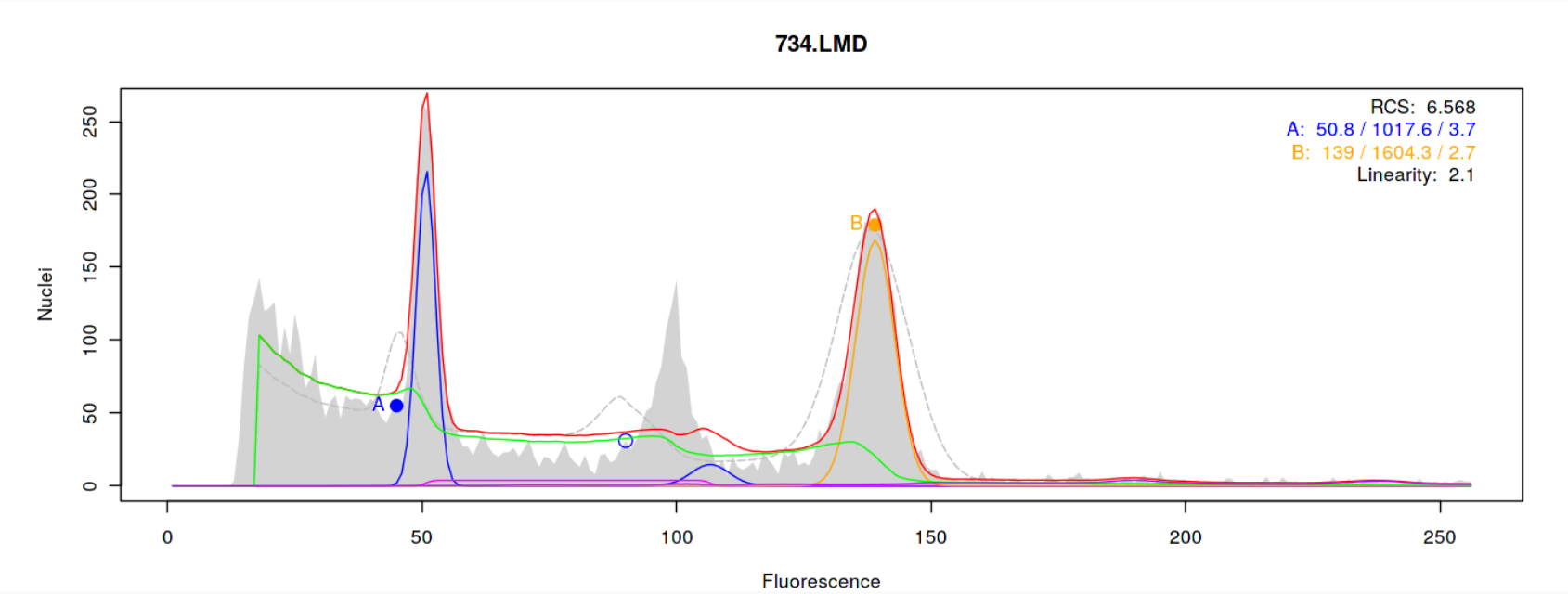
Local Minima

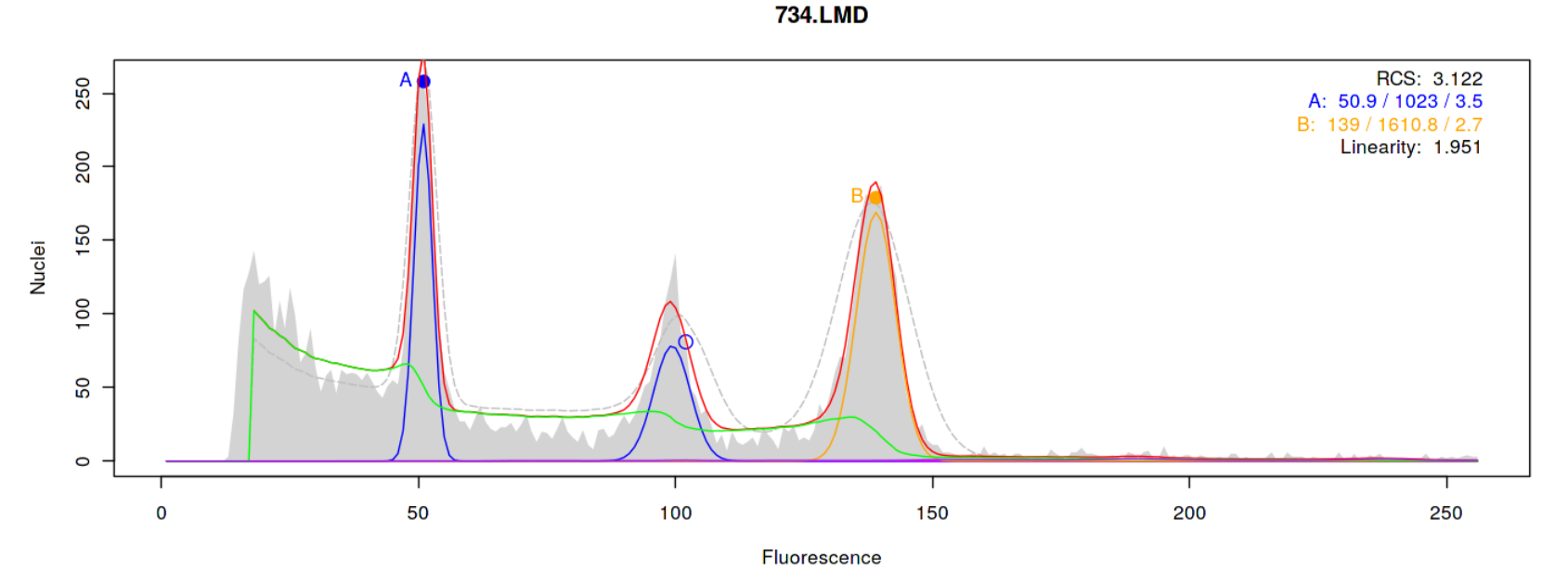
There is no direct solution for non-linear regression

Sometimes the algorithm gets stuck in a local minima:



Local Minima





Exporting Results

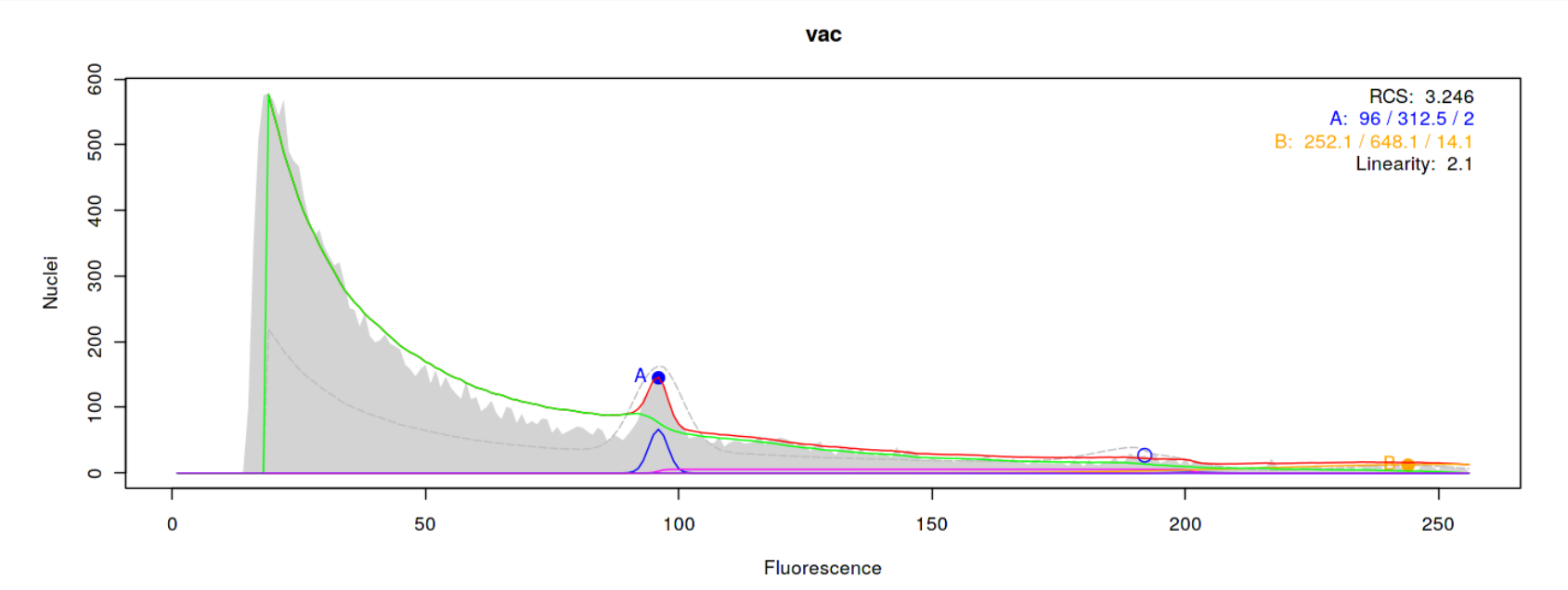
Save to file or use directly in R:

```
tabulateFlowHist(batch1)
```

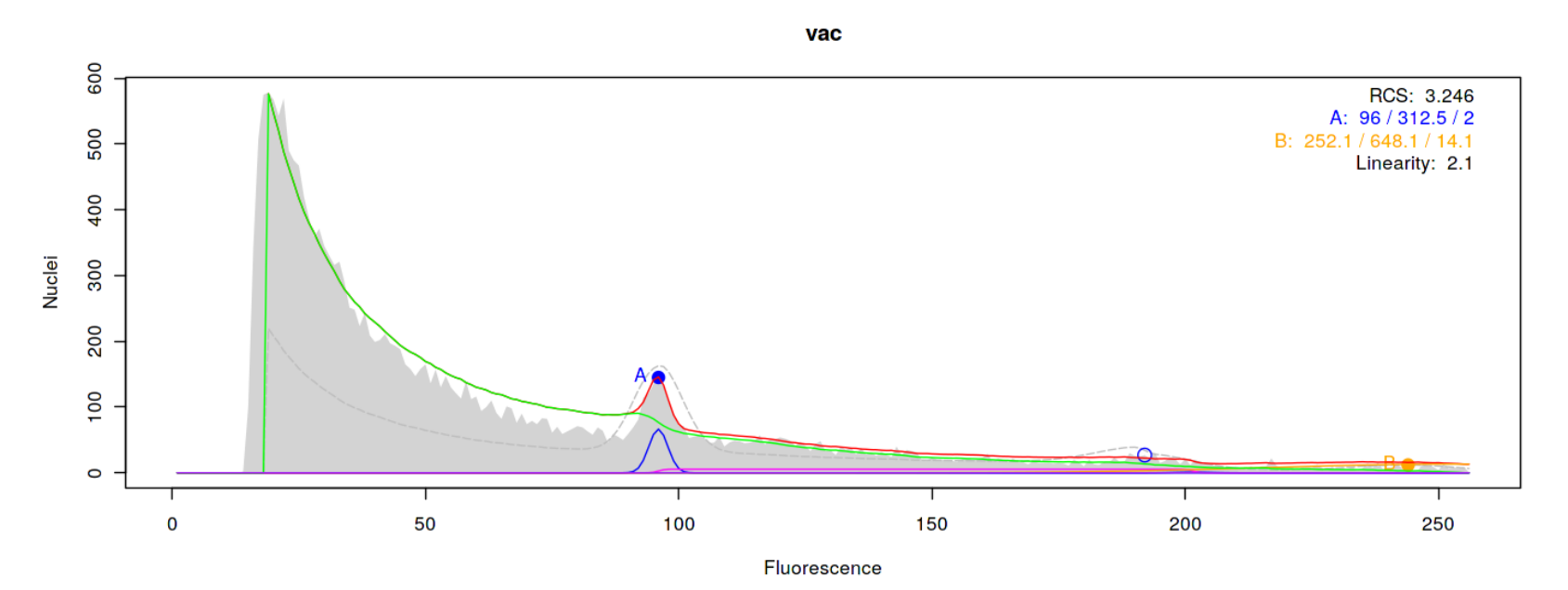
	countsA	sizeA	cvA	AB
188-15.LMD	1440.229	99.034	0.022	0.727
240-4-2+rad.LMD	449.525	64.598	0.028	0.597
248+S.LMD	2651.879	77.773	0.027	0.395

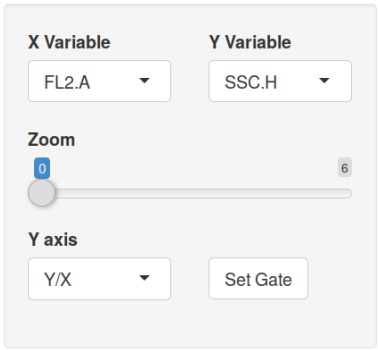
Gating

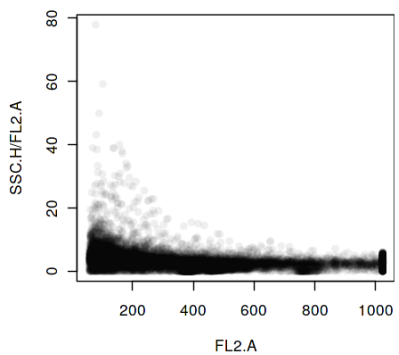
Flow data is not always pretty

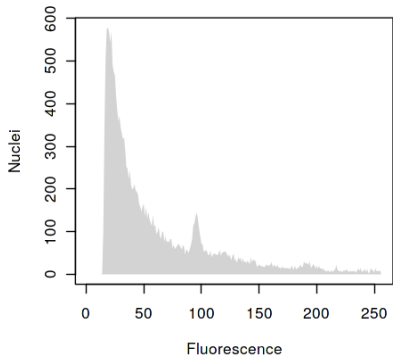


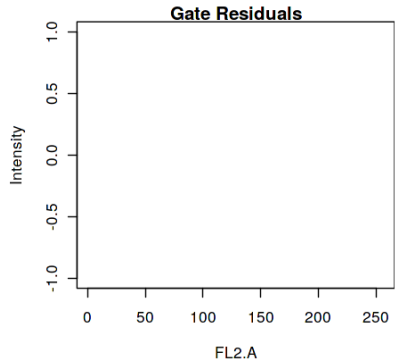
Gating


vac
RCS: 3.246
A: 96 / 312.5 / 2
B: 252.1 / 648.1 / 14.1
Linearity: 2.1
Nuclei
Fluorescence
A
B
0
100
200
300
400
500
600
0
50
100
150
200
250

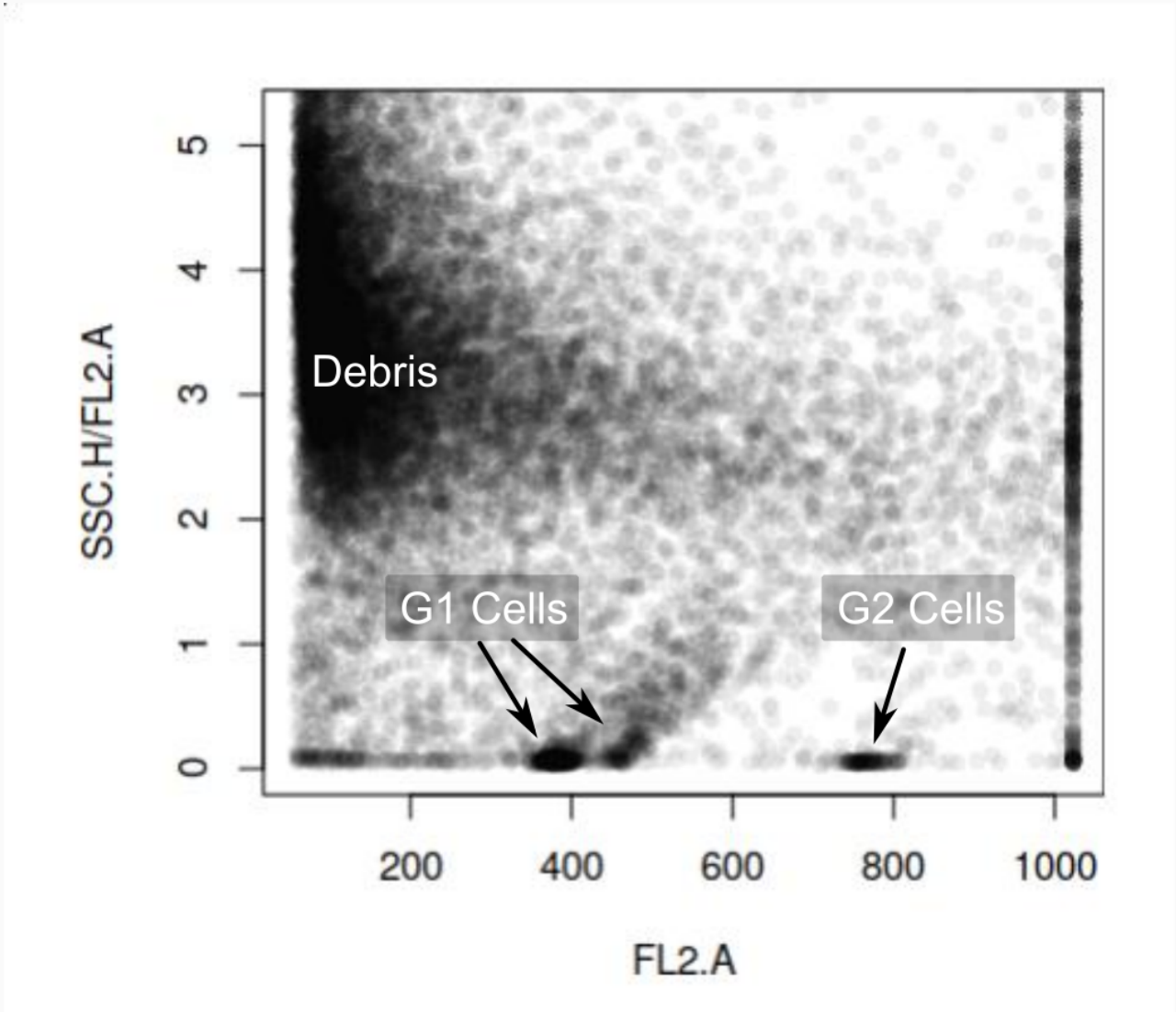

X Variable
FL2.A
Y Variable
SSC.H
Zoom
0
6
Y axis
Y/X
Set Gate


SSC.H/FL2.A
FL2.A
0
20
40
60
80
200
400
600
800
1000

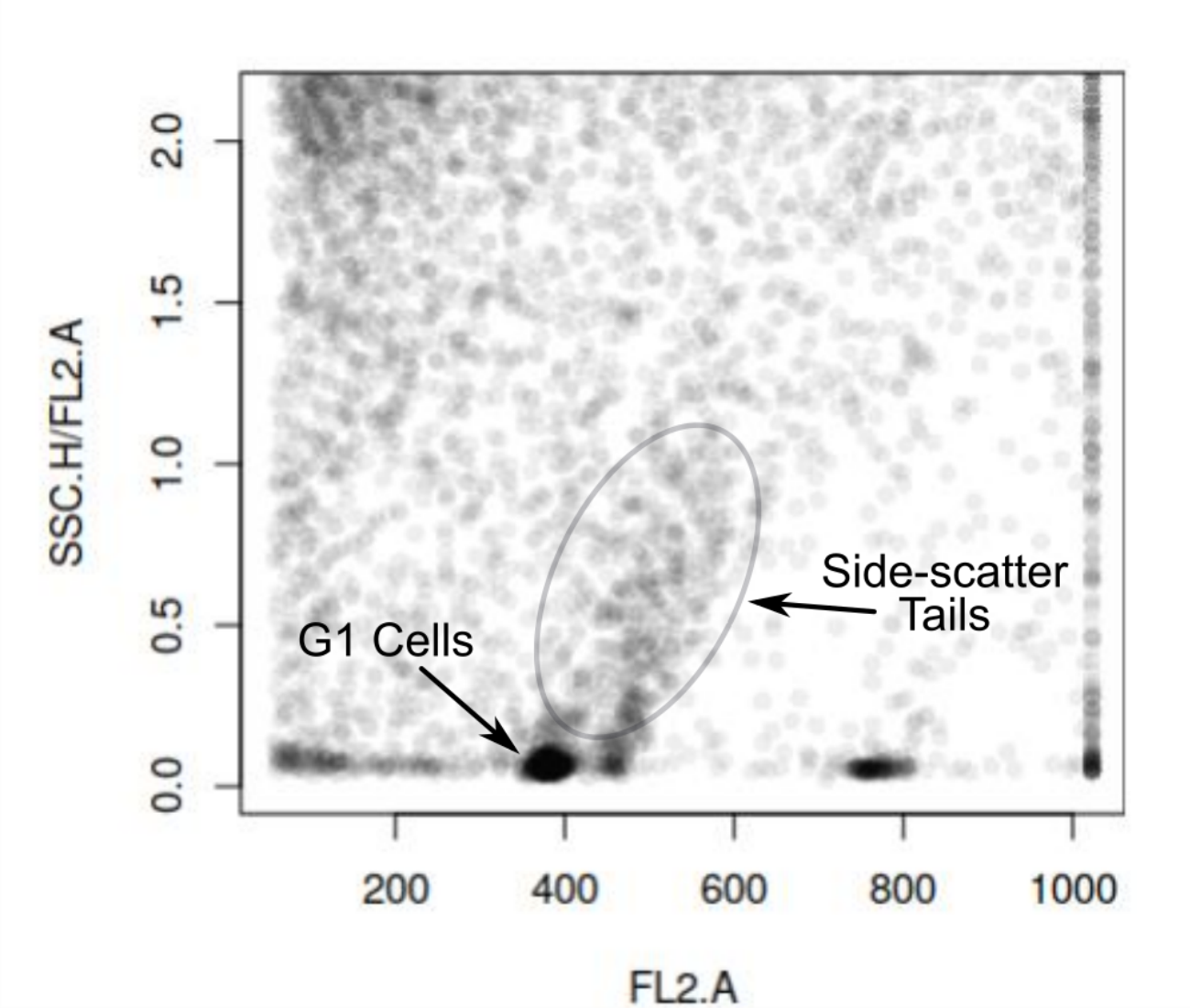

Nuclei
Fluorescence
0
100
200
300
400
500
600
0
50
100
150
200
250


Gate Residuals
Intensity
FL2.A
-1.0
-0.5
0.0
0.5
1.0
0
50
100
150
200
250

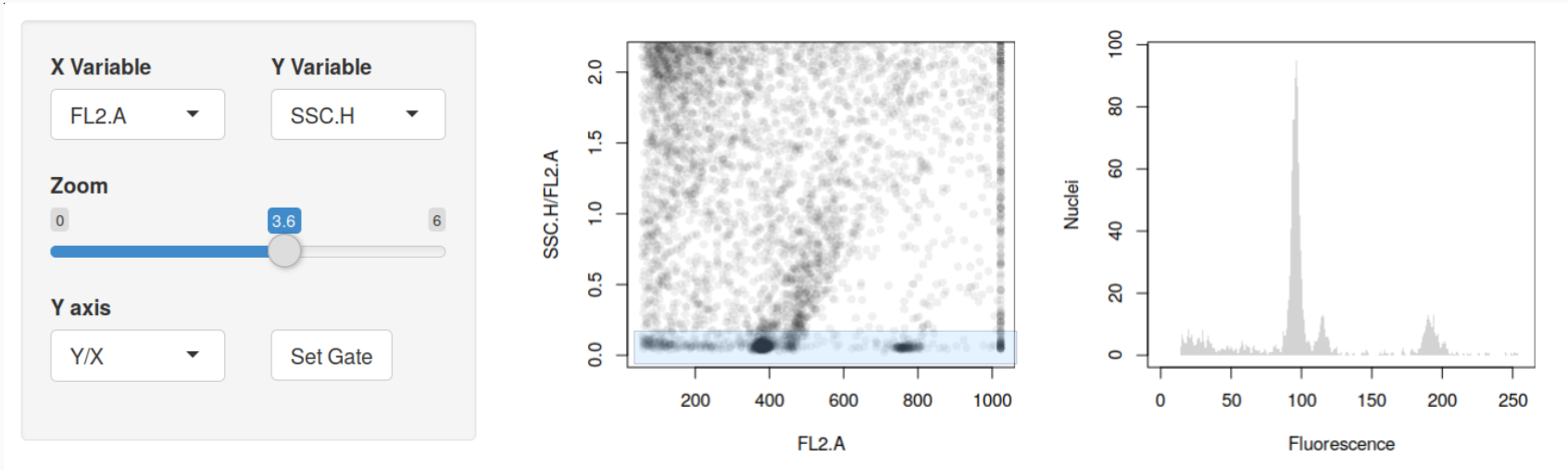
Gating


Debris
G1 Cells
G2 Cells
SSC.H/FL2.A
FL2.A
0
1
2
3
4
5
200
400
600
800
1000

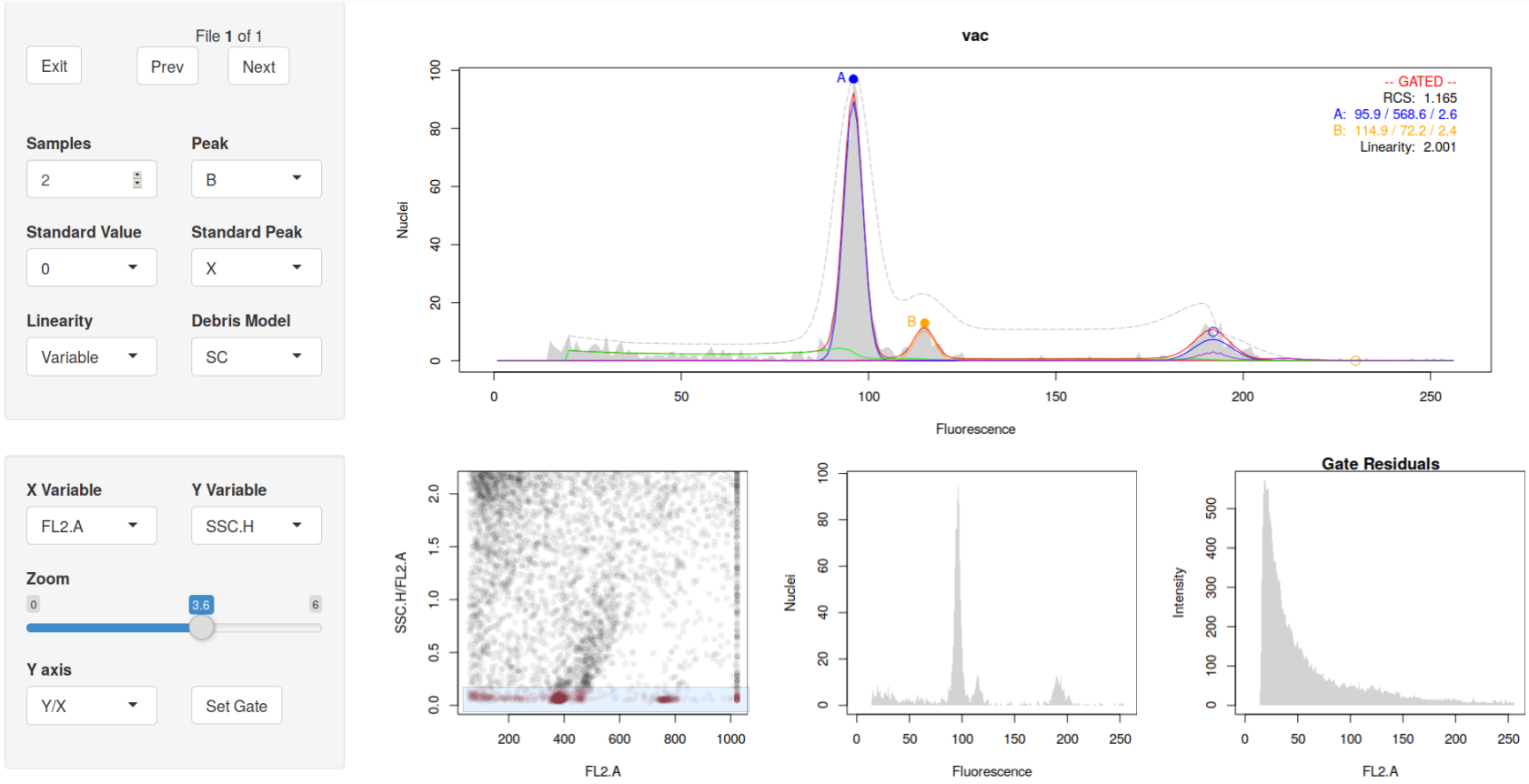
Gating


SSC.H/FL2.A
0.0
0.5
1.0
1.5
2.0
200
400
600
800
1000
FL2.A
G1 Cells
Side-scatter
Tails

Gating


X Variable
FL2.A
Y Variable
SSC.H
Zoom
0
3.6
6
Y axis
Y/X
Set Gate
SSC.H/FL2.A
0.0
0.5
1.0
1.5
2.0
200
400
600
800
1000
FL2.A
Nuclei
0
20
40
60
80
100
0
50
100
150
200
250
Fluorescence

Gating


File 1 of 1
Exit
Prev
Next
Samples
2
Peak
B
Standard Value
0
Standard Peak
X
Linearity
Variable
Debris Model
SC
X Variable
FL2.A
Y Variable
SSC.H
Zoom
0
3.6
6
Y axis
Y/X
Set Gate
vac
-- GATED --
RCS: 1.165
A: 95.9 / 568.6 / 2.6
B: 114.9 / 72.2 / 2.4
Linearity: 2.001
Nuclei
Fluorescence
SSC.H/FL2.A
FL2.A
Nuclei
Fluorescence
Gate Residuals
Intensity
FL2.A

`flowPloidy` Summary

Compared to modFit:

- parameter estimates within 1%
- simpler interface
- integration with R
- cost

Compared to manual analysis:

- objective and repeatable

Gating:

- less sensitive to subjective gating decisions

`flowPloidy` Future Development

Gating

- better define or automate best practice
- more sophisticated options for pulse analysis
- impact of gating on theoretical model components

Thank You!

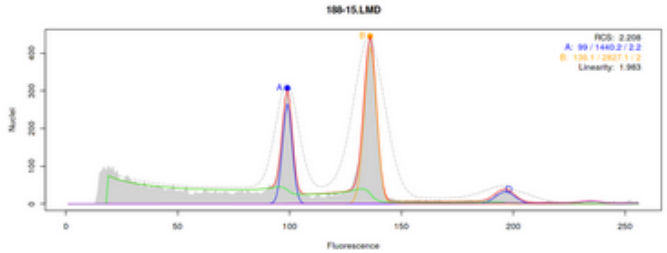
tyler@plantarum.ca

plantarum.github.io/flowPloidy

flowPloidy

An R package for flow cytometry histogram analysis

View the Project on GitHub



Introduction

A tutorial overview of `flowPloidy` is available on the Bioconductor website. This vignette is provided with the package, so once you have `flowPloidy` installed you can access it from with R (see below).