A DYNAMIC GRAPHICS APPROACH TO EXPLORING HCS DATA

The Chi-Square Works, Inc. (http://chi-square-works.com)

Abstract

HCS is multiplexed analysis of responses of individual cells in cell populations to drug treatments. While the potential of HCS in providing content-rich information is well accepted, its data analysis aspect, i.e., how to easily extract as much information out of the giant HCS data sets, remains the least addressed and yet probably the most important issue. Here we demonstrate a dynamic graphics approach to integrate many data visualization techniques long treasured by statisticians to explore HCS data sets. Two HCS data sets are used in this poster. One is from an experiment of cellular responses to a drug affecting cell cycle progression; the other is from an RNAi experiment, where the ability to examine subpopulations and correlating them to other parameters are crucial. This dynamic graphics approach empowers HCS users to examine easily any part of their data according to the biological context they know best.
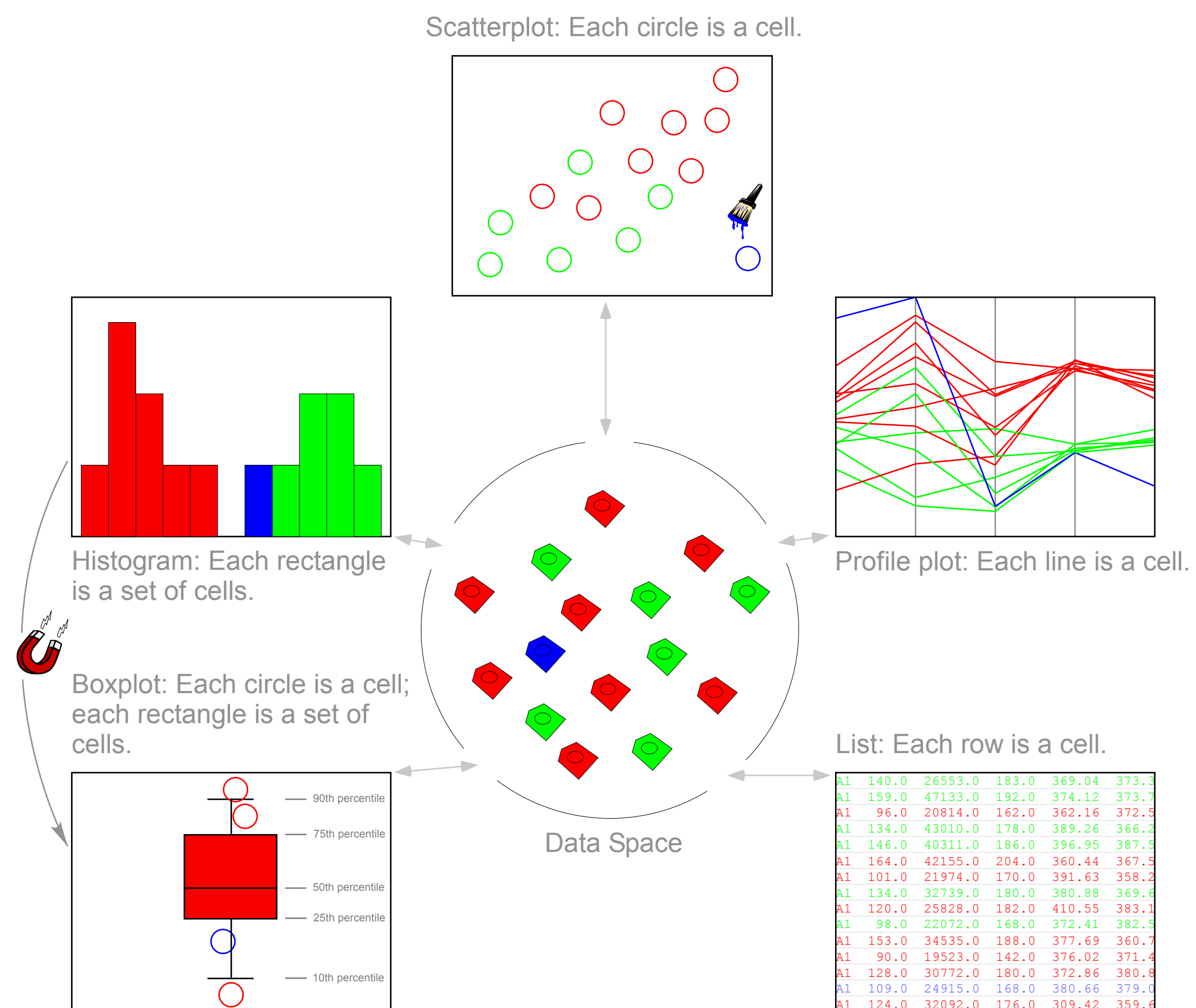
Introduction

- Data of high-content analysis/screening (HCA/HCS):

DNA   Target 1 Intensity   Target 1 Translocation   Target 2 Intensity
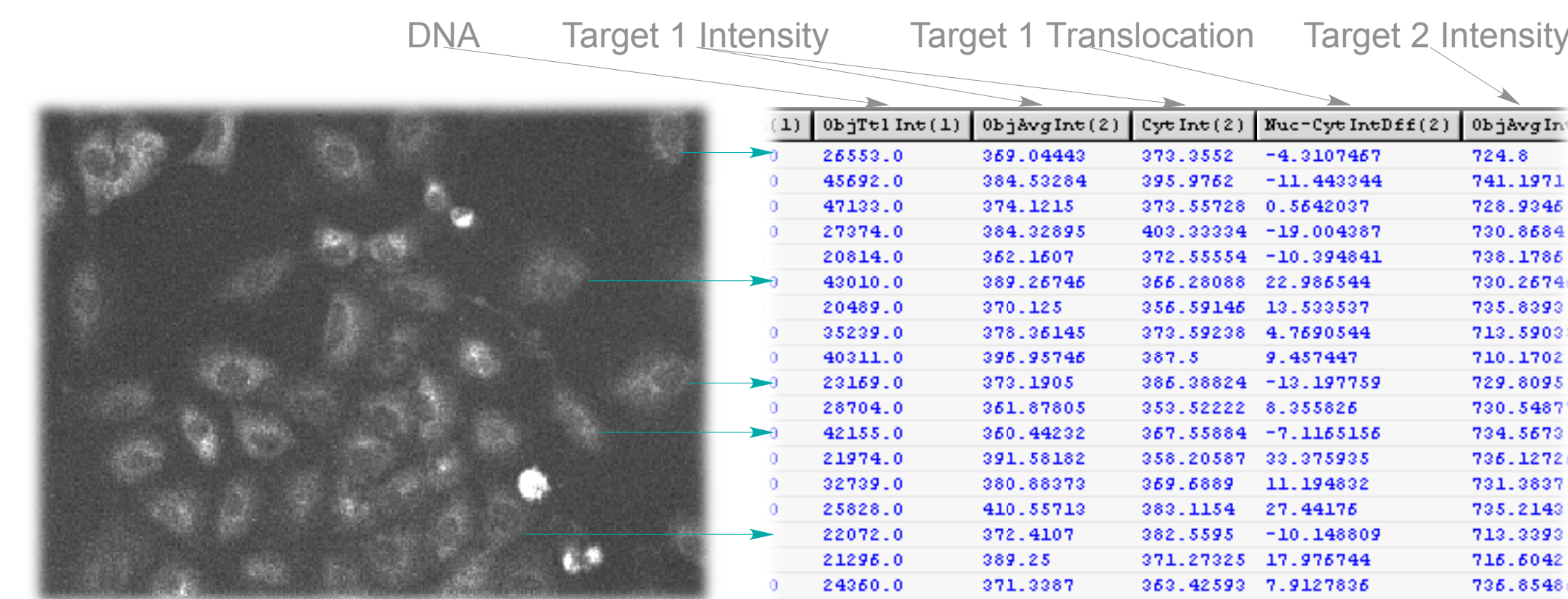
- HCA/HCS: Many parameters (or "features") are reported for each identified cell.
- In each well of microplates, many cells (up to thousands) are imaged in multiple channels and analyzed.
- Well-level summaries of features miss a lot of hidden content, such as:
  – The inherent hetrogeneity of cell populations
  – The hetrogeneity in cellular responses to stimuli
  – Sub-populations due to experimental procedures (transfection efficiency, mixed cell populations)
- Need to be able to EASILY examine HCS data at cellular level in the right biological context.
- HCS users, who have the best understanding of the context of biology and experimental design, should direct the process of information extraction by visualization, the most powerful and intuitively clear way to gain insights from any data sets.
- Traditional visualization techniques (such as histograms and scatterplots in Excel) are **static** and serve as a **one-way** communication link from computers to users. There is little users can communicate back to computers via these static plots.

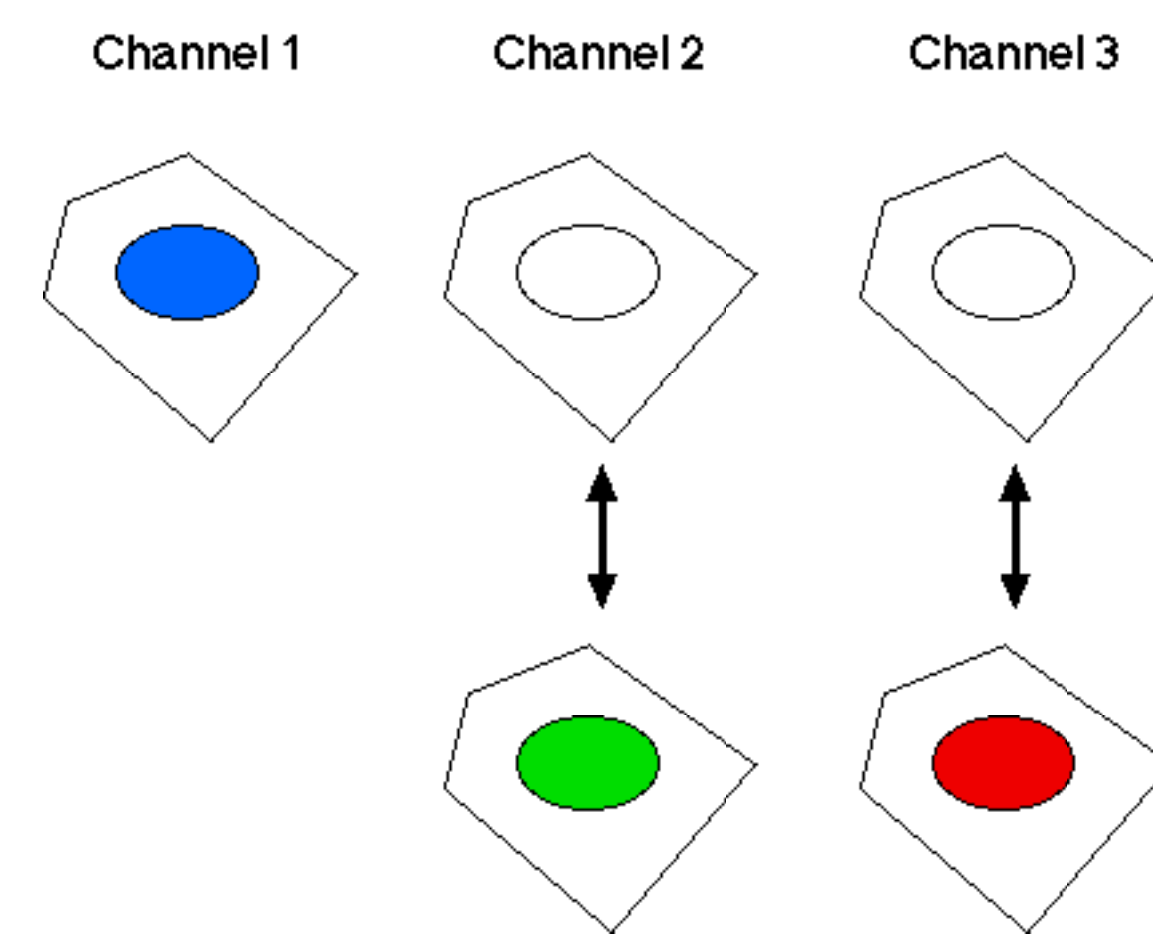Hard to integrate information in individual plots
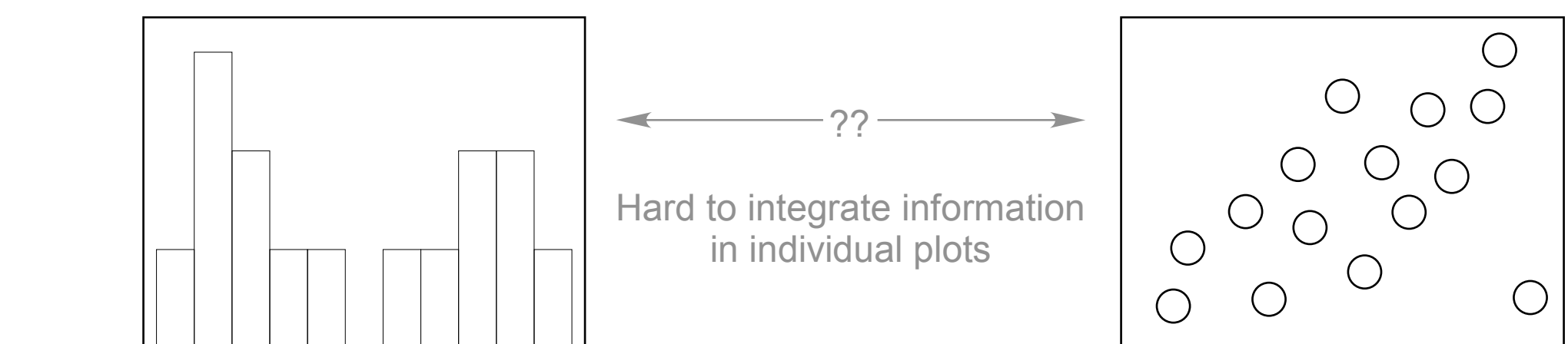
- **Dynamic** graphics treats plots as **two-way** communication links between computers and users. Computers use plots to tell users what data look like; users tell computers what actions to take via plots. Each plot is a **viewport** into the data space being explored. Users can not only look at data through various viewports but also can stick their "hands" into the data space through the viewports to manipulate the data.

Scatterplot: Each circle is a cell.

Histogram: Each rectangle is a set of cells.

Profile plot: Each line is a cell.

Boxplot: Each circle is a cell; each rectangle is a set of cells.

Data Space
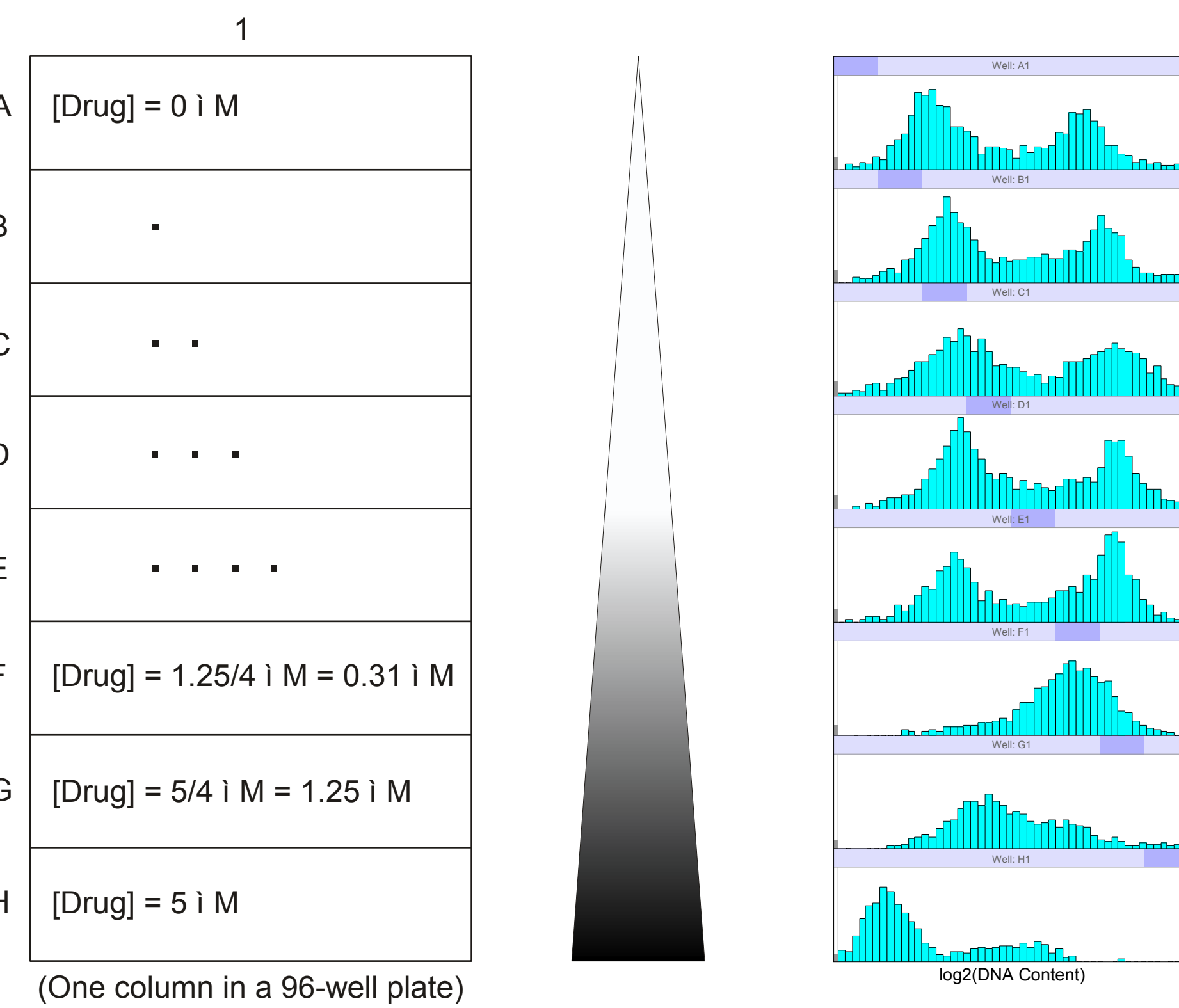
List: Each row is a cell.

Data Set I: Effect of Camptothecin on Cell Cycle Progression

- Cells treated with camptothecin for 24 hours.
- 3 cellular parameters are monitored:
  – Channel 1: DNA stain (nuclear stain intensity; DNA content)
  – Channel 2: p53 (nuclear stain intensity)
 – Channel 3: pRb (nuclear stain intensity)

Channel 1   Channel 2   Channel 3

- A trellis graph of DNA content:

	1
A	[Drug] = 0 ì M
B	.
C	. . .
D	. . .
E	. . . . .
F	[Drug] = 1.25/4 ì M = 0.31 ì M
G	[Drug] = 5/4 ì M = 1.25 ì M
H	[Drug] = 5 ì M

(One column in a 96-well plate)

log2(DNA Content)

Camptothecin has different effects on cell cycle progression at different concentration.

- Boxplots of target intensity in nuclei:

p53   pRb

log2(p53 (nuc))   log2(pRb (nuc))

Camptothecin Conc.   Camptothecin Conc.
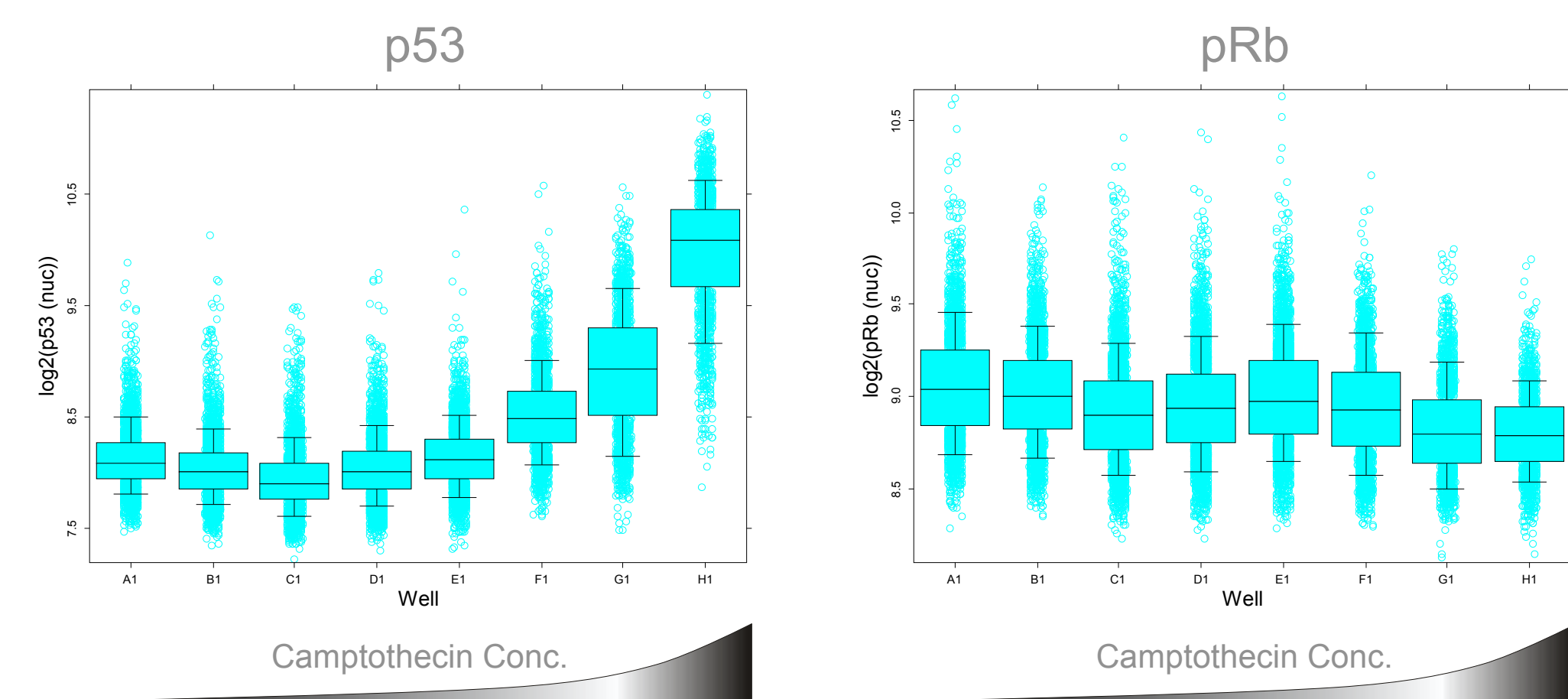
Dramatic increase in p53 stain intensity in nuclei at higher concentration. pRb stain intensity in nuclei decreases gradually as camptothecin concentration increases.

Point to ponder: If the expression LEVEL or phosphorylation STATE of a cellular TARGET changes in different phases of CELL CYCLE, does it make sense not to examine the target protein in terms of the phases of cell cycle?
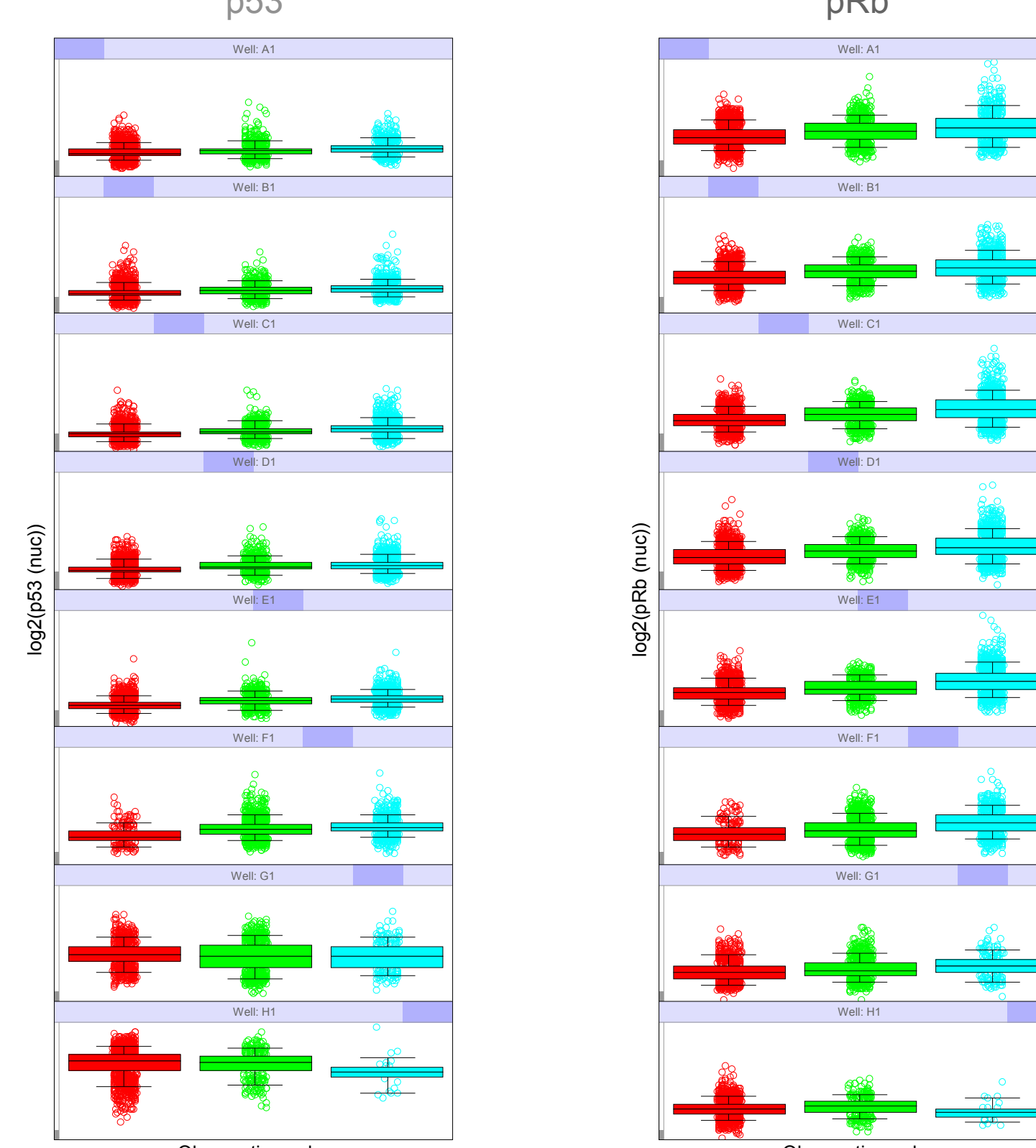
- Users can paint the histogram of DNA content into different colors:

Paint

DNA Content   DNA Content
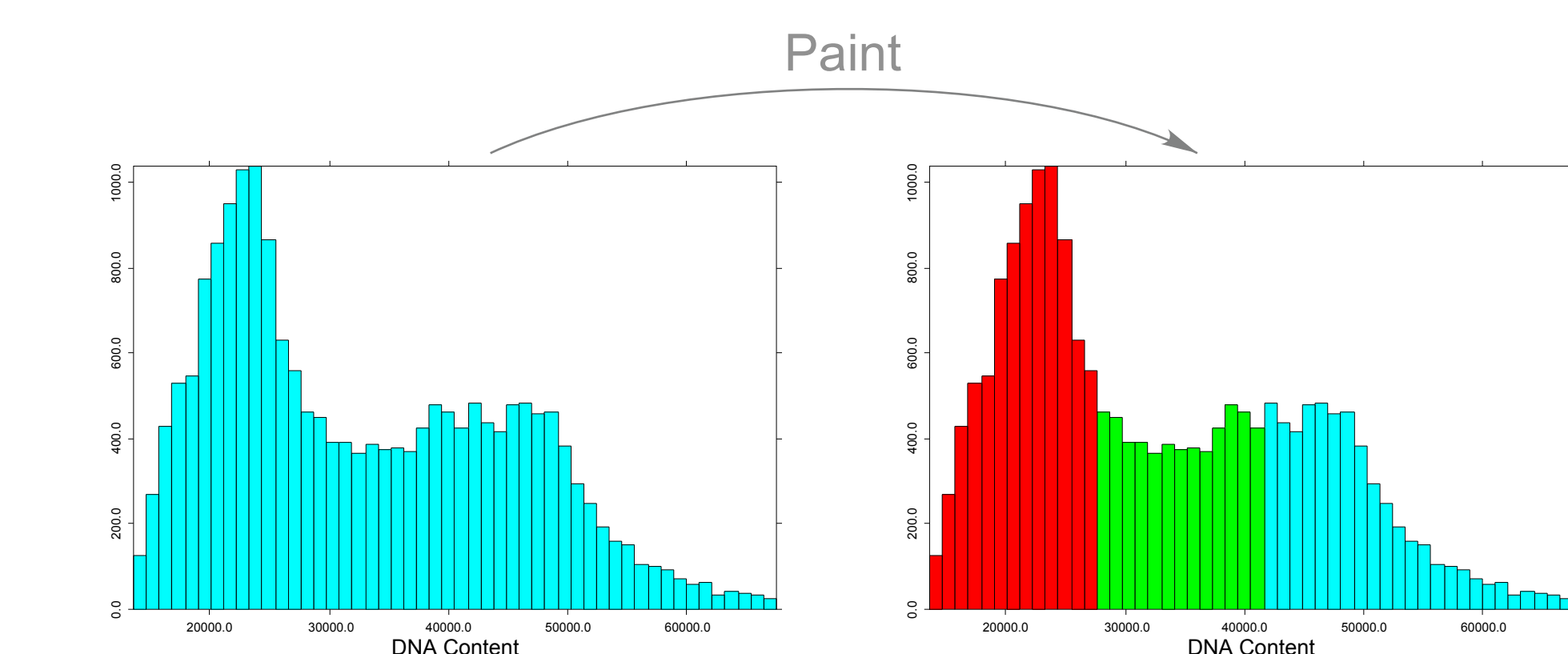
The earlier trellis graph also becomes painted:

Well A1
...
log2(DNA Content)
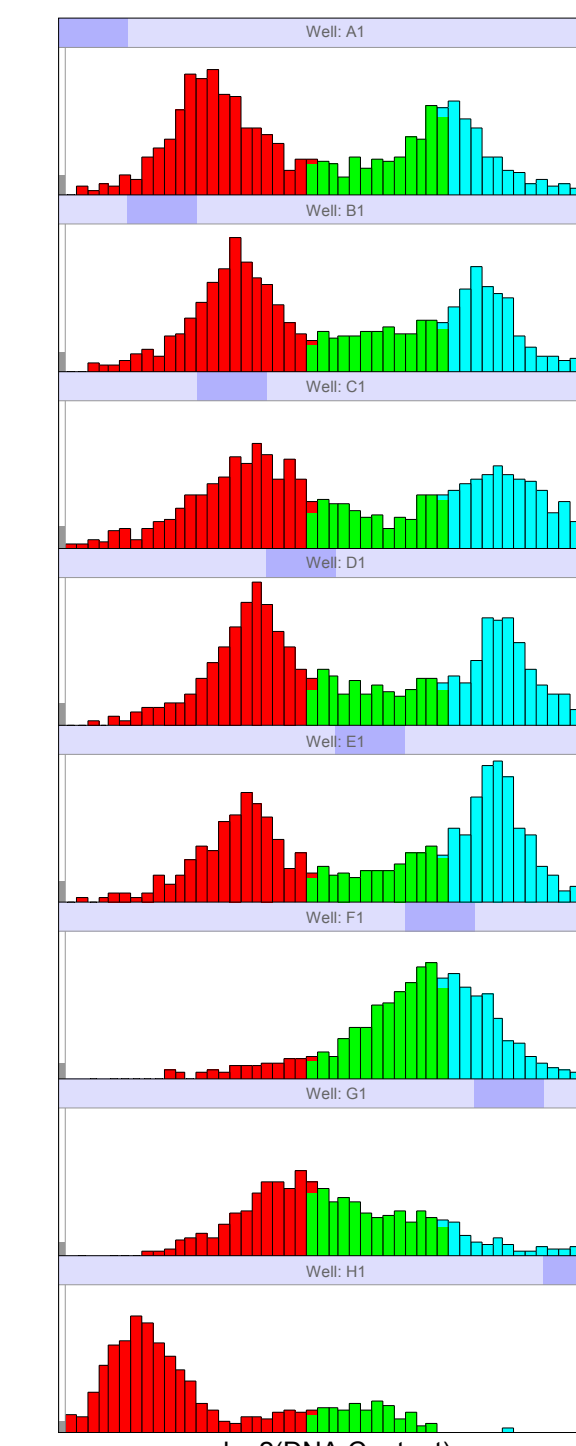
- Using color as a separator allows users easily to create 2 trellis graphs of boxplots:

p53   pRb

log2(p53 (nuc))   log2(pRb (nuc))

Observation color   Observation color

Note the differences of target intensity in cells in the 2N, 4N, and S phases of cell cycles in each of the 2 trellis graphs. Also note the trend of differences reverses in higher camptothecin concentrations.
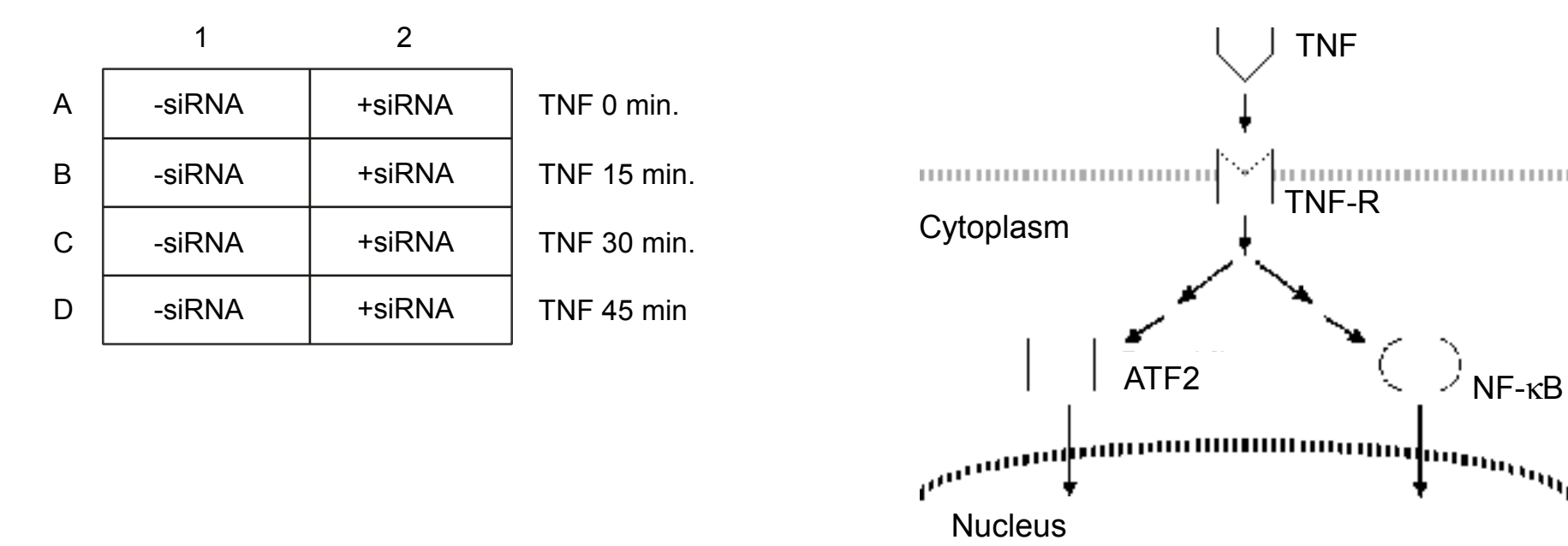
- Exploring the relationship between p53 and pRb in different parts of cell cycles: A trellis graph of scatterplots of pRb versus p53 with color as the separator.

pRb (nuc)

p53 (nuc)
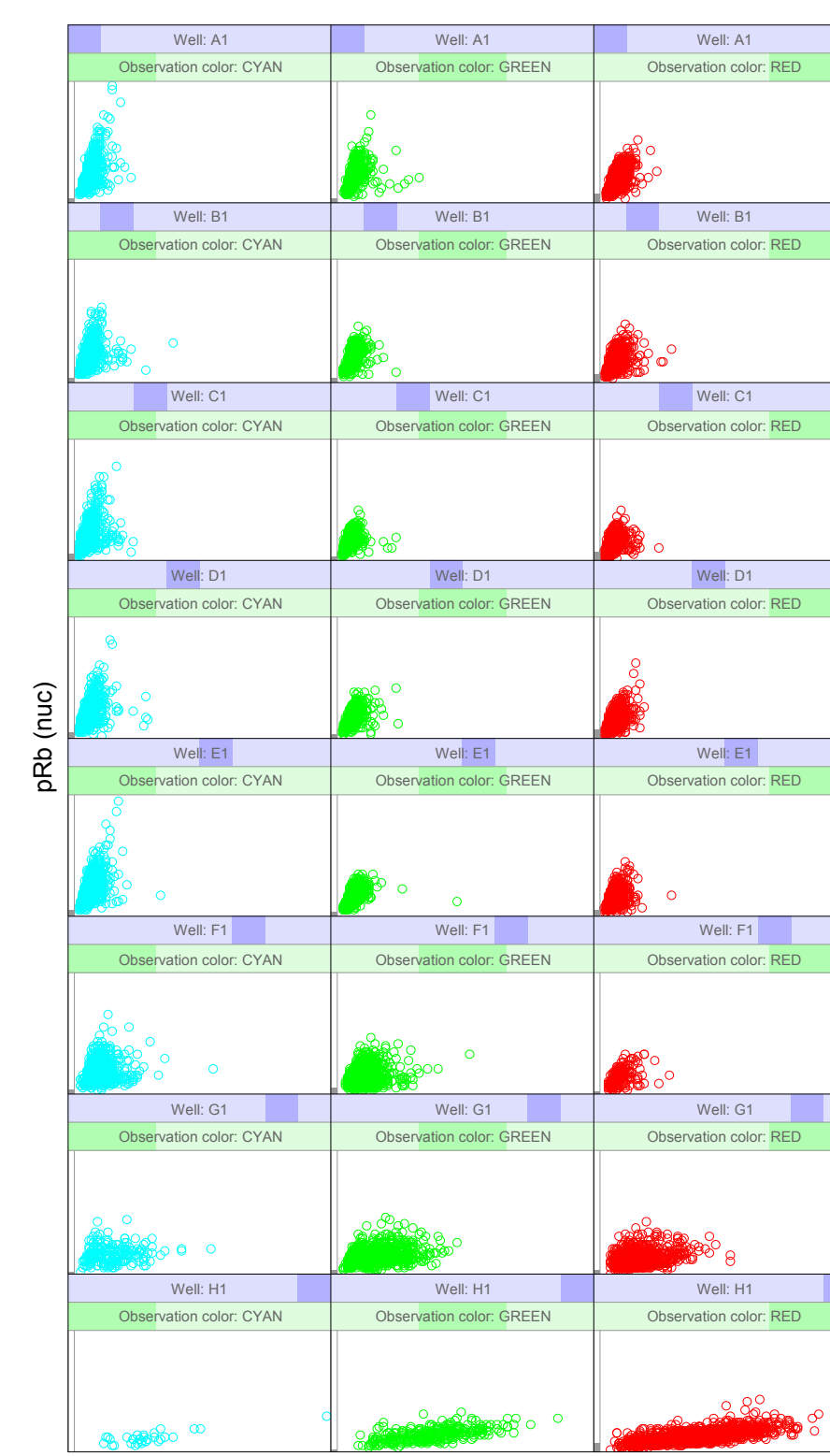
- Actually, with dynamic graphics and the right user interface, users can examine any parts of any graphs and extract data from them simply by painting the regions of interest.
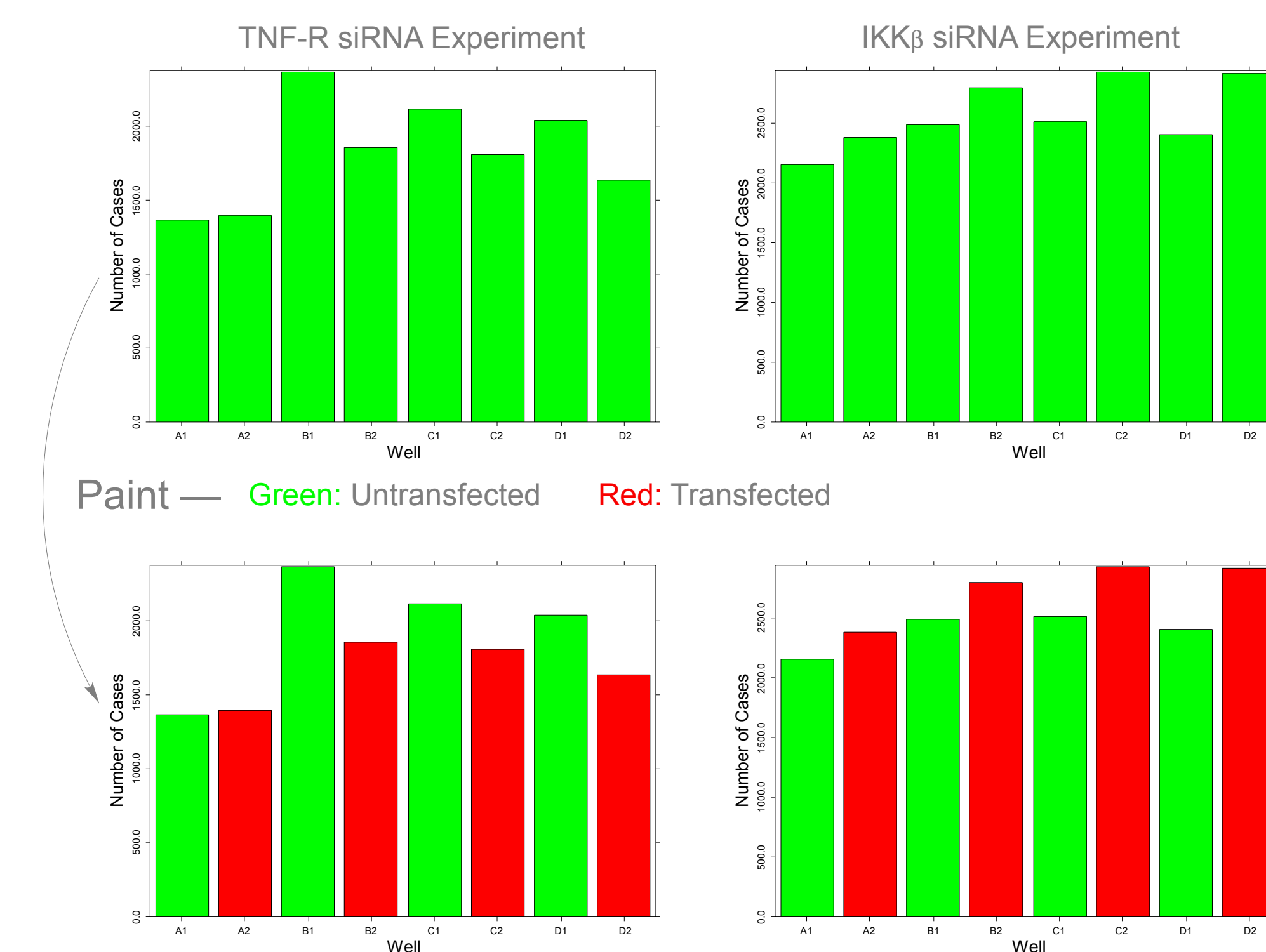
Date Set 2: siRNA Gene Knockdown in TNF-R/NF-κB Pathway

- A549 cells transiently transfected with siRNA specific to IKKβ in one experiment and TNF-R in the other experiment.

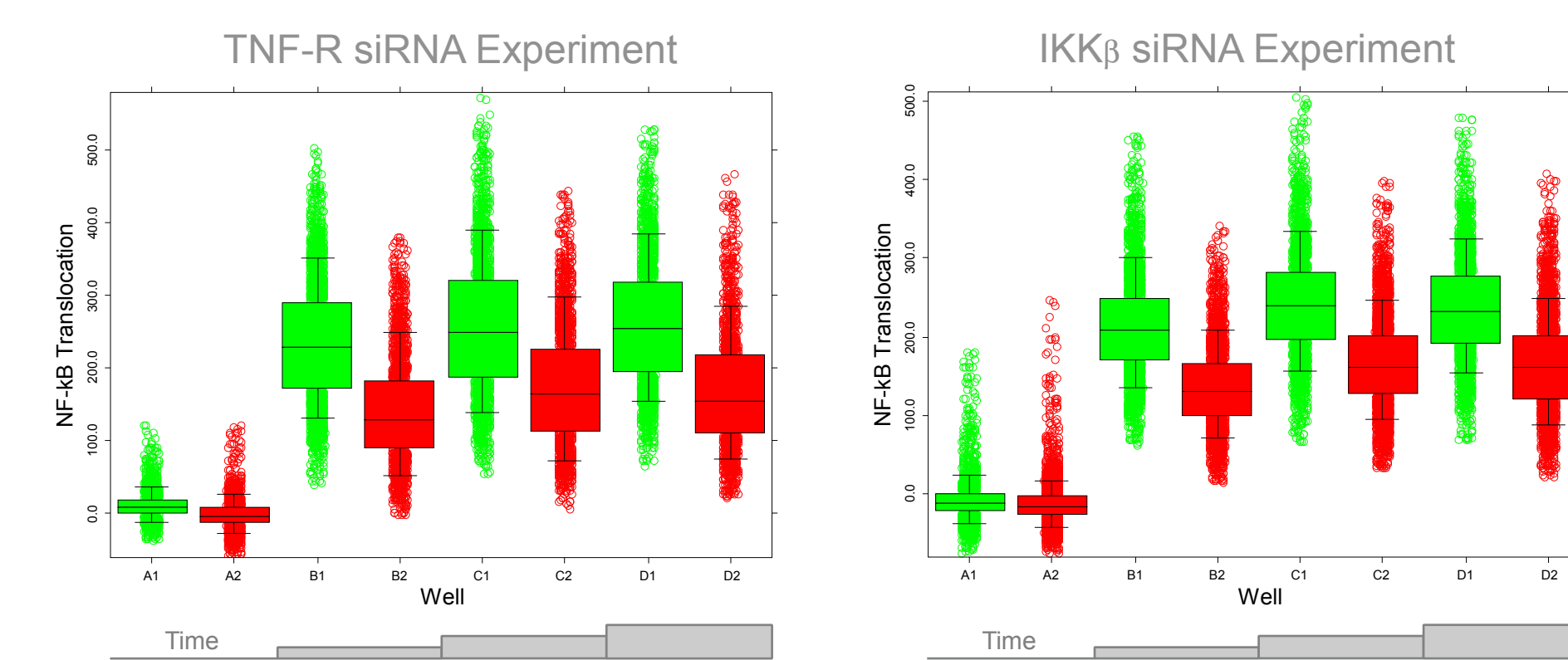|	1	2	
A	-siRNA	+siRNA	TNF 0 min.
B	-siRNA	+siRNA	TNF 15 min.
C	-siRNA	+siRNA	TNF 30 min.
D	-siRNA	+siRNA	TNF 45 min.

- Cellular parameters monitored:
  - Channel 1 — DNA stain
  - Channel 2 — NF-κB (cytoplasm-to-nucleus translocation)
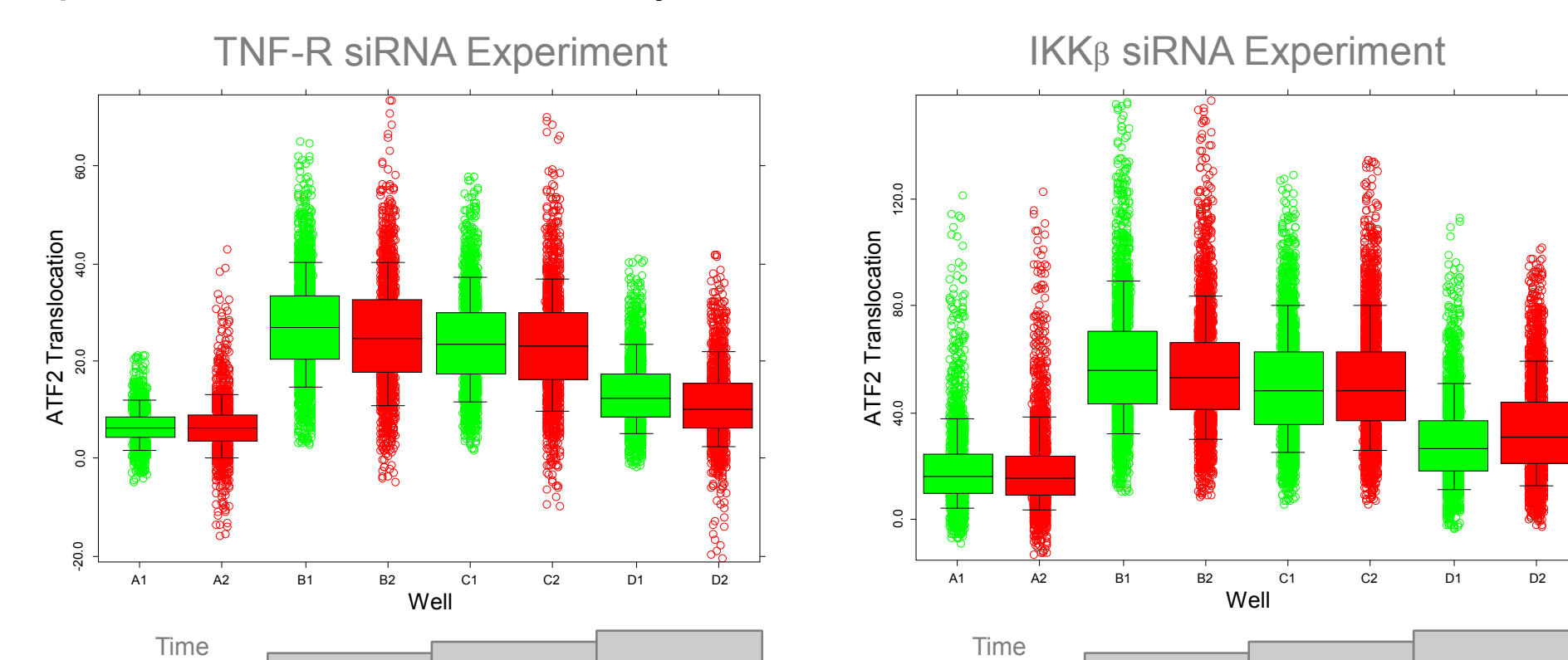 - Channel 3 — ATF2 (cytoplasm-to-nucleus translocation)

- Barplots to show numbers of cells in wells:

TNF-R siRNA Experiment   IKKβ siRNA Experiment

Number of Cases   Well

Paint —   Green: Untransfected   Red: Transfected

- Boxplots of NF-κB activation by TNF treatment:

TNF-R siRNA Experiment   IKKβ siRNA Experiment

NF-κB Translocation   Time

- Boxplots of ATF2 activation by TNF treatment:

TNF-R siRNA Experiment   IKKβ siRNA Experiment

ATF2 Translocation   Time

Note the ineffectiveness of IKKβ siRNA in reducing ATF2 activation.

- Exploring the relationship between NF-κB and ATF2 translocation:

TNF-R siRNA Experiment   IKKβ siRNA Experiment

NF-κB Translocation   ATF2 Translocation
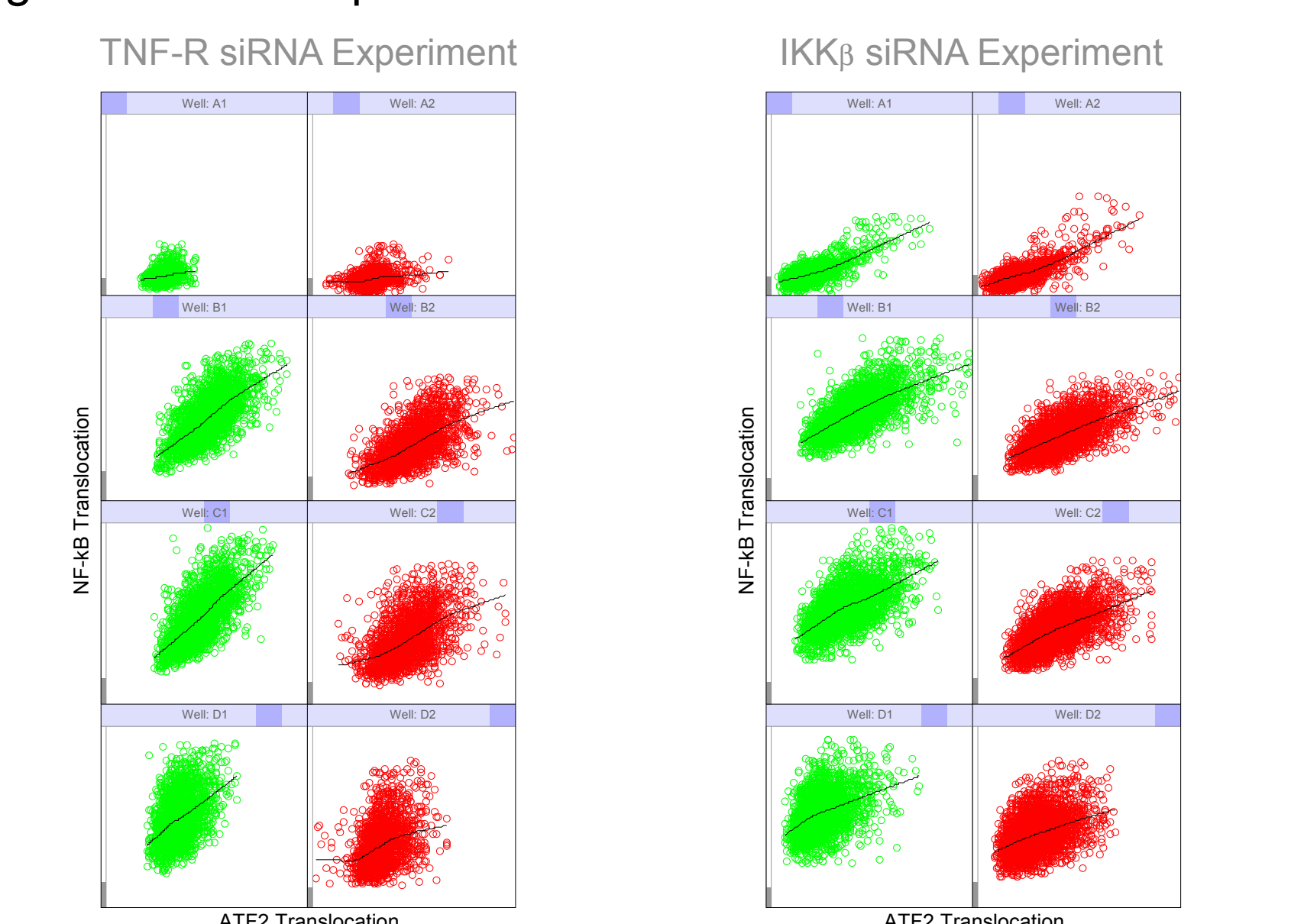
Note the consistent "unparallel" trends across rows.   Note the consistent "parallel" trends across rows.

- Any difference in DNA profiles between cells with high NF-κB and ATF2 translocation and those with low NF-κB and ATF2 translocation?
  – Pick the B2 well in the above left scatterplot trellis to explore.
  – Click the B2 panel to retrieve B2 cells. Fig. A is the result of this data retrieval.

Figure A   Figure B

NF-κB Translocation   ATF2 Translocation
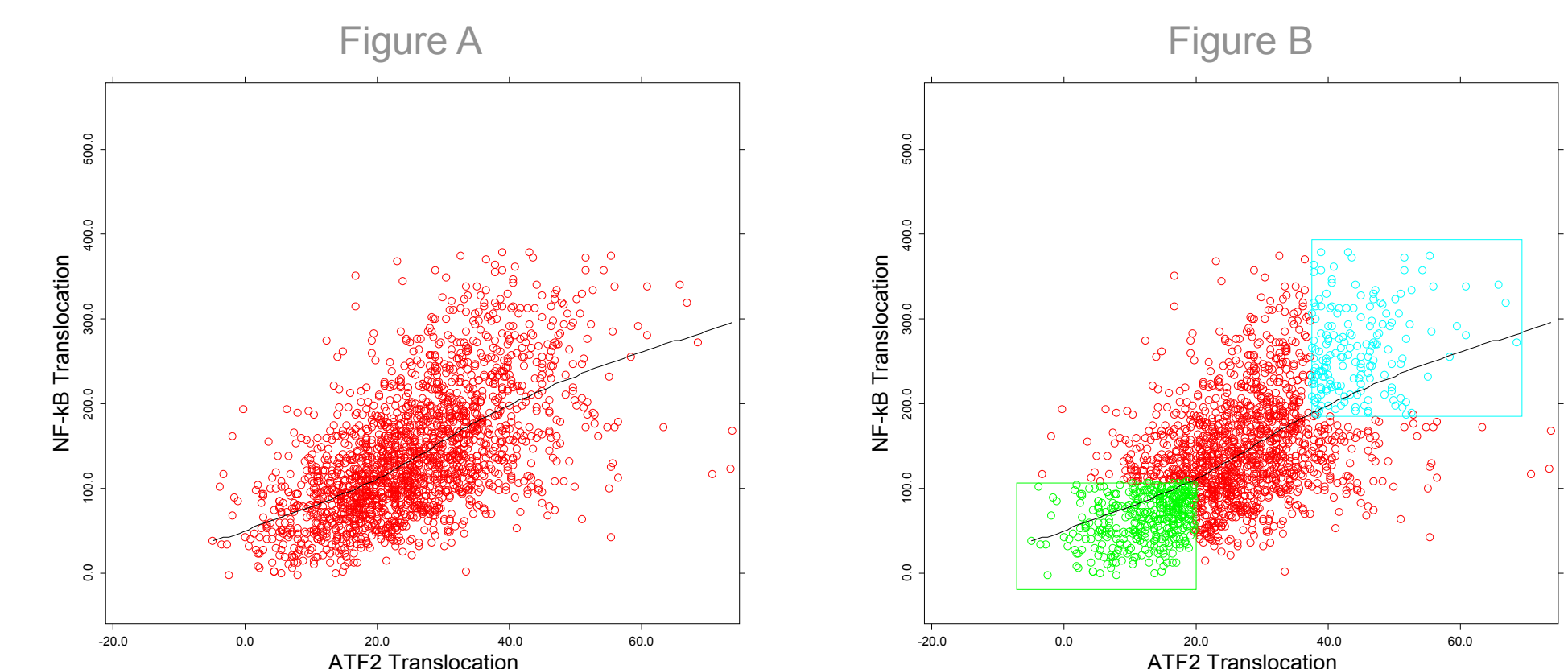
– Open 2 viewports zooming into the lower-left and the upper-right part of the point cloud in Fig. A. Paint cells in these 2 viewports green (Fig. C) and cyan (Fig. D), respectively. Fig. A changes into Fig. B.

Figure C   Figure D

NF-κB Translocation   ATF2 Translocation
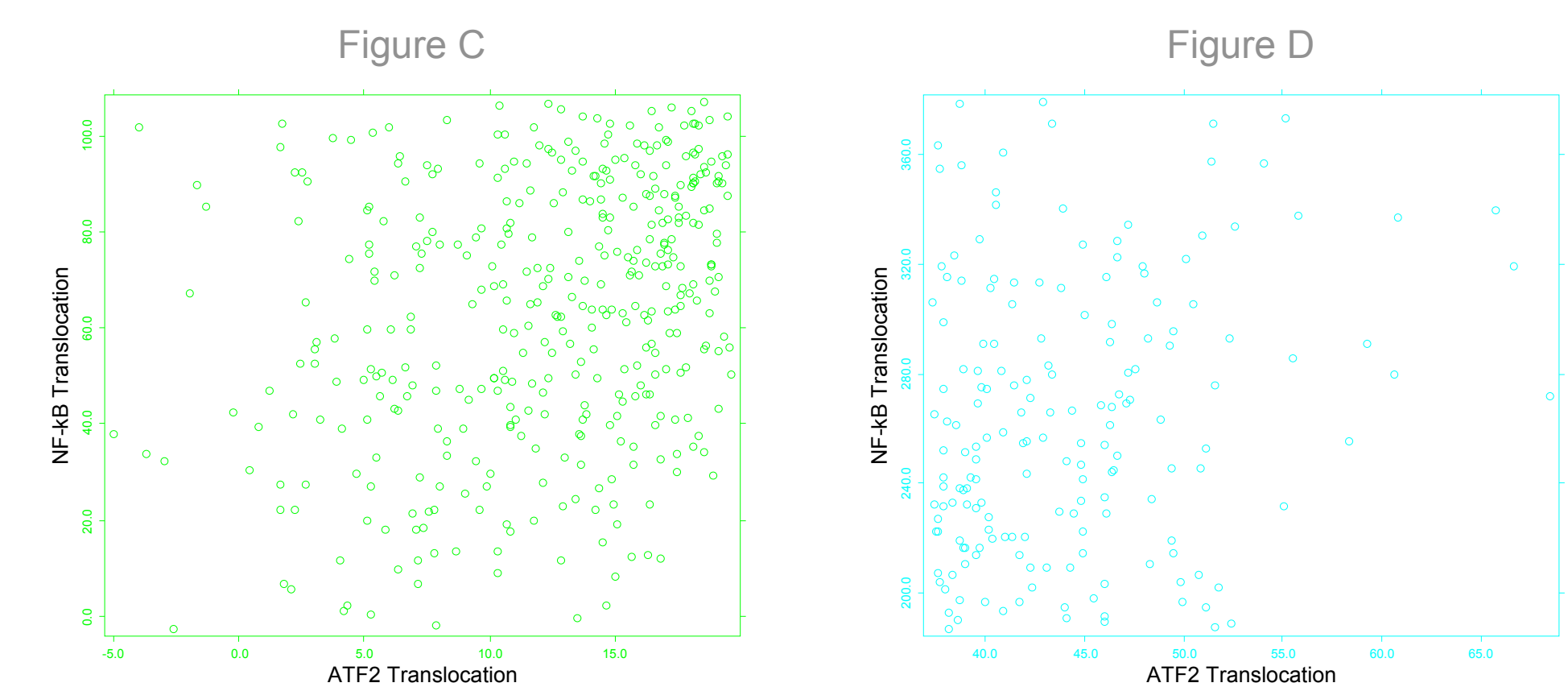
– Extract the green and cyan cells from Fig. B and do a trellis of DNA profiles with color as the separator.

DNA Content

Cyan: Cells with high NF-κB and ATF2 translocation
Green: Cells with low NF-κB and ATF2 translocation
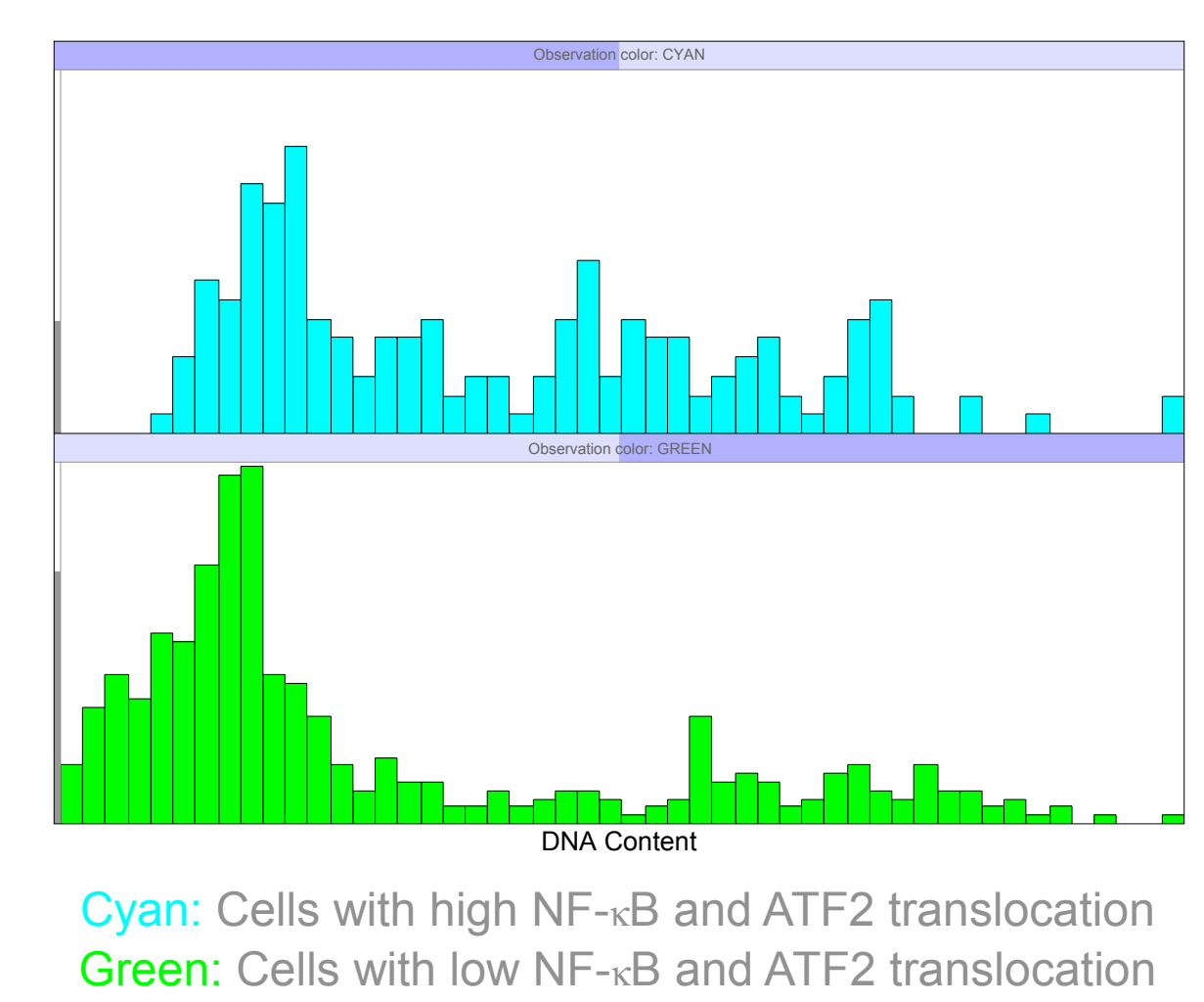
Summary

- HCS readers generate tremendous amount of numbers.
- Numbers are not information. Information (i.e.,content) is not available until uncovered.
- Visualization is the key method for uncovering hidden content in complex data sets.
- Graphics lets data talk for themselves and is the best way to explore data. It is also the first thing experienced statisticians do during data analysis. But statisticians do not necessarily know biology.
- Dynamic graphics coupled with carefully designed user interface can empower biologists to play a leading role in exploring HCA/HCS data in the biological context they know best.
- With dynamic graphics and the right user interface,
  – Users can easily choose and extract at will any subset of data in any combination to examine.
  – The result of data retrieval and analysis can be in the form of plots again.
  – Any plot generated during data exploration can be used as a new starting point to explore any other parameters in the data set.
  – Information in individual plots is integrated by painting to form a coherent image of the data as a whole.
- All the statistical plots in this poster are generated with Panmo, *the* data exploration and visualization system based on the dynamic graphics approach.