Interuniversity Institute for Biostatistics
and statistical Bioinformatics

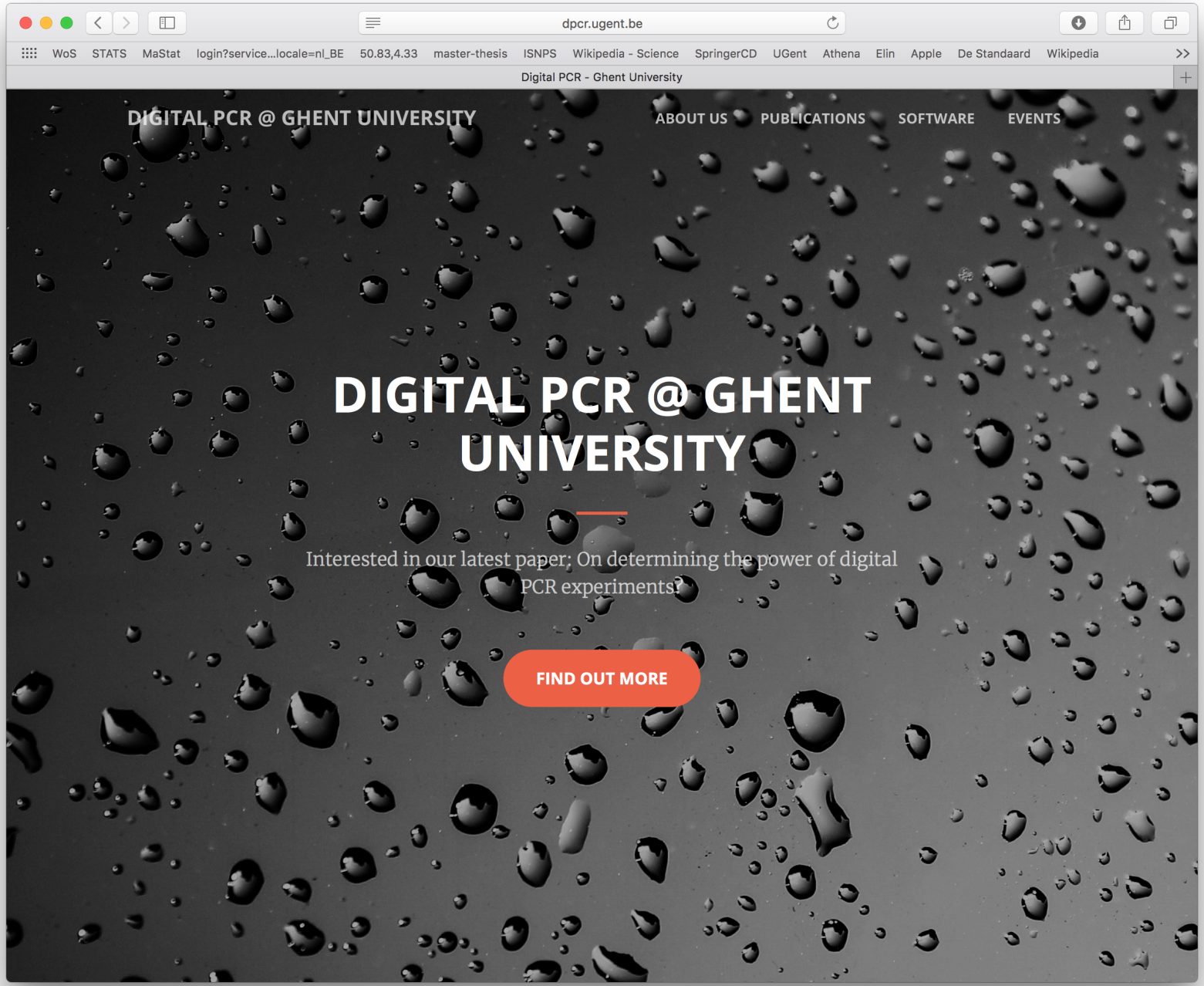
Data Analysis Pipelines and Tools for dPCR Experiments

Olivier Thas

Center for Statistics, Hasselt University, Belgium
Department of Data Analysis and Mathematical Modelling,
Ghent University, Belgium
National Institute for Applied Statistics Research Australia (NIASRA),
University of Wollongong, Australia

Joint work with

- Matthijs Vynck
- Jo Vandesompele
- Ward De Spiegelaere
- Wim Trypsteen
- ...


dpcr.ugent.be
WoS STATS MaStat login?service...locale=nl_BE 50.83,4.33 master-thesis ISNPS Wikipedia - Science SpringerCD UGent Athena Elin Apple De Standaard Wikipedia
Digital PCR - Ghent University
DIGITAL PCR @ GHENT UNIVERSITY
ABOUT US PUBLICATIONS SOFTWARE EVENTS
DIGITAL PCR @ GHENT UNIVERSITY
Interested in our latest paper; On determining the power of digital PCR experiments?
FIND OUT MORE

In this talk ...

- Threshold setting
- Quality Control
- Analysis of complex experiments (absolute quantification and CNV)

Threshold setting

Anal Bioanal Chem (2015) 407:5827–5834
DOI 10.1007/s00216-015-8773-4

RESEARCH PAPER

ddpcRquant: threshold determination for single channel droplet digital PCR experiments

Wim Trypsteen[1] · Matthijs Vynck[2] · Jan De Neve[2] · Pawel Bonczkowski[1] · Maja Kiselinova[1] · Eva Malatinkova[1] · Karen Vervisch[1] · Olivier Thas[2,3] · Linos Vandekerckhove[1] · Ward De Spiegelaere[1]

Threshold setting: Need for a threshold

The calculation of dPCR data starts from 0/1 outcomes, but measurements consists of intensity measurement.

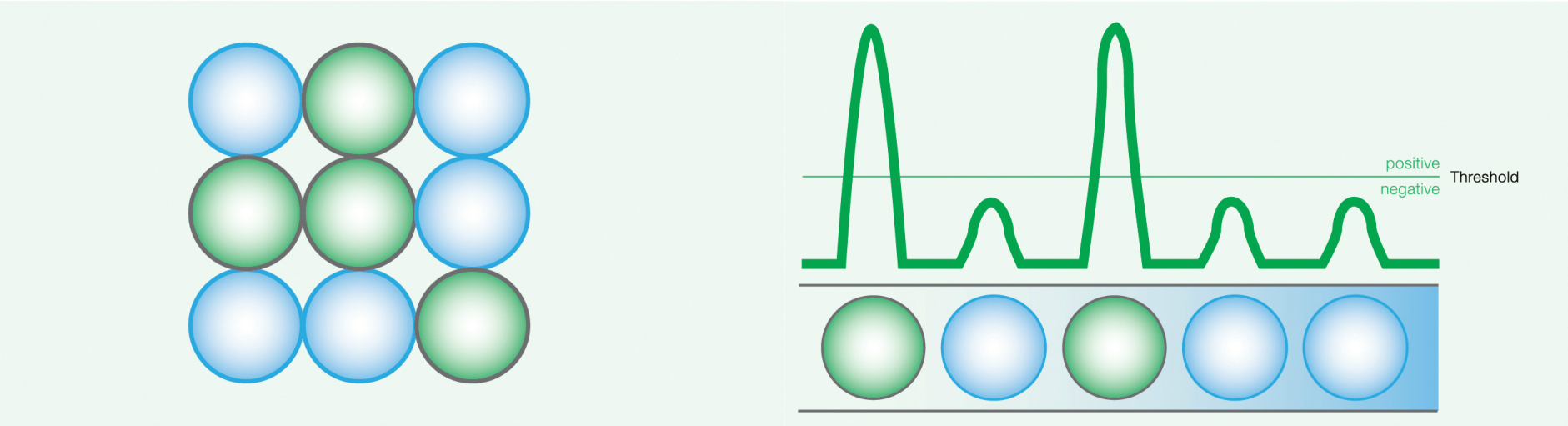


Figure: Measuring presence/absence (Bio-Rad Laboratories)

Even negative droplets show an intensity signal.

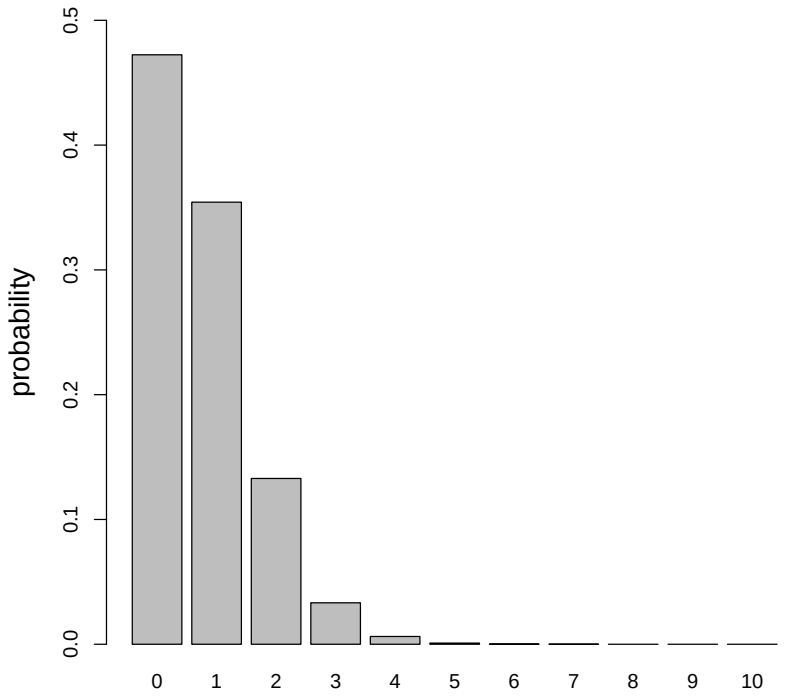
$$Y_j = \begin{cases} 0 & \text{if intensity if droplet } j < \text{threshold (negative droplet)} \\ 1 & \text{if intensity if droplet } j > \text{threshold (positive droplet)} \end{cases}$$

Threshold setting: Classical approach to quantification

dPCR data analysis methods rely on the Poisson assumption,

$$Y_j^* \sim \text{Poisson}(\lambda) \quad j = 1, \ldots, N,$$

where $\lambda = \text{E}\left\{ Y_j^* \right\}$ is the average number of target copies in a droplet.



Threshold setting: Classical approach to quantification

dPCR data analysis methods rely on the Poisson assumption,

$$Y_j^* \sim \text{Poisson}(\lambda) \quad j = 1, \ldots, N,$$

where $\lambda = \text{E}\left\{ Y_j^* \right\}$ is the average number of target copies in a droplet.

If λ can be estimated, say $\hat{\lambda}$, the target concentration can be estimated as

$$\hat{c} = \frac{\hat{\lambda}}{V_{\text{droplet}}}$$

where V_{droplet} is the average droplet volume (e.g. 0.85 nL).

Threshold setting: Classical approach to quantification

What we observe is the digital outcome

$$Y_j = \min(Y_j^*, 1) = \begin{cases} 0 & \text{if } Y_j^* = 0 \text{ (negative droplet)} \\ 1 & \text{otherwise (positive droplet).} \end{cases}$$

The distribution function of the Poisson distribution is given by

$$\mathrm{P}\{Y^* = y\} = \frac{\lambda^y}{y!} \exp(-\lambda).$$

Hence, for the observable Y, the probability of a negative droplet is

$$\mathrm{P}\{Y = 0\} = \mathrm{P}\{Y^* = 0\} = \frac{\lambda^0}{0!} \exp(-\lambda) = \exp(-\lambda).$$

This gives an expression for λ,

$$\lambda = -\ln \mathrm{P}\{Y = 0\}.$$

Threshold setting: Classical approach to quantification

The probability $\mathrm{P}\{Y = 0\}$ is easily estimated as

$$\widehat{\mathrm{P}\{Y = 0\}} = \frac{\text{number of negative droplets}}{\text{total number of droplets}}.$$

The Poisson parameter can thus be estimated as

$$\hat{\lambda} = -\ln\left(\frac{\text{number of negative droplets}}{\text{total number of droplets}}\right)$$

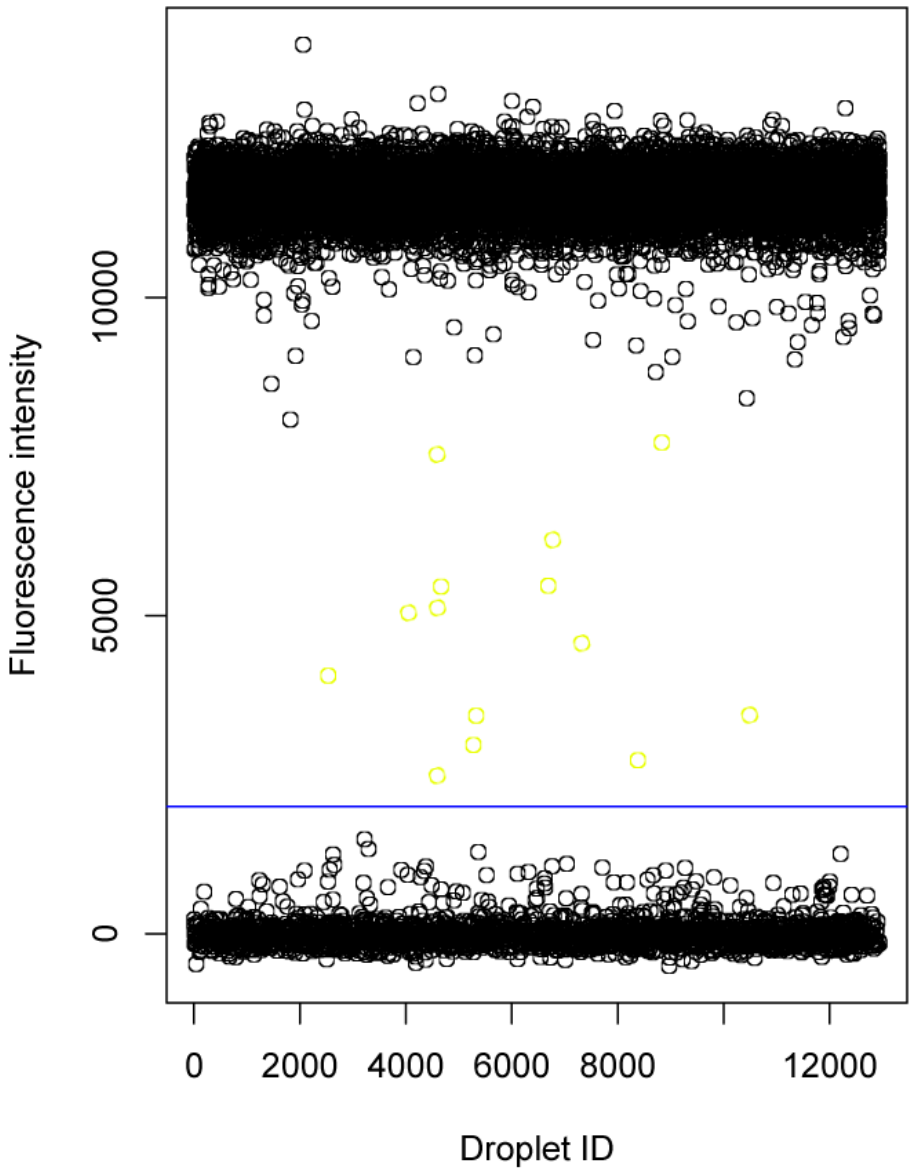
The concentration (target molecues/volume) is then estimated as:

$$\hat{c} = \frac{\hat{\lambda}}{V_{droplet}}$$

with $V_{droplet}$ a known constant (e.g. 0.85 nl).

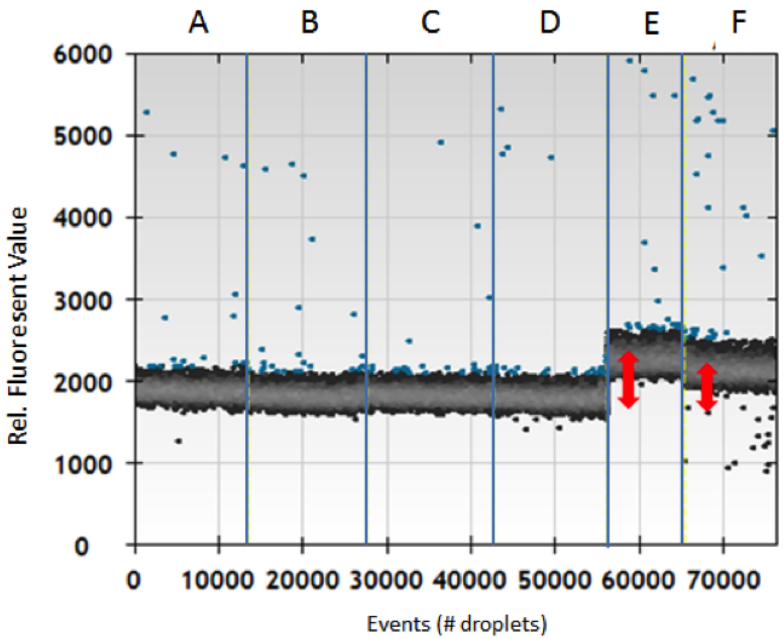
Threshold setting: Rain

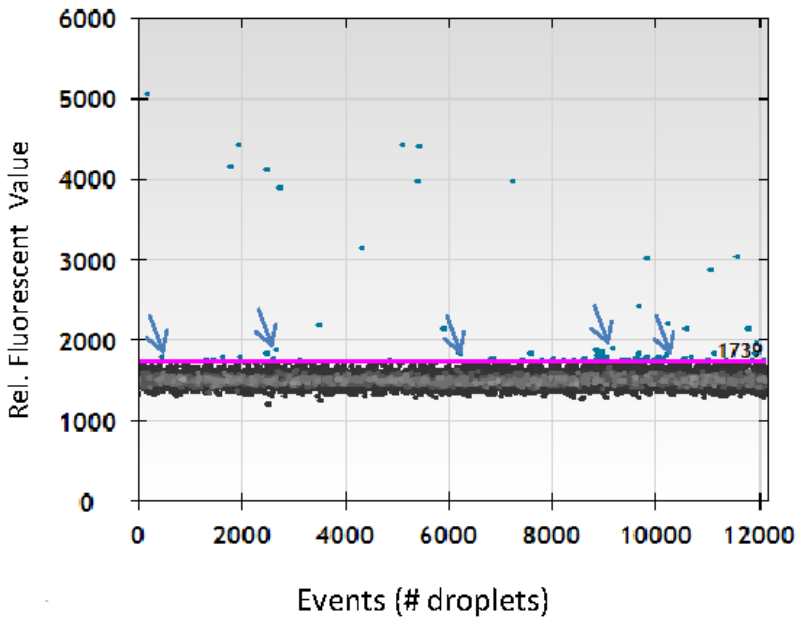
- positive/negative droplets are not always well separated: rain
- setting of threshold should control probabilities of false positives/negatives
- bad choice of threshold may cause biased copy number estimates
- No-template control (NTC, negative control) is typically used for setting threshold



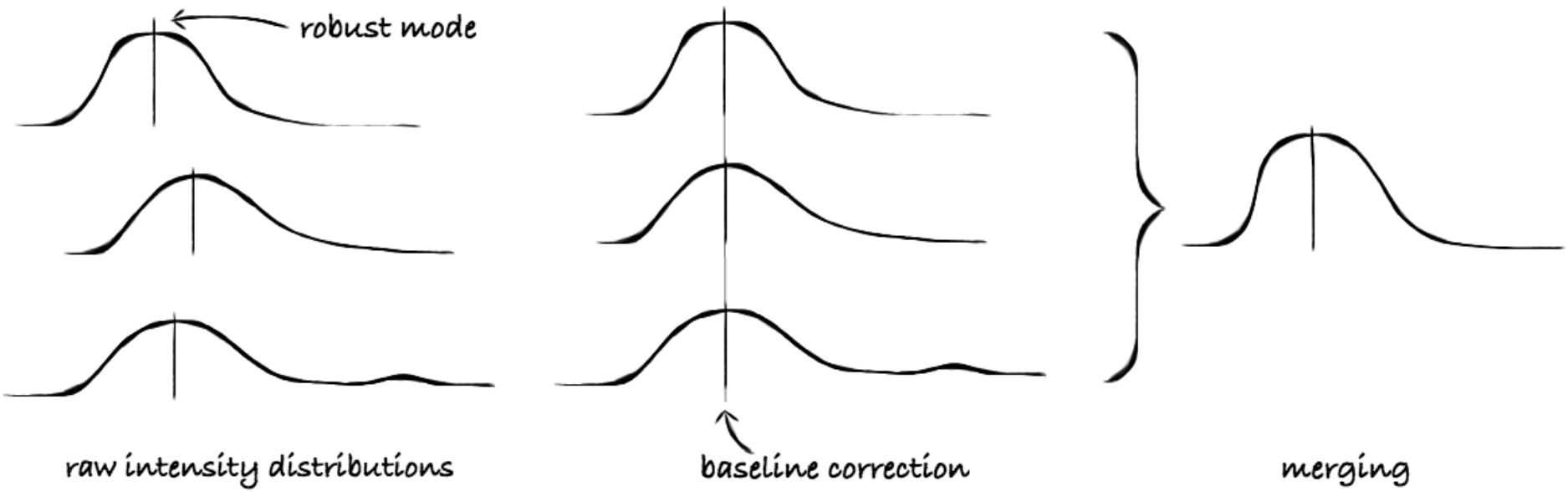
Threshold setting: Sample heterogeneity

- Baseline shifts between replicates
- NTC replicates cannot be pooled; normalisation is needed



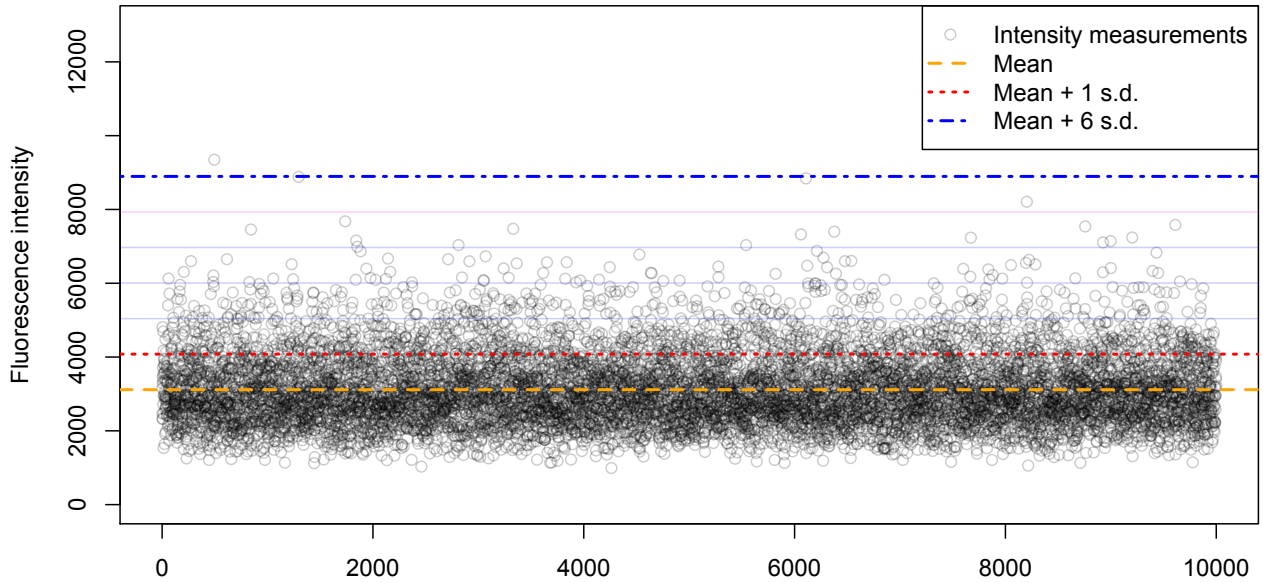


Threshold setting: Baseline correction



Threshold setting: Normality Assumption

Thresholds defined as mean $\pm$ $X\times$ s.d. rely on normality

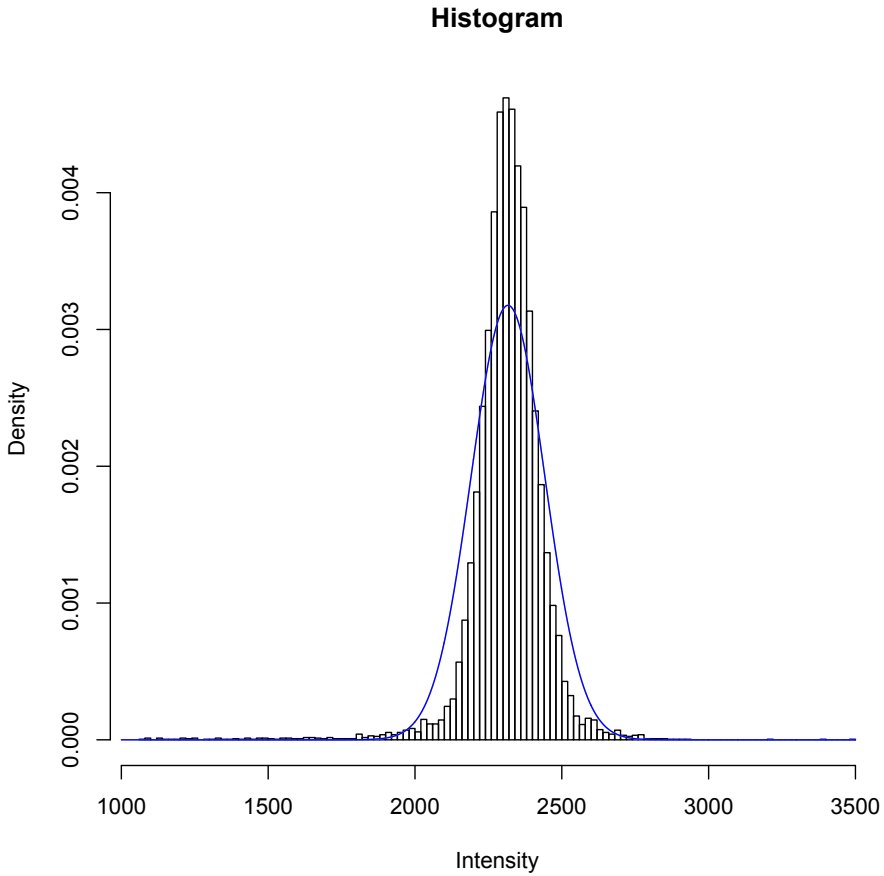


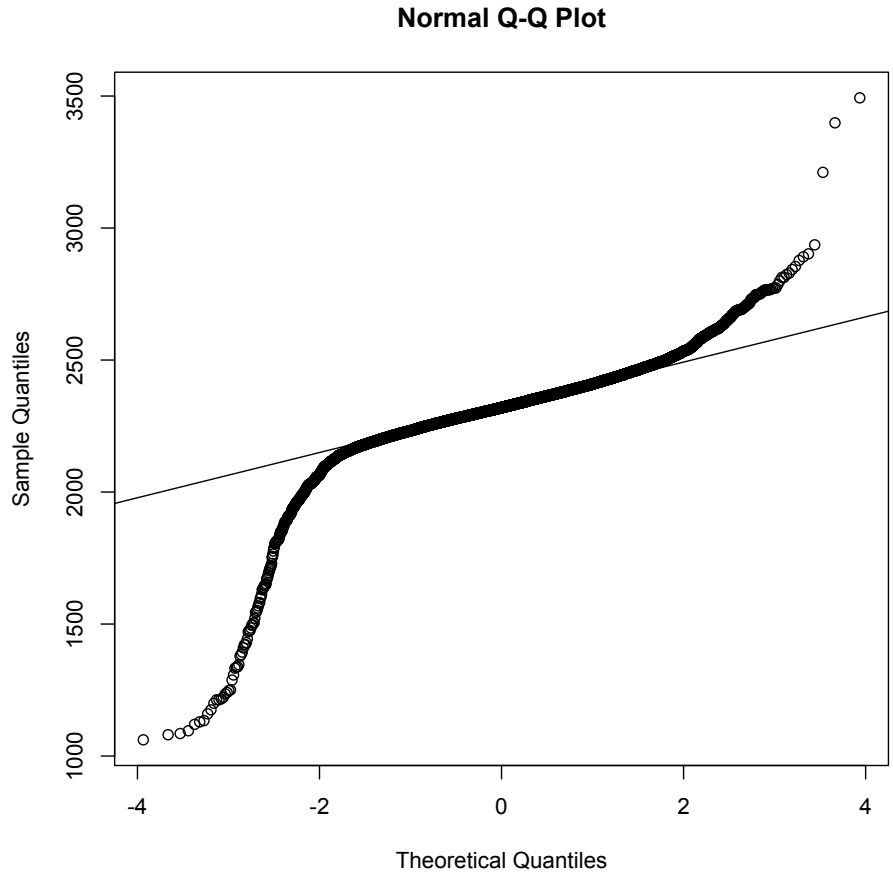
If normal assumption holds: false positive rate $\approx 10^{-9}$.

If normal assumption does not hold: false positive rate not controlled

Threshold setting: Normality Assumption

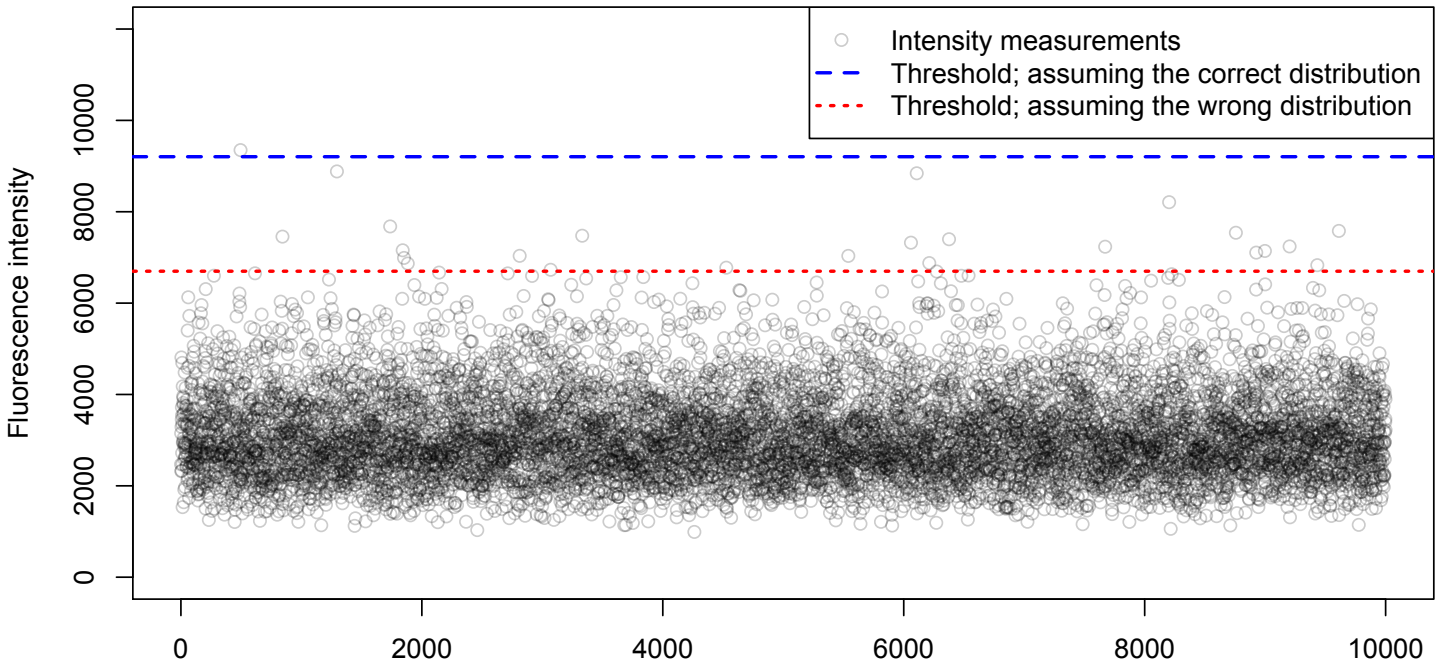
Thresholds defined as mean $\pm$ $X\times$ s.d. rely on normality





Intensities are not normally distributed!

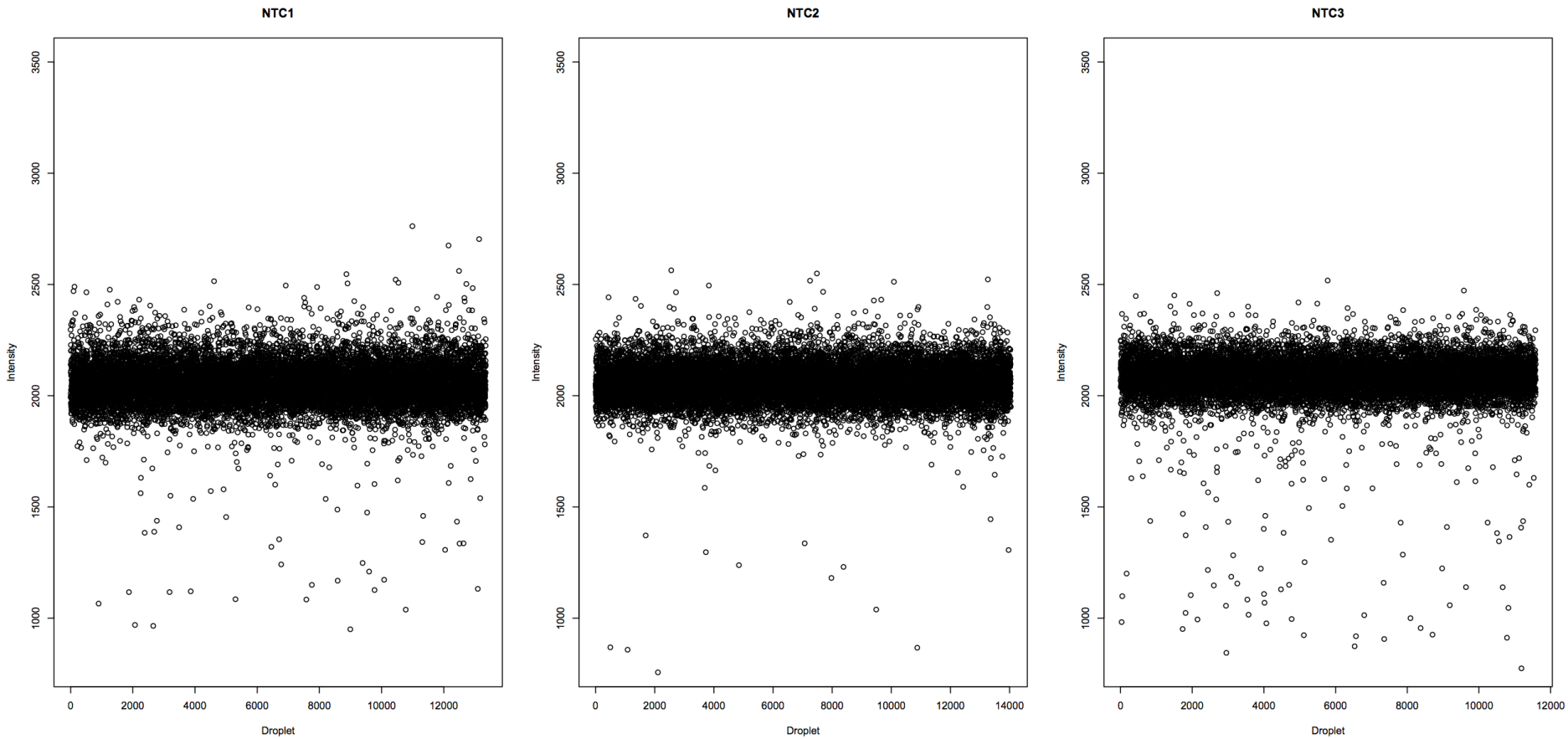
Threshold setting: Normality Assumption



$\longrightarrow$ threshold sensitive to distributional assumption

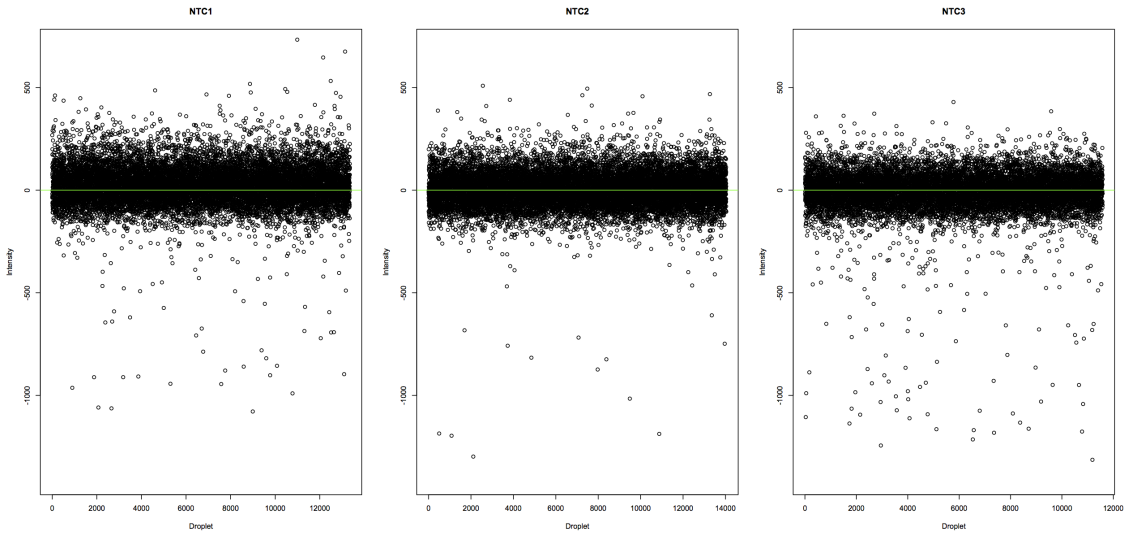
Threshold setting: Extreme Value Procedure

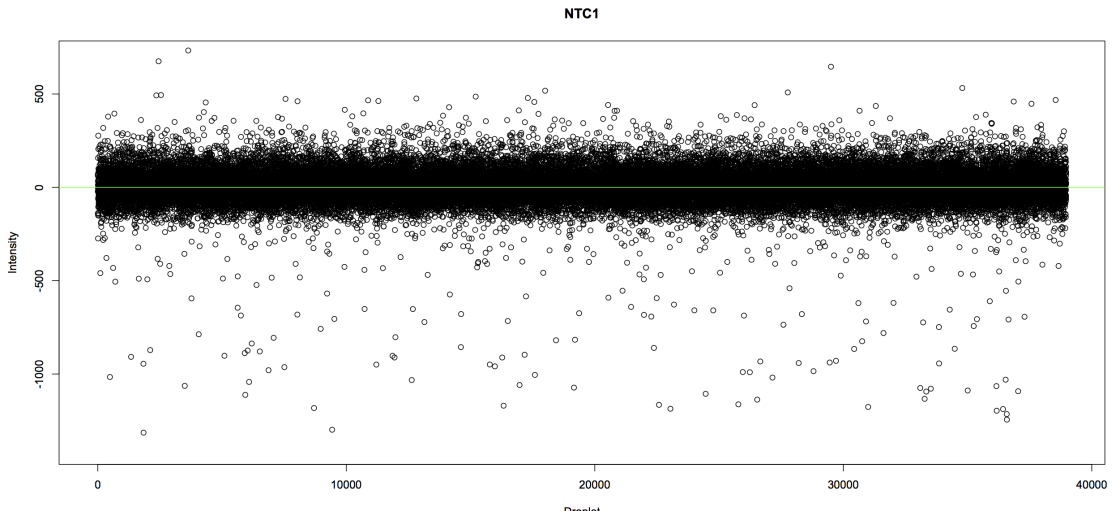
Step 0: Raw fluorescence measurements of no template controls (NTC)



Threshold setting: Extreme Value Procedure

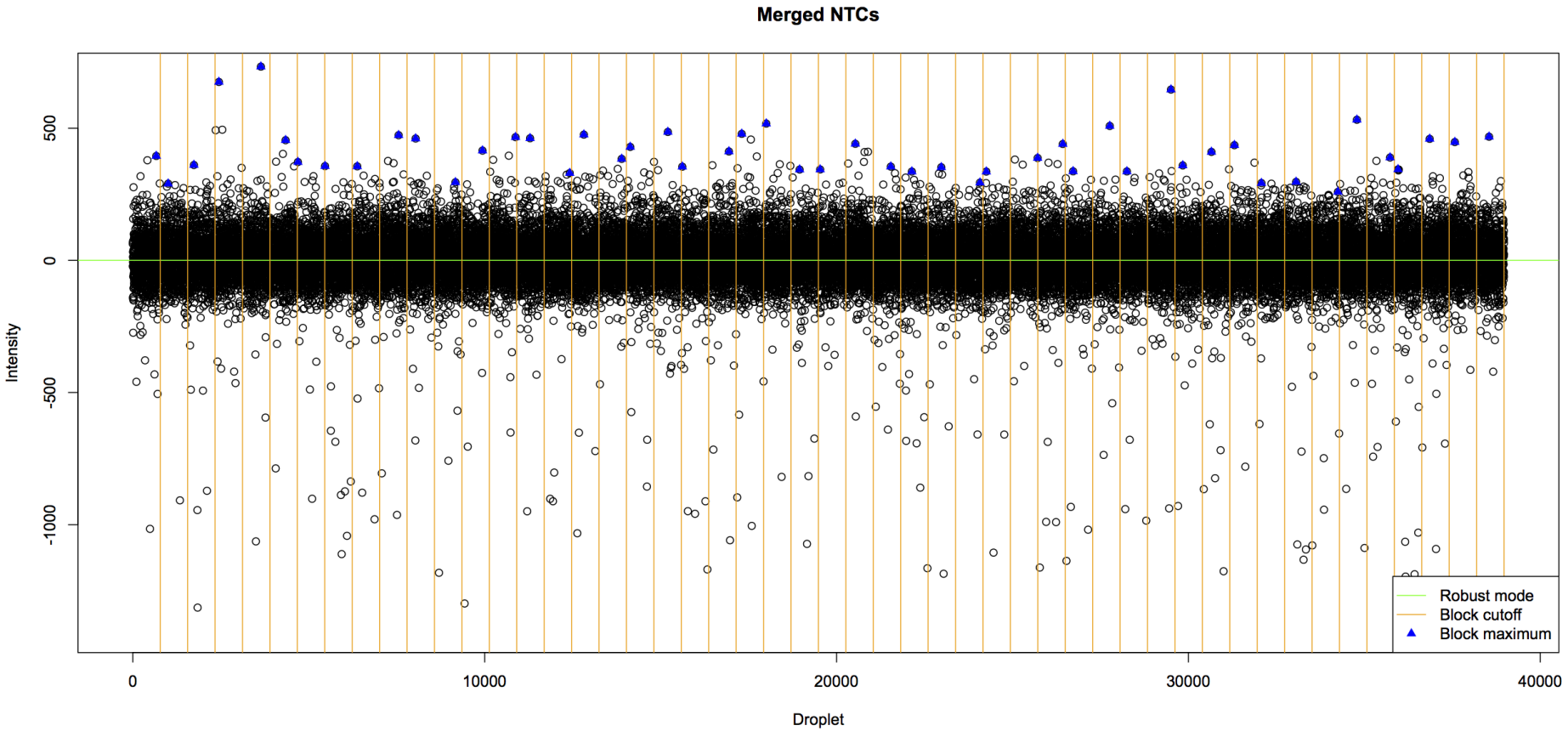
Step 1: Baseline correction and merging





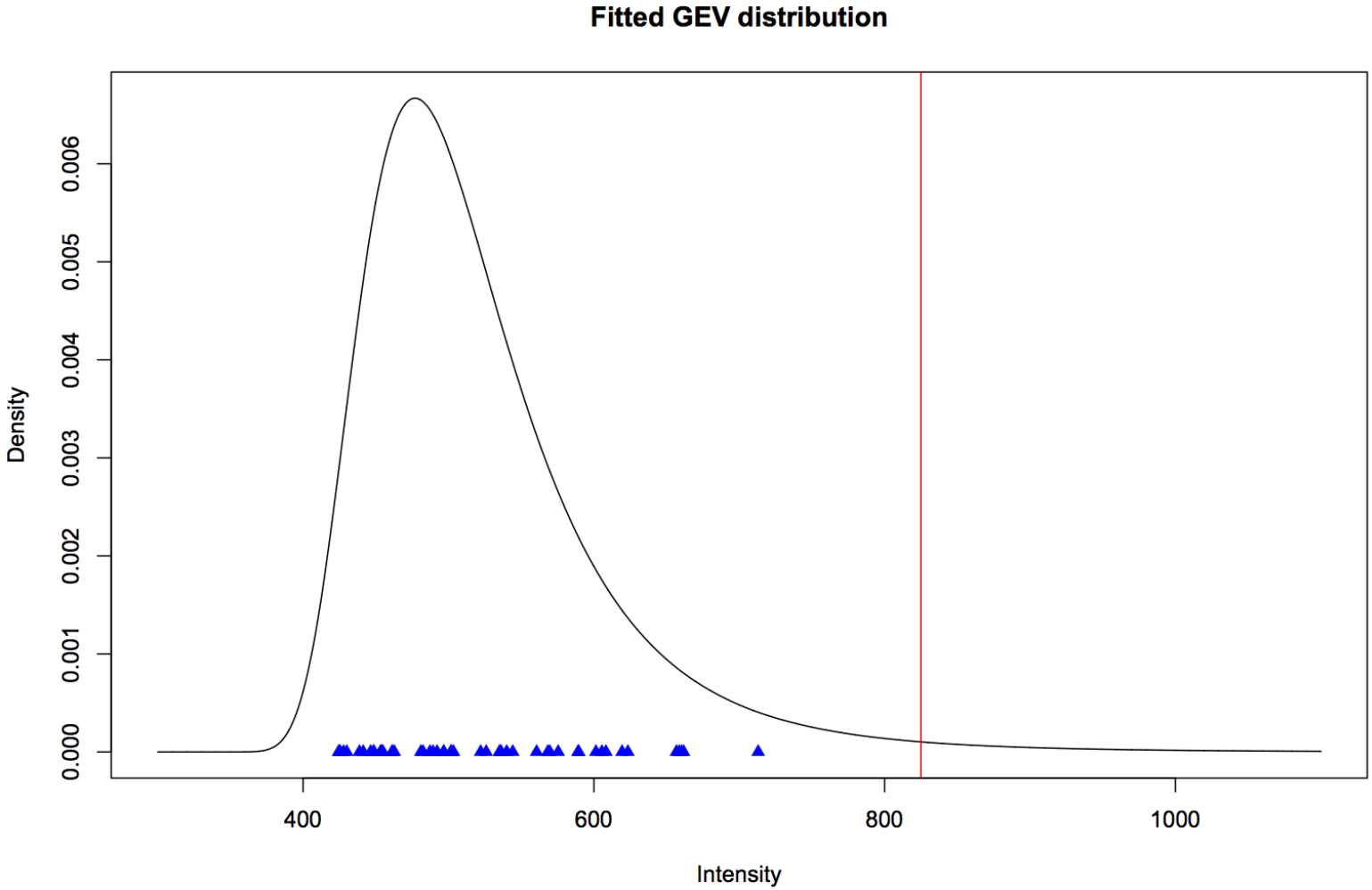
Threshold setting: Extreme Value Procedure

Step 2: Create blocks (randomly) and determine block maxima



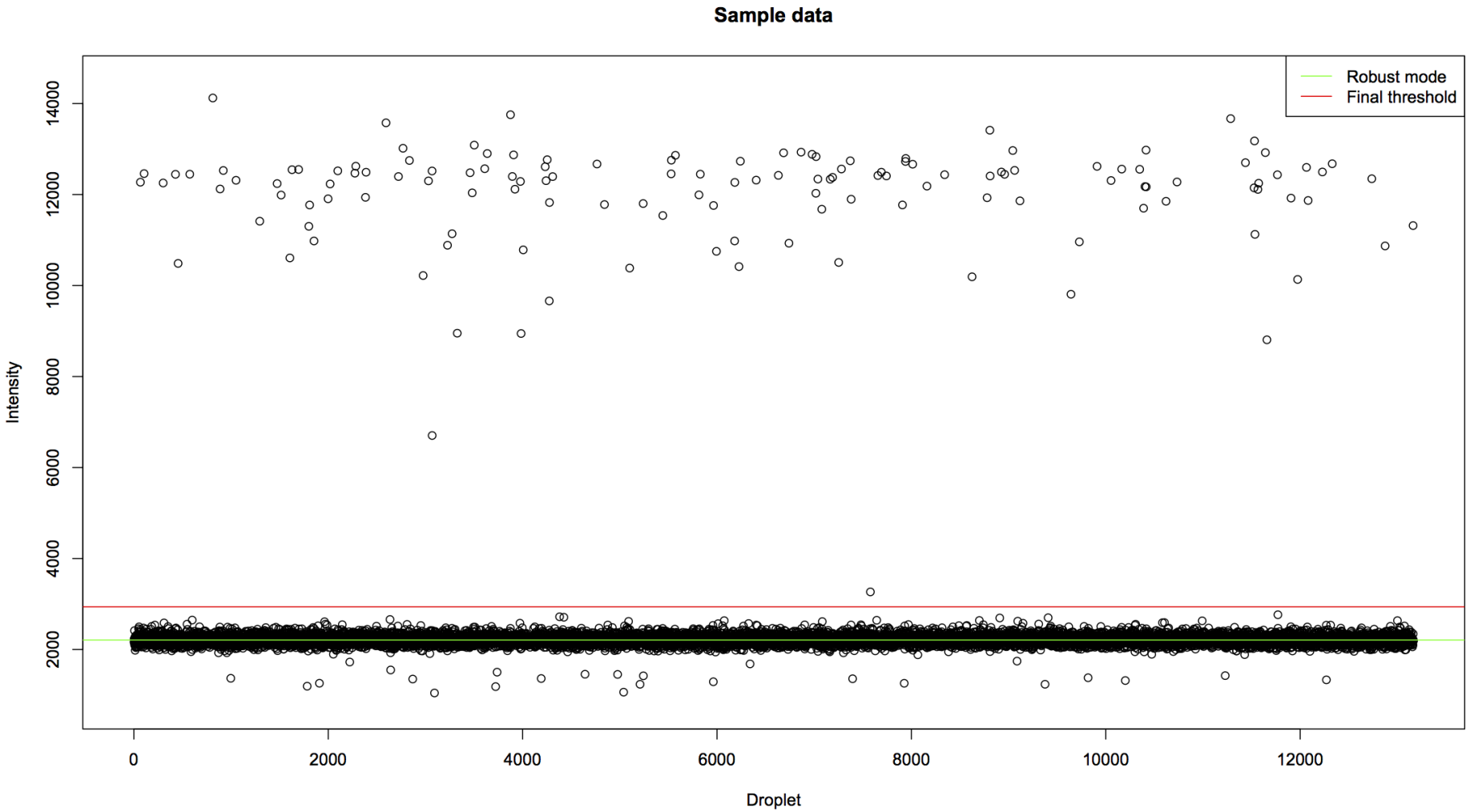
Threshold setting: Extreme Value Procedure

Step 3: Fit the GEV distribution to the block maxima and determine threshold

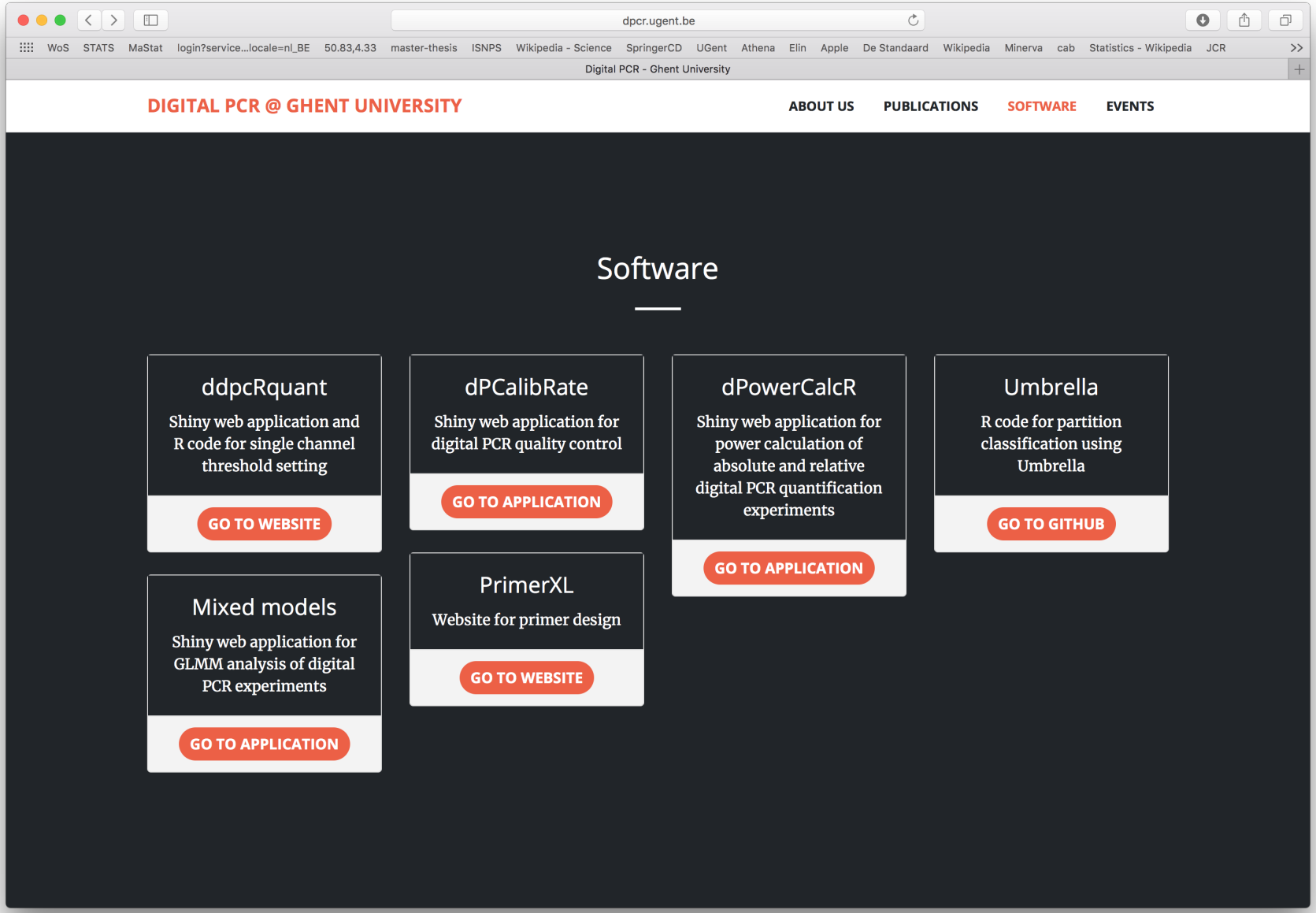


Threshold setting: Extreme Value Procedure

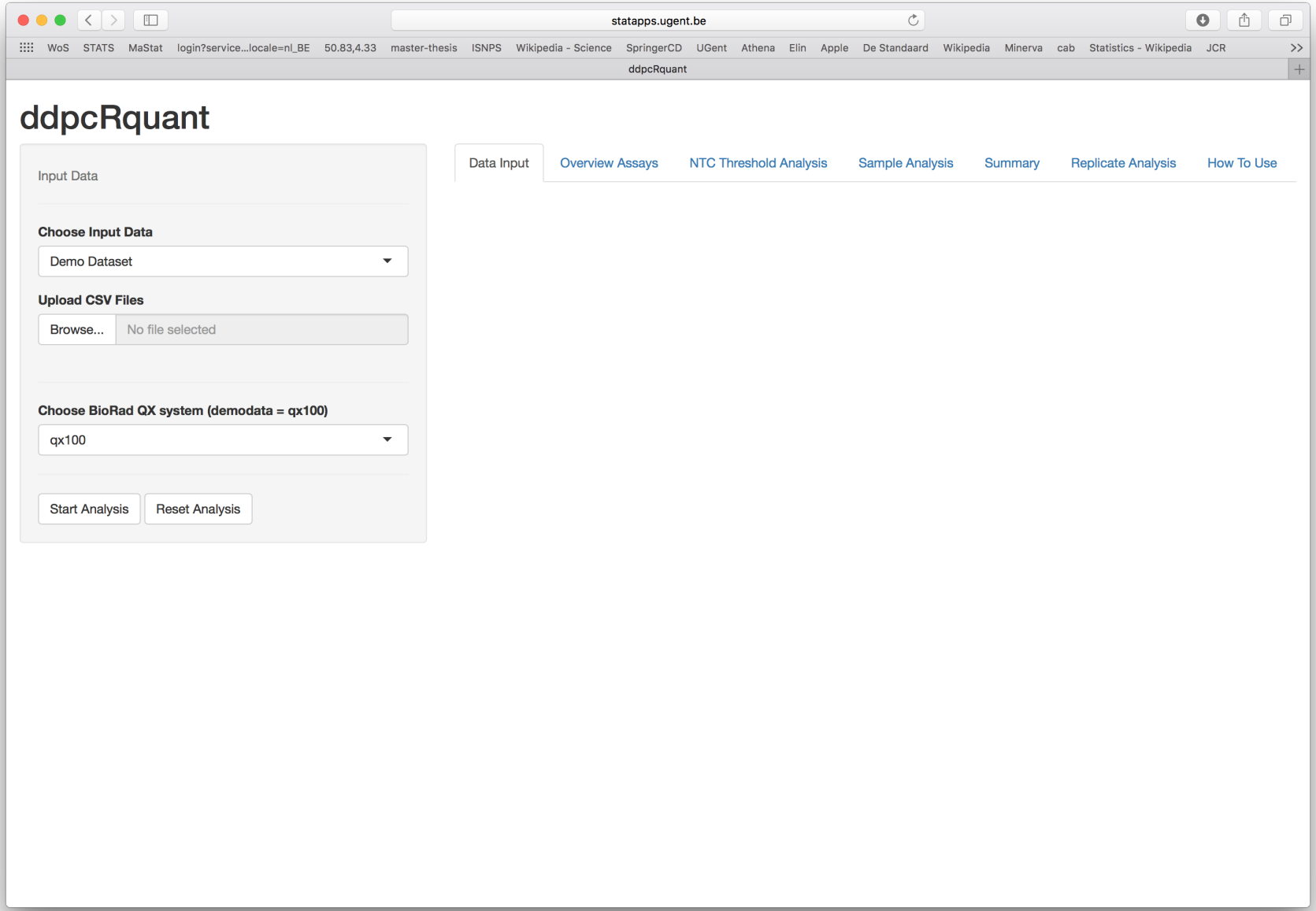
Step 4: Process samples



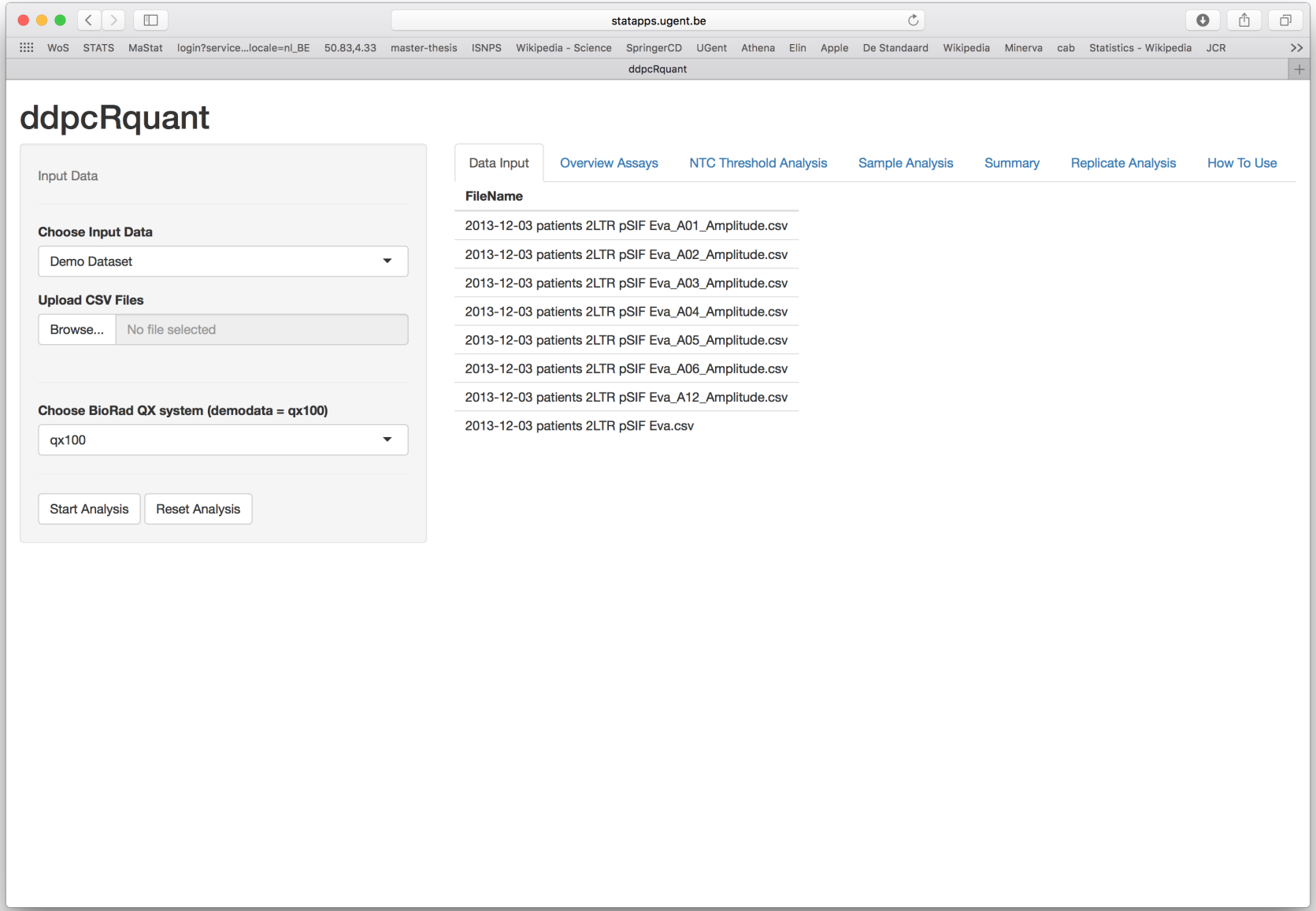
Threshold setting: Shiny App



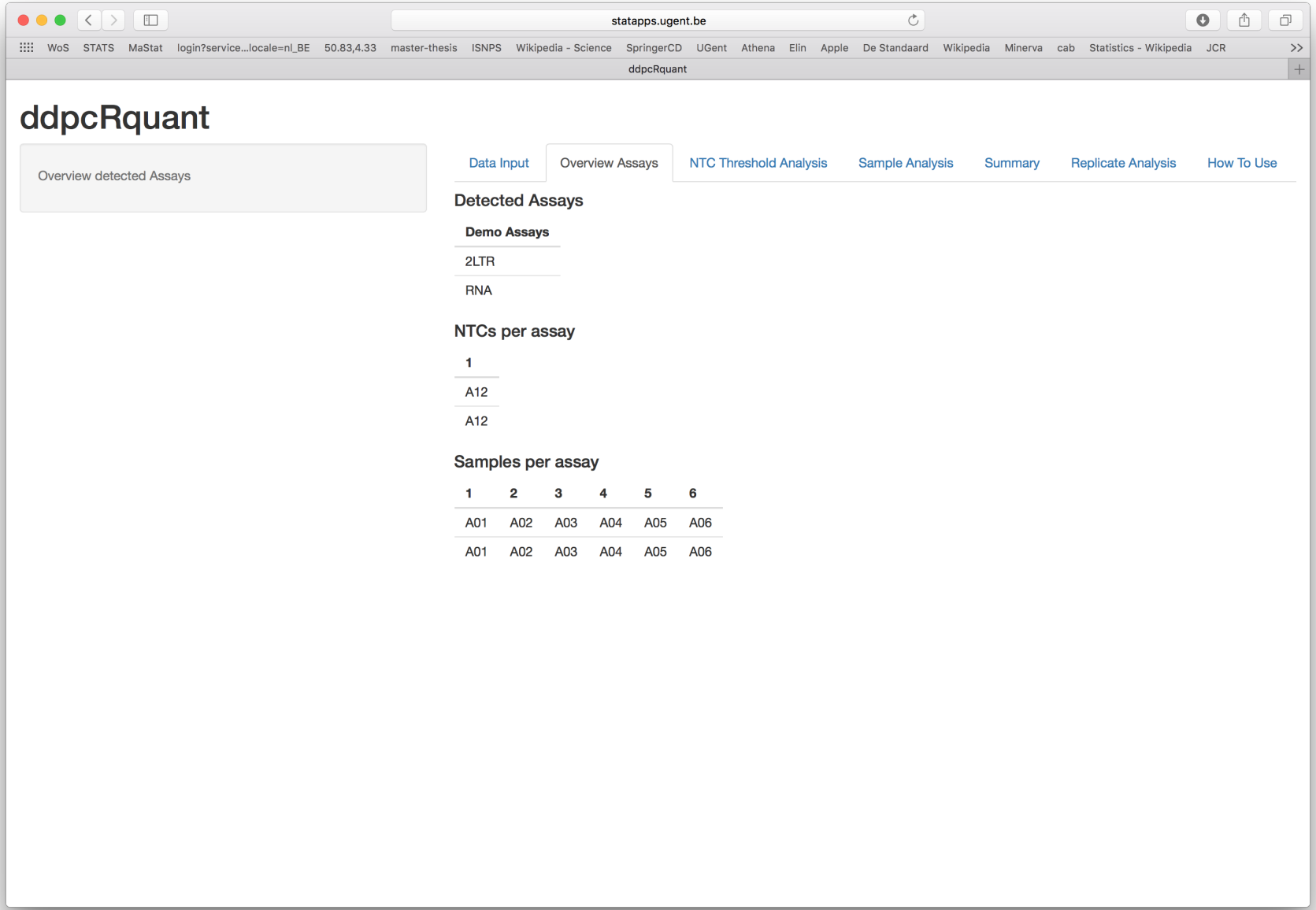
Threshold setting: Shiny App



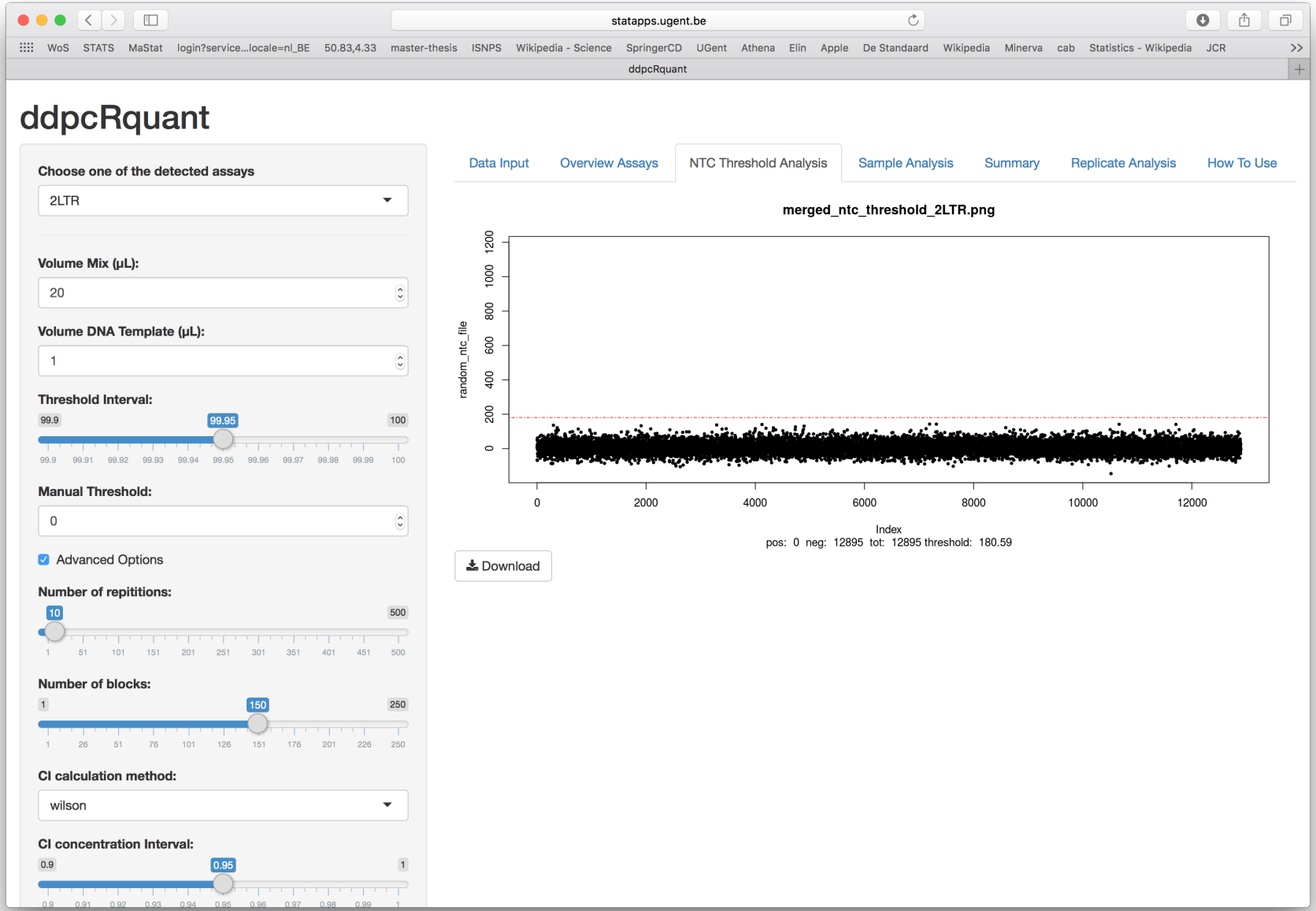
Threshold setting: Shiny App



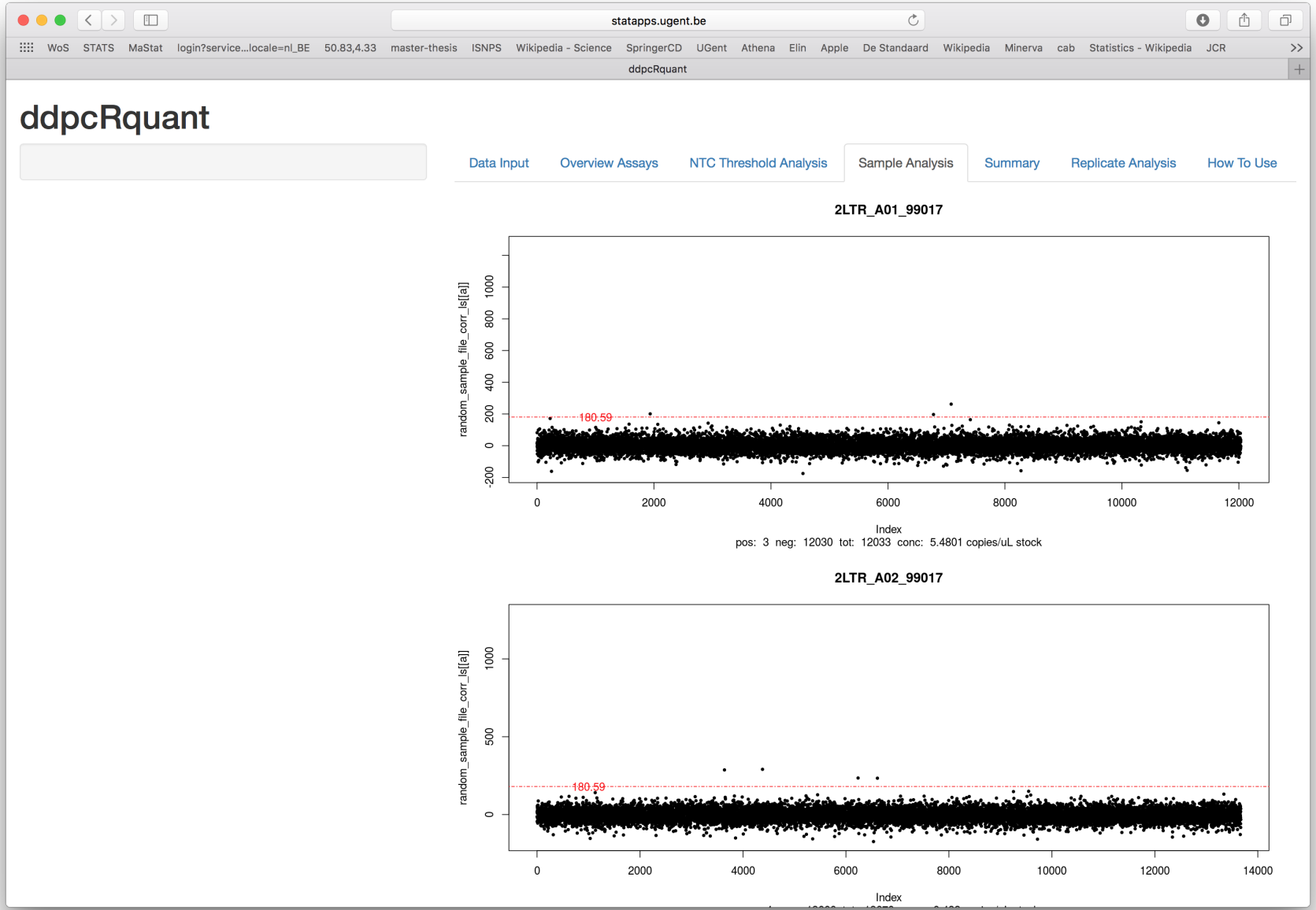
Threshold setting: Shiny App



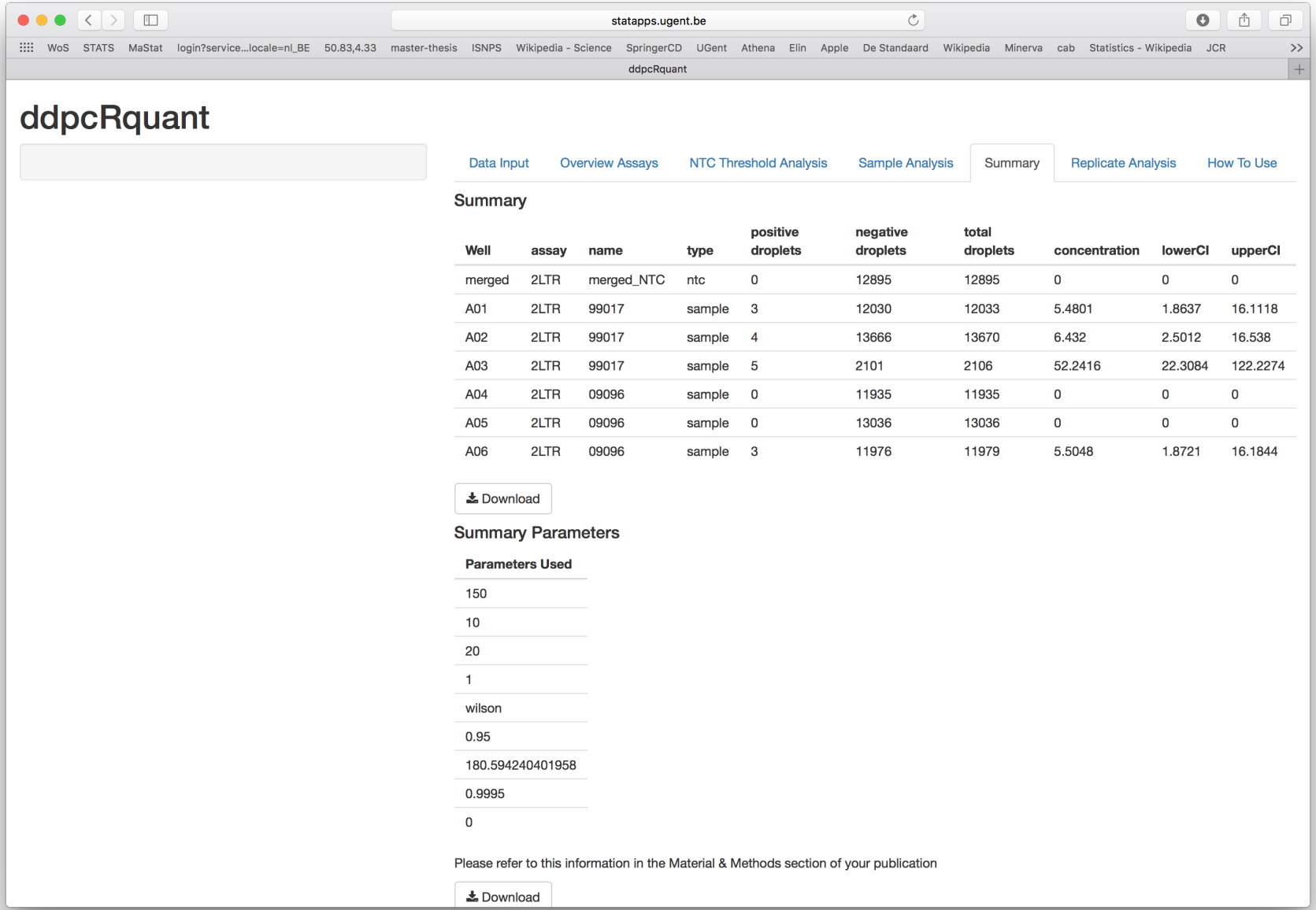
Threshold setting: Shiny App



Threshold setting: Shiny App



Threshold setting: Shiny App



Quality Control

Anal Bioanal Chem
DOI 10.1007/s00216-017-0538-9

PAPER IN FOREFRONT

Quality control of digital PCR assays and platforms

Matthijs Vynck[1] · Jo Vandesompele[2,3,4] · Olivier Thas[1,5,6]

Quality Control

For qPCR the MIQE guidelines say that

- calibration curve based on dilution series is needed
- R^2 for the linear fit should be reported (QC)

For dPCR no calibration curve is needed for absolute quantification, but it can be useful for QC:

- linear dynamic range
- accuracy
- precision

Quality Control

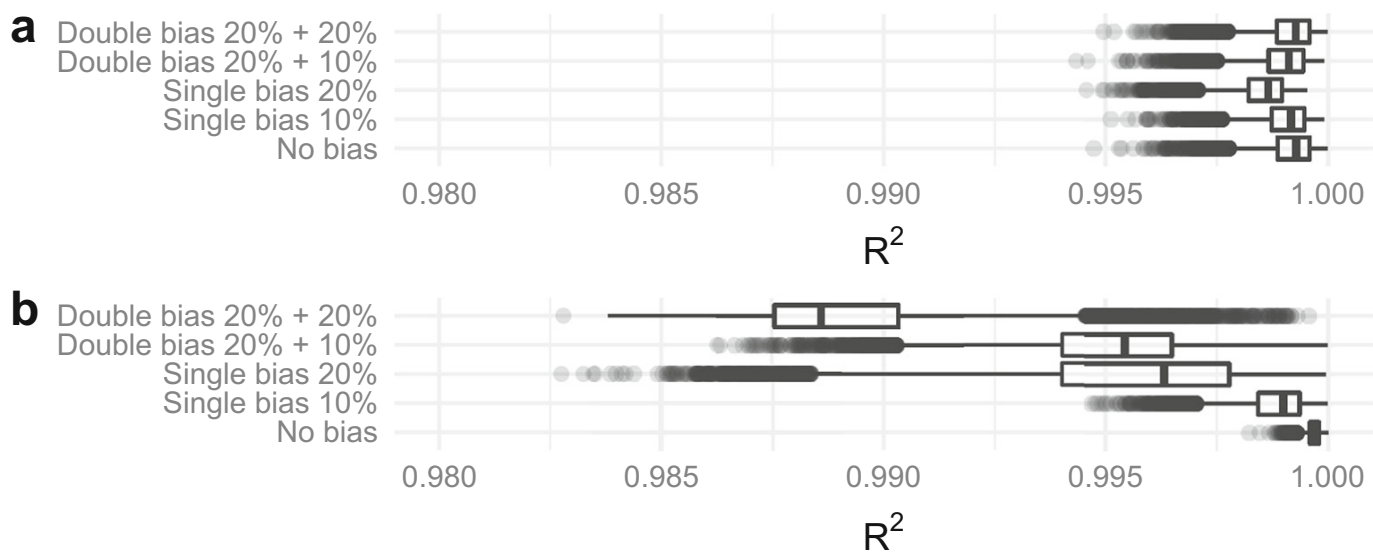
Often a large R^2 is considered as proof of

- linear dynamic range spanning the concentrations used for constructie the calibration curve
- good accuracy (i.e. estimated concentration = true concentration, i.e. slope of calibration curve is 1)
- good precision (small uncertainty on estimated concentration)

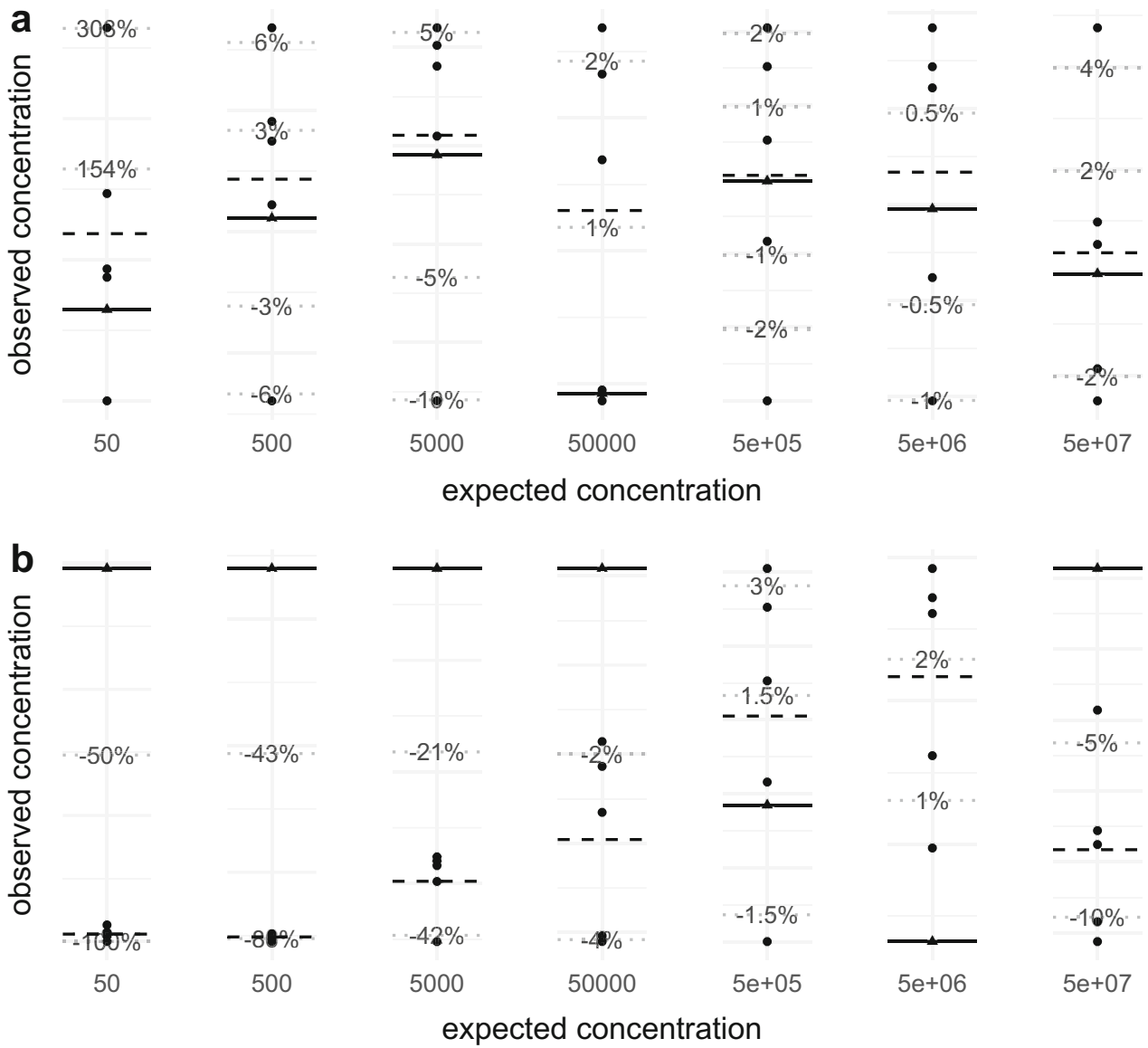
These considerations hold for qPCR and dPCR.

Quality Control: R^2

Fig. 1 R^2 values as obtained from data generated under different scenarios for (**a**) ordinary least squares, (**b**) robust weighted least squares. Single bias scenarios introduce a 10 or 20% downward bias at the highest concentration level. Double bias scenarios introduce a 20% downward bias at the highest concentration level and a 10 or 20% downward bias at the second highest concentration level



Quality Control: New plot



Quality Control: Linearity

5×10^7	5×10^6	5×10^5	5×10^4	5×10^3	5×10^2	5×10^1
-+-++	-+++-	++--+	-++++	+++-+	++-++	++++-
-----	+++++	+++-+	-----	-----	-----	-----

Quality Control: Linearity

Linearity can also be assessed by fitting a quadratic model and testing whether the quadratic term adds significant improvement ($H_0 : \beta_2 = 0$) to the model fit (t-test).

$$\text{observed dPCR conc} = \beta_0 + \beta_1 \times \text{expected conc} + \beta_2 \times (\text{expected conc})^2$$

Test based on RWLS performs best.

Quality Control: Accuracy

Deviation from expected model is plotted:

$$\text{observed dPCR conc} = 0 + 1 \times \text{expected conc.}$$

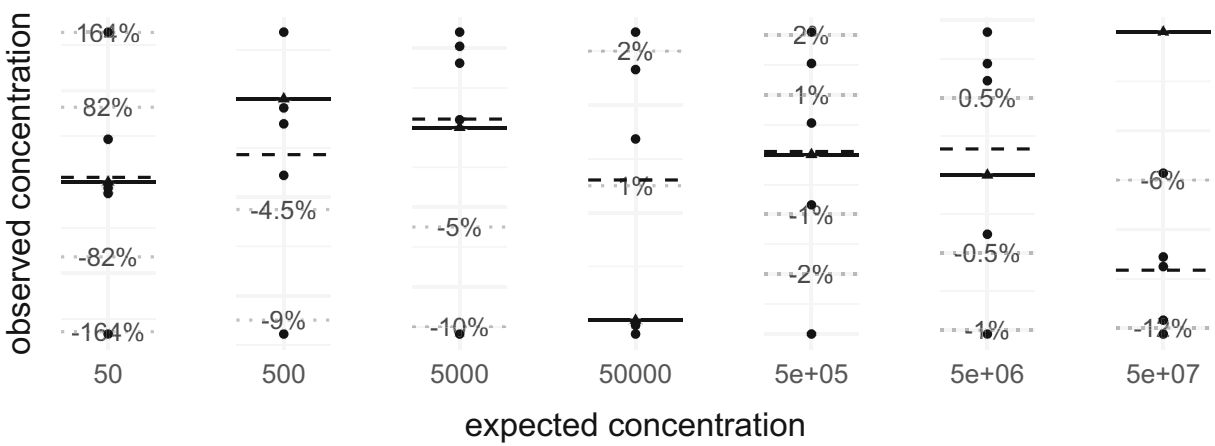
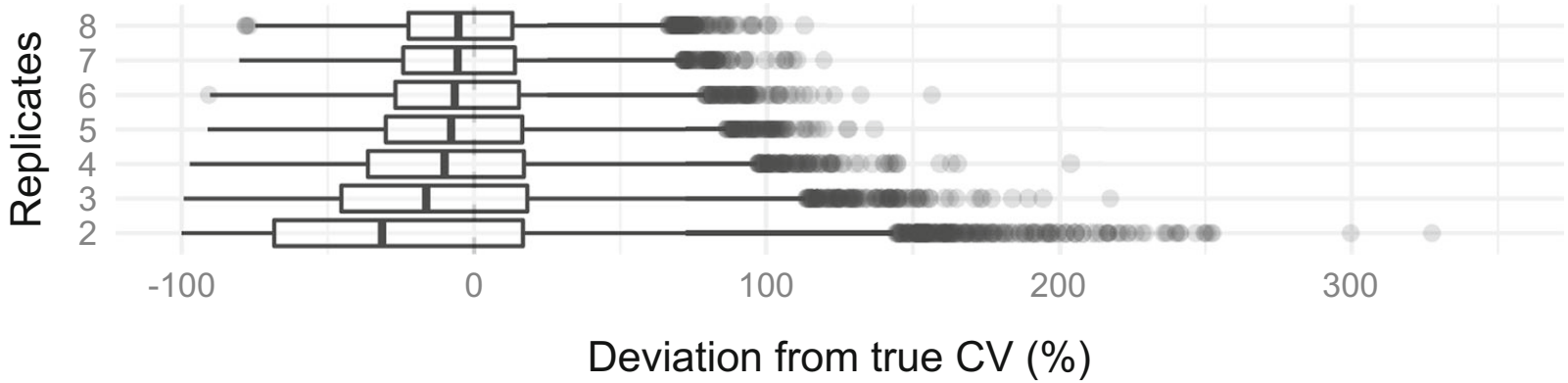


Fig. 4 Accuracy assessment plot: A bias of 10% is introduced at the highest concentration. *Triangles* indicate the expected concentration for perfect accuracy, i.e., corresponding to a slope of 1 and an intercept at 0. *Dots* indicate the observed concentrations. *Percentages* indicate deviation from perfect accuracy

Quality Control: Precision

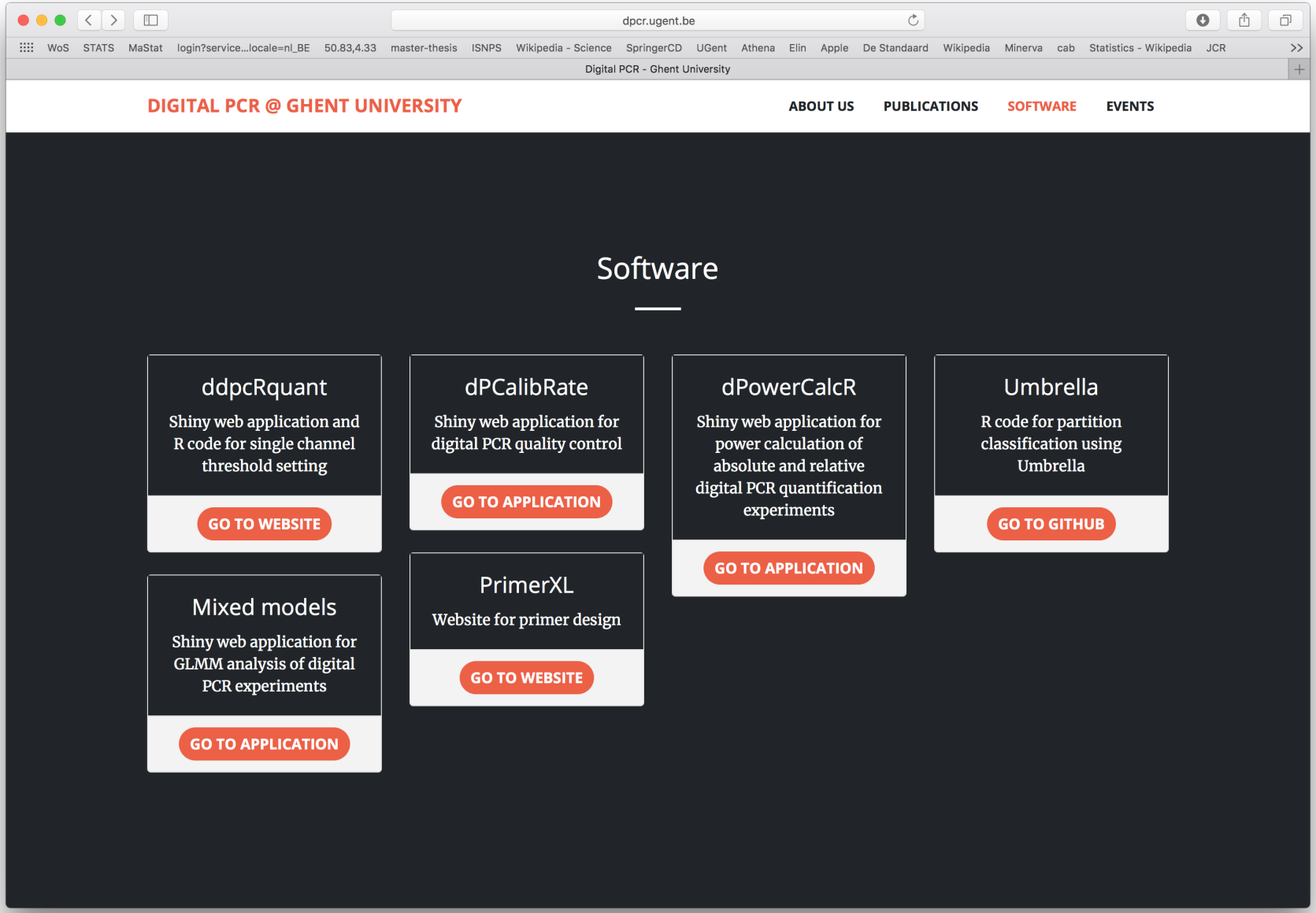
Precision can be measured by the Coefficient of Variation (CV)



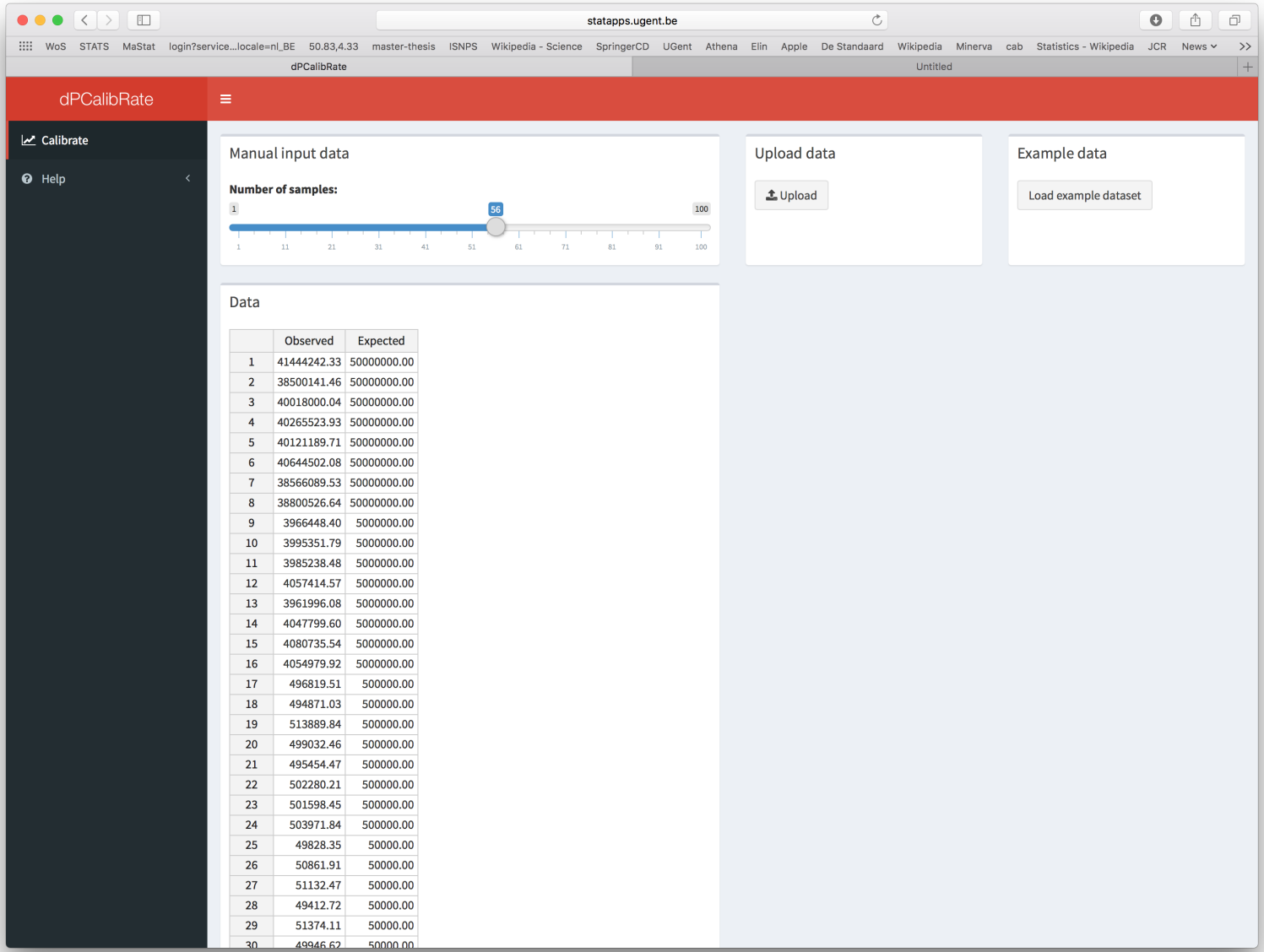
Quality Control: Recommendations

1. Robust weighted least squares (RWLS) is preferred over LS for fitting the calibration curve
2. Linearity: t-test in quadratic regression is preferred over the runs test, unless the number of replicates is very small
3. Number of replicates at each concentration should be at least 5. Otherwise the power of the tests is too small for detecting relevant deviations.
4. If linearity holds, R^2 may still be used to compare experiments in terms of precision
5. Graphical inspection is very informative.

Quality Control: Shiny app



Quality Control: Shiny app



	Observed	Expected
1	41444242.33	50000000.00
2	38500141.46	50000000.00
3	40018000.04	50000000.00
4	40265523.93	50000000.00
5	40121189.71	50000000.00
6	40644502.08	50000000.00
7	38566089.53	50000000.00
8	38800526.64	50000000.00
9	3966448.40	5000000.00
10	3995351.79	5000000.00
11	3985238.48	5000000.00
12	4057414.57	5000000.00
13	3961996.08	5000000.00
14	4047799.60	5000000.00
15	4080735.54	5000000.00
16	4054979.92	5000000.00
17	496819.51	500000.00
18	494871.03	500000.00
19	513889.84	500000.00
20	499032.46	500000.00
21	495454.47	500000.00
22	502280.21	500000.00
23	501598.45	500000.00
24	503971.84	500000.00
25	49828.35	50000.00
26	50861.91	50000.00
27	51132.47	50000.00
28	49412.72	50000.00
29	51374.11	50000.00
30	49946.62	50000.00

Analysis of complex experiments

Examples:

- Absolute quantification with technical replicates
- CNV
  - with/without technical replicates
  - singleplex / duplex / multiplex
  - one or more references

Biomolecular Detection and Quantification 9 (2016) 1–13

Contents lists available at ScienceDirect

Biomolecular Detection and Quantification

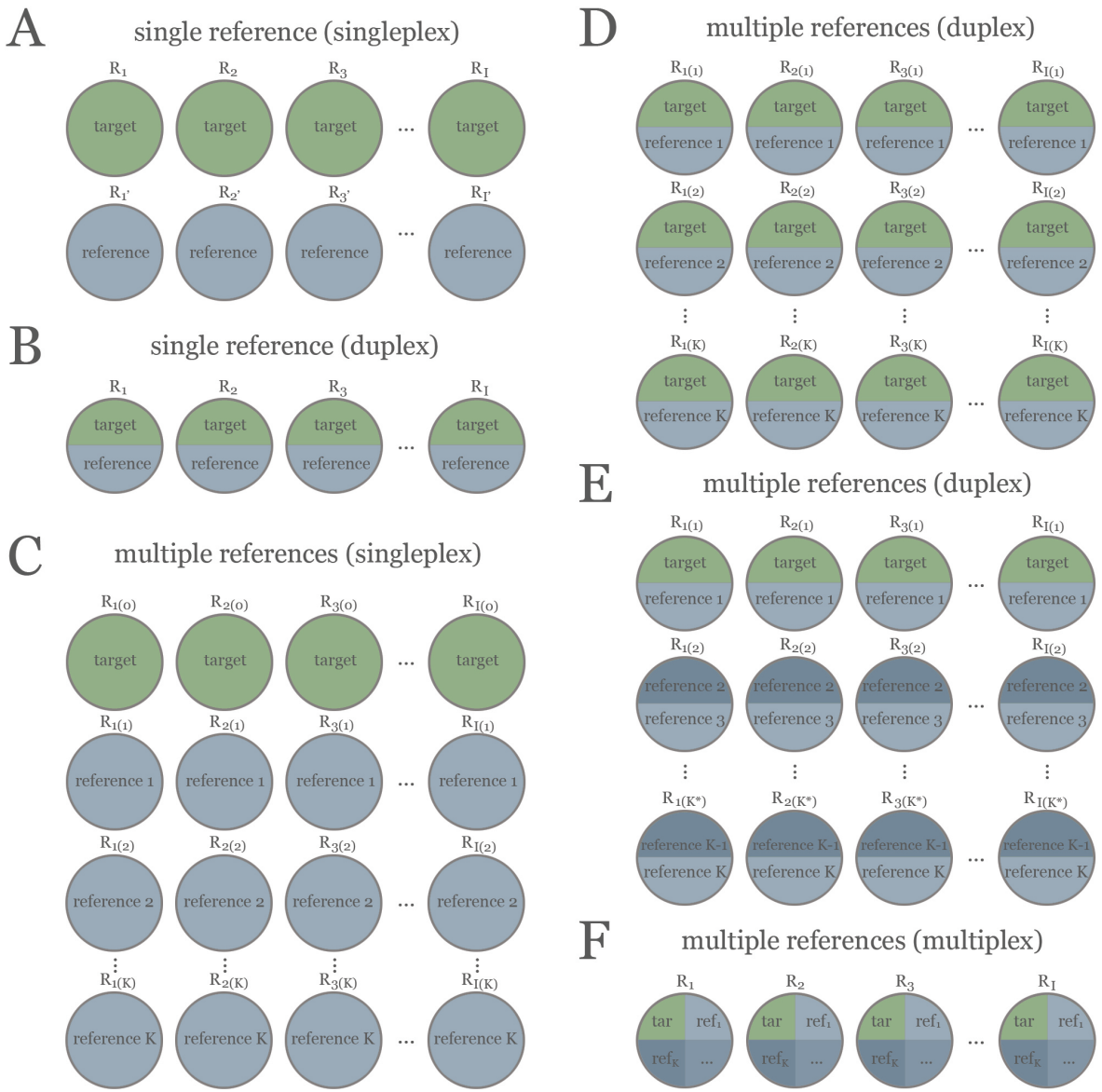
journal homepage: www.elsevier.com/locate/bdq

Research Paper

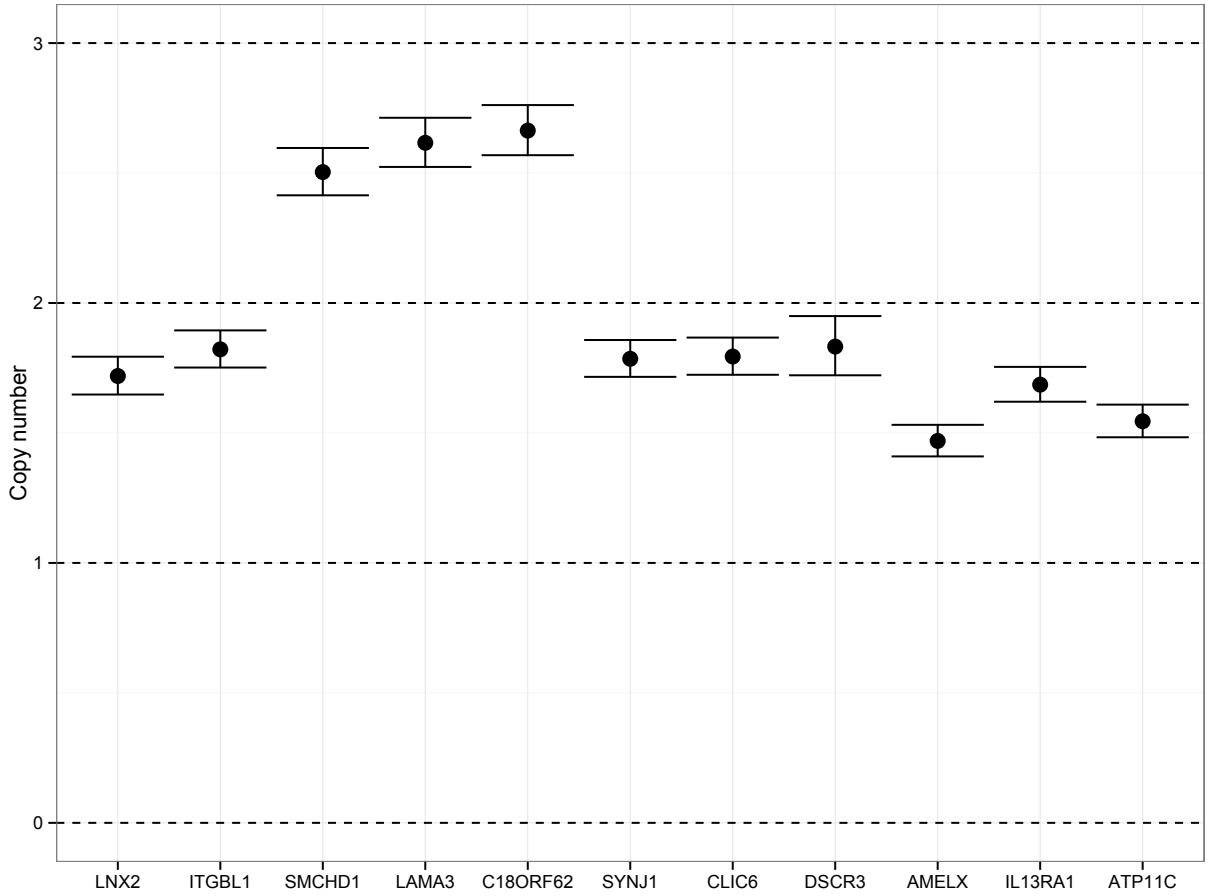
Flexible analysis of digital PCR experiments using generalized linear mixed models

Matthijs Vynck[a,*], Jo Vandesompele[b,c,d], Nele Nijs[d], Björn Menten[b,c], Ariane De Ganck[d], Olivier Thas[a,e]

Analysis of complex experiments: Designs CNV

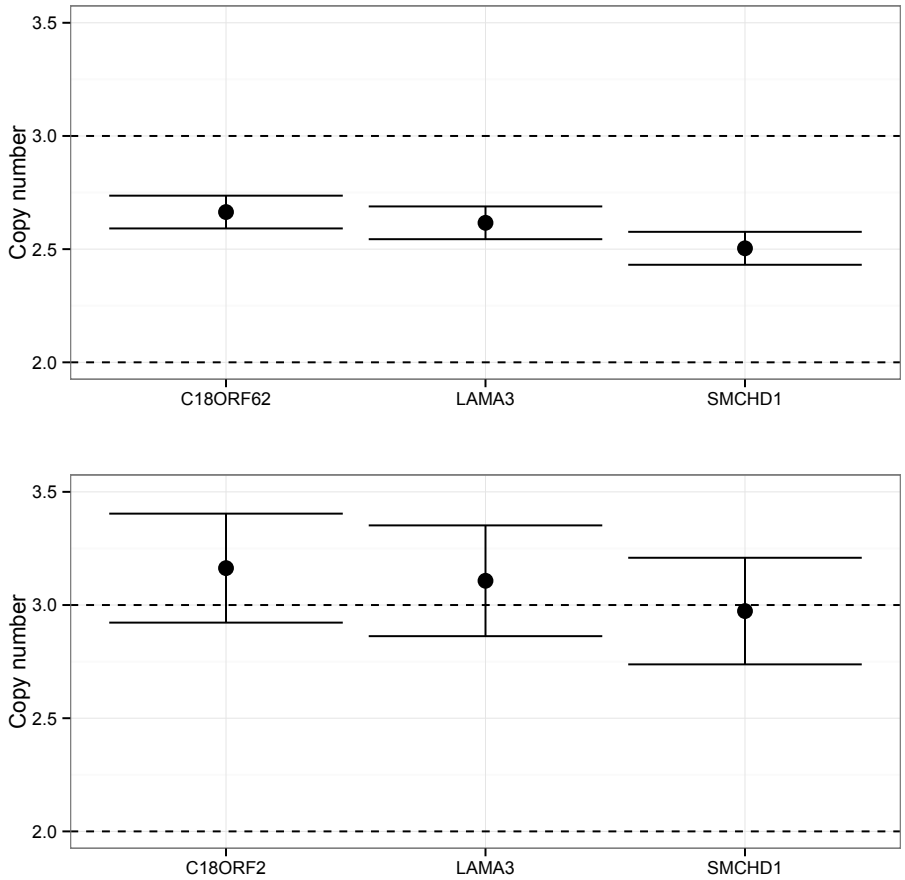


Analysis of complex experiments: CNV with single versus multiple ref. genes



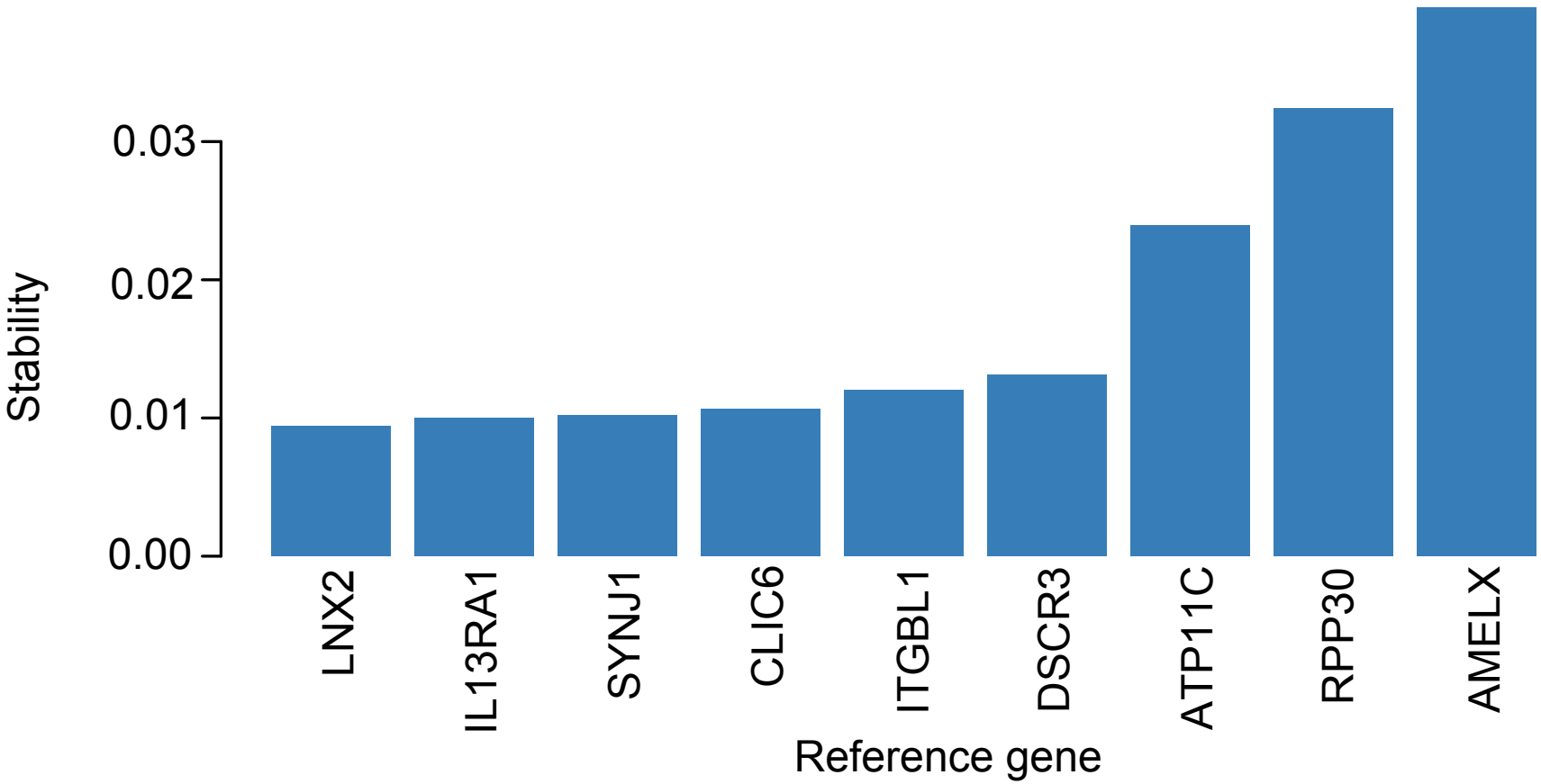
$\longrightarrow$ bias caused by single reference gene

Analysis of complex experiments: CNV with single versus multiple ref. genes

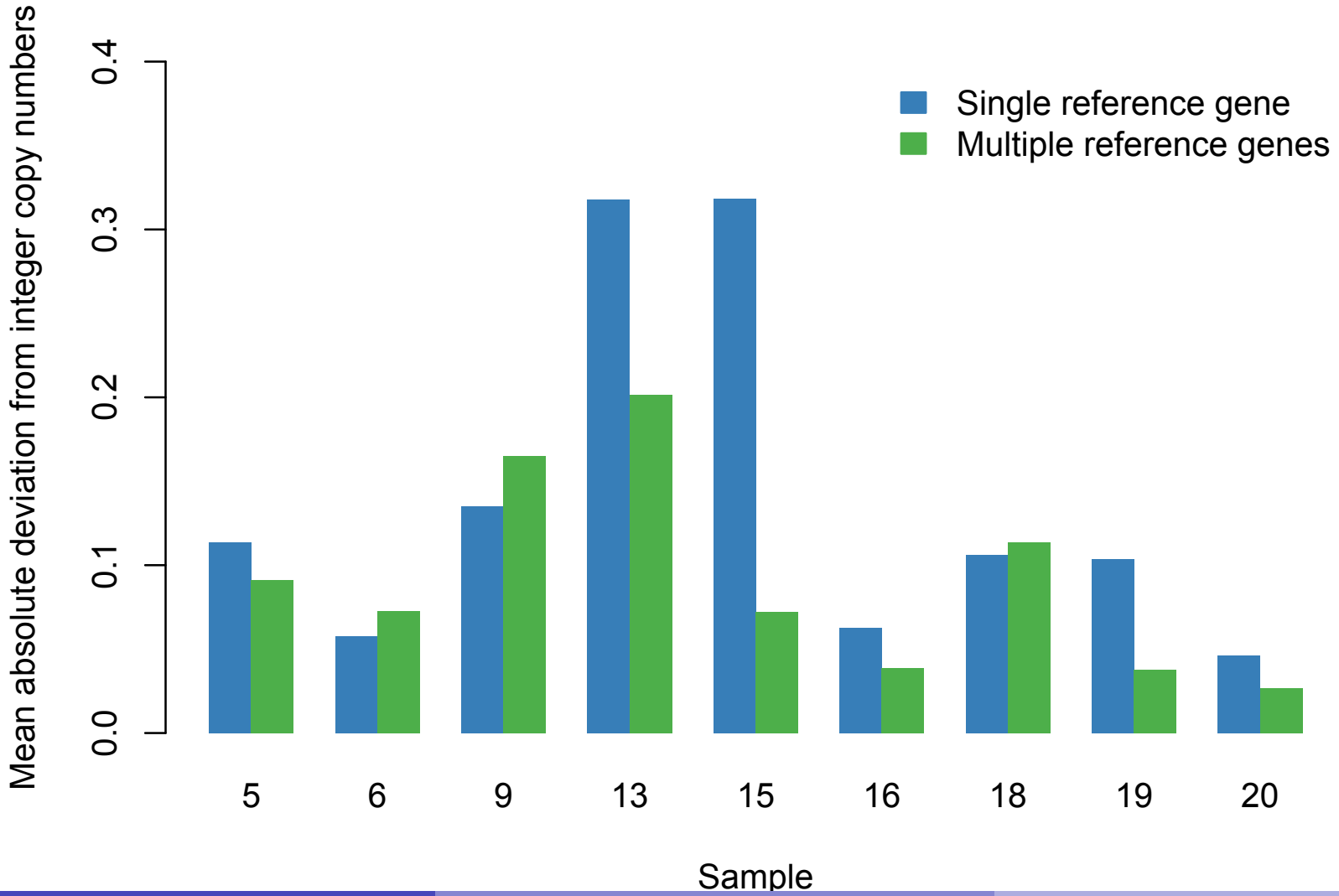


$\longrightarrow$ bias resolved by using multiple reference genes

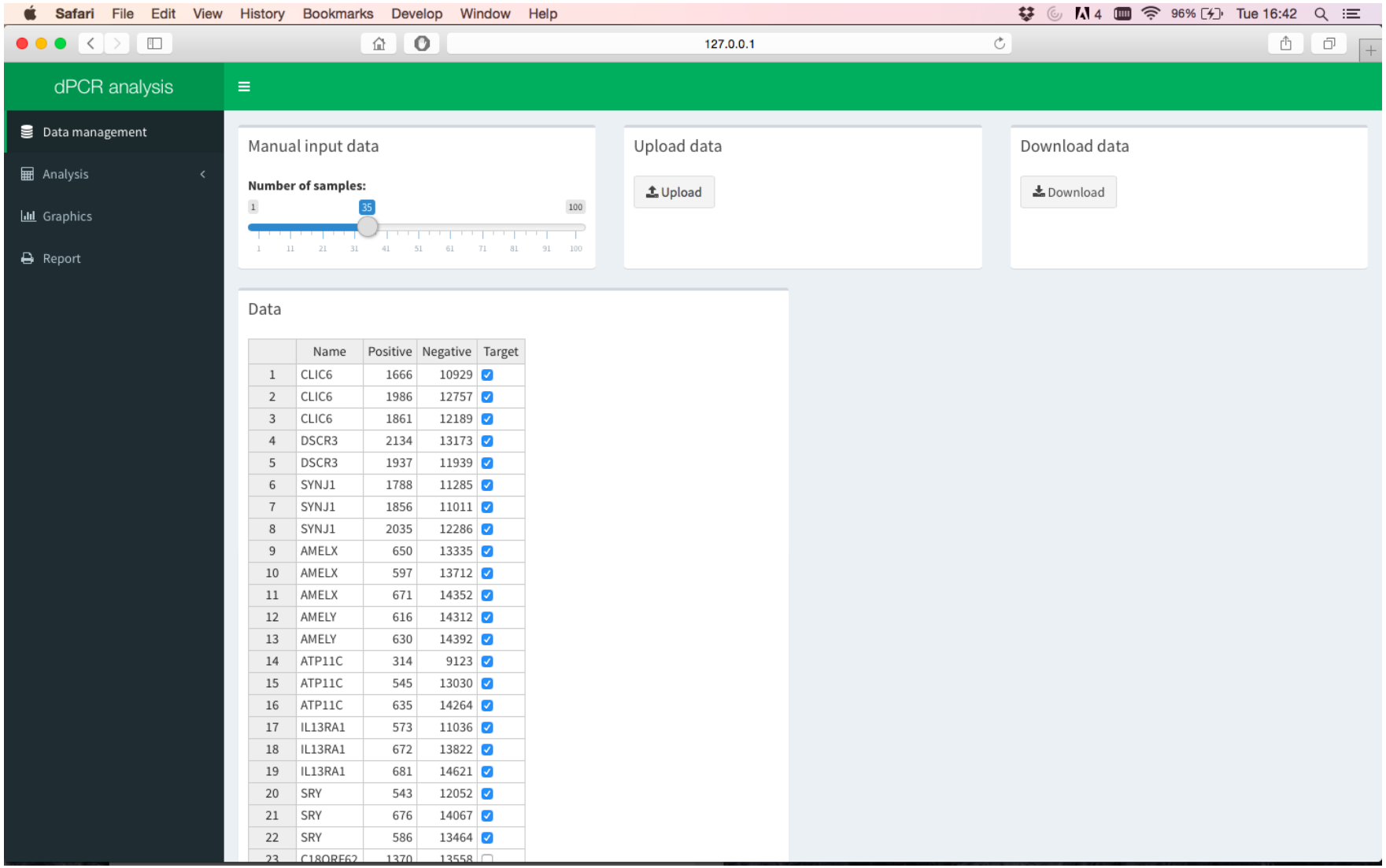
Analysis of complex experiments: Reference gene stability



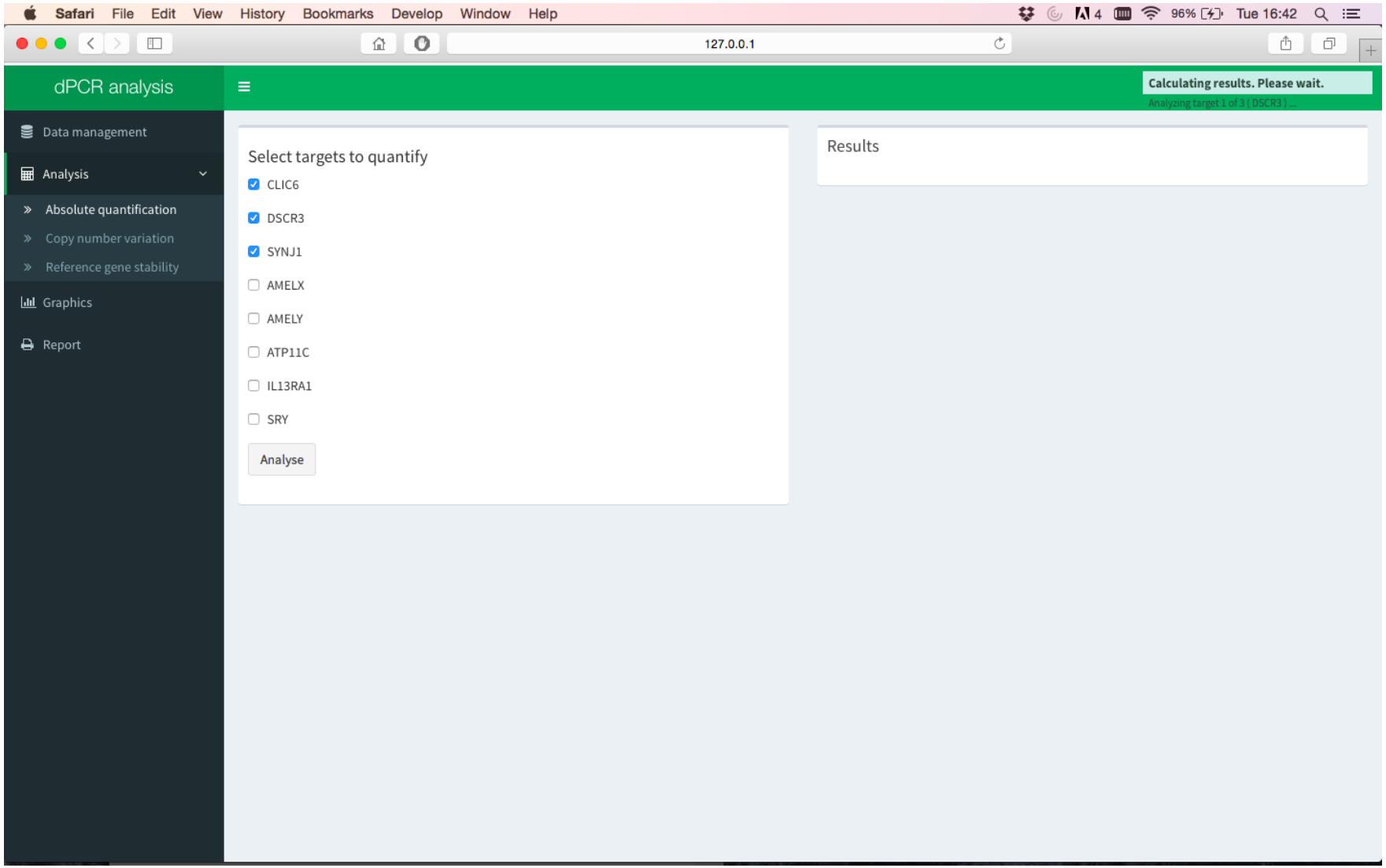
Analysis of complex experiments: CNV with single versus multiple ref. genes



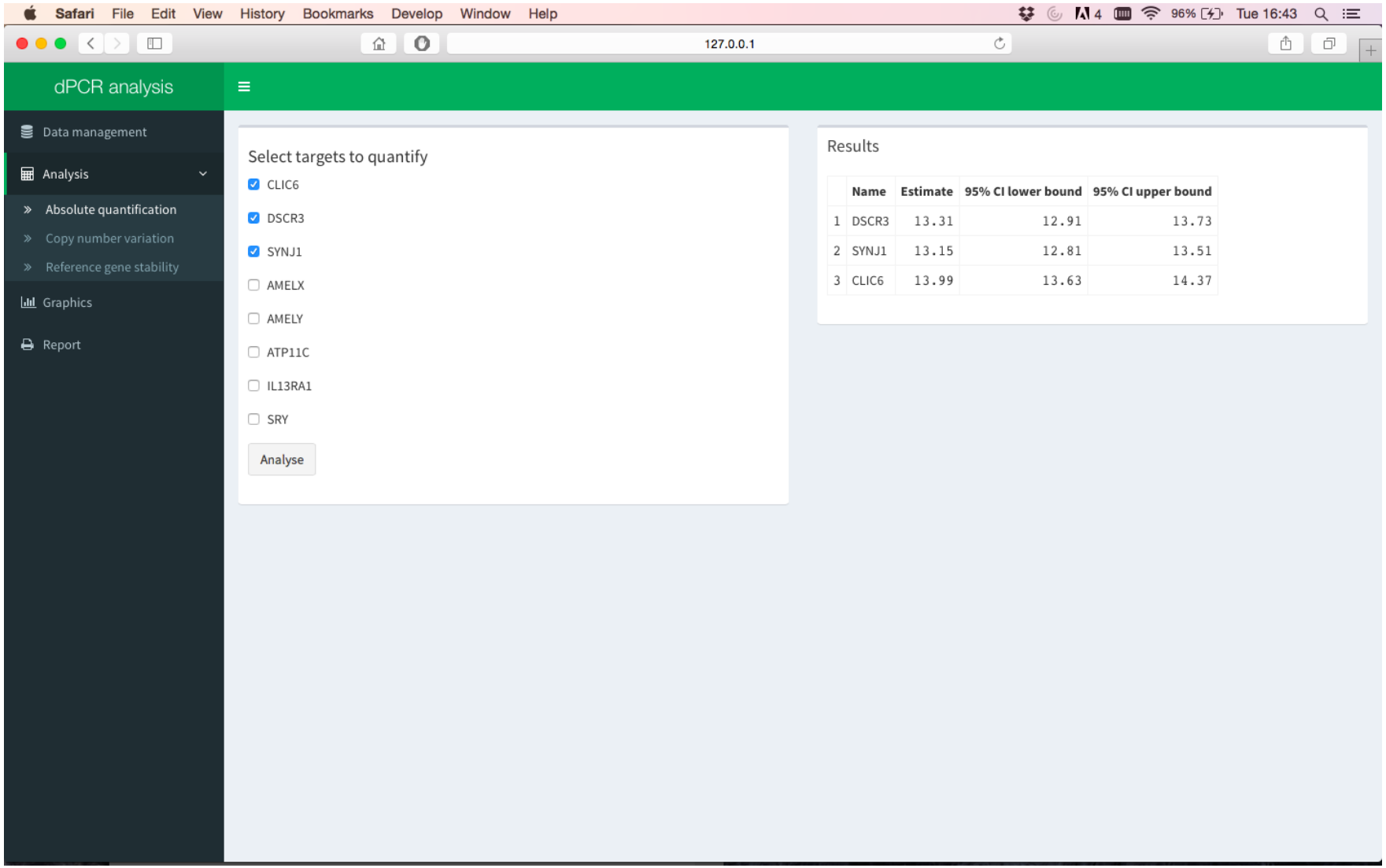
Analysis of complex experiments: Shiny App



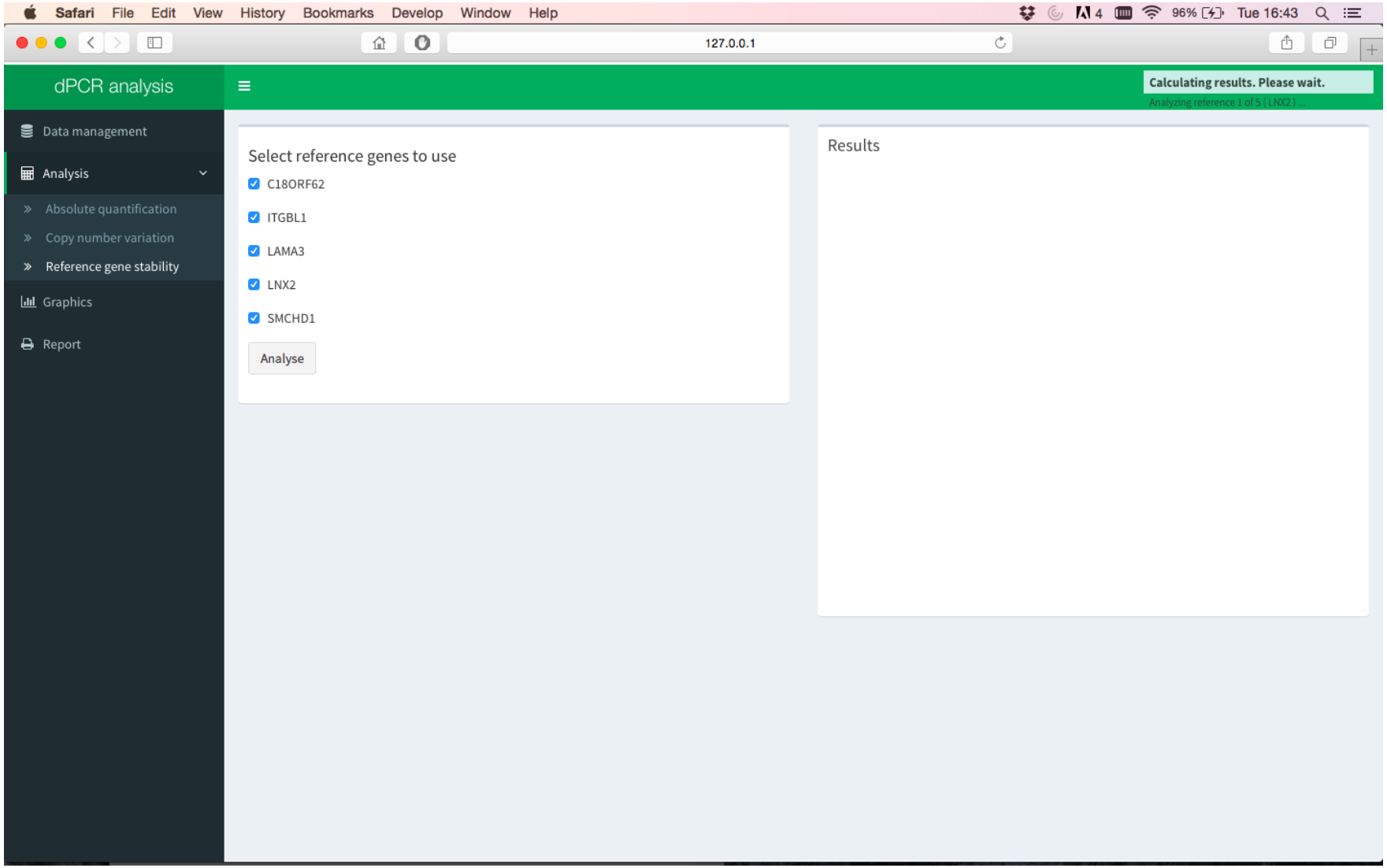
Analysis of complex experiments: Shiny App



Analysis of complex experiments: Shiny App



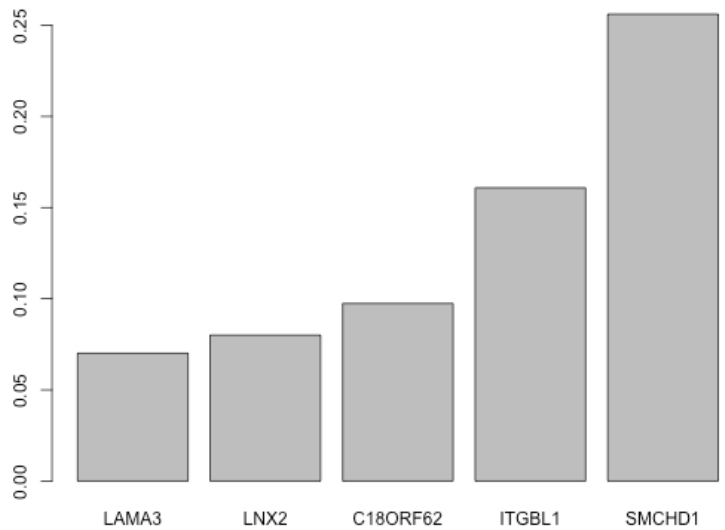
Analysis of complex experiments: Shiny App



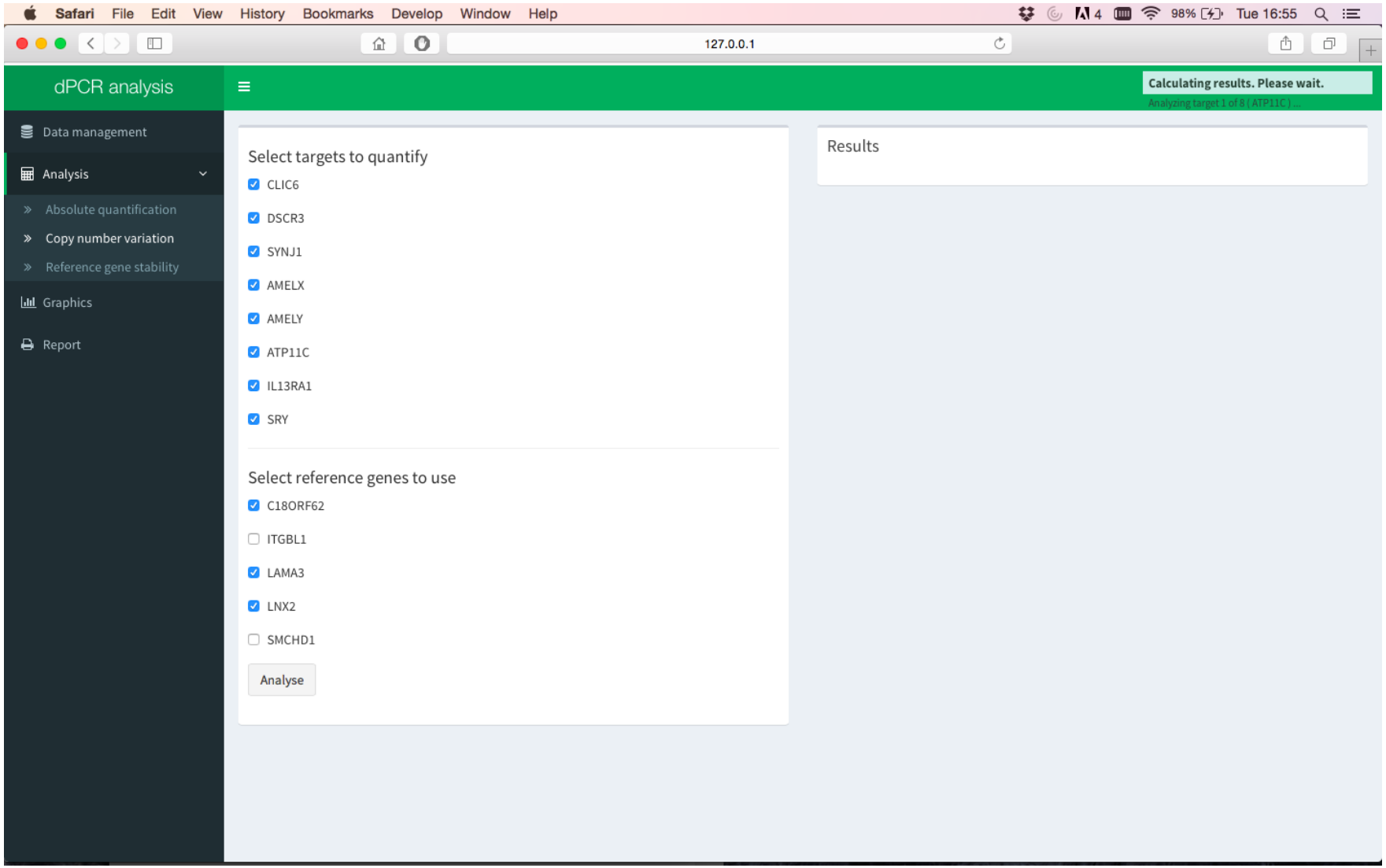
Analysis of complex experiments: Shiny App

Results

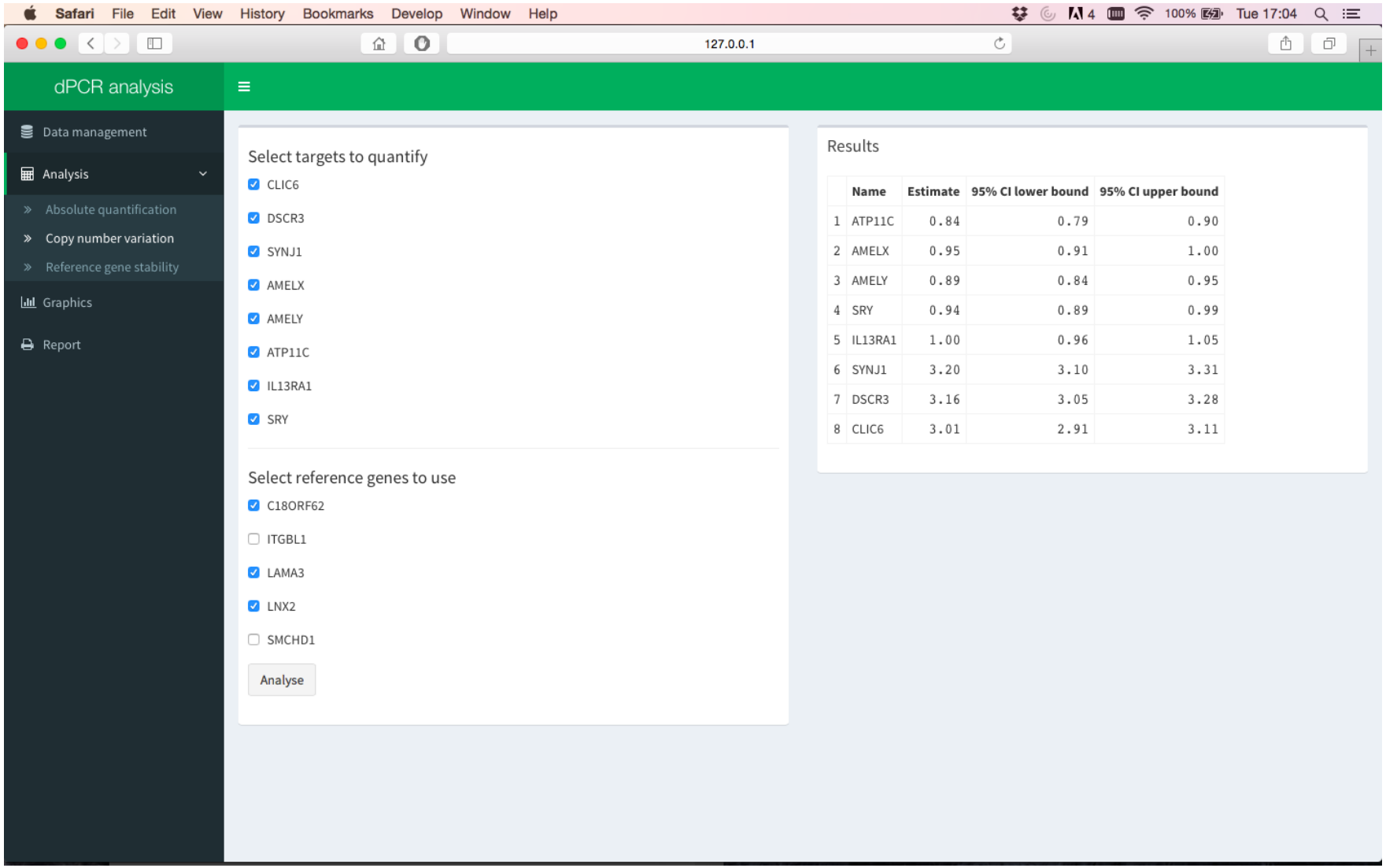
	Gene	Mean	Deviation
1	LNX2	2.05	0.08
2	SMCHD1	1.75	0.26
3	ITGBL1	2.16	0.16
4	LAMA3	2.00	0.07
5	C18ORF62	2.08	0.10



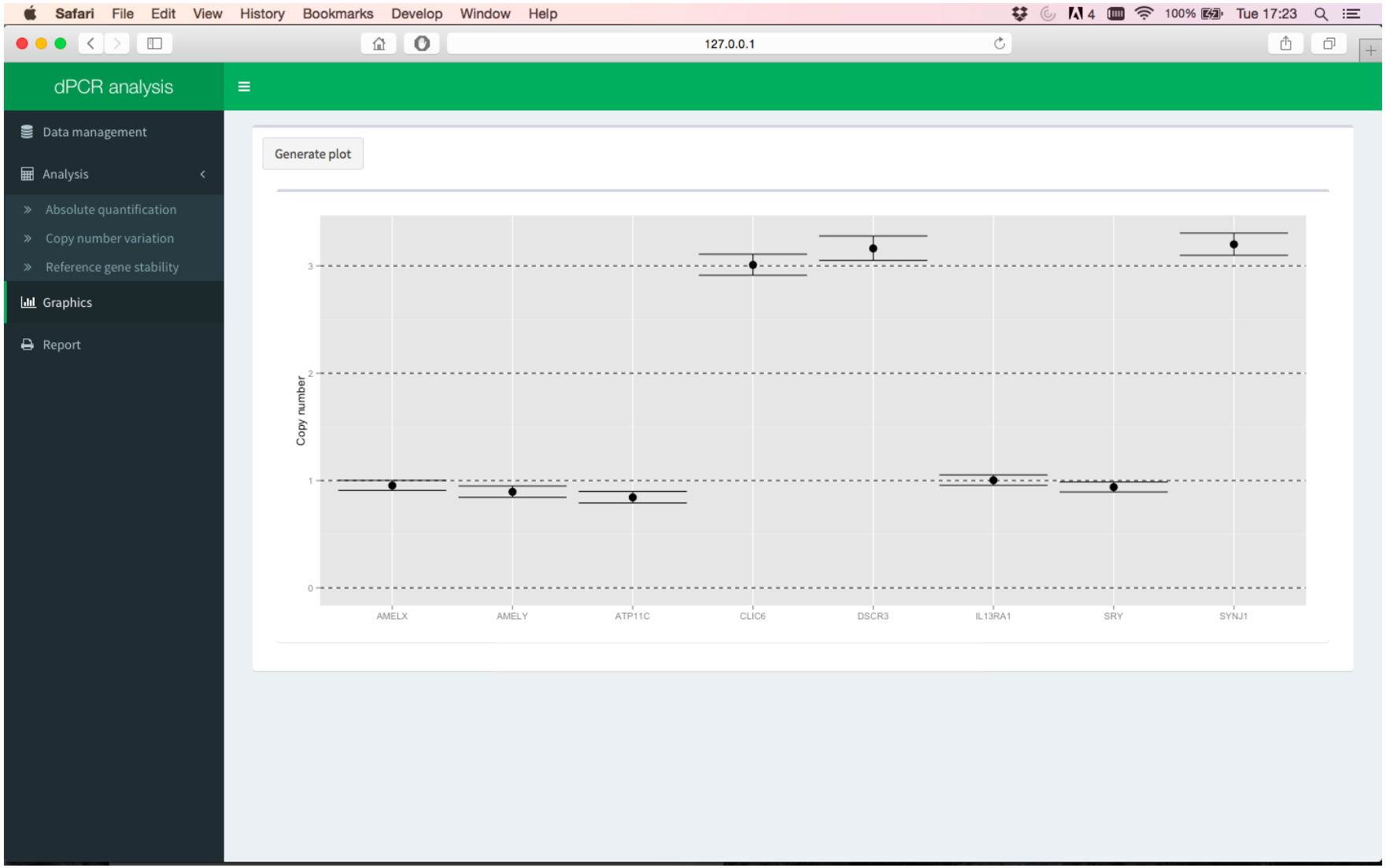
Analysis of complex experiments: Shiny App



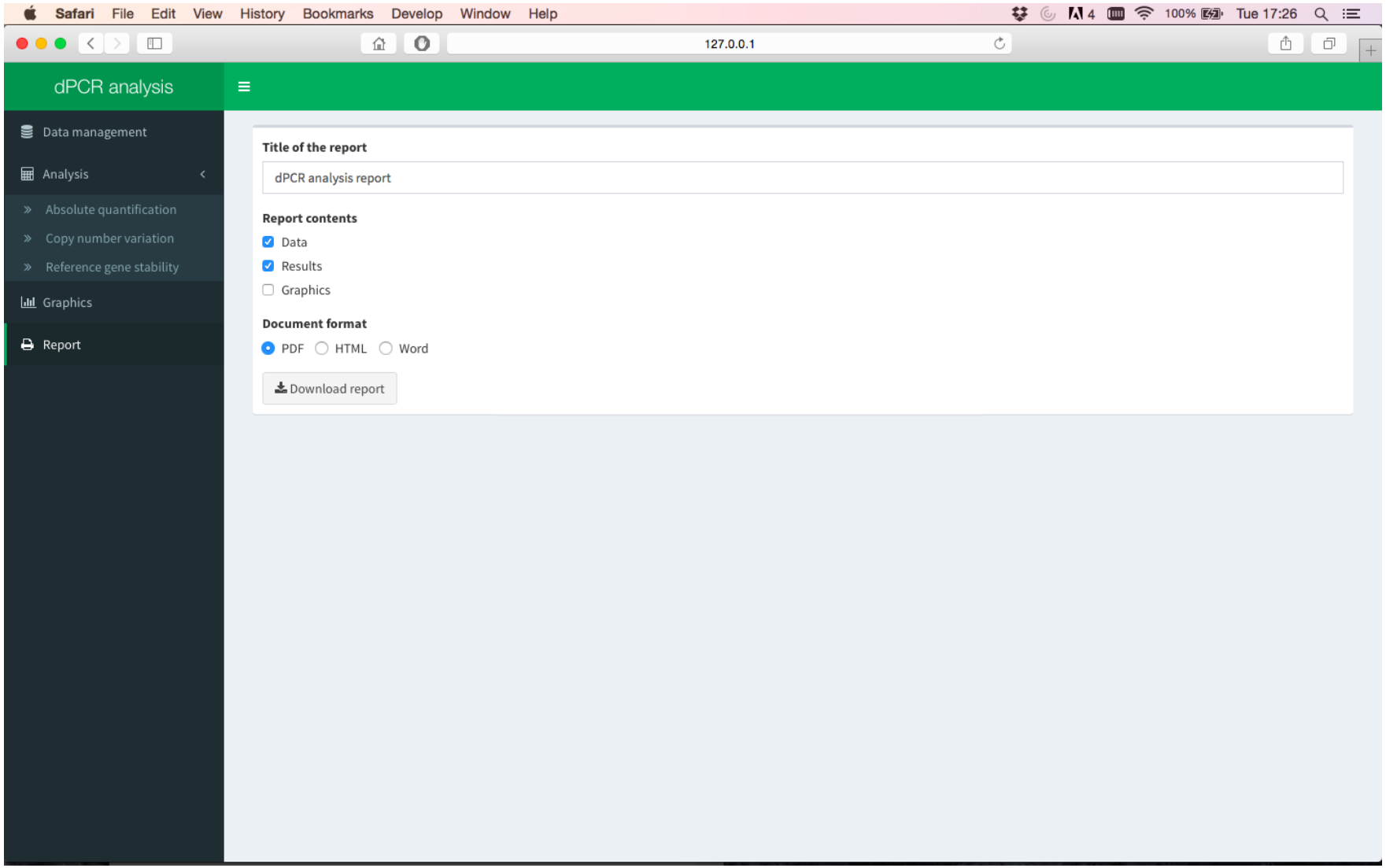
Analysis of complex experiments: Shiny App

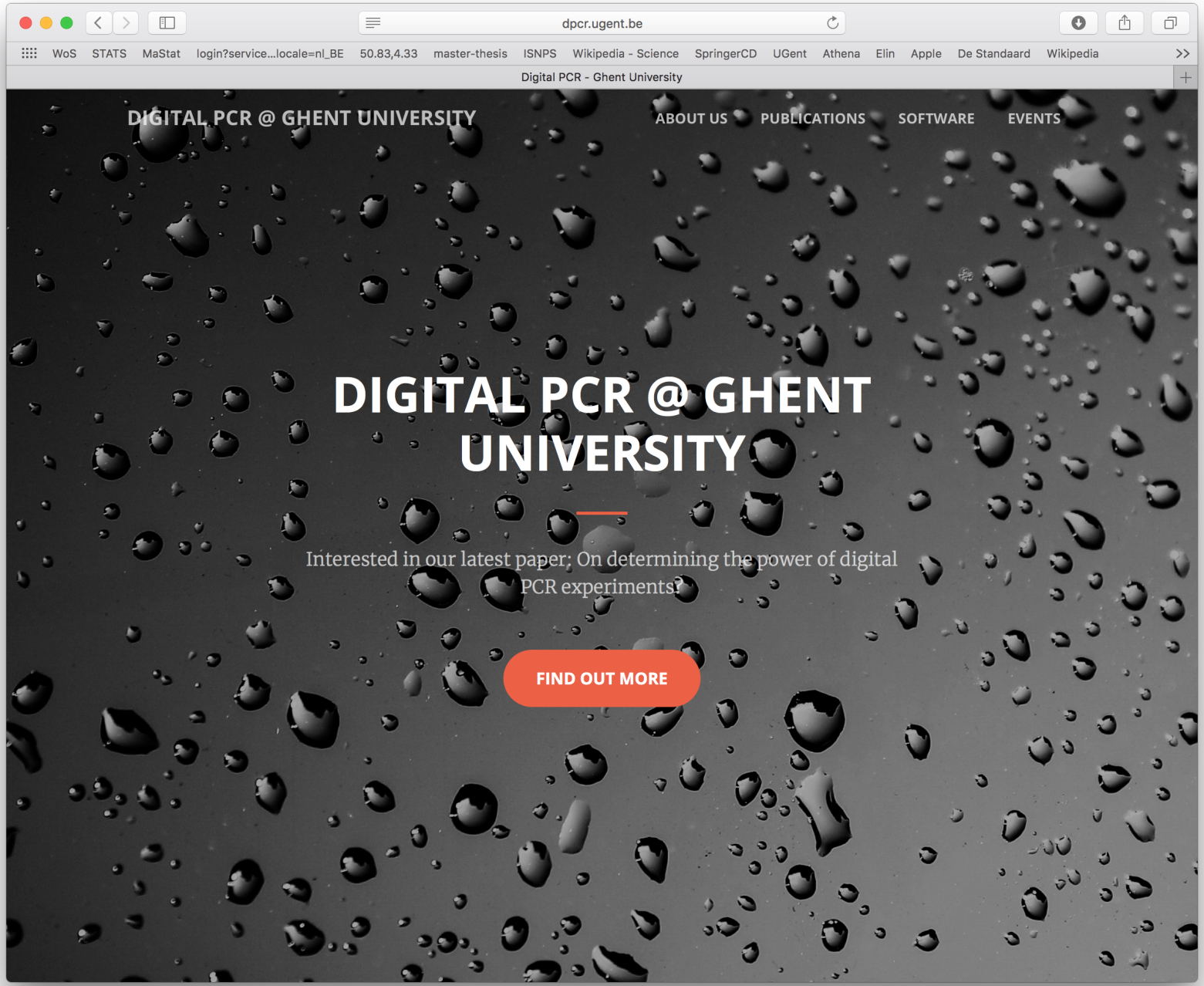


Analysis of complex experiments: Shiny App



Analysis of complex experiments: Shiny App




dpcr.ugent.be
WoS STATS MaStat login?service...locale=nl_BE 50.83,4.33 master-thesis ISNPS Wikipedia - Science SpringerCD UGent Athena Elin Apple De Standaard Wikipedia
Digital PCR - Ghent University
DIGITAL PCR @ GHENT UNIVERSITY
ABOUT US PUBLICATIONS SOFTWARE EVENTS
DIGITAL PCR @ GHENT UNIVERSITY
Interested in our latest paper; On determining the power of digital PCR experiments?
FIND OUT MORE