



Interuniversity Institute for Biostatistics
and statistical Bioinformatics



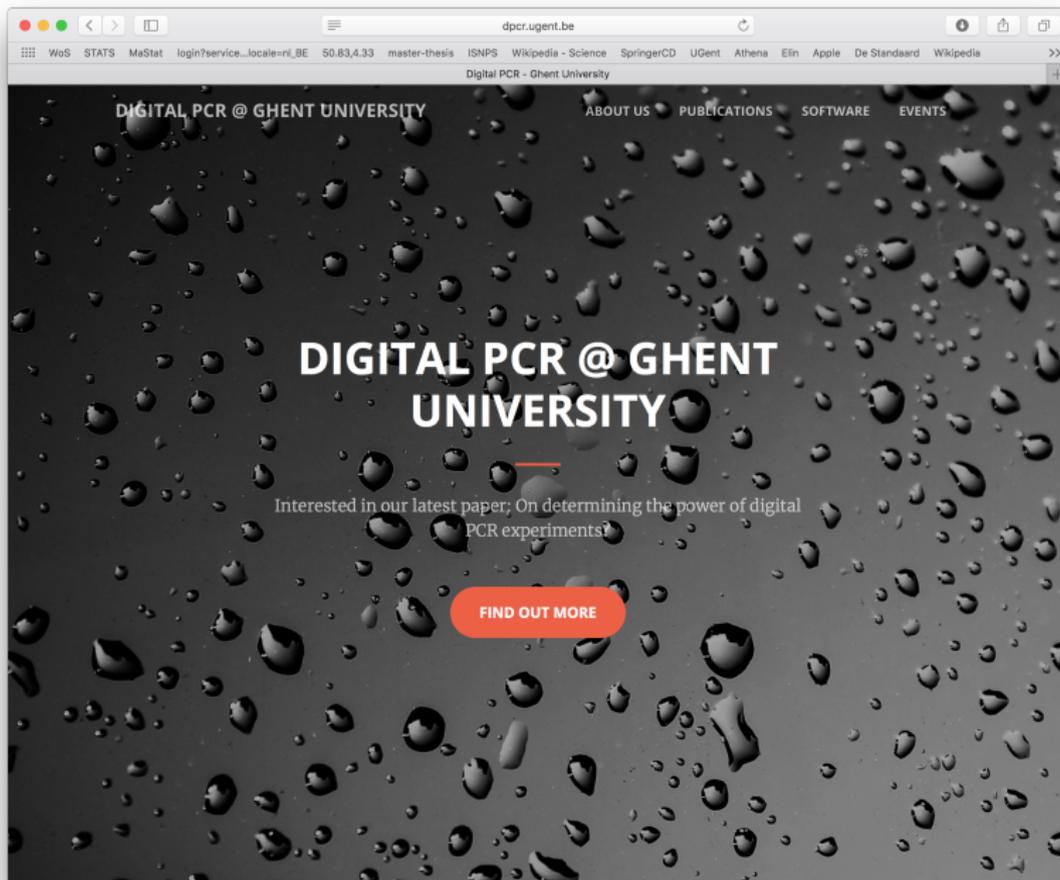
Data Analysis Pipelines and Tools for dPCR Experiments

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
- ...



In this talk ...

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

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 Trypsteen¹ · Matthijs Vynck² · Jan De Neve² · Pawel Bonczkowski¹ ·
Maja Kiselinova¹ · Eva Malatinkova¹ · Karen Vervisch¹ · Olivier Thas^{2,3} ·
Linus Vandekerckhove¹ · Ward De Spiegelaere¹

Threshold setting: Need for a threshold

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

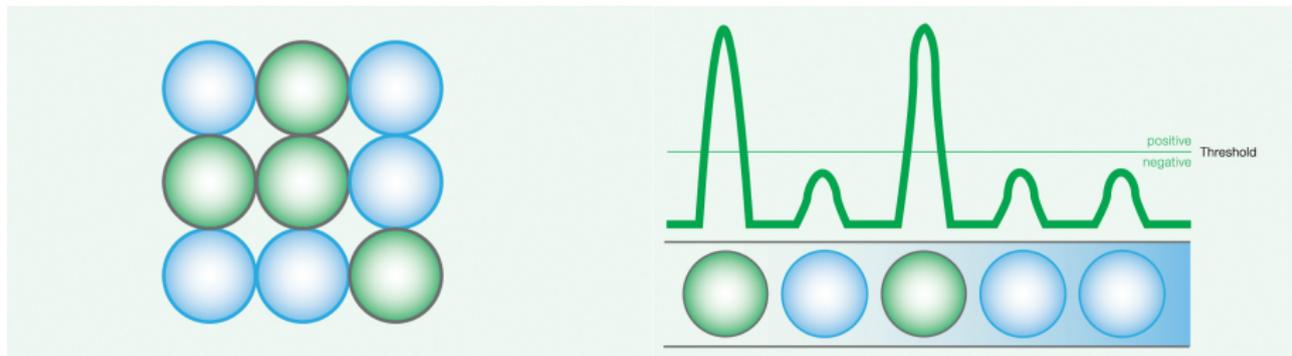


Figure: Measuring presence/absence (Bio-Rad Laboratories)

Even negative droplets show an intensity signal.

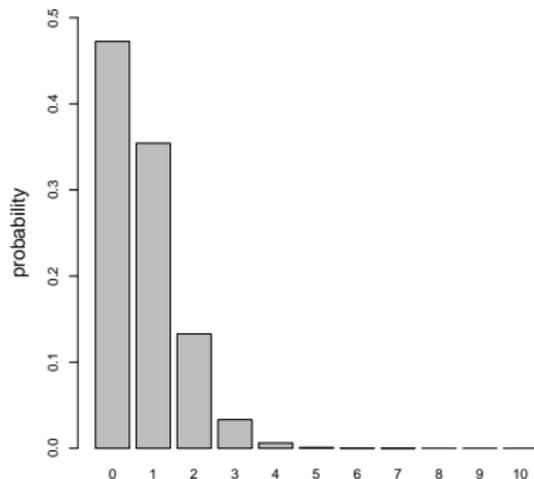
$$Y_j = \begin{cases} 0 & \text{if intensity if droplet } j < \text{threshold (negative droplet)} \\ 1 & \text{if intensity if droplet } j > \text{threshold (positive droplet)} \end{cases}$$

Threshold setting: Classical approach to quantification

dPCR data analysis methods rely on the **Poisson assumption**,

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

where $\lambda = E \{ Y_j^* \}$ 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**,

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

where $\lambda = E \{ Y_j^* \}$ 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

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

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

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

This gives an expression for λ ,

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

Threshold setting: Classical approach to quantification

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

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

The Poisson parameter can thus be estimated as

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

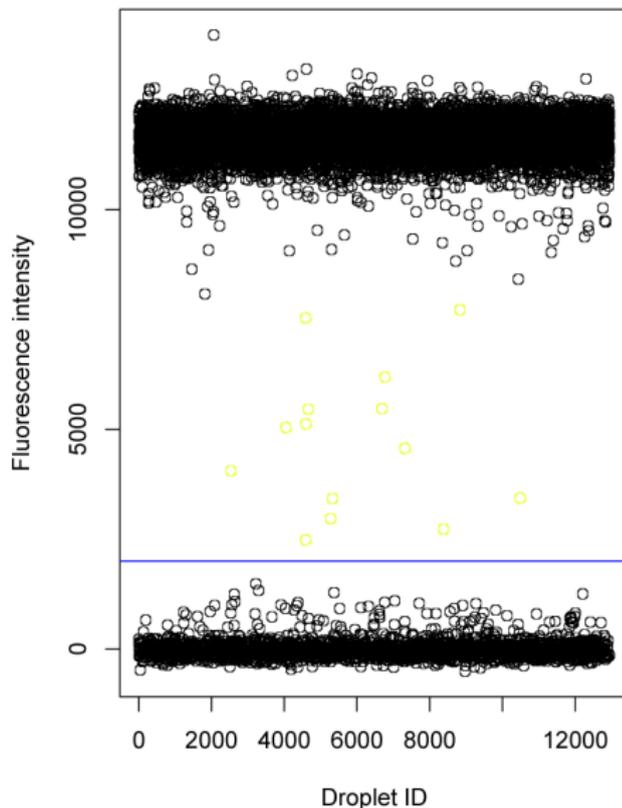
The concentration (target molecules/volume) is then estimated as:

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

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

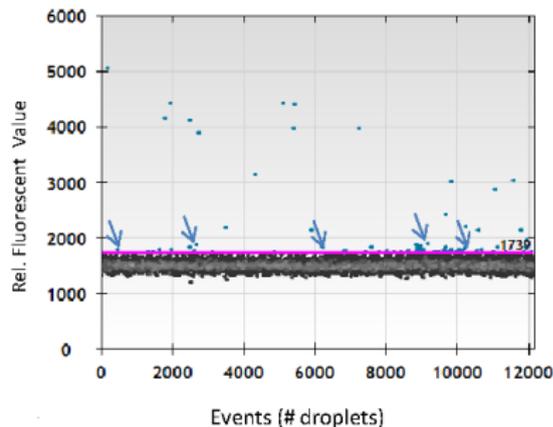
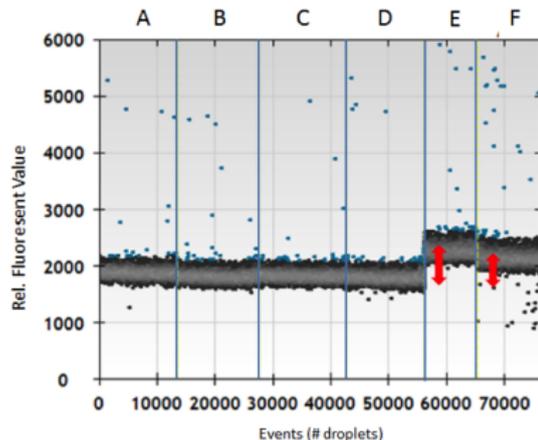
Threshold setting: Rain

- 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

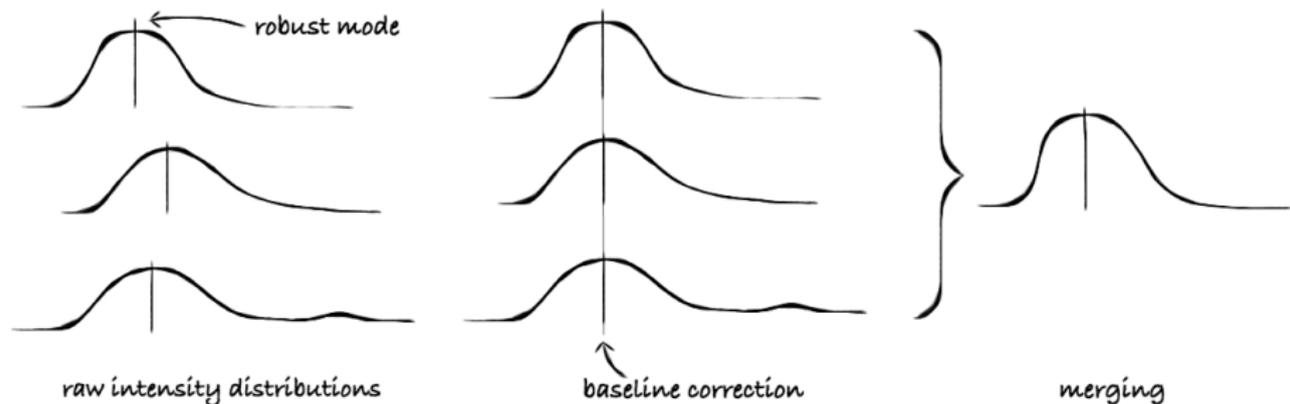


Threshold setting: Sample heterogeneity

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

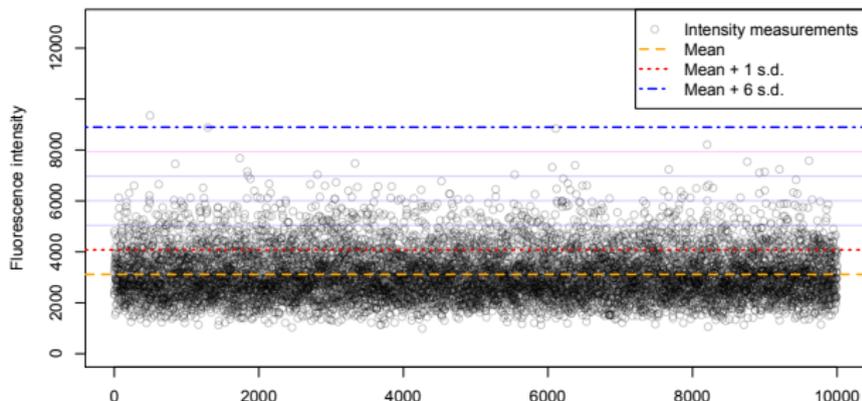


Threshold setting: Baseline correction



Threshold setting: Normality Assumption

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

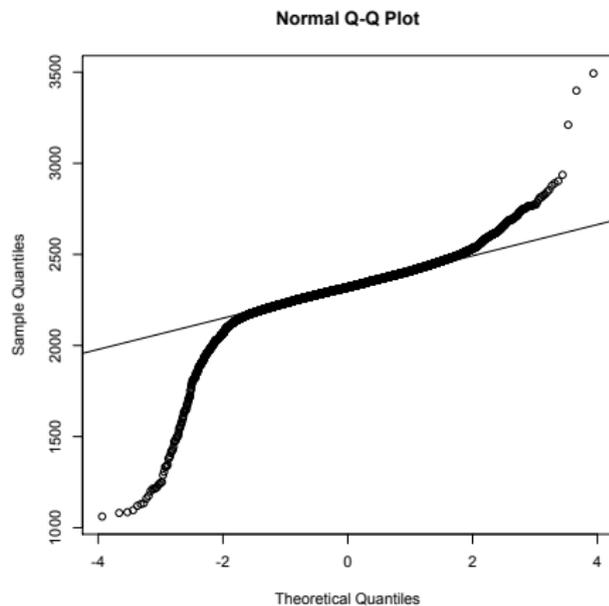
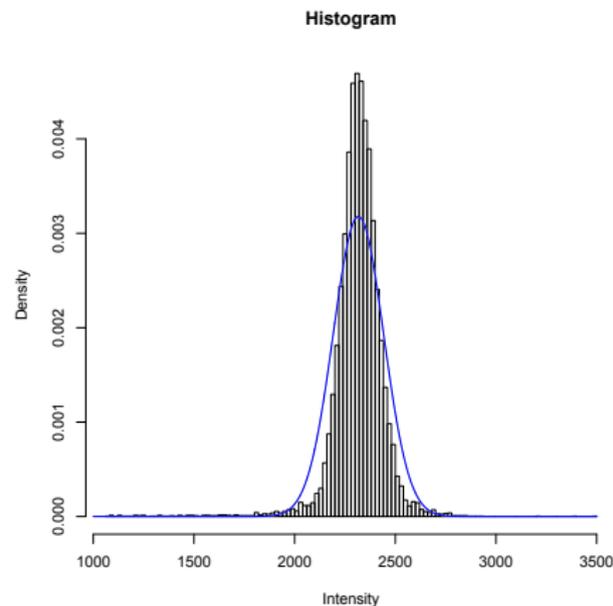


If normal assumption holds: false positive rate $\approx 10^{-9}$.

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

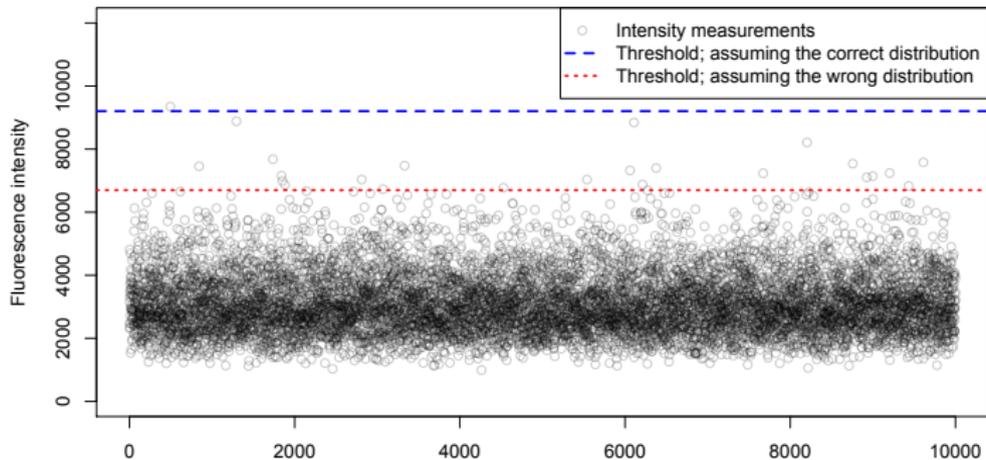
Threshold setting: Normality Assumption

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



Intensities are not normally distributed!

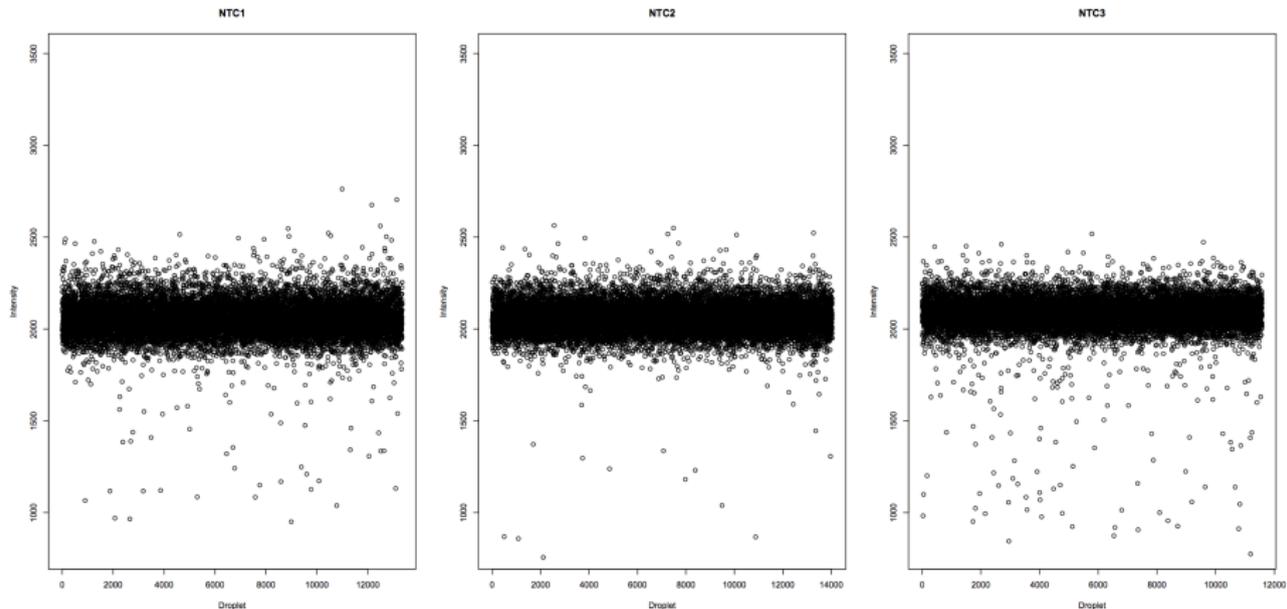
Threshold setting: Normality Assumption



→ threshold sensitive to distributional assumption

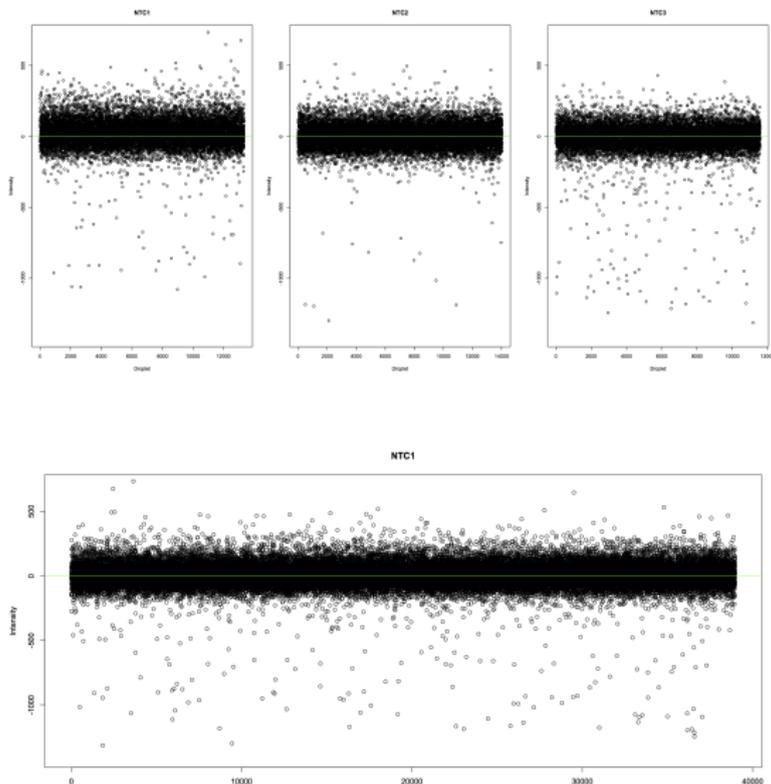
Threshold setting: Extreme Value Procedure

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



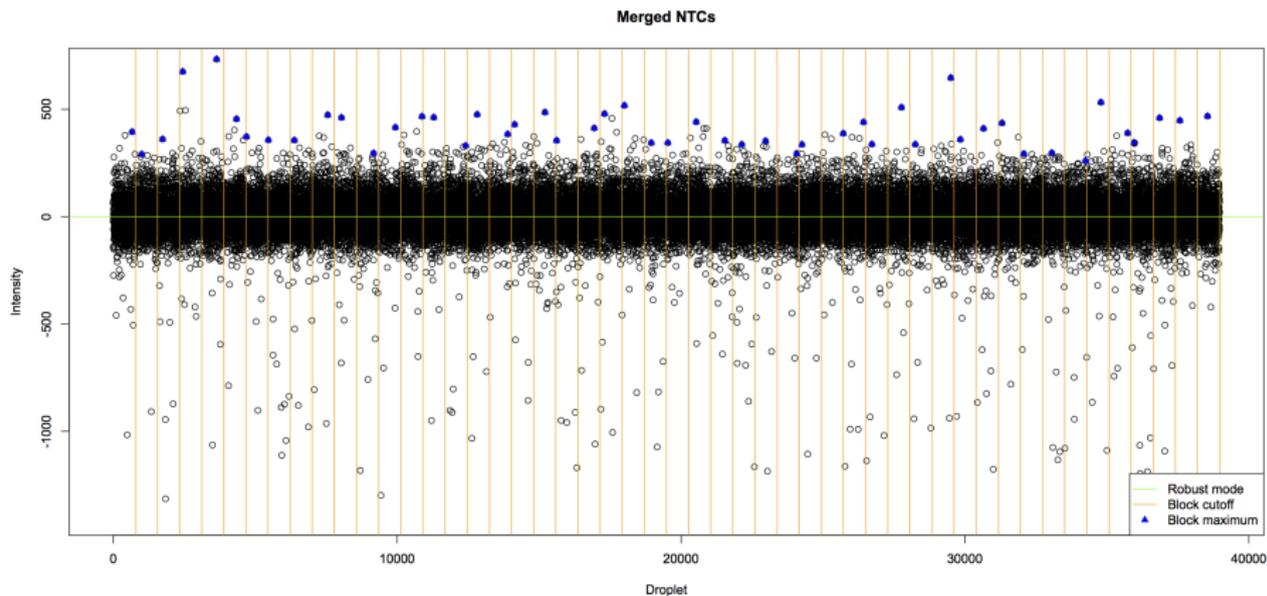
Threshold setting: Extreme Value Procedure

Step 1: Baseline correction and merging



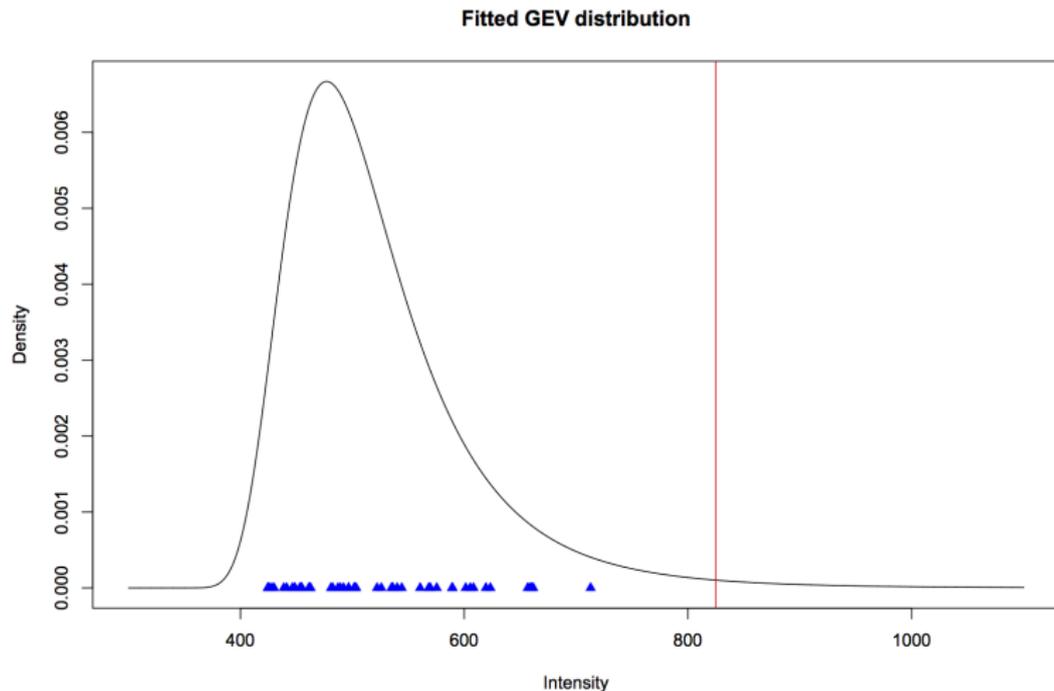
Threshold setting: Extreme Value Procedure

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



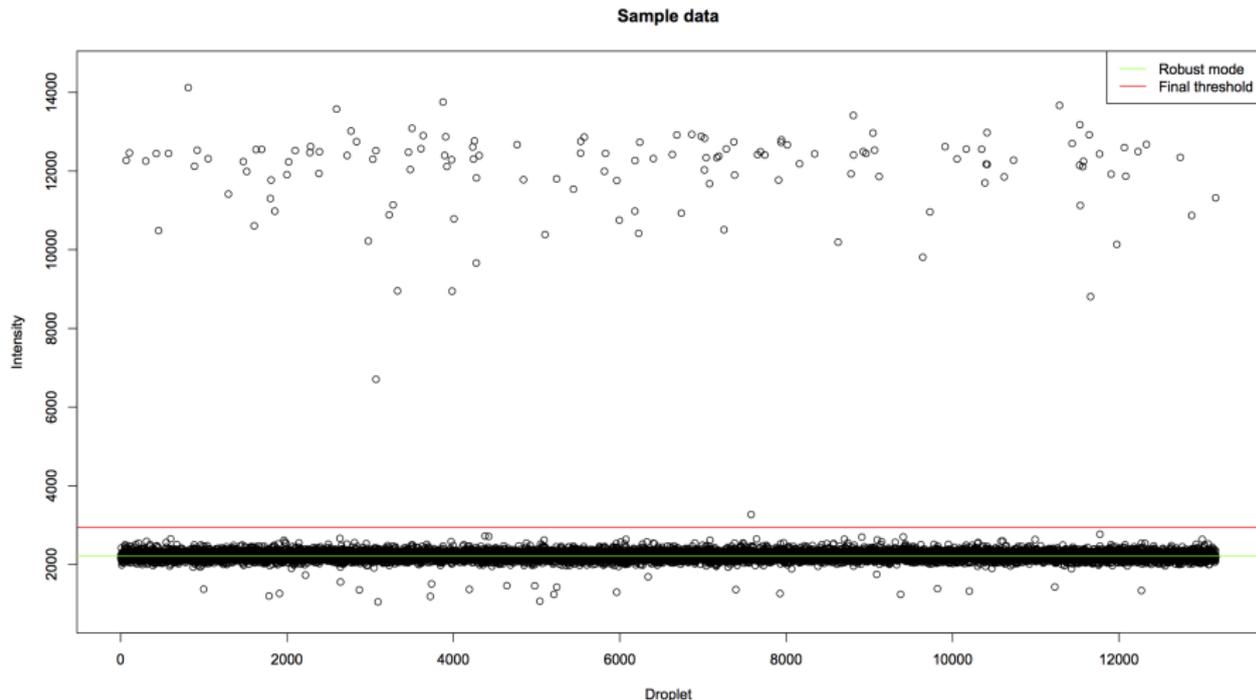
Threshold setting: Extreme Value Procedure

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

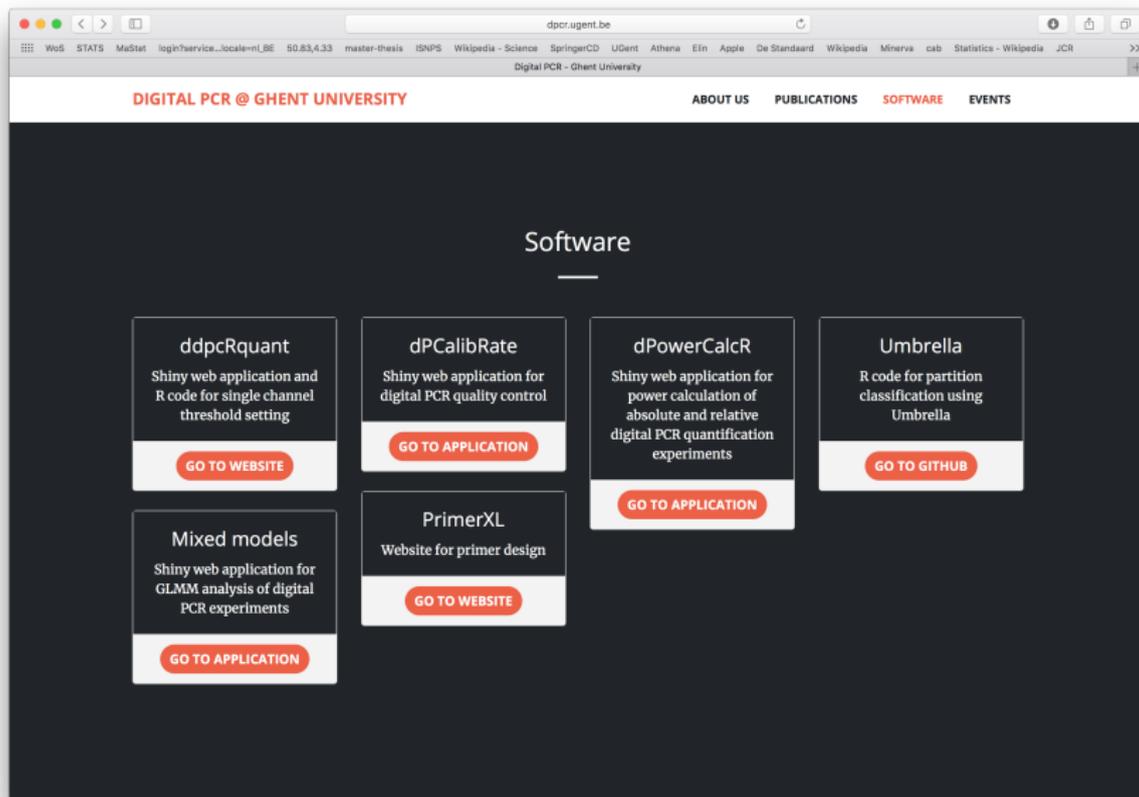


Threshold setting: Extreme Value Procedure

Step 4: Process samples



Threshold setting: Shiny App



The screenshot shows a web browser window displaying the 'Software' page of the Digital PCR @ Ghent University website. The page features a dark background with a white header and navigation menu. The main content area is titled 'Software' and contains six cards, each representing a different Shiny application or tool. Each card includes a title, a brief description, and a button to access the application or website.

Application Name	Description	Action Button
ddpcrQuant	Shiny web application and R code for single channel threshold setting	GO TO WEBSITE
dPCalibRate	Shiny web application for digital PCR quality control	GO TO APPLICATION
dPowerCalcR	Shiny web application for power calculation of absolute and relative digital PCR quantification experiments	GO TO APPLICATION
Umbrella	R code for partition classification using Umbrella	GO TO GITHUB
Mixed models	Shiny web application for GLMM analysis of digital PCR experiments	GO TO APPLICATION
PrimerXL	Website for primer design	GO TO WEBSITE

Threshold setting: Shiny App

The screenshot shows a web browser window with the URL `statapps.ugent.be`. The page title is **ddpCRquant**. A navigation menu at the top includes: [Data Input](#), [Overview Assays](#), [NTC Threshold Analysis](#), [Sample Analysis](#), [Summary](#), [Replicate Analysis](#), and [How To Use](#). The **Data Input** section contains the following controls:

- Input Data**
- Choose Input Data**: A dropdown menu currently showing "Demo Dataset".
- Upload CSV Files**: A "Browse..." button next to the text "No file selected".
- Choose BioRad QX system (demodata = qx100)**: A dropdown menu currently showing "qx100".
- Two buttons: "Start Analysis" and "Reset Analysis".

Threshold setting: Shiny App

The screenshot shows a web browser window displaying the ddpCRquant Shiny application. The browser's address bar shows the URL `statapps.ugent.be`. The application title is "ddpCRquant".

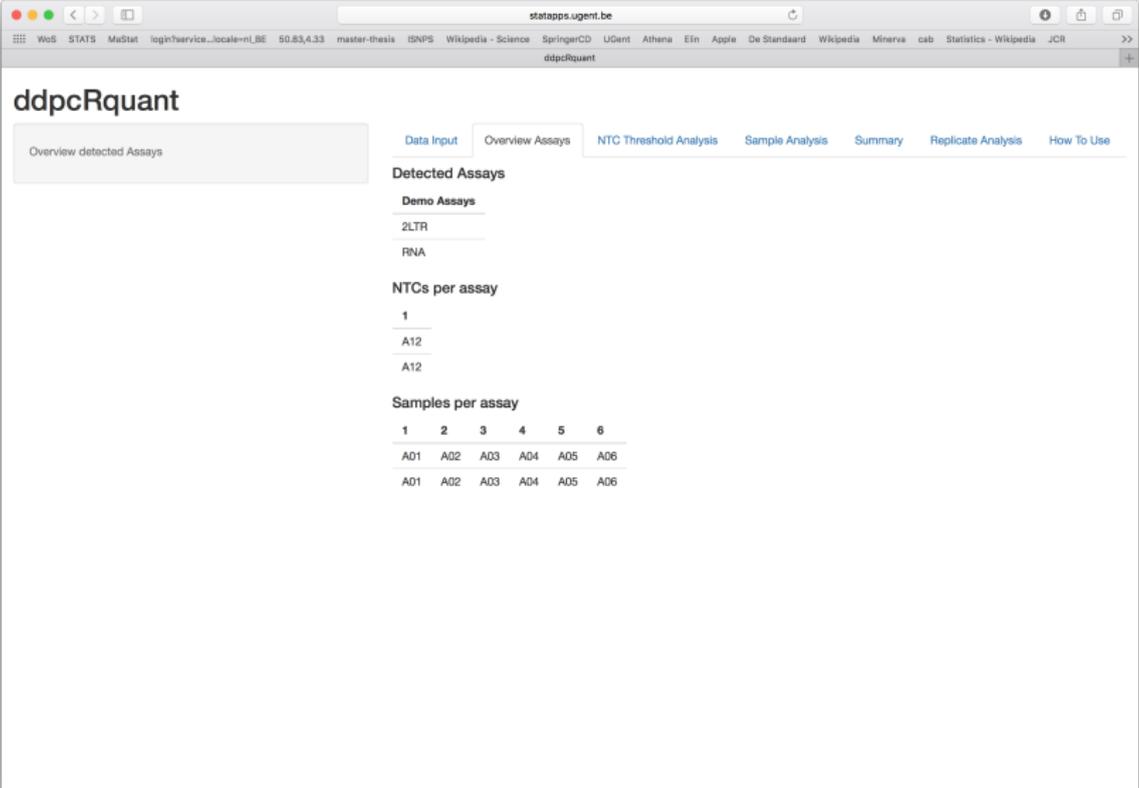
The interface is divided into two main sections:

- Input Data:** This section contains three main controls:
 - Choose Input Data:** A dropdown menu currently showing "Demo Dataset".
 - Upload CSV Files:** A "Browse..." button next to a grey box indicating "No file selected".
 - Choose BioRad QX system (demodata = qx100):** A dropdown menu currently showing "qx100".At the bottom of this section are two buttons: "Start Analysis" and "Reset Analysis".
- Assays:** A horizontal menu with tabs: "Data Input" (selected), "Overview Assays", "NTC Threshold Analysis", "Sample Analysis", "Summary", "Replicate Analysis", and "How To Use".

Under the "Data Input" tab, there is a table listing files:

FileName
2013-12-03 patients 2LTR pSIF Eva_A01_Amplitude.csv
2013-12-03 patients 2LTR pSIF Eva_A02_Amplitude.csv
2013-12-03 patients 2LTR pSIF Eva_A03_Amplitude.csv
2013-12-03 patients 2LTR pSIF Eva_A04_Amplitude.csv
2013-12-03 patients 2LTR pSIF Eva_A05_Amplitude.csv
2013-12-03 patients 2LTR pSIF Eva_A06_Amplitude.csv
2013-12-03 patients 2LTR pSIF Eva_A12_Amplitude.csv
2013-12-03 patients 2LTR pSIF Eva.csv

Threshold setting: Shiny App



The screenshot shows a web browser window with the URL `statapps.ugent.be`. The page title is **ddpCRquant**. A navigation bar includes links for [Data Input](#), [Overview Assays](#) (which is active), [NTC Threshold Analysis](#), [Sample Analysis](#), [Summary](#), [Replicate Analysis](#), and [How To Use](#). On the left, there is a button labeled "Overview detected Assays". The main content area displays the following information:

Detected Assays

Demo Assays

2LTR

RNA

NTCs per assay

1

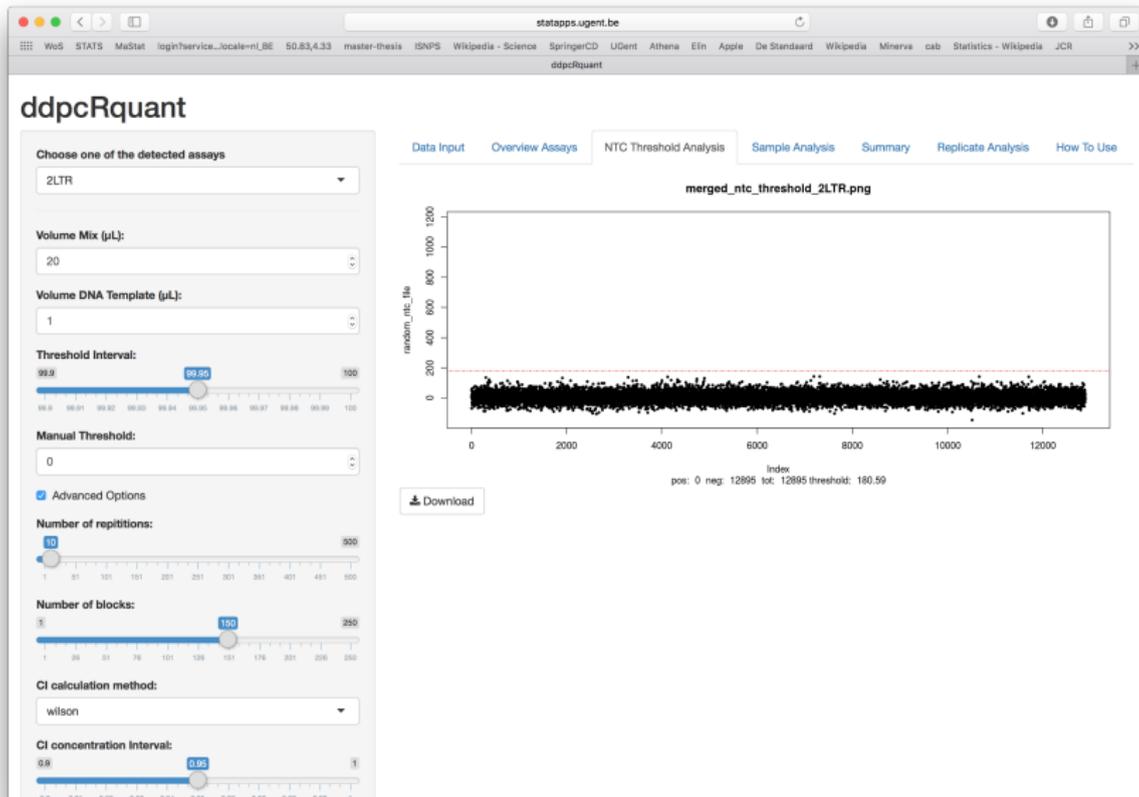
A12

A12

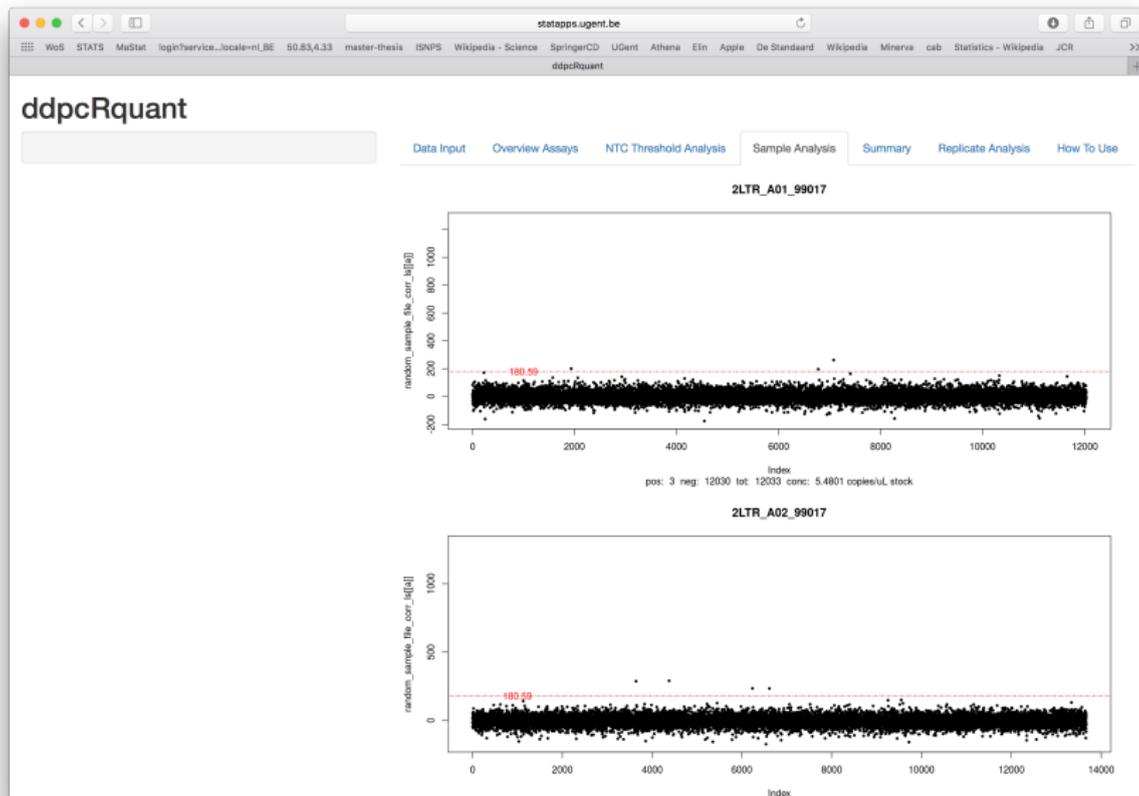
Samples per assay

1	2	3	4	5	6
A01	A02	A03	A04	A05	A06
A01	A02	A03	A04	A05	A06

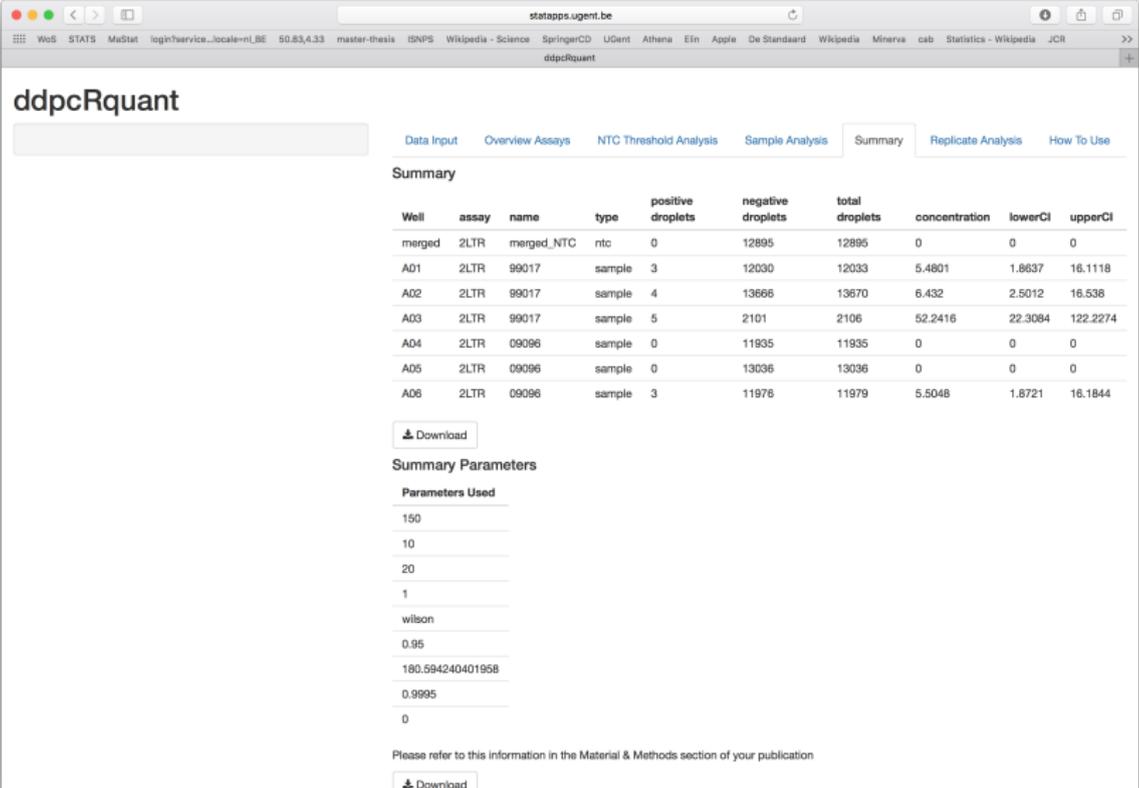
Threshold setting: Shiny App



Threshold setting: Shiny App



Threshold setting: Shiny App



The screenshot shows a web browser window displaying the ddpCRquant Shiny application. The browser's address bar shows the URL `statapps.ugent.be`. The application title is "ddpCRquant". The navigation menu includes "Data Input", "Overview Assays", "NTC Threshold Analysis", "Sample Analysis", "Summary" (which is selected), "Replicate Analysis", and "How To Use".

Summary

Well	assay	name	type	positive droplets	negative droplets	total droplets	concentration	lowerCI	upperCI
merged	2LTR	merged_NTC	ntc	0	12895	12895	0	0	0
A01	2LTR	99017	sample	3	12030	12033	5.4801	1.8637	16.1118
A02	2LTR	99017	sample	4	13666	13670	6.432	2.5012	16.538
A03	2LTR	99017	sample	5	2101	2106	52.2416	22.3084	122.2274
A04	2LTR	09096	sample	0	11935	11935	0	0	0
A05	2LTR	09096	sample	0	13036	13036	0	0	0
A06	2LTR	09096	sample	3	11976	11979	5.5048	1.8721	16.1844

[Download](#)

Summary Parameters

Parameters Used

- 150
- 10
- 20
- 1
- wilson
- 0.95
- 180.594240401958
- 0.9995
- 0

Please refer to this information in the Material & Methods section of your publication

[Download](#)

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



PAPER IN FOREFRONT

Quality control of digital PCR assays and platforms

Matthijs Vynck¹ · 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

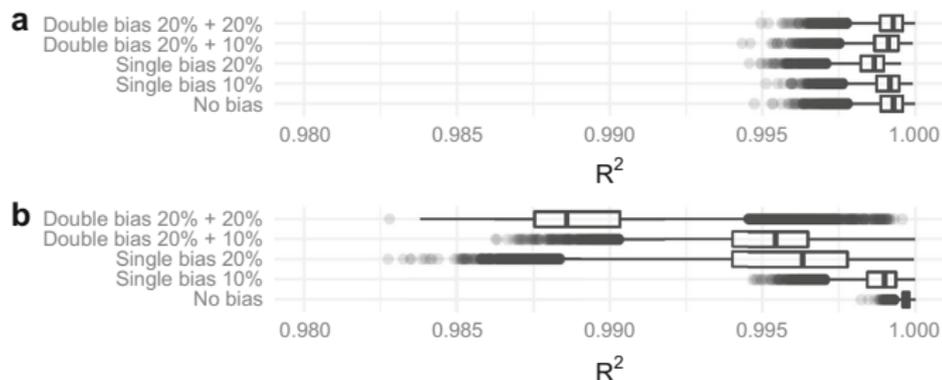
Often a large R^2 is considered as proof of

- **linear dynamic range** spanning the concentrations used for constructing 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: 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).

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

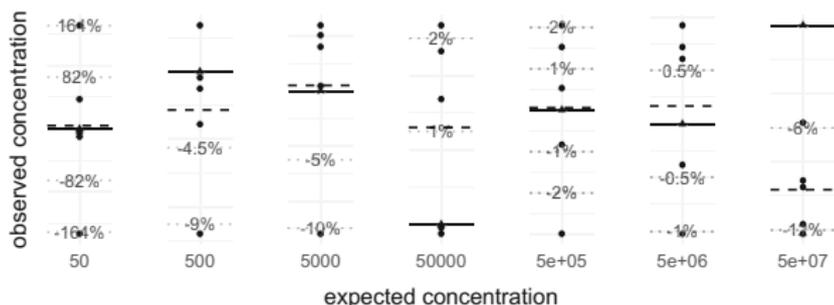
Test based on RWLS performs best.

Quality Control: Accuracy

Deviation from expected model is plotted:

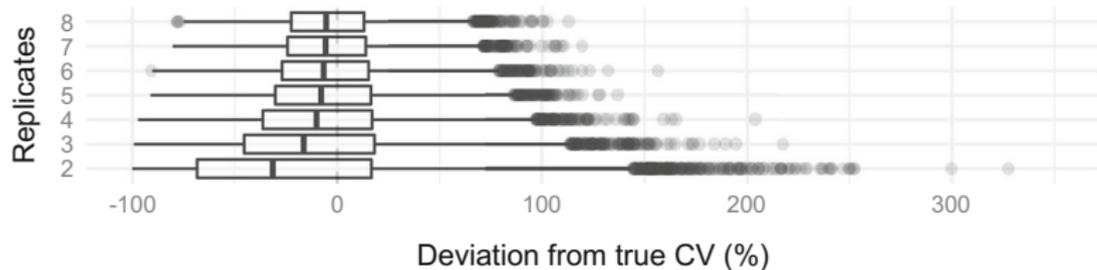
$$\text{observed dPCR conc} = 0 + 1 \times \text{expected conc.}$$

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

The screenshot shows a web browser window displaying the 'Digital PCR @ Ghent University' website. The page has a dark theme with a white navigation bar at the top. The main content area is titled 'Software' and features six application cards arranged in two rows. Each card contains the application name, a brief description, and a button to access the application or website.

DIGITAL PCR @ GHENT UNIVERSITY ABOUT US PUBLICATIONS SOFTWARE EVENTS

Software

- ddpcrQuant**
Shiny web application and R code for single channel threshold setting
[GO TO WEBSITE](#)
- dPCalibRate**
Shiny web application for digital PCR quality control
[GO TO APPLICATION](#)
- dPowerCalcR**
Shiny web application for power calculation of absolute and relative digital PCR quantification experiments
[GO TO APPLICATION](#)
- Umbrella**
R code for partition classification using Umbrella
[GO TO GITHUB](#)
- Mixed models**
Shiny web application for GLMM analysis of digital PCR experiments
[GO TO APPLICATION](#)
- PrimerXL**
Website for primer design
[GO TO WEBSITE](#)

Quality Control: Shiny app

Manual input data

Number of samples: 50

Upload data

Upload

Example data

Load example dataset

Data

	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	38566069.53	50000000.00
8	38820526.64	50000000.00
9	396448.40	50000000.00
10	3995351.79	50000000.00
11	3985238.48	50000000.00
12	4057414.57	50000000.00
13	3961996.08	50000000.00
14	4047799.60	50000000.00
15	4080735.54	50000000.00
16	4054979.92	50000000.00
17	496819.51	50000000.00
18	494871.03	50000000.00
19	513889.84	50000000.00
20	499032.46	50000000.00
21	495454.47	50000000.00
22	502289.21	50000000.00
23	501598.45	50000000.00
24	503971.84	50000000.00
25	49828.35	50000000.00
26	50861.91	50000000.00
27	51132.47	50000000.00
28	49412.72	50000000.00
29	51374.11	50000000.00
30	49565.62	50000000.00

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



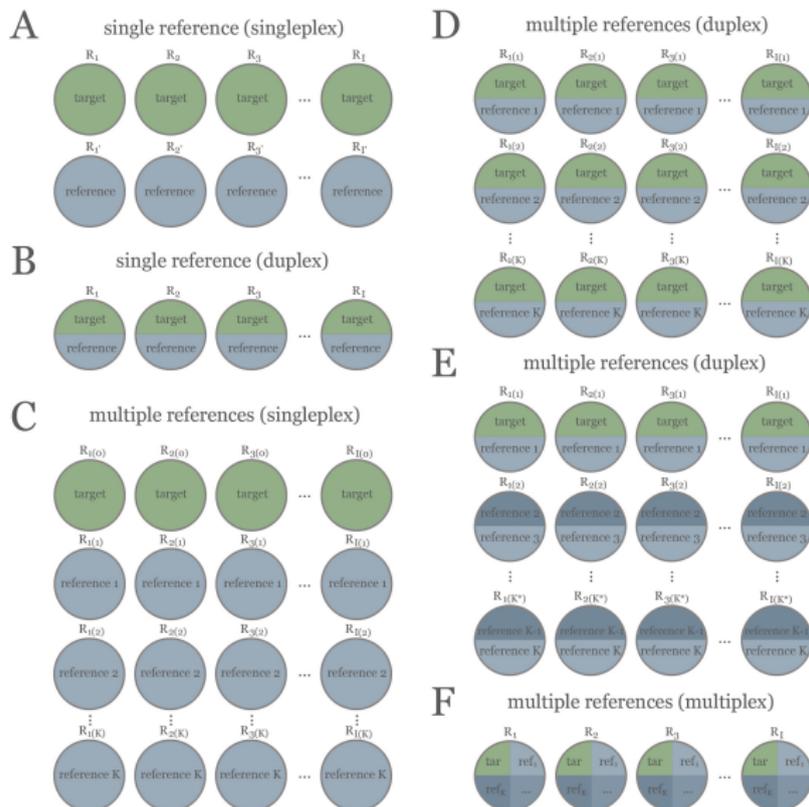
Research Paper

Flexible analysis of digital PCR experiments using generalized linear mixed models

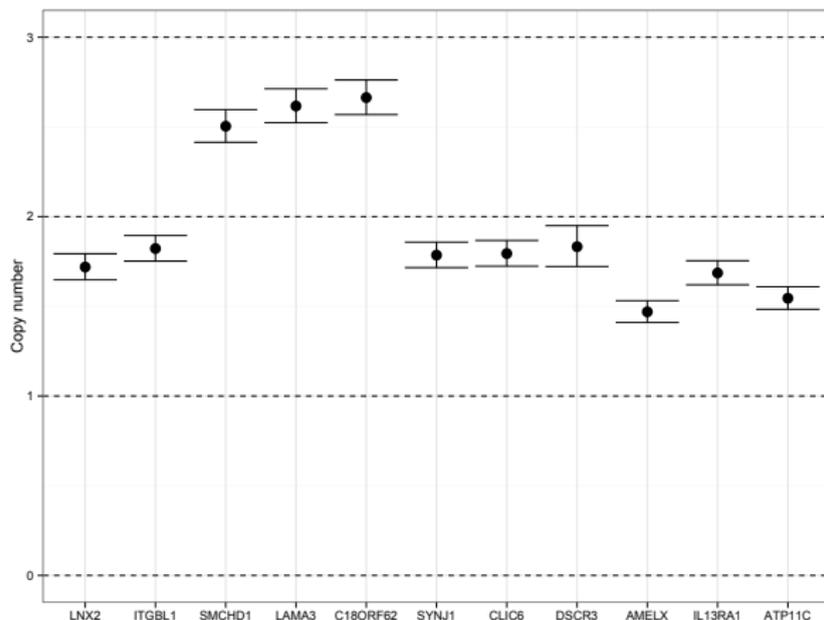
Matthijs Vynck^{a,*}, Jo Vandesompele^{b,c,d}, Nele Nijs^d, Björn Menten^{b,c}, Ariane De Ganck^d, Olivier Thas^{a,e}



Analysis of complex experiments: Designs CNV

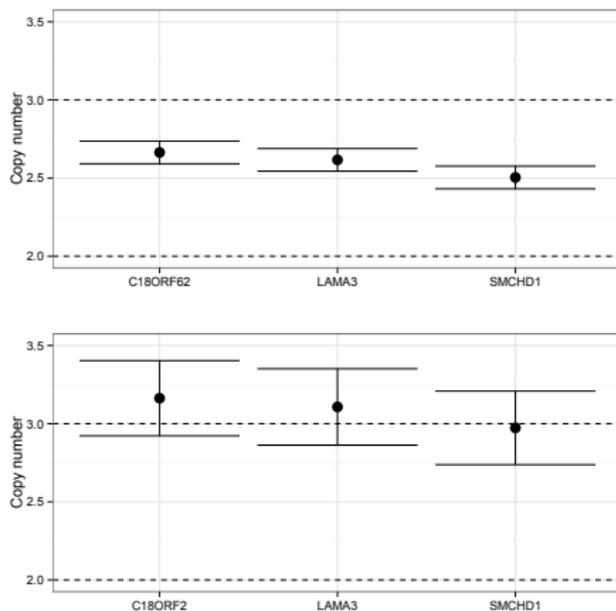


Analysis of complex experiments: CNV with single versus multiple ref. genes



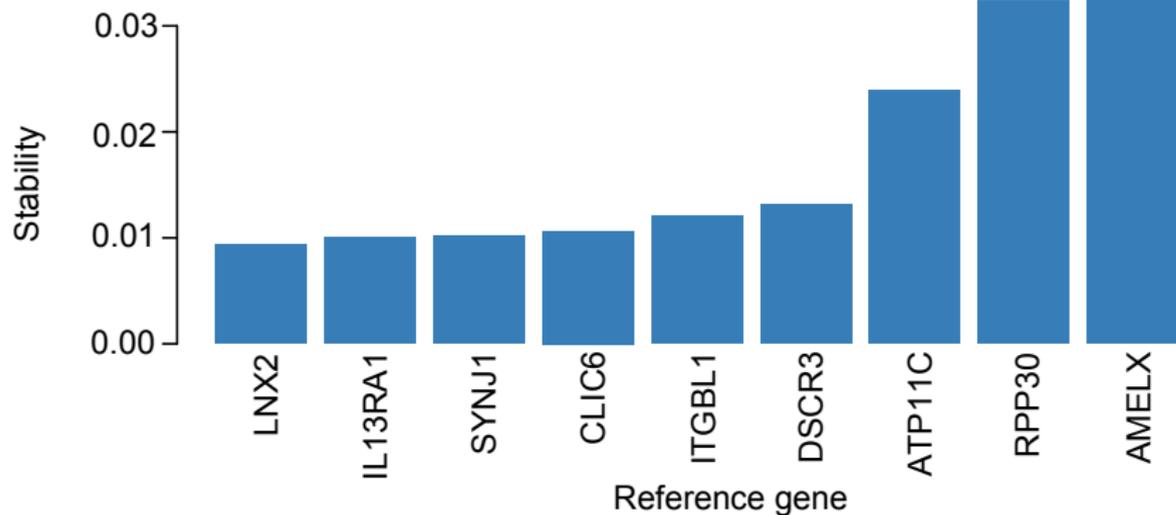
→ bias caused by single reference gene

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

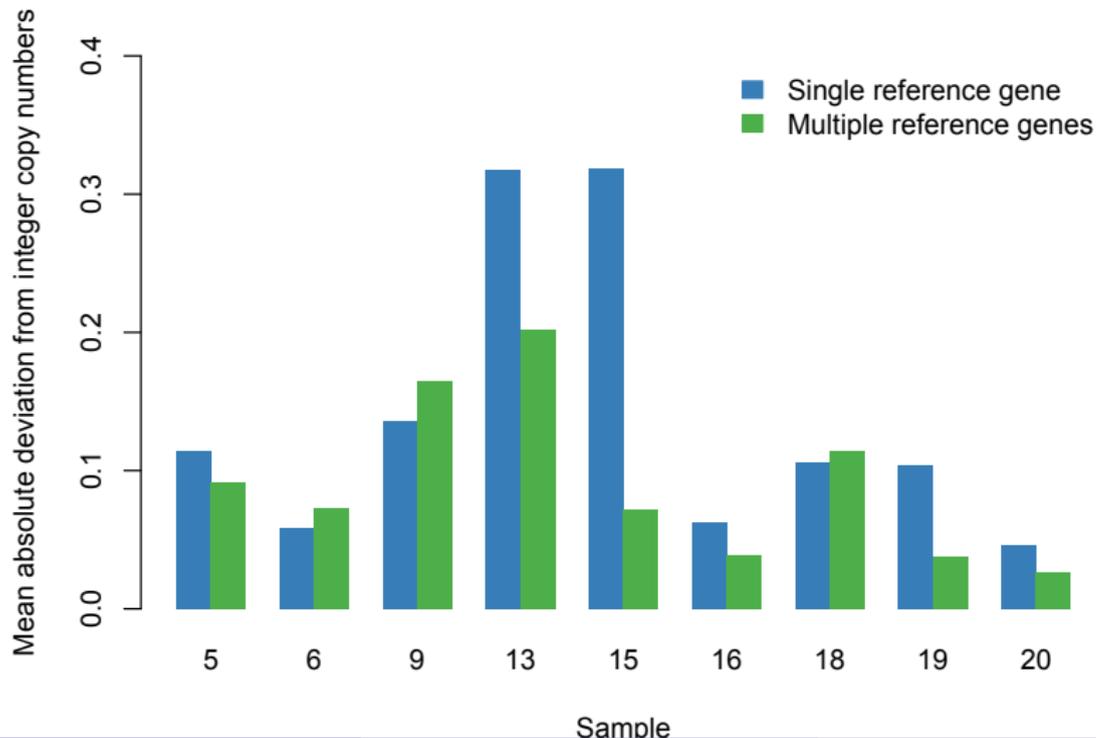


→ bias resolved by using multiple reference genes

Analysis of complex experiments: Reference gene stability



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



Analysis of complex experiments: Shiny App

The screenshot shows a web browser window displaying a Shiny application titled "dPCR analysis". The browser's address bar shows "127.0.0.1". The application interface has a green header bar with the title and a hamburger menu icon. On the left, there is a dark sidebar with navigation options: "Data management", "Analysis", "Graphics", and "Report". The main content area is divided into several sections:

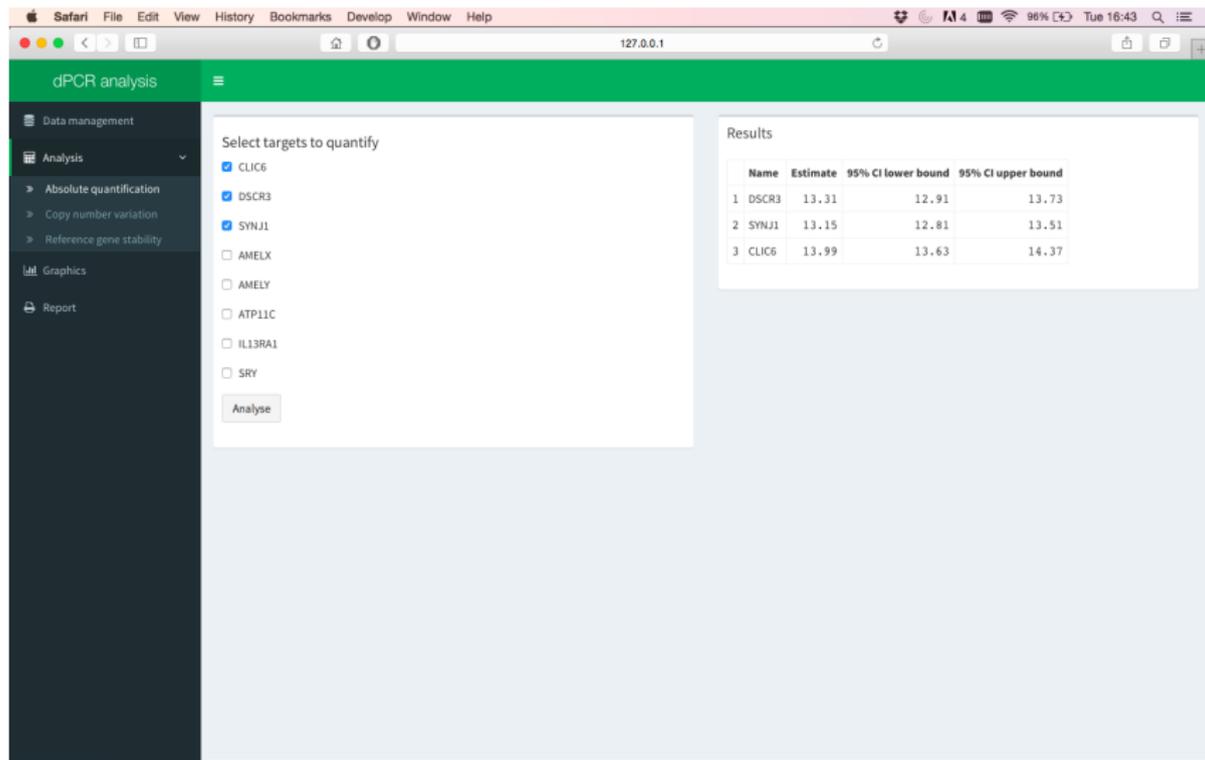
- Manual input data:** A section with a slider control for "Number of samples". The slider is currently set to 35, with a range from 1 to 100.
- Upload data:** A section with an "Upload" button.
- Download data:** A section with a "Download" button.
- Data:** A table displaying the results of the analysis. The table has columns for "Name", "Positive", "Negative", and "Target".

	Name	Positive	Negative	Target
1	CLIC6	1666	10929	<input checked="" type="checkbox"/>
2	CLIC6	1986	12757	<input checked="" type="checkbox"/>
3	CLIC6	1861	12189	<input checked="" type="checkbox"/>
4	DSCR3	2134	13173	<input checked="" type="checkbox"/>
5	DSCR3	1937	11939	<input checked="" type="checkbox"/>
6	SYNJ1	1788	11285	<input checked="" type="checkbox"/>
7	SYNJ1	1856	11011	<input checked="" type="checkbox"/>
8	SYNJ1	2035	12286	<input checked="" type="checkbox"/>
9	AMELX	650	13335	<input checked="" type="checkbox"/>
10	AMELX	597	13712	<input checked="" type="checkbox"/>
11	AMELX	671	14352	<input checked="" type="checkbox"/>
12	AMELY	616	14312	<input checked="" type="checkbox"/>
13	AMELY	630	14392	<input checked="" type="checkbox"/>
14	ATP11C	314	9123	<input checked="" type="checkbox"/>
15	ATP11C	545	13030	<input checked="" type="checkbox"/>
16	ATP11C	635	14264	<input checked="" type="checkbox"/>
17	IL13RA1	573	11036	<input checked="" type="checkbox"/>
18	IL13RA1	672	13822	<input checked="" type="checkbox"/>
19	IL13RA1	681	14621	<input checked="" type="checkbox"/>
20	SRY	543	12052	<input checked="" type="checkbox"/>
21	SRY	676	14067	<input checked="" type="checkbox"/>
22	SRY	586	13464	<input checked="" type="checkbox"/>
23	C18ORF62	1370	13558	<input type="checkbox"/>

Analysis of complex experiments: Shiny App

The screenshot shows a web browser window displaying a Shiny application titled "dPCR analysis". The browser's address bar shows the URL "127.0.0.1". The application interface has a green header bar with the text "Calculating results. Please wait." on the right. On the left, there is a dark sidebar with a menu containing "Data management", "Analysis", "Absolute quantification", "Copy number variation", "Reference gene stability", "Graphics", and "Report". The "Analysis" menu is expanded, showing sub-options. The main content area is divided into two panels. The left panel, titled "Select targets to quantify", contains a list of target genes with checkboxes: CLIC6 (checked), DSCR3 (checked), SYNJ1 (checked), AMELX (unchecked), AMELY (unchecked), ATP11C (unchecked), IL13RA1 (unchecked), and SRY (unchecked). Below the list is an "Analyse" button. The right panel, titled "Results", is currently empty.

Analysis of complex experiments: Shiny App



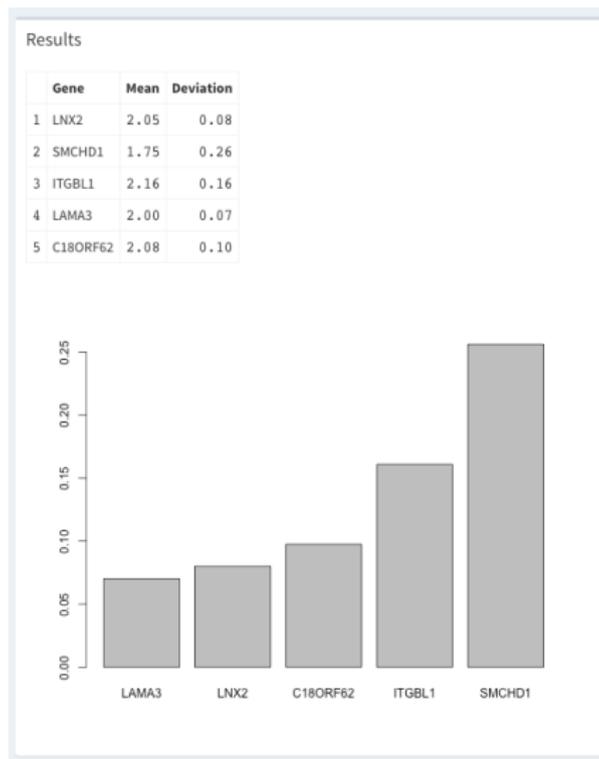
The screenshot shows a web browser window displaying a Shiny application titled "dPCR analysis". The browser's address bar shows the URL "127.0.0.1". The application interface is divided into a dark sidebar on the left and a main content area on the right. The sidebar contains a menu with the following items: "Data management", "Analysis" (expanded), "Absolute quantification", "Copy number variation", "Reference gene stability", "Graphics", and "Report". The "Analysis" section is active, showing a "Select targets to quantify" panel with a list of genes: CLIC6, DSCR3, SYNJ1, AMELX, AMELY, ATP11C, IL13RA1, and SRY. The first three genes (CLIC6, DSCR3, and SYNJ1) are selected with blue checkboxes. Below the list is an "Analyse" button. To the right of the selection panel is a "Results" panel containing a table with the following data:

	Name	Estimate	95% CI lower bound	95% CI upper bound
1	DSCR3	13.31	12.91	13.73
2	SYNJ1	13.15	12.81	13.51
3	CLIC6	13.99	13.63	14.37

Analysis of complex experiments: Shiny App

The screenshot shows a web browser window displaying a Shiny application titled "dPCR analysis". The browser's address bar shows the URL "127.0.0.1". The application interface features a green header bar with the text "Calculating results. Please wait." on the right. A dark sidebar on the left contains a navigation menu with the following items: "Data management", "Analysis" (expanded), "Absolute quantification", "Copy number variation", "Reference gene stability", "Graphics", and "Report". The main content area is divided into two panels. The left panel, titled "Select reference genes to use", contains a list of five genes with checkboxes: "C18ORF62", "ITGBL1", "LAMA3", "LNX2", and "SMCHD1". All checkboxes are checked. Below the list is an "Analyse" button. The right panel, titled "Results", is currently empty.

Analysis of complex experiments: Shiny App



Analysis of complex experiments: Shiny App

The screenshot shows a web browser window displaying a Shiny application titled "dPCR analysis". The browser's address bar shows the URL "127.0.0.1". The application interface features a dark sidebar on the left with navigation options: "Data management", "Analysis" (expanded), "Absolute quantification", "Copy number variation", "Reference gene stability", "Graphics", and "Report". The main content area is divided into two sections. The left section, titled "Select targets to quantify", contains a list of target genes with checkboxes: CLIC6, DSCR3, SYNJ1, AMELX, AMELY, ATP11C, IL13RA1, and SRY. The right section, titled "Select reference genes to use", contains a list of reference genes with checkboxes: C18ORF62, ITGBL1, LAMA3, LNX2, and SMCHD1. Below this list is an "Analyse" button. On the right side of the application, there is a "Results" box, and a status bar at the top right indicates "Calculating results. Please wait."

Analysis of complex experiments: Shiny App

dPCR analysis

Data management

Analysis

- Absolute quantification
- Copy number variation
- Reference gene stability

Graphics

Report

Select targets to quantify

- CLIC6
- DSCR3
- SYNJ1
- AMELX
- AMELY
- ATP11C
- IL13RA1
- SRY

Select reference genes to use

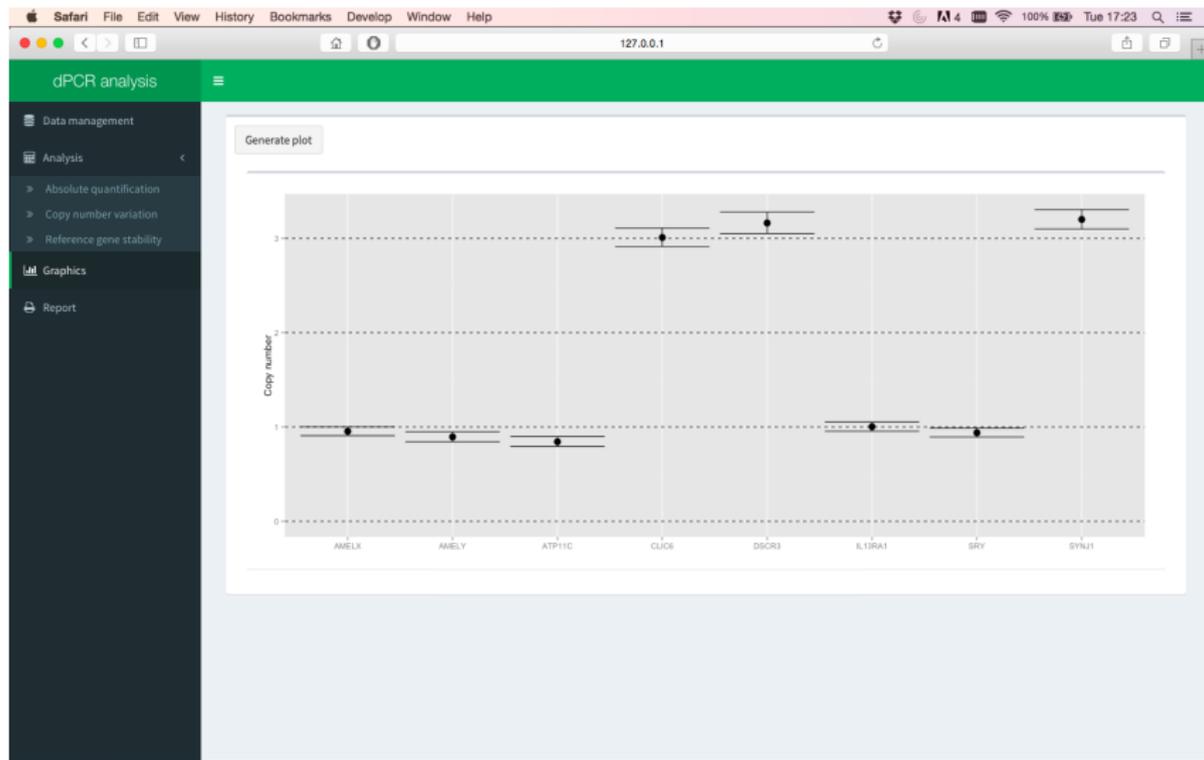
- C18ORF62
- ITGBL1
- LAMA3
- LNX2
- SMCHD1

Analyse

Results

	Name	Estimate	95% CI lower bound	95% CI upper bound
1	ATP11C	0.84	0.79	0.90
2	AMELX	0.95	0.91	1.00
3	AMELY	0.89	0.84	0.95
4	SRY	0.94	0.89	0.99
5	IL13RA1	1.00	0.96	1.05
6	SYNJ1	3.20	3.10	3.31
7	DSCR3	3.16	3.05	3.28
8	CLIC6	3.01	2.91	3.11

Analysis of complex experiments: Shiny App



Analysis of complex experiments: Shiny App

The screenshot shows a web browser window displaying a Shiny application titled "dPCR analysis". The browser's address bar shows the URL "127.0.0.1". The application interface features a dark grey sidebar on the left with a menu containing "Data management", "Analysis", "Graphics", and "Report". The "Analysis" section is expanded, showing sub-options: "Absolute quantification", "Copy number variation", and "Reference gene stability". The main content area is white and contains a form for generating a report. The form has three sections: "Title of the report" with a text input field containing "dPCR analysis report"; "Report contents" with three checkboxes: "Data" (checked), "Results" (checked), and "Graphics" (unchecked); and "Document format" with three radio buttons: "PDF" (selected), "HTML", and "Word". Below these sections is a "Download report" button with a download icon.

