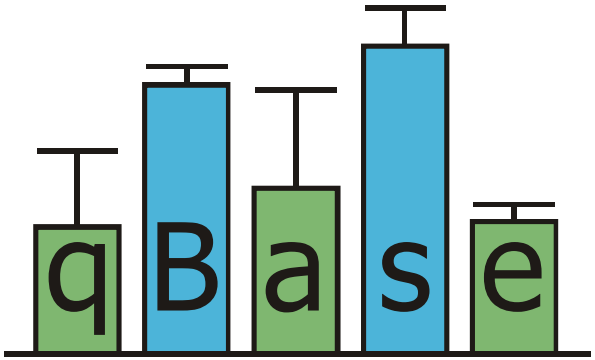
qBase

management and automated analysis of qPCR data



Jan Hellemans
Center for Medical Genetics
Ghent University Hospital, Belgium

Center for Medical Genetics

Introduction

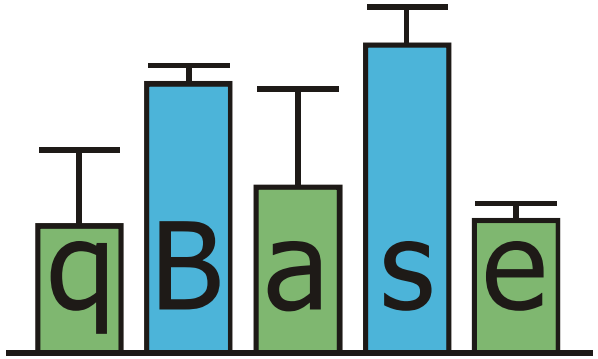
- Nucleic acid quantification: increasingly important in
  - Biological research
  - Clinical decision making
- Practical performance feasible for most labs because of
  - Improved detection chemistry & design guidelines
  - Advanced instruments
- Remaining problems
  - Accurate and straightforward data processing
  - Management of ever growing data sets
  - Need for flexible & user friendly software:

    SoFar, REST, Q-gene, DART, GENEX (BioRad), …

Limitations current qPCR gene expression analysis tools

- Dedicated for one instrument, cumbersome data import
- Limited number of samples or genes (<1run / plate)
- Fixed number of replicates
- Only one reference gene
- Lack of data quality controls
  - ***Replicate variability***
  - ***Standard curve***
  - ***NTC control***
  - ***Reference gene stability***
- Inaccurate error propagation / quantification
- Limited visualisation / rescaling
- Lack of experiment archive
- Closed architecture

Our solution: qBase



- Excel workbooks + VBA scripts
- Free – open source (Artistic license)
- Management & automated analysis

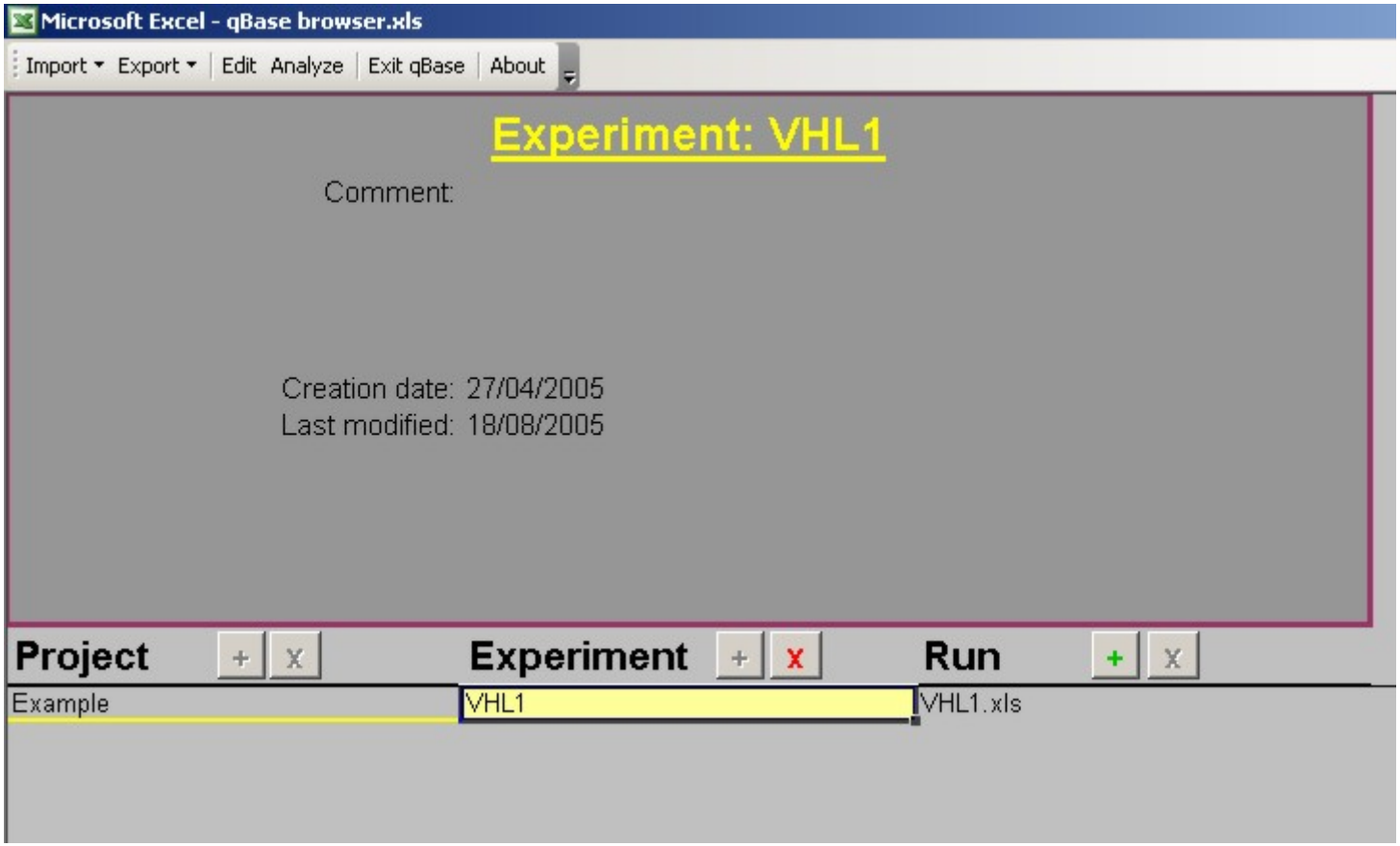
→ Analyzer module

→ Browser module

Organize data on three levels

- Run (plate)
- Experiment: group of runs that need to be processed as a whole
  - Wells with identical sample and gene name are considered replicates
  - Sample results are relative to those of all the other samples
  - 2 independent (e.g. biological) replicates = 2 different experiments results are relative within an experiment but not between the two
- Project: group of related experiments

qBase Browser



Menu

Annotation

Add & remove

Browser

- Manage data
- Annotate projects / experiments / runs
- Add & remove projects / experiments / runs
- Import & export projects / experiments (data exchange)
- Import run data
- Start analysis

- qBase format (.xls)
  - Well
  - (sample) type
  - Sample (name)
  - Gene (name)
  - Ct / Cp
  - Quantity
- Conversion of proprietary formats into qBase format
  - Many formats ➔ Lots of work
  - Difficulties – errors
  - Continued adaptation for newer formats
- Universal XML (work in progress)

XML

- XML = Extensible Markup Language
- Open format (extensible with future additions)
- Identical for all qPCR users
- Facilitates data exchange between
  - Users
  - Software packages (e.g. data collection software vs. qBase)
- Can be read by everyone
- Allows submission of raw data along with publication
- Format is still open for discussion …

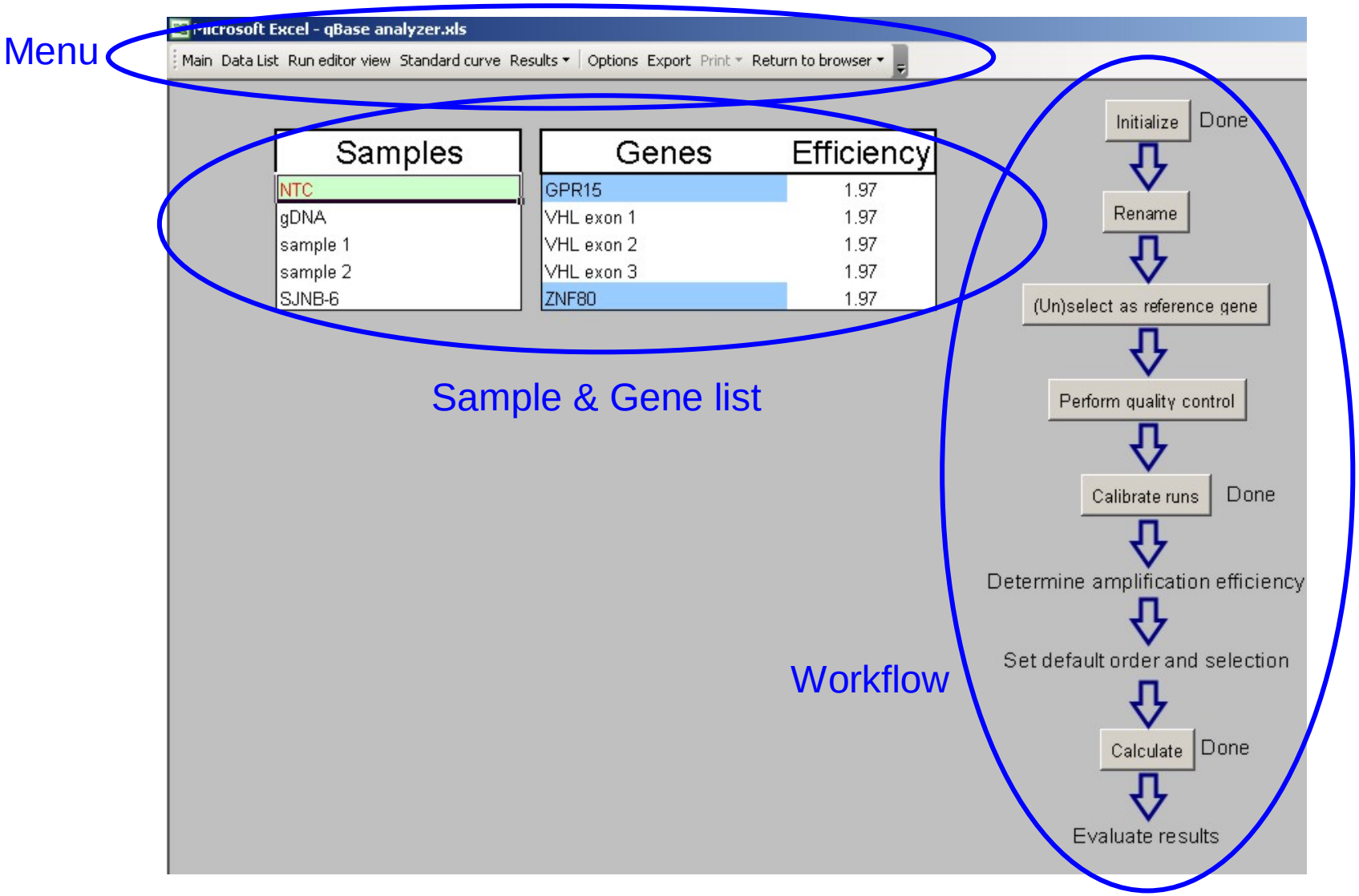
  qBase@medgen.ugent.be

XML-format

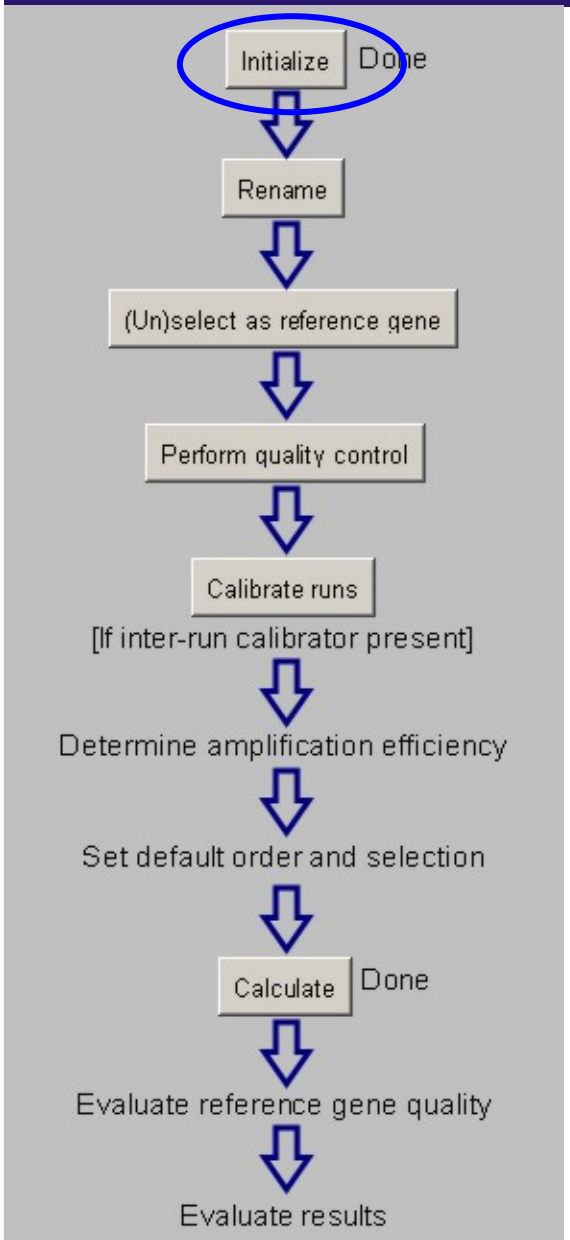
- Experiment
  - (Experiment name)
  - (Experiment description)
  - Run [1]
    - (Run name)
    - (Run Description)
    - Well
      - Well Name (or number)
      - Sample type {UNKN, NTC, NAC, STD}
      - Sample name
      - Gene name
      - Ct value
      - (Quantity)
      - Exclusion
    - …
    - Well
  - …
  - Run [n]

- Project level (under construction)
  - Statistical modules for analysis of normalized relative quantities
- Experiment level
  - Calculation of normalized relative quantities

Experiment analyzer



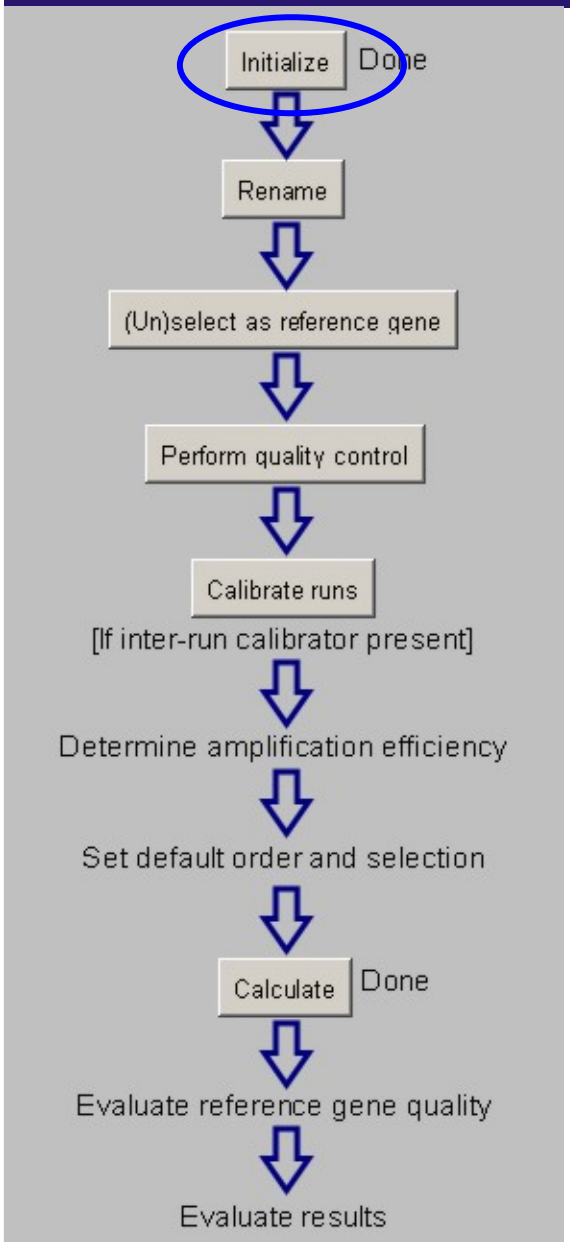
Workflow



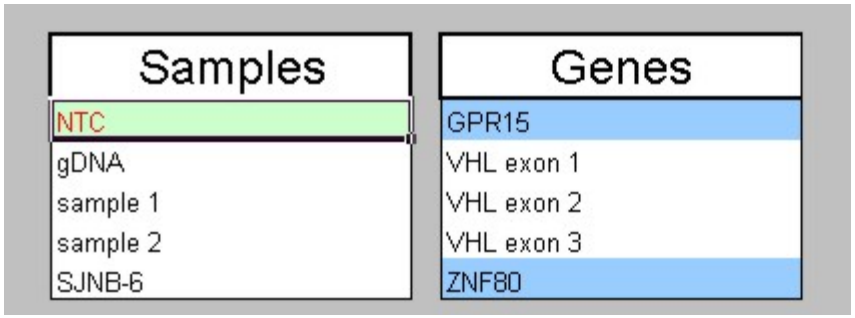
- Fetches data from source file
- Creates a
  - Data list

Run	Well	Type	Name	Gene	Ct	Quant
1	G2	UNKN	SJNB-6	VHL exon 1	29.1	
1	H1	UNKN	SJNB-6	VHL exon 1	28.8	
1	H2	UNKN	SJNB-6	VHL exon 1	28.9	
1	A3	UNKN	gDNA	VHL exon 2	23.3	
1	A4	UNKN	gDNA	VHL exon 2	23.2	
1	B3	UNKN	gDNA	VHL exon 2	23.2	
1	B4	UNKN	gDNA	VHL exon 2	23.1	
1	B11	NTC	NTC	VHL exon 2		
1	B12	NTC	NTC	VHL exon 2		

Workflow

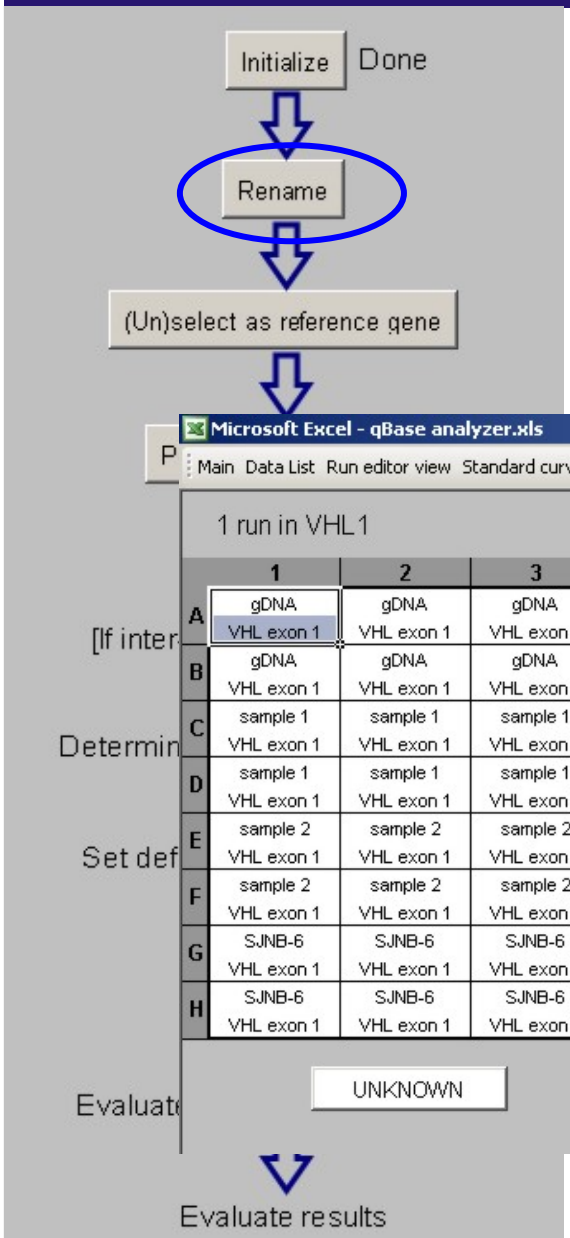


- Fetches data from source file
- Creates a
  - Data list
  - Sample list
  - Gene list



Center for Medical Genetics

Workflow



- Renames all the instances of a name (sample or gene)
- Renaming of specific wells
  ➔ Run editor

Microsoft Excel - qBase analyzer.xls

Main | Data List | Run editor view | Standard curve | Results | Options | Export | Print | Return to browser

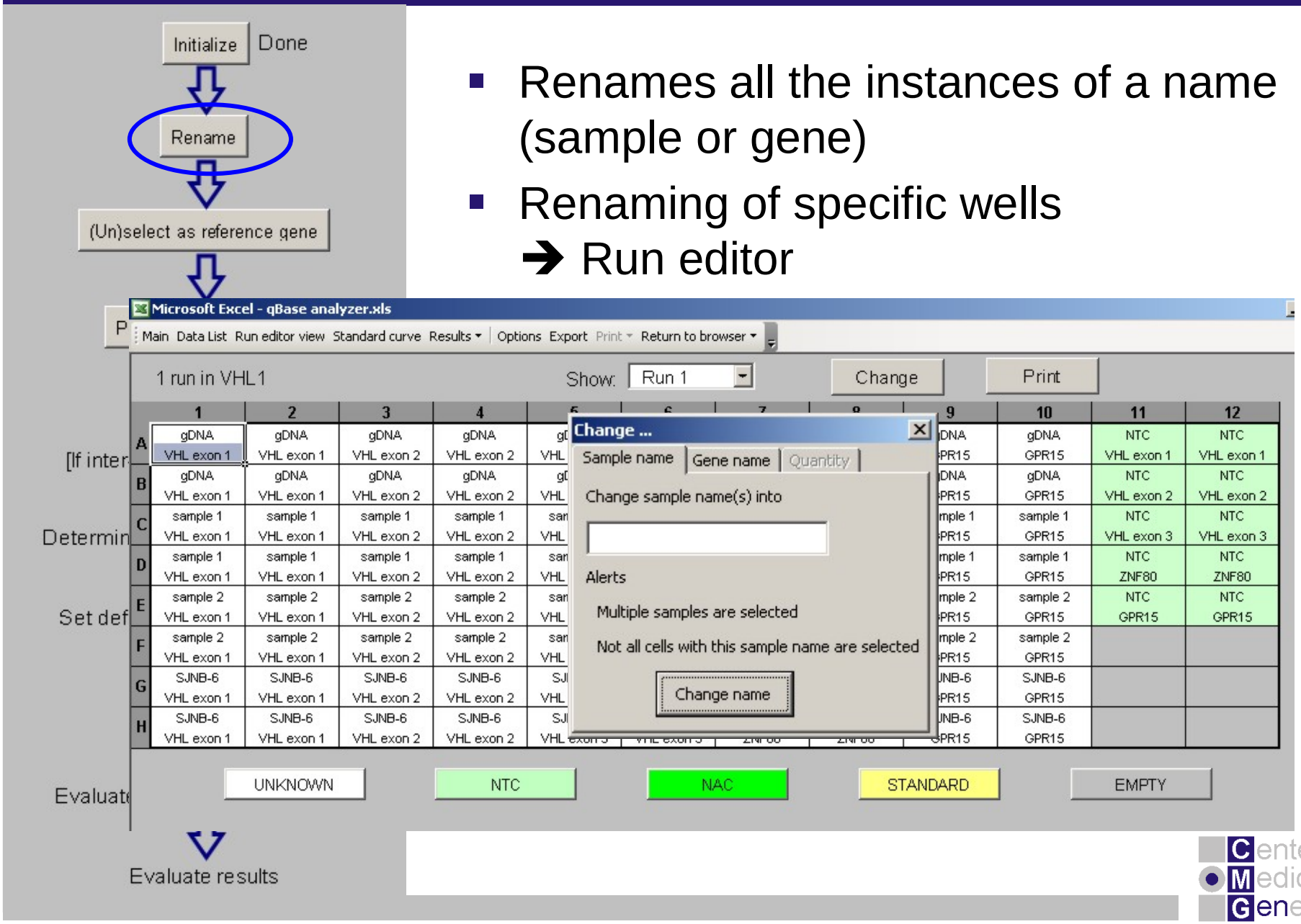
1 run in VHL1 — Show: Run 1 — Change — Print

	1	2	3	4	5	6	7	8	9	10	11	12
A	gDNA VHL exon 1	gDNA VHL exon 1	gDNA VHL exon 2	gDNA VHL exon 2	gDNA VHL exon 3	gDNA VHL exon 3	gDNA ZNF80	gDNA ZNF80	gDNA GPR15	gDNA GPR15	NTC VHL exon 1	NTC VHL exon 1
B	gDNA VHL exon 1	gDNA VHL exon 1	gDNA VHL exon 2	gDNA VHL exon 2	gDNA VHL exon 3	gDNA VHL exon 3	gDNA ZNF80	gDNA ZNF80	gDNA GPR15	gDNA GPR15	NTC VHL exon 2	NTC VHL exon 2
C	sample 1 VHL exon 1	sample 1 VHL exon 1	sample 1 VHL exon 2	sample 1 VHL exon 2	sample 1 VHL exon 3	sample 1 VHL exon 3	sample 1 ZNF80	sample 1 ZNF80	sample 1 GPR15	sample 1 GPR15	NTC VHL exon 3	NTC VHL exon 3
D	sample 1 VHL exon 1	sample 1 VHL exon 1	sample 1 VHL exon 2	sample 1 VHL exon 2	sample 1 VHL exon 3	sample 1 VHL exon 3	sample 1 ZNF80	sample 1 ZNF80	sample 1 GPR15	sample 1 GPR15	NTC ZNF80	NTC ZNF80
E	sample 2 VHL exon 1	sample 2 VHL exon 1	sample 2 VHL exon 2	sample 2 VHL exon 2	sample 2 VHL exon 3	sample 2 VHL exon 3	sample 2 ZNF80	sample 2 ZNF80	sample 2 GPR15	sample 2 GPR15	NTC GPR15	NTC GPR15
F	sample 2 VHL exon 1	sample 2 VHL exon 1	sample 2 VHL exon 2	sample 2 VHL exon 2	sample 2 VHL exon 3	sample 2 VHL exon 3	sample 2 ZNF80	sample 2 ZNF80	sample 2 GPR15	sample 2 GPR15		
G	SJNB-6 VHL exon 1	SJNB-6 VHL exon 1	SJNB-6 VHL exon 2	SJNB-6 VHL exon 2	SJNB-6 VHL exon 3	SJNB-6 VHL exon 3	SJNB-6 ZNF80	SJNB-6 ZNF80	SJNB-6 GPR15	SJNB-6 GPR15		
H	SJNB-6 VHL exon 1	SJNB-6 VHL exon 1	SJNB-6 VHL exon 2	SJNB-6 VHL exon 2	SJNB-6 VHL exon 3	SJNB-6 VHL exon 3	SJNB-6 ZNF80	SJNB-6 ZNF80	SJNB-6 GPR15	SJNB-6 GPR15		

UNKNOWN | NTC | NAC | STANDARD | EMPTY

Center for Medical Genetics

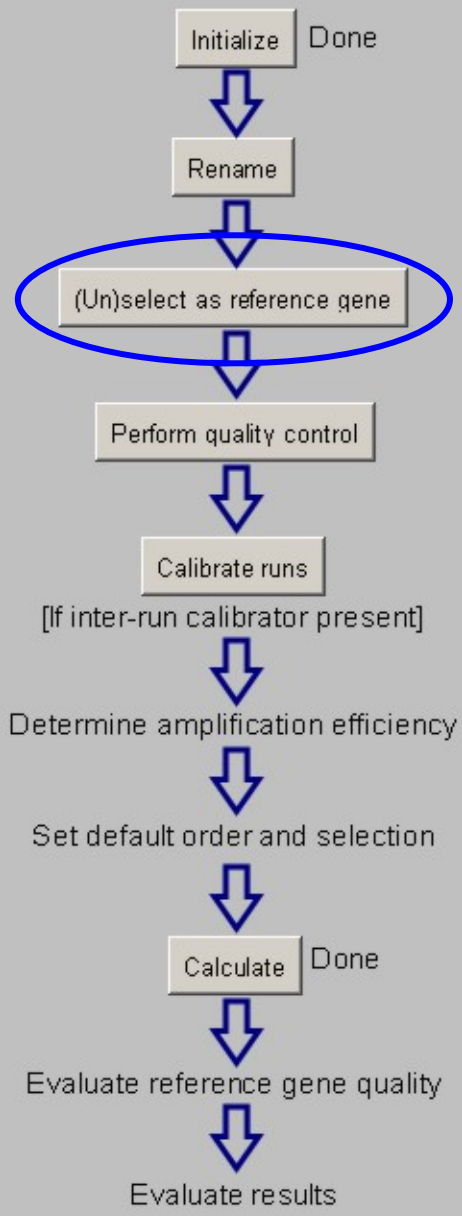
Workflow



- Renames all the instances of a name (sample or gene)
- Renaming of specific wells
  ➔ Run editor

Center for Medical Genetics

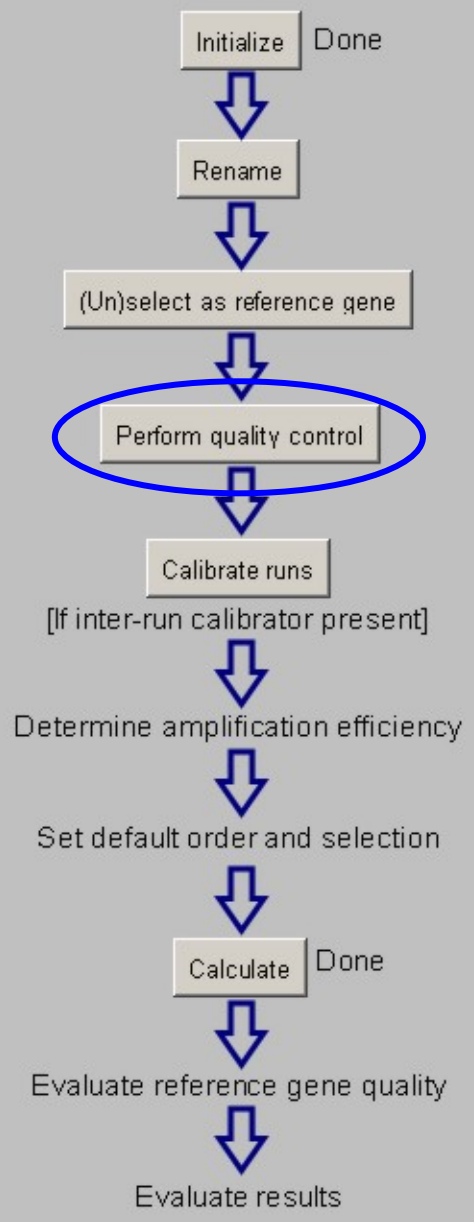
Workflow



- Select up to 5 reference genes

Center for Medical Genetics

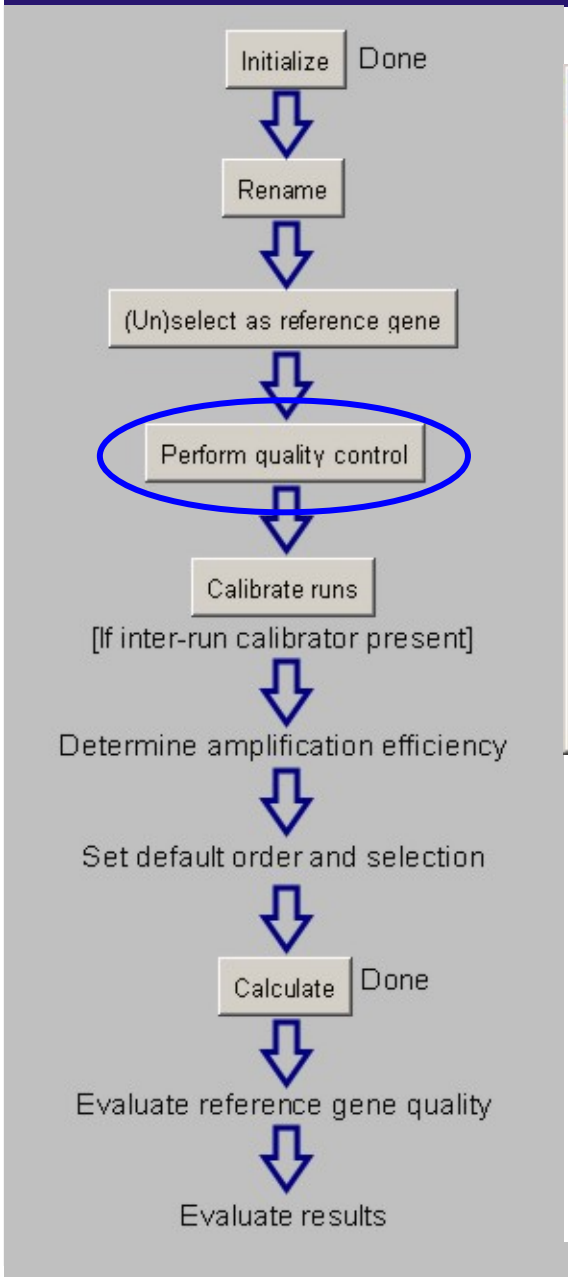
Workflow

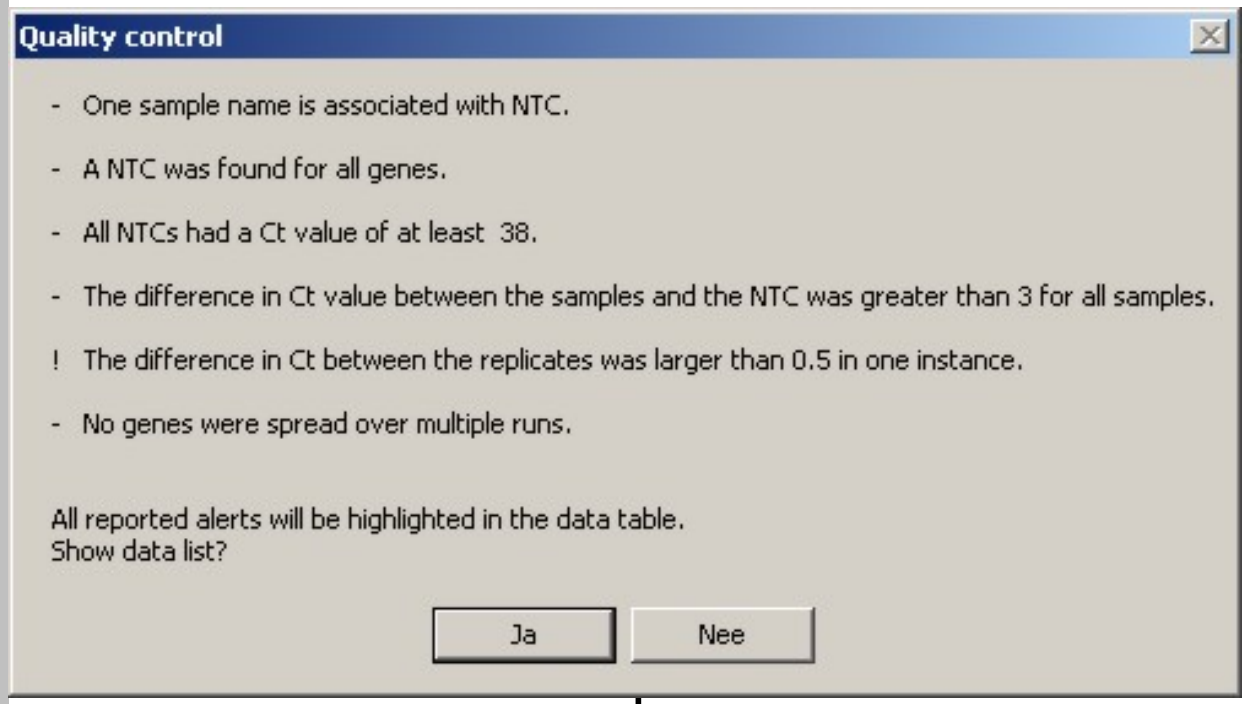


- NTC present?
- Minimal value for NTC
- Maximum difference between replicates
- Minimum difference between UNKN and NTC
- Gene spread over multiple runs
- Thresholds in ‘Options’

Center for Medical Genetics

Workflow





- Summary of quality control
- Review data in highlighted list

Center for Medical Genetics

Data list

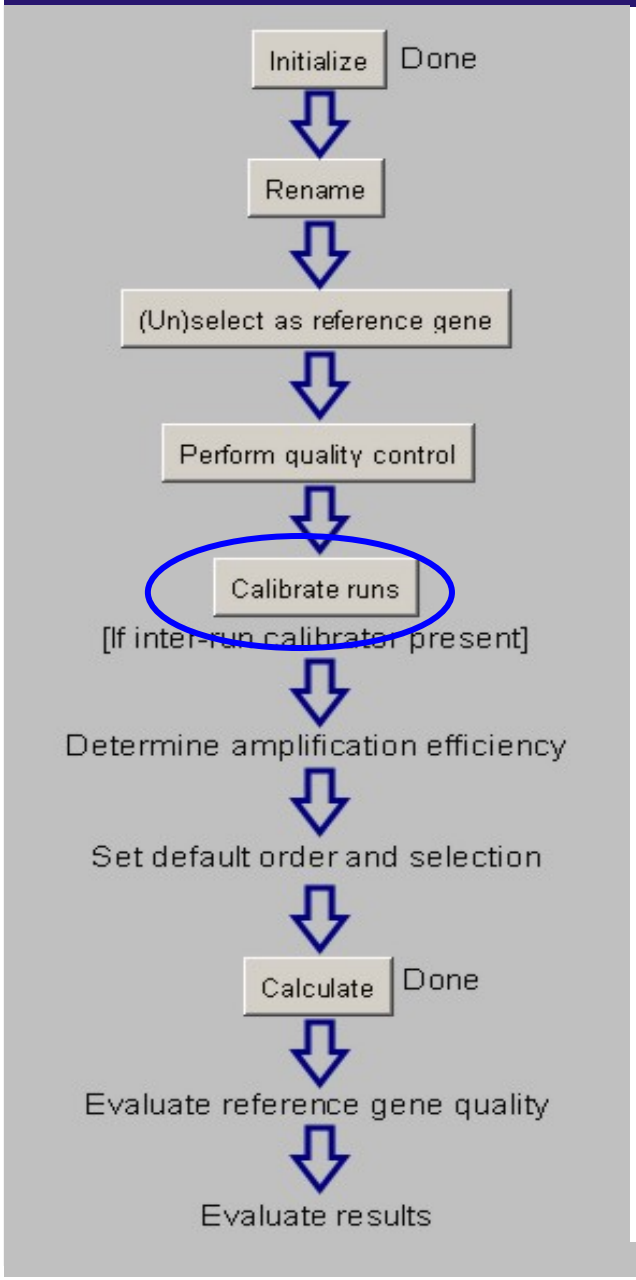
Microsoft Excel - qBase analyzer.xls

Main Data List Run editor view Standard curve Results Options Export Print Return to browser

Run	Well	Type	Name	Gene	Ct	Quant	ΔCt (NTC) test	ΔCt (replicates) test	Exclude
1	G2	UNKN	SJNB-6	VHL exon 1	29.1				
1	H1	UNKN	SJNB-6	VHL exon 1	28.8				
1	H2	UNKN	SJNB-6	VHL exon 1	28.9				
1	A3	UNKN	gDNA	VHL exon 2	23.3				
1	A4	UNKN	gDNA	VHL exon 2	23.2				
1	B3	UNKN	gDNA	VHL exon 2	23.2				
1	B4	UNKN	gDNA	VHL exon 2	23.1				
1	B11	NTC	NTC	VHL exon 2					
1	B12	NTC	NTC	VHL exon 2					
1	C3	UNKN	sample 1	VHL exon 2	25				
1	C4	UNKN	sample 1	VHL exon 2	24.8				
1	D3	UNKN	sample 1	VHL exon 2	25.3				
1	D4	UNKN	sample 1	VHL exon 2	25.2				
1	E3	UNKN	sample 2	VHL exon 2	25.3			Replicate problem	
1	E4	UNKN	sample 2	VHL exon 2	25.2			Replicate problem	
1	F3	UNKN	sample 2	VHL exon 2	24			Replicate problem	Excluded
1	F4	UNKN	sample 2	VHL exon 2	24.2			Replicate problem	Excluded
1	G3	UNKN	SJNB-6	VHL exon 2	28.9				
1	G4	UNKN	SJNB-6	VHL exon 2	29.1				
1	H3	UNKN	SJNB-6	VHL exon 2	29.1				
1	H4	UNKN	SJNB-6	VHL exon 2	29.2				
1	A5	UNKN	gDNA	VHL exon 3	23.5				
1	A6	UNKN	gDNA	VHL exon 3	23.4				
1	B5	UNKN	gDNA	VHL exon 3	23.5				
1	B6	UNKN	gDNA	VHL exon 3	23.4				
1	C11	NTC	NTC	VHL exon 3					
1	C12	NTC	NTC	VHL exon 3					
1	C5	UNKN	sample 1	VHL exon 3	24.5				

'Delete'

Workflow



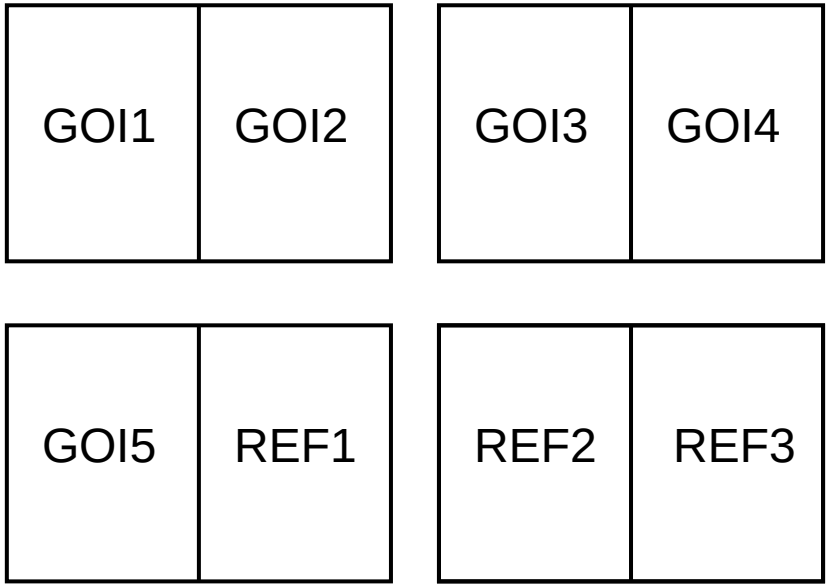
- Inter-run calibration for experiments with genes spread over multiple runs in the presence of inter-run calibrators

Center for Medical Genetics

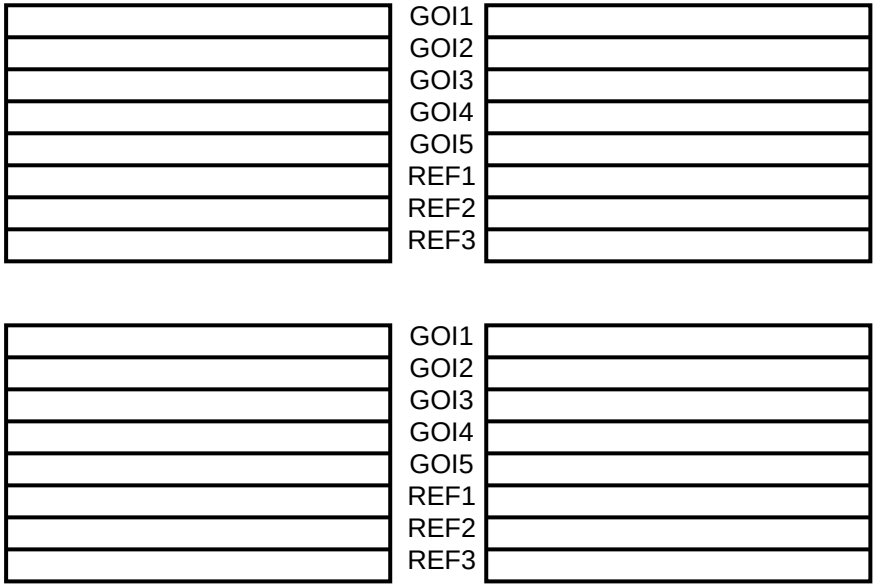
Experiment setup

[5 GOI + 3REF] X [20 samples + 1 NTC] X [2 replicates] ➔ 4 runs

Maximize number of samples ⟷ Maximize or fix number of genes



48 wells
2X (20 samples + NTC)



12 wells
2X (5 samples + NTC)

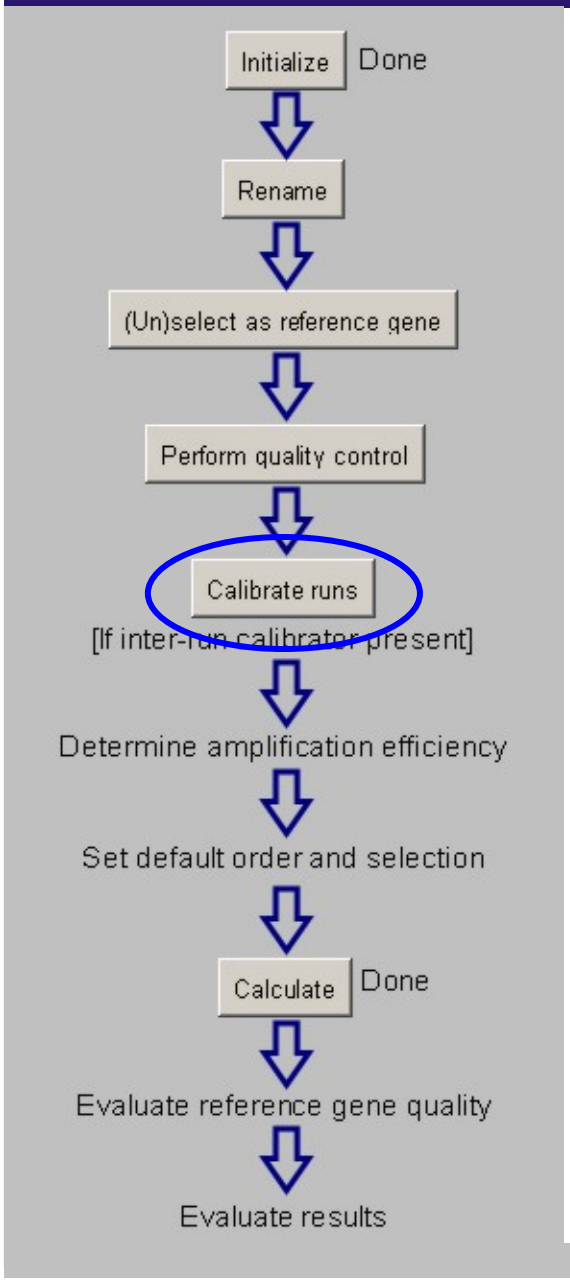
Most common use: relative quantification, comparison of gene expression levels between samples

⬇

Minimize variation NOT caused by differential expression

- Differences in cDNA starting concentration and quality
  - ➔ Normalization
- Differences in amplification conditions
  - ➔ Good equipment with minimal intra-run variation
  - ➔ Use ‘Sample maximalisation’
  - ➔ Inter-run calibration (if required)

Workflow



- Inter-run calibration for experiments with genes spread over multiple runs in the presence of inter-run calibrators
- Automated
  - Identification of inter-run calibrators
  - Calibration (gene specific)
    - Calculation of ΔCt between the runs for all inter-run calibrators
    - Calculate the average ΔCt
    - Adjust Ct values with this average ΔCt
- Quality control
  - Avoid calibration with bad calibrators
  - Allows user intervention

Center for Medical Genetics

Inter-run calibration

Microsoft Excel - qBase analyzer.xls

Main Data List Plate view Standard curve Results Options Export Print Return to browser

Calibrating run1 vs run2 for gene ctgf

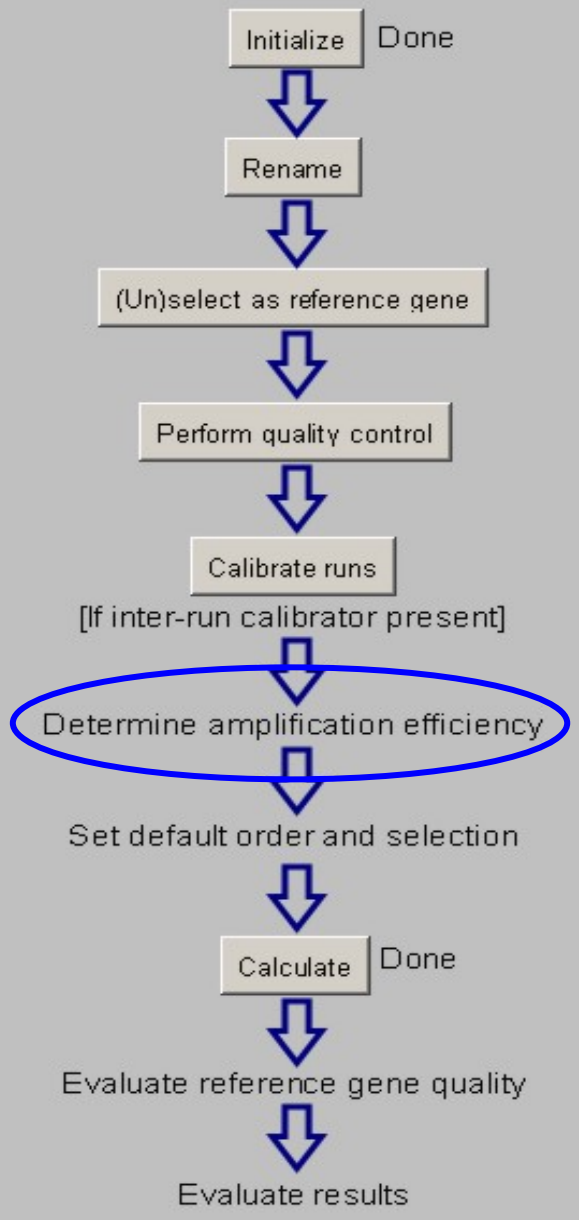
Inter-run calibrator	Ct Run1	Ct Run2	dCt
affected+TGFb	18.69	16.5	
	18.59	30	
	18.64	16.5	2.14
normal	20.04	18.036	
	19.71	17.739	
	19.875	17.8875	1.9875

Continue

average dCt --> 2.06375

ddCt --> 0.1525

Workflow



- Set default ampl. eff. for all genes
- Specify a gene specific ampl. eff.
- Determine ampl. eff. with standard curve

Center for Medical Genetics

Standard curve

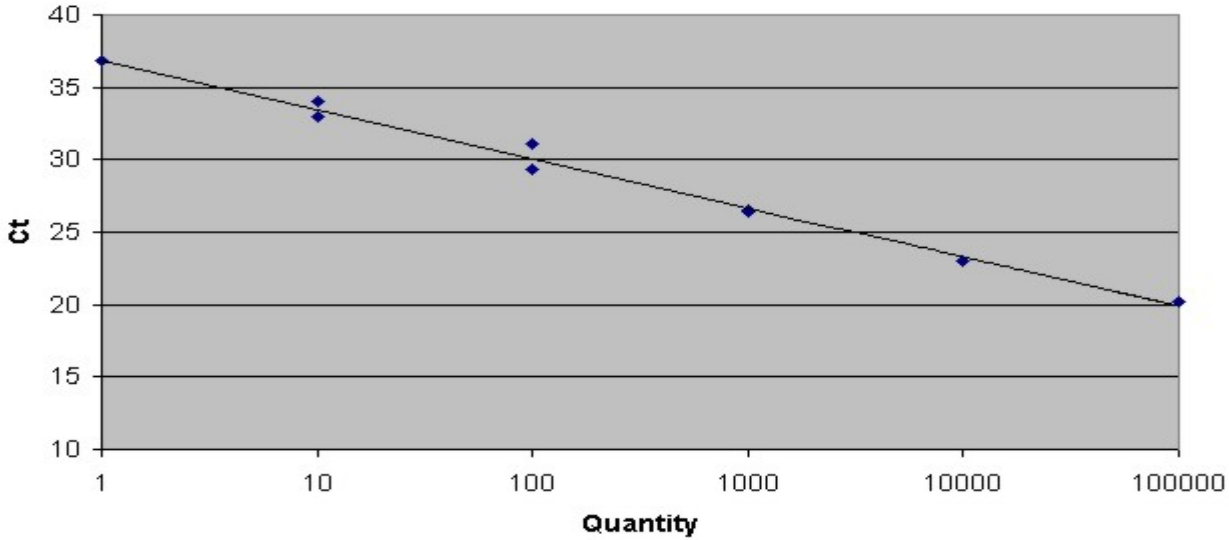
Microsoft Excel - qBase analyzer.xls

Main Data List Plate view Standard curve Results Options Export Print Return to browser

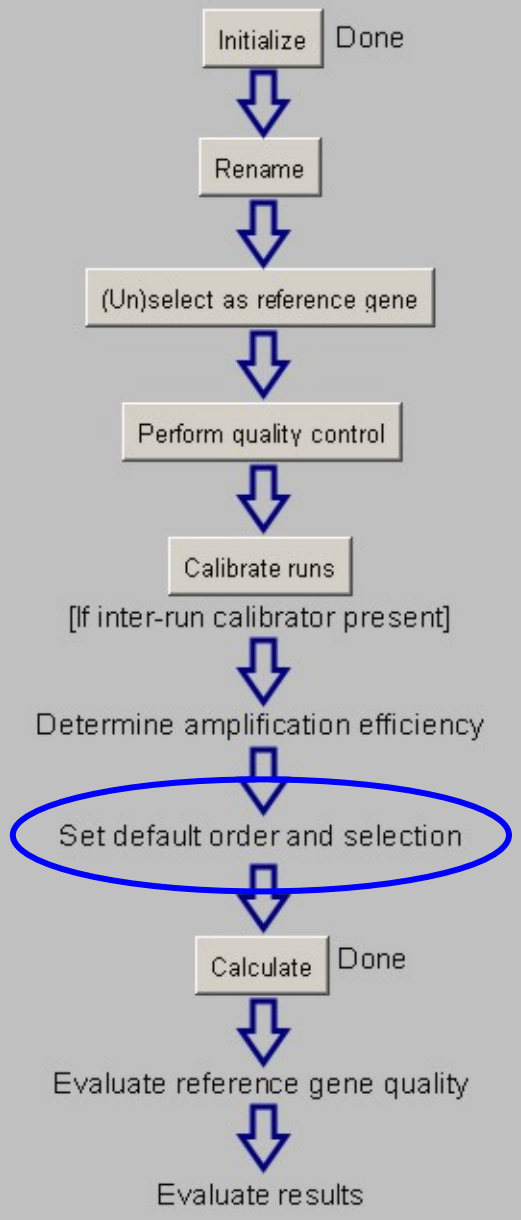
NF1skip43

Sample	Quantity	Ct
STD1	1	36.86
STD10	10	33.99
STD10	10	32.96
STD100	100	29.3
STD100	100	31.11
STD1000	1000	26.54
STD1000	1000	26.42
STD10000	10000	22.99
STD10000	10000	23
STD100000	100000	20.16
STD100000	100000	20.23

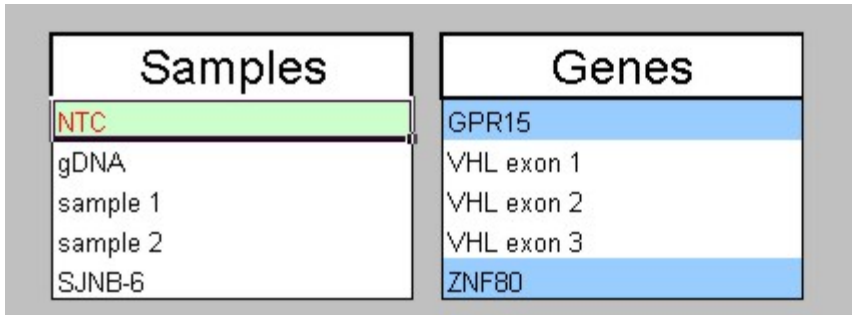
Number of points:	11	
R^2:	0.991877	
Slope:	-3.38271	
Base for exponential amplification:	1.975245	==> Use this value as gene specific amplification efficiency
Amplification efficiency:	97.5%	



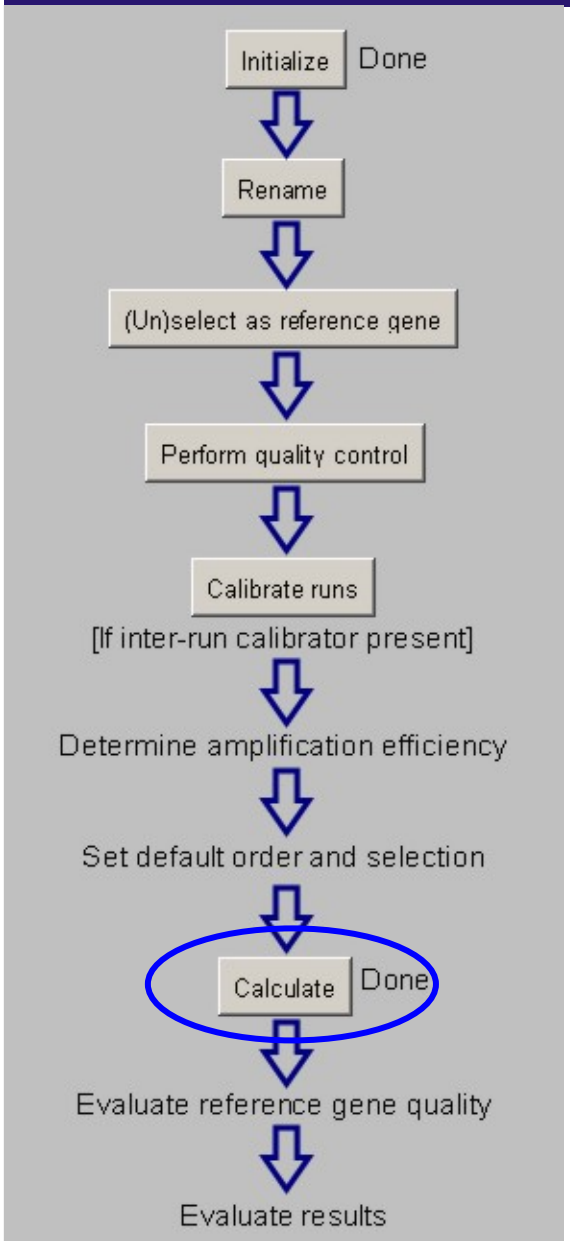
Workflow



- Use keyboard button ‘del’ to exclude samples in the results
- Use keyboard buttons ‘up’ and ‘down’ to reorder samples



Workflow

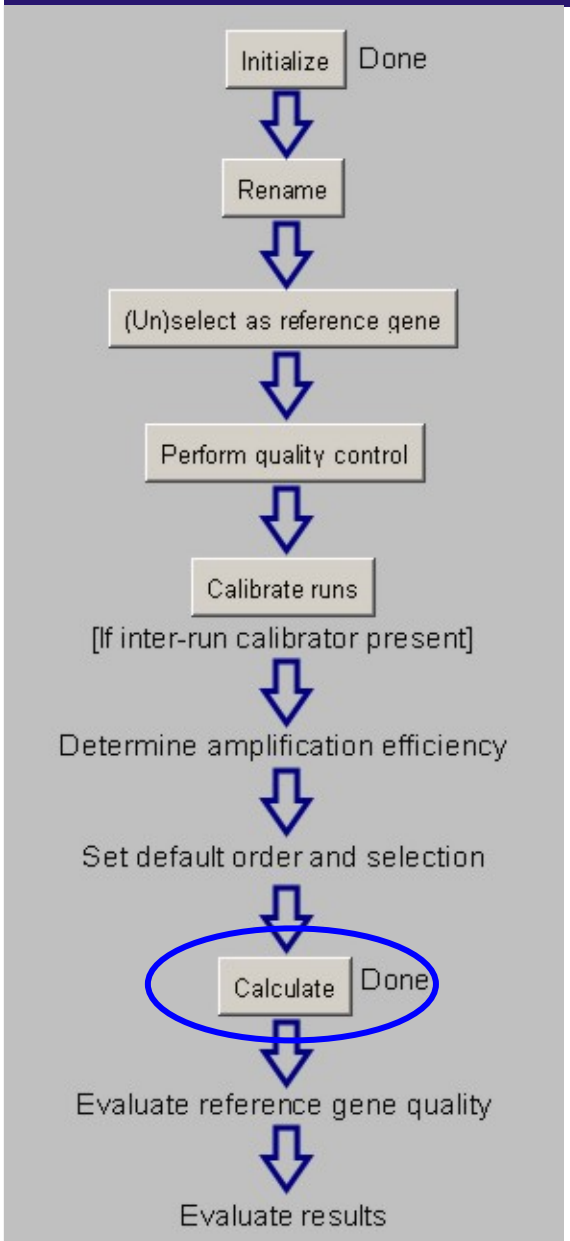


Three generations of quantification models

1. Livak and Schmittgen (2001)
   100% PCR efficiency, 1 reference gene

$$\boldsymbol{NRQ} = 2^{\Delta\Delta \boldsymbol{Ct}}$$

Workflow



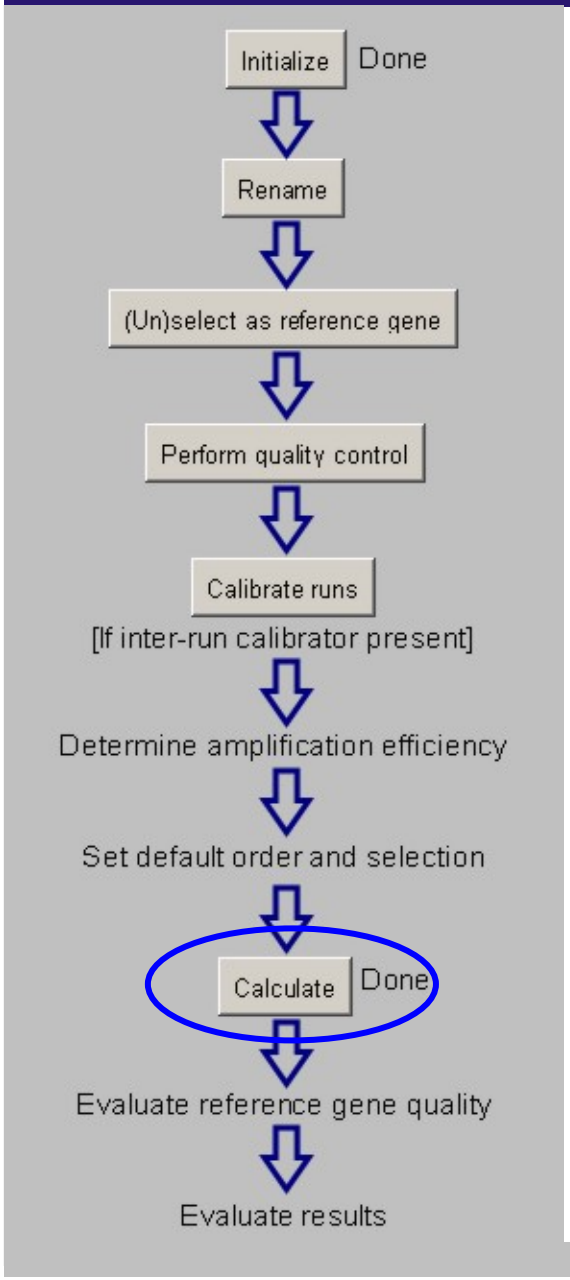
Three generations of quantification models

1. Livak and Schmittgen (2001)
   100% PCR efficiency, 1 reference gene
2. Pfaffl (2001)
   adjusted PCR efficiency, 1 ref. gene

$$NRQ = \frac{E_{goi}^{\Delta Ct,goi}}{E_{ref}^{\Delta Ct,ref}}$$

Center for Medical Genetics

Workflow



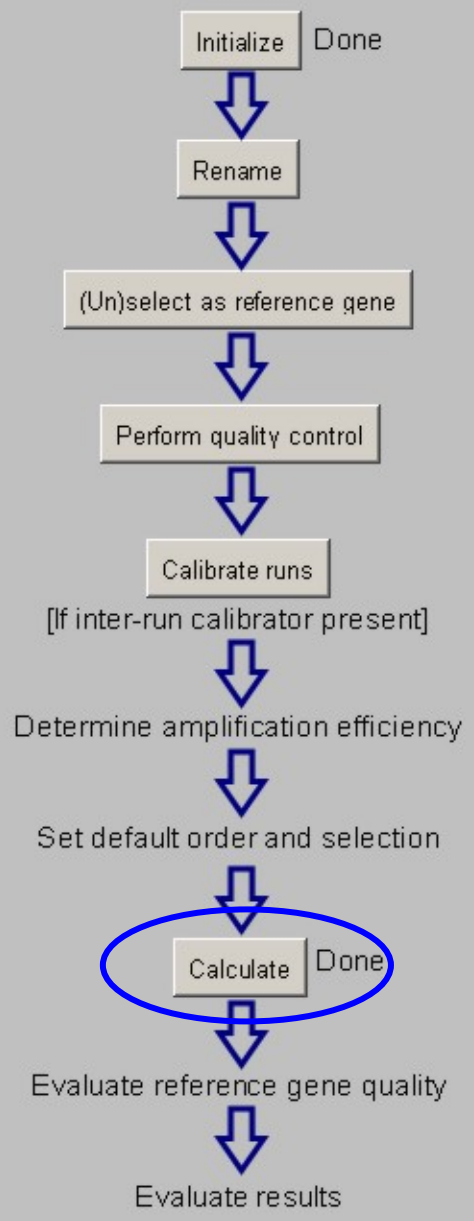
Three generations of quantification models

1. Livak and Schmittgen (2001)
   100% PCR efficiency, 1 reference gene
2. Pfaffl (2001)
   adjusted PCR efficiency, 1 ref. gene
3. Unpublished new model
   adjusted PCR eff. & multiple ref. genes

$$NRQ = \frac{E_{goi}^{\Delta Ct,goi}}{\sqrt[n]{\prod_{i}^{n} E_{ref_i}^{\Delta Ct,ref_i}}}$$

Center for Medical Genetics

Workflow

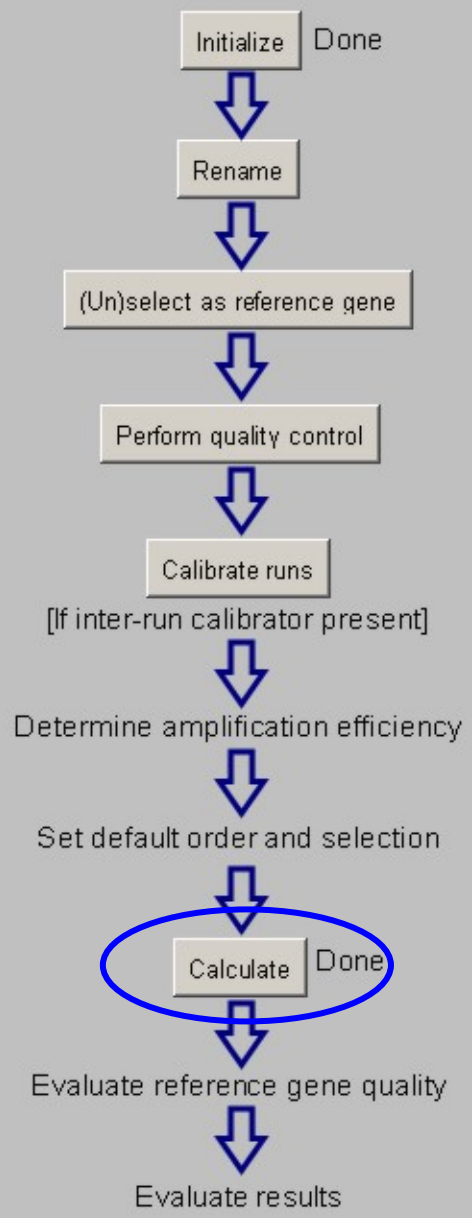


- Average Ct values of replicates

$$\overline{Ct}_{jk} = \frac{\sum_{i=1}^{n} Ct_{ijk}}{n}$$

$$SD(\overline{Ct}_{jk}) = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(Ct_{ijk} - \overline{Ct}_{jk}\right)^2}$$

Workflow



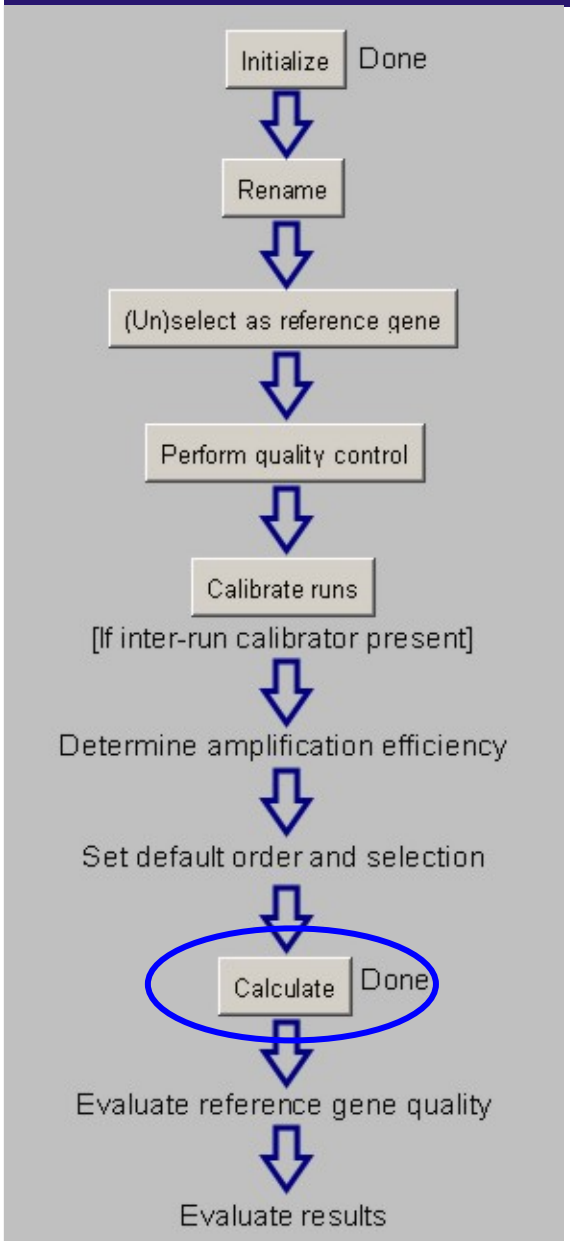
- Average Ct values of replicates
- Convert Ct values into relative quantities

$$RQ_{jk} = E_j^{(\overline{Ct}_{control,j} - \overline{Ct}_{jk})}$$

$$SD(RQ_{jk}) = RQ_{jk}.\ln(E_j).SD(\overline{Ct}_{jk})$$

Center for Medical Genetics

Workflow



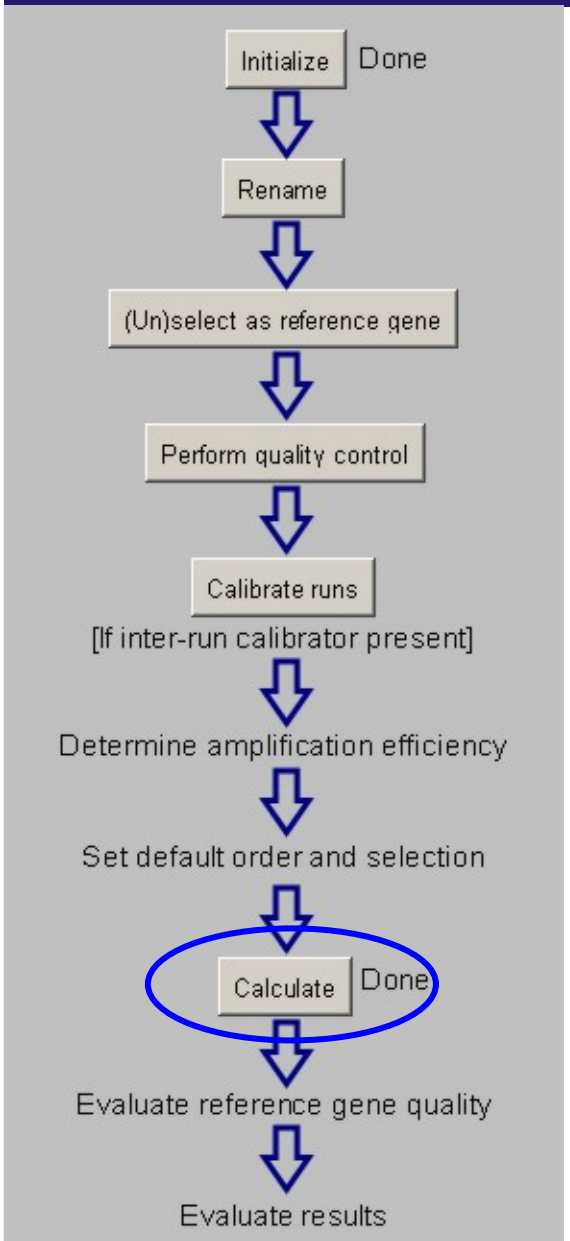
- Average Ct values of replicates
- Convert Ct values into relative quantities
- Calculate the normalization factor

$$NF_k = \sqrt[f]{\prod_{o=1}^{f} RQ_{ref,ok}}$$

$$SD(NF_k) = NF_k \sqrt{\sum_{o=1}^{f} \left(\frac{SD(RQ_{ref,ok})}{f.RQ_{ref,ok}} \right)^2}$$

Center for Medical Genetics

Workflow



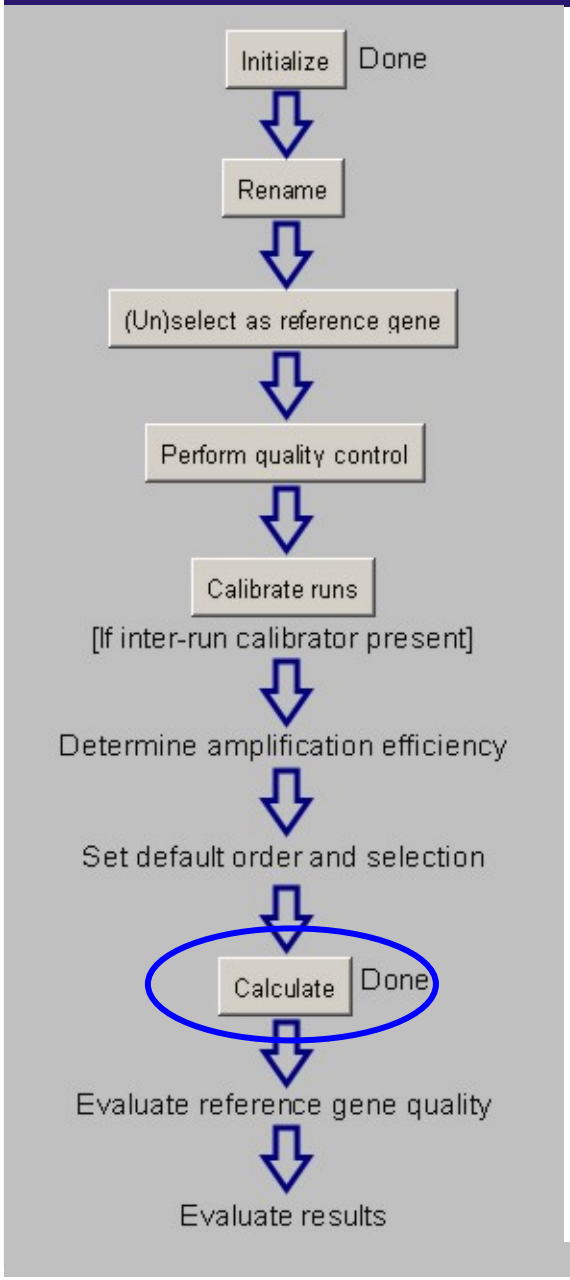
- Average Ct values of replicates
- Convert Ct values into relative quantities
- Calculate the normalization factor
- Perform normalization

$$NRQ_{jk} = \frac{RQ_{jk}}{NF_k}$$

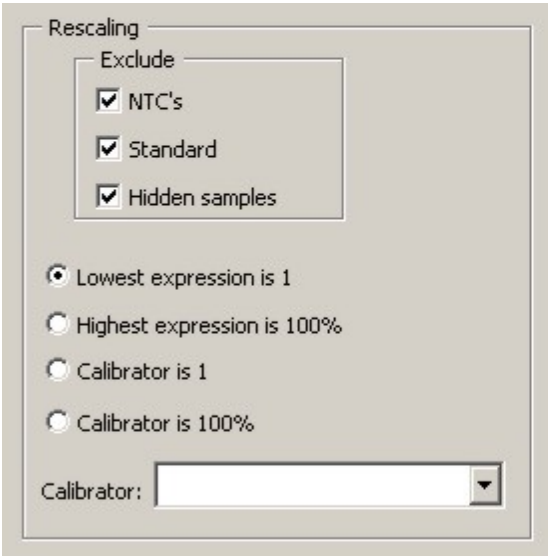
$$SD(NRQ_{jk}) = NRQ_{jk}\sqrt{\left(\frac{SD(NF_k)}{NF_k}\right)^2 + \left(\frac{SD(RQ_{jk})}{RQ_{jk}}\right)^2}$$

Center for Medical Genetics

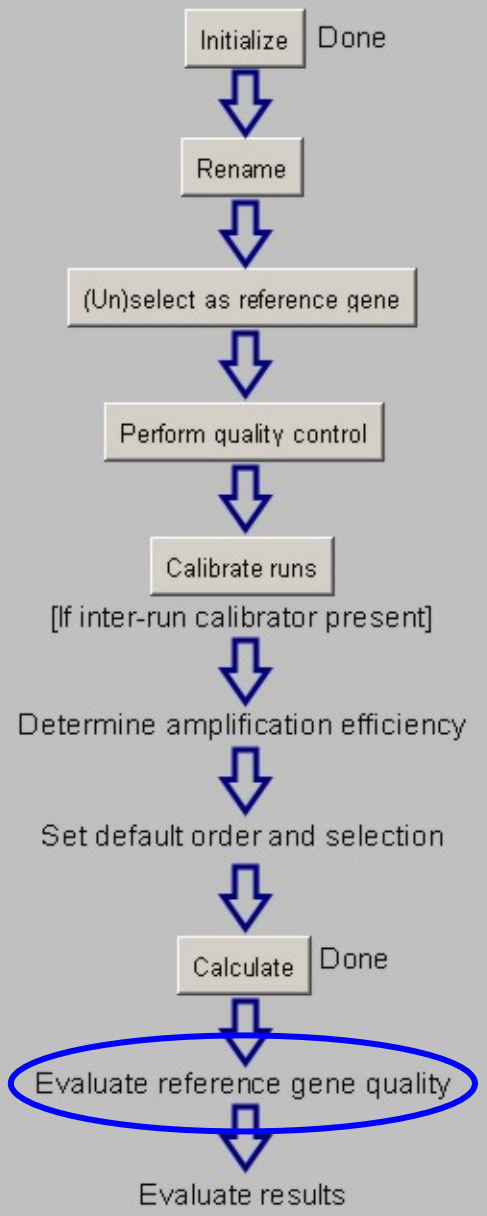
Workflow



- Average Ct values of replicates
- Convert Ct values into relative quantities
- Calculate the normalization factor
- Perform normalization
- Rescale data (Options)



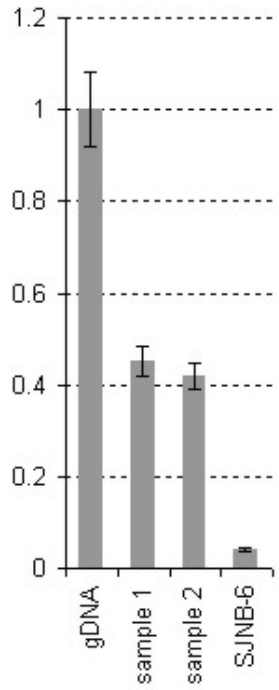
Workflow



- Only applicable with >1 reference gene
- Normalization factor histogram (abnormalities?)

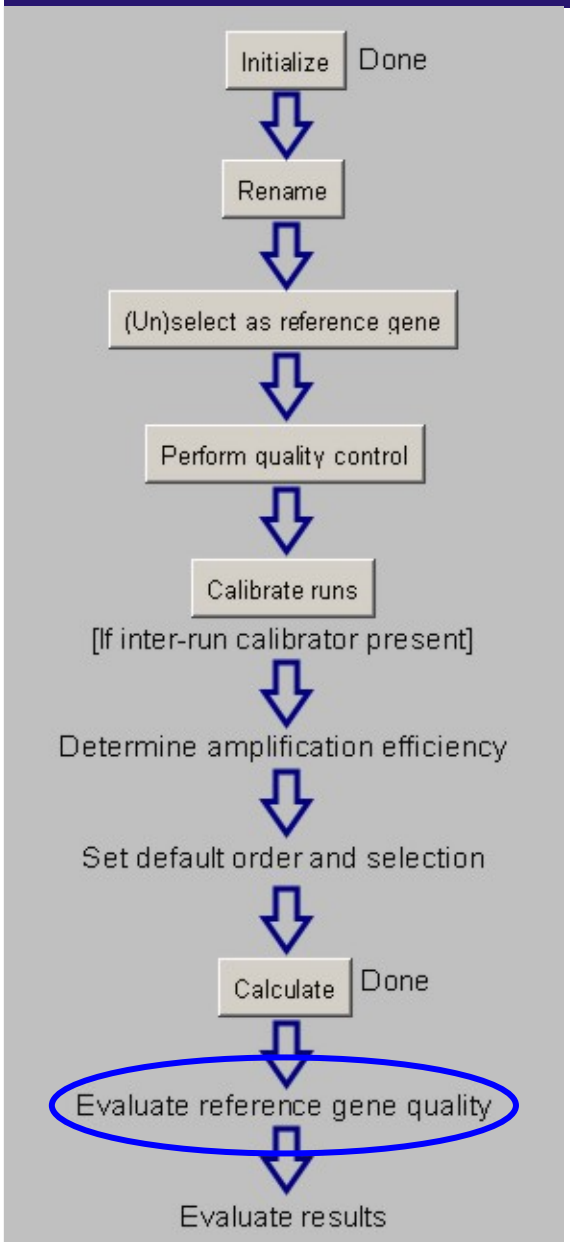
NF histogram

Sample	NF	StdDev
gDNA	1	0.080037742
sample 1	0.450625231	0.033129679
sample 2	0.420448208	0.028529638
SJNB-6	0.038808057	0.004090611

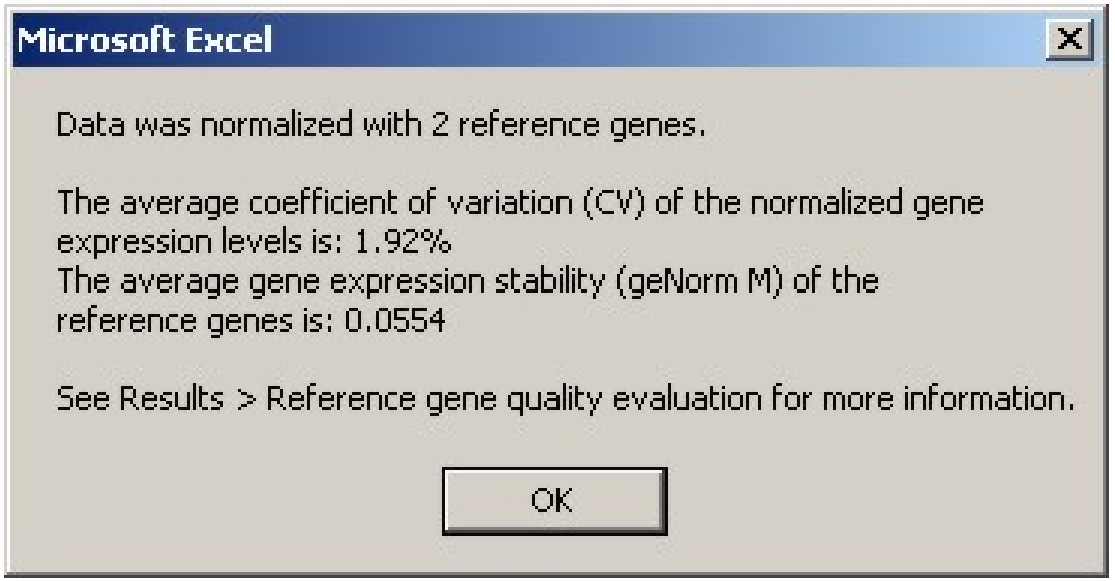


Center for Medical Genetics

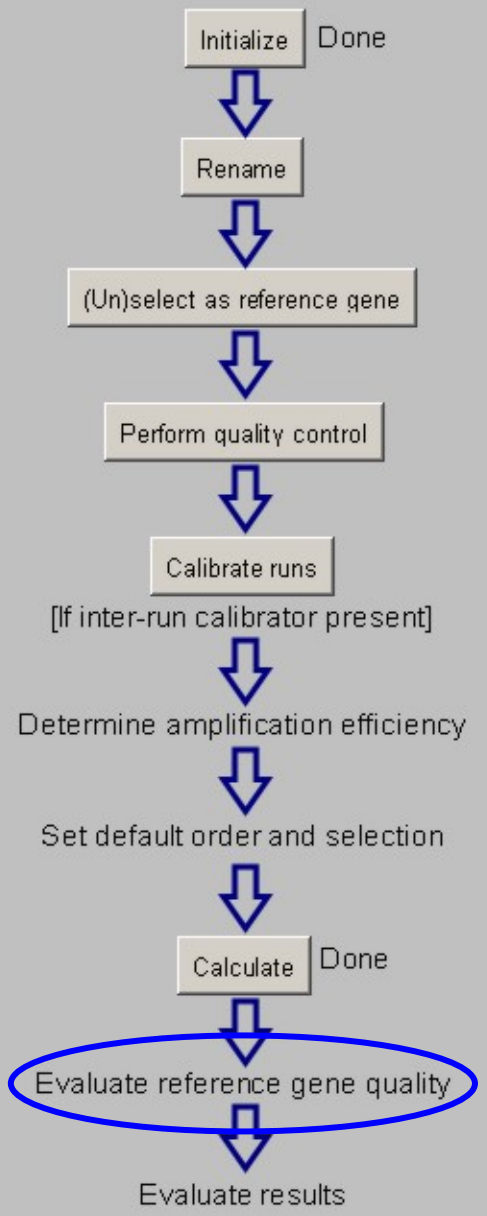
Workflow



- Only applicable with >1 reference gene
- Normalization factor histogram (abnormalities?)
- CV & M



Center for Medical Genetics

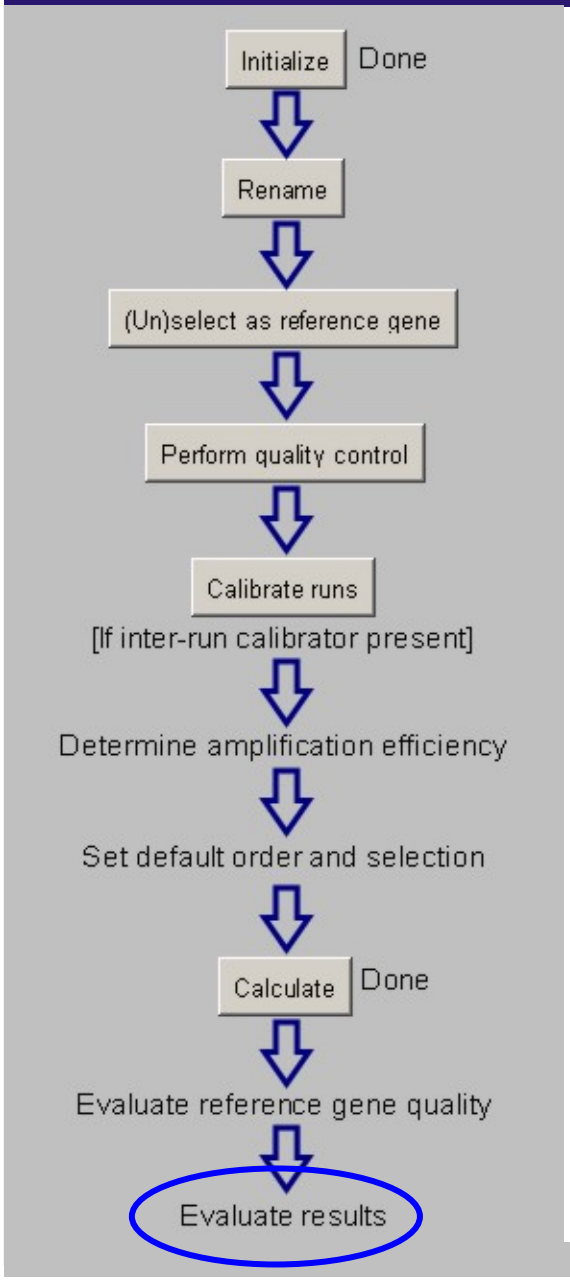


- Only applicable with >1 reference gene
- Normalization factor histogram (abnormalities?)
- CV & M

Reference gene quality evaluation

	CV	M (geNorm)
GPR15	1.92%	0.0554
ZNF80	1.93%	0.0554
Total	1.92%	0.0554

Workflow



- Table

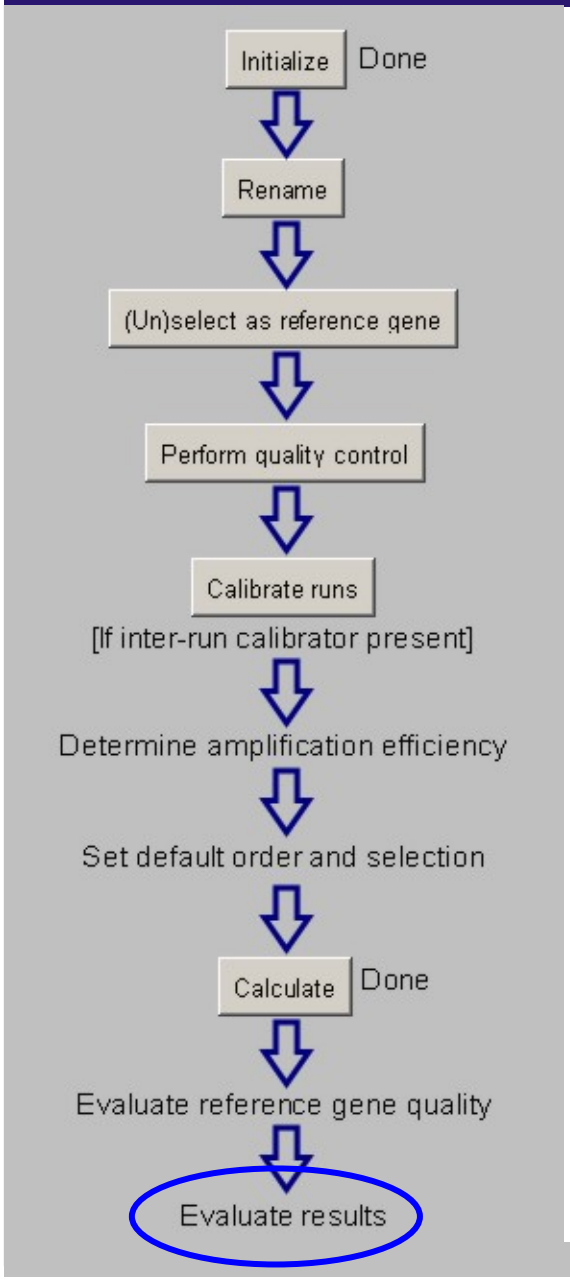
Microsoft Excel - qBase analyzer.xls

Main Data List Plate view Standard curve Results Options Export Print Return to browser

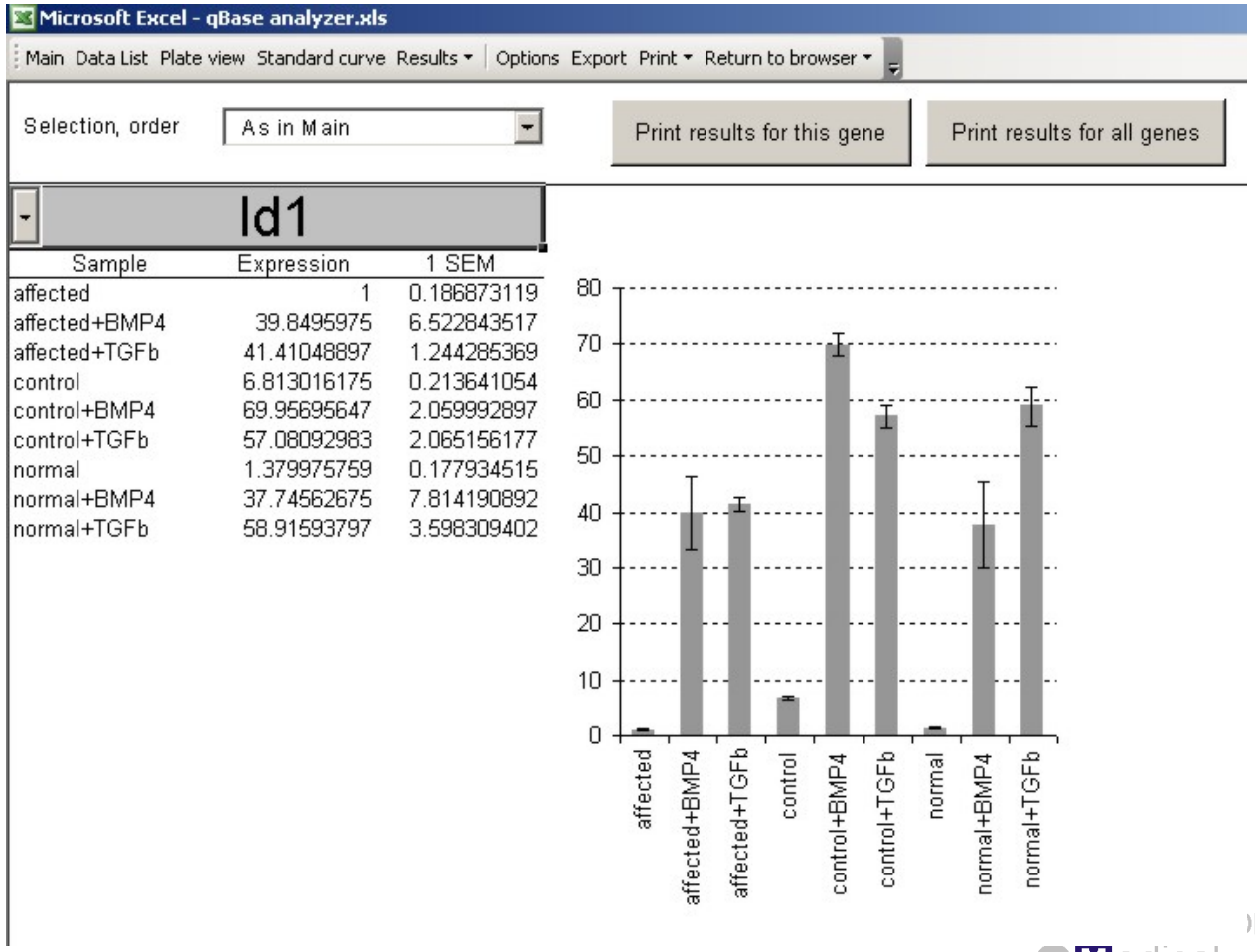
Sample/Gene	GPR15	1 SEM	VHL exon	1 SEM	VHL exon	1 SEM	VHL exon	1 SEM
gDNA	1	0.074868	1	0.065605	1	0.049013	1	0.044742
NTC	1.035265	0	1.898684	0	1.802501	0	2.143547	0
sample 1	1.01748	0.081754	0.949342	0.04404	0.604997	0.051538	1.053361	0.079215
sample 2	0.982821	0.061011	0.458502	0.025746	0.574349	0.027856	1.071773	0.070722
SJNB-6	1.026334	0.091698	0.513167	0.035104	0.439063	0.030035	0.513167	0.033959

Center for Medical Genetics

Workflow

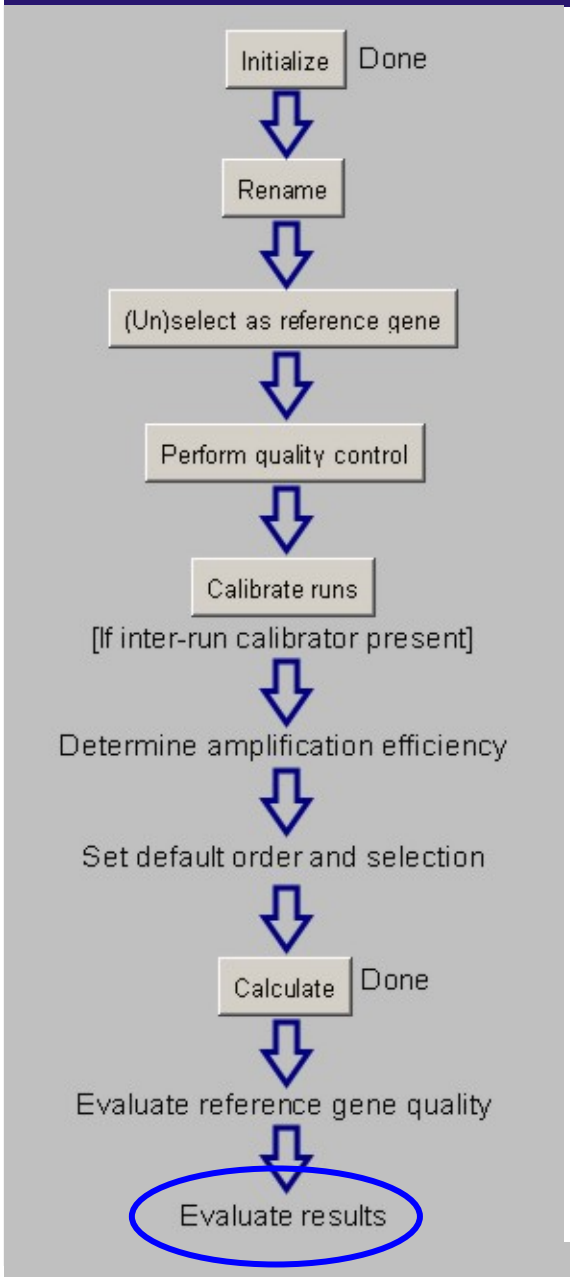


- Table
- Single gene histogram

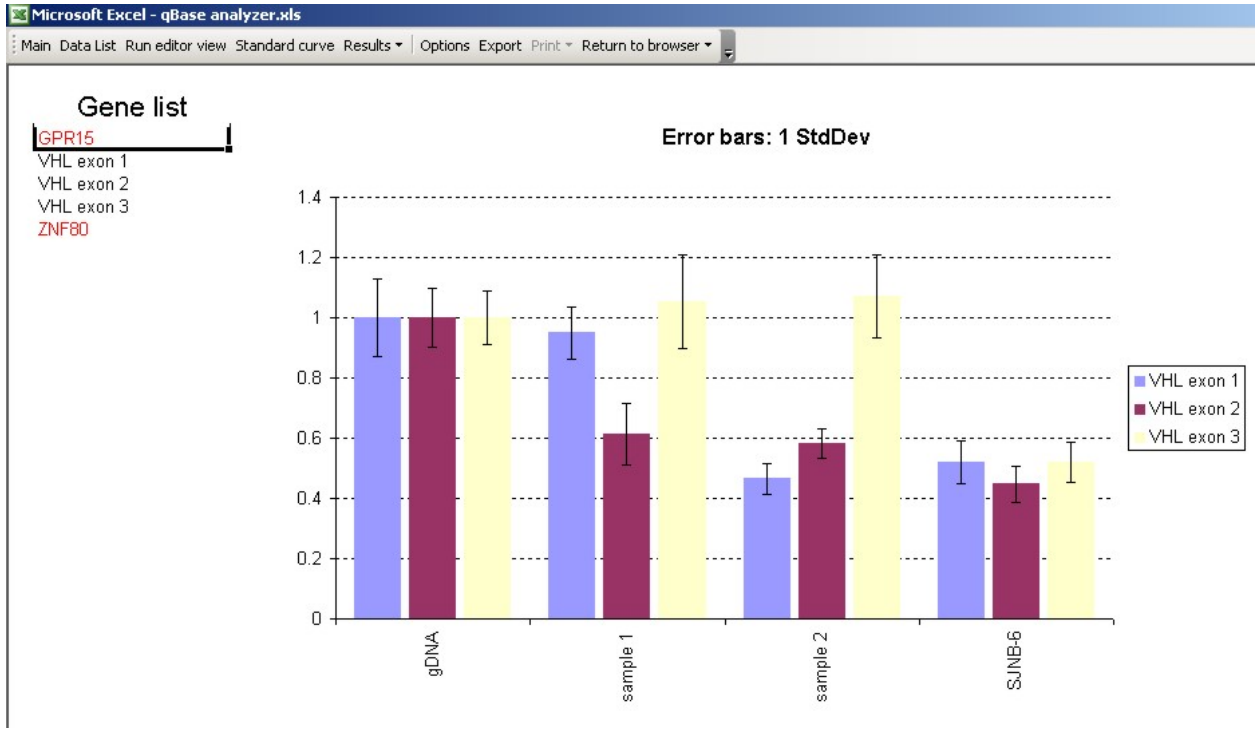


Sample	Expression	1 SEM
affected	1	0.186873119
affected+BMP4	39.8495975	6.522843517
affected+TGFb	41.41048897	1.244285369
control	6.813016175	0.213641054
control+BMP4	69.95695647	2.059992897
control+TGFb	57.08092983	2.065156177
normal	1.379975759	0.177934515
normal+BMP4	37.74562675	7.814190892
normal+TGFb	58.91593797	3.598309402

Workflow

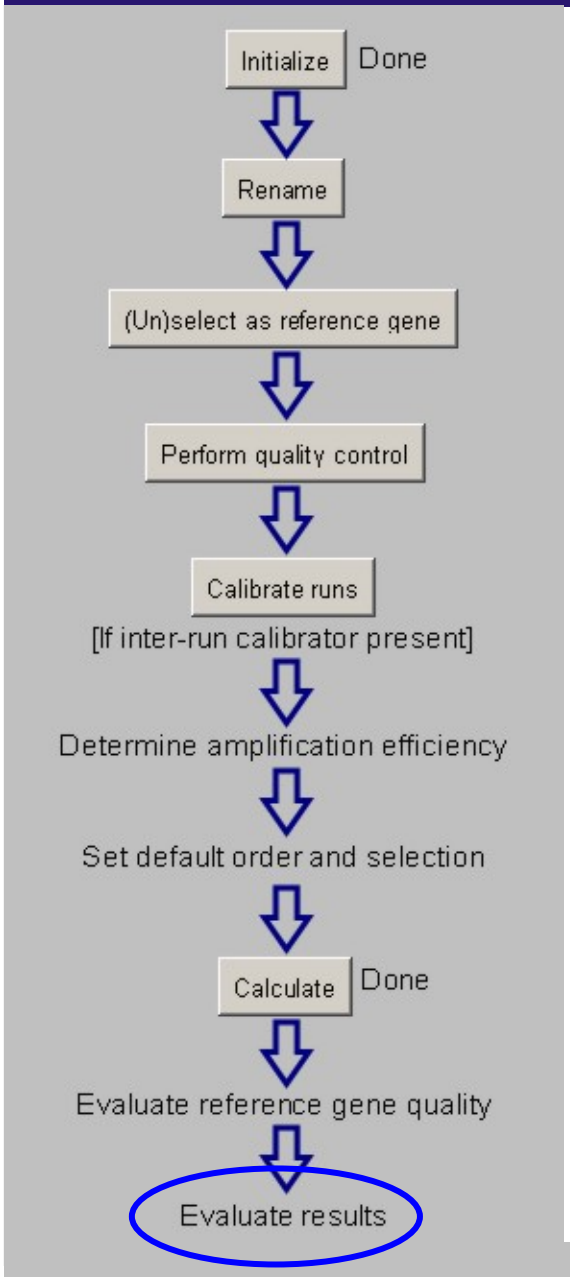


- Table
- Single gene histogram
- Multi gene histogram



Center for Medical Genetics

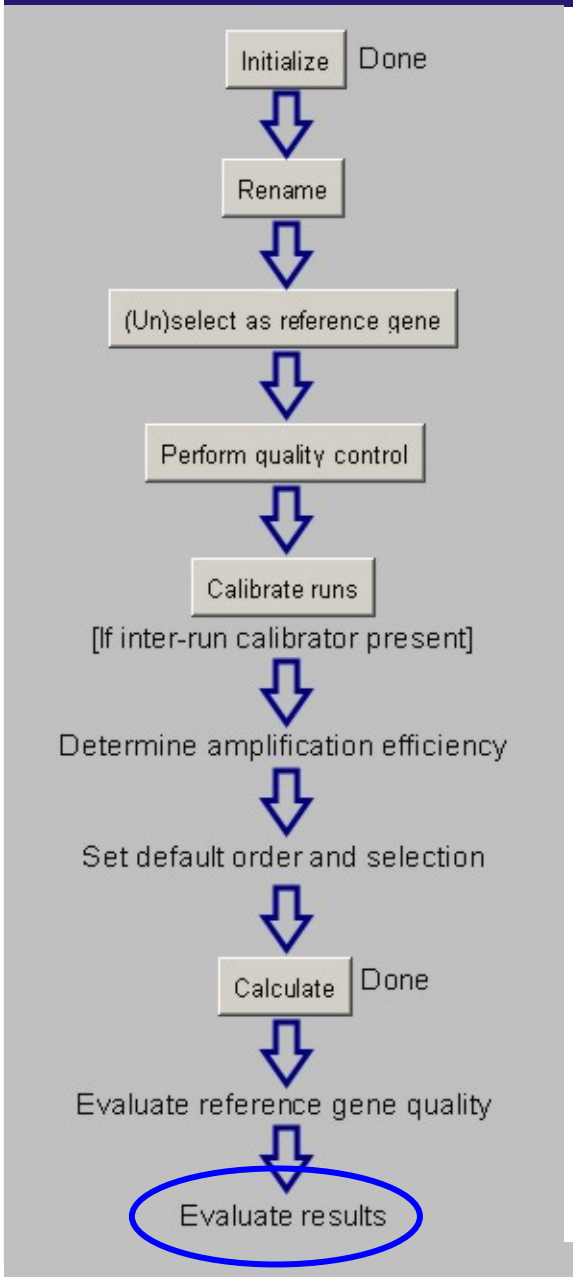
Workflow

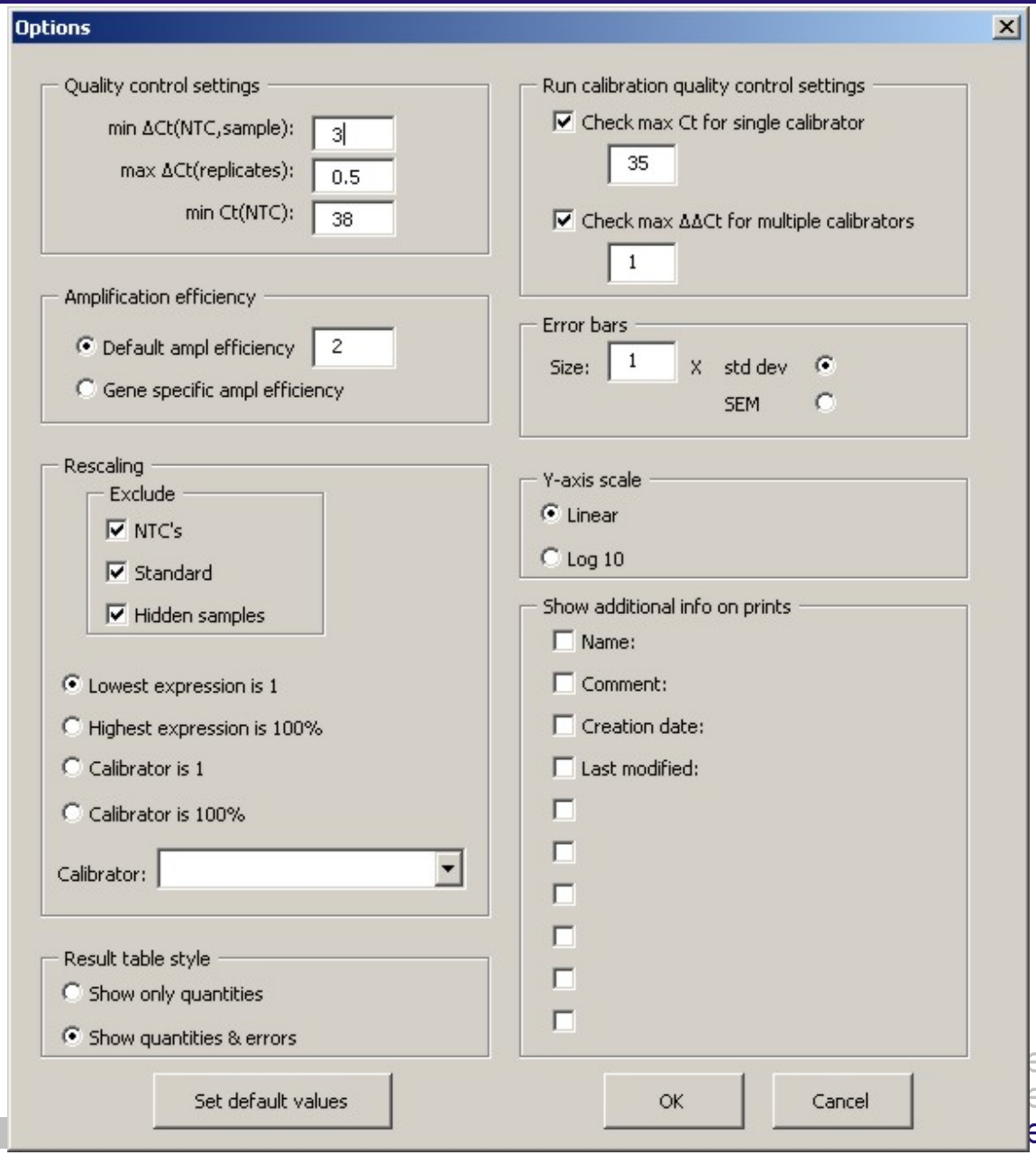


- Table
- Single gene histogram
- Multi gene histogram
- Options
- Export
  - Results (NRQ + SD or SEM)
  - geNorm input file

Center for Medical Genetics

Workflow





enter for
edical
enetics

Summary

- qBase Browser
  - Import raw data in different formats (proposal for a universal XML format)
  - Manage, organize and annotate data
- qBase Analyzer, experiment level
  - Quality control of raw data
  - Inter-run calibration
  - Correct conversion of Ct values into relative quantities using efficiencey correction
  - Normalization with up to 5 reference genes
  - Reference gene quality evaluation
  - User friendly results viewing
- qBase Analyzer, project level (under construction)
  - Statistical processing of normalized relative quantities
- Easy data exchange
- Free – open source

Acknowledgement

- Geert Mortier
- Frank Speleman
- Anne De Paepe
- Jo Vandesompele

- Colleagues at CMGG for testing and evaluation
- All users for suggestions and bug reports

qBase@medgen.ugent.be
http://medgen.ugent.be/qbase

Center for Medical Genetics