qPCR course

Alyona Minina
Department of Plant Biology
SLU

2014

What do you want to know?

Data Analysis and interpretation

Application of qPCR

Statistical analysis of qPCR

Basics

Lecture 1

How qPCR works

Normalization

Troubleshooting

Everything

More

Results of the quiz

1. Did you learn anything new from the lecture

A lot	yes	a bit
2	10	4

2. If yes, do you think you will remember it tomorrow?

no, unless I refresh it	I hope	yes
2	5	9

3. Would you prefer to have more information in the lecture, keep it the same, reduce it?

Keep the same	More information
15	1

concerning the prints:
how bad do you want them?

4. Were the tasks too easy, ok, or too hard?

too hard	ok	too easy
3	11	2

comment: the tasks were diffuse

8. Would you suggest any changes?

it was a bit too fast

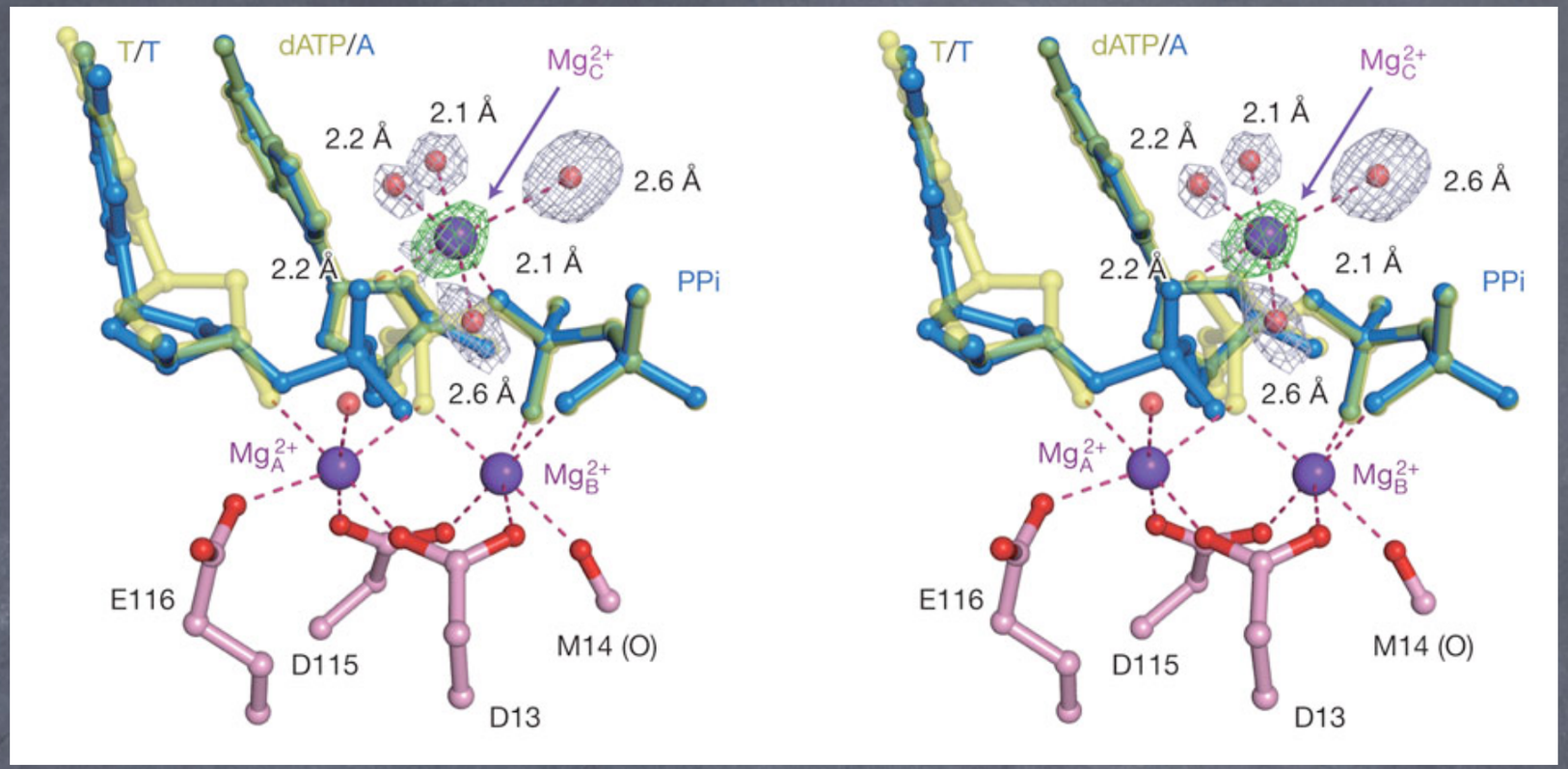
the course website info

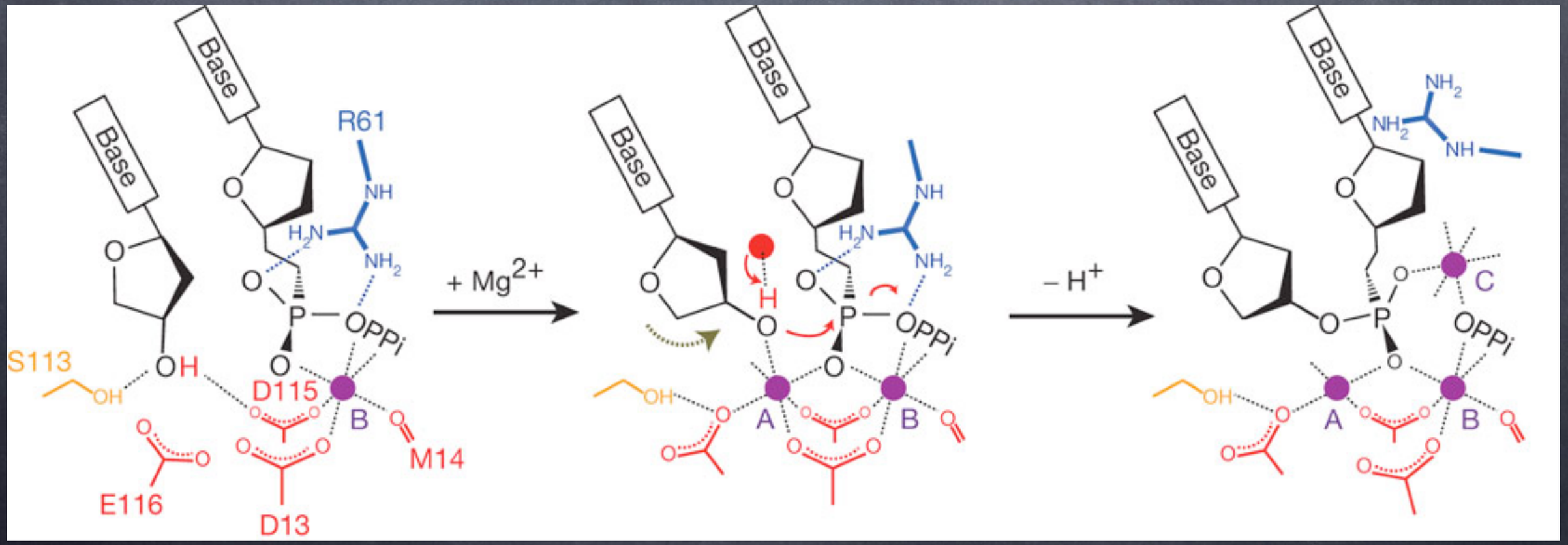
warm up questions from Lecture 1

Intro with definition of technique and application

DNA replication should be summarised on one slide

Mg2+ in the active center of pol





Nakamura et al., 2012. Nature

from the Lecture 1

what is DNA made of?

is it water soluble?

will it dissolve in 70% Ethanol?

what are 3′ and 5′ ends?

how does DNA pol add nucleotides to the growing strand?

what is 3 step PCR?

Reverse transcription?

RT-qPCR

Real time

quantitative

end point PCR vs real time PCR

End point PCR

mix PCR

↓

run PCR until the very last cycle

↓

mix with fluorescent dye (EtBr/GelRed etc)

↓

use UV lamp and your eyes to see fluorescence on a gel

Real time PCR

mix PCR together with fluorescent dye

↓

run PCR and make detection after EACH cycle

what is fluorescence?

"glowing under UV light", George Stokes 19th century

what is fluorophore?

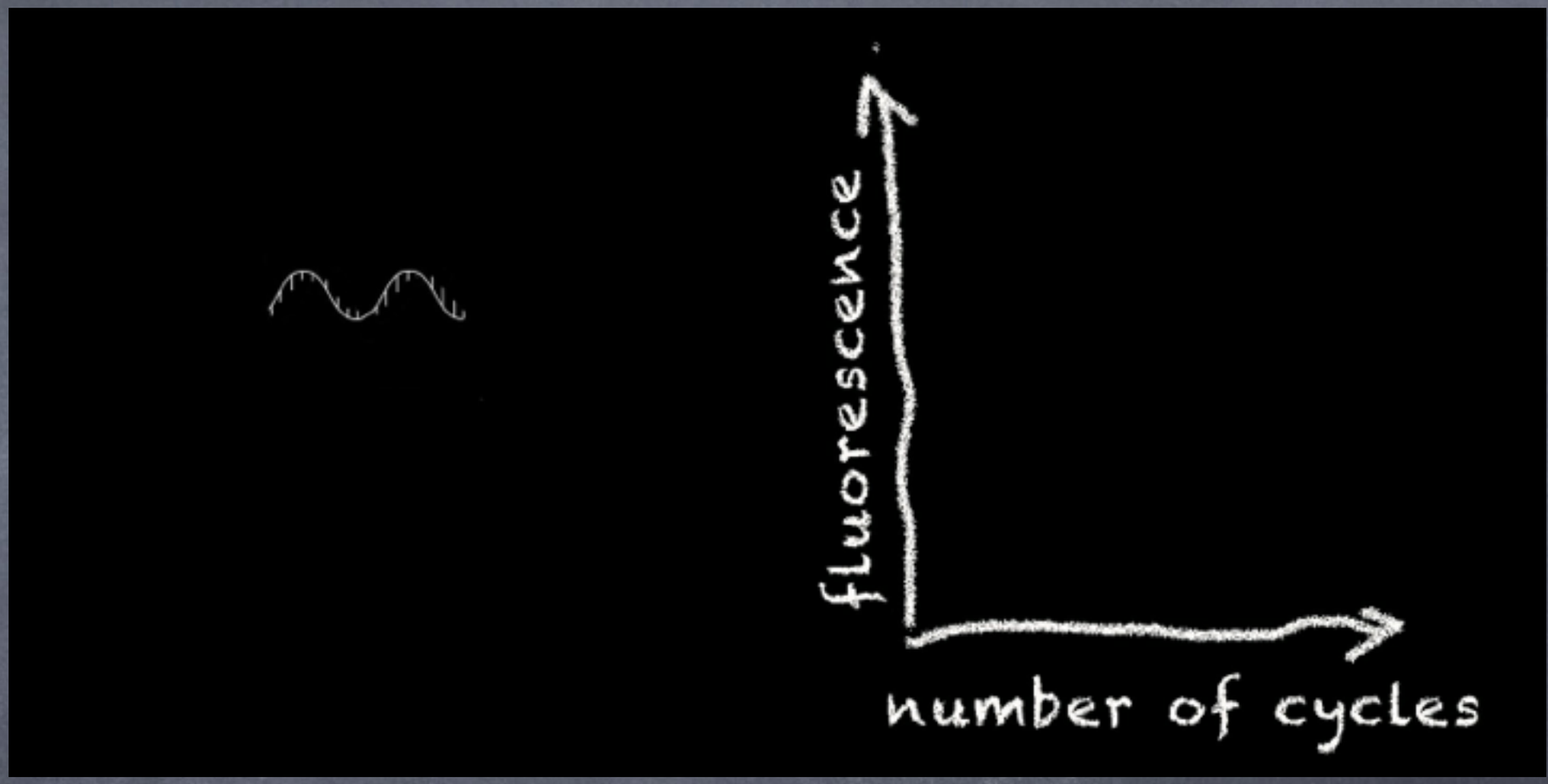
φορος (-phoros, "bearing", "carrying")

detection in end point PCR

mix PCR

↓

run PCR until the very last cycle

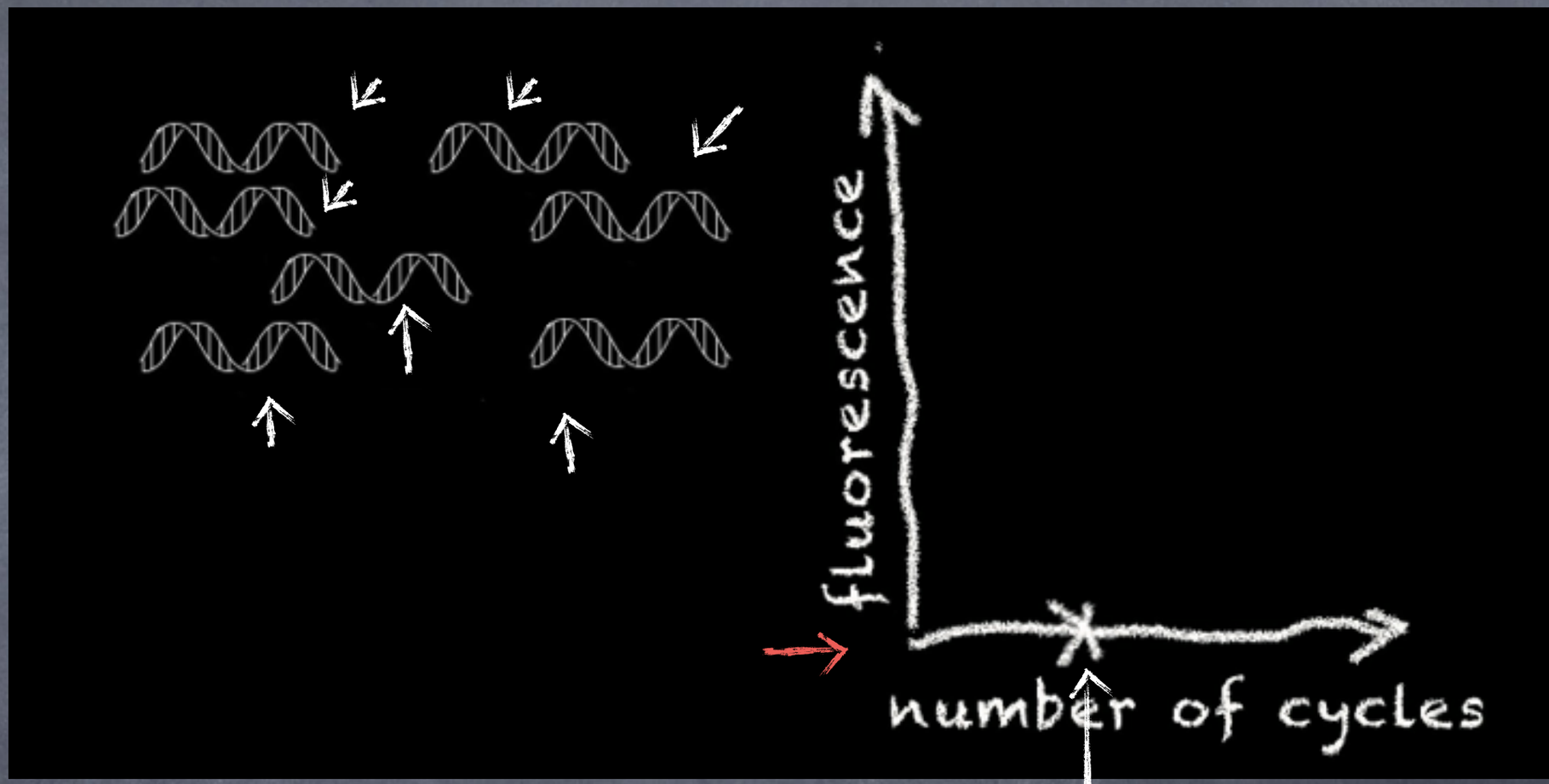


detection in end point PCR

mix PCR

↓

run PCR until the very last cycle



detection in end point PCR

mix PCR

↓

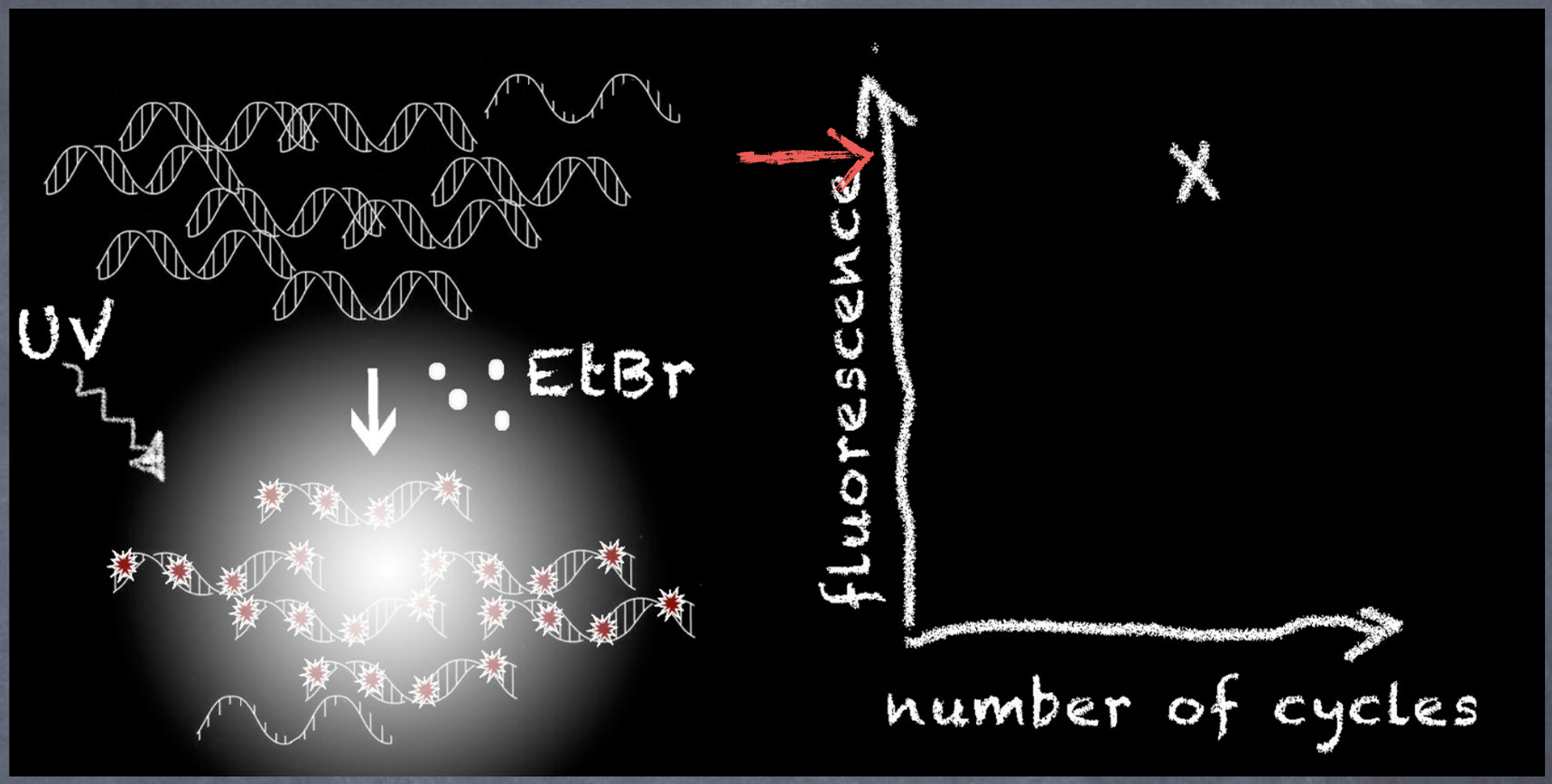
run PCR until the very last cycle

↓

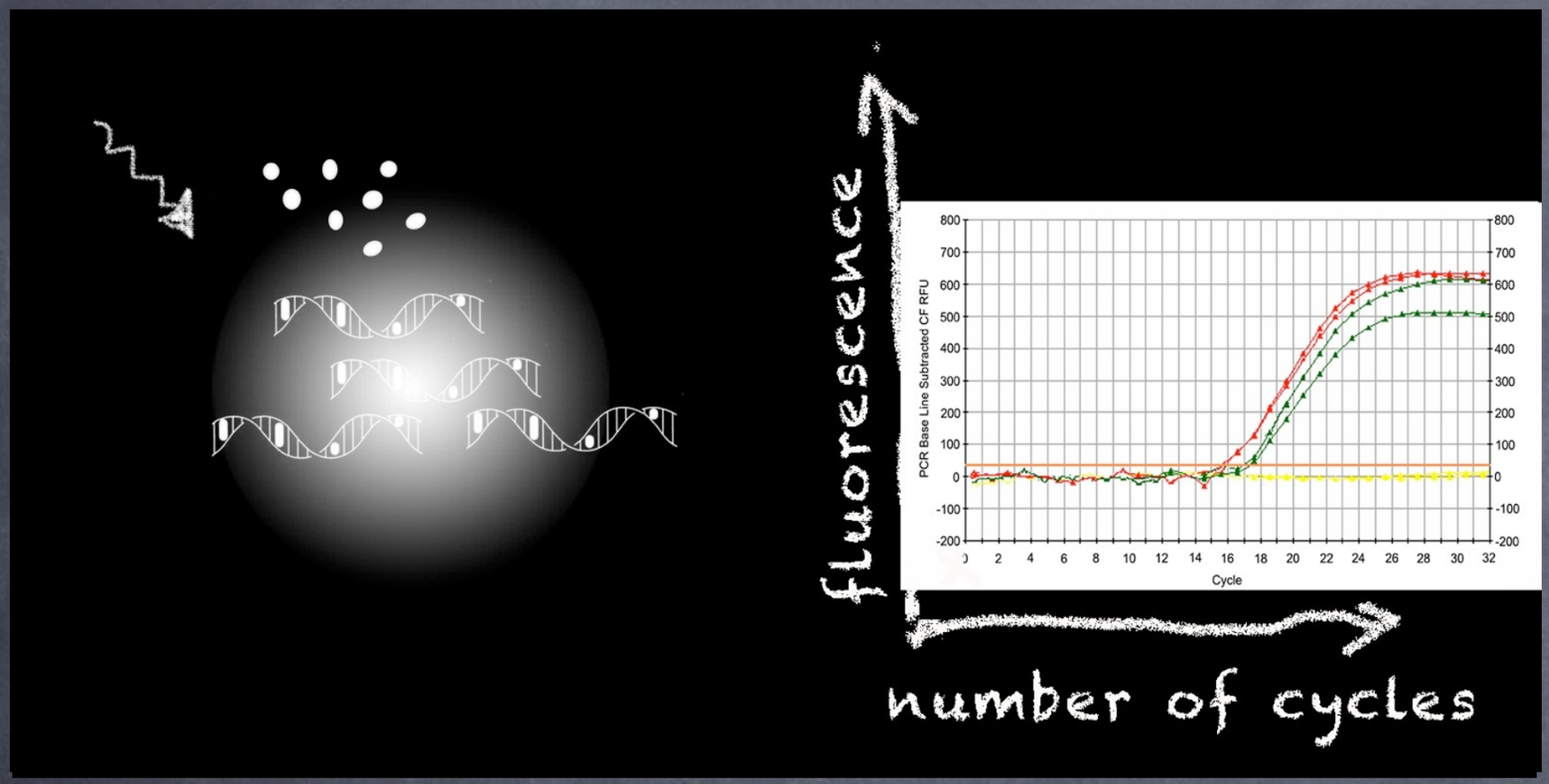
mix with fluorescent dye

↓

use UV lamp to induce fluorescence

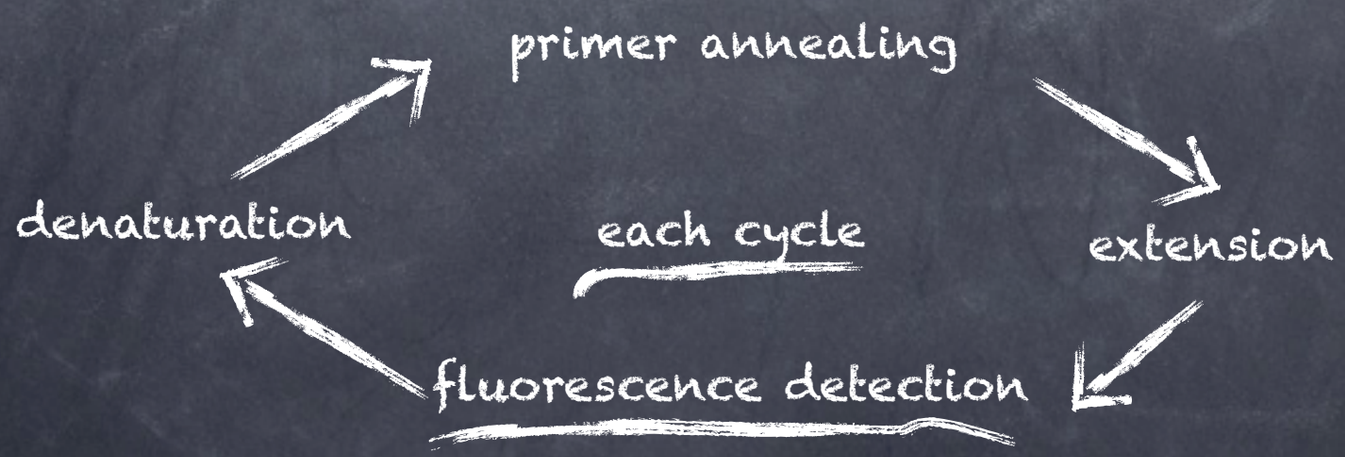


what is "real time" in the RT-PCR

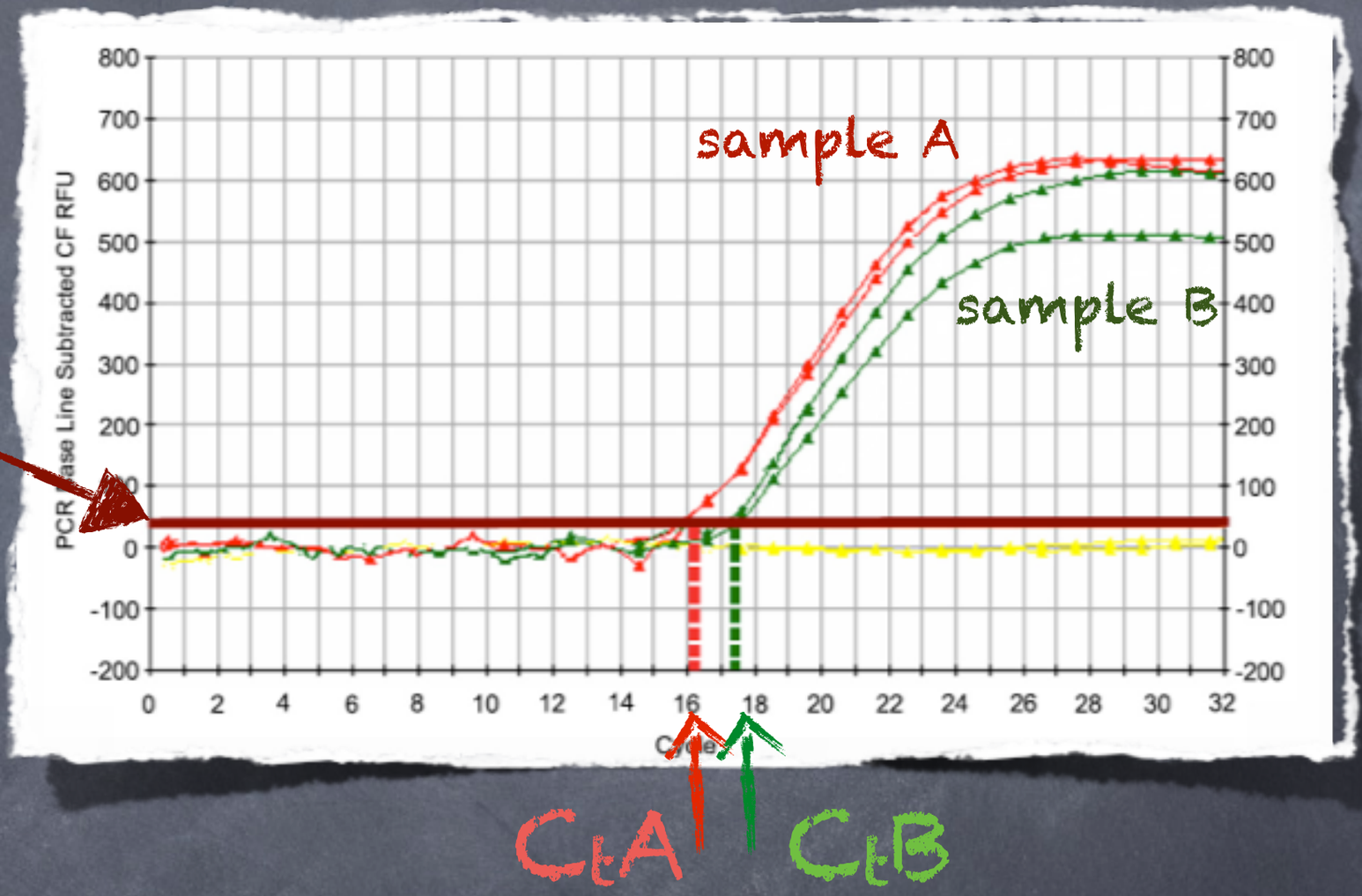


mix PCR together with fluorescent dye

run PCR and make detection after EACH cycle



threshold level of fluorescence
(fluorescence level is above
background => detectable)



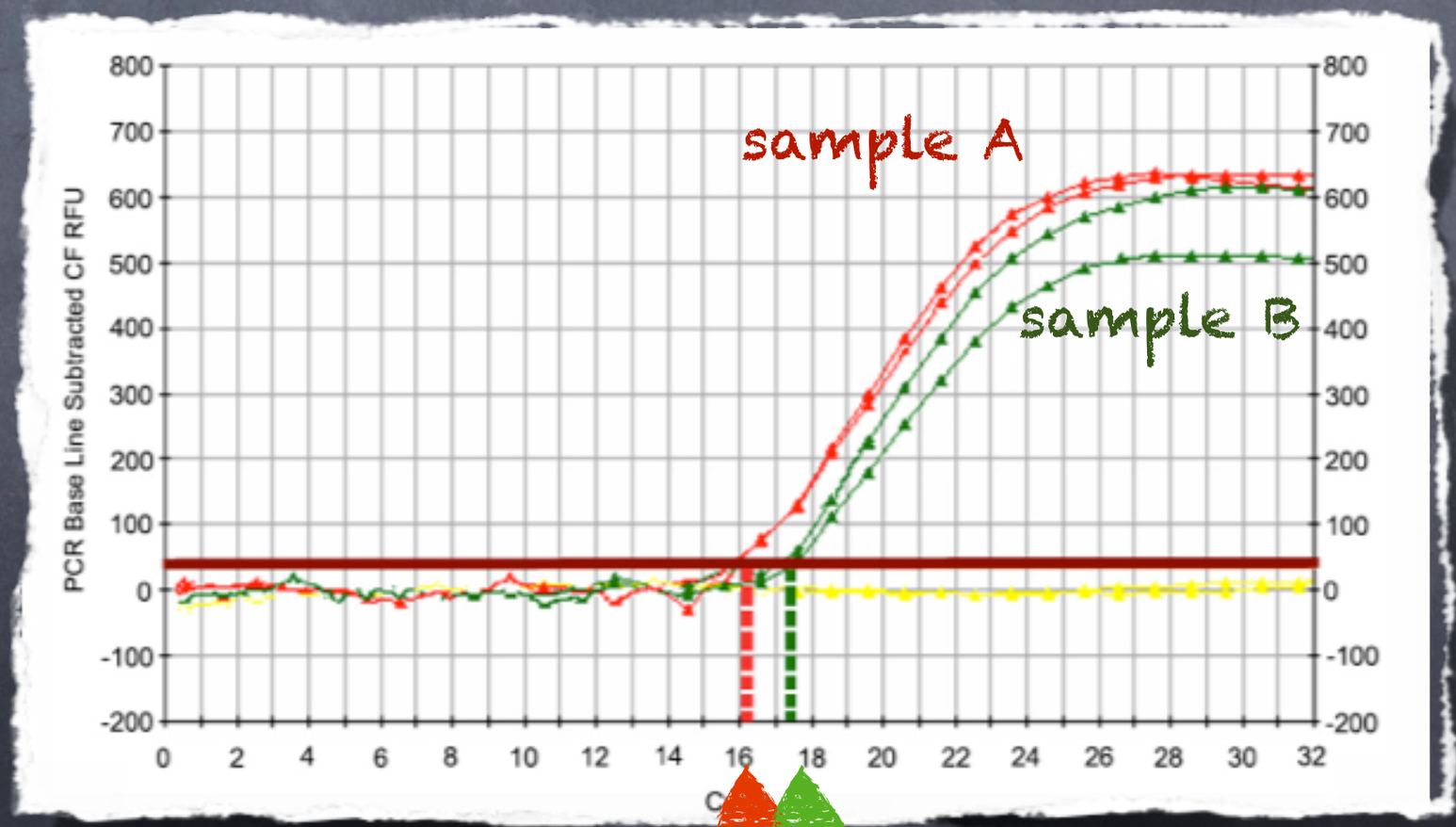
C_t= threshold cycle = C_q = quantitative cycle

C_t value represents how many PCR cycles are required for the sample fluorescence to reach the threshold level.

quantitative → qPCR

C_t value is determined by initial amount of a template in a sample

Lower Ct => higher amount of the template



Higher Ct => lower amount of the template

two types of quantification

Absolute

how many copies

is estimated using a standard curve

Relative

how much more/less

requires use of reference genes

Why qPCR?

- It is accurate (readings during exponential phase)
- It is quantitative
- Is much more sensitive than regular PCR, (detection at a few pg)
- Broad dynamic range (> 9 orders of magnitude)
- Fully integrated (amplify->detect and calculate) => allows sample throughput

For what qPCR?

- Transcription research/verification of RNAseq and microarray
- Diagnostic research: how much virus is swimming in your blood?
- Forensics: is it your DNA on the crime scene?
- Detect amount of GMO in stuff
- Detection of gene duplication or deletion
- Allelic discrimination assay
- Detection of % methylation of specific regions
- High resolution melt curve (SNP detection)
- etc.

Different types of qPCR

I'm teaching you yesterdays high tech

- Based on DNA-binding fluorescent dyes
- Based on fluorescent primers/probes:
  - TaqMan hydrolysis probe
  - Molecular beacons hairpin probe
- Digital PCR
- etc.

MIQE guidelines

Table 1. MIQE checklist for authors, reviewers, and editors.[a]

Item to check	Importance	Item to check	Importance
Experimental design		qPCR oligonucleotides	
Definition of experimental and control groups	E	Primer sequences	E
Number within each group	E	RTPrimerDB identification number	D
Assay carried out by the core or investigator's laboratory?	D	Probe sequences	D[d]
Acknowledgment of authors' contributions	D	Location and identity of any modifications	E
Sample		Manufacturer of oligonucleotides	D
Description	E	Purification method	D
Volume/mass of sample processed	D	qPCR protocol	
Microdissection or macrodissection	E	Complete reaction conditions	E
Processing procedure	E	Reaction volume and amount of cDNA/DNA	E
If frozen, how and how quickly?	E	Primer, (probe), Mg^{2+}, and dNTP concentrations	E
If fixed, with what and how quickly?	E	Polymerase identity and concentration	E
Sample storage conditions and duration (especially for FFPE[b] samples)	E	Buffer/kit identity and manufacturer	E
Nucleic acid extraction		Exact chemical composition of the buffer	D
Procedure and/or instrumentation	E	Additives (SYBR Green I, DMSO, and so forth)	E
Name of kit and details of any modifications	E	Manufacturer of plates/tubes and catalog number	D
Source of additional reagents used	D	Complete thermocycling parameters	E
Details of DNase or RNase treatment	E	Reaction setup (manual/robotic)	D
Contamination assessment (DNA or RNA)	E	Manufacturer of qPCR instrument	E
Nucleic acid quantification	E	qPCR validation	
Instrument and method	E	Evidence of optimization (from gradients)	D
Purity (A_{260}/A_{280})	D	Specificity (gel, sequence, melt, or digest)	E
Yield	D	For SYBR Green I, C_q of the NTC	E
RNA integrity: method/instrument	E	Calibration curves with slope and *y* intercept	E
RIN/RQI or C_q of 3′ and 5′ transcripts	E	PCR efficiency calculated from slope	E
Electrophoresis traces	D	CIs for PCR efficiency or SE	D
Inhibition testing (C_q dilutions, spike, or other)	E	r^2 of calibration curve	E
Reverse transcription		Linear dynamic range	E
Complete reaction conditions	E	C_q variation at LOD	E
Amount of RNA and reaction volume	E	CIs throughout range	D
Priming oligonucleotide (if using GSP) and concentration	E	Evidence for LOD	E
Reverse transcriptase and concentration	E	If multiplex, efficiency and LOD of each assay	E
Temperature and time	E	Data analysis	
Manufacturer of reagents and catalogue numbers	D	qPCR analysis program (source, version)	E
C_qs with and without reverse transcription	D[c]	Method of C_q determination	E
Storage conditions of cDNA	D	Outlier identification and disposition	E
qPCR target information		Results for NTCs	E
Gene symbol	E	Justification of number and choice of reference genes	E
Sequence accession number	E	Description of normalization method	E
Location of amplicon	D	Number and concordance of biological replicates	D
Amplicon length	E	Number and stage (reverse transcription or qPCR) of technical replicates	E
In silico specificity screen (BLAST, and so on)	E	Repeatability (intraassay variation)	E
Pseudogenes, retropseudogenes, or other homologs?	D	Reproducibility (interassay variation, CV)	D
Sequence alignment	D	Power analysis	D
Secondary structure analysis of amplicon	D	Statistical methods for results significance	E
Location of each primer by exon or intron (if applicable)	E	Software (source, version)	E
What splice variants are targeted?	E	C_q or raw data submission with RDML	D

[a] All essential information (E) must be submitted with the manuscript. Desirable information (D) should be submitted if available. If primers are from RTPrimerDB, information on qPCR target, oligonucleotides, protocols, and validation is available from that source.
[b] FFPE, formalin-fixed, paraffin-embedded; RIN, RNA integrity number; RQI, RNA quality indicator; GSP, gene-specific priming; dNTP, deoxynucleoside triphosphate.
[c] Assessing the absence of DNA with a no–reverse transcription assay is essential when first extracting RNA. Once the sample has been validated as rDNA free, inclusion of a no–reverse transcription control is desirable but no longer essential.
[d] Disclosure of the probe sequence is highly desirable and strongly encouraged; however, because not all vendors of commercial predesigned assays provide this information, it cannot be an essential requirement. Use of such assays is discouraged.

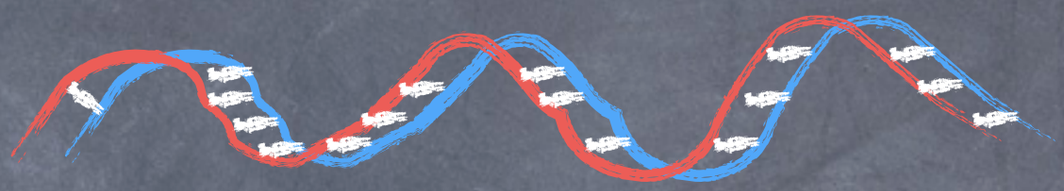
Efficiency of PCR

if PCR works ideally (E=100%) then with each cycle the amount of amplicon is doubled, i. e.

1 cycle of PCR

Efficiency of PCR

if PCR works ideally (E=100%) then with each cycle the amount of amplicon is doubled, i. e.



1 template gives rise to 2 copies

grey is the newly synthesised strand

Efficiency of PCR

if PCR works ideally (E=100%) then with each cycle the amount of amplicon is doubled, i. e.

N-number of template copies
n-number of PCR cycles

happens in an ideal world

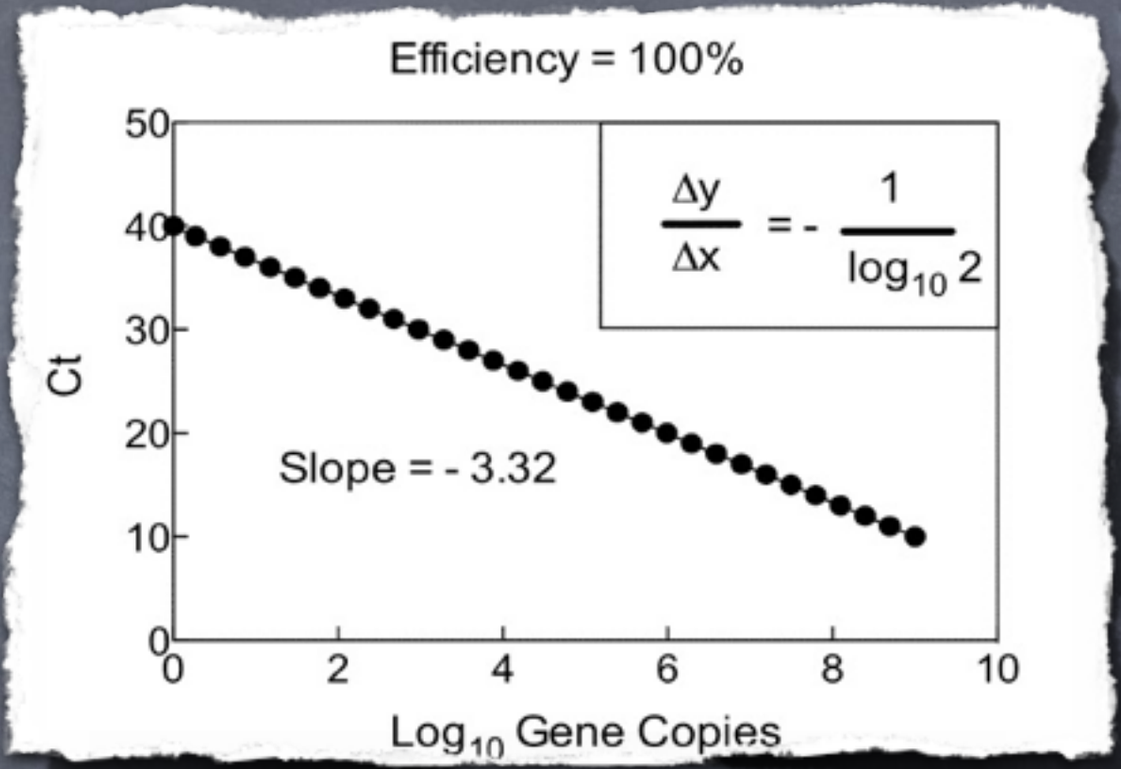
$$N_{final} = N_{initial} \times 2^n$$

this is the data you get in a form of the threshold fluorescence

this is what you want to know

Efficiency is estimated by running PCR with template dilution series

- template concentration is hard to get -> use dilutions
- The efficiency of a qPCR reaction is 100% when the DNA doubles with each cycle.
- adding twice less template will increase the Ct value by 1 cycle



$$E = 10^{-1/slope}$$

$$\% \text{ Efficiency} = (E - 1) \times 100\%$$

E = 100%

DNA 16 → C_t=20

↓ dilute twice

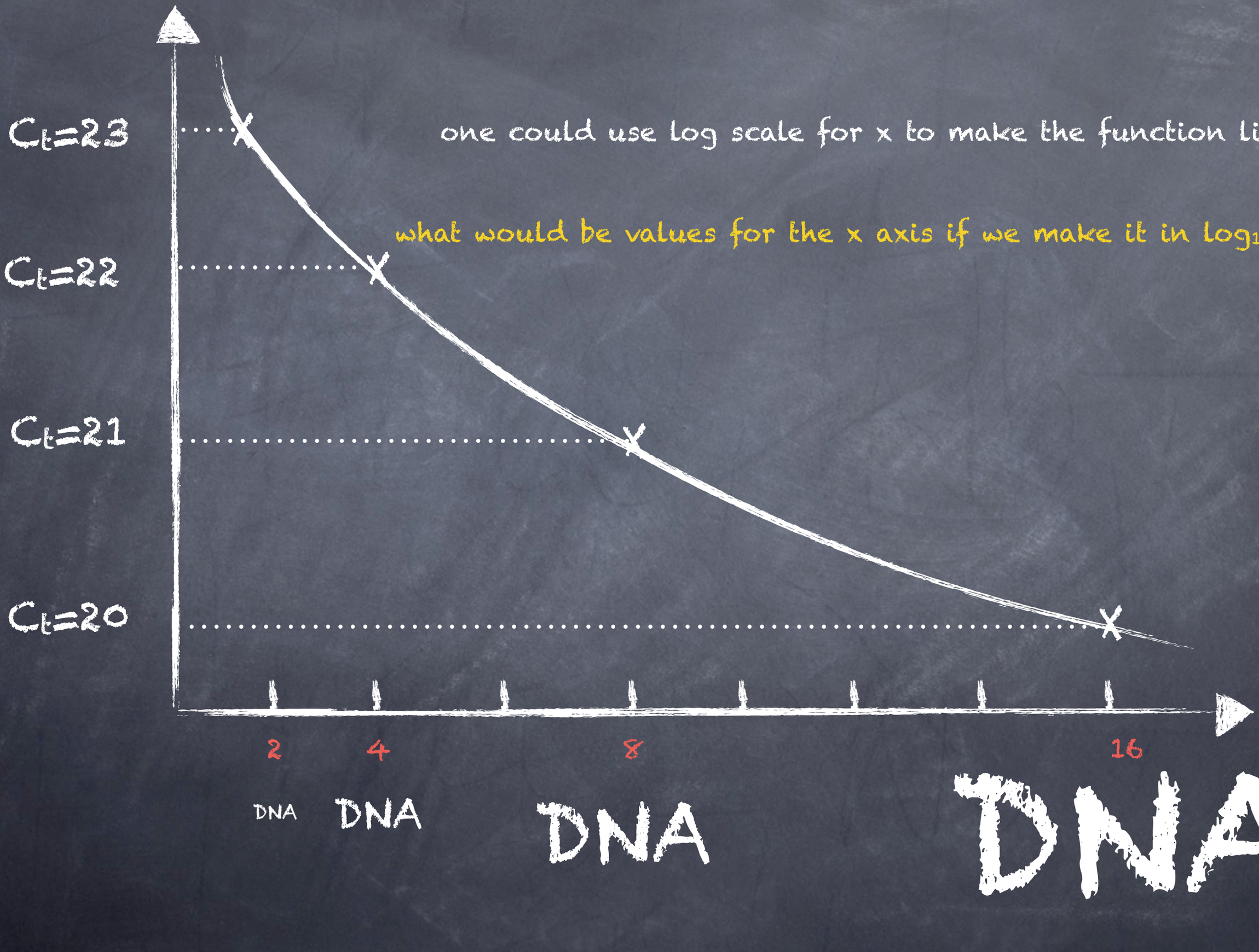
DNA 8 → C_t=?

↓ dilute twice

DNA 4 → C_t=?

↓ dilute twice

DNA 2 → C_t=?



one could use log scale for x to make the function linear

what would be values for the x axis if we make it in $\log_{10}$ scale?

what is log?

$\log_a b=c \quad \Rightarrow \quad a^c=b$

$\log_{10}1000=? \quad \Rightarrow \quad 10^3=1000$

$\log_{10}100=? \quad \Rightarrow \quad 10^2=100$

$\log_{10}10=? \quad \Rightarrow \quad 10^1=10$

$\log_{10}1=? \quad \Rightarrow \quad 10^0=1$

$\log_{10}16=? \quad \Rightarrow \quad 10^{1.2}=16$

$\log_{10}8=? \quad \Rightarrow \quad 10^{0.9}=8$

$\log_{10}4=? \quad \Rightarrow \quad 10^{0.6}=4$

$\log_{10}2=? \quad \Rightarrow \quad 10^{0.3}=2$

Δx = dilution

$\Delta y = \Delta C_t$

$\Delta y / \Delta x$ = slope

C_t=23

C_t=22

C_t=21

C_t=20

$E = 10^{-1/slope}$

% Efficiency = (E - 1) x 100%

0.3

0.6

0.9

1.2

DNA

DNA

DNA

DNA

DNA	→	E = ? Ct=20	E = ? Ct=19	E = ? Ct=20	E = ? Ct=33
↓ dilute five times					
DNA	→	Ct=22	Ct=22	Ct=23	Ct=35
↓ dilute five times					
DNA	→	Ct=24	Ct=25	Ct=24	Ct=36.8
↓ dilute five times					
DNA	→	Ct=26	Ct=27	Ct=25	Ct=39
		group 1	group 2	group 3	group 4

What can mess up efficiency of amplification?

E > 100%

low primer specificity (melt curve)

primer dimers (melt curve)

inhibitors (dilution series)

E < 100%:

bad primers/primer concentration

dead polymerase

R^2

coefficient of determination

= how well does the model predict future outcomes

should I really trust this?

if R^2 is > 0.98, then you can trust it

$R^2 = 1 - (SS_{err}/SS_{tot})$

you want R^2 to be as close to 1 as possible.

what will happen to R^2 if you use 2 dilutions for your curve?

3 basic methods to quantify relative qPCR data

Normalized expression = $E^{-\Delta Ct}$

- Pfaffl method
- ΔΔCt (Livak method) = Pfaffl method when E=100%
- ΔCt

A^x

$A^{-x} = ?$

$A^x * A^y = ?$

$A^x / A^y = ?$

$A^x * B^y = ?$

$A^x / B^y = ?$

A gene control sample
A gene treated sample
Reference gene control sample
Reference gene treated sample

what are the units for this?

what is this?

why does this look so bad?

what's this?

what are the units for this?

fluorescence

20

25 27

27.5

cycles

what are these?

CtAc CtAt CtRc CtRt

N-number of template copies
n-number of PCR cycles

happens in an ideal world

$$N_{final} = N_{initial} \times 2^n$$

this is the data you get in a form of the threshold fluorescence

this is what you want to know

$$N_{initial} = N_{final} / E^n$$

how is this expressed in the qPCR data?
How do you see amount of DNA?

A gene control sample
A gene treated sample
Reference gene control sample
Reference gene treated sample

where are N_{final} and $N_{initial}$?

$$N_{initial\ control} = N_{final\ control} / E^{Ct\ control}$$

$$N_{initial\ treatment} = N_{final\ treatment} / E^{Ct\ treatment}$$

how much more/less DNA we have in the treatment?

$$\frac{N_{initial\ treatment}}{N_{initial\ control}} = \frac{N_{final\ treatment} / E^{Ct\ treatment}}{N_{final\ control} / E^{Ct\ control}}$$

$$\frac{N_{initial\ treatment}}{N_{initial\ control}} = \frac{E^{Ct}}{E^{Ct}} = E^{(Ct-Ct)} = E^{-(Ct-Ct)}$$

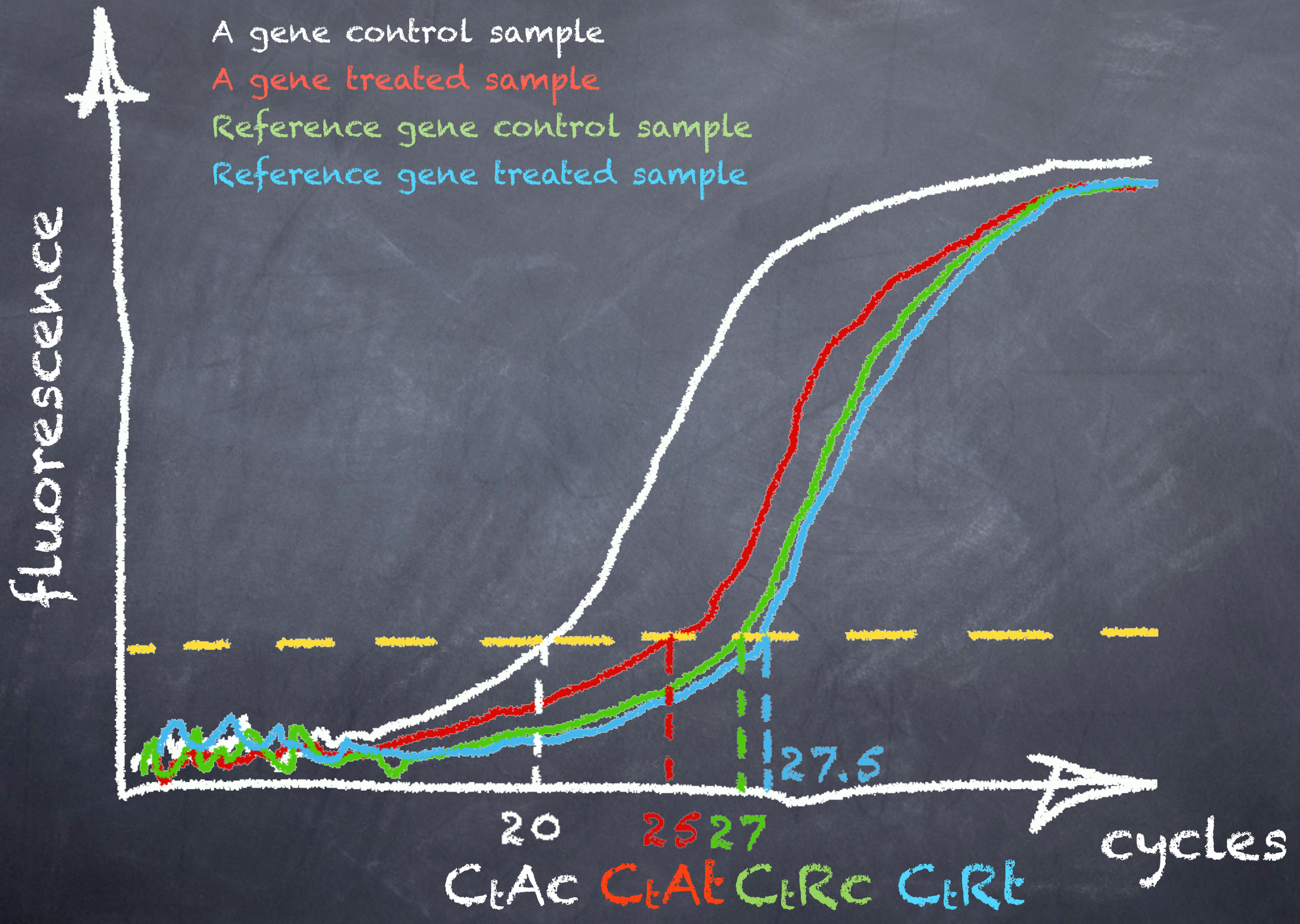
Normalized expression = $E^{-(Ct-Ct)} = E^{-\Delta Ct}$

A gene control sample

A gene treated sample

Reference gene control sample

Reference gene treated sample



Pfaffl method
(most reliable)

1. Normalize expression of gene of interest in control and treatment

$$E_{geneA}^{-\Delta C_t geneA} = E_{geneA}^{-(C_t Ac - C_t At)}$$

2. Normalize expression of reference gene control and treatment

$$E_{ref}^{-\Delta C_t ref} = E_{ref}^{-(C_t Rc - C_t Rt)}$$

3. Calculate expression ratio of genA and reference gene

$$E_{geneA}^{-\Delta C_t geneA} / E_{ref}^{-\Delta C_t ref}$$

ΔΔCt

$2^{-\Delta\Delta Ct}$ method, Livak method

is the same as Pfaffl method, but requires 100% efficiency,

or equal efficiency of all primers you use (you put it instead of 2)

1. normalize gene of interest to the reference gene

$$\Delta Ct control = CtAc - CtRc$$

$$\Delta Ct treatment = CtAt - CtRt$$

2. normalize test to control

$$\Delta\Delta Ct = \Delta Ct treatment - \Delta Ct control$$

3. expression ratio

$$\text{Normalized expression ratio} = 2^{-\Delta\Delta Ct}$$

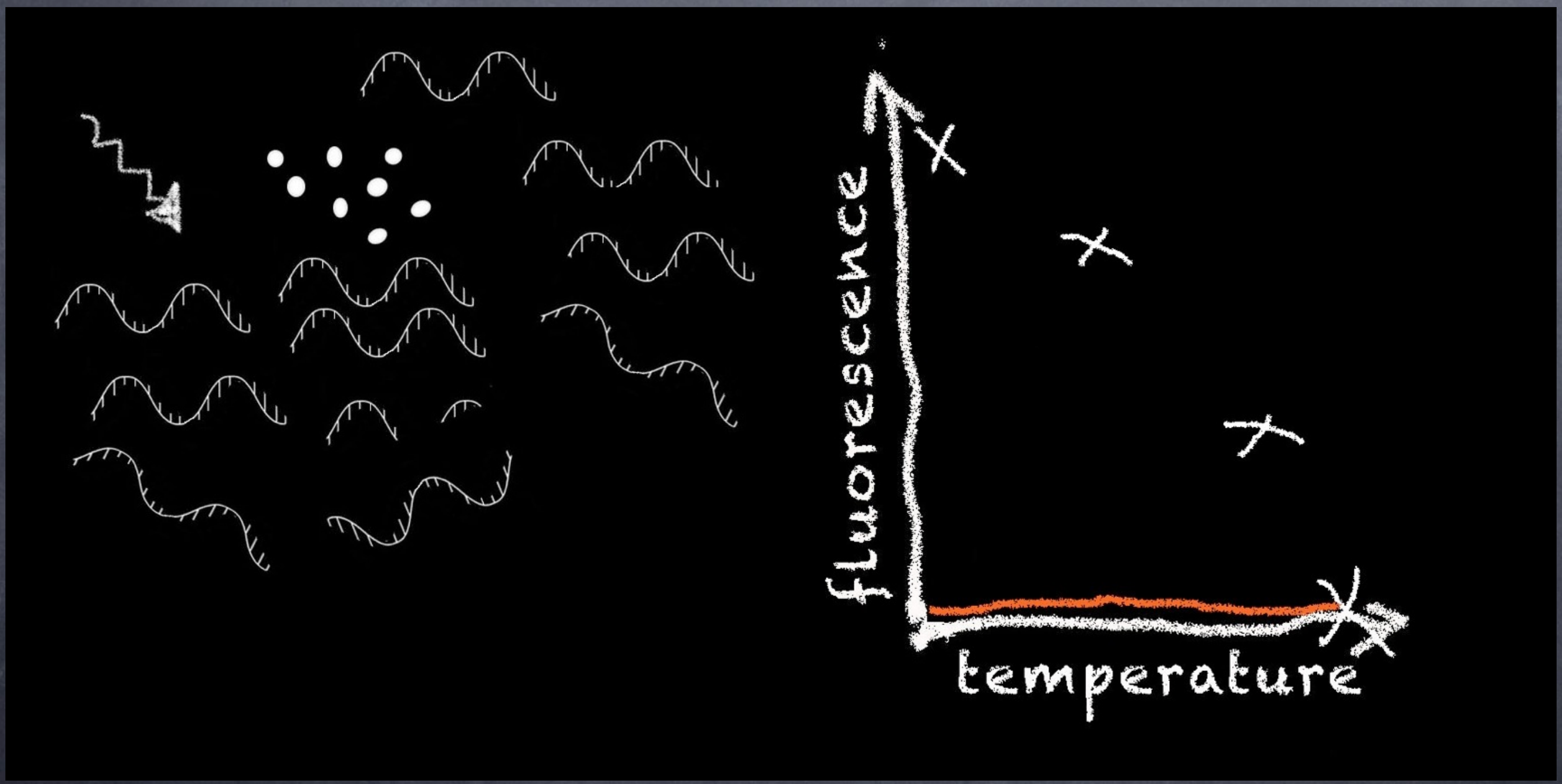
why can't you use Livak's method if E for genes are not the same?

ΔCt method

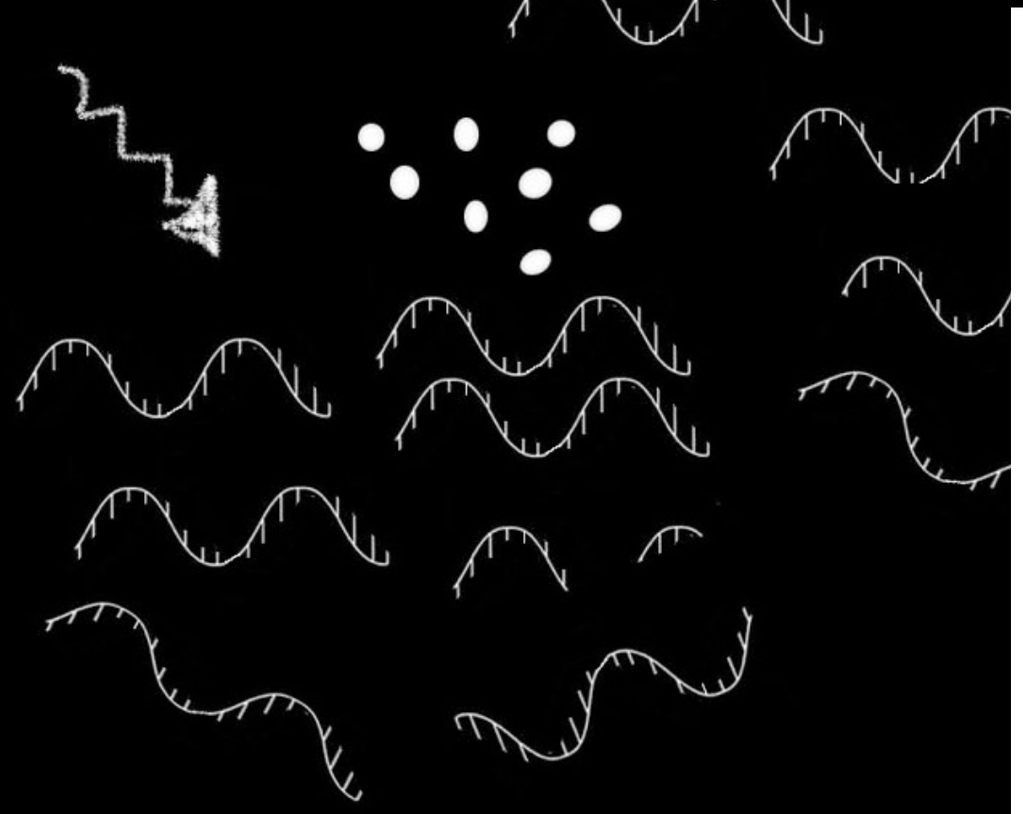
no reference gene is used, instead samples were equilibrated using another parameter (DNA concentration, amount of cells etc.)

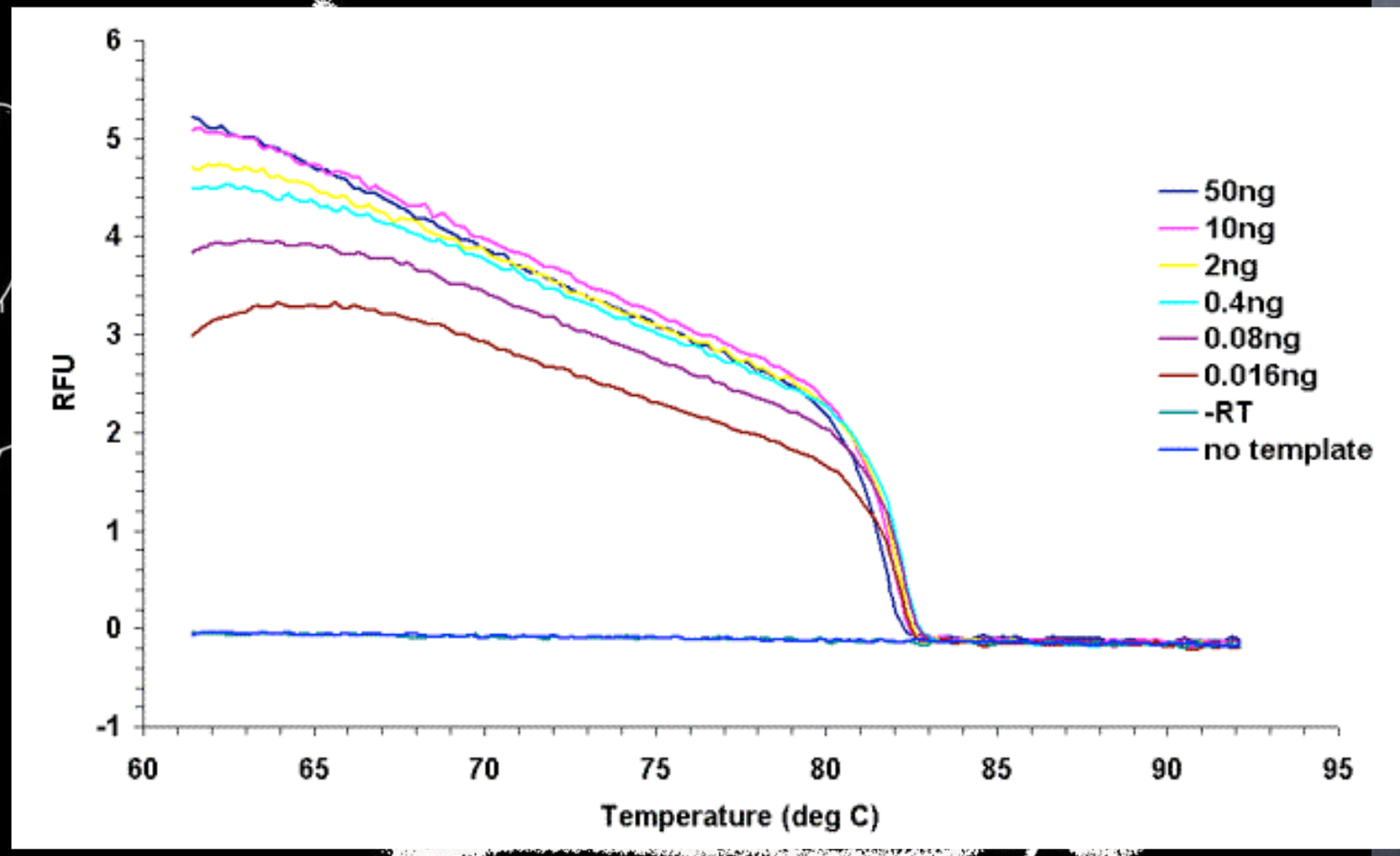
Normalized expression ratio = $E^{\Delta Ct} = E^{(Ct\ control\ -Ct\ treatment)}$

melt curve



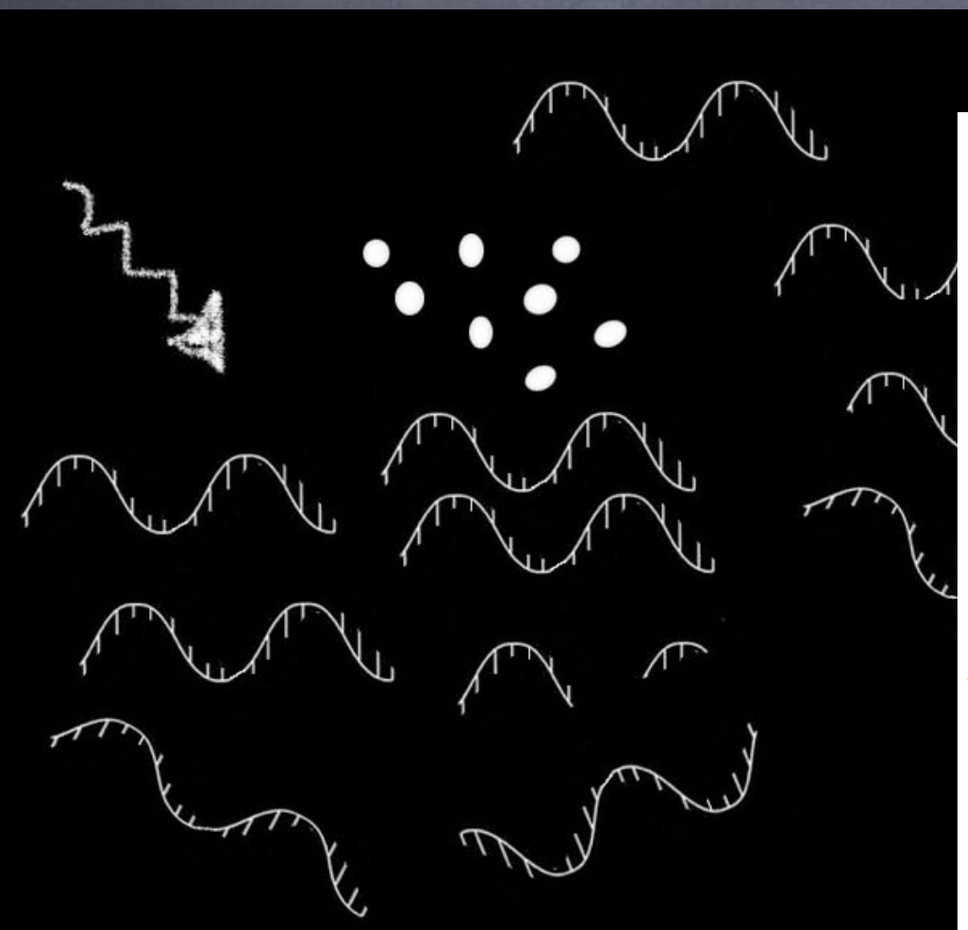
melt curve

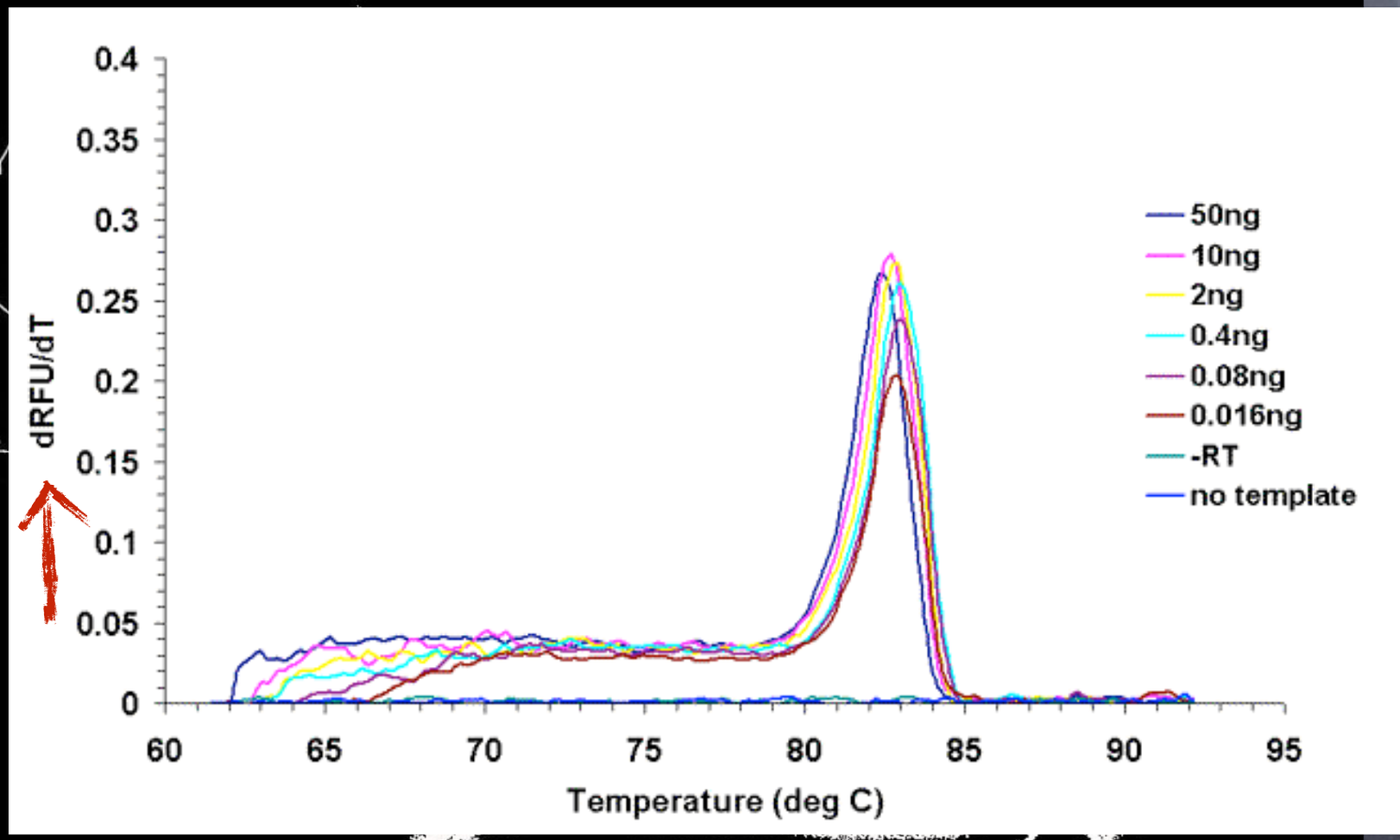




temperature

melt curve





temperature

Homework will be uploaded on the site