

**MIQE - Minimum standard for the
provision of information in quantitative
PCR**

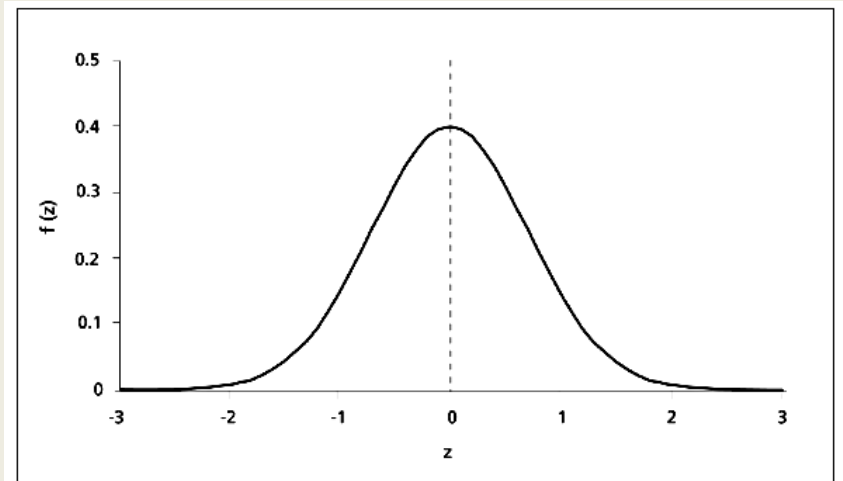
Statistické vyhodnocení

Nulová hypotéza (není rozdíl), kterou se test buď potvrdí nebo vyvrátí

Základní statistické parametry – průměr \pm SD nebo medián \pm x-percentil

Parametrické testy

- Normální rozložení
- Stejný rozptyl
- **Dva vzorky** t-test (one/two tailed)
- **Různé nezávislé proměnné** - ANOVA (i tehdy, není-li rozložení úplně Gaussovské)



Genová exprese (expresní poměry) mívá obvykle normální rozložení, pokud je vyjádřena v log scale.

Statistické vyhodnocení

Neparametrické testy

- Neznáme parametry rozložení
- Testování rozdílů mezi nezávislými skupinami (Independent samples)
 - dva vzorky**, u kterých porovnááme průměry některé z proměnných

- Mann-Whitey U test; Kolmogorov-Smirnov test

více skupin

- Kruskal-Wallis test; Mediánový test

-Testování rozdílů mezi závislými skupinami

- Porovnávání proměnných, zjišťovaných na jednom vzorku
 - Wilcoxonův test (parametrická alternativa – t-test/ANOVA)
 - Hodnoty typu „mRNA přítomná/nepřítomná“ (dichotomické hodnoty) – McNemarův χ^2 test

-Testování vztahů mezi proměnnými

- Regrese a korelace (Spearmanův/Pearsonův korelační koeficient)
 - Standardní křivky

Kvantifikační strategie

Statistické vyhodnocení

Kdy použít který test

Parametrické vs. neparametrické testy

RT-PCR

Obvykle malý počet hodnot, s velkým rozptylem, většinou nesledujících normální rozložení

-> **neparametrické testy**

V případě většího počtu hodnot (>100), lze použít parametrické testy

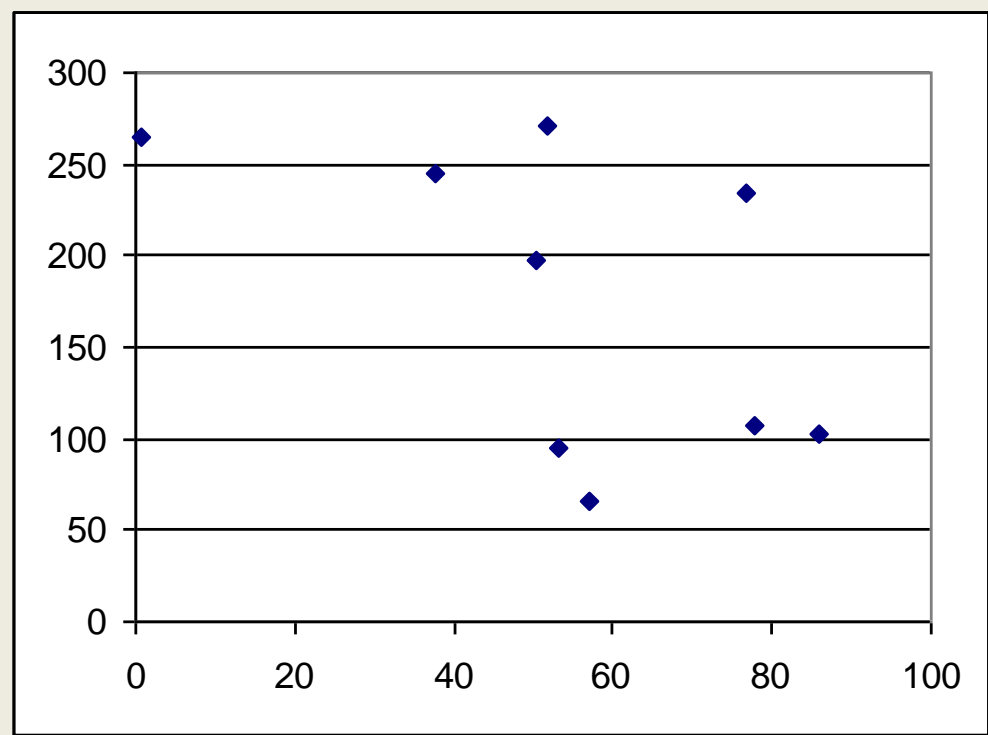
Neparametrické testy jsou méně náchylné k α -chybám (nesprávné zamítnutí nulové hypotézy), ale jsou méně citlivé než parametrické (např. srovnání p u para < p u neparametrických testů), jako signifikantní označí větší rozdíl než parametrické testy.

Kvantifikační strategie

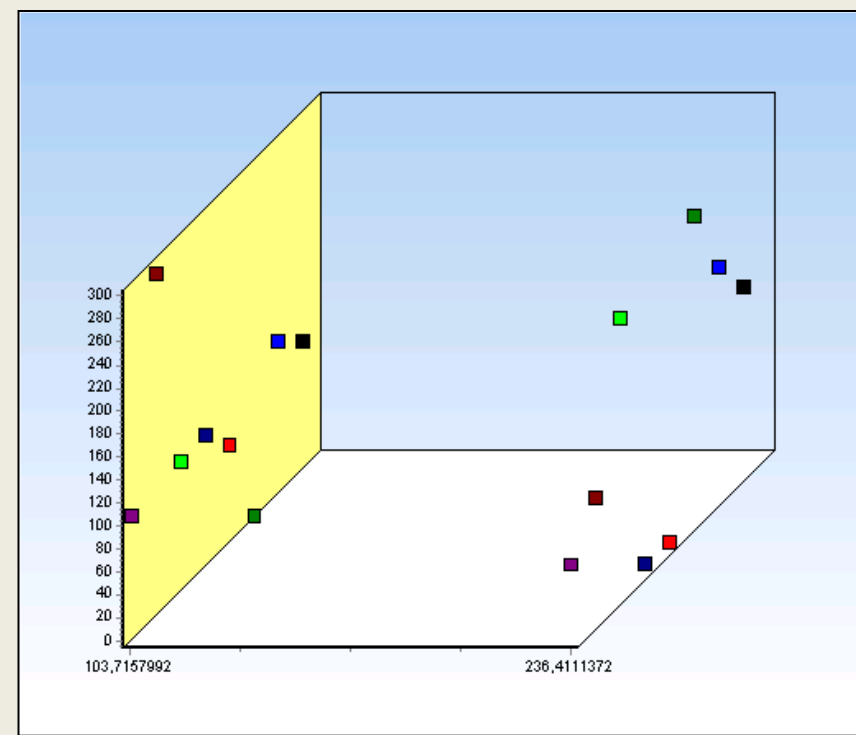
Statistické vyhodnocení

Analýza více genů/vzorků (clustering)

Dvourozměrný graf



Třírozměrný graf



Kvantifikační strategie

Statistické vyhodnocení

Analýza více genů/vzorků (hledání trendů, clustering)

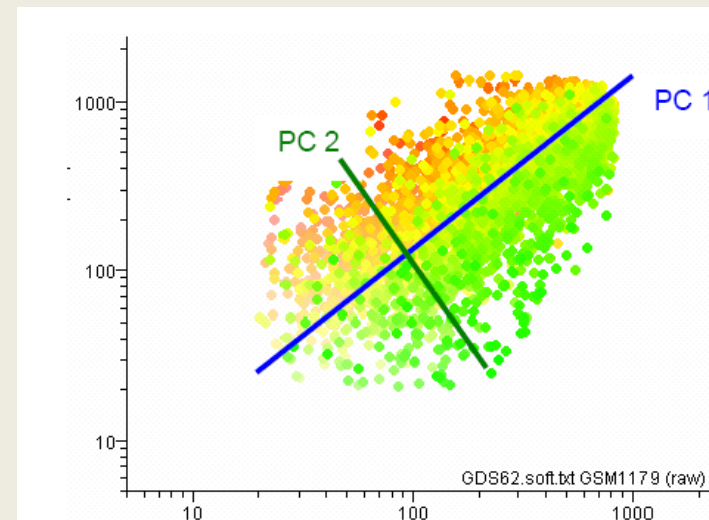
N různých genů – n různých proměnných – n rozměrný graf ?

Př. 10000 genů... (microarray)

Principal component analysis (PCA)

Redukce počtu rozměrů (dimenzionality) na základě výpočtu kovariance mezi jednotlivými vzorky.

Původní osy jdou nahrazeny tzv. komponentami



Kritická místa statistického vyhodnocení

Kontrola dat (outliers)

Úprava efektivity PCR

Kompenzace variability mezi jednotlivými PCR (inter-plate calibration)

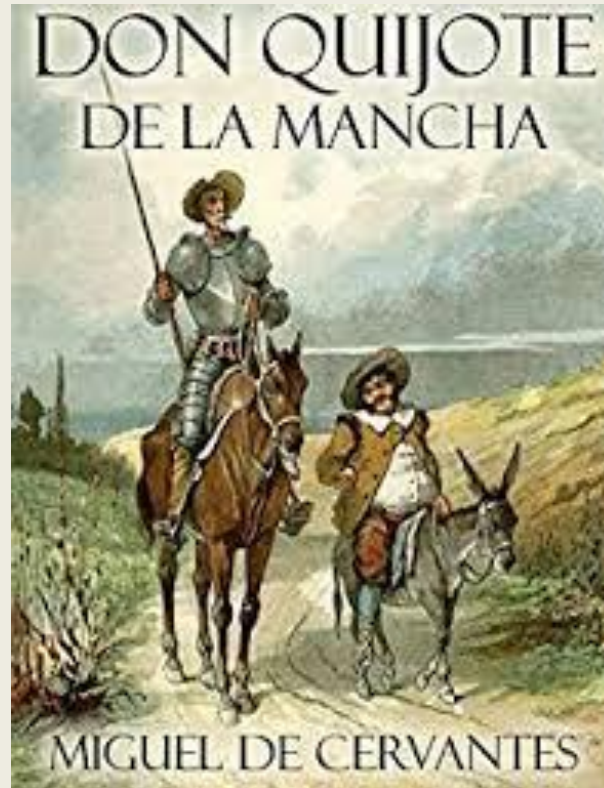
Normalizace na stejné množství vzorku (RNA/DNA)

Průměrování technických replikátů

Výpočet množství/poměrů

Vlastní statistická analýza

MIQE guidelines



Standardy MIQE

PCR & qPCR

jednoduchá, snadná, rychlá, citlivá metoda

→ **Popularita ve vědecké komunitě**

adaptace, úpravy, specifikace protokolů...

→ **Publikace dat s různou kvalitou**

problematická data zpochybňují i oprávněné interpretace

→ **Nutnost standardů**

...there is a lamentable lack of transparency of reporting, with the material and methods sections of many publications, especially those with high impact factors, not fit for the purpose of evaluating the quality of any reported qPCR data. This poses a challenge to the integrity of the scientific literature with serious consequences not just for basic research but potentially calamitous implications for drug development and disease monitoring....

Bustin 2009

Variabilita a její příčiny

Biologická variabilita

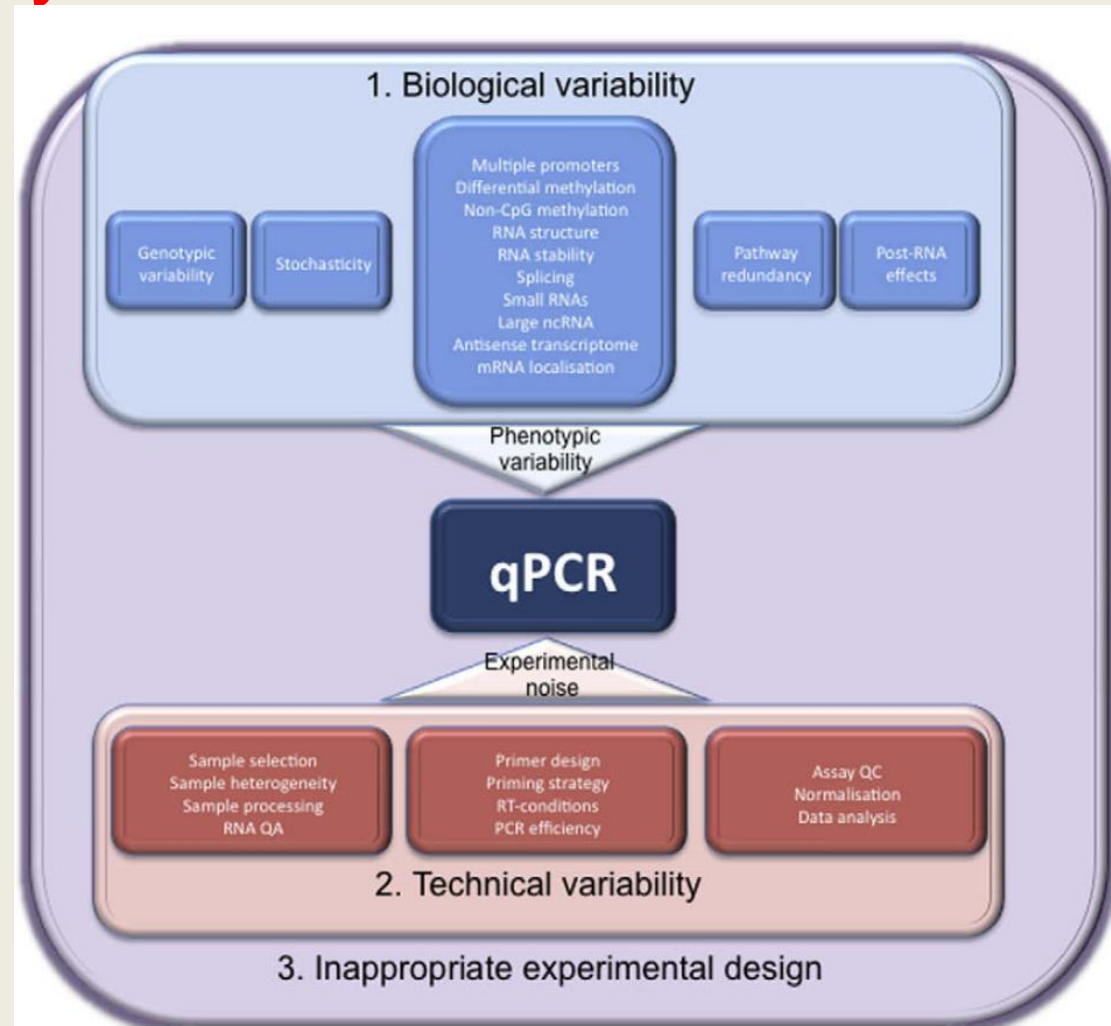
– experimenty odrážejí různorodost reality – nebudou nikdy identické

Technická variabilita

– chyby v měření, znemožňující adekvátní popis reality

Experimentální design

– chybná hypotéza vedoucí k výsledkům platným pouze v rámci experimentu
– biologické nebo klinické významnosti
– data overestimation



Biologická variabilita

- hallmark of life
- kombinace genotypových a fenotypových variací mezi jedinci
- tkáně a buňky – dynamické systémy se schopností komplexní adaptace a reakce na různé podmínky

Genetická variabilita

polymorfismy, copy-number variace, alternativní splicing, posttranskripční a posttranslační regulace, epigenetické modifikace

Fenotypová variabilita

environmentální interakce, intra- a extraindividuální faktory (věk, životní/reprodukční cyklus, pohlaví, čas, nutriční stav...)

Biologická variabilita

Stochastická variabilita na úrovni kinetického šumu
biochemických reakcí uvnitř jediné buňky

= dynamické chování jediné buňky není přesně reprodukovatelné



Expresní profily jednotlivých buněk **se liší**,
i v rámci homogenní kultury

Interakce mezi regulačními molekulami a DNA
Lokalizace mRNA a proteinů v rámci buňky
Epigenetické modifikace



**I geneticky identické buňky ve stejném prostředí
mohou mít různý fenotyp**

Biologická variabilita

Biologické informační dráhy jsou robustní, redundantní, závislé na buněčném, tkáňovém i environmentálním kontextu



Biologicky relevantní interpretace pozorovaného jevu a jeho odlišení od přirozené variability a heterogenity v daném systému vyžaduje správný experimentální design, analytické metody a vhodný statistický model



Slepě nepřejímat cizí protokoly, přemýšlet 😊

Technická variabilita

- Mnoho protokolů, které se liší v každém kroku = rozdíly ve výsledcích
- Meticulous attention to sample isolation, storage and preparation, množství replikátů, design assays, samotný qPCR proces, normalizace a statistická analýza
- Sample size je relativně malá, statistické analýzy mají malou sílu, fold change je málo precizní a zvýšená pravděpodobnost falešně pozitivních výsledků roste

Proč je to problém?

- Původní jednoduchý protokol se změnil na mnoho složitějších špatně popsaných protokolů, které je těžké zkontrolovat
- qPCR = neadekvátní standardizace, komplexní a nekonsistentní technika vedoucí ke špatným výsledkům

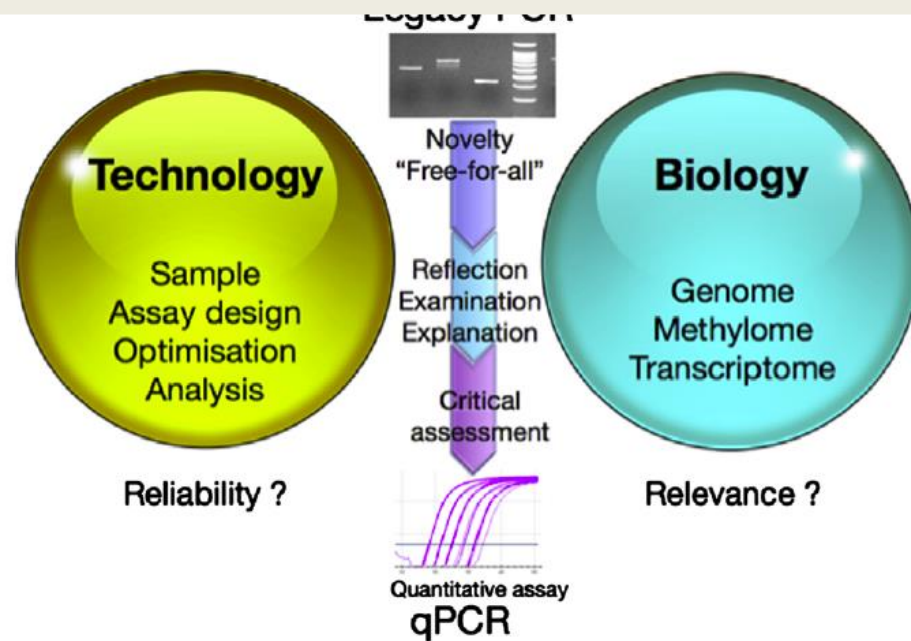


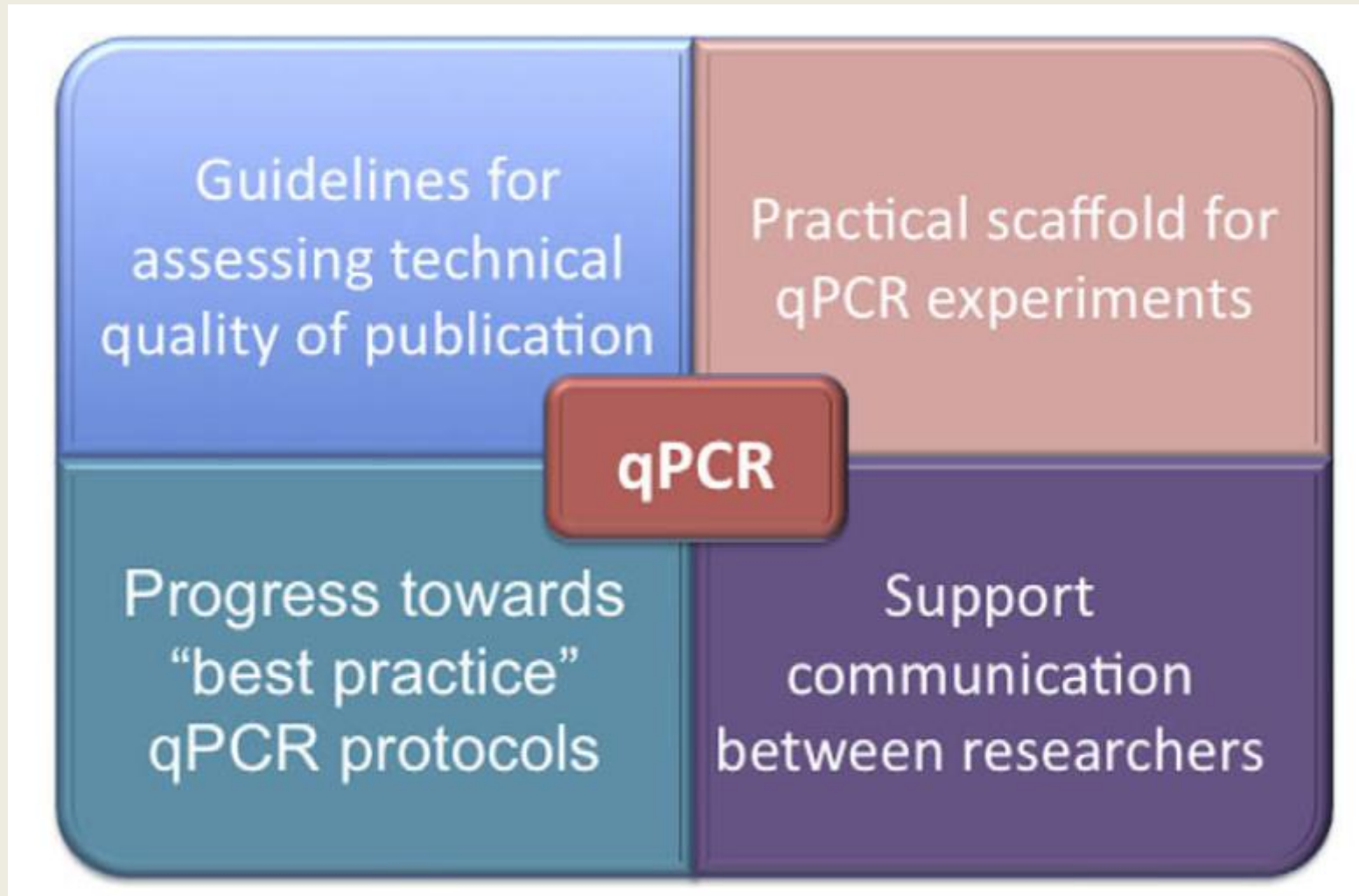
Fig. 2. PCR evolution. The evolution of legacy, gel-based PCR, established as a powerful method for the qualitative detection of nucleic acids into the quantitative qPCR assay of to-day proceeded through a protracted phase of trial and error. This scrutiny resulted in a more detailed understanding of both the technological as well as biological limitations and challenges of this technology. The serious question marks surrounding both the reliability and relevance of qPCR data contributed heavily to the development of the MIQE guidelines.

Technická variabilita



Publication Parameter	A	B	C	D
RNA QA	Agilent	gel	✗	Agilent
Primers probes	Assay on demand	✓	✓	✓
RT	✓	✓	✓	✓
cDNA priming	oligo-dT	✗	random hexamers	✗
Temp/vol/time	✓/✓/✗	✓/✓/✗	✗/✓/✗	✗/✗/✗
PCR efficiency	✗	✗	✗	✗
Normalisation	single, unvalidated RG	single, unvalidated RG	unvalidated 18S rRNA	single, unvalidated RG
Biological replicates	✗	✗	✗	✗

Cíle standardizace



Technická variabilita

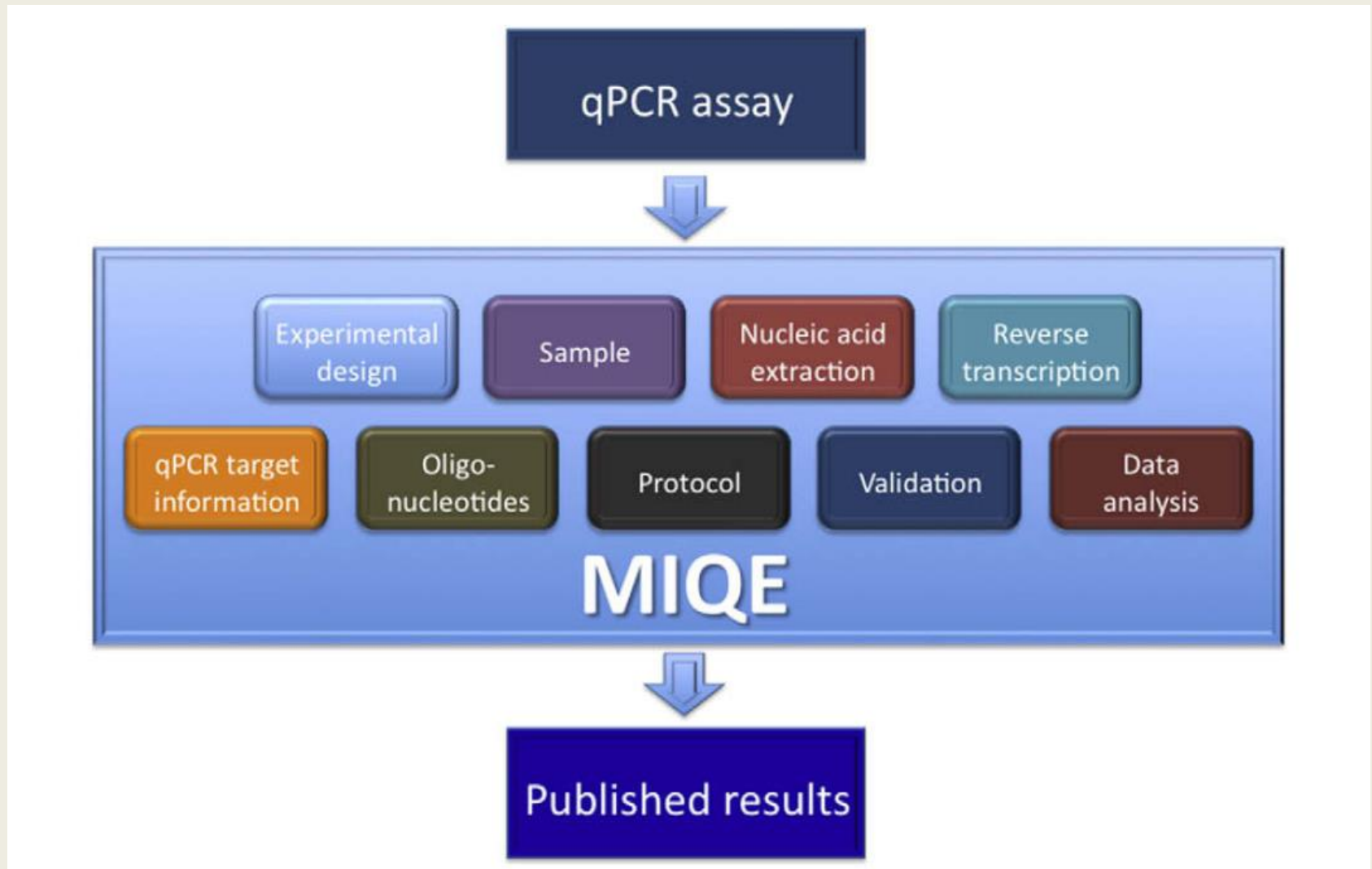


Table 1. MIQE checklist for authors, reviewers, and editors.^a

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 ^b
Acknowledgment of authors' contributions	D	Location and identity of any modifications	E
Sample		Manufacturer of oligonucleotides	
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	
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		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	CI _s 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	
Complete reaction conditions	E	C _q variation at LOD	E
Amount of RNA and reaction volume	E	CI _s 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 _q s 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	
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

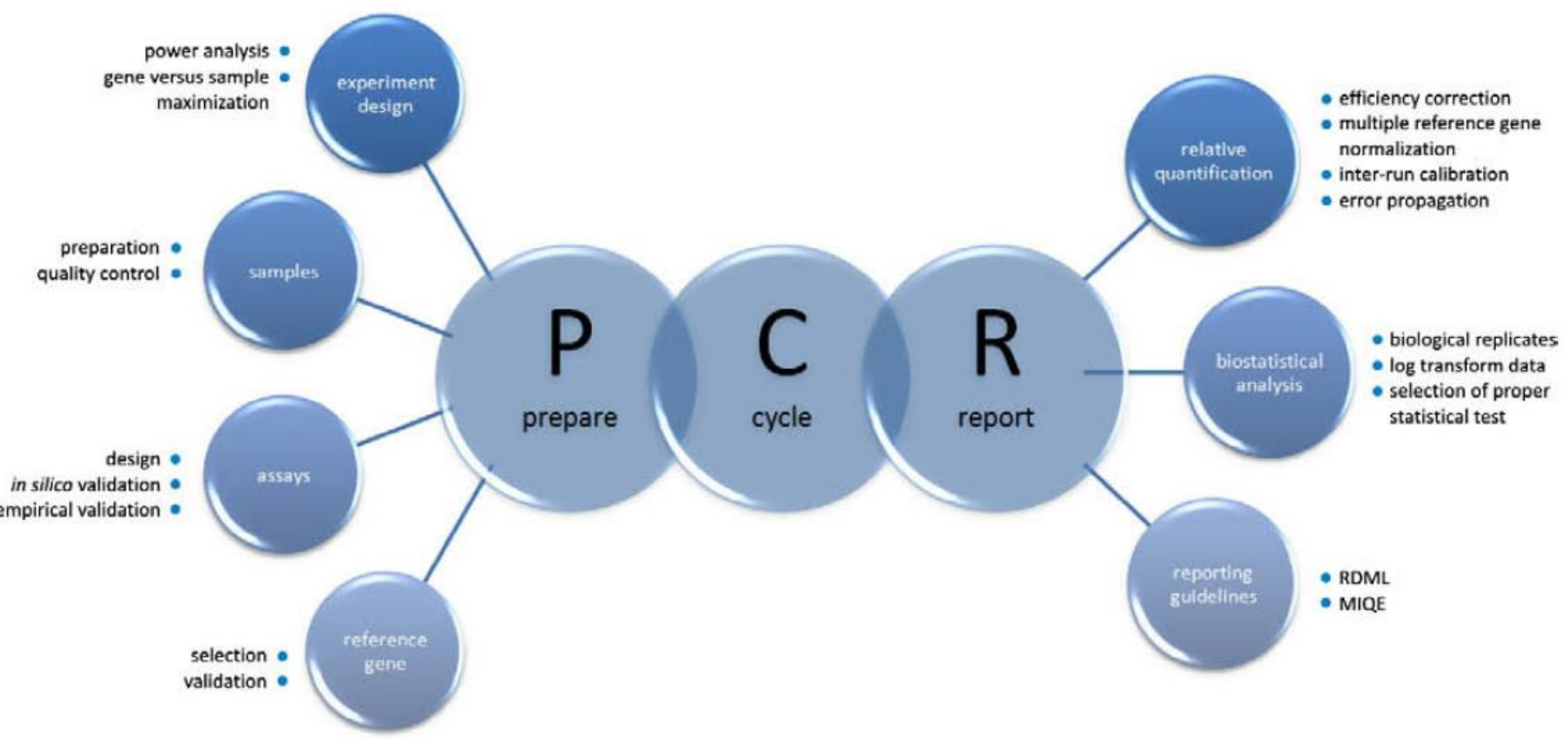
^a All essential information (E) must be submitted with the manuscript. Desirable information (D) should be submitted if available. If primers are from RTPrimerDB, information on qPCR target, oligonucleotides, protocols, and validation is available from that source.

^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 DNA free, inclusion of a no-reverse transcription control is desirable but no longer essential.

Bustin et al 2009

Experimentální design



Typický experiment

1. Plán
2. Izolace RNA
3. RT
4. qPCR

Minimalizace propagace technických chyb (errors of measurement) v experimentu

Gene vs. sample maximization

Technické chyby jsou nezávislé v každém experimentálním kroku a aditivní

Plánování experimentu – příklad: exprese jednoho genu ve dvou myších:

Schéma: např. 2 x 3 x 3 x 3

Subjekty	2	myši	2
Vzorky	3	3 odběry vzorků a izolace RNA	6
RT	3	3 RT ze každého vzorku	18
PCR	3	3 PCR replikáty z každé RT	54

Design and Optimization of Reverse-Transcription Quantitative PCR Experiments

Ales Tichopad,^{1,2*} Rob Kitchen,³ Irmgard Riedmaier,¹ Christiane Becker,¹ Anders Ståhlberg,^{2,4} and Mikael Kubista^{2,5}

Table 1. Variance in biological experiments.^a

	Confounding variance		Studied variance
	Intersubject variance	Processing noise	Treatment effect
Source	Different baseline expression	Sampling	The difference between groups induced by treatment
	Different responses to treatment	RT qPCR	
Intervention	Randomize	Replicates	Maximize effect (e.g., dose selection)
	Large N	Normalization to internal standard or spike	
	Paired measures		

^a The confounding variance consists of the intersubject variance and the processing variance. To maximize the resolution of the effect, the confounding variance must be substantially lower than the effect.

Design and Optimization of Reverse-Transcription Quantitative PCR Experiments

Ales Tichopad,^{1,2*} Rob Kitchen,³ Irmgard Riedmaier,¹ Christiane Becker,¹ Anders Ståhlberg,^{2,4} and
Mikael Kubista^{2,5}

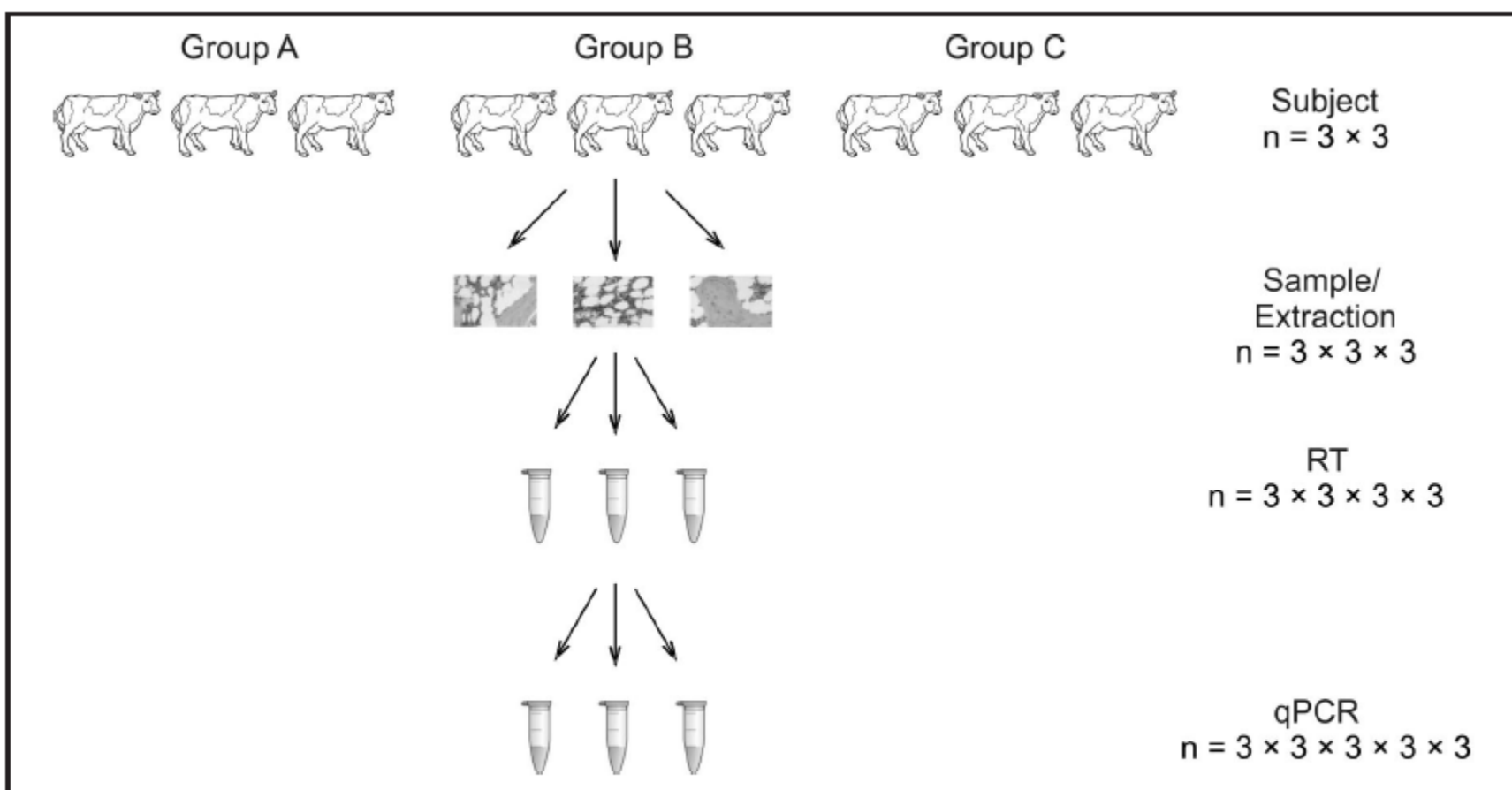


Fig. 1. Comparison of 3 groups with a nested experimental design.

Each group consists of 3 subjects, from which 3 samples are collected and extracted. The extracted and then split into 3 RT reactions, which are then finally into 3 qPCRs. The nested design is $3 \times 3 \times 3 \times 3 \times 3$ and produces a total of 81 Cq values.

Conclusion

Jak navrhnout správný experiment?

- Definovat jej před vlastním začátkem experimentu
- Brát v úvahu hypotézu
- Být maximálně jednoduchý (co nejméně komplexní)
- Maximálně kontrolovatelný
- Technicky a ekonomicky proveditelný, statisticky vyhodnotitelný

Variabilita dat

Nežádoucí

Technická: zpracování vzorku (sampling, izolace, RT-PCR)

Řešení: replikáty, normalizace k internímu standardu

Biologická: rozdíly mezi vzorky (bazální exprese, odpověď na treatment)

Řešení: opakovaná měření, normalizace ke kontrolní skupině

Hledaná

Rozdíly mezi testovanými skupinami

Náhodný sampling, velký soubor

Hodně štěstí