

ÚVOD DO KVANTITATIVNÍ REAL-TIME PCR

9. *Úvahy nad experimentálním designem* *Statistické minimum* & MIQE

Minimum standard for the provision
of information in quantitative PCR

Kvantifikační strategie

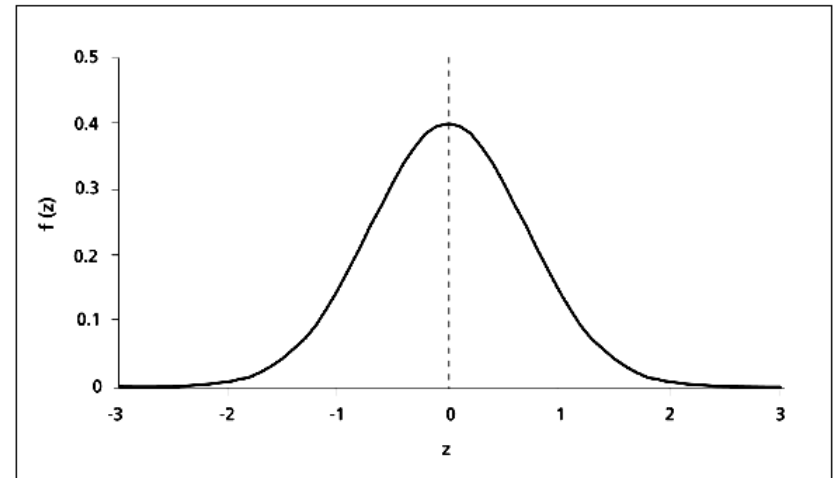
Statistické vyhodnocení

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.

Kvantifikační strategie

Statistické vyhodnocení

Neparametrické testy

- Neznáme parametry rozložení
- Testování rozdílů mezi nezávislými skupinami (Independent samples)
 - dva vzorky**, u kterých porovnáваме 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

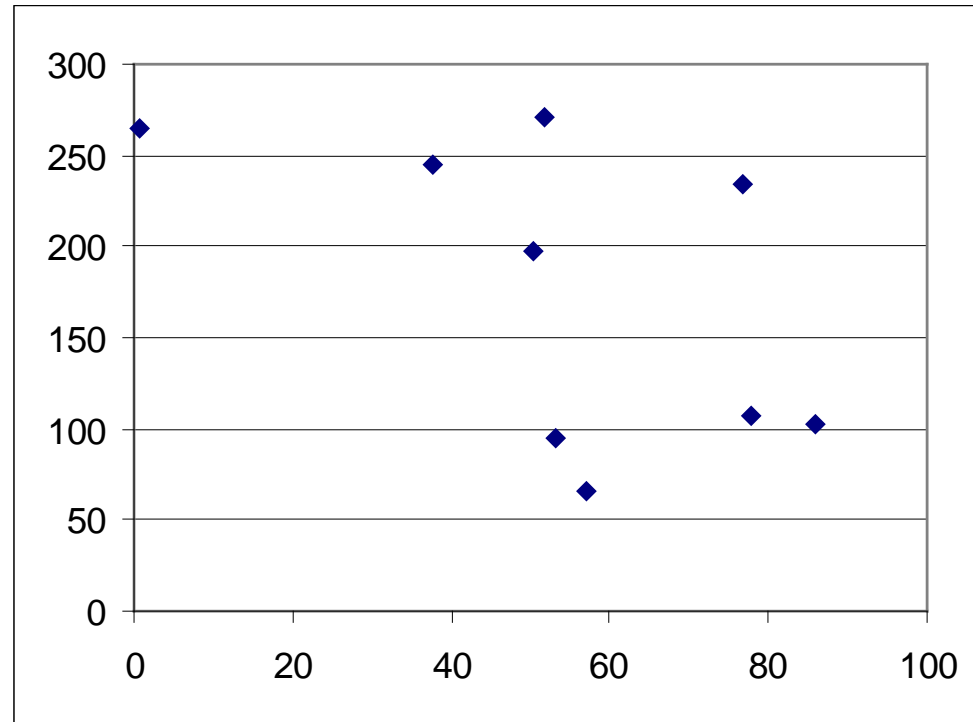
Neparametrické testy jsou méně náchylné k α -chybám (nesprávné zamítnutí nulové hypotézy), ale jsou méně citlivé než parametrické (např. srovnání p u para < p u neparametrických testů), jako signifikantní označí větší rozdíl než parametrické testy.

Kvantifikační strategie

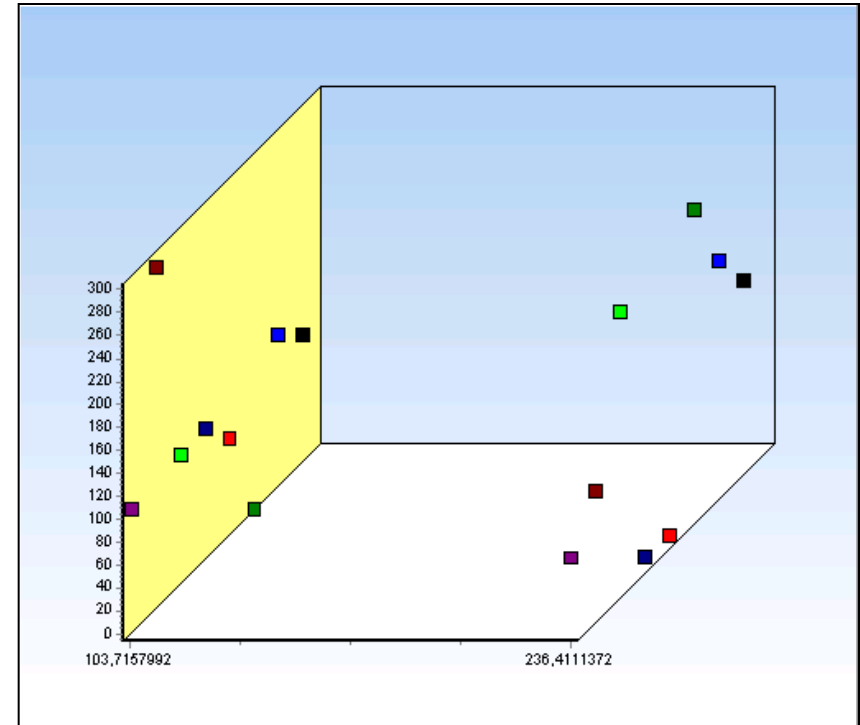
Statistické vyhodnocení

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

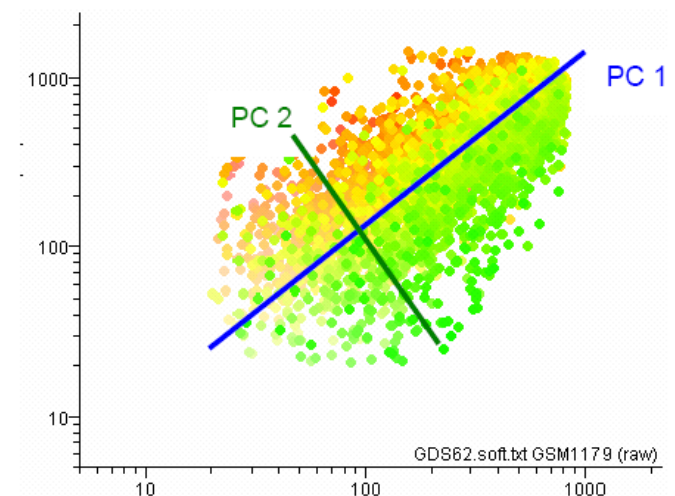
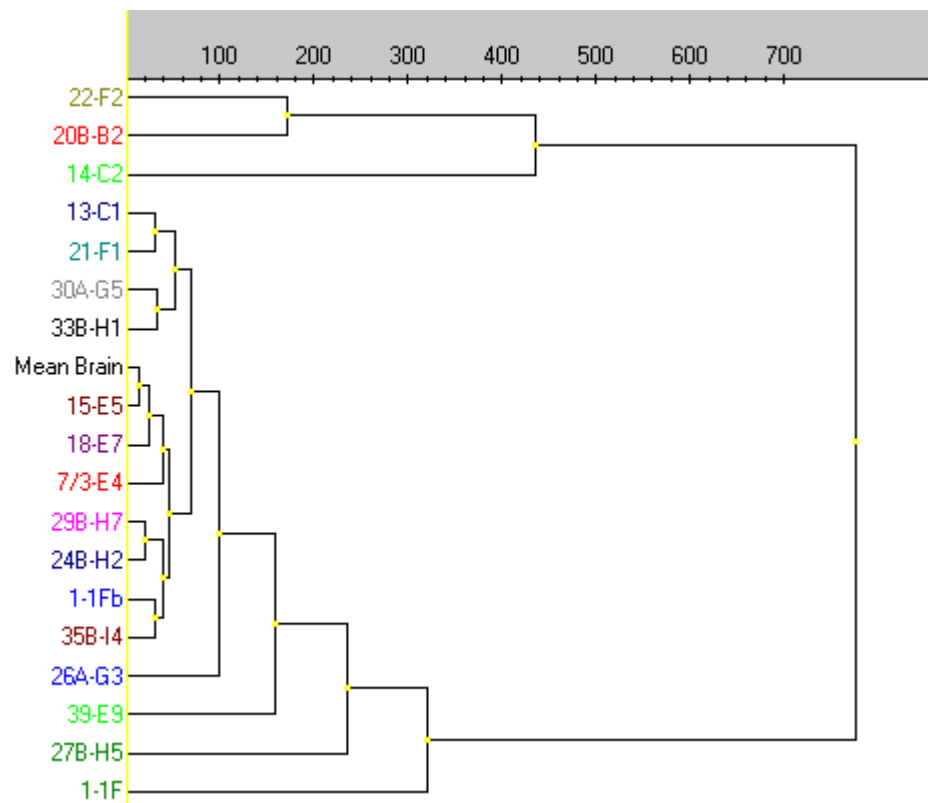
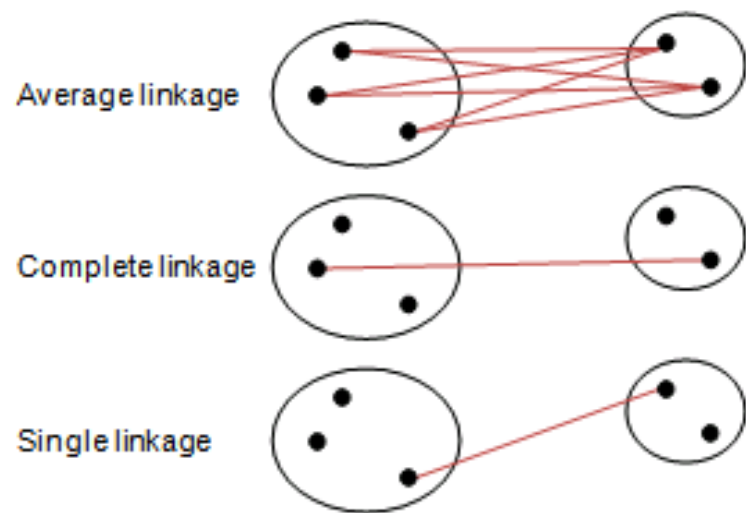


Fig 1: Football-shaped data set with two main components.

Statistické vyhodnocení – shluková (clusterová) analýza

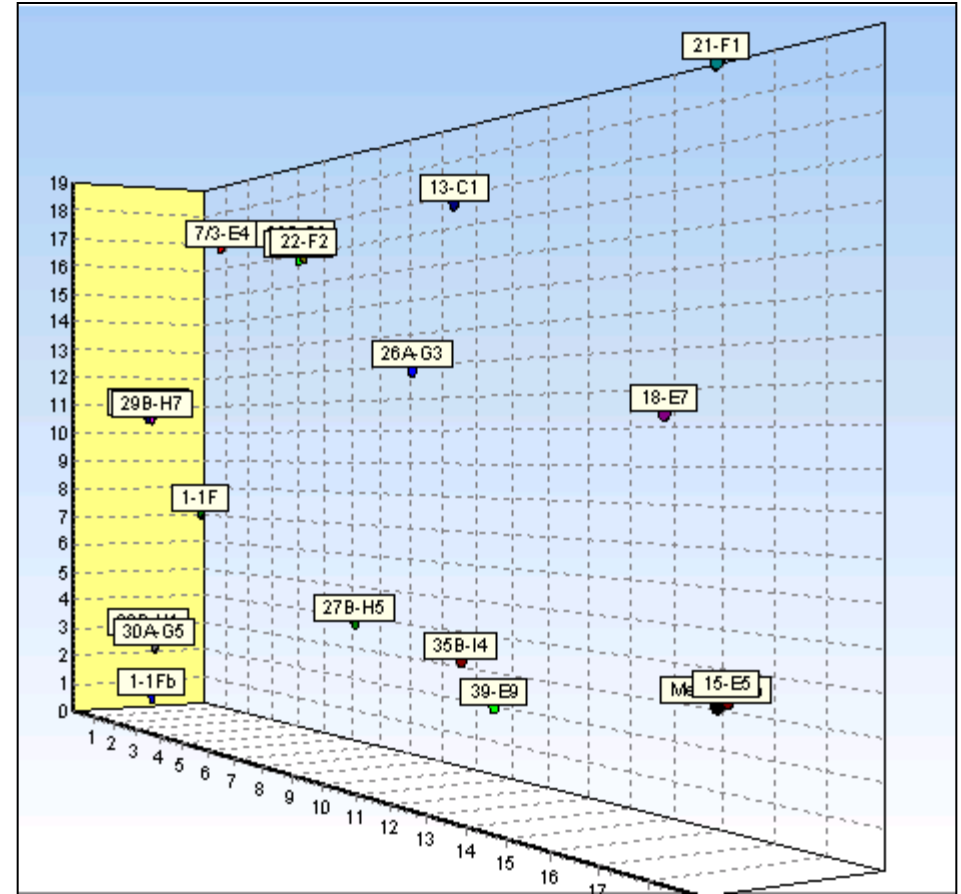
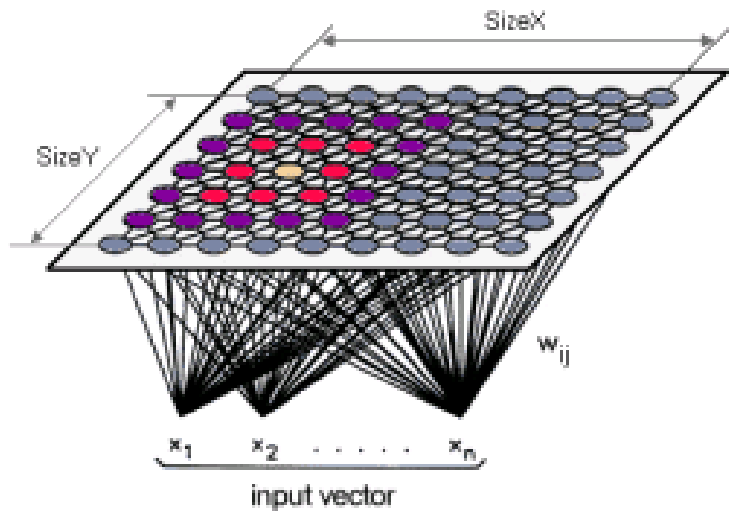
Dendrogramy



Kvantifikační strategie

Statistické vyhodnocení – shluková (clusterová) analýza

Self organizing maps (ANN, Kohonen)



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

Variabilita a její příčiny

Biologická variabilita

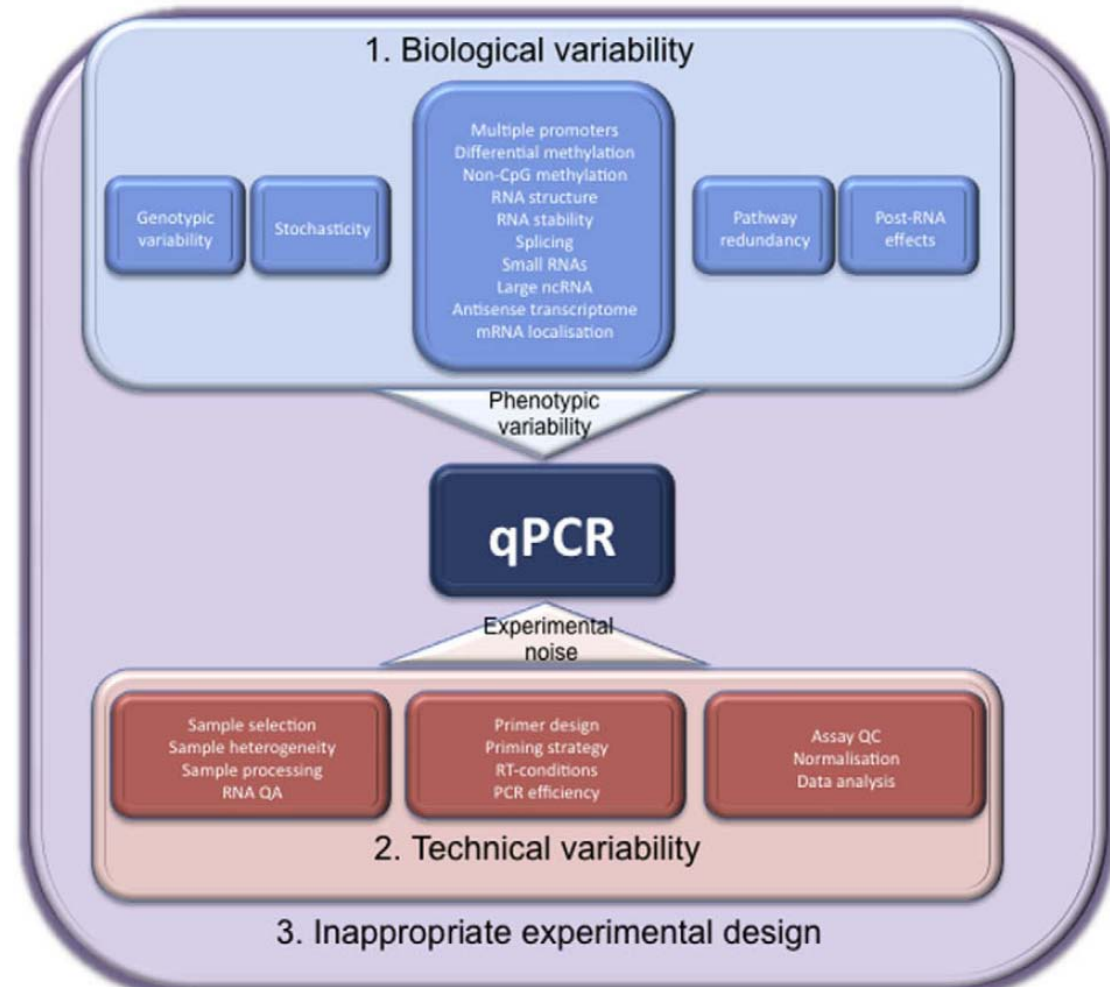
– experimenty odrážejí různorodost reality – nebudou nikdy identické

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



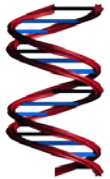


Biologická variabilita



Biologická variabilita

- 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

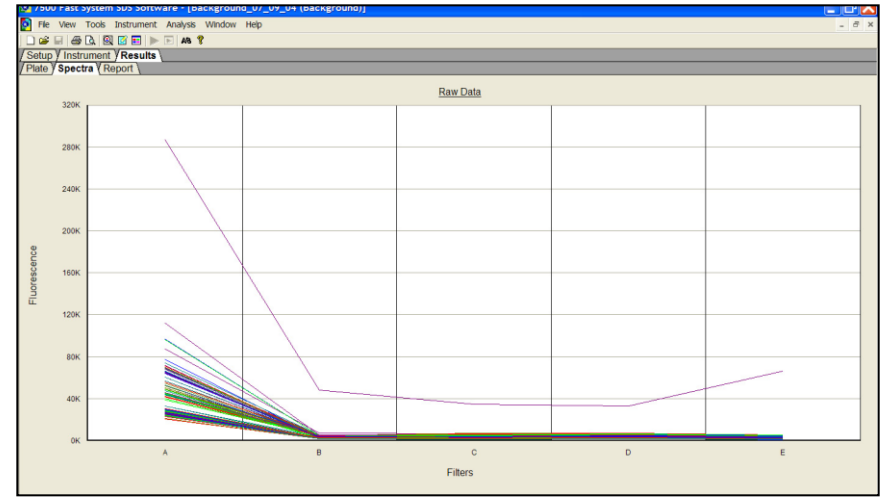
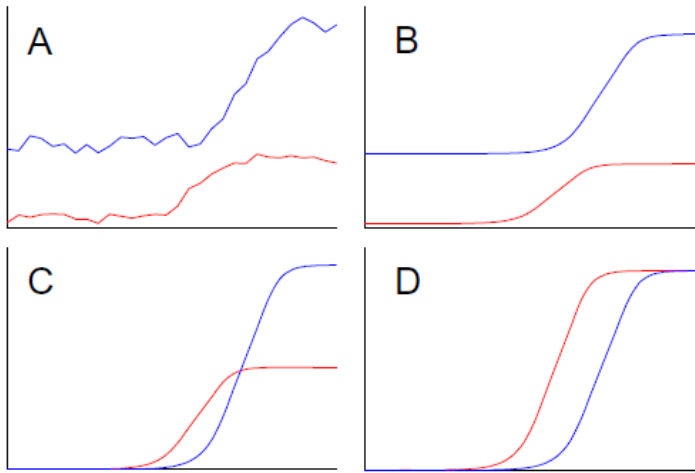
Biologické informační dráhy jsou robustní, redundantní, závislé na buněčném, tkáňovém i environmentálním kontextu



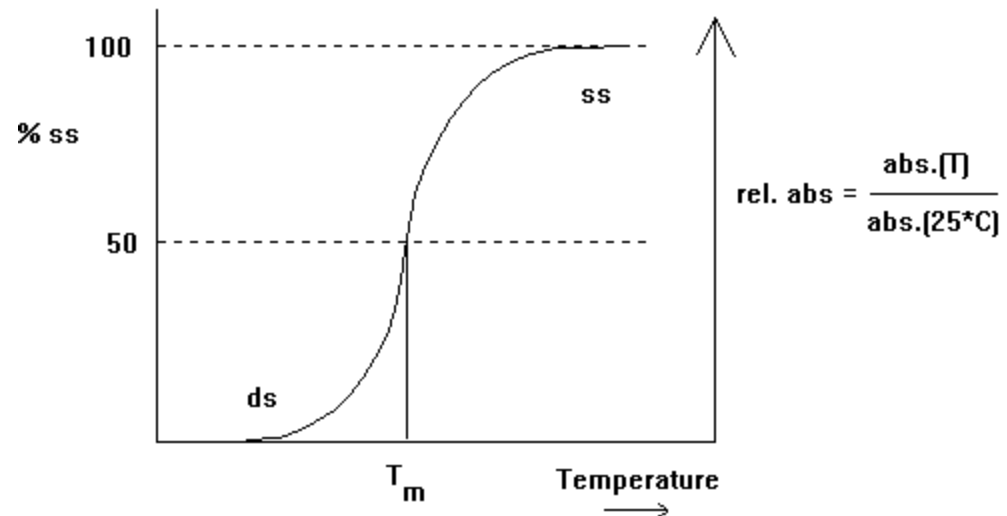
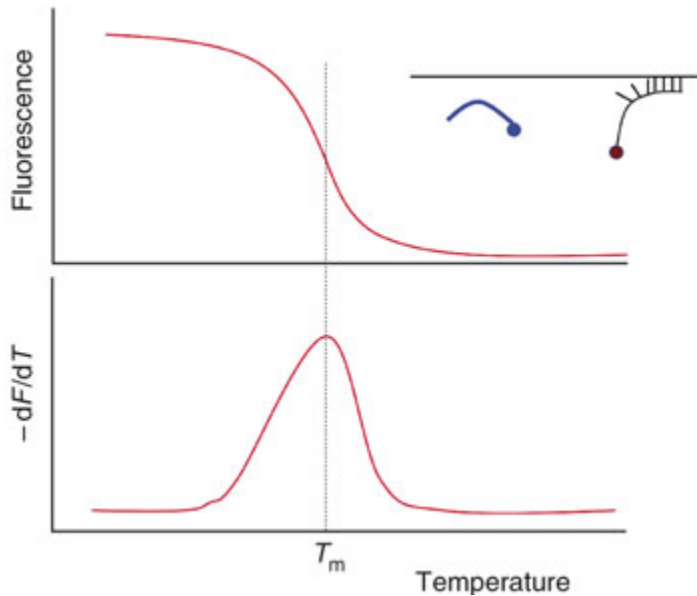
Biologicky relevantní interpretace pozorovaného jevu a jeho odlišení od přirozené variability a heterogenity v daném systému vyžaduje správný experimentální design, analytické metody a vhodný statistický model



Slepě nepřejímat cizí protokoly, přemýšlet 😊



Technická variabilita



Technická variabilita a 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

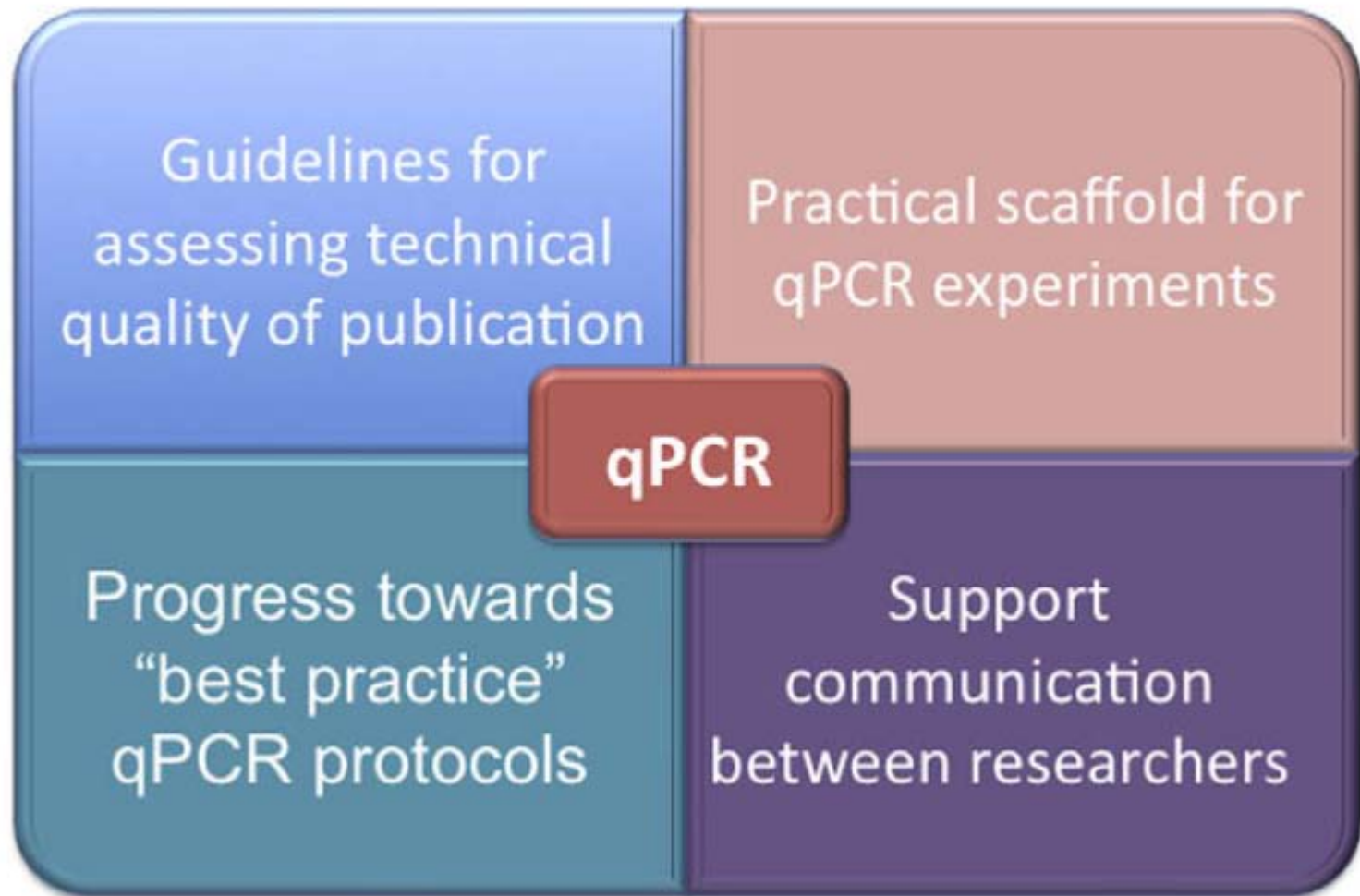
→ **Nutnost standardů**

Technická variabilita

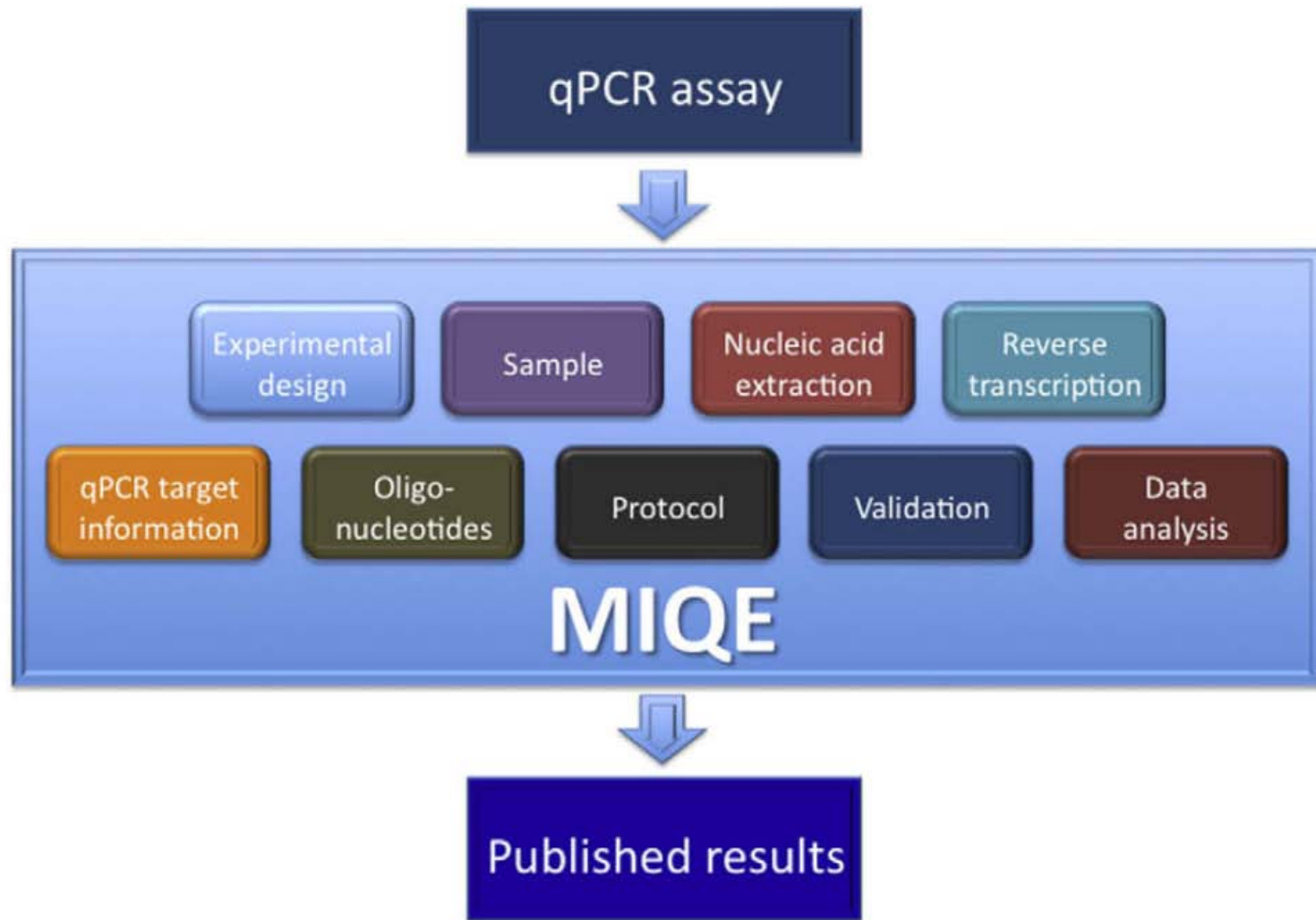


Publication Parameter	A	B	C	D
RNA QA	Agilent	gel	×	Agilent
Primers probes	Assay on demand	✓	✓	✓
RT	✓	✓	✓	✓
cDNA priming	oligo-dT	×	random hexamers	×
Temp/vol/time	✓/✓/×	✓/✓/×	×/✓/×	×/×/×
PCR efficiency	×	×	×	×
Normalisation	single, unvalidated RG	single, unvalidated RG	unvalidated 18S rRNA	single, unvalidated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

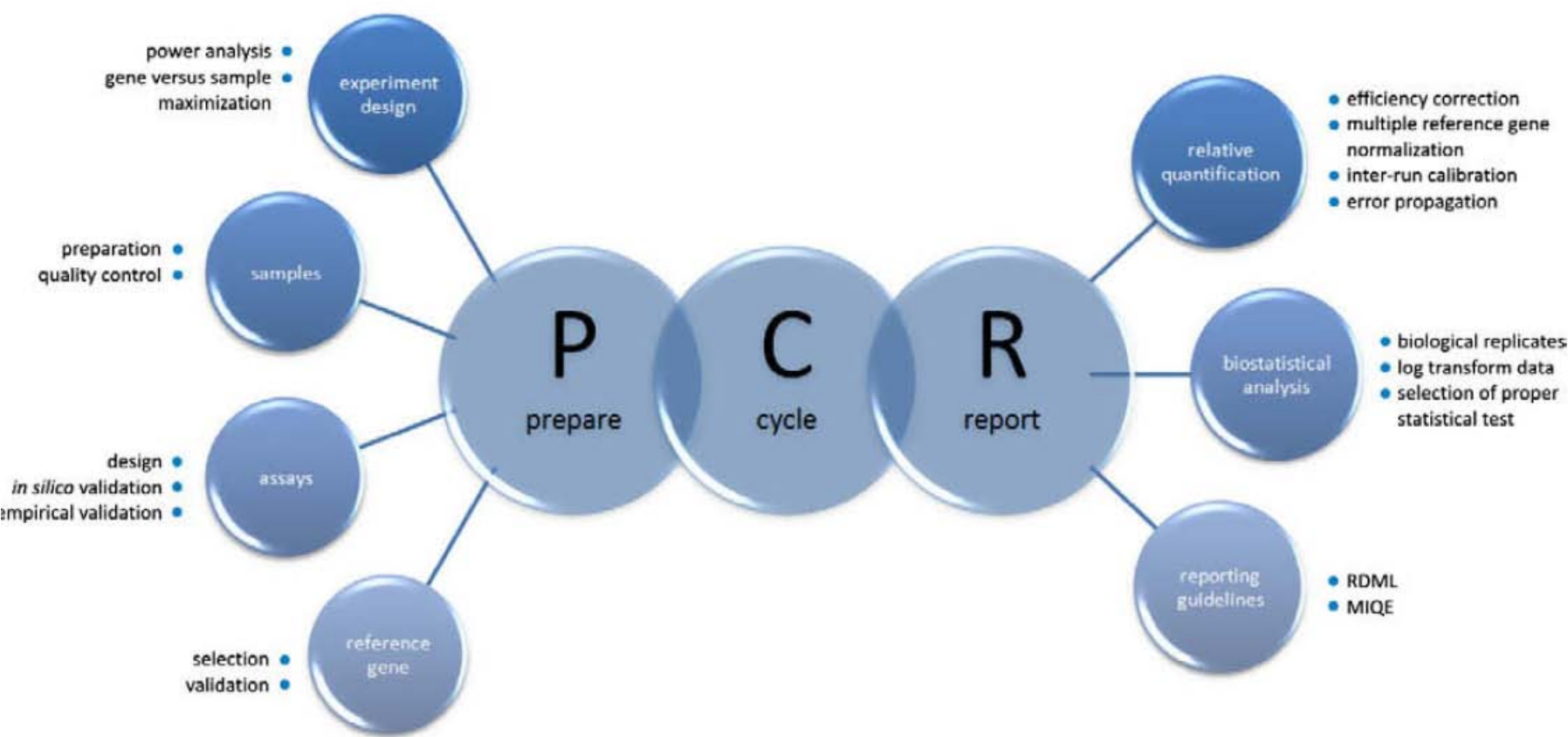
Cíle standardizace



Technická variabilita



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}

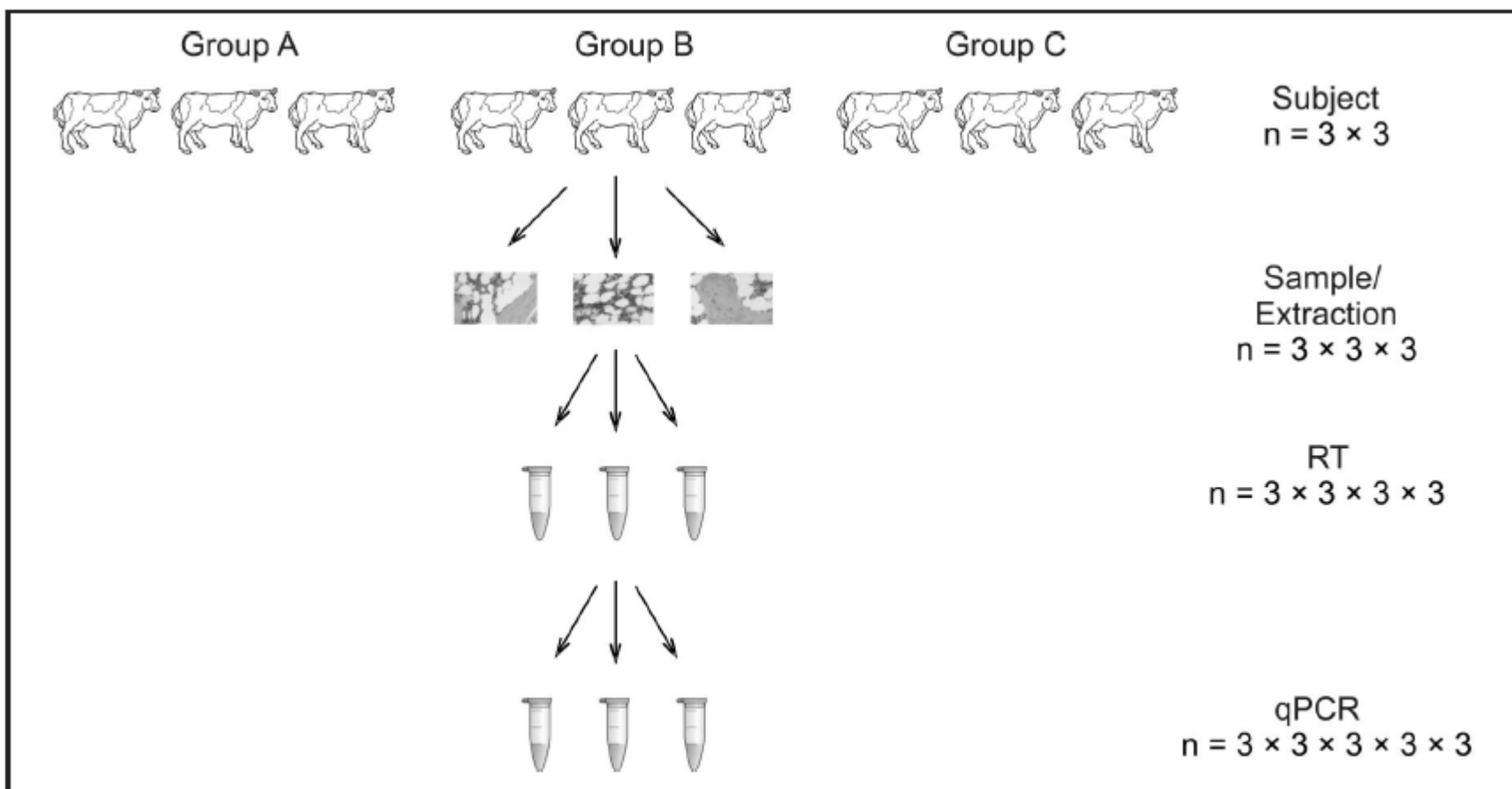


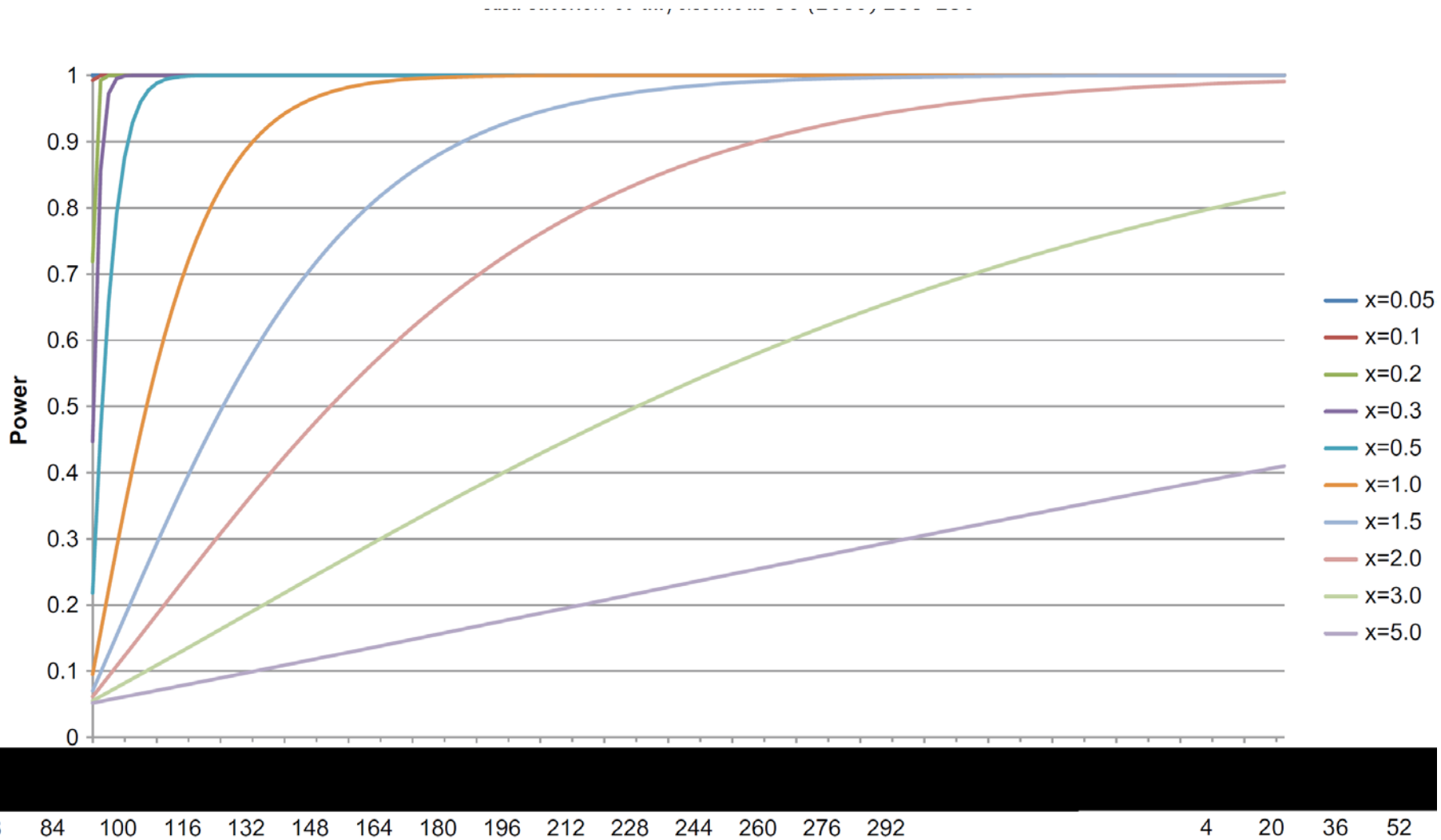
Fig. 1. Comparison of 3 groups with a nested experimental design.

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.

Každý krok vnáší do analýzy variabilitu,
kterou lze charakterizovat

$$\sigma_{Cq}^2 = \sigma_g^2 + \sigma_i^2 + \sigma_j^2 + \sigma_k^2 + \sigma_l^2$$

Určení faktorů , které omezují variabilitu systému umožňuje kalkulovat náklady



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

KONEC



Merry Christmas

