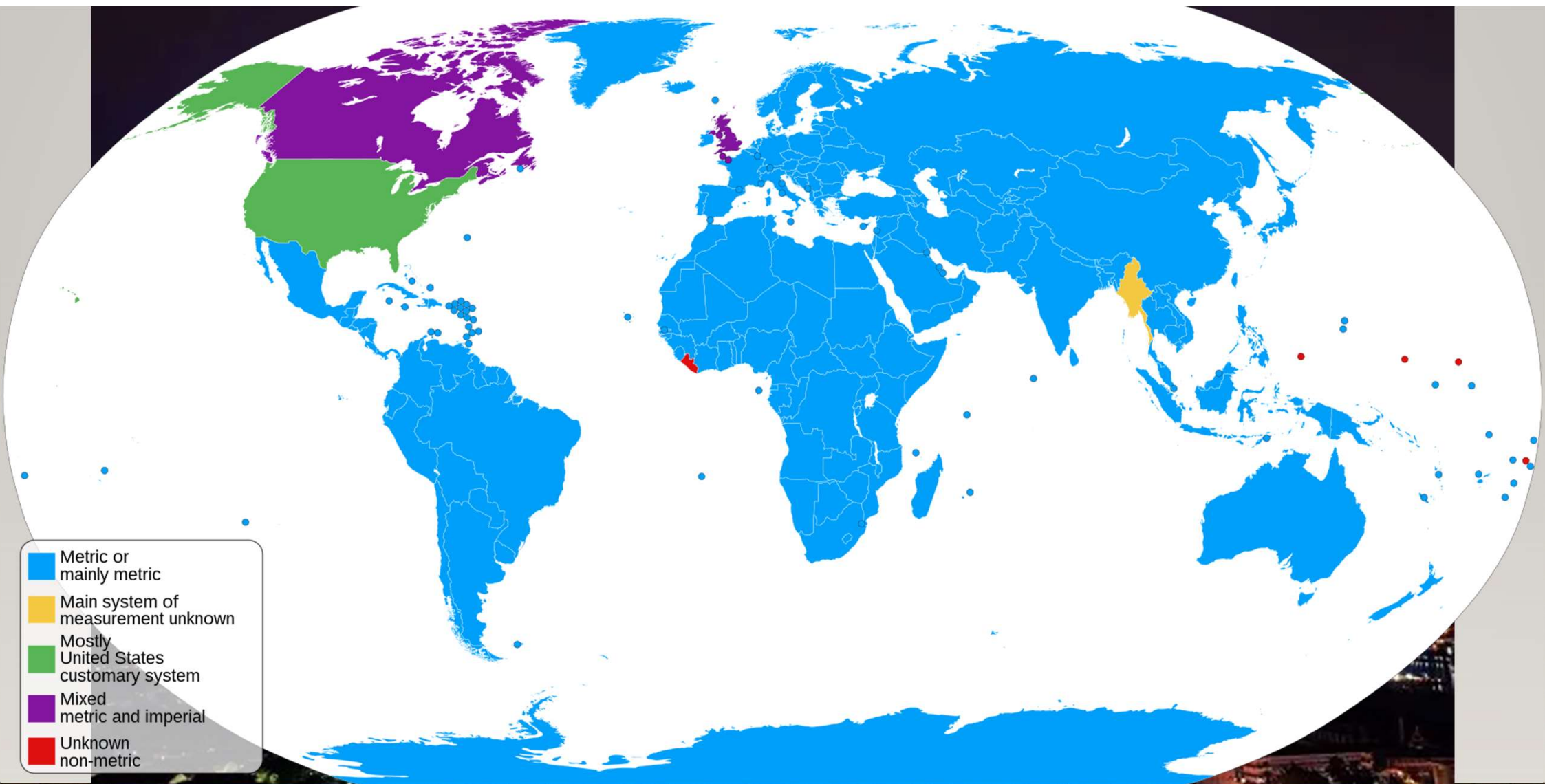


ANIMAL MODELS: RNA EXPRESSION ANALYSES AS AN IMPORTANT SOURCE OF IRREPRODUCIBILITY AND LOW TRANSLATABILITY

ALEŠ BELIČ, ANA UNKOVIČ, EMANUELA BOŠTJANČIČ, MARTINA PERŠE

THINGS TO DO:

- Normalization to reference genes (What are the units now?)
- Along comes probability with statistics
- A real case



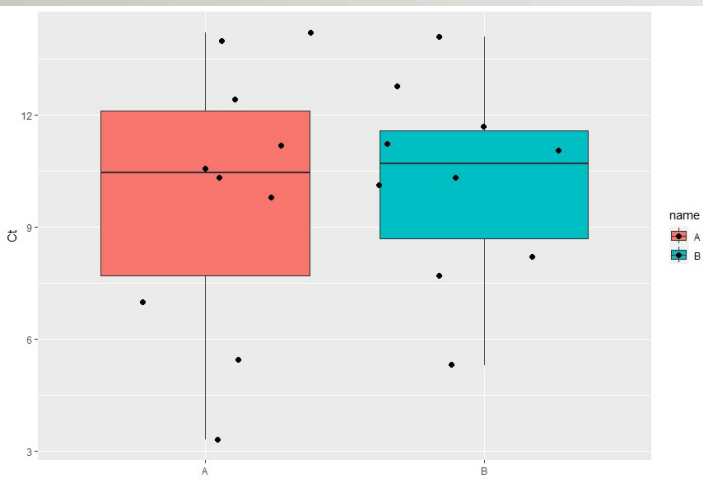
WHICH ARE GOOD REFERENCE GENES?

- One-size-fits-all approach is discouraged
 - Complex process of expression may be different in different tissues
 - Different questions may require different reference genes
- Each reference gene candidate should be tested for stability of expression
 - Minimal variance
- Combination of genes is better than a single gene
 - Mean or geomean of gene expressions (depends on expression measure: Ct, conc., ...)
 - Minimal variance of the combination

SIMULATION CASE

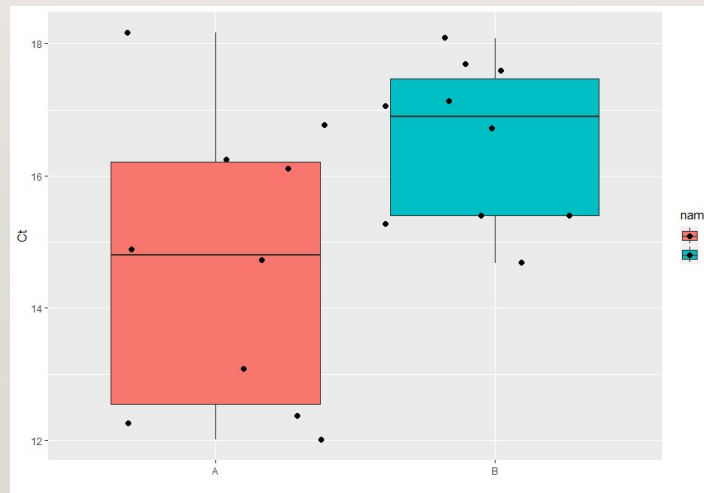
- Target genes expression is affected by two factors A and B
 - each factor contributes some mean value plus a random value ($A = 15 \pm 1 \text{ Ct}$, $B = 16 (18) \pm 1 \text{ Ct}$)
- Metabolism effect is estimated by reference gene and contributes to target and reference gene expression
 - Common contribution to reference and target genes: $5 \pm 2 \text{ Ct}$
 - Additional variability of the reference gene: $0 \pm 1 \text{ Ct}$
- Variability of analytical method: 20%
- 10 (20) samples for each factor taken
- Question: If we repeat the the study 100000 times, how often will the 1Ct (1SD) or 3Ct (3SD) difference between A and B be detected?

HOW DO SELECTION PRECISION AND STABILITY OF REFERENCE GENE EFFECT THE CONCLUSIONS?

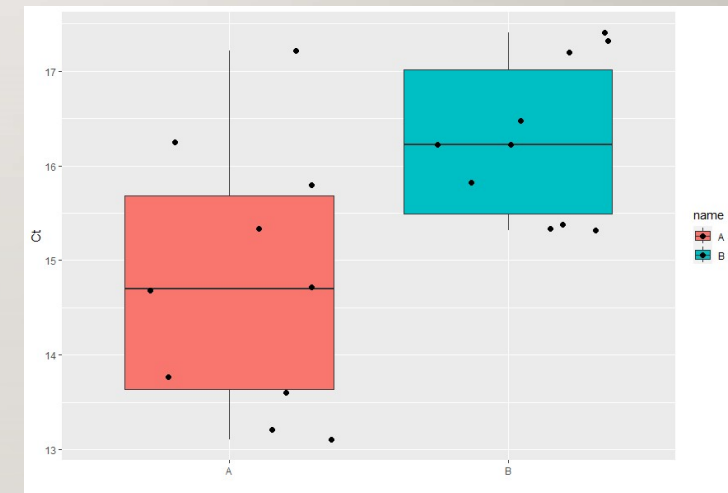


Factor B casuses ISD larger Ct than factor A

Probability to detect difference on non-normalized data: 15.27 %



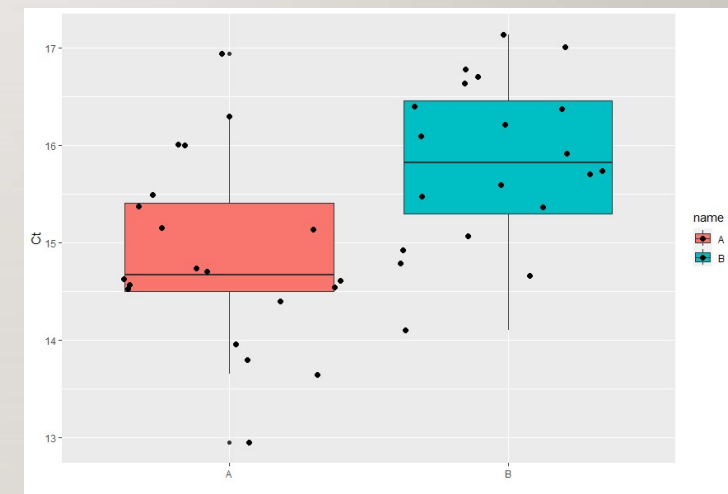
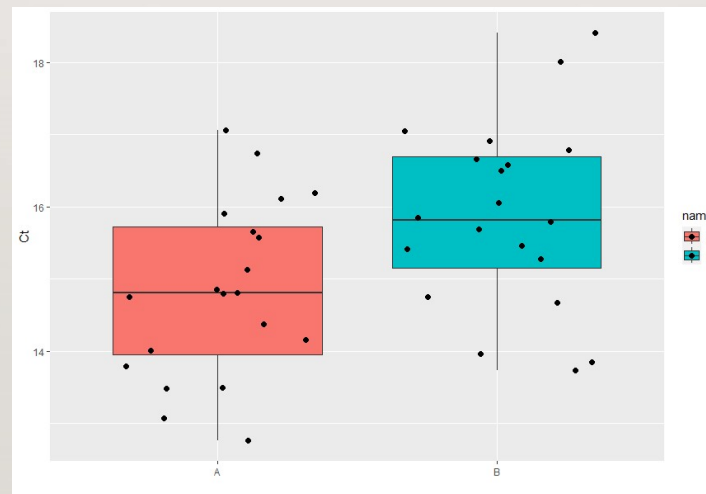
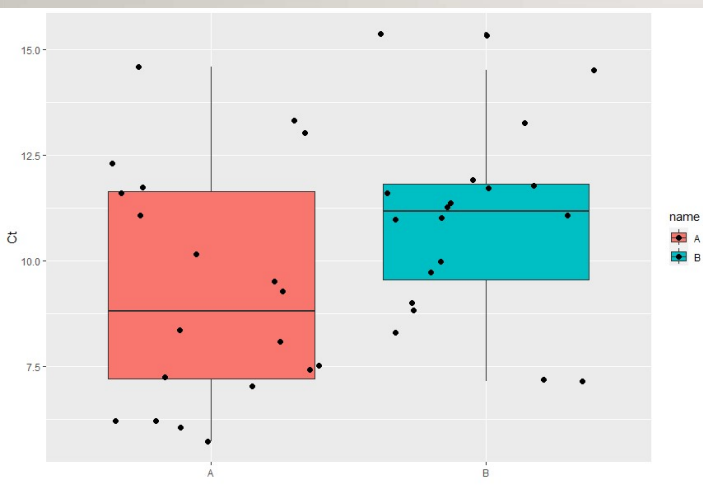
Probability to detect difference on normalized data: 30.73 %



Probability to detect difference on true normalized data: 53.95 %

SD = true variability of target gene response, simulation repeated 100000x, 10 samples drawn from each group

HOW DO SELECTION PRECISION AND STABILITY OF REFERENCE GENE EFFECT THE CONCLUSIONS?



Factor B causes ISD larger Ct than factor A

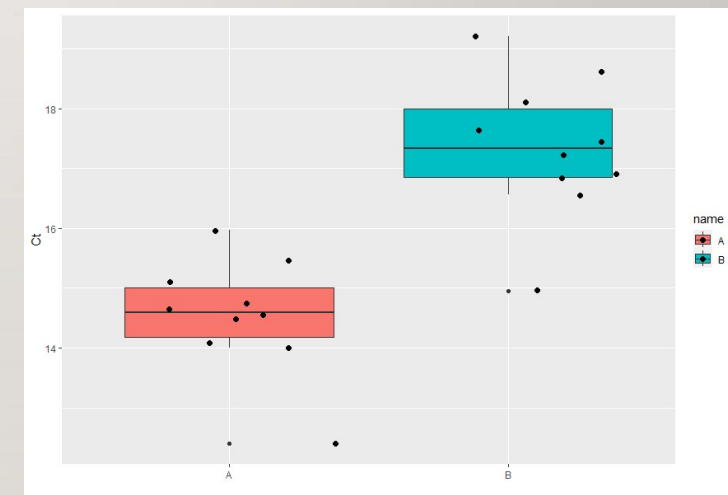
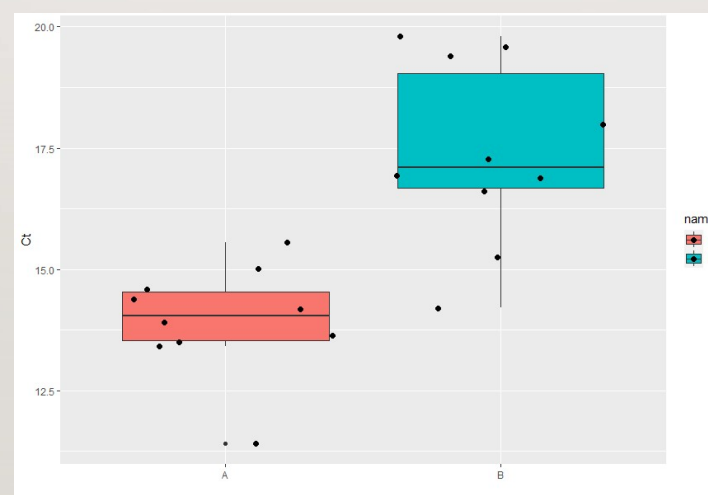
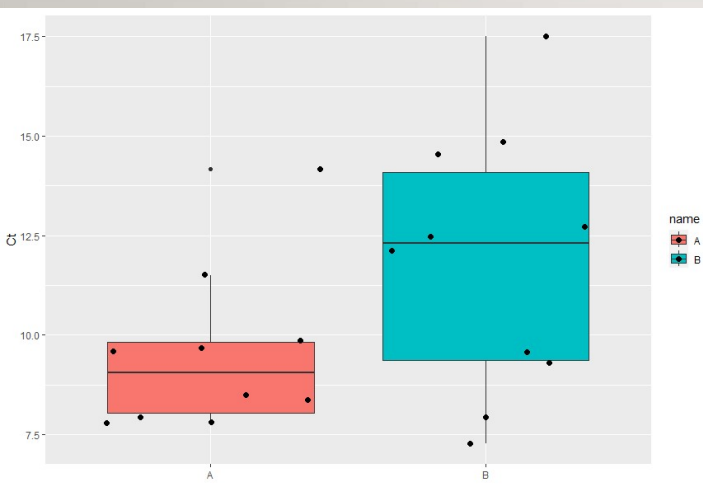
Probability to detect difference on non-normalized data: 28.6 %

Probability to detect difference on normalized data: 57.44 %

Probability to detect difference on true normalized data: 85.59 %

SD = true variability of target gene response, simulation repeated 100000x, 20 samples drawn from each group

HOW DO SELECTION PRECISION AND STABILITY OF REFERENCE GENE EFFECT THE CONCLUSIONS?



Factor B causes 3SD larger Ct than factor A

Probability to detect difference on non-normalized data: 80.36 %

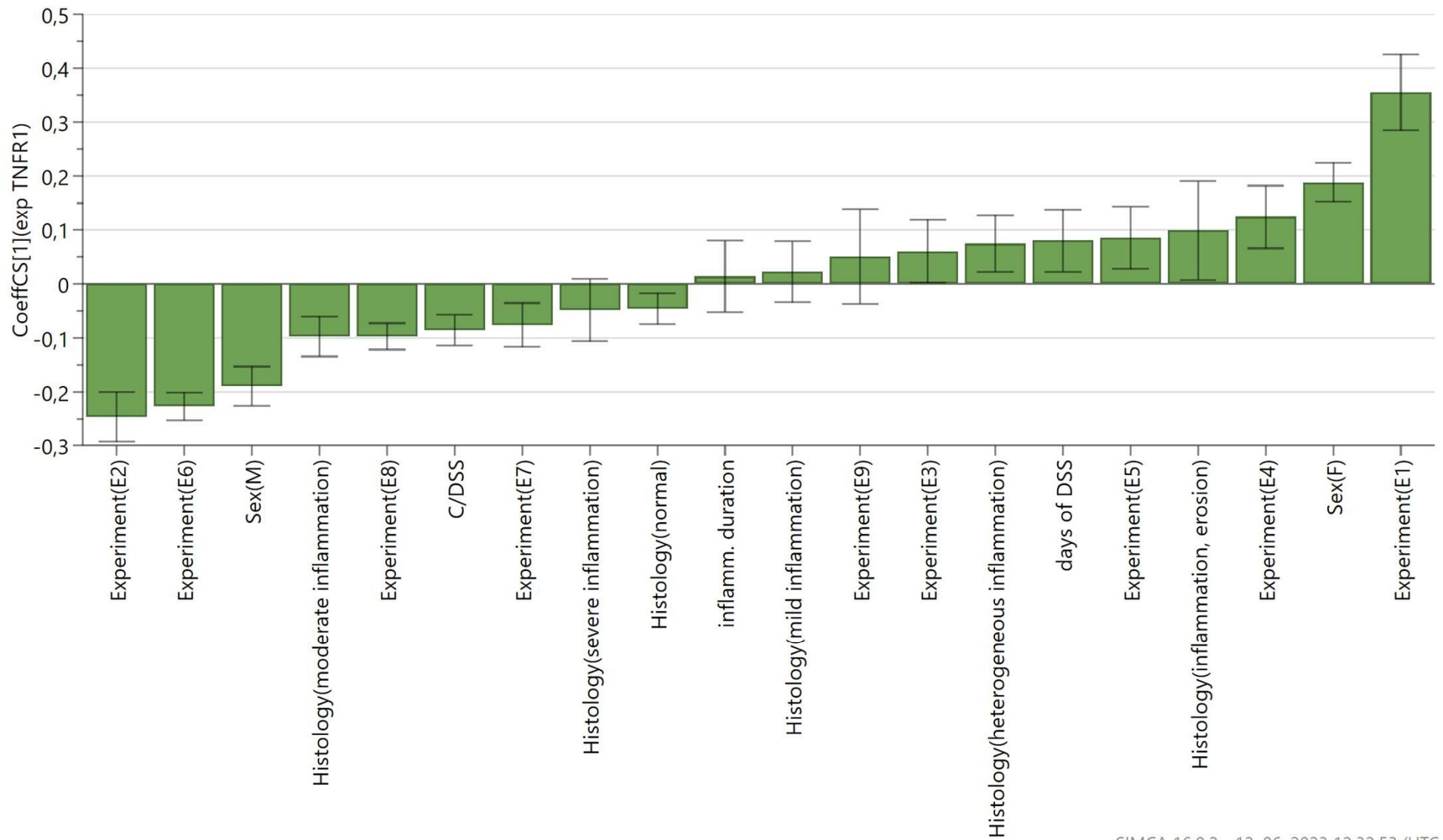
Probability to detect difference on normalized data: 99.24 %

Probability to detect difference on true normalized data: 100 %

SD = true variability of target gene response, simulation repeated 100000x, 10 samples drawn from each group

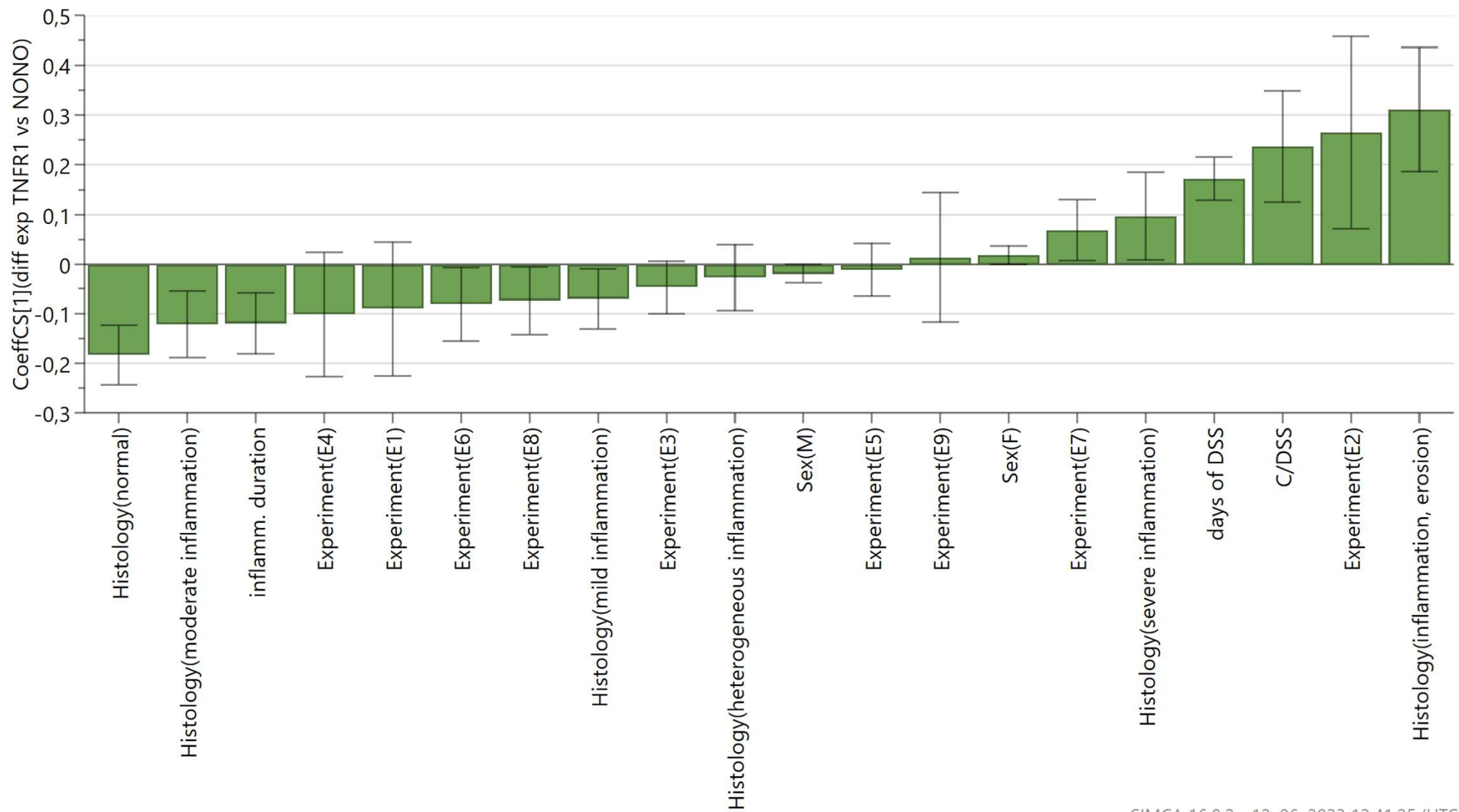
REAL CASE: FFPE COLON SAMPLES WITH VARIOUS LEVELS OF DSS INDUCED INFLAMMATION

- Samples were obtained from several experiments (with as decorrelated experimental factors as possible)
- Experimental factors: histological picture (mild-severe inflammation, mucosal/transmural, erosion), DSS induction protocol, non-specific effects of experiment and sex
- Target genes: TNFR1
- Analysis method: Multivariate analysis using PLS method
 - advantages over multiple regression when experimental factors are not completely uncorrelated
 - testing for significant contributions of experimental factors to target gene expression



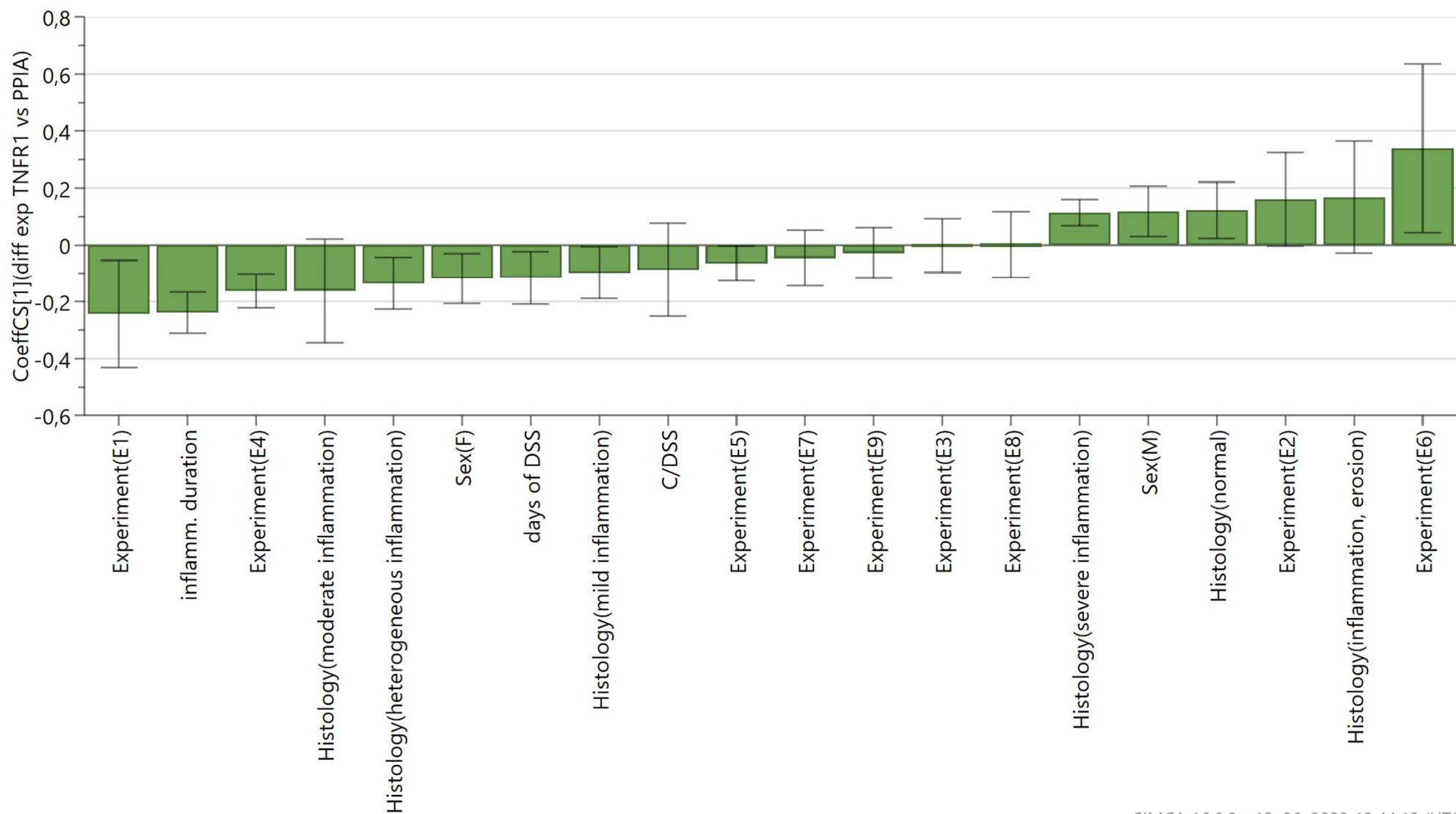
SIMCA 16.0.2 - 12. 06. 2023 12:32:53 (UTC+2)

No normalization to reference gene



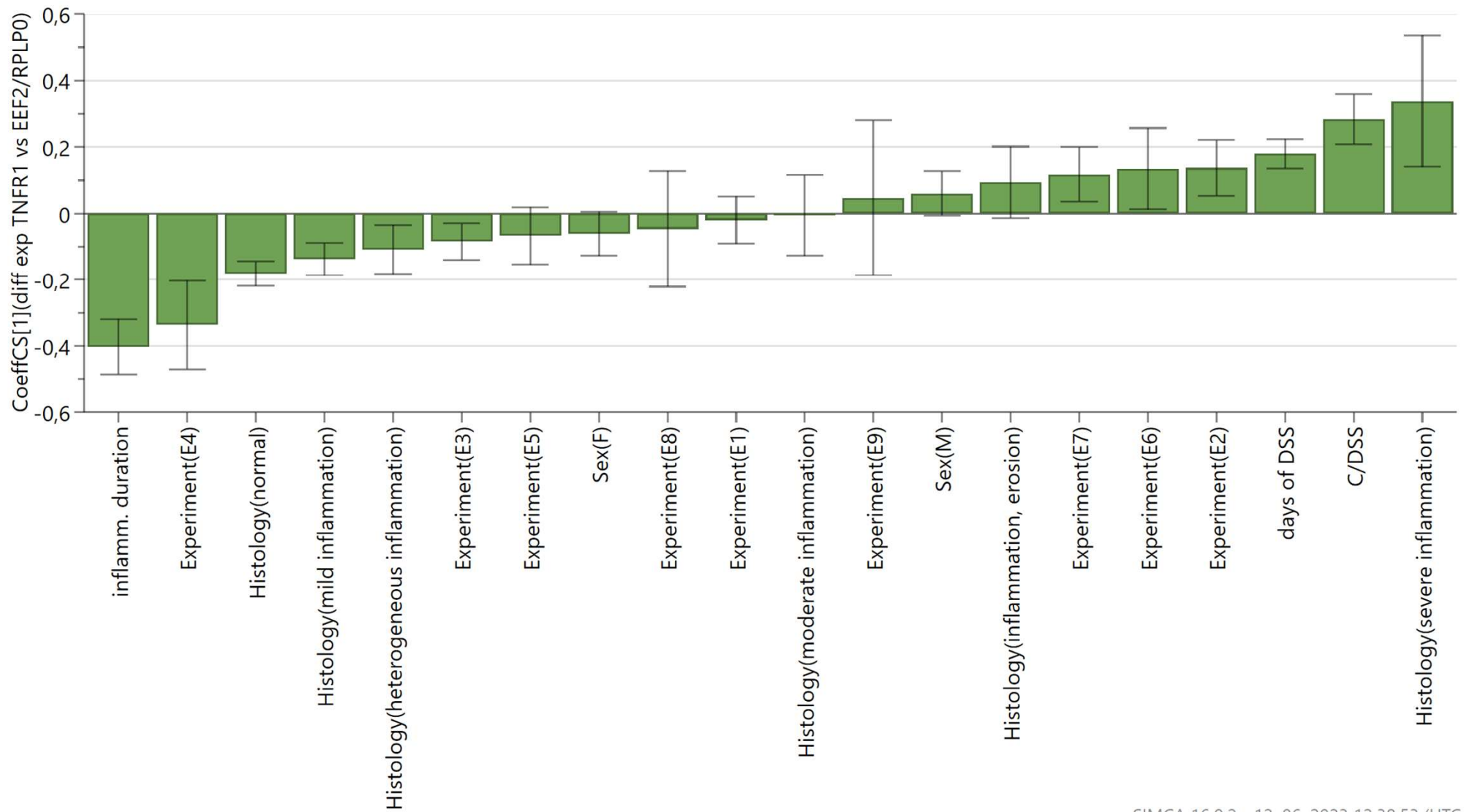
SIMCA 16.0.2 - 12. 06. 2023 12:41:25 (UTC+2)

Normalization to NONO



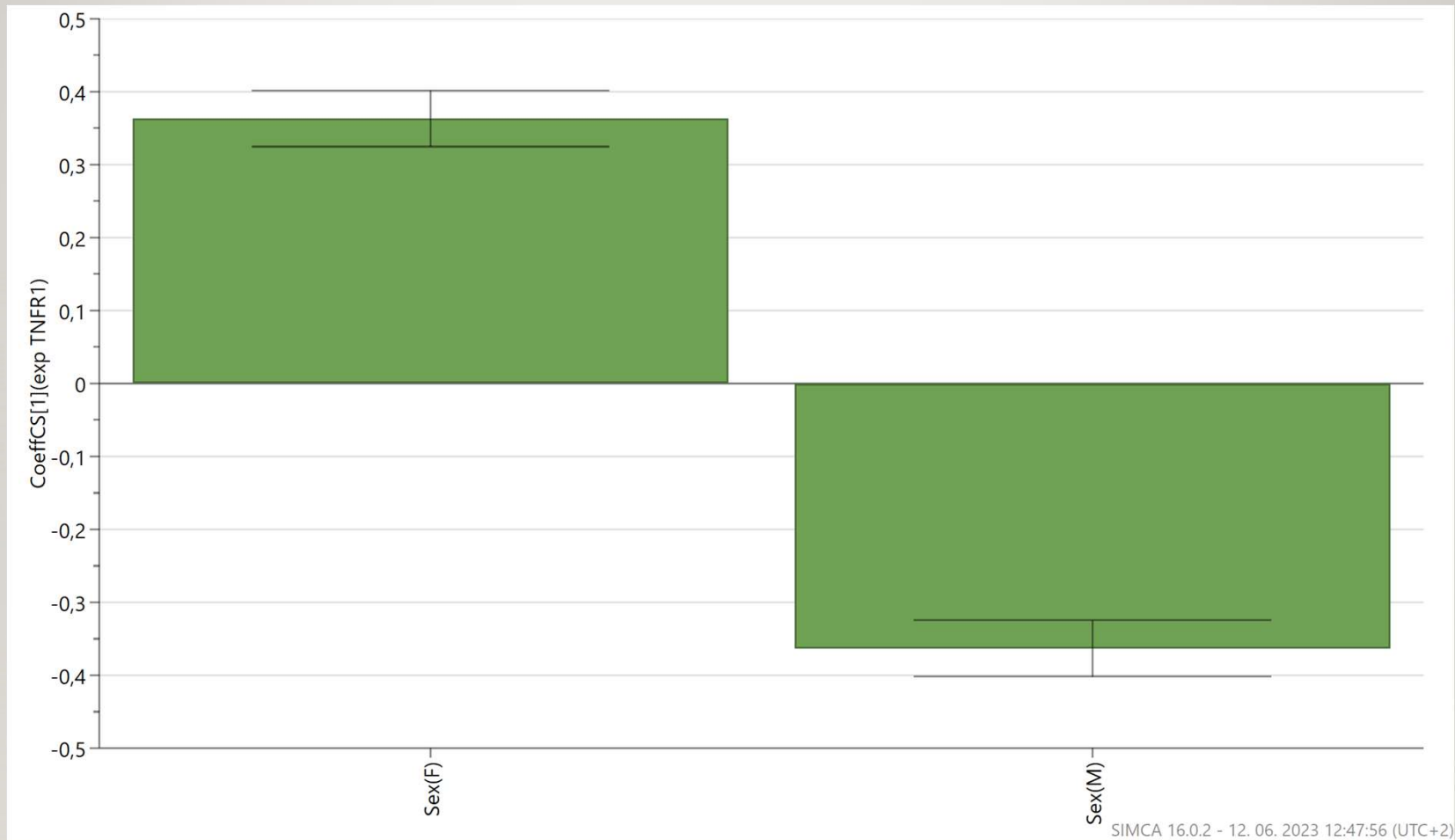
SIMCA 16.0.2 - 12. 06. 2023 12:44:12 (UTC+2)

Normalization to PPIA



SIMCA 16.0.2 - 12. 06. 2023 12:38:53 (UTC+2)

Normalization to EEF2/RPLP0



Normalization to EFF2/RPLP0 ... univariate view

CONCLUSIONS

- Selection of reference genes is not a trivial task
- As analytical methods are becoming more accurate the choice of suitable reference genes becomes even more important
 - large effects can mostly be detected even without normalization
 - smaller and smaller effects can be observed if data is correctly normalized
- Poor choice of reference genes can severely impair quality and reproducibility of study outcomes